# Banking Across America: Distance and Branch Use*

Jung Sakong† Alexander K. Zentefis‡

September 1, 2021

**Abstract**

We use location data from millions of mobile devices to infer three features of income and racial patterns in bank branch access and use throughout the United States. First, residents from poorer block groups are 7.2 percentage points less likely to visit a branch in a year than residents from richer block groups. Likewise, residents from block groups with higher Black population shares are 6.2 percentage points less likely to visit a branch relative to residents from block groups with higher White population shares. Survey evidence suggests that these lower visitation rates are not recouped by greater use of mobile or online banking. The drop-off in visitation by income steepens in large Metropolitan core areas. These urban cores also observe the highest segregation of branch goers by race and income. The urbanized Northeast is more segregated than the rural South. Second, residents from block groups with larger Black population shares on average live farther from their nearest branch and travel farther when visiting banks. A gravity equation model demonstrates that Black residents living farther from bank branches explains roughly 25-33% of the Black-White gap in branch use across the country and 72-86% of the gap within Metro cores. Black residents' far greater remoteness from banks in big cities explains why distance plays such a major role in affecting their branch use there. Third, a policy of postal banking that adds banking services to Post Office branches would relieve some distance costs, and we estimate it would decrease the mean distance to banks in Metro cores by 11.5%. But the policy would close only 6% of the Black-White gap in bank branch use within these areas. The modest effect is due in part to residents of Metro cores living relatively farther away from Post Office branches than they do from private banks, thus making it difficult for postal banking to overcome the distance barriers.

**JEL classification:** G21, J15, L87, R22
**Keywords:** banking, inequality, location economics, spatial analysis, public postal service

*We thank Gary Gorton for his very generous comments. This work would not have been possible without extraordinary research assistance from Gen Li, Lizzie Tong, and Layla Yi. The views expressed in this paper are those of the authors and do not reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System. All errors are our own.

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

# 1 Introduction

Lack of access to financial services has long been considered a leading driver of persistent inequality (Claessens and Perotti, 2007; Beck, Demirgüç-Kunt and Honohan, 2009). But data limitations have routinely compelled economists to rely on aggregate measures, like the number of loans per capita or deposits-to-GDP ratio, when estimating the relations between financial access, the use of financial services, and socioeconomic outcomes (Honohan, 2005; Claessens, 2006). In this article, we capitalize on micro-level observations of individual use of a ubiquitous financial service: bank branches. We show that traveling distance to branches poses a significant access barrier that can explain a substantial amount of the disparity in branch use across the United States.

We characterize bank branch access and use by relying on anonymous location data from millions of mobile devices throughout the US from 2018-2019. We infer income and racial patterns based on demographic characteristics of the Census block groups where the owners of the devices reside. We organize our analysis into three parts.

Section 3 contains the first part. There, we present several new facts on actual branch use across the country. We measure branch use as visits to bank branches. In our core analysis, we take the set of branches as all businesses in the commercial banking industry for which we have visitor data and whose brands are also listed in the 2019 vestige of the Federal Deposit Insurance Corporation's Summary of Deposits (FDIC SOD). Nationwide, we find that for every doubling in a block group's median household income, branch use among its residents increases by roughly 7.2 percentage points per year. This income gradient in branch use is large, as the unconditional likelihood of a mobile device visiting a bank branch during the year is about 72 percentage points. Controlling for income and age, we also find that residents from block groups with higher Black population shares are roughly 6.2 percentage points less likely to visit a branch relative to residents from block groups with higher White population shares.

Our estimates of branch use are consistent with national representative survey responses of banked and unbanked households. Among respondents to the most recent version of the "FDIC's Survey of Household Use of Banking and Financial Services," almost 81% answer having visited a bank branch in the past twelve months, and almost 30% admit having visited a bank branch ten or more times. Hence, despite the growing popularity of mobile and online banking, visiting a branch is still a popular method of accessing bank services. The fraction of households that respond having visited a branch in the past year increases in reported income. And a substantially smaller fraction of Black respondents admit visiting a branch in the past year compared to the fraction of White respondents.

Survey responses further suggest that lower-income and Black households do not make up their lesser branch use with greater use of alternative banking methods like online or mobile banking. Among respondents who are banked and make less than $15,000 in household income annually, only 32% say that online or mobile banking is their most common bank access method. In contrast, 69% of respondents having at least $75,000 in household income cite mobile and online banking as their most common access method. Similarly, 47% of Black respondents name online/mobile banking as their most common access method, compared to 56% of White respondents.

Visits to bank branches as a whole is just one measure of disparity in branch use. Another measure is the extent to which different groups visit different branches. That is, do residents with different income or racial profiles choose different *menus* of bank branches? To answer this question, in Section 3.2 we present new estimates of both income and racial segregation among bank branch visitors across the country. We find that racial segregation at bank branches nationwide is lower than estimates of residential and school segregation (e.g., Massey and Denton, 1988; Cutler and Glaeser, 1997; Clotfelter, 1999; Frankel and Volij, 2011). This finding is consistent with recent work documenting relatively lower segregation in daily-life activities, like restaurant dining (Davis, Dingel, Monras and Morales, 2019) or experienced interactions (Athey, Ferguson, Gentzkow and Schmidt, 2020). We further find that income segregation at bank branches is relatively lower than income segregation in residential areas (e.g., Reardon and Bischoff, 2011; Reardon, Bischoff, Owens and Townsend, 2018) or schools (e.g., Owens, Reardon and Jencks, 2016).

In Section 3.3, we explore the geography of bank branch segregation across US counties. We document three spatial patterns in the two segregation measures. First, racial and income segregation at bank branches tend to be positively correlated. For example, Essex County, NJ ranks first in income segregation and fourth in racial segregation. Second, segregation varies substantially across the country. Both segregation measures are highest in the Northeast, the Midwest (east of the Great Plains), the Southwest, and the Pacific Coast. The South and the Mountain West observe lower bank branch segregation. 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. Third, major cities see the highest segregation. Returning to the previous two examples, Essex County, NJ contains Newark, and

Wayne County, MI contains Detroit. Even in the South, where bank branch segregation is generally lowest, high segregation pockets are seen in and around large cities like Atlanta, Houston, Jackson, and Miami. Nearly 40% of the variation in income segregation and 20% of the variation in racial segregation across counties can be explained by a county's urban share. In fact, urban and rural differences in bank branch segregation are so stark, we estimate that a county which switches from fully rural to fully urban jumps from the 10th to 90th percentile in both racial and income segregation.

In the second part of the article, we investigate the degree to which distance from bank branches is an access barrier to branch use. In Section 4.1, we measure the distance between residents and their nearest branch. We measure the distances using the haversine formula, which accounts for the curvature of the Earth. We find that distance to the nearest branch is a sensible metric to use for appraising access via the distance channel, as nearly 50% of bank branch goers visit their nearest branch rather than ones farther away. The fraction visiting each subsequently ranked branch declines rapidly, with just over 20% visiting their second nearest branch and 10% visiting the third nearest one.

In evaluating how distance affects branch use, we calculate national statistics, but also zero in on parts of the country with Black population shares that resemble the nationwide average. In particular, we examine large Metropolitan core areas, which are high population centers that contain the primary commuting flows within urbanized areas, as these areas match the country most closely by their share of Black residents. Most of the US population resides in these areas as well.

We find that the relation between distance and income is positive: Across the country and within Metro cores, residents of higher income block groups live farther away from their nearest bank branch. Residents from block groups with higher Black population shares also live farther from their nearest branch. Residents of these block groups live about 27% farther away nationwide from the nearest branch compared to residents from block groups with higher White population shares.

In Metro core areas, the Black-White gap in distance from the nearest branch is even larger. We find that residents of block groups in Metro cores with higher Black population shares live about 52% farther from the nearest branch. The positive income gradient and the positive Black-White gap in branch distance, both nationwide and in Metro cores, is highly robust to alternative specifications.

Section 4.2 estimates the distances that branch goers actually travel nationally and within Metro cores. Branch goers from richer block groups travel farther than residents from poorer block groups. A doubling of a block group's median household income is associated with its residents traveling roughly 17% farther. Residents from block groups with larger Black population shares also travel farther by about 9%. In Metropolitan core areas, these relations remain. In these areas, a doubling of a block group's median household income implies that its residents travel between 17% and 19% farther to their branches. Depending on the specification, residents of block groups with higher Black population shares in Metro core areas travel between 6% to 15% farther than corresponding residents of block groups with higher White population shares.

To formally evaluate how distance affects branch use, we use a gravity equation model in Section 4.3. Estimates from the gravity model help us determine the fraction of the disparity in branch visitation that can be explained by distance. Our baseline gravity equation regresses visitor flows from home block groups to visited branches across time on the distances between block group-branch pairs. We include fixed effects for origins and destinations interacted with time to capture (i) characteristics of block groups that contribute to its residents visiting any branch and (ii) characteristics of branches that make them attractive to visit among residents of any block group. Using origin and destination fixed effects dates back to Harrigan (1996) and has become standard practice in the trade literature when estimating gravity relations.

