Supplementary Material to **Bank Branch Access: Evidence from Geolocation Data**[*]

Jung Sakong[†] Alexander K. Zentefis[‡]

April 7, 2023

[*]These views expressed in this paper are those of the authors and do not reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

[†]Federal Reserve Bank of Chicago (email: jung.sakong@chi.frb.org)
[‡]Yale School of Management (email: alexander.zentefis@yale.edu)

G Bank Branch Segregation

In this section, we examine the extent to which different groups choose different *menus* of branches. In other words, do Black, Hispanic, and White branch visitors sort into distinct branches or do they commingle at the same branches? Likewise, do high-income and low-income bran goers separate in the branches they visit? A natural way to investigate these questions is to estimate measures of segregation among bank branch visitors.

The topic of ethnic and racial segregation began absorbing the energies of researchers decades ago. Over the intervening years, a sweeping library of articles has emerged, seeking to measure the amount of segregation and to estimate its consequences for human welfare.[1] For the most part, the literature has focused on residential or school segregation. We present new segregation estimates among visitors to bank branches across the United States. By evaluating the extent to which people sort ethnically, racially, or by income in their routine visits to banks, our work here is similar to research that estimates segregation not according to neighborhoods, but activity in daily life (e.g., Davis et al. 2019; Athey et al. 2021).

Examining segregation among bank branch visitors is important for multiple reasons. First, branch visits engender chance encounters with others, and contacting dissimilar people over the course of the day enriches the human experience and promotes progress (see Sunstein 2001 for a forceful argument of this thesis). Second, bank branches are heterogeneous from many aspects, such as in their product menus, interest rates, and promotions; staff quality; and loan approval proclivity. Populations that stay separate in their branch visits might mean some groups are deprived of valuable offerings available to others. Third, bank branch visits involve personal savings and investments, and effects from branch heterogeneity can compound over time and contribute to long run wealth inequality.

Because we do not know the demographic attributes of an individual branch visitor—instead, assigning characteristics based on each visitor's home Census block group—our measures of segregation are slightly different in concept from standard segregation estimates that have access to individual attributes. With this caveat, Table C.1 presents several segregation measures at the national level. Our three main segregation measures are (i) racial dissimilarity, (ii) racial entropy, and (iii) income entropy.

G.1 Racial Dissimilarity Index

We begin by estimating racial segregation using the dissimilarity index developed by Jahn, Schmid, and Schrag (1947), which measures the differential distribution of a population. A minority group is considered segregated according to this measure if the group is unevenly separated over spatial areas (Massey and Denton 1988). Elaborating on this index, we may suppose an area is partitioned into N sections. Following Echenique and Fryer Jr. (2007), the dissimilarity index between Black residents and non-Black residents in the area is

$$\text{Dissimilarity Index} = \frac{1}{2}\sum_{i=1}^{N}\left|\frac{\text{Black}_i}{\text{Black}_{\text{total}}} - \frac{\text{Non-Black}_i}{\text{Non-Black}_{\text{total}}}\right|, \tag{C.74}$$

where Black_i is the number of Black residents in section i, $\text{Black}_{\text{total}}$ is the total number of Black residents in the area, Non-Black_i is the number of non-Black residents in the section, and $\text{Non-Black}_{\text{total}}$ is the total number of non-Black residents in the area.

Conceptually, the dissimilarity index measures the fraction of a group's population that would need to change sections for each section's fraction of that group to match the group's overall share in the area. In our application, a section is a discrete bank branch, and we measure the dissimilarity index at the national level. Our dissimilarity index value is thus the fraction of bank branch visitors who are Black that would need to visit a different branch so that each branch would have the same fraction of Black visitors as the overall share of Black visitors to banks in the country. The measure ranges from 0 to 1 and reaches the highest value (maximal segregation) if no bank branch had both Black and non-Black visitors.

We evaluate the racial dissimilarity index in Eq. (C.74) for bank branch visitors by estimating each component. Let N be the total number of branches in the country in a year-month, ignoring any time notation for simplicity. The value $\widehat{\text{Black}}_i$ is an estimate of the expected number of branch i's visitors who are Black. We calculate this value by (i)

[1] Too many papers exist on segregation and its ramifications to give proper credit to all. Just a few examples include early work by Duncan and Duncan (1955); Kain (1968); Wilson (1987); Case and Katz (1991); Cutler and Glaeser (1997); later papers by Echenique and Fryer Jr. (2007); Iceland and Scopilliti (2008); Card, Mas, and Rothstein (2008); Ananat (2011); Billings, Deming, and Rockoff (2014); and recent papers by Logan and Parman (2017); Fogli and Guerrieri (2019); Akbar, Li, Shertzer, and Walsh (2020); Cook, Jones, Rosé, and Logan (2020); Logan, Foster, Xu, and Zhang (2020).

multiplying the visitor count from each home Census block group with travelers to the branch by the block group's Black population share from the 2019 5-yr. ACS, and (ii) summing these block-group-visitor-count × Black-share products together. In symbols, let $n_{j,i}$ denote the number of visitors from block group j to branch i, and let π_j denote the Black population share of block group j. The estimate

$$\widehat{\text{Black}}_i = \sum_j n_{j,i}\pi_j. \tag{C.75}$$

The value $\widehat{\text{Black}}_{\text{total}}$ is an estimate of the expected total number of Black visitors to banks in the country. We compute this estimate as follows. Relying on the notation established, let $N_i = \sum_j n_{j,i}$ be the total number of visitors (whose home block group we know) who visit branch i. Let $\hat{\Pi}_i$ denote the estimated expected share of branch i's visitors who are Black. This share is computed as

$$\hat{\Pi}_i = \sum_j \left(\frac{n_{j,i}}{N_i}\right)\pi_j. \tag{C.76}$$

The estimate of the expected total number of Black branch goers in the country is

$$\widehat{\text{Black}}_{\text{total}} = \sum_i N_i\hat{\Pi}_i. \tag{C.77}$$

The estimates $\widehat{\text{Non-Black}}_i$ and $\widehat{\text{Non-Black}}_{\text{total}}$ are computed identically as their counterparts, but with the Black population share replaced with the non-Black population share from the 2019 5-year ACS. The national dissimilarity index estimate considers all branches in our core sample. In the calculation, visitor home Census block groups with zero population according to the 2019 5-year ACS are dropped from the calculation. The national index is computed month-by-month, and the number in Table C.1 is a simple average over the core sample period. The monthly estimates are quite stable, and they are provided in Online Table C.3.

The national estimated Black/non-Black dissimilarity index is 0.447. In Table C.1, we also provide comparison estimates of Black/non-Black dissimilarity from several other research papers across different settings. Bank branch dissimilarity is lower than residential dissimilarity as estimated by Massey and Denton (1988) (0.597), Cutler and Glaeser (1997) (0.586), and Iceland and Scopilliti (2008) (0.674). The spatial unit for these other dissimilarity estimates is a Census tract. Cutler and Glaeser (1997) report an average measure that spans 209 MSAs with at least 100,000 total residents and at least 10,000 Black residents as of the 1990 Census. Iceland and Scopilliti (2008) provide a population-weighted average of the dissimilarity index across 84 Metropolitan Areas (MAs) that contained at least 1,000 Black residents, and the authors' estimate is derived from the 2000 Census. Massey and Denton (1988) supply a population-weighted mean across the 60 largest MSAs as of the 1982 Census. Their measure combines dissimilarity estimates for Hispanics, Blacks, and Asians, using non-Hispanic Whites as the comparison racial group in each case. Although their estimate is not for a strictly Black/non-Black index, we include it as comparison because of the paper's ubiquity in the segregation literature.

Davis et al. (2019) present a measure of dissimilarity in urban consumption. The spatial unit of analysis is a restaurant venue in New York City, and the authors use Yelp reviews between 2005 and 2011 to infer restaurant trips. A discrete choice model is used to produce the measure of consumption segregation. The value reported in the table is the authors' model-based estimate when all factors entering a consumer's choice are operational. Urban consumption dissimilarity by their estimate of 0.352 is moderately lower than our estimate of banking dissimilarity. Moving to school segregation, we report dissimilarity estimates from Clotfelter (1999) and Billings et al. (2014), who both use as their spatial units a public school within a district. Examining K-12 schooling across school districts in Washington, D.C. during the 1994-1995 school year, Clotfelter (1999) presents an estimated dissimilarity value of 0.550, which is slightly higher than our national estimate of banking dissimilarity. One caveat here is that Clotfelter (1999) uses Whites and non-Whites as the two racial groups. Finally, Billings et al. (2014)'s measure of dissimilarity in K-5 schooling across the state of North Carolina of 0.300 is mildly lower than our estimate of banking dissimilarity. Their sample covers the period 2008-2012, it includes 115 public school districts, and the estimate reported in the table is the unweighted sample mean across districts.

