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

This research was supported by a research grant from National Institutes of Health (1R21HD078762-01A1) 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.

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
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. We re-evaluate the trend. Because the effective sample for the 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 effect of different sampling rates on the observed measure of inequality, drawing from a population at a single point in time so that there is no change in actual income segregation. We find considerable variation in estimates across samples taken from the same population, particularly for smaller samples. The difference between the median estimate using sampling rates comparable to Census 2000 and the ACS is as large as the observed changes since 2000. We propose alternative approaches to calculate unbiased estimates of class segregation.

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

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

Jun Ke
Department of Biostatistics
Brown University
Providence, RI 02912
jun_ke@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 (Rickets 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).

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' 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 income segregation.

The actual measures used in income segregation studies are varied and complex. For several different measures we assess how the observed estimates could have been affected by sample size by analyzing data from a historical census for which 100% microdata are available – the 1940 Census of Population. This is the first census in which wage data were collected, allowing us to calculate tract-level segregation in income for major cities. We draw many samples of varying sizes from these data. Because of course the "real" income variation across tracts is constant (all drawn from data for the city in the same year), any variations in results can be attributed to differences in sample sizes. We show that there is systematic bias. The smaller the sampling proportion, the greater the estimated income segregation across tracts. We then compare the size of these effects with the actually observed changes between Census 2000 and the 2007-2011 ACS for several metropolitan regions.

Where the observed changes show increases, these increases are in the same range as the changes that could be expected from the reduction in sample size in the ACS. We cannot demonstrate that the shift in sampling between the 2000 Census and the ACS is responsible for the observed results, but this research provides evidence that it may well have been. More work will be needed before social scientists can be sure of what really happened between 2000 and 2009. In the final section of the paper we propose alternative approaches to this task. Two of these are based not on the aggregated tract-level data routinely provided by the Census Bureau but rather on the underlying individual-level sample data. We also provide a method to estimate the bias associated with one type of measure (measures using rank-ordered income data). This correction partly compensates for the distortion caused by the smaller sample size in the ACS, and it can be applied to calculations from the published tract-level income distributions.

Patterns and trends in income segregation

Past studies of trends in income segregation have dealt with two main substantive questions: has income segregation been on the rise, and is it rising the same for whites and minorities? Because the current study is about methodology, we highlight how strongly the technical choices made by researchers affect what they find. These studies all use the same underlying data – the income distribution within census tracts from the one-in-six sample count data of decennial censuses through 2000 and more recently from the pooled 2007-2011 samples from the American Community Survey. Researchers have applied a variety of measures of income-based sorting to these data and relied on varying samples of metropolitan regions. These choices lead to differing conclusions about what happened in each decade since 1970. For example, one research team has reported contradictory findings for a single decade, showing declining isolation of high and low income families in the 1990s (Bischoff and Reardon 2014) but (using a different measure) increasing segregation by income in that decade (Reardon and Bischoff (2011).

Regarding the whole population, what these studies have in common is the finding that the 1980s was the decade in which income segregation increased the most, while there were smaller increases and possibly actual declines, in other decades.

One early study (Abramson et al 1995) reported increasing segregation of poor from non-poor Americans from 1970 through 1990, and he found that poor people were also living on average in high-poverty census tracts by 1990 (referred to as isolation; see also Massey and Eggers 1993, Massey 1996). Two studies (Mayer 2001, Watson 2009) distinguished between the 1970s and 1980s, finding that income segregation declined slightly or stayed the same in the 1970s but then rose substantially in the 1980s. Using different measures, Reardon and Bischoff (2011) reported steadily increasing segregation in both of these decades. Watson (2009) extended her analysis through the 1990s, reporting little change in that decade. However, Jargowsky (1996), who agreed that segregation rose in the 1980s, described an actual "dramatic decline" in concentrated poverty in the 1990s (Jargowsky 2003). And Reardon and Bischoff (2011) reported a small increase in that decade. Most recently two studies (Bischoff and Reardon 2014, Reardon, Fox and Townsend 2015) have brought the analysis up to date with data from the American Community Survey centered on 2009. These two studies used different measures of income segregation and different samples of metropolitan regions. Bischoff and Reardon (2014) find that high-income and low-income families experienced declining isolation through the 1990s, but income segregation then rose rapidly from 2000 to 2009. Yet Reardon, Fox and Townsend (2015, p. 89) report that households at the 10th and 50th percentiles of income on average lived in neighborhoods with about the same median income in 1990, 2000, and 2009; increasing segregation was found only for households at the 90th percentile, who significantly improved the income level of their neighborhoods between 2000 and 2009.

Several studies have distinguished trends by race. Such analyses are substantively interesting because U.S. urban neighborhoods are more highly segregated by race than by social class, and there are questions about how recent middle-class residential mobility has affected class segregation among blacks. Race-specific analyses raise special methodological questions because whites are well represented in all metropolitan regions, but blacks have relatively small populations in many of them, precluding analysis of sorting by income. Typically researchers study only metros meeting a specified criterion of subgroup size.

Reported results again differ in decade-to-decade detail depending on the methods used. Three studies provide results for 1970-2000 (Reardon and Bischoff 2011, Massey and Fischer 2003, and Watson 2009). Combining results from Jargowsky (1996) and Yang and Jargowsky (2006) also covers these three decades. All four studies agree on one point, mirroring results for the whole population: income segregation jumped strongly for both whites and blacks in the

1980s. Otherwise they differ. First, how did segregation compare between whites and blacks at the beginning of the time series? Jargowsky and Massey/Fischer agree that blacks had higher income segregation, but Watson and Reardon/Bischoff report modestly higher segregation for whites. Second, what happened in the 1990s, following the large increases in the 1980s? Watson finds little change for either group. Yang/Jargowsky and Massey/Fischer find declines for both whites and blacks. Reardon/Bischoff report a small decline for blacks but a continuing increase for whites. Third, how did the groups compare by 2000? Yang/Jargowsky, Reardon/Bischoff, and Watson report higher income segregation in 2000 for blacks, while Massey/Fischer finds slightly higher segregation for whites.

Two reports update these mixed findings beyond 2000. Bischoff and Reardon (2014), using the ACS from 2007-2011, find that there was a small increase in income segregation for whites that continued the upward trend since 1980. For blacks there was a very sharp increase, even greater than experienced in the 1980s. Reardon, Fox and Townsend (2015, p. 89), again using different samples and measures, report little change in the neighborhood income level of blacks at the 10^{th} or 50^{th} percentile of income. But blacks at the 90^{th} percentile experienced a substantial upward shift, from living in neighborhoods at the 50.5 percentile in 2000 to ones at the 53.0 percentile in 2009. It appears to be high-income blacks, not whites, for whom the trend of pulling away from less affluent households was accentuated after 2000.

This summary of past findings has three implications for our planned analysis. First, it is essential to maintain a fixed sample of study sites. Second, the analysis should be based on at least two different measures of income segregation, since findings may vary across measures. Third, we should be especially attentive to differences between the whole population and subgroup populations such as black families.

How observed trends could be affected by sampling variation and sample size

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 as a result 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 will 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).

The potential impact on the estimate of variance across tracts of increasing the error in tract estimates is intuitively straightforward. Let us use y_i to represent the estimate of the mean income in a tract, which is a combination of the real mean income x_i and an error ε_i due to sampling variation ($y_i = x_i + \varepsilon_i$). Then the formula for the sample estimate of the variance (disregarding n, the number of census tracts) is

$$s^2 = \sum[(x_i + \varepsilon_i) - \bar{y}]^2$$

We take x_i to be unbiased, so ε_i on average will be 0. However in calculating the variance, two kinds of cases will count more in the calculation of s^2: those cases where x_i is greater than the mean and ε_i is positive (i.e., x_i is an overestimate) and those cases where x_i is lower than the mean and ε_i is negative (i.e., x_i is an underestimate). These over- and under-estimates are squared, so both kinds of cases will tend to increase disproportionately the estimate of s^2. When applied to variance in income, 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.

We suspect that other measures of income segregation will be similarly affected. If so, we infer two hypotheses that we wish to evaluate here:

H1. Observed increases in income segregation may result from comparing measures in 2000 with measures in 2007-2011 based on smaller sample sizes, which exaggerate variation across tracts (upward bias) and result in less reliable estimates.

H2. Measures for subgroup populations such as blacks or immigrants are particularly prone to both upward bias and variable estimates across samples, because the group-specific income data at the tract level are based on much smaller samples than income data for the full population.

Research design

Our procedure is to draw samples from the 1940 Census with varying sample proportions and examine how the sampling rate is associated with the measure of class segregation. Results will be compared with national and city-specific trends in measures of income segregation of families between 2000 and 2007-2011 using the rank-order H measures (as reported by Reardon and Bischoff 2014) and the Neighborhood Sorting Index (NSI). We have selected six large metropolitan areas in the contemporary period with substantial minority populations and large numbers of census tracts from which to compute segregation indices for the total population and for the black population separately. These metropolitan regions are Chicago, Cleveland, Detroit, Los Angeles, Philadelphia, and Pittsburgh. For 1940 we draw data from their central cities, which were fully tracted at that time. For various sampling rates in 1940 we calculate indices for the whole population as well as for foreign-born whites whom we use to represent a subgroup comparable in size to the contemporary black population.

Sampling proportions in the decennial census and 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%.

Measures of income segregation

As already noted, researchers have employed several different measures of income segregation. Even in the simpler case of black-white segregation, where there is a simple dichotomy, there are multiple ways to conceptualize and measure segregation. With income, which the census reports in multiple categories of income, there are more alternatives. 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).[1] 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. Variants of this index are H10 and H90, the extent to which the lowest-earning families (bottom 10%) or highest earning families (top 10%) live separately from the remaining families. Hence these two measures, like the Index of Dissimilarity, divide the population into just two categories. In principle the rank-order H is similar to the Centile Gap measure used by Watson (2009), which converts income levels into percentiles within a metropolitan area. In this measure, the percentile ranks of families in a census tract are compared to the percentile rank of the median income family in the tract. If there is little spread within a tract, that is an indication of high segregation across tracts. Both of these measures have the property of being unaffected by any rank-preserving change in the income distribution. In other words, they are not sensitive to the extent of income inequality in the region.