Our estimated gravity coefficient on distance is highly robust. The value is virtually the same for all block group-branch pairs across the country and those pairs limited to Metro core areas. We also consider specifications that include racial population shares of block groups interacted with distance. We find that the coefficients on the interacted terms are fairly small, which suggests that the relation between distance and visitor flows is universal across racial groups. Even so, the coefficient values of the interaction terms are informative. The coefficient on the interaction term with the Black share is positive and precisely estimated. The positive value is consistent with residents from block groups with larger Black population shares having more inelastic demand for banking at branches (and a lower elasticity of substitution with other banking methods). The higher inelasticity implied in the estimate coincides with responses from the FDIC survey, which finds a larger fraction of Black respondents citing bank tellers and ATMs as their most common bank access method compared to White respondents.

A simple method to assess the impact of distance on Black residents' branch visitation is to multiply the estimated gravity coefficient by Black residents' distance to bank branches relative to White residents. Measuring relative distance is challenging, as it ought to consider all branches that Black residents might reasonably entertain visiting. Distances to all these branches would then need to be weighted appropriately to form a single relative distance. Instead, we proxy for this relative distance to all branches with the relative distance to the *nearest* branch. Our finding that a very high fraction of branch goers visit the nearest branch supports the viability of the proxy.

Performing the calculation over our primary sample of bank branches with visitor information, we find that roughly 25% of the Black-White gap in branch use cross-country can be explained by barriers from distance. When considering all branch locations from the FDIC Summary of Deposits in calculating the distance to the nearest branch, we find that distance can explain about 33% of the Black-White gap cross-country.

Metro core areas observe the largest differences between Blacks and Whites in proximity to the nearest branch. The calculation using our primary set of bank branches demonstrates that about 72% of the Black-White gap in branch use can be explained by distance within Metro cores. Expanding the set of bank branches eligible to those in the Summary of Deposits, the fraction explained is roughly 86%.

The national and Metro core gravity coefficient estimates are roughly identical, but the Black-White gap in branch use within Metro cores is over twice as large as the gap nationally. Meanwhile, residents of Metro core block groups with higher Black population shares live relatively farther away from branches compared to their counterparts nationally. This greater remoteness from banks explains why distance plays a much greater role in affecting Black residents' branch use in these big cities.

Finally, in the third part of the article, we evaluate a policy proposal that might improve banking access: postal banking. A Postal Savings System existed in the United States beginning in 1911, but eventually was phased out by Congress in 1966 (O'Hara and Easley, 1979; Shaw, 2018). Re-instituting the Postal Savings System has been a policy proposed by members of Congress (Warren, 2014; Gillibrand, 2021; Sanders, 2021) and parts of academia (Baradaran, 2013; Johnson, 2017).

With our data, we can approximate how a Postal Banking System—which would extend checking, savings, and possibly credit services to some or all Post Office branches—would affect branch use by altering the distance between consumers and their (private and public) bank branches. We use our gravity and distance estimates to measure a partial policy impact of postal banking. This kind of analysis, which keeps the gravity estimates fixed under the new policy and ignores general equilibrium effects under the policy change, is akin to what the trade literature calls a "partial trade impact" of a policy change in trade costs, such as tariffs (Head and Mayer, 2014).

In Section 5, when Post offices are included in calculating the distance between home block groups and their nearest branch, the mean distance nationwide declines from 1.98 miles (when only private bank branches are considered) to 1.45 miles, a roughly 26.7% drop. In Metro core areas, the mean distance from the nearest branch declines from 1.21 miles to 1.07 miles, about a 11.5% decline.

Knowing how distance to the nearest branch would alter under postal banking, we use the gravity equation coefficient to estimate how visitor flows to physical banking services would change under the policy. According to the drop in distance, we estimate that overall visitor flows to physical banking services would increase by roughly 1.38% per month nationally under postal banking.

Breaking down this increase by demographic variables, we show that the income gradient on branch use would flatten by roughly 2% nationally. In other words, postal banking would reduce distances to the nearest bank branches and thus be associated with higher visitation rates nationwide, and poorer block groups would observe a 2% greater increase in bank visitation than richer block groups. We further estimate that the change in the Black-White gap in branch use nationwide would actually *increase* by roughly 1%. The decline in distance would benefit branch visits across racial groups, but block groups with higher Black population shares would observe a roughly 1% smaller increase in visitation compared to block groups with higher White population shares.

In Metro core areas, we estimate that the drop in distance would be associated with visitor flows to bank branches rising by roughly 0.61% per month. We calculate that the income gradient of branch use in Metro cores would flatten by about 1%, which is similar to the national case. The Black-White gap in branch use within Metro cores would decline by roughly 6%.

This last finding implies that postal banking would have the biggest impact on branch use in Metro core areas rather than nationwide, and its effects would be more powerful in narrowing the Black-White gap in branch use rather than compressing the income gradient. Nevertheless, our estimates suggest that any encouragement that postal banking might bring to branch use in Metro cores strictly by relaxing distance costs would be modest. The relatively small impact is due to residents in these large urban areas living relatively farther away from Post Office branches than they do from private banks. Postal banking, however, might encourage bank branch or general banking use through other means, such as improving confidence or trust in financial products, which we do not examine in this article.

**Contribution to the Literature.** First, this article contributes to the literature that takes advantage of mobile device data to answer economic questions. An early example of in this area is Chen and Rohla (2018), who examine how political partisanship affects time spent together during Thanksgiving dinner. Athey, Blei, Donnelly, Ruiz and Schmidt (2018) study consumer choice of restaurant dining. Chen, Haggag, Pope and Rohla (2019) look at racial disparities in vote waiting times. Athey et al.

(2020) develop a measure of segregation based on where people actually visit over the course of a day. Kreindler and Miyauchi (2021) infer the spatial distribution of income from commuting flows in Sri Lanka and Bangladesh. Miyauchi, Nakajima and Redding (2021) measure consumption access and agglomeration of economic activity from consumption and commuting trips in Japan. Many researchers have also used mobile device data to explore topics related to the Covid-19 pandemic (e.g., Coven, Gupta and Yao, 2020; Almagro, Coven, Gupta and Orane-Hutchinson, 2021; Goolsbee and Syverson, 2021; Couture, Dingel, Green, Handbury and Williams, 2021; Chen, Chevalier and Long, 2021). No paper has used this kind of data to examine banking use and access across the US.

Second, this article contributes to the vast array of work that investigates financial access and use and their joint relation with inequality. See Claessens (2006) and Claessens and Perotti (2007) for surveys. Much of this research has examined differences in access and use around the globe. Beck, Demirguc-Kunt and Peria (2007) develop indicators of banking sector outreach across 98 countries (e.g., the number of ATMs per capita, or the number of loans per capita), and they show that these indicators are correlated with factors that influence financial sector depth (e.g., degree of credit information sharing, or the development of physical infrastructure). Beck, Demirgüç-Kunt and Martinez Peria (2008) measure bank access barriers (e.g., minimum account and loan balances, or account fees) across 62 countries. They find that access barriers are higher in countries with sharper restrictions on bank entry, less media freedom, and greater government-owned banking systems. A limitation of these studies has been the reliance on aggregated data in measuring banking access and use. We instead take advantage of detailed individual data to evaluate access and use, which lets us establish that distance to bank branches poses a significant access barrier for certain groups in the US.

Third, this article contributes to the large literature in regional and urban economics on commuting flows and the spatial arrangement of economic activity. Much of this work has focused on either firm and household location decisions (e.g., Lucas and Rossi-Hansberg, 2002; Ahlfeldt, Redding, Sturm and Wolf, 2015) or agglomeration effects (e.g., Dekle and Eaton, 1999; Rosenthal and Strange, 2004). We are the first to use micro-level observations of actual travel behaviors to infer spatial patterns in banking acti branchy.

Finally, this article contributes to empirical work that has examined postal banking systems in the US (e.g., O'Hara and Easley, 1979; Schuster, Jaremski and Perlman, 2020) and around the world (Cargill and Yoshino, 2003). We contribute to the discussion of re-establishing postal banking in the US by estimating its potential impact on branch use via the distance channel.

# 2 Data

We use mobile device data from SafeGraph from January 2018 through December 2019. The data include bank branch locations and information about branch visitors. A visitor is identified by a mobile device, and one device is treated as one visitor. Visitor information includes the total number of visitors to each branch, the median dwell time spent at the branch, the median distance travelled from home, and visitors' home Census block groups. A device must spend at least 4 minutes at the branch to qualify as a visitor. Unless specified otherwise, the data are monthly. Appendix A provides background information on the SafeGraph data and a detailed explanation of the way we construct our primary analysis sample. Here we give a brief summary.[1]

## 2.1 Primary Bank Branch Sample

Our primary (core) data set consists of bank branches across all fifty states and the District of Columbia. SafeGraph categorizes businesses by their six-digit NAICS codes. To ensure that we only analyze depository institutions in the SafeGraph data, we take advantage of information from the FDIC's Summary of Deposits (SOD), which is an annual survey of bank branches for all FDIC-insured institutions.

In our core sample, we include only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) whose brands are also listed in the 2019 vestige of the SOD. For example, Wells Fargo & Company and SunTrust Banks, Inc. are two bank brands with branch locations in SOD. We therefore include all Wells Fargo and SunTrust Bank branch locations in SafeGraph. The physical locations of bank branches are identified by SafeGraph's geographic coordinates for them, rather than the SOD's, as we found that SafeGraph's coordinates typically were more accurate.[2]

Our core sample is confined to bank branches for which SafeGraph has visitor information. Many

[1]SafeGraph asks all researchers who use the company's data to include the disclaimer: "SafeGraph is a data company that aggregates anonymized location data from numerous applications in order to provide insights about physical places, via the Placekey Community. To enhance privacy, SafeGraph excludes census block group information if fewer than two devices visited an establishment in a month from a given census block group." The documentation to the SafeGraph data is here: SafeGraph Documentation.