G.2 Racial Entropy Index

The dissimilarity index is disadvantaged by restricting analysis to just two groups. An alternative segregation index, the information entropy (H) index introduced in Theil (1972), measures segregation among multiple groups. Like the dissimilarity index, the entropy index measures "evenness," or the extent to which groups are evenly distributed among spatial areas (Iceland 2004b). Entropy in this context is a measure of racial/ethnic diversity, and it is greatest when each group is equally represented in the area. We compute the entropy index considering four mutually exclusive and exhaustive racial/ethnic groups: Hispanics, non-Hispanic Whites, non-Hispanic Blacks, and others.

Suppose again that the country has N bank branches in a period. Let π_s denote the fraction of total bank branch visitors in the country who belong to group s. The entropy of the groups of branch visitors across the country is $E = \sum \pi_s \ln\left(\frac{1}{\pi_s}\right)$. Similarly, the entropy of the groups of visitors to bank branch i is $E_i = \sum \pi_{s,i} \ln\left(\frac{1}{\pi_{s,i}}\right)$, where $\pi_{s,i}$ is the fraction of branch i's visitors who belong to group s.[2]

Following Reardon and Firebaugh (2002), we write the entropy segregation index as

$$\text{Entropy Index} = \sum_{i=1}^{N} \frac{\text{visitors}_i}{\text{visitors}_{\text{total}}} \left(1 - \frac{E_i}{E}\right), \tag{C.78}$$

where visitors_i denotes the number of visitors to branch i and $\text{visitors}_{\text{total}}$ denotes the total number of visitors to bank branches in the country.

Conceptually, the entropy index calculates the difference in racial/ethnic diversity between sections of an area and the area as a whole. In our application, the index is maximized at $H = 1$ (where segregation is highest) when each branch observes visitors from one group only, making $E_i = 0$ for all branches. The index is minimized at $H = 0$ when each branch shares the same racial/ethnic composition as the composition of all branch visitors throughout the country, so that $E_i = E$ across branches.

The only terms in Eq. (C.78) that require estimation are the fractions of branch visitors belonging to a group, both for individual branches ($\pi_{s,i}$) and across the country (π_s). We estimate $\pi_{s,i}$ in an identical fashion as $\hat{\Pi}_i$ in Eq. (C.76) in the previous section, which uses information about the number of visitors from different home Census block groups to branch i, the total number of visitors to the branch, and the population shares of the four racial/ethnic groups from the 2019 5-yr. ACS.[3] Each group has its own estimate, denoted $\hat{\Pi}_{s,i}$. The estimate for π_s is computed similarly as Eq. (C.77) of the previous section. Specifically, let $N = \sum_i N_i$ denote the total number of bank branch visitors in the country, where N_i is branch i's total visitors. The estimate of the share of branch visitors from each group throughout the country is

$$\hat{\Pi}_s = \sum_i \left(\frac{N_i}{N}\right) \hat{\Pi}_i. \tag{C.79}$$

In Table C.1, the national estimated racial/ethnic entropy index is 0.204. (Estimates per month over the core sample period are provided in Table C.3.) Compared to other papers, this value is lower than residential segregation measures based on racial entropy. Massey and Denton (1988)'s estimate of 0.267 is computed over slightly different racial groups than ours (Hispanics, Blacks, Asians, and non-Hispanic Whites). Iceland (2004a)'s estimate is 0.247. He calculates the measure with 2000 Census data and uses six racial categories: non-Hispanic Whites, non-Hispanic African Americans, non-Hispanic Asians and Pacific Islanders, non-Hispanic American Indians and Alaska Natives, non-Hispanics of other races, and Hispanics. Like Massey and Denton (1988), Iceland's spatial unit is a Census tract, but he spans 325 MAs in the United States. Finally, moving to public schooling, we report the entropy-based racial segregation estimate from Frankel and Volij (2011) for K-12 public schools during the 2007-2008 school year. Their racial groups are Asians, non-Hispanic Whites, non-Hispanic Blacks, and Hispanics, and they include all U.S. public schools that report a positive number of students in the Common Core of Data. Frankel and Volij (2011)'s segregation estimate of 0.422 is substantially higher than both our estimate of bank branch segregation and the other entropy-based residential segregation estimates. Overall, both the racial dissimilarity index and the racial entropy index estimates imply lower levels of racial/ethnic segregation that people experience at bank branches than they do in housing or in schools.

G.3 Income Entropy Index

An entropy-based measure can be used to examine income segregation among bank branch visitors as well, where we turn next. We adopt the rank-order income segregation measure from Reardon (2011), which accounts for the

[2] Note that if a group does not visit an individual branch at all (i.e., $\pi_{s,i} = 0$), the group's value in the entropy formula is evaluated as $0 \cdot \ln\left(\frac{1}{0}\right) = \lim_{\pi \to 0}\left(\pi \ln\left(\frac{1}{\pi}\right)\right) = 0$. In addition, it clearly is assumed that some racial/ethnic heterogeneity exists among branch visitors in the country so that $E \neq 0$.

[3] As before with the dissimilarity index, visitor home Census block groups with zero population according to the 2019 5-year ACS are dropped from the calculation.

natural numeric ordering of income. In our application, this measure estimates the extent to which households of different incomes are evenly distributed during their branch trips throughout the country. The measure is independent of the degree of income inequality in the population. The income segregation index is highest at 1 when, within each branch, all visitors have identical incomes. It is lowest at 0 when the income distribution of visitors at each branch matches the overall income distribution of branch visitors across the country.

Constructing the index starts by calculating the segregation of visitors at each branch using a two-group entropy index. The two groups are visitors with incomes below the p-th percentile of the income distribution and visitors with incomes above the p-th percentile. The entropy of the two income groups is $E(p) = p \ln \frac{1}{p} + (1-p) \ln \frac{1}{1-p}$, and the pairwise segregation measure $H(p)$ of the two income groups is determined using the formula in Eq. (C.78) from before. Pairwise segregation measures can extend to comparing the remaining percentiles of the income distribution to form the income segregation index. With this in mind, one defines the income segregation index as

$$\text{Income Segregation Index} = 2 \ln(2) \int_0^1 E(p) H(p) \, dp. \tag{C.80}$$

Conceptually, the income segregation index is a weighted average of the pairwise segregation measures $H(p)$ across all percentiles p, with greater weight assigned to the middle of the income distribution, where entropy $E(p)$ is highest and where two randomly drawn branch visitors are more likely to have their incomes positioned. We compute Eq. (C.80) using income data from the 2019 5-year ACS, which provides 16 binned categories. We estimate $H(p)$ at each of the thresholds using the procedure described in Reardon (2011), and we replace the racial/ethnic population shares from the ACS used in the previous section with the population income shares. Branch visitor home Census block groups with zero population according to the 2019 5-year ACS are dropped from the calculation. We provide a step-by-step guide in Appendix H of the Supplementary Material.

In Table C.1, the national estimated income entropy index is 0.059. (Estimates per month over the core sample period are provided in Table C.3.) Our estimate is lower than other measures of income segregation in the literature. Using Census tracts as their spatial unit of analysis in computing income entropy based on residence, Reardon and Bischoff (2011) report a value of 0.157; Bischoff and Reardon (2014), a value of 0.148; and Reardon, Bischoff, Owens, and Townsend (2018), a value of 0.115. All three papers use family instead of household income. Reardon and Bischoff (2011)'s estimate spans the 100 largest MAs as of the 2000 Census; Bischoff and Reardon (2014)'s, the 117 largest MAs according to the 2011 5-year ACS; and Reardon et al. (2018)'s, the 116 largest MAs according to the 2016 5-year ACS. The value from Reardon et al. (2018) reported in the table is the measure of income entropy-based segregation that attempts to correct for sampling bias. Finally, Owens, Reardon, and Jencks (2016) estimates income segregation among families with children in K-12 public schools across the 100 largest MAs. Relying on the 2012 5-year ACS, they estimate the average family income segregation between school districts to be 0.089, still higher than our national estimate of household income segregation among branch visitors.