The second measure is the correlation ratio, which Jargowsky (1996) refers to as the Neighborhood Sorting Index (NSI). The NSI is the square root of the between-tract variance in income divided by the total variance of income. 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" (1996, p. 998). It differs from the rank-order version of H because it gives greater weight to tracts that differ greatly from the mean. A difficulty in computing this index from income data grouped into categories, such as tract-level or metropolitan-level census data, is that there is no information about the distribution of incomes within the top income category. Extremely high incomes in the top category have disproportionate influence on the total variance in income. We used census microdata from Census 2000, ACS 2010, and the Current Population Survey in 2000 and 2010 to provide estimates of the maximum and mean values in the top income category for the metropolitan areas in the study. Then, following other researchers, we assumed a Pareto distribution of incomes in the lower and upper income categories in order to calculate the total variance in income.[2] We calculate measures for all families and for black-headed families.

Data 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. Household income (or alternatively, family income) is substantively preferable to individual income for research on income inequality. Note, though, that it reduces the number of cases on which statistics are computed. All of the income segregation measures can be calculated directly from these microdata. In addition we compute segregation indices for a subgroup whose population share in 1940 was comparable to the share of black residents in the contemporary data: foreign-born whites. Households were categorized by 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 each of the sample cities (we did not attempt to measure family income because family relationships are not clearly defined in the 1940 data). We draw samples from the full population at varying sampling proportions from 1% to 20%. We pay special attention to results in the vicinity of 15% and 5%, which we understand to be the approximate sampling rates for the Census 2000 long form and ACS 2007-2011, respectively. We repeat the procedure 100 times for every level of the sampling proportion and calculate every segregation measure for every draw. Non-residential tracts (10 households or less) are omitted; a minimum of one household was sampled in every tract. This yields a sampling distribution of estimated income segregation. 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.

Results

Observed change in income segregation

We begin with the observed changes in segregation between 2000 and 2007-2011, including the six metropolitan areas that we will use as examples and the average and standard deviation of values for the largest metropolitan regions in the country. Table 1 presents results for the total population and, to represent a population subgroup, the non-Hispanic black population.[3] Blacks comprised between 8.9% and 22.7% of the metropolitan populations in 5 of the regions, but 41.4% in Detroit.

Table 1 about here

The observed patterns of change vary depending on the measure and population that are used. 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 .5 standard deviations). These results support reports of increasing income segregation. However NSI declined in 5 of the 6 example metros, and in the 6th case (Detroit) it increased by only .005. The average NSI in the largest metros declined slightly from .218 to .202. By this measure income segregation was moving in the opposite direction from what is indicated by the H measures.

Results for the black population show a different pattern, a more uniform increase regardless of the measure. The rank-order H measures increased more for blacks than for the total population. 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). Similarly NSI rose in four of six metros, and the average value in metros with large black populations increased slightly from .252 to .260.

These results raise two kinds of questions about what actually happened to income segregation. The first, accepting the sample results as valid, is why the rank-order H measures and the NSI show different trends based on exactly the same data. We can answer this question only in general terms based on their formulae; because they measure different aspects of income distribution, they need not necessarily move in the same direction. The NSI is calculated from the variance in mean incomes across census tracts. It is especially sensitive to changes at the top and bottom of the distribution, and it will tend to decline if the mean incomes in very high and very low income tracts move toward the overall mean. If there is some increasing spread between tracts with values closer to the overall mean, changes in that part of the distribution will not be weighted as heavily. The rank-order H, on the other hand, treats changes in the center of the distribution (where very small absolute changes can appear as larger changes in rank order) the same as changes at the extremes (where values are more widely spread and it takes a larger absolute change in income to alter the rank order). Possibly, then, in the post-2000 decade there were different kinds of changes in different parts of the income distribution across tracts – some convergence toward the overall mean at the high and low end, but at the same time some increasing spread in the center.

We will not pursue this question further; our only point is that researchers need to be aware of differences in what aspect of income is being measured by each alternative index. Our main purpose is to examine a different question, which is how results with either measure may have been affected by the sharp reduction in sample size between the 2000 Census and the 2007-2011 ACS. Based on the reasoning presented above, we wish to consider the possibility that in the full population – not in the samples enumerated by the Census Bureau – there actually was no change or even a decline in both types of income segregation measures. Further, because the calculations for the black population were based on samples that were typically less than half as large as those for the total population, and often only 10-20% as large, the apparently more consistent pattern of increases in black income segregation may also be illusory.

Simulations with full-count historical census data

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.[4] We carry out this exercise for the total population in six cities. In addition we repeat the analysis for white foreign-born households in order to illustrate the greater difficulty of estimating income segregation for subpopulations. Foreign-born whites ranged from 12.5% to 19.8% of all households in our six cities, comparable to the black share in most of the example metro areas that we studied with contemporary data.

In each of six cities, we plotted the estimates based on all 100 subsamples drawn at each sampling rate between 1% and 20%. For parsimony we present here only the plots for Chicago. We found greater sampling variation for H10 and H90 than for H. This could be expected from the fact that H uses data from every decile in the income distribution, whereas H10 and H90 compare the tails (the bottom and top deciles) to the rest of the population. Because samples from a single decile are smaller, measures based on them can be expected to be less reliable. We rely on results for H90 in comparison to H to reveal this pattern.

We begin with the plots for Chicago that show estimates for varying sampling proportions for H, H90, and NSI for the whole population and the smaller foreign-born white population (Figure 1). These plots show the distribution of estimates from different sample draws: the median, the values at the 25th (Q_1) and 75th (Q_3) percentiles, and the minimum and maximum estimate. The vertical axis represents the value of segregation (from 0 to 1.0). The horizontal axis represents the sampling rate as a percentage of the population, ranging from 1% to 20%. At each sampling rate the graph displays the distribution of values from the 100 samples that were drawn: the maximum and minimum values and (within the limits of the resolution of the figure) the median value, the value at the 25th percentile (Q_1) and the value at the 75th percentile (Q_3). This figure is useful as a visualization of the differences across measures, and these plots are most helpful in illustrating the bias in median estimates. The detailed statistics on distributions are presented below in tabular form for all six cities. The plot for H, total population, shows that in Chicago there is considerable upward bias for sampling rates as low as 1% or 2%, but estimates level out at the actual 100% population value of .060 and vary little across samples when the sampling proportion is higher than 3%. Estimates of H for the foreign-born population are more biased at low sampling rates and do not converge as quickly with increasing sampling rate. The true value is .043, but at a 5% sampling rate the median estimate is more than twice that. H90 behaves somewhat less consistently than H for the total population. Its true value is .063, but the median estimate at 5% is .092 and even with a 20% sample the estimate is upwardly biased at .070. Again the results are more biased for the foreign-born, not reaching a plateau until the sampling rate is over 10%. Finally the figure shows much greater variability across samples (at the same sampling rate) for NSI, as well as a slower convergence to the true value. We reach two tentative conclusions: 1) the upward bias that we suspected is a greater problem for subgroups than for the total population, presumably because data for subgroups come from numerically smaller samples, and 2) bias is smallest for H, larger for H90, and largest for NSI.

Figure 1 about here

We now turn to tables that summarize the results for all six cities for each measure. Our purpose is to quantify the patterns shown in Figure 1 and to confirm whether similar patterns are found in different cities.

1. Sample draw 1: H for the total population

Table 2 reports estimated values of H (total population) for sampling rates of 5%, 10%, 15%, and 20%, and in addition the table lists the actual full-population values of H for all six cities. In Chicago, as already displayed in Figure 1, the median estimate with a 5% sample is .078, which is nearly a third higher than the true value of .060. The median estimate from a 15% sample is considerably better, .066, although it also is biased upward. We find a similar pattern for all the cities. The median estimate at 5% is higher than the estimate at 15% by a margin of between .009 and .015. And even with a 15% sample, all median estimates are higher than the true value.

Table 2 about here

When social scientists rely on census data, of course, they do not work with a "median" sample. There is only one sample, drawn with methodologies that are designed to be unbiased but that are of course subject to sampling variation. Fortunately it appears that this variation has less impact on the results than do differences in the sampling rate. Although the maximum and minimum values are quite different, most values fall within a narrower range. Consider again the Chicago example. At a 15% sampling rate, the interquartile range of estimated H (between the 25th and 75th percentile) is from .065 to .066, just a little above or below the median estimate. This interquartile range is consistently higher when the sampling rate is lower (from .076 to .079 at 5%).

2. Sample draw 2: H for the white foreign-born population

Table 3 summarizes results for estimates of H for the foreign-born. As displayed above, Chicago's median estimate at 5% is much higher (.095) than the median estimate at 15% (.060), and both are higher than the true value (.043). In Chicago as well as in other cities the absolute size of this variation is greater than for the total population, ranging from .031 in the case of Cleveland (.085 vs. .054, with a true value of .039) to .065 in the case of Pittsburgh (.075 vs. .140, with a true value of .043). Not only is the upward bias greater for the foreign-born estimates, but the variability across samples is also larger. For 5% samples this variability is most pronounced for Los Angeles, where the interquartile range is from .109 to .133.

Table 3 about here

3. Sample draw 3: H90 for the total population

Table 4 switches to measures of H90. We again begin with the example of Chicago, where the actual value of H90 is .063. The median estimate from a 15% sample (comparable to Census 2000) is .073, and estimates range from .069 to .076. The median estimate from a 5% sample (comparable to the ACS) is .092, and estimates range from .086 to .098. Due solely to the difference in sampling rates, one median estimate is 26% higher than the other.

Table 4 about here

How large is the difference between the median estimate for a 15% sample and a 5% sample in other cities? The Chicago case turns out to be average in this respect, as differences range between .014 (Los Angeles) and .026 (Pittsburgh). We can compare this to the observed increase in income segregation (H90) from 2000 to 2007-2011. Here the average change (as shown in Table 1) was from .185 to .200, a difference of .015. In other words, it appears that the apparent increase in the last decade was no larger – and in most cases smaller – than could be expected on the basis of the reduction in sample size between Census 2000 and the ACS.