[2]For most branches, the geographic coordinates in SafeGraph and the SOD matched. When the two sources disagreed, a Google Maps search of a branch address in the SOD often confirmed that no physical place existed at that address. (The place's absence was not due to a branch closing.) A higher quality set of geographic coordinates from SafeGraph should come at little surprise, as the success of the company's business relies in part on providing highly accurate location information.

bank locations that are recorded in SafeGraph lack such information, as it is often difficult to attribute mobile device visits to particular branches. There are two main reasons. First, in dense environments (such as multi-story buildings or malls), SafeGraph might not be confident about the geometric boundary of a place, which makes attributing visits to a unique place that is part of a shared space awfully difficult. To reduce false attributions, SafeGraph instead often allocates visits to the larger "parent" space. Second, and related, a bank branch might be entirely enclosed indoors within a parent location (i.e., a customer must enter the parent's structure to reach the branch). Because mobile device GPS data accuracy deteriorates severely within indoor structures, SafeGraph is reluctant to assign visits to an enclosed branch. Instead, those visits are aggregated to the level of the parent location. For example, many Woodforest National Bank branches are enclosed in Walmart Supercenters. (Walmart partners with Woodforest to provide the retail company's banking services.) Visits to these enclosed branches cannot be separated from visits to Walmart, and so, these branches are deprived of visit data.

The SOD registers 86,374 bank branch locations as of 2019. While SafeGraph can account for 71,468 branches according to our core sample definition (83% coverage), only 51,369 of these places have visitor information and constitute our core sample. Our core sample thus covers around 60% of bank branches in the United States.[3]

## 2.2 Assigning Demographic Attributes to Individual Visitors

In using anonymous mobile device data, we face a limitation: We do not know the precise demographic attributes of an individual bank branch visitor. Instead, we must assign attributes to visitors according to the demographic characteristics of their home Census block groups. Inferring individual attributes or behavior from aggregate data is a well-studied area in social science known as ecological inference (King, 1997; King, Tanner and Rosen, 2004).

The information lost in the aggregation makes ecological inference challenging. Aggregate demographic characteristics of a block group, such as the median household income or Black population share, might not necessarily fit an individual branch goer or even the average one. For example, we will observe in the data that the expected probability of a person visiting a bank branch increases in the median household income of the person's home block group. Based on this finding, a resident from a low-income block group who visits a bank branch is more likely to earn higher income than her

[3] Online Figure A.2 presents a time-series of the number of branches per month in our core sample. The number of recorded branches per month is fairly stable and averages around 38,000.

average neighbor.

We have an advantage in that our spatial unit of observation is a Census block group, which is typically quite small in geographic area. Differences in demographic attributes among residents of block groups is narrower than differences over larger spatial units, such as zip codes. Hence, inferring individual behavior from grouped data over these smaller areas has less error. In addition, the heterogeneity in attributes within a block group is also smaller than the heterogeneity across block groups, which is the variation we exploit in estimating spatial patterns of branch use.

Even so, benefiting from block-group-level information does not mean we escape entirely from the ecological inference problem. Focusing on household income, Online Figure A.1 Panel A presents the percentiles of the distribution of individual-level household income and block-group-level median household income. The percentiles of the two distributions are quite close from the 50th percentile and below. This close alignment of the two distributions over these percentiles suggests that individual-level behavior based on income can be inferred quite accurately from the grouped data over this income range. As the percentiles get farther above the median, however, the gap between the two distributions grows substantially. Individual-level household income at the top percentiles is over twice as large as block-group-level median household income. This divergence is unsurprising, as calculating the median household income naturally compresses the distribution across block groups.

When faced with an ecological inference problem, how then can one interpret our coefficients from linear regressions of variables of interest on demographic attributes? First, in the strictest sense, the interpretation must be restricted to associating the dependent variable of interest with characteristics of block groups. For example, suppose that the shares of mobile devices that visit a branch are regressed on block-group racial population shares (with the White population shares omitted). And suppose that the regression produces a coefficient estimate of $-x$ on the Black population share, which is one of our key independent variables of interest. The strict interpretation would be: "A 1% increase in the Black population share of residents in a block group is associated with a $x$% lower share of residents from that block group visiting a bank branch."

A second, looser interpretation would express a more global effect. Although the linear coefficients measure local, incremental changes, one can extrapolate the estimated effects to a global change. One can do so with more confidence if the independent variable fully spans its domain across block groups. Online Figure A.1 Panel B plots the distribution of the Black population shares across block groups. Block groups in our cross section span a range from having a 0 percent to nearly 100 percent Black population share. Therefore, an extrapolated interpretation such as the following is more plausible in our setting: "A block group with a 100% Black population share observes a $100x\%$ decline in its share of residents visiting a branch, compared to a block group with a 100% White population share."

The third, and loosest, interpretation of our coefficients is to ignore the ecological inference problem entirely and interpret individual-level behavior from the grouped data. Our small geographic units of observation, the proximity of the block-group income distribution to the individual-level income distribution for nearly all but the top percentiles, and the spanning of the domain in the Black population share gives more credence to this interpretation than otherwise. Such an individual-level interpretation would be: "A Black resident is estimated as $100x\%$ less likely to visit a branch than a White resident."

## 2.3 Descriptive Statistics

Table I reports descriptive statistics from our core sample of bank branches. The typical branch has 40 visitors a month on average and an interquartile range of 5 to 48 visitors. The median distance that visitors travel from home to a branch is 5 miles on average, and the 90th percentile is only 9 miles. The median dwell time that visitors spend at a branch is 49 minutes on average, though it ranges from 6 minutes (10th percentile) to 2.5 hours (90th percentile). Finally, of the 36.5 million total mobile devices recorded in our core sample with information on the type of device, 52% are iOS and 46% are Android.

In Table II, we compare demographic characteristics of the geographic areas represented in our core sample with geographic areas represented by all bank branches in the SOD. Demographic characteristics in the table are taken from the 2019 5-year American Community Survey (ACS) and are averaged at the level of the Census Bureau's zip code tabulation area (ZCTA). In areas represented by the SOD, the fraction of white households is 80.5%, which aligns closely with the 79.9% share of white households in areas represented by our core sample. The SOD and core sample are also similar according to the percentage of black households (9.5% in SOD vs. 10.3% in our core sample) and the percentage of Hispanic households (10.6% vs. 10.9%). Median household income in areas covered by our sample is just over $500 (1%) higher on average than median household income in areas covered by the SOD. Urban areas are over-represented by about 3% compared to the SOD, which coincides with the greater mobile coverage in urban over rural areas. The differences in demographic attributes between the two samples are precisely estimated and significant, but overall, the economic magnitudes of the differences are small relative to the mean values across areas.

# 3 Bank Branch Use

We begin our empirical analysis by describing bank branch use in the United States. We first relate the share of bank branch visitors to household income, race, and ethnicity. These shares give an estimate of a person's expected likelihood of visiting a bank branch according to the person's demographic attributes. We then measure the level of segregation at bank branches and compare it to estimates of residential segregation in the literature. We present statistics for both the US as a whole and across different areas. [4]

## 3.1 National Bank Branch Visitation

In our primary analysis, we use our core sample of bank branches (all branches in SafeGraph with visitor data whose bank brands are present in the 2019 SOD). Figure I presents a binned scatter plot of the share of bank branch visitors by household income. Our variable for household income is the median household income of a visitor's home Census block group, as measured in the 2019 5-year ACS. To construct this panel, we divide the horizontal axis into 100 equal-sized (percentile) bins and plot the mean annual share of residents visiting a bank branch versus the mean household income within each bin. Each point represents a nonparametric estimate of the expected likelihood that a person visits a bank branch over the past year, conditional on the person's household income.

To support the reliability of our estimates, we compare branch visitation patterns based on the mobility data to reported bank visits from a national survey. Every two years in June, the FDIC fields a survey on households' use of banking and financial services.[5] The most recent survey was conducted in 2019, and in that year, 80.9% of all households responded having visited a bank branch in the past 12 months, and just over 29.7% admitted having visited a branch ten or more times. Traveling to a

[4] Regarding visits, if a bank branch closed and SafeGraph were aware of its closure, any visits to the building (say, if a new business opened there) would no longer be attributed to the branch. Likewise, if a branch opened and SafeGraph were aware of it, visits would start being attributed to the branch. Nevertheless, if SafeGraph is unaware of a branch's opening or closing, visits would be incorrectly attributed and count toward measurement error.

[5] The survey is a supplement to the US Census Bureau's Current Population Survey, which covers a representative sample of households in the US each month. The FDIC survey queries both banked and unbanked households, and the most recent survey collected responses from almost 33,000 households. More information on the FDIC survey and the survey's latest findings are here: "2019 FDIC Survey of Household Use of Banking and Financial Services".

branch is the primary (i.e., most common) method of bank use among 23% of banked respondents. Mobile banking is more frequently cited as a primary method of use (31.4%) for banked households, but even in this group of respondents, 81.2% admitted visiting a branch over the past year and about 1 in 5 in that group visited ten or more times. Overall, despite the growing popularity of mobile and online banking, visiting a branch is still a common and popular method of accessing bank services.