G.4 Geography of Bank Branch Segregation

In this section, we draw attention to spatial variation in bank branch segregation. We focus on both the racial and income entropy segregation measures, and we compute them at the county level in the same manner described in Appendix G.2 and Appendix G.3 of the Supplementary Material. Bank branches are assigned to counties according to their locations in SafeGraph. We again calculate segregation indices month-by-month, but now, to aggregate across time, we weight each year-month by its total branch visitors whose home Census block group we know. We do this to account for the noticeable variation in visitor counts through time in the smaller-population counties.[4]

Fig. C.1, Panel A presents a heatmap of income segregation estimates by county, whereas Panel B presents a heatmap of racial segregation by county. Counties colored darker in the greenscale are estimated as more segregated in their branch visitors.[5]

[4]The entropy-based measure of racial segregation is highly correlated with the dissimilarity measure at the county level. For our core sample of bank branches, that correlation is 75.72%.

[5]Visitor home Census block groups with zero population according to the 2019 5-year ACS are dropped from the county calculations. Counties with less than 2 branches in each month, for which we cannot compute a segregation index, and counties without 24 months of visitors in the core sample (Jan. 2018 - Dec. 2019), for which we have inadequate data to estimate segregation, are shaded white in the figures. Our two filters remove 983 counties. Of the 33.5 million total branch visitors over the sample period for whom we have home Census block group information, dropping these counties omits 500 thousand visitors (around 1.5%). The minimum visitor count per month across counties under these filters is 509.

Three spatial patterns are visible from the figure. First, racial and income segregation in banking are positively correlated. Areas of the country where racial segregation is high also tend to observe high income segregation. The correlation between the two segregation measures is 72.78%. Table C.2 presents the top-50 U.S. counties ranked by income and racial segregation, which displays the positive relation. For example, Essex County, NJ ranks first in income segregation and fourth in racial segregation. Wayne County, MI is fifth in income segregation and eighth in racial segregation.

Second, segregation varies substantially across regions of the country. Both segregation measures are highest in the Northeast, the Midwest (east of the Great Plains), the Southwest, and the Pacific Coast.[6] The South and the Mountain West observe lower bank branch segregation. The Great Plains broadly lacks sufficient data to make reliable segregation estimates. There is substantial within-region variation as well. Weighted county-level regressions of segregation on state fixed effects estimate that 28 percent of cross-county variance in racial segregation and 18 percent of income segregation cross-county variance is within states. Similar analysis using the four Census regions shows that 14.6 percent of the cross-county variance in racial segregation and 7.11 percent of cross-county variance in income segregation is within regions.

Third, major urban cores see the highest segregation. Returning to the previous two examples, Essex County, NJ contains Newark and Wayne County, MI contains Detroit. Cook County, IL, which contains Chicago, ranks highly, as does St. Louis County, which borders the city of St. Louis. Even in the South, where bank branch segregation is generally lowest, high segregation pockets are seen in big cities like Atlanta, Houston, Jackson, and Miami. Fig. C.4 presents binned scatter plots of the segregation estimates by counties' urban area shares, along with best-fit lines from OLS regressions. Nearly 40% of the variation in income segregation and 20% of the variation in racial segregation across counties can be explained by the urban share. The estimated coefficient of 0.047 for the income segregation regression is also roughly the same as the 10 to 90 percentile range of income segregation values across all counties. Hence, extrapolation of the coefficient implies that a county that switches from fully rural to fully urban jumps from the left to the right side of the distribution of income segregation. Similarly, the estimated coefficient of 0.076 for the racial segregation regression is just short of the 10 to 90 percentile range of racial segregation values across all counties. Fig. C.5 compares segregation values by RUCA classifications. Presented are coefficients from county-level OLS regressions of the income and racial segregation estimates on county population shares that reside in each area type. Both racial and income bank segregation increases the most when transitioning into a Metropolitan core, with the change more than doubling the effects from switching into a Metropolitan suburb, a Micropolitan/Small town core, or a Rural area.

H Income Segregation Computational Steps

This section presents the steps to compute the income entropy segregation indices of Appendix G.3. The steps follow closely with those outlined in Reardon (2011), but they are applied to our banking context. The formula for income segregation IS we want to estimate is

$$IS = 2\ln 2 \int_0^1 E(p)\, H(p)\, dp, \tag{C.81}$$

where p is percentile and $E(p)$ is the entropy of the percentile:

$$E(p) = p \ln\left(\frac{1}{p}\right) + (1-p)\ln\left(\frac{1}{1-p}\right). \tag{C.82}$$

H.1 Preliminaries

There are 16 household income ranges registered in the 2019 5-year ACS, which implies that there are $K = 16$ ranges of income. Call an example range $k \in \{1, \ldots, 16\}$. For instance, $k = 1$ is $< \$10,000$, $k = 2$ is $\$10,000 - \$15,000$, and $k = K$ is $> \$200,000$.

We use the $k \in \{1, 2, ..., K-1\}$ ranges, and the last k that we use is $k = K - 1 = \$150,000 - \$200,000$. We do not use the range $k = K$ ($> \$200,000$) because we already know its percentile, which is equal to 1.

[6] Two counties stand out in the Southwest: Apache County and Navajo County in Arizona. Both counties are home to large Indian Reservations. Based on the 2010 Census, the Native American population share in Apache County is 72.9%, whereas the share in Navajo County is 43.4%.

The percentile p_k for $k \in [1, 2, ..., K-1]$ is the cumulative proportion of people with household income at or below the right point of the range k. For example, for $k = 1 = (< \$10,000)$, p_k is the share of households with income $< \$10,000$. For $k = 2 = \$10,000 - \$15,000$, p_k is the share of households with income $< \$15,000$ (the right point of the range), which is the sum of the shares of the first two income ranges. For $k = 15 = \$150,000 - \$200,000$, p_k is the share of households with income $< \$200,000$, which is the cumulative share of all but the last income range in the ACS.

H.2 Step 1: Calculate $E(p_k) \equiv E_k$ of all percentiles across all branches in the spatial unit (national or county)

To explain these steps, we take the spatial unit to be the entire U.S., but the same logic applies for the county analysis we present as well. We start by dropping all home block groups that have zero population according to the ACS.

Suppose the country has N branches in the period. Let p_k denote the cumulative share of total branch visitors in the country with income in the k-th income range and below. We estimate this share in the exact same manner as we explained earlier on estimating the share of all branch visitors in the country who are part of a particular race group. (See Appendix G.2.) There, we used the notation π_s for the share belonging to race group s. Here, we use p_k for the share of visitors at or below the right point of a particular ACS household income range.

Using Eq. (C.82), we compute the entropy of this percentile is

$$E(p_k) \equiv E_k = p_k \ln\left(\frac{1}{p_k}\right) + (1 - p_k) \ln\left(\frac{1}{1 - p_k}\right). \tag{C.83}$$

We calculate this entropy estimate for each of the k ranges at the national level, which delivers 15 E_k values.

H.3 Step 2: Calculate $E(p_{k,i}) \equiv E_{k,i}$ of all percentiles for each individual branch in the spatial unit

Here, we perform the same calculation of the entropy value, but at the individual branch level. We follow the same procedure as we did for racial entropy, where we used the notation $\pi_{s,i}$ (See Appendix G.2.) For example, consider branch i. The entropy of the two income-percentile-defined groups of visitors to the branch is

$$E(p_{k,i}) \equiv E_{k,i} = p_{k,i} \ln\left(\frac{1}{p_{k,i}}\right) + (1 - p_{k,i}) \ln\left(\frac{1}{1 - p_{k,i}}\right),$$

where $p_{k,i}$ is the fraction of branch i's visitors who have income at or less than threshold k. If $p_{k,i} = 0$ at a particular branch, then $E_{k,i} = 0 \ln\left(\frac{1}{0}\right) + (1 - 0) \ln\left(\frac{1}{1}\right) = 0$. These calculations produce $N \times (K - 1)$ values for $E_{k,i}$ (i.e., 15 values per branch).