There also seems to be little likelihood of avoiding an upward bias by chance. In none of the six cities does the interquartile range at a 5% sampling rate overlap with the interquartile range at a 15% sampling rate. Further, in only one instance is the **minimum estimate** as low as the actual value of H90 (this is found at the 20% sampling rate in Pittsburgh). These observations reinforce our concern with the implications of measuring change with data sets that have markedly different sample sizes and sampling rates.

4. Sample draw 4: H90 for the white foreign-born population

The estimates of H90 for the white foreign-born population are summarized in Table 5. Using Chicago as an example, the median estimate of H90 is .073 at a 15% sampling rate but much higher (.127) at a 5% sampling rate. Both values are well above the true foreign-born white value of .046. Even at a 15% rate this difference is .027. In our parallel analysis of H90 for the total population (Table 4 above) Note that at this same sampling rate the difference between the true and median sample estimates was only .010. Why is the upward bias so much larger for the foreign-born white subgroup? Apparently (as we have confirmed with separate analyses based on average tract sample sizes, not shown here) both the sampling rate and the absolute size of the samples affect these estimates. For samples drawn from the total population, Chicago had valid income data for over a million households in 900 tracts. The average tract had 1123 households, so that a 5% sample would yield 56 households and a 15% sample would yield 169. For samples drawn from the foreign-born white population, Chicago had valid income data for only about 340,000 households in 852 tracts, a reduction of approximately two-thirds. The average tract had 399 households, so that a 5% sample would yield 19 households and a 15%

sample would yield 60. Further, since the foreign-born white population was still highly residentially segregated in 1940 (though not as segregated as the black population is today), one could expect more variation across tracts in actual sample sizes for foreign-born whites than for total population.

Table 5 about here

Chicago is not an anomaly. Among all six sampled cities, the actual value of H90 for foreign-born white households was below the value for the total population. Yet the median estimate was higher at every sampling rate, and the disparity was substantially greater at lower sampling rates. It appears that sample estimates of income segregation for subgroups of the population are subject to greater upward bias than estimates for the total population, especially at sampling rates as low as 5%.

5. Sample draws 5-6: NSI for the total population and white foreign-born population

We now repeat these analyses for another measure of income segregation, the NSI. The pattern is similar to what we observed with H and H90: estimates trend downward as the sampling rate increases, and the downward slope is more pronounced for foreign-born whites than for the total population. There is one noticeable difference in the NSI results: greater variation in estimates across samples than found even for H90 at the same sampling rate. In this respect estimates for NSI are less stable than estimates for H90, and much less stable than estimates for H.

Tables 6-7 summarize these results. Turning first to the Chicago case in Table 6 (total population) where the true value is .241, note the very large difference in the median estimates at a 15% sampling rate (.256) and at 5% (.286). The latter is 12% higher than the former, and both are higher than the true value. There are large differences in every city between the median estimates at 5% and 15%. Notice also the interquartile range in the estimates. At a 5% sampling rate, it is .038 in Chicago and it ranges up to .071 in the other five cities. Finally the upward bias is pronounced. In every case even the Q1 estimate with a 5% sample is considerably higher than the true value.

Tables 6-7 about here

As we saw with the rank-order measures H and H90, the results for the foreign-born white population show greater bias and more sampling variability than those for the total population. In the Chicago case, for example, with a true value of .227, the median estimate of NSI with a 5% sample (.326) is about 23% higher than with a 15% sample (.265). The interquartile range at a 5% sampling rate is .013 in Chicago and it ranges up to .027 in the other five cities. The upward bias in these estimates is greater than in the estimates of NSI for the total population – for four of the six cities, even at a 20% sample, the **minimum estimate** is higher than the true value.

Correcting the upward bias and unreliability in sample data

The standard approach to measuring income segregation is to begin with estimates of the income distribution within individual census tracts. These estimates are then manipulated in various ways to produce a measure of the disparities across tracts. We have shown that such measures are biased upward when based on sample data, and they are more strongly biased upward when calculated from samples in the range of sizes that are now available from the American Community Survey than in the range previously provided by the decennial census. They are also less reliable.

We now present two alternative approaches that make use of the original census or ACS sample data but avoid estimating mean values for individual tracts.[5] Each method requires access to data at the level of families or households, which has been facilitated by the expansion in the Census Bureau's system of Federal Statistical Research Data Centers (RDCs). One method uses these unit data to estimate the variance between tracts, a value that can be used to calculate the Coefficient of Variation which is a plausible measure of class segregation. The other estimates the variances within tracts, which can summed to the city or metro level and converted directly into the NSI. Application of this approach to ranked-data yields the rank-order relative diversity index (R). Finally, we propose a method to estimate (and therefore to correct for) the upward bias in H (or H90, etc.) that is due to smaller sample sizes.

None 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). However knowledge about this sampling variation gleaned from the 1940 simulations offers a possibility of establishing confidence intervals around the estimates.

Restricted Maximum Likelihood estimation (REML)

Direct estimation of variance components through Restricted Maximum Likelihood estimation (REML) produces superior estimates of the variance across local areas, which is a potential measure of class segregation. This estimate of variance is indirectly related to the NSI. Recall that the NSI is simply the square root of the ratio of the variance between areas to total variance in the population. However the variance used for the numerator of the NSI (based on the squared difference between each tract mean and the overall mean) is weighted by the number of observations in the tract. Large tracts can contribute more to the variance than small tracts. In the REML estimate of variance, every tract is weighted equally. Nevertheless it can potentially serve as an indicator of income segregation, perhaps in its standardized form as the Coefficient of Variation (the standard deviation divided by the overall mean).[6]

REML is a variant of older Maximum Likelihood (ML) techniques that have been applied to estimating the variance components in a standard linear model (Harville 1977). A problem with ML estimators is that they do not take into account the loss in degrees of freedom resulting from the estimation of the model's fixed effects. The "restricted" ML approach deals with this problem by transforming the original data set into a set of contrasts calculated from the data. The likelihood function is then calculated from the probability distribution of these contrasts. Although REML does not in principle produce unbiased estimates of variance components, it is less biased than ML. We use REML procedures available in R (Schnabel, Koontz, and Weiss 1985) to calculate estimates from the 1940 data.

Figure 2 reports a comparison between the performance of direct estimates and REML estimates of the variance across tracts in median family incomes in Chicago in 1940. It shows the distribution of estimates from 100 sample draws at each sampling rate. The solid horizontal black line indicates the actual variance. There is a familiar pattern for direct estimates: at low sampling rates (up to about 7%) the estimates are much higher than the actual value. Also there is a very wide variation in results even at a 20% sampling rate. In contrast the REML estimates are close to the actual value even at the lowest sampling rates, although they tend to be underestimates. The main advantage of the REML approach is that most estimates are within a narrow range. Table 8 reports results for all six cities in 1940 for selected sampling rates. This table reports the average bias in these estimates, defined as the difference between the true value and the median estimate from 100 samples at a given sampling rate. Results vary across cities. In general the bias from either approach is lowest in Los Angeles and Pittsburgh but much higher in Philadelphia. REML results understate the variance (though often by smaller amounts) while direct estimates overstate it. But although REML results are more reliable, the average bias in REML estimates is larger than the average bias of direct estimates for Chicago and Philadelphia.

Figure 2 and Table 8 about here

Sparse-Sampling Variance Decomposition (SSVD)

Another approach to estimating class segregation is to begin with the within-tract variances in household income and then make use of the fact that the within and across variation must sum to the total variation. We presume that the total variance in income can be reliably estimated from the sample data in large cities and metropolitan areas. If we can estimate the total within variance, the between variance follows directly. For convenience we refer to this approach as Sparse-Sampling Variance Decomposition (SSVD). Although to our knowledge this has not been done before, the source of its efficacy is parallel to the use of "small t, large n" panel data methods (Mundlak 1978) where one takes advantage of the fact that a large number of census tracts can be leveraged to average-out errors in the estimation of the distribution within each tract. Not only can this approach be used to obtain unbiased estimates of the NSI in sparse samples but, as illustrated below, it also can be used for other variance-based estimates of segregation such as the rank-order variance ratio index R (Reardon 2011).[7]

The expected total variation of income within tracts for the city is the average of the track-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 populated-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.[8] 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. This logic is expressed more formally in Appendix I.

Figure 3 compares estimates of the NSI using the SSVD and direct approaches in the case of Chicago. Again the figure is based on 100 simulations for each sampling rate. For comparability we have used the same vertical scale as the other NSI figures. The direct estimates are biased upward, close to the true value only at around 15% and above. The direct estimates are also quite variable even at the higher sampling rates. In contrast the SSVD estimates are very close to the true value, with a slight positive bias for the lowest sampling probabilities. However, the variability of sample estimates for the SSVD is similar to variability for direct estimates (except at the smallest sampling rates).

Figure 3 about here

Table 9 summarizes results for all six cities in terms of bias, using the same measure of bias as in Table 8. Both measures have some upward bias, much larger for the direct measure. At the 5% sampling rate, the bias for the SSVD estimate is about one third or half as large as for the direct estimate for Cleveland, Detroit, Los Angeles, Philadelphia, and Pittsburgh, and it is almost nil for Chicago. Bias is generally smaller for both estimates at a higher sampling rate. At 15% the bias of the SSVE estimate is close to zero for Chicago, Detroit, and Pittsburgh but still appreciable for the direct estimate. In Cleveland, Los Angeles, and Philadelphia the bias in the SSVD estimate is about two-thirds as large as for the direct estimate.[9]

Table 9 about here

Compared to the direct estimates of NSI, the SSVD approach has the advantage of much less bias, although variability across samples could affect the accuracy of estimates for particular cities. Fortunately, the average NSI for a set of cities and its trajectory over time are likely to be much more reliably estimated by this approach in spite of the sampling variability for estimates for each individual city.