Behind the binned scatter plot in Figure I, we insert as a bar chart the 2019 FDIC survey responses across the five income buckets available in the survey. The survey response is the share of households (among both banked and unbanked) that acknowledged visiting a bank branch within the past 12 months (i.e., between July 2018 and June 2019). To coincide with the 12-month span of the FDIC survey, we measure the annual share of actual branch visitors in the binned scatterplot over that same period.[6]

The comparison of the FDIC's survey responses to the visitation patterns observed in SafeGraph is not perfect. The survey responses measure whether a respondent visited any US bank branch (i.e., the extensive margin across all branches), whereas SafeGraph measures whether a person visited a *particular* branch (i.e., the extensive margin between branches). SafeGraph distinguishes visits from visitors, and we use visitor values in Figure I. The same person visiting the same branch multiple times in the month would count as one visitor, but the same person traveling to multiple branches in the same month would count as distinct visitors. The SafeGraph values in the figure would exactly match the survey responses if (i) SafeGraph included all bank branches in the United States, (ii) it recorded every branch visitor without error, (iii) it separated out visitors to multiple branches, (iv) branch visits were independent month-to-month, (v) we knew the household income of individual visitors rather than only the median income of their home block groups, and (vi) survey respondents answered accurately.

Notwithstanding these imperfections, relating the FDIC survey responses to the visitation patterns in SafeGraph is useful and reveals a strong resemblance between the two sources. Both reported branch visitor shares from the FDIC survey and actual branch visitor shares from the mobility data are increasing and concave in household income. Around 63% of respondents with household income less

[6] To compute this annual share of branch visitors, we first divide the total branch visitors in each Census block group by the total recorded devices residing in the block group per month. This ratio gives an estimate of the probability that a device from each home block group visits a bank branch at least once during the month. Let this estimated branch visitor probability for block group $j$ in month $t$ be denoted $p_{j,t}$. Not every block group has a visitor probability each month; so, let $k_j$ denote the number of months for which block group $j$ has observations. The annual branch visitor share $s_j$ for block group $j$ is computed as $s_j = 1 - \prod_{t=1}^{12/k_j} (1 - p_{j,t})^{12/k_j}$. Each home Census block group thus has an annual branch visitor share, and we then categorize block groups by household income, measured from the 2019 5-year ACS.

than \$15,000 reported having visited a branch over the past year, whereas 86% of those with income \$75,000 and above reported has having visited. Using the mobility data, the actual visitor share is 59% for households earning around \$12,000 and 71% for households earning around \$206,000.

Despite the two sources displaying similar relations between household income and a person's expected likelihood of visiting a bank branch, the FDIC survey responses and SafeGraph visitor shares differ from two important aspects. First, the SafeGraph shares are systematically below the corresponding shares from the FDIC survey. These lower values are most likely due to our core sample omitting many US bank branches (and their visitors). Another contributing explanation is SafeGraph entirely missing some visitors to branches, either from errors in attributing a mobile device to a branch or from short duration trips that are not counted as a visit. Second, our estimated expected likelihood of visiting a branch for every additional thousand dollars in household income rises at a slower pace than the survey responses suggest. To understand this muted slope, recall that income is measured as the median household income of a visitor's home Census block group rather than the person's individual income. Because the likelihood of visiting a bank increases in income, branch visitors from lower-income block groups are more likely to earn income above their block group's median. The most likely explanation of the difference in slopes is this measurement error that inflates the visitor shares at the bottom of the income distribution. Another possibility, though, is that SafeGraph regularly misses branch visitors from higher income block groups, which would understate the visitor shares at the top of the income distribution and compress the slope.

We transition now to evaluating the statistical relation between visitor's demographic attributes and their branch visitation. We start by examining cross-sectional patterns across the entire US using our core sample. Because a central focus of our study is answering how bank branch use and access differs among Blacks compared to Whites, we also zero in on local parts of the country that present Black population shares close to the national average. In these areas, we can make meaningful comparisons in branch use between Blacks and Whites. To identify these areas, we partition the US into the 10 Rural-Urban Commuting Areas (RUCAs) classified by the US Department of Agriculture's Economic Research Service. RUCA codes separate census tracts by their urban/rural status and their commuting relationships with other areas using Census measures of population density, levels of urbanization, and daily home-to-work commuting.[7]

Online Table A.2 presents household counts and Black shares throughout the US and within each

[7] The RUCA data are available here: RUCA classification.

RUCA. The national Black share is 12%. The commuting areas with figures closest to these national shares are Metropolitan area core (Metro core), having a 15% Black share, and Micropolitan area core (micro core), having a 9% Black share. Metro core areas are census tract equivalents of urbanized areas, which themselves are urban areas with populations of 50,000 or more. Micro cores are census tract equivalents of large urban clusters, which themselves are urban areas with populations between 10,000 and 49,999.[8] Despite the similarity in racial shares, Metro core areas vastly outnumber micro core areas in household counts (99.5 million vs. 8.5 million), and Metro core areas capture roughly 72% of the 138.9 million total households in the US. For this reason, we supplement our national estimates of bank branch use with local estimates in Metro core areas.

Table III presents weighted OLS regressions of bank branch visitor shares by demographic attributes. Observations are at the level of a home Census block group per year-month over the core sample period and are weighted by the number of mobile devices residing in the block group in the year-month. Standard errors are clustered at the block-group level. Visitor shares are calculated in the same way as in Figure I, except not annualized. Independent variables are population-based shares from the 2019 5-year ACS. The five racial/ethnic groups used in the regressions are non-Hispanic Asian, non-Hispanic Black, non-Hispanic White, non-Hispanic Other Races, and Hispanic.

In columns (1)-(6), the dependent variable is the monthly visitor share times 100. The unconditional likelihood of a mobile device visiting a bank branch during the month is 9.99%. Column (1) reports coefficients of visitor shares on household income and race. The coefficients in the column define a "branch use gradient" as a function of demographic attributes. Residents of higher income block groups are more likely to visit a branch (about 0.29 percentage points higher for every doubling in median household income), whereas residents of block groups with larger Black or Hispanic population shares are less likely. Extrapolating from the coefficients on racial shares, we estimate that Black residents are 1.696 percentage points less likely to visit a branch per month compared to White residents, and Hispanic residents are 1.515 percentage points less likely.[9]

Column (2) adds county and year-month fixed effects to the income-race/ethnic specification of

[8]In this context, "urban areas" follows the Census Bureau's definition provided here: Urban Area Definition.

[9]These estimates, once annualized, are also comparable to the FDIC survey. Among all White survey respondents (banked and unbanked), 84.26% answered having visited a bank branch in the past year. Among Black respondents, 69.1% reported having done so, and among Hispanic respondents, it was 71.85%. In unreported regressions with only racial shares as regressors, we find that the monthly visitor share among White residents is 11.05%, the share among Black residents is 9.13%, and the share among Hispanic residents is 9.36%. On an annual basis, these visitor shares are 75.47%, 68.29%, 69.26%, respectively.

column (1). The positive relation between income and visitation strengthens substantially (from 0.29 to 1.018 percentage points), but the coefficients on race/ethnicity drop by about half (to -0.803 and -0.712 percentage points, for Black and Hispanic population shares, respectively), though the coefficients are still precisely estimated. Even controlling for differences in county visitation patterns or possible seasonality in branch visits, the positive relation between income and visit likelihood remains, as does the negative relation with Black and Hispanic population shares.

An important factor that might drive branch visits are differences in financial savvy or technical sophistication from disparities in age or education (Caskey and Peterson, 1994; Caskey, 1994; Hogarth and O'Donnell, 1997; Hogarth, Anguelov and Lee, 2005; Blank and Barr, 2009). To evaluate this avenue, in column (3) we add age shares, and in column (4) we additionally include education shares. The coefficients on income and both Black and Hispanic population shares remain roughly the same with the additional controls. Block groups with greater shares of 15- to 34-year-olds observe the lowest visitation, and older home block groups (55+) see the highest visitation (between 6-7.6 percentage points higher). Similarly, block groups with higher shares of college graduates and post-college degree earners see lower visitation. These findings are consistent with relatively younger, more educated, and financially sophisticated residents opting for online and mobile banking over visiting branches; and older, less educated, and less technically savvy cohorts relying on face-to-face interactions with bankers and tellers over mobile and online banking.

On an annualized basis, estimates from column (3) imply that residents from poorer block groups are 7.2 percentage points less likely to visit a branch in a year than residents from richer block groups. Likewise, residents from block groups with higher Black population shares are 6.2 percentage points less likely to visit a branch relative to residents from block groups with higher White population shares.

A natural question might be whether residents of poorer block groups or block groups with higher Black population shares substitute their lower branch visitation rates with greater use of alternative bank methods, such as online or mobile banking. Responses from the FDIC survey suggest not. Among respondents who are banked and make less than $15,000 in household income annually, only 31.7% admit that online or mobile banking is their most common bank access method. In contrast, 68.9% of respondents having at least $75,000 in household income cite mobile and online banking as their most common access method. Similarly, 46.5% of Black respondents name online/mobile banking as their most common access method, compared to 55.9% of White respondents.