H.4 Step 3: Calculate the entropy index across all branches in the spatial unit

The entropy index aggregates information across branches in the country. We calculate it for each k, hence producing 15 values. The entropy index formula is

$$\text{Entropy Index}_k \equiv H_k = \sum_{i=1}^{N} \frac{\text{visitors}_i}{\text{visitors}} \left(1 - \frac{E_{k,i}}{E_k}\right).$$

For the term visitors_i in the formula, we use the sum of visitors to branch i whose home block group we know. The term visitors in the formula is the sum of visitors_i across all branches.

Each value of H_k represents the pairwise segregation of branch visitors with income at the $100 \times p_k$-th percentile and the $100 \times (1 - p_k)$-th percentile. Fig. C.2 plots the 15 values of H_k against their corresponding percentiles for the single month of September 2019, which provides a sense of what the complete function $H(p)$ in Eq. (C.81) looks like. At least in this month, among branch visitors in the U.S., the sequence of H_k values is monotonically increasing.

H.5 Step 4: Estimate the function $H(p)$ in Eq. (C.81)

The function $H(p)$ is unknown, but it can be estimated using the $K-1$ (i.e., 15) values $H(p_k) \equiv H_k$ that can be measured. The intuition for this process is that the collection of H_k points, when plotted against their corresponding p_k points as in Fig. C.2, produces a function that can be fitted with a polynomial of some order $M \leq K-2 = 14$.

We fit the polynomial using weighted least squares in which each point is weighted by E_k^2, which itself is taken from Eq. (C.83). Weighting the regression by the square of the entropy value minimizes the weighted squared errors and ensures that the fitted polynomial will fit best for p_k near ½, where H_k is weighted most.

The choice of polynomial order is at the discretion of the researcher, and it should balance parsimony and precision. To select an appropriate order, we estimated the country-wide income segregation index for the month of September 2019 using polynomial orders 1-8. We then plot the 95% confidence intervals around each point estimate. (Obtaining the standard error of the estimate is described below). The plot is provided in Fig. C.3. The standard errors shrink significantly and the estimates stabilize beginning with polynomial order 4. For that reason, we use this polynomial order in our estimation.

To fit the values H_k, we run a single WLS regression:

$$H_k = \beta_0 + \beta_1 p_k + \beta_2 p_k^2 + \beta_3 p_k^3 + \ldots + \beta_M p_5^5 + e_k,$$

where, we weight the points by E_k^2. Let the vector of coefficients be denoted $B = \left(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_4, \hat{\beta}_5\right)'$ and let the variance-covariance of the estimated coefficients be denoted V.

H.6 Step 5: Compute the estimated Income Segregation Index $\hat{IS}$

Finally, the estimate for income segregation, denoted $\hat{IS}$, is computed as

$$\hat{IS} = \Delta \cdot B,$$

which is the dot product between the vector of coefficients from the WLS regression and a vector of parameters $\Delta = (\delta_1, \delta_2, \ldots, \delta_M)$ provided in Reardon (2011). He shows that for income entropy, the parameters δ_m can be evaluated as

$$\delta_m = \frac{2}{(2+m)^2} + 2\sum_{n=0}^{m} \frac{(-1)^{m-n}\binom{m}{n}}{(m-n+2)^2}, \tag{C.84}$$

where $\binom{m}{n} = \frac{m!}{n!(m-n)!}$ is the combinatorial function. The number m is the chosen polynomial order, which in our case is 4.

The 5 values for δ_m that we require are $\left(1, \frac{1}{2}, \frac{11}{36}, \frac{5}{24}, \frac{137}{900}\right)$. The measure of uncertainty about the estimated income segregation is $\text{Var}\left(\hat{IS}\right) = \Delta' V \Delta$, which we use to compute the 95% confidence intervals in Fig. C.3.

Supplementary Material References

Akbar, P. A., S. Li, A. Shertzer, and R. P. Walsh (2020): "Racial segregation in housing markets and the erosion of black wealth," Working paper no. 25805. National Bureau of Economic Research, Cambridge, MA.

Ananat, E. O. (2011): "The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality," *American Economic Journal: Applied Economics*, 3, 34–66.

Athey, S., B. Ferguson, M. Gentzkow, and T. Schmidt (2021): "Estimating experienced racial segregation in US cities using large-scale GPS data," *Proceedings of the National Academy of Sciences*, 118, 1–9.

Billings, S. B., D. J. Deming, and J. Rockoff (2014): "School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg," *Quarterly Journal of Economics*, 129, 435–476.

Bischoff, K. and S. F. Reardon (2014): "Residential segregation by income, 1970–2009," *Diversity and disparities: America enters a new century*, 43.

Card, D., A. Mas, and J. Rothstein (2008): "Tipping and the dynamics of segregation," *Quarterly Journal of Economics*, 123, 177–218.

Case, A. C. and L. F. Katz (1991): "The company you keep: The effects of family and neighborhood on disadvantaged youths," Working paper no. 3705. National Bureau of Economic Research, Cambridge, MA.

Clotfelter, C. T. (1999): "Public school segregation in metropolitan areas," *Land Economics*, 487–504.

Cook, L. D., M. E. Jones, D. Rosé, and T. D. Logan (2020): "The Green Books and the geography of segregation in public accommodations," Working paper no. 26819. National Bureau of Economic Research, Cambridge, MA.

Cutler, D. M. and E. L. Glaeser (1997): "Are ghettos good or bad?" *Quarterly Journal of Economics*, 112, 827–872.

Davis, D. R., J. I. Dingel, J. Monras, and E. Morales (2019): "How segregated is urban consumption?" *Journal of Political Economy*, 127, 1684–1738.

Duncan, O. D. and B. Duncan (1955): "A methodological analysis of segregation indexes," *American Sociological Review*, 20, 210–217.

Echenique, F. and R. G. Fryer Jr. (2007): "A measure of segregation based on social interactions," *Quarterly Journal of Economics*, 122, 441–485.

Fogli, A. and V. Guerrieri (2019): "The end of the American Dream? Inequality and segregation in US cities," Working paper no. 26143. National Bureau of Economic Research, Cambridge, MA.

Frankel, D. M. and O. Volij (2011): "Measuring school segregation," *Journal of Economic Theory*, 146, 1–38.

Iceland, J. (2004a): "Beyond black and white: Metropolitan residential segregation in multi-ethnic America," *Social Science Research*, 33, 248–271.

——— (2004b): "The multigroup entropy index (also known as Theil's H or the information theory index)," *US Census Bureau*, 31, 10.

Iceland, J. and M. Scopilliti (2008): "Immigrant residential segregation in US metropolitan areas, 1990–2000," *Demography*, 45, 79–94.

Jahn, J., C. F. Schmid, and C. Schrag (1947): "The measurement of ecological segregation," *American Sociological Review*, 12, 293–303.

Kain, J. F. (1968): "Housing segregation, negro employment, and metropolitan decentralization," *Quarterly Journal of Economics*, 82, 175–197.

Logan, J. R., A. Foster, H. Xu, and W. Zhang (2020): "Income Segregation: Up or down, and for whom?" *Demography*, 57, 1951–1974.

Logan, T. D. and J. M. Parman (2017): "Segregation and homeownership in the early twentieth century," *American Economic Review*, 107, 410–14.

Macartney, H. and J. D. Singleton (2018): "School boards and student segregation," *Journal of Public Economics*, 164, 165–182.

Massey, D. S. and N. A. Denton (1988): "The dimensions of residential segregation," *Social Forces*, 67, 281–315.

Owens, A., S. F. Reardon, and C. Jencks (2016): "Income segregation between schools and school districts," *American Educational Research Journal*, 53, 1159–1197.

Reardon, S. F. (2011): "Measures of income segregation," Working Paper. Stanford Center for Education Policy Analysis, Stanford, CA.

Reardon, S. F. and K. Bischoff (2011): "Income inequality and income segregation," *American Journal of Sociology*, 116, 1092–1153.

Reardon, S. F., K. Bischoff, A. Owens, and J. B. Townsend (2018): "Has income segregation really increased? Bias and bias correction in sample-based segregation estimates," *Demography*, 55, 2129–2160.