A critique of the NSI is that it is affected by changes in the distribution of income even when the ranking is preserved. While we do not believe that the SSVD, or a related procedure, can be applied to entropy based estimates of segregation such as H, the approach can be used for other rank-based measures. Of particular note is R^R, the rank-order variance ratio index, "which can be interpreted as a measure of the average variance of the neighborhood cumulative percentile density function" (Reardon 2011 p26). The idea is simply to 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]

Figures 4-5 about here

Figures 4 and 5 show the SSVD and Direct estimates for R^R and R^R_{90} for different sampling weights as in Figure 1 (we plot the square-root of each for comparability with the NSI) in Chicago. The picture is a familiar one. The mean direct estimates of $\sqrt{R^R}$ are about 45% too big for a 1 percent sample and about 10% too big for a 5% sample. On the other hand the mean SSVD estimate is very close to the population value throughout the whole range of the sample and is quite precisely estimated. The standard deviation at 1% is just 0.008, falling to 0.001 for a 20% sample. The fact that these estimates are more precise than NSI is not surprising; the process of scaling the incomes by the cumulative distribution substantially reduces the relative influence of very high incomes. The results for R^R_{90} in Figure 5 show the same pattern. The SSVD estimate follows the population estimate extremely closely while the direct estimate is substantially biased at low sampling rates. In addition to showing the generalizability of the SSVD concept, these estimates illustrate the particular value of variance-based measures, given that these measures share some desirable properties of rank-order entropy-based measures (Reardon 2011).

Correcting the bias in H

Above we provide a new method to construct an unbiased estimate of the NSI as well as several other variance-based measures of segregation, but this method cannot be applied to entropy based measures such as H and H90. We have, however, derived a convenient and feasible approximation of the bias in such measures 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 yields an estimate of the bias that can be then subtracted from the sample estimate to get an approximate true estimate. The procedure 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. These conditions appear to be met in the 5-15% range of tract sampling rates given average U.S. tract sizes.

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

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

A second order approximation to the sample entropy is

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

Taking expectations with respect to s_i yields

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

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. The formal expression for the expected approximate bias for $H90$ adds up the tract-specific terms:

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

In order to assess the accuracy of the approximation we need to provide bounds on O_j. We have derived a complex formula for O_j, available on request, that can be applied for any tract population, size, and income distribution. As there are only M_j+1 possible values of p_j for any tract it is possible to check every possible value to assess the maximum approximation bias. 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, it seems 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:

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

The procedure provides a good fit to Table 4. Based on the 1940 census data, a 5% sample for all tracts in Chicago leads to an approximate expected bias of .026. The difference between the actual and the median estimate for $H90$ for Chicago in Table 4 is in fact .029. For 15% sampling the average expected bias is .008. In Table 4, the actual bias is .010. Figure 6 shows the relationship for all sampling rates and cities in Table 4 as well as a 45 degree line.

While the predicted bias is a bit higher overall than that in Table 4, the relationship between the two measures is extremely close.

Figure 6 about here

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,

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

which takes the value 0.017 for a 5% sample in Chicago. The difference between the true and the median 5% sample in Table 2 is .018. For a 15% sample the estimated bias is .006. For Chicago at 15% sampling, the true and median differences are both .006. Figure 7 shows the relationship for all sampling rates and cities in Table 2 along with a 45-degree line. Again, the fit is extremely close.

Figure 7 about here

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. Analyses based on the 2000 Census and 2007-2011 ACS offer mixed results on these points, depending on the measure of income segregation. Based on the rank-order measures used by Bischoff and Reardon there was a tendency for segregation to increase modestly for the total population and greatly for African American families. Based on the NSI used by Jargowsky (which we have updated here) there was a modest decline for the total population and a modest increase for black families. These findings suggest that more attention needs to be given to the characteristics of alternative measures that reflect different aspects of income segregation, and they attenuate the strong conclusions in recently published research.

We draw attention to a different concern. It is well known that the reliability of local area estimates is conditional on sampling rates and sample sizes. This fact matters here because the Bureau of the Census has shifted from the one-in-six sampling of "long form" data used in recent decennial censuses through 2000 to smaller annual samples in the American Community Survey. The net result is that tract estimates that were previously based on approximately 15% sampling rates are now based on samples of around 5%, even when data are pooled over five years. We have asked how these smaller samples affect estimates of income segregation that are aggregated from sample data for many tracts in every metropolitan region.

Our approach has been to exploit the available 100% samples that have been made available recently by the Minnesota Population Center for 1940, the first year when the census collected data on wage and salary income. For six major cities we drew random samples for the whole population and for the foreign-born population (a minority category that is roughly comparable in population share to today's black population in these cities). In this way we have been able to calculate the same measures of income segregation used by contemporary researchers and compare results for samples of varying sizes. We believe this is a reasonable simulation with real data of how measures based on a 15% sample in 2000 would compare to a 5% sample in 2007-2011 in an urban area where there had been no change in income segregation. The answer is clear and consistent. The change in samples would result in measured increases in segregation for the whole population that are as large as or larger than those that have been observed in the last decade. Further, the effect of sampling rate on the foreign-born population is considerably stronger, and it may be sufficient in itself to account for the higher and more rapidly rising values of minority income segregation that have been reported, even prior to implementation of the ACS.

We do not conclude that the reported increase in income segregation is necessarily illusory. The single sample that is available for 2000 and the single sample for 2007-2011 may faithfully represent the population, and we have no direct method to calculate confidence intervals around their results. However a neutral observer will suspect that the confidence intervals for segregation estimates are likely to be large, especially for the ACS data, and may conclude that a null hypothesis of no change over time – and no difference between the total population and the minority population – cannot be rejected.

Given the substantive importance of trends in class segregation, we have put forward three alternative approaches that are less susceptible to the Census Bureau's changes in data collection. We have shown that a measure based on the unweighted variance across tracts can be more reliably measured by Restricted Maximum Likelihood estimation than by direct calculation from sample data. Another approach based on estimating variance within tracts results in estimates of the NSI and variants of H and H90 that are less biased and more reliable than direct estimates from aggregated tract data. Researchers with access to confidential census data centers can take advantage of the original individual-level data to assess trends in these alternative measures of income segregation. A third approach corrects H and H90 (calculated from published tract-level income distributions) for small-sample bias using a second order approximation. If results based on these methods and measures consistently showed increasing segregation, we would have more confidence in previously reported conclusions. If not, we should turn our attention to the question of what factors other than income inequality itself have been holding segregation steady or even reducing it.

Footnotes

1. 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

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

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."

2. 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:

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

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

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

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:

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

3. 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 area 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, H10, and H90 in Table 1 are based on data provided by them. We calculated the NSI for a comparable sample: all metros with 500,000 or more residents in 2007-2011 (n=117), and black NSI for those metros with more than 30,000 black families in 2007-2011 (n=66). The means and standard deviations in Table 1 are unweighted.

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 in each city for average sample sizes per tract of 50, 100, 150, 200, and 250. In these cities 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. Here we present statistics on the sampling distributions of estimates for 5% and 15% samples in order to focus on the approximate difference between what might be expected between the 2000 Census and 2007-2011 ACS.

5. We also evaluated another alternative that was devised by Sean Reardon and Kendra Bischoff and described to us in personal communications. They applied this method in comparisons involving ACS data beginning with Bischoff and Reardon (2015), though it is not documented in published work. They describe the procedure as follows: "We took the published (sample-based) tabulations of the income distribution in each tract and created a simulated data set containing one observation for each family in the tabulation … Then we drew a sample, without replacement, of size 50*T (where T is the number of tracts in the metro) from this simulated metro population. This results in a data set with an average of 50 families per tract (though larger tracts will tend to have more, and smaller tracts fewer, in proportion to their population size). We then compute income segregation based on this sampled data set. We repeat this process 100 times … then compute the average income segregation across the resulting estimates. The idea is that income segregation measures are biased upward when within-tract samples are small, so we wanted to keep the within-tract samples the same size, on average, across metros/year/groups. Because we base our estimates on samples of size 50, they will be

biased upwards, but our reasoning was they would be biased upwards the same amount everywhere (so trends and comparisons would be valid)."

We examined the consequences of this sampling procedure using 1940 income data and found that measures from subsamples based originally on a 5% sample are more biased upwards than measures from subsamples based originally on a 15% sample. Results are reported in Appendix II.

6. Firebaugh (1999) describes the use of the squared Coefficient of Variation (V^2) as a measure of income inequality across nations. Especially relevant here is his discussion of the differences between measures that are weighted by national population size and those that are unweighted.

7. As a referee has correctly pointed out, the SSVD is related to de-biasing procedures developed to look at the extent to which relatively productive workers are sorted into high productivity firms. In particular, it is recognized that due to limited mobility over time of workers across firms, worker and firm effects yields biased estimates of the population variance in worker and firm productivity as well as the correlation between worker and firm productivity (Andrews and Gill 2008; Card, Heining and Kline 2012, Peterson, Penner and Hogsnes 2014). Andrews and Gill develop a feasible procedure to correct the bias. However, their formula is solving a somewhat different problem in that it uses worker mobility over time to separately identify worker and firm productivity, whereas we are focused on sorting of households across neighbors using cross-sectional data. The key difference is that the theoretical bias in the estimate of NSI from a sparse sample depends on the population variance of income in each neighborhood and thus a feasible procedure must account for the fact that this within variance is noisily estimated when sampling is sparse. This is precisely the problem solved by the SSVD.

8. It can also be shown that the estimated NSI is asymptotically normal with root n convergence

in the number of tracts. This provides a guide to the relative precision of the estimates. For example, Chicago has almost 1000 tracts. Standard errors would be approximately 41 percent larger in a city with half as many tracts. Simulations using a random sample of Chicago's tracts (not shown) confirm this conclusion.

9. An unusual result for Philadelphia is that the bias in both estimates does not decline much with a higher sampling rate. Table 6 similarly showed surprisingly high bias for both estimates of the variance in this case. This pattern raises an additional methodological concern. On closer inspection of the Philadelphia income data, we noticed several outliers with extremely high incomes. When these households happened to be drawn in a sample, which was more common at higher sampling rates, their inclusion greatly affected the estimates of class segregation. To gauge this effect, we reran the direct and REML estimates of between-variance and the direct and SSVD estimates of NSI after top-coding data at the 99.9th percentile for every city (about $350,000 in 2016 dollars). In all cases the direct estimates are more precise but still quite biased; REML and SSVD estimates are much improved, and they remain superior to the direct estimates. Decisions about how to deal with very high reported incomes have to be made by individual researchers. Note that when dealing with grouped data published by the Census Bureau the problem presents itself differently – the top value and the average value for the top category are not given but have to be imputed from the distribution.