Moving away from nationwide estimates of branch use to local ones, we next focus on large urban centers in column (5). Here we restrict observations to block groups classified as Metropolitan area cores (MC). Regressing visitor shares on income and the racial/ethnic categories with county and year-month fixed effects produces a positive coefficient on income roughly the same as in column (2), which used all block groups. The coefficients on Black and Hispanic population shares remain negative, though are roughly half in magnitude in Metro cores (-0.426 and -0.383, respectively). Column (6) again focuses on Metro core areas, but adds age shares. Here, the magnitude of the Black population share coefficient is further reduced (to -0.257 percentage points), and the effect of the Hispanic population share on visitation is no longer precisely estimated.

To evaluate how the flow of branch visitors correlates with the demographic attributes of home block groups, we use the natural logarithm of the number of branch visitors as the dependent variable in columns (7) and (8). In both specifications, we control for the natural logarithm of the number of mobile devices residing in the block group. Across the country and in Metro core areas, between 24-25% more devices visit a bank branch per month for every doubling in median household income. Block groups with higher Hispanic population shares observe around 10% fewer branch visitors per month. Similarly, block groups with larger Black population shares see around 2.5% fewer visitors per month in Metro core areas and 5.7% fewer visitors cross-country. These sensitivities in visitor flows will be helpful when analyzing the effects of branch distance on visitation in Section 4.3.2.[10]

## 3.2 National Bank Branch Segregation

Having established that racial and income groups vary significantly in the amount of bank branch services they use, we turn next to examining the extent to which different groups choose different *menus* of branches. In other words, do Blacks, Hispanics, and Whites sort into distinct branches or do they commingle at the same branches? Likewise, do the rich and the poor separate in the branches they visit? A natural way to investigate these questions is to estimate measures of segregation among bank branch visitors.

[10]Note that the coefficients in columns (7) and (8) are internally consistent with their counterparts in columns (3) and (6). In particular, the coefficient of -0.057 on the Black population share is roughly the same as the counterpart coefficient value of -0.536 in column (3) divided by the value of the constant 7.775 in column (1). Likewise, the coefficient of 0.244 on log income in column (7) is about the same as the coefficient value of 0.622 in column (3) divided by the constant value of 7.775. Similar calculations reveal comparable values when comparing the coefficients in column (8) to those in column (6). These relations between the specifications are not surprising, as they describe the same visitation patterns, except that coefficients in columns (1)-(6) are expressed in units of percentage points, whereas those in columns (7) and (8) are expressed in percent.

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.[11] For the most part, the literature has focused on residential or school segregation. In this section, we present new segregation estimates among visitors to bank branches across the US. 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., 2020).

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 IV presents several segregation measures at the national level. Our three main segregation measures are (i) racial dissimilarity, (ii) racial entropy, and (iii) income entropy.

### 3.2.1 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 the measure if the group is unevenly separated over spatial areas (Massey and Denton, 1988). Elaborating on this index, suppose an area is partitioned into $N$

[11] 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).

sections. Following Echenique and Fryer Jr. (2007), the dissimilarity index between Black residents and non-Black residents in the area is

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

where $\text{Black}_i$ is the number of Black residents in section $i$, $\text{Black}_{\text{total}}$ is the total number of Black residents in the area, $\text{Non-Black}_i$ 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. (1) for bank branch visitors by estimating each component. Let $N$ be the total number of branches in the country. 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) 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 $\times$ Black-share products together. In symbols, let $n_{j,i}$ denote the number of visitors from block group $j$ to branch $i$, and let $\pi_j$ denote the Black population share of block group $j$. The estimate

$$\widehat{\text{Black}}_i = \sum_j n_{j,i}\pi_j. \tag{2}$$

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) that 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

$$\hat{\Pi}_i = \sum_j \left(\frac{n_{j,i}}{N_i}\right)\pi_j. \tag{3}$$

The estimate of the expected total Blacks visiting banks in the country is

$$\widehat{\text{Black}}_{\text{total}} = \sum_i N_i \hat{\Pi}_i. \tag{4}$$

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 IV is a simple average over the core sample period. The monthly estimates are quite stable, and they are provided in Online Table A.3.

From the table, the national estimated Black/non-Black dissimilarity index is 0.447. In the table, we also provide comparison estimates of Black/non-Black dissimilarity from several other research papers across several contexts. 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 they 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 consumers 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, DC 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.

### 3.2.2 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. Let $\pi_s$ 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 $s$.[12]

Following Reardon and Firebaugh (2002), the entropy segregation index is

$$\text{Entropy Index} = \sum_{i=1}^{N} \frac{\text{visitors}_i}{\text{visitors}_{\text{total}}} \left(1 - \frac{E_i}{E}\right), \tag{5}$$

where $\text{visitors}_i$ 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

[12]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 $E \neq 0$.

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. (5) that require estimation are the fractions of branch visitors belonging to a group, both for individual branches $(\pi_{s,i})$ and across the country $(\pi_s)$. We estimate $\pi_{s,i}$ in an identical fashion as $\hat{\Pi}_i$ in Eq. (3) 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.[13] Each group has its own estimate, denoted $\hat{\Pi}_{s,i}$. The estimate for $\pi_s$ is computed similarly as Eq. (4) of the previous section. Specifically, let $N = \sum_i N_i$ denote the total number of bank branch visitors in the country, where, again, $N_i$ is branch $i$'s total visitors. The estimate for the share of branch visitors from each group throughout the country is

$$\hat{\Pi}_s = \sum_i \left(\frac{N_i}{N}\right) \hat{\Pi}_i. \tag{6}$$

From the table, the national estimated racial/ethnic entropy index is 0.204. (Estimates per month over the core sample period are provided in Online Table A.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, and 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 US. 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 US 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.

[13] Like 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.

### 3.2.3 Income Entropy Index

An entropy-based measure can be used to examine income segregation among bank branch visitors as well, which is where we turn next. We adopt the rank-order income segregation measure from Reardon (2011), which accounts for the 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. (5) 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, the income segregation index is defined as

$$\text{Income Segregation Index} = 2\ln(2)\int_0^1 E(p)H(p)\,dp. \tag{7}$$

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. (7) 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 Online Appendix B.

From the table, the national estimated income entropy index is 0.059. (Estimates per month over the core sample period are provided in Online Table A.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 et al. (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 et al. (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.

## 3.3 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 Sections 3.2.2 and 3.2.3. Bank branches are assigned to counties according to their location 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.[14]

Figure II 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.[15]

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%. Online

[14] 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%.

[15] 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.

Table A.1 presents the top-50 US 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.[16] 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. Online Figure A.5 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. Online Figure A.6 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

[16] 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%.

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.

# 4 Bank Branch Access

We move now to examining bank branch access across the United States. We think of *access* as the ease or availability to obtain or make use of banking services at a branch. We start by measuring the distances between residents and their nearest bank branches according to income and race. We then estimate actual distances traveled to branches by race and income. Finally, we present gravity equations according to these estimates to infer which fraction of the Black-White gap in branch use can be explained by distance barriers.

## 4.1 Distance from Nearest Branch

We start by examining the distances between residents and their nearest bank branch. Provided that residents bear any kind of transportation costs, remoteness from physical bank locations interferes with branch access. Several researchers have studied how farther distance from vital products, such as grocery stores and hospitals, depresses use of those products and disrupts welfare (Yantzi, Rosenberg, Burke and Harrison, 2001; Inagami, Cohen, Finch and Asch, 2006; Nicholl, West, Goodacre and Turner, 2007). Furthermore, distance to these products have been found to vary by socioeconomic status (Currie and Reagan, 2003; Hamrick, Hopkins et al., 2012). Here, we assess how distance from the nearest branch varies by income, race, and ethnicity. We later use these estimates to evaluate the degree to which branch distance influences branch use.

Distance from the nearest branch alone is not the only measure of proximity we could have chosen. We omit other branches that could be within reasonable reach to households. We could instead, say, examine a weighted average distance from all branches within a fixed radius. By selecting the nearest branch, we also restrict attention to a single bank brand in the area, which might not be the most popular among households who live there. A Bank of America branch could be the nearest around, but if surrounding residents primarily are customers of a remote Sandhills Bank branch, the distance to the single nearest branch would misrepresent branch access.

To assess our decision to use the nearest branch, we examine the shares of visitors who travel to their nearest branch, their next nearest branch, their next, next nearest branch, and so on. These calculations produce an empirical distribution of visitor shares to surrounding branches by ranked distance. Figure III presents this distribution. Each share in the figure is an estimate of the expected likelihood that a randomly drawn visitor travels to a branch according to the branch's ranked distance from home. The figure indicates that nearly 50% of bank branch goers visit their nearest branch rather than ones farther away. The fraction visiting each subsequently ranked branch declines rapidly, with just over 20% visiting their second nearest branch and 10% visiting the third nearest one. This fast drop off implies that distance from the nearest branch is a sensible metric to use for appraising access.

We consider multiple sets of US branches when measuring distance. The first set is all branches having visitor information in our core SafeGraph sample. In the next section, we will compare estimates from this set with estimates of actual distances traveled by branch visitors. The second set is all branches available to residents according to the 2019 Summary of Deposits. This set delivers a fuller picture of access to local branches across the country. The remaining branch sets are those located in Metropolitan core areas within SafeGraph and the Summary of Deposits. We measure distance from the nearest branch to the population-weighted center of visitors' home block groups.[17]

Table V presents weighted OLS regressions of log distance between home block groups and their nearest branch by demographic attributes. The specifications mirror those from earlier in Table III, where the branch visitor share was used as the dependent variable. As before, observations are at the level of a home Census block group per year-month over the core sample period, they are weighted by the number of mobile devices residing in the block group in the year-month, and standard errors are clustered at the block-group level. We use panel regressions (instead of cross-sections) to account for changes over time in the menu of banks available to households from branch openings and closings.