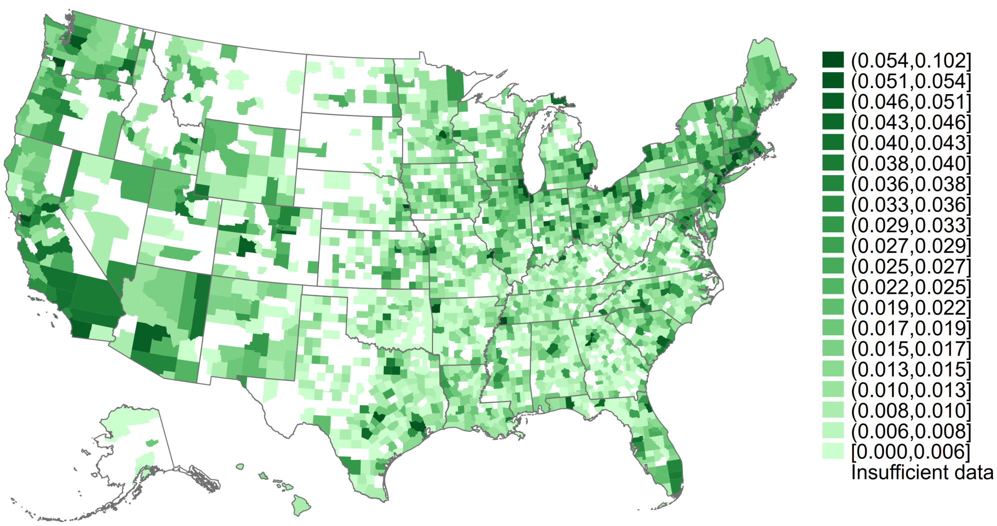
Reardon, S. F. and G. Firebaugh (2002): "Measures of multigroup segregation," *Sociological methodology*, 32, 33–67.

Sunstein, C. (2001): *Republic.com*, Princeton University Press.

Theil, H. (1972): *Statistical decomposition analysis: With applications in the social and administrative sciences*, vol. 14, North-Holland Publishing Company.

Wilson, W. J. (1987): *The truly disadvantaged: The inner city, the underclass, and public policy*, University of Chicago Press.

(A) Income Segregation by County



(B) Racial Segregation by County

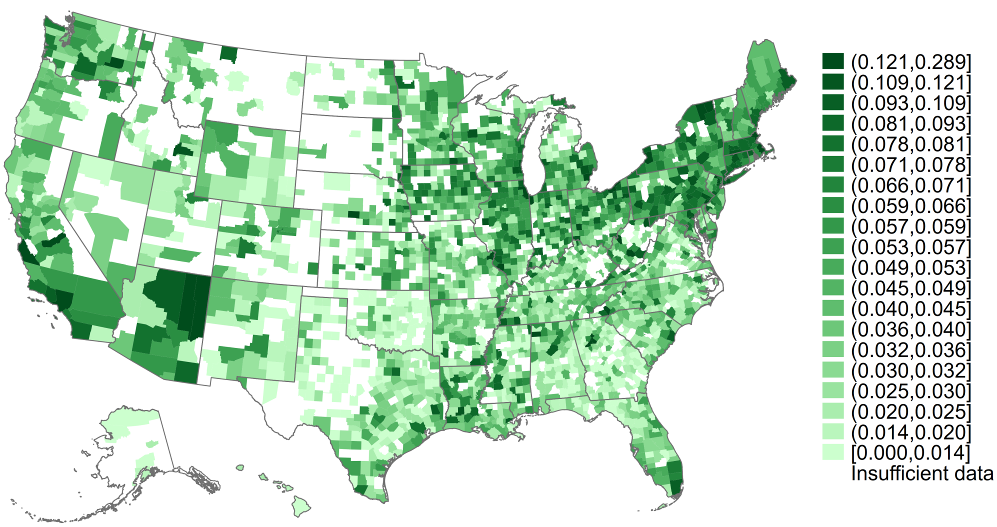


FIGURE C.1

GEOGRAPHY OF BANK BRANCH SEGREGATION

The figure presents heatmaps of income and racial segregation at U.S. bank branches, where segregation is measured by the entropy index per county. The figure is based on our core SafeGraph sample of bank branches; i.e., only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits. The income entropy segregation index values portrayed in Panel A are estimates of Eq. (C.80), made using the procedure described in Reardon (2011). The racial entropy segregation index values portrayed in Panel B are estimates of Eq. (C.78). Branches are assigned to counties based on their locations in SafeGraph. Visitor home Census block groups with zero population according to the 2019 5-year American Community Survey are dropped from the calculations. Values are calculated month-by-month for each county, and the figure presents weighted monthly averages, where each month's weight is its share of total visitors (whose home block groups we know) to branches in the county over the core sample period (January 2018 - December 2019). The maps are constructed by grouping counties into 20 vigintiles and shading the areas so that darker tints in the greenscale imply higher segregation index values. Counties with less than 2 branches in each month, for which we cannot compute a meaningful segregation index, and counties without 24 months of visitors in the core sample (Jan. 2018 - Dec. 2019), for which we have inadequate data to estimate segregation, are shaded white.

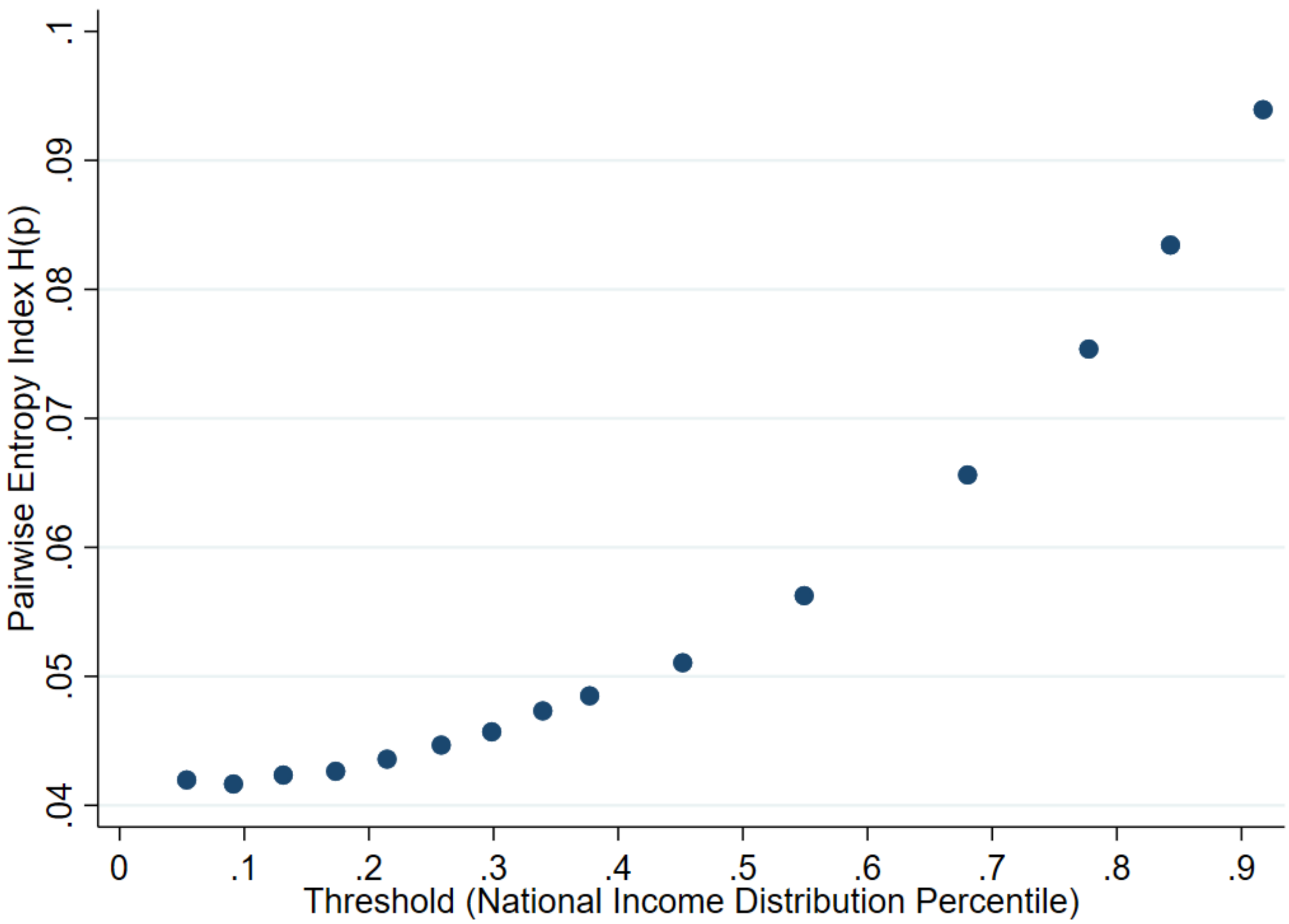


Figure C.2
Pairwise Income Segregation Profiles - Sept. 2019

The figure presents the pairwise household income segregation profiles (based on the entropy index) for September 2019 using our core sample. The core sample includes only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits. The pairwise income segregation profiles are the 15 values of H_k, calculated using the steps described in Supplementary Material Appendix H. Each value measures the pairwise income segregation of branch visitors with income at the $100 \times p_k$-th percentile and the $100 \times (1 - p_k)$-th percentile.