10. 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

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

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

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

where , and f_{ij}^s is the cumulative percentile of a household in the sampled city population, accounting for any differential sampling weights by tract, S_j indexes the sample from tract *j* ordered such that if $i \le i^*$ then $f_{ij}^s \le f_{i^*j}^s \,\forall i, i^* \in S_j$, and $N_j = |S_j|$ is the number of sampled households in that sample. The number 6 comes from the fact that $\int_{s-0}^{1} p(1-p)dp = 1/6$ 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.

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

Andrews and Gill 2008

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.

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.

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://cepa.stanford.edu/content/measures-income- segregation (accessed September 1, 2014).

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: Stability of estimates from SSVD

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 , S_j the set of households in tract j that are in the sample, $N_j = \| S_j \|$ and $M_j = \| P_j \|$.

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

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

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

Denote the sample analogs of W and T as

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

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

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:

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

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

Appendix II: Resampling procedure 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, 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*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, the 5 and 15 percent samples for the Bischoff and Reardon procedure (“N*50 Cen”) do not yield similar biases. 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 increase 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

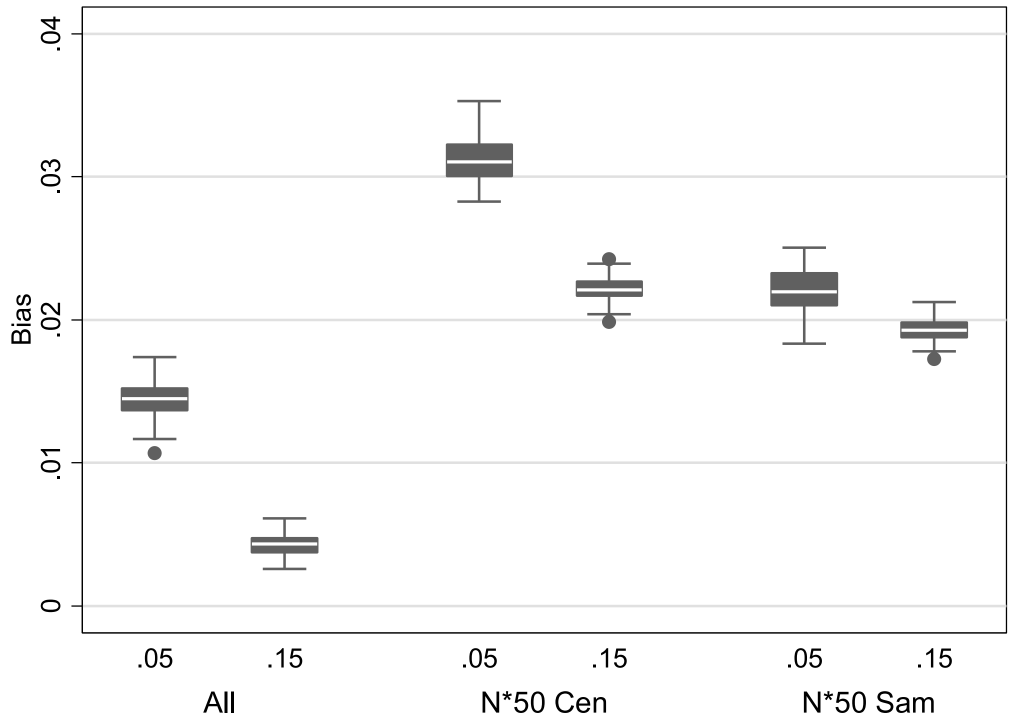

Table 1. Observed income segregation in metropolitan regions, 2000 and 2007-2011				
	Total		**Black**	
	2000	**2007-2011**	**2000**	**2007-2011**
H				
Chicago	0.164	0.168	0.147	0.186
Cleveland	0.158	0.172	0.160	0.202
Detroit	0.162	0.194	0.094	0.141
Los Angeles	0.174	0.179	0.177	0.268
Philadelphia	0.189	0.207	0.127	0.175
Pittsburgh	0.114	0.130	0.184	0.254
Mean large metros [1]	*0.134*	*0.147*	*0.169*	*0.251*
SD	*0.027*	*0.027*	*0.055*	*0.082*
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
Mean large metros [1]	*0.146*	*0.163*	*0.171*	*0.270*
Standard deviation	*0.031*	*0.029*	*0.054*	*0.093*
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
Mean large metros [1]	*0.185*	*0.200*	*0.231*	*0.341*
SD	*0.036*	*0.036*	*0.079*	*0.096*
NSI				
Chicago	0.241	0.228	0.257	0.319
Cleveland	0.275	0.270	0.314	0.347
Detroit	0.254	0.259	0.256	0.370
Los Angeles	0.340	0.286	0.379	0.349
Philadelphia	0.266	0.247	0.268	0.343
Pittsburgh	0.282	0.243	0.432	0.428
Mean large metros [1]	*0.218*	*0.202*	*0.252*	*0.260*
Standard deviation	*0.045*	*0.040*	*0.075*	*0.058*

[1] See footnote 1 for selection of large metros and metros with large black populations.

Table 2. Distribution of H (total population): true value and estimates at varying sampling rates					
	Minimum	Q_1	Median	Q_3	Maximum
Chicago H=.060					
5% Sample	0.074	0.076	0.078	0.079	0.082
10% Sample	0.066	0.068	0.069	0.069	0.071
15% Sample	0.064	0.065	0.066	0.066	0.069
20% Sample	0.062	0.064	0.064	0.065	0.066
Cleveland H=.059					
5% Sample	0.068	0.073	0.075	0.077	0.083
10% Sample	0.061	0.065	0.067	0.068	0.070
15% Sample	0.061	0.063	0.064	0.065	0.068
20% Sample	0.060	0.062	0.063	0.064	0.065
Detroit H=.058					
5% Sample	0.067	0.072	0.073	0.075	0.078
10% Sample	0.061	0.064	0.065	0.066	0.069
15% Sample	0.060	0.062	0.063	0.064	0.066
20% Sample	0.059	0.061	0.062	0.062	0.064
Los Angeles H=.042					
5% Sample	0.047	0.053	0.055	0.057	0.063
10% Sample	0.043	0.047	0.049	0.050	0.055
15% Sample	0.041	0.045	0.046	0.047	0.051
20% Sample	0.041	0.044	0.045	0.046	0.048
Philadelphia H=.056					
5% Sample	0.066	0.069	0.070	0.072	0.076
10% Sample	0.059	0.062	0.063	0.064	0.068
15% Sample	0.058	0.060	0.061	0.061	0.064
20% Sample	0.057	0.059	0.060	0.060	0.062
Pittsburgh H=.054					
5% Sample	0.068	0.074	0.076	0.077	0.087
10% Sample	0.054	0.063	0.064	0.066	0.070
15% Sample	0.055	0.059	0.061	0.063	0.067
20% Sample	0.053	0.057	0.059	0.060	0.066

Table 3. Distribution of H (white foreign population): true value and estimates at varying sampling rates					
	Minimum	**Q_1**	**Median**	**Q_3**	**Maximum**
Chicago H=.043					
5% Sample	0.089	0.093	0.095	0.097	0.104
10% Sample	0.066	0.068	0.069	0.070	0.073
15% Sample	0.057	0.059	0.060	0.061	0.064
20% Sample	0.052	0.055	0.056	0.057	0.058
Cleveland H=.039					
5% Sample	0.071	0.081	0.085	0.087	0.094
10% Sample	0.053	0.060	0.061	0.063	0.071
15% Sample	0.047	0.052	0.054	0.055	0.061
20% Sample	0.046	0.049	0.050	0.051	0.057
Detroit H=.040					
5% Sample	0.084	0.087	0.090	0.092	0.099
10% Sample	0.058	0.063	0.064	0.066	0.070
15% Sample	0.051	0.055	0.056	0.057	0.060
20% Sample	0.048	0.051	0.052	0.053	0.055
Los Angeles H=.041					
5% Sample	0.068	0.109	0.123	0.133	0.176
10% Sample	0.054	0.075	0.084	0.094	0.108
15% Sample	0.048	0.065	0.071	0.077	0.092
20% Sample	0.051	0.060	0.065	0.069	0.089
Philadelphia H=.035					
5% Sample	0.076	0.082	0.085	0.088	0.097
10% Sample	0.053	0.057	0.059	0.061	0.067
15% Sample	0.045	0.050	0.051	0.052	0.057
20% Sample	0.043	0.046	0.047	0.048	0.052
Pittsburgh H=.043					
5% Sample	0.115	0.133	0.140	0.147	0.169
10% Sample	0.074	0.085	0.089	0.095	0.108
15% Sample	0.063	0.071	0.075	0.079	0.089
20% Sample	0.052	0.062	0.066	0.069	0.076

Table 4. Distribution of H90 (total population): true value and estimates at varying sampling rates

	Minimum	Q_1	Median	Q_3	Maximum
Chicago H90=.063					
5% sample	0.086	0.090	0.092	0.093	0.098
10% sample	0.073	0.076	0.077	0.079	0.082
15% sample	0.069	0.072	0.073	0.074	0.076
20% sample	0.067	0.069	0.070	0.071	0.074
Cleveland H90=.061					
5% sample	0.078	0.087	0.089	0.092	0.104
10% sample	0.063	0.072	0.075	0.077	0.082
15% sample	0.064	0.068	0.070	0.072	0.076
20% sample	0.062	0.066	0.067	0.068	0.072
Detroit H90=.053					
5% sample	0.069	0.076	0.079	0.081	0.088
10% sample	0.059	0.064	0.066	0.067	0.072
15% sample	0.056	0.060	0.061	0.063	0.066
20% sample	0.055	0.058	0.060	0.061	0.063
Los Angeles H90=.062					
5% sample	0.075	0.080	0.083	0.084	0.092
10% sample	0.066	0.071	0.073	0.074	0.077
15% sample	0.065	0.068	0.069	0.071	0.074
20% sample	0.063	0.066	0.067	0.068	0.071
Philadelphia H90=.056					
5% sample	0.071	0.076	0.079	0.081	0.087
10% sample	0.060	0.066	0.067	0.069	0.074
15% sample	0.059	0.062	0.064	0.064	0.067
20% sample	0.057	0.060	0.062	0.063	0.066
Pittsburgh H90=.084					
5% sample	0.102	0.114	0.121	0.126	0.140
10% sample	0.091	0.098	0.101	0.106	0.116
15% sample	0.085	0.092	0.095	0.099	0.104
20% sample	0.083	0.090	0.093	0.095	0.101