Costs of travel vary substantially throughout the diverse US landscape. A mile in downtown Chicago is much costlier to traverse in a car, train, or bus than a mile in the surrounding Cook County suburbs.

[17] The centers of population are computed using population counts from the 2010 Census and are found here: 2010 Census Centers of Population. We measure distance using the haversine formula, which accounts for the curvature of the Earth. The haversine distance between two latitude-longitude coordinates $(lat_1, long_1)$ and $(lat_2, long_2)$ is $2r \arcsin\left(\sqrt{h}\right)$, where $r$ is the Earth's radius and $h = \text{hav}(lat_1 - lat_2) + \cos(lat_1)\cos(lat_2)\,\text{hav}(long_2 - long_2)$. The haversine function $\text{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right)$. We take the Earth's radius to be 3,956.5 miles, which is midway between the polar minimum of 3,950 miles and the equatorial maximum of 3,963 miles. The haversine formula treats the Earth as a sphere and is less precise than other measures that consider the Earth's ellipticity, such as Vincenty's formula. Yet another alternative that is more representative of actual traveling distance is the road driving distance between locations. Nevertheless, the haversine formula is simple, fairly accurate, and convenient to compute, unlike these other measures that involve iterative methods, potentially enormous computational resources, or reliance on proprietary algorithms.

Because transportation costs might differ even within counties, county fixed effects are insufficient as controls. To control for variation in traveling times and make the notion of "distance" as comparable as possible across different types of areas (urban, rural, and suburban), we add RUCA fixed effects to our specifications in the table.

Across the country, residents are on average 1.98 miles away from their nearest branch. The median distance cross-country is 1.07 miles. In these calculations, the nearest branch distance per block group is weighted by the block group's population count. Column (1) conditions the log distance regressions on just household income and race. Residents of richer block groups tend to live farther away from their nearest branch. For every doubling in median household income, the nearest branch is roughly 7% farther away. Upon the inclusion of county, year-month, and RUCA fixed effects in column (2), the positive gradient of distance on income is much sharper (a doubling in household income is associated with the nearest branch being nearly 43% farther away). Indeed, all specifications in columns (1)-(8) demonstrate that residents of higher income block groups live farther away from their nearest bank branch.

Residents of block groups with higher Black population shares are also farther from the nearest branch once controlling for the county, year-month, and RUCA status in column (2). Extrapolating from the coefficients on racial shares, we estimate that Black residents live roughly 39.2% farther away from their nearest branch compared to White residents. Controlling for age in column (3) as well as education in column (4) still preserves the positive relation between the share of Black population and distance to the nearest branch, though the magnitude drops. Controlling for age, a Black resident is 27.2% farther from the nearest branch and roughly 9.5% farther once additionally controlling for education.

In Metropolitan core areas, residents are on average 1.21 miles away from their nearest branch. The median distance in Metro cores is 0.91 miles. Column (5) restricts the sample to block groups in these areas. There, the coefficients on both Black and Hispanic shares are strongly positive. Once controlling for age in column (6), extrapolations of the coefficients imply that a Black resident in a Metro core block group lives about 35.4% farther from the nearest branch than a nearby White resident. Similarly, an Hispanic resident in a Metro core block group lives roughly 1.9% farther away. In a binned scatter plot, Figure IV illustrates this clear positive relation between Black or Hispanic population shares in Metro core areas and distance from the nearest bank branch.

Finally, in columns (7) and (8), we expand the sample of bank branches to reflect all locations in the 2018 and 2019 FDIC Summary of Deposits. Column (7) includes distances from that set of branches to all home block groups for which we have SafeGraph visitor data, and column (8) restricts the set of home block groups to those with visitor data in Metro core areas. The positive relation between the Black population share and distance (with higher magnitudes in Metro core areas) is actually stronger under this expanded set of bank branches. When all SOD branches are included in the set available to residents, Blacks live about 35.5% farther away from the nearest branch compared to Whites cross-country, controlling for age. This national estimate is roughly 1.3 times higher than the 27.2% coefficient value in column (3), which has the same specification over the smaller core set of branches. In Metro core areas, Black residents live about 42.2% farther when controlling for age and including the set of SOD branches. This Metro core estimate is about 1.2 times higher than the 35.4% estimate in column (6), which used just our core sample of branches. The Black-White gap in distance to the nearest branch is thus highly robust to specifications and the set of bank branches considered.

Overall, results from Table V show that (i) across the country, residents of higher income block groups reside farther from their nearest bank branches, and (ii) residents of block groups with greater Black population shares reside farther from their nearest branch, particularly in Metro core areas.

## 4.2 Distance Traveled to Branches

In the previous section we asked: Which types of residents on average live closer or farther from their nearest bank branch? Those estimates provide one assessment of the "supply" of bank branches available to residents according to their demographic attributes. We turn next to estimating how the actual distances that residents journey to branches vary with those attributes.

Actual distance traveled is a person's equilibrium choice. It combines both supply factors (e.g., available bank brands in the area or the distances to nearby branches) and demand factors (e.g., immediate need for banking services, transportation costs, or the elasticity of substitution between branches). To account for as many of these factors as possible in assessing residents' branch visitation decisions, we consider all branches visited by a block group and weight the distances traveled by the number of visitors.

Table VI presents weighted OLS regressions of weighted average log distances between home block groups and visited bank branches by demographic attributes. The specifications mirror those from earlier in Table V. As before, observations are at the level of a home Census block group per year-month over the core sample period, they are weighted by the number of mobile devices residing in the block group in the year-month, standard errors are clustered at the block-group level, and both county and RUCA fixed effects are added.

Cross-country, residents travel 27.17 miles from their block groups to bank branches on average. The median distance traveled is 3.98 miles. Actual travel distances are thus significantly larger than distances to the nearest branch. These actual distances traveled include outlier cases in which a resident might have journeyed to another part of the country on a short trip and incidentally visited a bank branch. We do not explicitly filter out exceedingly large distances traveled from home because these instances are likely rare and carry low weight in our regressions.

Column (2) demonstrates that branch goers from richer block groups travel farther than residents from poorer block groups. A doubling of median household income is associated with traveling roughly 17.1% farther on average. Residents from block groups with larger Black population shares also travel farther, which coincides the early finding that residents from these block groups also live farther from the nearest branch. Extrapolation of the coefficient implies that a Black branch goer travels about 9% farther than a White branch goer. In Metropolitan core areas, branch visitors travel on average 26.49 miles from their home block groups to bank branches. The median distance traveled in Metro cores is 3.24 miles. In these areas, a doubling of a block group's median household income implies that its residents travel between 16.6% to 18.8% farther to their branches. Extrapolation from the coefficients implies that Black residents in Metro core areas travel between 6.3% to 13.5% farther than White residents.

## 4.3 Gravity Model

Thus far, we have established several new facts about branch use and branch distance by income and race. Regarding income, (i) residents of poorer block groups visit bank branches less frequently than residents of richer block groups, (ii) they live closer to the nearest branch, and (iii) among those visiting branches, they travel shorter distances.

Regarding race, (i) block groups with higher Black population shares visit branches less frequently, (ii) they live farther from the nearest bank branch, and (iii) among those who visit branches, they travel farther. These last three findings in combination imply that distance might be an access barrier to branch use by race that is worth further investigation. In this section, we estimate the share of the

Black-White gap in branch use that can be explained by distance. To do so, we use a gravity equation model.

#### 4.3.1 Gravity estimates

A gravity equation is a useful framework to examine how distance sways branch goers to visit certain branches over others. After decades of development in their micro-foundations, carefully specified gravity equations produce coefficient estimates that are consistent with theory, even if no structural model is proposed (Head and Mayer, 2014). Estimates from the gravity model will help us determine the extent to which differences in branch visitation (i.e., the branch use gradient) is due to distance.

Our baseline gravity equation for visitor flows from block group $i$ to branch $j$ in year-month $t$ is

$$\log\left(\text{No. of visitors}_{ijt}\right) = \gamma_{it} + \lambda_{jt} + \beta \log\left(\text{Distance}_{ijt}\right) + \varepsilon_{ijt}, \tag{8}$$

where $\gamma_{it}$ is a block-group by year-month fixed effect, capturing all characteristics of block group $i$ that contribute to residents visiting any branch in year-month $t$; and $\lambda_{jt}$ is a branch by year-month fixed effect, capturing all characteristics of branch $j$ that make it a destination for any block group in year-month $t$. Using origin and destination fixed effects dates back to Harrigan (1996) and has become standard practice in the trade literature instead of using "mass" variables like country GDP, or the number of devices in our setting. Because we estimate a panel, time fixed effects are appended to the cross-sectional fixed effects.

To preserve user privacy, SafeGraph censors data from home block groups having very few visitors. If three or four devices from a block group visits a branch, the number of visitors is rounded up to four. This bottom coding is clearly observable in Figure V Panel A, which presents a binned scatter plot of the log number of visitors from block groups to visited branches by the log distance between the origin and destination. In the panel, all block group-branch pairs are included. A clear negative relation between distance and visitation is visible, but the relation begins to flatten out when the log number of visitors approaches 1.4, which corresponds to 4 visitors. Because the bottom censoring distorts the gravity coefficient estimates, we run two sets of regressions: those including all block group-branch pairs and those that limit pairs with more than four branch visitors.