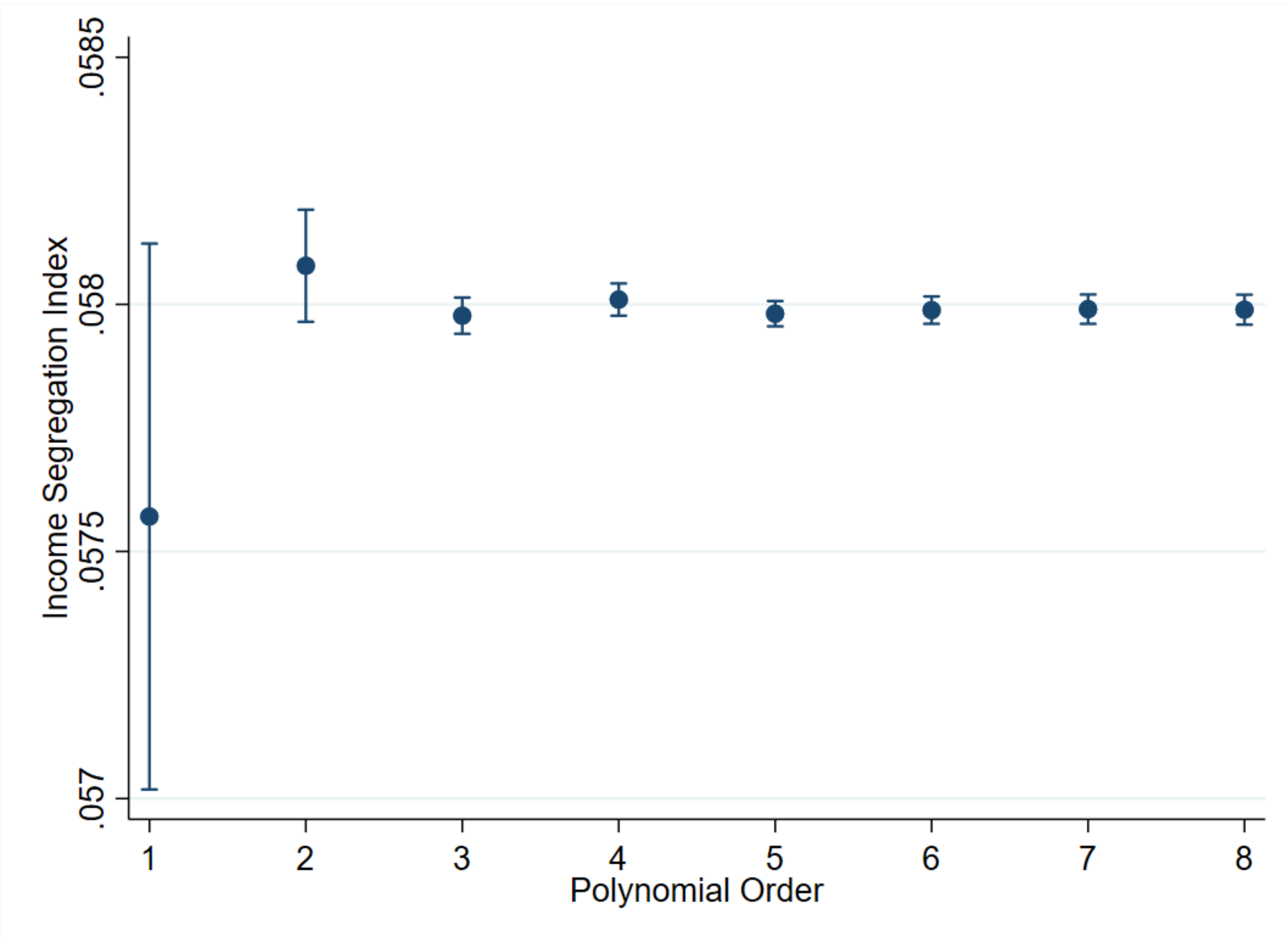


Figure C.3

Estimated Income Segregation by Polynomial Order - Sept. 2019

The figure presents national income segregation estimates and 95% confidence intervals by different polynomial orders for September 2019 using our core sample. The core sample includes only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits. The polynomial orders stand for the orders of the polynomials that fit the 15 values of pairwise income segregation H_k, which themselves are calculated using the steps described in Supplementary Material Appendix H. The method for computing the standard errors for the income segregation estimates are also described in that appendix.