Table 5. Distribution of H90 (white foreign population): true value and estimates at varying sampling rates					
	Minimum	**Q_1**	**Median**	**Q_3**	**Maximum**
Chicago H90=.046					
5% sample	0.118	0.124	0.127	0.130	0.139
10% sample	0.077	0.085	0.086	0.089	0.093
15% sample	0.067	0.072	0.073	0.074	0.079
20% sample	0.062	0.065	0.066	0.067	0.071
Cleveland H90=.034					
5% sample	0.091	0.104	0.110	0.114	0.133
10% sample	0.062	0.068	0.071	0.076	0.092
15% sample	0.051	0.055	0.059	0.062	0.070
20% sample	0.044	0.051	0.052	0.055	0.062
Detroit H90=.033					
5% sample	0.106	0.112	0.116	0.120	0.138
10% sample	0.061	0.070	0.074	0.076	0.087
15% sample	0.051	0.058	0.059	0.061	0.068
20% sample	0.044	0.051	0.053	0.055	0.062
Los Angeles H90=.055					
5% sample	0.133	0.146	0.152	0.159	0.169
10% sample	0.089	0.101	0.105	0.108	0.119
15% sample	0.079	0.085	0.088	0.091	0.097
20% sample	0.069	0.077	0.079	0.081	0.087
Philadelphi H90=.034					
5% sample	0.097	0.108	0.113	0.119	0.130
10% sample	0.064	0.070	0.073	0.076	0.082
15% sample	0.053	0.058	0.060	0.063	0.069
20% sample	0.045	0.051	0.053	0.055	0.061
Pittsburgh H90=.064					
5% sample	0.179	0.202	0.210	0.219	0.247
10% sample	0.116	0.132	0.138	0.144	0.169
15% sample	0.095	0.110	0.114	0.119	0.138
20% sample	0.084	0.097	0.100	0.104	0.119

Table 6. Distribution of NSI (total population): true value and estimates at varying sampling rates					
	Minimum	Q_1	Median	Q_3	Maximum
Chicago NSI=.241					
5% sa	0.211	0.267	0.286	0.305	0.349
10% s	0.207	0.249	0.263	0.274	0.311
15% s	0.213	0.241	0.256	0.269	0.306
20% s	0.221	0.240	0.251	0.262	0.281
Cleveland NSI=.234					
5% sa	0.183	0.265	0.292	0.308	0.333
10% s	0.178	0.228	0.271	0.285	0.307
15% s	0.172	0.238	0.260	0.278	0.297
20% s	0.183	0.234	0.263	0.276	0.296
Detroit NSI=.183					
5% sa	0.160	0.206	0.245	0.276	0.301
10% s	0.150	0.189	0.208	0.229	0.264
15% s	0.147	0.184	0.204	0.225	0.271
20% s	0.156	0.186	0.203	0.219	0.250
Los Angeles NSI=.168					
5% sa	0.135	0.213	0.228	0.242	0.270
10% s	0.117	0.189	0.207	0.219	0.241
15% s	0.118	0.191	0.205	0.213	0.232
20% s	0.120	0.187	0.199	0.210	0.229
Philadelphia NSI=.191					
5% sa	0.225	0.280	0.293	0.311	0.448
10% s	0.214	0.261	0.274	0.287	0.411
15% s	0.210	0.254	0.267	0.275	0.307
20% s	0.190	0.245	0.255	0.267	0.296
Pittsburgh NSI=.261					
5% sa	0.233	0.285	0.313	0.325	0.356
10% s	0.212	0.276	0.294	0.304	0.324
15% s	0.228	0.263	0.278	0.292	0.311
20% s	0.222	0.254	0.273	0.287	0.310

Table 7. Distribution of NSI (white foreign population): true value and estimates at varying sampling rates					
	Minimum	Q_1	Median	Q_3	Maximum
Chicago NSI=.227					
5% sample	0.288	0.317	0.326	0.330	0.365
10% sample	0.250	0.274	0.282	0.287	0.298
15% sample	0.225	0.257	0.265	0.271	0.282
20% sample	0.228	0.246	0.254	0.260	0.271
Cleveland NSI=.216					
5% sample	0.235	0.293	0.304	0.312	0.345
10% sample	0.221	0.256	0.264	0.271	0.288
15% sample	0.209	0.244	0.251	0.258	0.277
20% sample	0.211	0.236	0.243	0.249	0.264
Detroit NSI=.188					
5% sample	0.272	0.300	0.309	0.323	0.589
10% sample	0.238	0.260	0.265	0.275	0.422
15% sample	0.215	0.237	0.247	0.256	0.430
20% sample	0.203	0.223	0.237	0.246	0.363
Los Angeles NSI=.120					
5% sample	0.236	0.298	0.314	0.325	0.353
10% sample	0.208	0.242	0.255	0.264	0.285
15% sample	0.192	0.214	0.235	0.248	0.275
20% sample	0.168	0.201	0.227	0.235	0.252
Philadelphia NSI=.197					
5% sample	0.215	0.291	0.308	0.318	0.361
10% sample	0.209	0.239	0.256	0.267	0.284
15% sample	0.197	0.226	0.234	0.245	0.262
20% sample	0.202	0.218	0.226	0.233	0.256
Pittsburgh NSI=.228					
5% sample	0.308	0.384	0.402	0.411	0.515
10% sample	0.257	0.319	0.332	0.340	0.391
15% sample	0.238	0.284	0.304	0.317	0.344
20% sample	0.222	0.266	0.287	0.299	0.356

Table 8. Bias of estimated variance from direct and REML methods (difference between the true value and the median estimate from 100 samples)							
Sampling Rate		**Chicago**	**Cleveland**	**Detroit**	**Los Angeles**	**Philadelphia**	**Pittsburgh**
0.01	**Direct**	356,170	209,694	210,471	132,665	301,523	311,511
	REML	-97,496	-36,349	-26,183	-3,772	-212,950	-2,379
0.03	**Direct**	106,030	87,089	88,107	49,870	130,133	135,362
	REML	-96,455	-24,173	-26,049	-3,458	-223,121	-2,132
0.05	**Direct**	51,079	48,703	103,174	30,999	107,445	82,785
	REML	-91,008	-20,298	-16,964	-1,161	-209,805	-1,362
0.07	**Direct**	44,653	39,351	59,549	21,506	129,583	64,281
	REML	-86,656	-15,574	-16,420	-1,980	-194,067	-752
0.09	**Direct**	28,312	38,049	49,060	16,999	106,003	56,647
	REML	-82,584	-14,166	-15,074	-2,432	-199,860	5,201
0.15	**Direct**	1,929	16,606	35,448	11,970	93,047	33,799
	REML	-73,107	-14,819	-9,557	-1,158	-157,321	821
0.19	**Direct**	-11,423	16,083	23,618	10,036	64,555	28,167
	REML	-72,429	-11,615	-13,017	-845	-147,350	1,083

Table 9. Bias of NSI from direct and SSVD methods (difference between the true value and the median estimate from 100 samples)							
Sampling Rate		**Chicago**	**Cleveland**	**Detroit**	**Los Angeles**	**Philadelphia**	**Pittsburgh**
0.01	**Direct**	0.161	0.153	0.197	0.150	0.207	0.170
	SSVD	0.027	0.035	0.067	0.049	0.077	0.017
0.03	**Direct**	0.071	0.083	0.105	0.079	0.122	0.072
	SSVD	0.009	0.042	0.042	0.039	0.066	0.016
0.05	**Direct**	0.046	0.062	0.067	0.057	0.101	0.051
	SSVD	0.006	0.033	0.022	0.033	0.063	0.015
0.07	**Direct**	0.030	0.048	0.052	0.050	0.091	0.040
	SSVD	0.005	0.026	0.017	0.032	0.062	0.012
0.09	**Direct**	0.025	0.038	0.037	0.042	0.078	0.028
	SSVD	0.005	0.022	0.011	0.027	0.051	0.008
0.15	**Direct**	0.016	0.031	0.017	0.033	0.070	0.023
	SSVD	0.003	0.021	0.001	0.024	0.054	0.009
0.19	**Direct**	0.010	0.023	0.011	0.029	0.067	0.016
	SSVD	-0.001	0.015	-0.004	0.022	0.054	0.005