Table VII presents the gravity regressions under the baseline specification of Eq. (8) and a specification in which racial shares interacted with distance are added as independent variables. From column

(1), in which all block group-branch pairs are included across the country, the coefficient estimate implies that a branch will experience 5.3% fewer visitors per month from a block group for every 1% increase in the block group's distance away. Structural models of commuting flows (e.g., Ahlfeldt et al., 2015) can produce gravity equations in which the loading on distance combines both per-unit travel costs and the elasticity of substitution between destinations (and in our setting, between alternative forms of banking services as well). Therefore, the 5.3% estimate should be thought of as reflecting both of these factors.

When racial shares are interacted with distance in column (2), the distance coefficient remains largely unchanged at -5.6%. The coefficients on the racial interaction terms are small, which suggests that the relation between distance and visitor flows is universal across racial groups. Even so, the coefficient values of the interaction terms are informative. The coefficient on the interaction term with the Black share is positive and precisely estimated. The positive value is consistent with residents from block groups with larger Black population shares having more inelastic demand for banking at branches (and a lower elasticity of substitution with other banking methods). The higher inelasticity suggested in the estimate coincides with responses from the FDIC survey. Among banked Black respondents, nearly 49.9% cite bank tellers and ATMs as their most common bank access method. Among banked White respondents, only 40.9% cite those methods as their most common. Instead, nearly 56.6% cite mobile and online banking.

In Metropolitan core areas, the baseline gravity coefficient estimate in column (3) is -5.1%, virtually unchanged from the national estimate. Adding the interaction terms in column (4) also does not disturb the estimate, as it becomes -5.0%. Overall, the regressions in columns (1)-(4) imply that the gravity estimates between visitor flows and distance are robust across the country and within Metro cores.

Columns (5)-(6) limits the sample to block group-branch pairs with greater than 4 branch visitors. Doing so removes any interference with SafeGraph's bottom coding. In column (5), the coefficient on distance sharply increases in magnitude to -28.3%. Figure V Panel B displays a binned scatter plot that corresponds with this specification. There is a steep negative relation between the number of visitors from block groups to branches and the distance between them. The flattening of the curve is now absent because of the minimum placed on the number of visitors. When racial shares are added as interaction terms in column (6), the coefficient on distance is largely stable at -25.8%. The sign on the interaction term with the Black population share is still positive, though no longer precisely estimated.

Focusing on Metro cores, column (7) presents a coefficient on distance of -31.1%, which is the largest estimated magnitude across all specifications. With racial shares added via interaction terms in column (8), the coefficient changes moderately to -28.4%. Overall, across columns (5)-(8), the estimated gravity coefficients are notably stable cross-country and in Metro cores.

### 4.3.2 Explaining the Black-White gap in branch use

Section 3.1 established that residents from block groups with higher Black population shares take advantage of bank branches relatively less frequently. Indeed, Table III showed that residents from these block groups see around 2.5% fewer bank branch visitors per month in Metro core areas and 5.7% fewer branch visitors cross-country compared to block groups with higher White population shares. The gravity equation coefficient estimates from the previous section can help evaluate the extent to which these two Black-White gaps in branch use are due to distance (i.e., Blacks living comparatively farther away from bank branches than Whites).

The gravity equation's coefficient on distance describes the sensitivity of branch visitor flows to an incremental change in branch distance. The coefficient estimate cross-country of -0.053 in column (1) of Table VII suggests that if a representative branch is located 1% farther away from a representative block group, the number of residents from that block group who travel to that branch will drop by 5.3%. Here, we consider the gravity estimate when all home block group-branch pairs are used (rather than just those pairs with greater than 4 visitors) because the regressions involving branch visitation in Table III included all home block groups.

A simple method to assess the impact of distance on Black residents' branch visitation is to multiply the gravity coefficient estimate by Black residents' distance to bank branches relative to White residents. Measuring relative distance is challenging, as it ought to consider all branches that Black residents might reasonably entertain visiting. Distances to all these branches would then need to be weighted appropriately to form a single relative distance. A sensible alternative is to proxy for this relative distance to all branches with the relative distance to the *nearest* branch. This proxy would align exactly with the true relative distance if all branch goers only visited their nearest branch. Figure III revealed that a very high fraction (roughly 50%) of branch visitors travel to their nearest branch, which supports the viability of the proxy. However, the share is not 100%, which makes the proxy imperfect.

With this caveat in mind, the gravity coefficient estimate of -0.053 can be multiplied with the coefficient on the Black population share from one of the specifications in Table V that regressed the distance to the nearest bank branch on demographic attributes. For visitation across the country, we start with the specification in column (3), which uses our core sample of branches and included controls for income and age. The coefficient on the Black population share in that column is 0.272. Multiplying this value by -0.053 produces -0.0144, which is roughly 25% of the Black-White gap in branch use cross-country. We also consider the expanded set of branches available in the Summary of Deposits to measure distance to the nearest branch. This sample is used in column (7) of Table V, and the coefficient on the Black population share is 0.355. Multiplying that value by the gravity coefficient of -0.053 gives -0.0188, which is roughly 33% of the Black-White gap cross-country. Hence, depending on the set of bank branches considered, we estimate that distance can explain between 25-33% of the national Black-White gap in branch use.

Metro core areas observe the largest differences in proximity to the nearest branch between Blacks and Whites. The gravity coefficient estimate for Metro core areas from Table VII, column (3) is -0.051. For distance to the nearest branch, in column (6) of Table V, the coefficient on the Black population share is 0.354 when the set of branches covers our core sample and only block groups in Metro cores are included. Multiplying this value by the gravity estimate produces -0.018, which is roughly 72% of the Black-White gap in branch use within Metro cores. Expanding the set of bank branches eligible to those in the Summary of Deposits, the coefficient on the Black population share in column (8) is 0.422. Multiplying this value by the gravity coefficient estimate gives -0.0215, which is roughly 86% of the Black-White gap. Hence, we estimate that distance can explain between 75-86% of the Black-White gap in branch use within Metro core areas.

The national and Metro-core gravity estimates are roughly identical, but the Black-White gap in branch use within Metro cores is over twice as large as the gap nationally. Meanwhile, Black residents in Metro cores live relatively farther away from branches compared to Black residents nationally. This greater remoteness from banks explains why distance plays a much greater role in affecting Black residents' branch use in these large urban areas.

# 5 Postal Banking

Having argued that distance can explain sizeable fractions of the Black-White gap in branch use across the country and especially in Metro core areas, we turn next to evaluating a policy proposal that might alleviate distance costs for Black branch goers. In particular, we study postal banking. A Postal

Savings System existed in the United States beginning in 1911, but eventually it was phased out by Congress in 1966 (O'Hara and Easley, 1979; Shaw, 2018). Originally promoted to reach the unbanked, the US Postal Savings System was initially used by non-farming immigrant populations for short-term savings and provided a partial substitute for private banks (Schuster et al., 2020). Re-instituting the Postal Savings System has been a policy proposed by members of Congress (Warren, 2014; Gillibrand, 2021; Sanders, 2021) and parts of academia (Baradaran, 2013; Johnson, 2017).

With our data, we are only in a position to asses how a Postal Banking System—which would extend checking, savings, and possibly credit services to some or all US Post Office branches—might affect branch use by altering the distance between consumers and their (private and public) bank branches. A Postal Banking System could just as easily influence branch use through other channels, such as providing a financial product that is considered more trustworthy than those issued by private banks or enabling economies of scope for consumers to spread out fixed costs of travel; i.e., accessing financial services when dropping off mail (Office of the USPS Inspector General, 2014; Baradaran, 2015). Whichever effects we estimate from postal banking will be limited to the distance channel alone.

With this caveat in mind, we evaluate the influence that postal banking might have on branch use by exploiting the gravity equation estimates from Section 4.3.1. We presume that the coefficients would remain the same if postal banking were introduced. This assumption renders our exercise a partial impact assessment of a postal banking policy that does not account for general equilibrium effects of adding postal banks. Such an exercise is akin to what Head and Mayer (2014) call in the trade literature a "partial trade impact" of a policy change in tariffs.

Because some Post Offices would convert into effective bank branches under a Postal Savings System, the distance between a typical resident and the nearest bank branch would automatically shorten (or remain unchanged) if postal banking were re-introduced. Nationwide, the median distance between the population weighted center of Census block groups and the nearest Post Office is 2.35 miles. In Metro cores, the median distance is 1.99 miles. These two figures are significantly higher than the distances reported earlier between residents and their nearest private bank branches nationwide (1.07 miles) and within Metro cores (0.91 miles).

To measure the extent to which distance to the nearest branch would change under postal banking, we re-run the nearest branch regressions from Table V, but this time, we also include the locations of all Post Office branches that are registered in SafeGraph when calculating distance to the nearest branch. By including Post Office branches together with private bank branches, we implicit assume that both types of banking services would be perfect substitutes.

We identify Post Office branches as all businesses in SafeGraph with NAICS codes equal to 491110 (Postal Services). Selection by this criterion is convenient, but it is possible that not all postal locations chosen are customer-facing (e.g., some facilities might be vehicle maintenance centers or administrative buildings). We therefore provide closer to an upper bound on the effect on branch use via the distance channel, as not all the postal locations we include would likely expand to feature banking services under the policy. One caveat is that SafeGraph likely does not register all Post Office locations in existence, which would have the opposite effect of underestimating the impact of postal banking on distance.