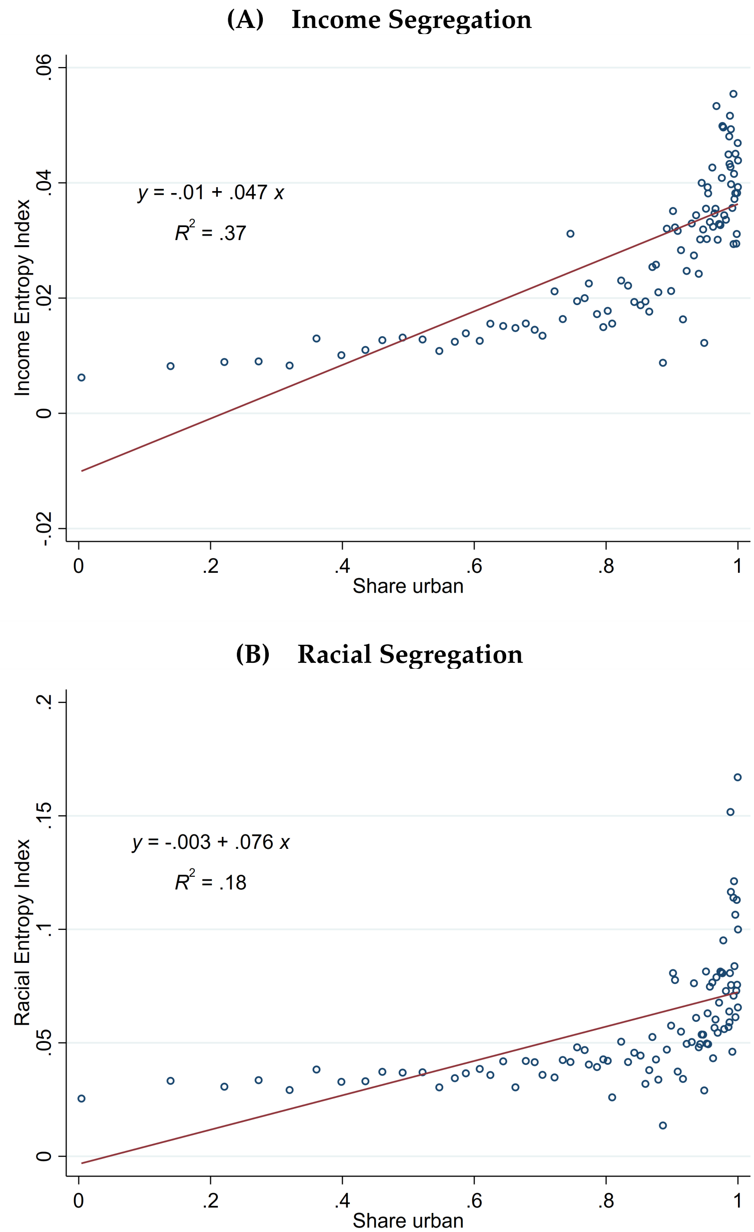


Figure C.4

Bank Branch Segregation by County's Urban Share

The figure presents binned scatter plots of within-county income and racial segregation estimates among bank branch visitors according to counties' urban area shares. Segregation estimates are based on entropy indices and are calculated using our core SafeGraph sample of bank branches; i.e., only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits. The income entropy index values are estimates of Eq. (C.80). The racial entropy index values are estimates of Eq. (C.78). Branches are assigned to counties based on their locations in SafeGraph. Visitor home Census block groups with zero population according to the 2019 5-year American Community Survey are dropped from the calculations. Values are calculated month-by-month for each county, and the segregation estimates are weighted monthly averages, where each month's weight is its share of total visitors (whose home block groups we know) to branches in the county over the core sample period (January 2018 - December 2019). Urban area shares are from the 2010 decennial Census. To construct the binned scatter plots, we divide the horizontal axes into 100 equal-sized (percentile) bins and plot the mean segregation estimate and the mean urban share within each bin. The slopes and best-fit lines are estimated using weighted OLS regressions of the county-level segregation estimates on the urban area shares. Observations are weighted by the counties' total branch visitors across the core sample period.

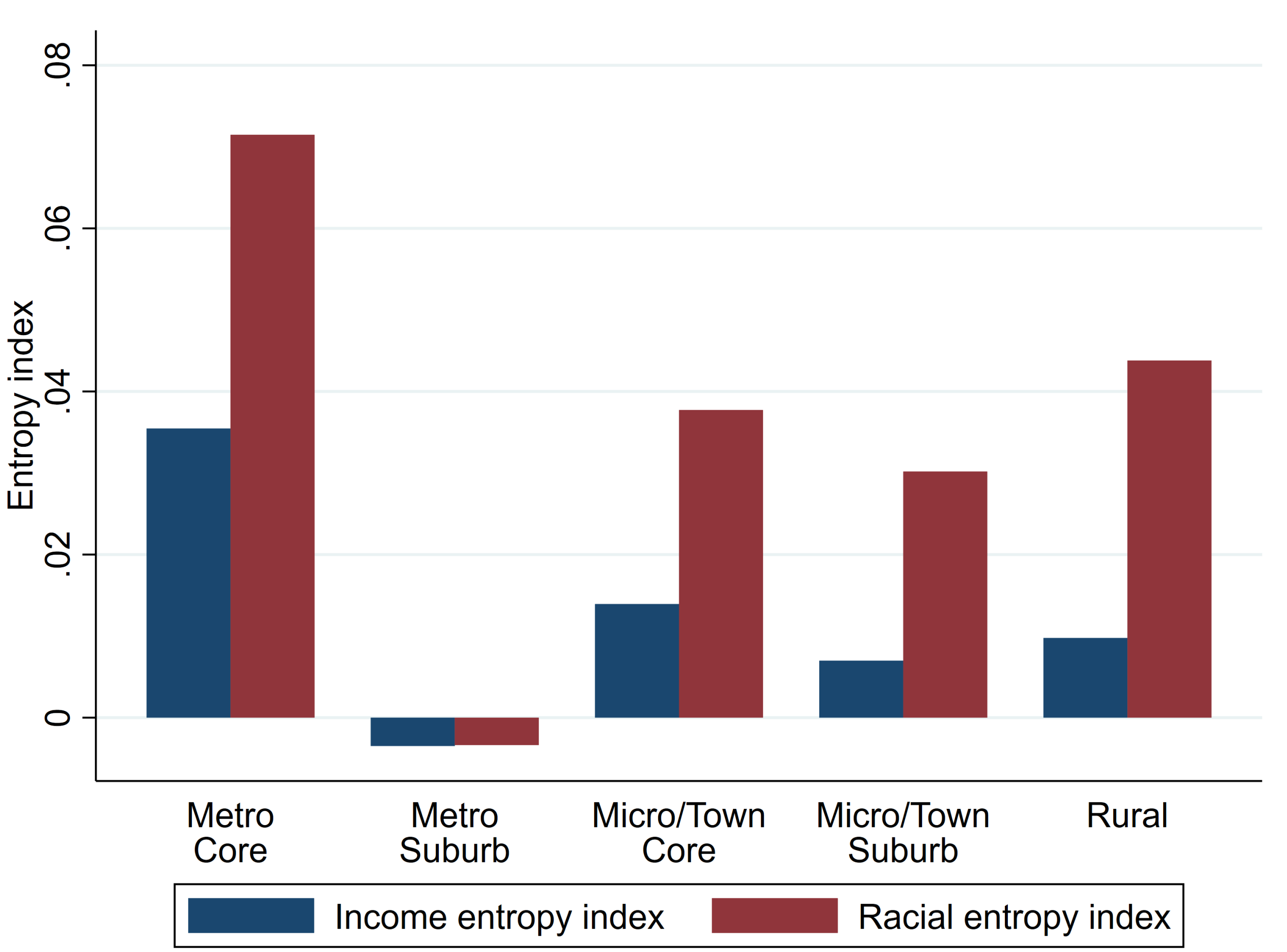


FIGURE C.5
BANK BRANCH SEGREGATION BY RUCA CLASSIFICATION

The figure presents the coefficients from two weighted OLS regressions of county-level income and racial bank branch segregation estimates on the primary Rural-Urban Commuting Area (RUCA) shares within counties. Observations are weighted by the counties' total branch visitors across the core sample period (January 2018 - December 2019). Per county, a RUCA's share is the fraction of the county's population living in the RUCA. *Metro Core* includes code 1 alone, *Metro Suburb* includes codes 2 and 3, *Micro/Town Core* includes codes 4 and 7, *Micro/Town Suburb* includes codes 5, 6, 8, and 9, and *Rural* includes code 10 alone. Segregation estimates are based on entropy indices and are calculated using our core SafeGraph sample of bank branches. The income entropy index values are estimates of Eq. (C.80). The racial entropy index values are estimates of Eq. (C.78). Branches are assigned to counties based on their locations in SafeGraph. Visitor home Census block groups with zero population according to the 2019 5-year American Community Survey are dropped from the calculations. Values are calculated month-by-month for each county, and the segregation estimates are weighted monthly averages, where each month's weight is its share of total visitors (whose home block groups we know) to branches in the county over the core sample period.

Table C.1
Bank Branch Visitor Segregation

Type	Index	Spatial Unit	Source
Racial Dissimilarity			
Banking	0.447	Branch	This paper
Residential	0.597	Census Tract	Massey and Denton (1988)
Residential	0.586	Census Tract	Cutler and Glaeser (1997)
Residential	0.674	Census Tract	Iceland and Scopilliti (2008)
Urban Consumption	0.352	Restaurant	Davis et al. (2019)
K-12 Public Schooling	0.550	School	Clotfelter (1999)
K-5 Public Schooling	0.300	School	Macartney and Singleton (2018)
Racial Entropy			
Banking	0.204	Branch	This paper
Residential	0.267	Census Tract	Massey and Denton (1988)
Residential	0.247	Census Tract	Iceland (2004a)
K-12 Public Schooling	0.422	School	Frankel and Volij (2011)
Income Entropy			
Banking	0.059	Branch	This paper
Residential	0.157	Census Tract	Reardon and Bischoff (2011)
Residential	0.148	Census Tract	Bischoff and Reardon (2014)
Residential	0.115	Census Tract	Reardon et al. (2018)
K-12 Public Schooling	0.089	School District	Owens et al. (2016)

The table reports national estimates of segregation among bank branch visitors. All values are based on our core sample of branch locations, which consists of only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits. The dissimilarity index in this paper is an estimate of Eq. (C.74), as described in Supplementary Material Appendix G.1. The two racial groups in the dissimilarity index computation are Black and non-Black. The racial entropy index is an estimate of Eq. (C.78), as described in Supplementary Material Appendix G.2. The four racial groups used in computing the racial entropy index are Hispanic, non-Hispanic White, non-Hispanic Black, and other. The income entropy index is an estimate of Eq. (C.80), as described in Supplementary Material Appendix G.3. The index comprises the fifteen income ranges provided in the 2019 5-year American Community Survey (ACS). Each bank branch segregation index is calculated using all bank branches available in our core sample. Visitor home Census block groups with zero population according to the 2019 5-year ACS are dropped from the calculations. Segregation values are calculated month-by-month, and the numbers in the table are simple averages over the core sample period (January 2018 - December 2019). Segregation index values from other research papers are organized by category in the table for comparison.