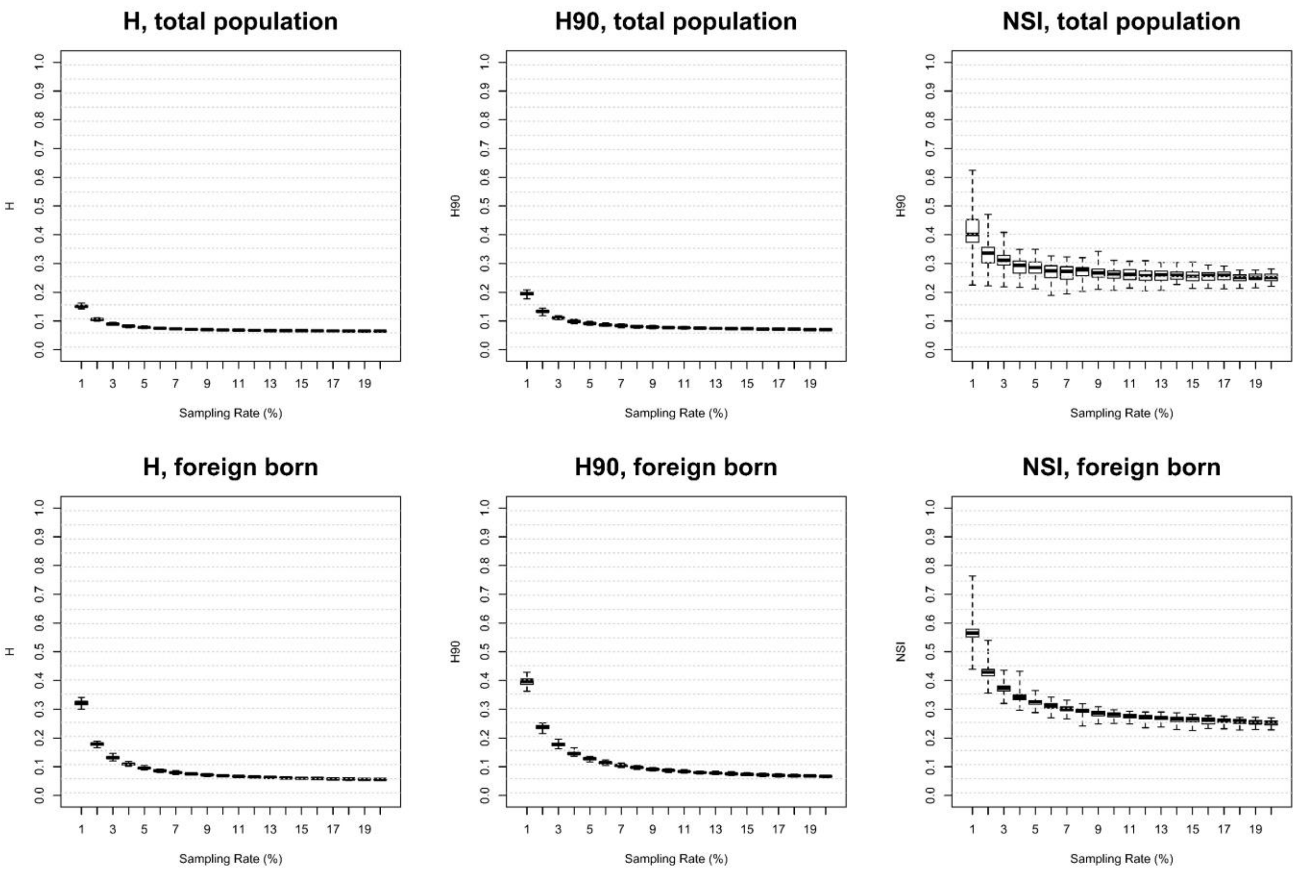

Figure 1. Relation between sampling rate and estimated values of H10, H90, and NSI (total households and white foreign-born households) based on draws from the 1940 population data for Chicago.

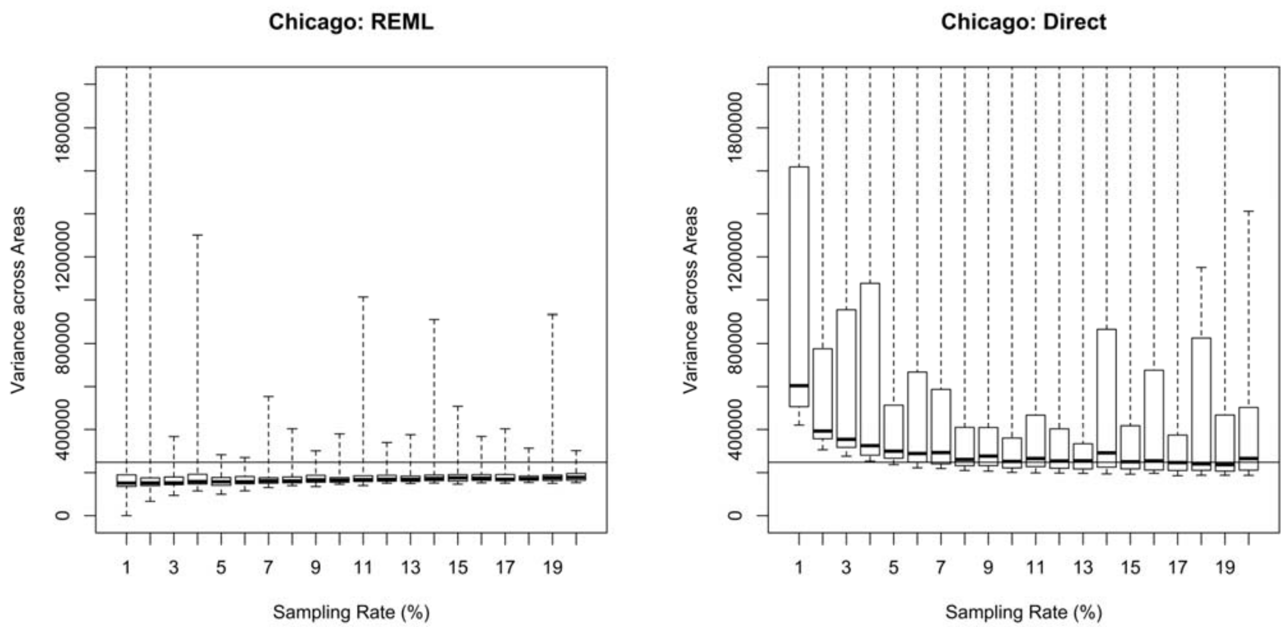

Figure 2. Relation between sampling rate and estimated values of variance in family income across areas based on draws from the 1940 population data for Chicago: direct and REML estimates.

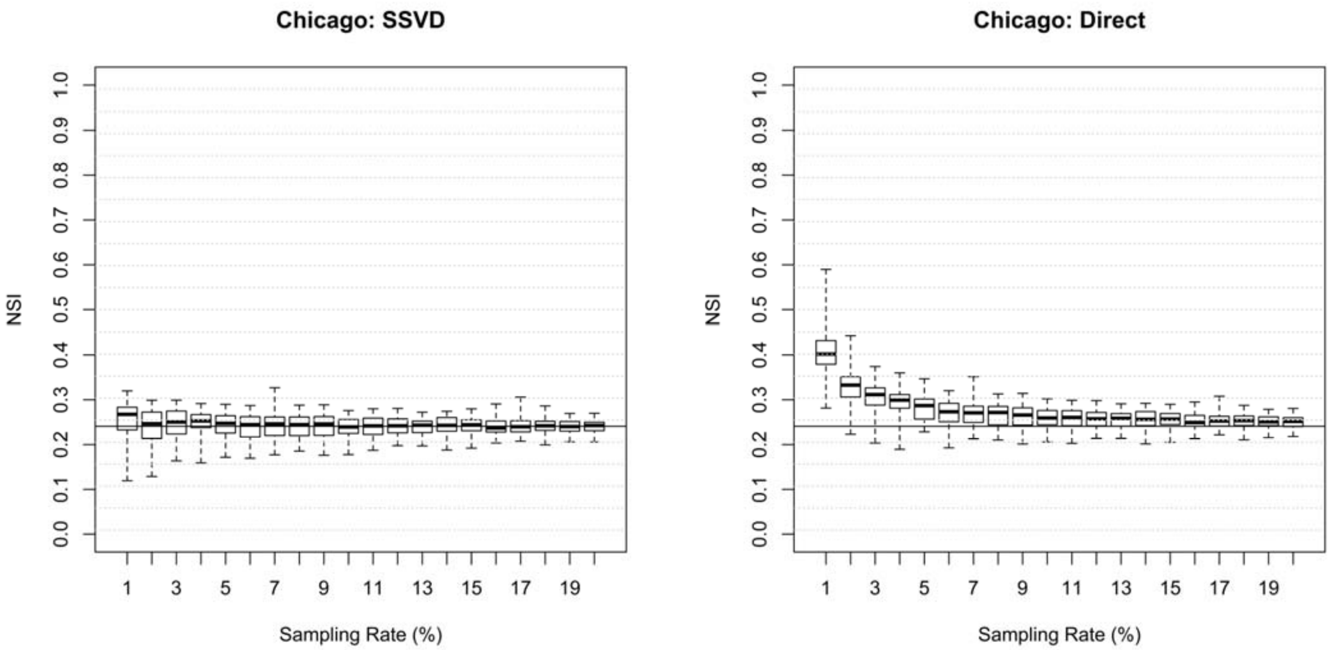

Figure 3. Relation between sampling rate and estimated values of NSI for family income across areas based on draws from the 1940 population data for Chicago: direct and SSVD estimates.

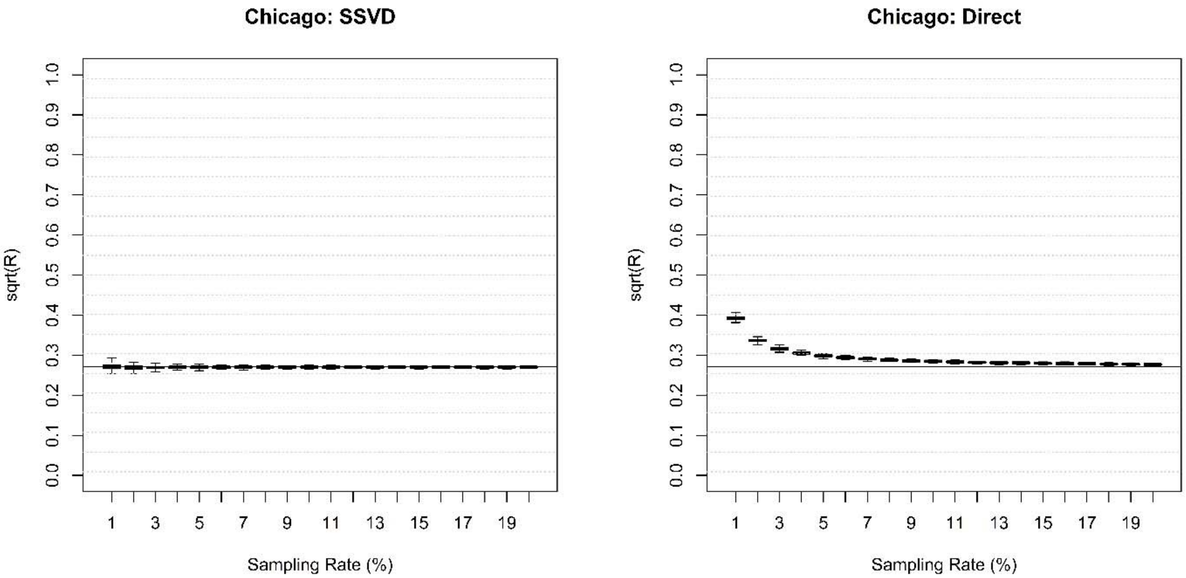

Figure 4. Relation between sampling rate and estimated values of rank-income variance ratio index $\sqrt{R^R}$ for family income across areas based on draws from the 1940 population data for Chicago: direct and SSVD estimates.