Table C.2
Top-50 Rank of US Counties by Income and Racial Segregation

	Income Segregation					Racial Segregation			
	County	State	# Visitors	Value		County	State	# Visitors	Value
1	Essex	NJ	62,988	0.103	1	Apache	AZ	1,016	0.304
2	Fulton	GA	144,629	0.073	2	St. Louis	MO	129,591	0.211
3	Union	NJ	64,363	0.072	3	Cook	IL	423,070	0.208
4	Franklin	OH	147,284	0.069	4	Essex	NJ	62,988	0.201
5	Wayne	MI	177,376	0.069	5	Fayette	WV	872	0.190
6	Westchester	NY	66,836	0.067	6	Dawson	NE	4,050	0.187
7	Cowlitz	WA	709	0.065	7	Navajo	AZ	2,398	0.187
8	Washington	AR	72,418	0.064	8	Wayne	MI	177,376	0.182
9	Cuyahoga	OH	87,139	0.062	9	Erie	NY	61,488	0.166
10	Hartford	CT	74,815	0.061	10	Fulton	GA	144,629	0.165
11	Douglas	NE	81,674	0.060	11	Kings	NY	62,034	0.159
12	St. Louis	MO	129,591	0.058	12	Cuyahoga	OH	87,139	0.158
13	Mercer	NJ	82,426	0.058	13	Madera	CA	5,984	0.150
14	Contra Costa	CA	90,859	0.058	14	Lake	IN	52,187	0.149
15	Passaic	NJ	32,739	0.058	15	Plymouth	MA	43,984	0.148
16	Lake	IL	80,174	0.057	16	Essex	MA	28,289	0.147
17	Shelby	TN	136,246	0.056	17	Franklin	NY	1,195	0.144
18	DC	DC	61,437	0.055	18	Monterey	CA	13,544	0.144
19	Cook	IL	423,070	0.054	19	Clinton	NY	1,558	0.137
20	King	WA	91,745	0.054	20	Adams	WA	621	0.136
21	Howard	MD	26,324	0.053	21	Randolph	IL	2,110	0.135
22	Bristol	MA	29,407	0.053	22	Passaic	NJ	32,739	0.132
23	Harris	TX	657,460	0.052	23	Delaware	PA	39,915	0.132
24	Travis	TX	116,400	0.052	24	Lake	OH	17,763	0.129
25	Hennepin	MN	109,782	0.052	25	DeKalb	GA	72,970	0.127
26	Geary	KS	434	0.051	26	Jackson	WV	917	0.126
27	Richmond	VA	6,645	0.051	27	Montgomery	OH	53,773	0.126
28	Dallas	TX	367,241	0.050	28	McDonough	IL	944	0.126
29	Montgomery	OH	53,773	0.050	29	Franklin	AL	5,482	0.124
30	Maricopa	AZ	446,571	0.050	30	Los Angeles	CA	607,978	0.122
31	Delaware	PA	39,915	0.050	31	Preston	WV	2,254	0.122
32	Boone	IN	5,985	0.050	32	Union	NJ	64,363	0.120
33	San Diego	CA	155,515	0.049	33	Milwaukee	WI	124,877	0.119
34	Philadelphia	PA	64,325	0.049	34	Hampden	MA	38,933	0.118
35	Fairfield	CT	68,785	0.049	35	Baltimore	MD	113,668	0.118
36	Lake	IN	52,187	0.048	36	Waukesha	WI	47,444	0.115
37	Arapahoe	CO	91,950	0.048	37	Luzerne	PA	24,962	0.115
38	Summit	OH	60,667	0.048	38	Jackson	NC	1,520	0.114
39	El Dorado	CA	8,597	0.048	39	Allegheny	PA	60,837	0.113
40	New Haven	CT	61,663	0.048	40	Hamilton	OH	80,514	0.113
41	Walton	FL	9,512	0.048	41	Philadelphia	PA	64,325	0.113
42	Jefferson	KY	120,277	0.048	42	Coconino	AZ	13,168	0.113
43	St. Johns	FL	27,653	0.047	43	Hartford	CT	74,815	0.113
44	Lorain	OH	22,580	0.047	44	Mahoning	OH	21,295	0.113
45	Berkeley	SC	10,430	0.047	45	Niagara	NY	6,886	0.112
46	Allegheny	PA	60,837	0.047	46	Queens	NY	64,630	0.112
47	Hamilton	OH	80,514	0.047	47	DC	DC	61,437	0.112
48	Baltimore	MD	38,808	0.047	48	Baltimore	MD	38,808	0.111

Table C.2 (CONTINUED)

	Income Segregation					Racial Segregation			
	County	**State**	**# Visitors**	**Value**		**County**	**State**	**# Visitors**	**Value**
49	Essex	MA	28,289	0.047	49	Oakland	MI	174,618	0.110
50	Washington	PA	5,514	0.047	50	Montgomery	PA	76,289	0.110

The table reports the top-50 U.S. counties ranked by their estimated bank branch income and racial segregation. Counties are sorted in descending order by segregation values, which are measured using entropy-based indices. The segregation values are computed over the core sample (only businesses in SafeGraph with NAICS codes equal to 522110, 522120, or 551111 for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits). Branches are assigned to counties based on their locations in SafeGraph. Segregation estimates are calculated according to the methods described in Supplementary Material Appendix G. Visitor home Census block groups with zero population according to the 2019 5-year American Community Survey are dropped from the calculations. Segregation values are calculated month-by-month for each county, and the table presents weighted monthly averages, where each month's weight is its share of total visitors (whose home block groups we know) to branches in the county over the core sample period (January 2018 - December 2019).

Table C.3
Segregation Index Estimates by Month

Year-Month	Racial Dissimilarity	Racial Entropy	Income Entropy
2018m1	0.4383	0.2022	0.0616
2018m2	0.4332	0.1990	0.0605
2018m3	0.4423	0.2033	0.0594
2018m4	0.4437	0.2060	0.0592
2018m5	0.4450	0.2052	0.0584
2018m6	0.4484	0.2040	0.0589
2018m7	0.4493	0.2034	0.0584
2018m8	0.4489	0.2030	0.0590
2018m9	0.4496	0.2051	0.0598
2018m10	0.4475	0.2047	0.0591
2018m11	0.4466	0.2040	0.0583
2018m12	0.4459	0.2015	0.0587
2019m1	0.4485	0.2046	0.0597
2019m2	0.4477	0.2071	0.0603
2019m3	0.4428	0.2027	0.0582
2019m4	0.4393	0.1988	0.0574
2019m5	0.4405	0.1989	0.0567
2019m6	0.4455	0.2001	0.0581
2019m7	0.4465	0.2012	0.0574
2019m8	0.4482	0.2011	0.0575
2019m9	0.4433	0.1990	0.0580
2019m10	0.4444	0.2042	0.0583
2019m11	0.4457	0.2065	0.0584
2019m12	0.4445	0.2031	0.0574

The table reports national estimates of segregation among bank branch visitors for each month of the core sample period. All values are based on our core sample of branch locations, which consists of only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits. The dissimilarity index in this paper is an estimate of Eq. (C.74), as described in Supplementary Material Appendix G.1. The two racial groups in the dissimilarity index computation are Black and non-Black. The racial entropy index is an estimate of Eq. (C.78), as described in Supplementary Material Appendix G.2. The four racial groups used in computing the racial entropy index are Hispanic, non-Hispanic White, non-Hispanic Black, and non-Hispanic Other Races. The income entropy index is an estimate of Eq. (C.80), as described in Supplementary Material Appendix G.3. The index comprises the fifteen income ranges provided in the 2019 5-year American Community Survey (ACS). Each bank branch segregation index is calculated using all bank branches available in our core sample. Visitor home Census block groups with zero population according to the 2019 5-year ACS are dropped from the calculations.