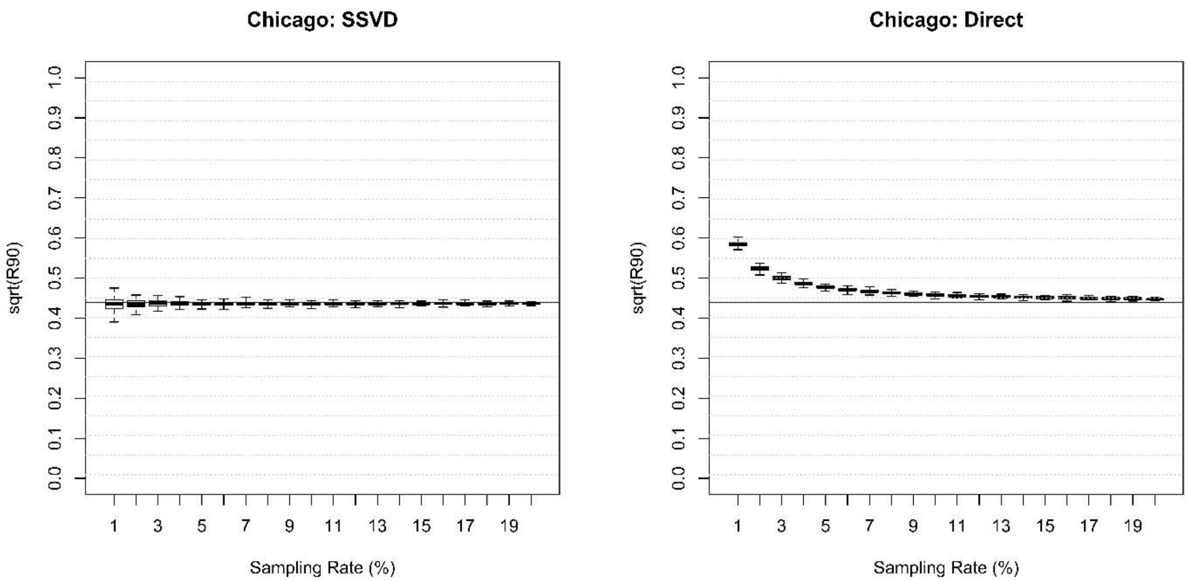

Figure 5. Relation between sampling rate and estimated values of 90 percentile variance ratio index $\sqrt{R_{90}^{R}}$ for family income across areas based on draws from the 1940 population data for Chicago: direct and SSVD estimates.

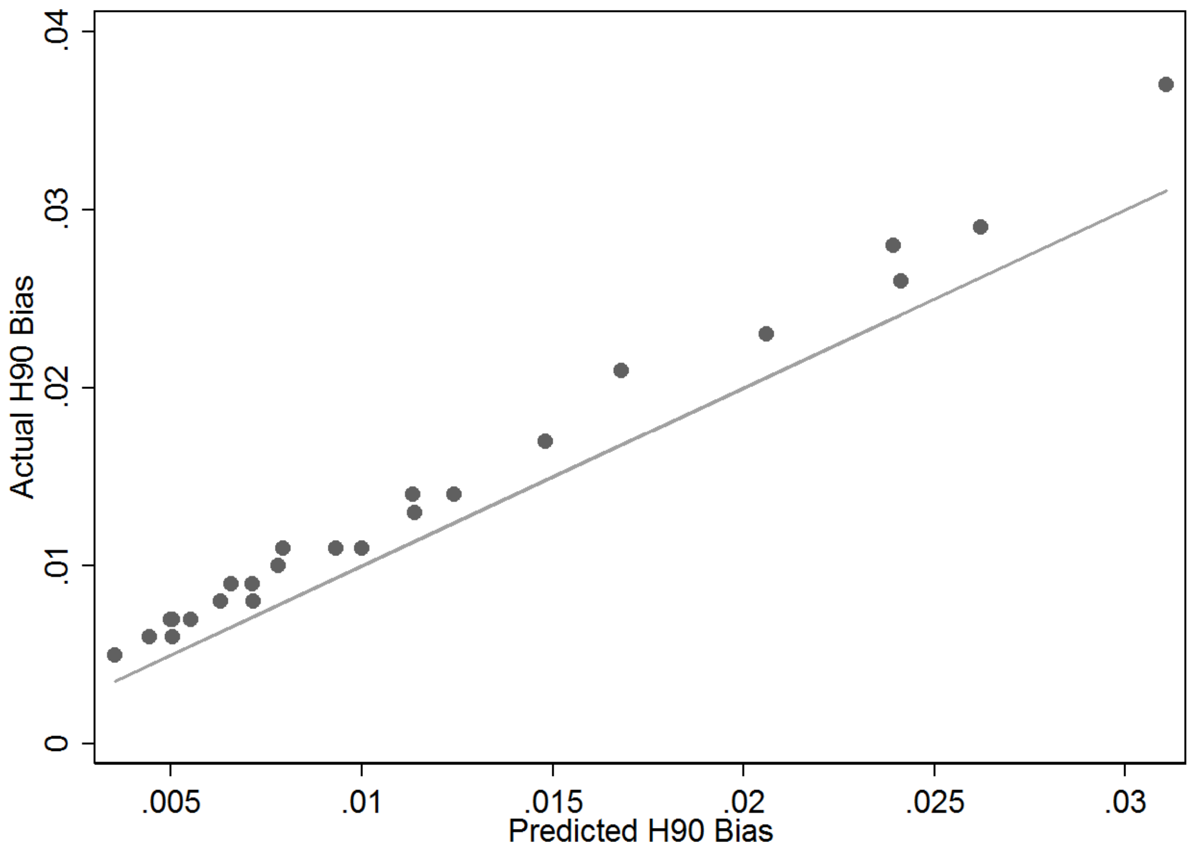

Figure 6. Relation between predicted bias in estimated H90 and actual bias from simulations using cities and sampling rates considered in Table 4; includes 45 degree line that would obtain if predicted and actual were equal

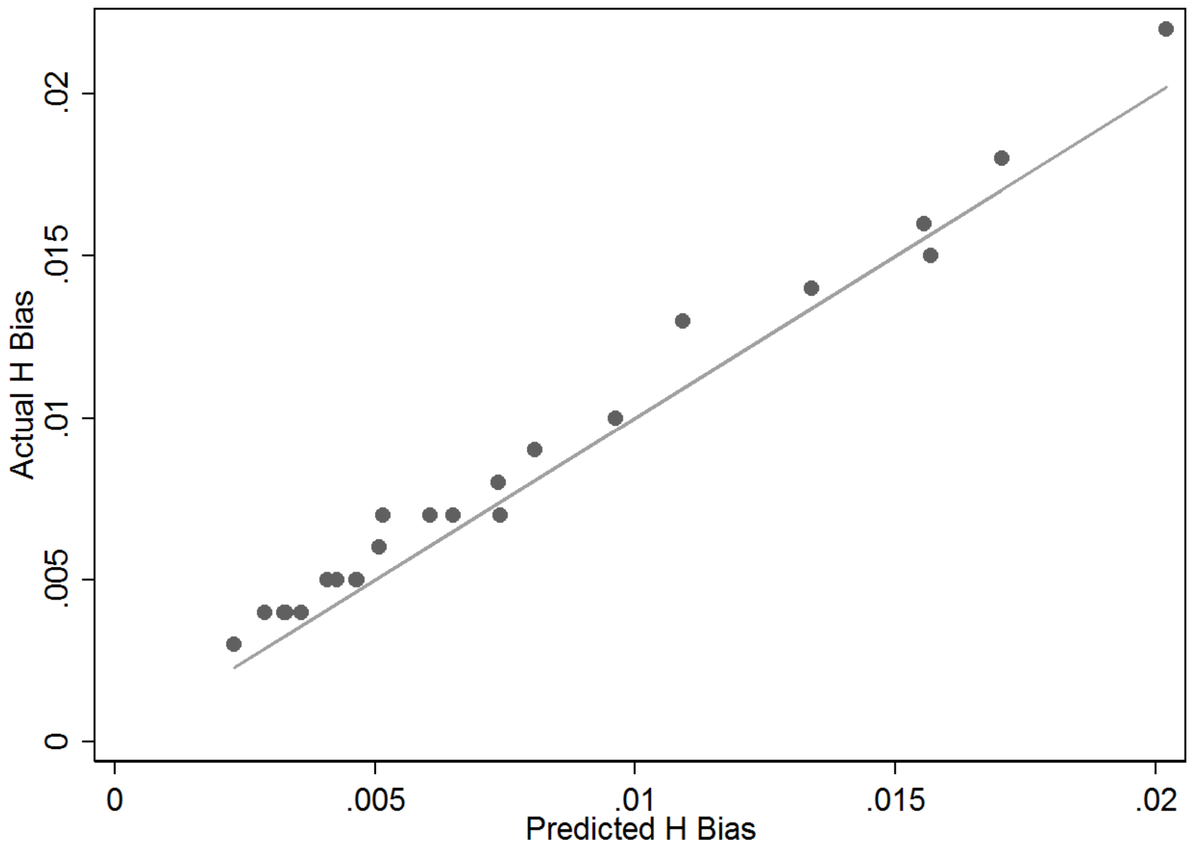

Figure 7. Relation between predicted bias in estimated H and actual bias from simulations using the cities and sampling rates considered in Table 2; includes 45 degree line that would obtain if predicted and actual were equal