When Post Offices are included in calculating the distance between home block groups and their nearest branch, the mean distance away nationwide declines from 1.98 miles (when only private bank branches are considered) to 1.45 miles, a roughly 26.7% drop. In Metro core areas, the mean distance from the nearest branch declines from 1.21 miles to 1.07 miles, about a 11.5% decline.

Table VIII presents weighted OLS regressions of log distance between home block groups and their nearest branch by demographic attributes when Post Office branches are and are not included in the distance calculations. Odd columns in the table exclude Post locations, whereas even columns include them. Columns (1)-(4) cover all block groups in the US, and columns (5)-(8) limit the sample to Metro core block groups.

In all cases, the coefficient on income increases when Post Office branches are included. For example, when controlling for racial and age shares, columns (3) and (4) indicate that a doubling in median household income is associated with residents living 30.9% farther away from the nearest private bank branch, but 39.9% from the nearest private bank or Post Office branch. The increase is roughly the same for Metro core areas in columns (7) and (8). This result implies that residents of poorer block groups not only live closer to their nearest Post Office branch than residents of richer block groups, but the extent to which poorer residents live closer to their nearest Post Office branch exceeds the extent to which they live closer to their nearest private bank branch. Therefore, when Post Office branches are included in calculating the distance to the nearest bank branch, the income gradient rises.

Columns (3) and (4) suggest that residents from block groups across the country with higher Black shares live relatively farther from their nearest Post Office than residents from block groups with higher

White population shares. Extrapolation from the coefficient on the Black population share suggests that Black residents live 27.2% farther from the nearest private bank branch than a White resident. When Post Office branches are included, the Black-White gap in distance to the nearest branch rises to 28.5%. In Metro core areas, the coefficients in columns (7) and (8) suggest that residents of block groups with higher Black population shares live comparatively closer to their nearest Post Office branches. When only private bank branches are considered, extrapolation from the coefficient on the Black population share implies that Black residents live 35.4% farther from the nearest bank branch than White residents. When Post Office branches are added, the estimate declines to 32.4%. Hence, the Black-White gap in distance to the nearest branch would mildly shrink in Metro core areas if postal banking were established.

Knowing how distance to the nearest branch would alter under postal banking, we next use the gravity equation coefficients to estimate how visitor flows to physical banking services would change under the policy. Across the country, the gravity coefficient on distance was -0.051. As mentioned earlier, the drop in the average distance to bank branches nationwide would be about 27% under postal banking. Multiplying this drop by the national gravity coefficient estimate implies that overall visitor flows to physical banking services would increase by roughly 1.38% per month. This increase can further be analyzed be demographic variables. Starting with income, the change in the income coefficient from column (3) (without postal banks) to (4) (with postal banks) in Table VIII is +0.09 (i.e., 0.399 – 0.309). Multiplying this change by the gravity coefficient and dividing the product by the coefficient on income from column (7) in the branch use regressions from Table III implies that the income gradient on branch use would flatten by roughly 2% nationally $\left(\text{i.e.,}\frac{-0.051\times0.09}{0.244}\right)$ if postal banking were re-instituted. In other words, postal banking would reduce distances to the nearest bank branches and thus be associated with higher visitation rates nationwide, and poorer block groups would observe a 2% greater increase in bank visitation than richer block groups.

Nationally, the estimated change in the Black-White gap in branch use can be calculated in a similar fashion. Multiplying the gravity estimate of -0.051 by the change in the coefficient on the Black population share from columns (3) to (4) in Table VIII when postal banks are added, and then dividing by the Black-White gap in branch use from column (7) in Table III generates an estimated *increase* in the Black-White gap of roughly 1% $\left(\text{i.e.,}\frac{-0.051\times(0.285-0.272)}{-0.057}\right)$. Hence, the decline in distance would benefit branch visits across racial groups, but block groups with higher Black population shares would observe a roughly 1% smaller increase in visitation compared to block groups with higher White population shares.

In Metro core areas, the average distance to the nearest branch under postal banking would decline by roughly 11.5%. Multiplying this change by the gravity coefficient in Metro cores (-0.053) implies that overall visitor flows to bank branches would rise by roughly 0.61% per month in these major urban areas. Repeating the similar computations as earlier to estimate changes in visitor flows within Metro cores by demographic characteristics reveals that the income gradient of branch use in Metro cores would flatten by about 1%, which was similar to the national case $\left(\text{i.e.,} \frac{-0.051\times(0.385-0.325)}{0.254}\right)$. The Black-White gap in branch use within Metro cores would decline by roughly 6% $\left(\text{i.e.,} \frac{-0.051\times(0.324-0.354)}{-0.025}\right)$. This last finding implies that postal banking would have the greatest impact on branch use in Metro core areas rather than nationwide, and its effects would be more powerful in narrowing the Black-White gap in branch use rather than compressing the income gradient.

Nevertheless, our estimates suggest that any encouragement that postal banking might bring to branch use in Metro cores strictly by relaxing distance costs would be modest. The relatively small impact is due to residents in these large urban areas living relatively farther away from Post Office branches than they do from private banks. Postal banking, however, might encourage bank branch or general banking use through other means, such as improving confidence or trust in financial products, which we do not examine in this article.

# 6 Conclusion

We use anonymous location data from millions of mobile devices to infer income and racial patterns in bank branch access and use throughout the United States. A key finding is that residents from poorer block groups across the country are 7.2% less likely to visit a branch in a year. Likewise, residents from block groups with higher Black population shares are 6.2% less likely to visit a branch relative to residents from block groups with higher White population shares. Survey evidence from the FDIC suggests that these groups do not substitute their lower branch visitation rates with higher use of mobile or online banking.

Differences in branch use sharpen in large Metropolitan core areas of the country. These urban cores also observe the highest segregation of branch goers by race and income. High levels of segregation imply that visitors differ not only in their use of branch services, but also in the menu of banks they choose. The Northeast overall, and especially its most urban parts, is more segregated than the South, particularly its most rural areas. Urban and rural differences in bank branch segregation are so stark, that a county which switches from rural to urban jumps from the 10th to 90th percentile in both racial and income segregation.

To partially explain the racial and income gaps in branch use, we study variation in distances to bank branches across the country. Controlling for income, residents from block groups with high Black shares on average live farther from their nearest branch and travel farther when visiting banks. This result implies that farther distances to physical banking services might explain the lower visitation among Black residents.

We implement a gravity equation model to estimate the role of distance in affecting branch use. We estimate that Black residents living farther from bank branches explains roughly 25-33 percent of the Black-White gap in branch use across the country and 72-86 percent of the gap within Metro cores. The national and Metro core gravity estimates are roughly identical, but the Black-White gap in branch use within Metro cores is over twice as large as the gap nationally. Meanwhile, Black residents in Metro cores live relatively farther away from branches compared to Black residents nationally. This greater remoteness from banks explains why distance plays a much greater role in affecting Black residents' branch use in these large urban areas.

Finally, we evaluate a policy of postal banking that adds banking services to Post Office branches. Increasing the number of bank branches would alleviate some distance costs and potentially close some of the Black-White gap in branch use. We analyze the partial impact of the policy, as we fix any general equilibrium effects from Post Office branches offering banking services. We estimate that postal banking would decrease the mean distance to banks in Metro cores by 11.5%. But the policy would close only about 6% of the Black-White gap in bank branch use within these large urban areas. The modest effect is due in part to residents of Metro cores living relatively farther away from Post Office branches than they do from private banks, thus making it difficult for postal banking to overcome the distance barriers.

# References

**Ahlfeldt, Gabriel M, Stephen J Redding, Daniel M Sturm, and Nikolaus Wolf**, "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 2015, *83* (6), 2127–2189.

**Akbar, Prottoy A, Sijie Li, Allison Shertzer, and Randall P Walsh**, "Racial segregation in housing markets and the erosion of black wealth," January 2020. Working paper no. 25805. National Bureau of Economic Research, Cambridge, MA.

**Almagro, Milena, Joshua Coven, Arpit Gupta, and Angelo Orane-Hutchinson**, "Racial disparities in frontline workers and housing crowding during COVID-19: Evidence from geolocation data," February 2021. Unpublished working paper. New York University, New York, NY.

**Ananat, Elizabeth Oltmans**, "The wrong side(s) of the tracks: The causal effects of racial segregation on urban poverty and inequality," *American Economic Journal: Applied Economics*, 2011, *3* (2), 34–66.

**Athey, Susan, Billy A Ferguson, Matthew Gentzkow, and Tobias Schmidt**, "Experienced segregation," July 2020. Working paper no. 27572. National Bureau of Economic Research, Cambridge, MA.

**_ , David Blei, Robert Donnelly, Francisco Ruiz, and Tobias Schmidt**, "Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data," in "AEA Papers and Proceedings," Vol. 108 2018, pp. 64–67.

**Baradaran, Mehrsa**, "It's time for postal banking," *Harvard Law Review Forum*, 2013, *127*, 165–175.

_ , *How the other half banks: Exclusion, exploitation, and the threat to democracy*, Harvard University Press, 2015.

**Beck, Thorsten, Asli Demirguc-Kunt, and Maria Soledad Martinez Peria**, "Reaching out: Access to and use of banking services across countries," *Journal of Financial Economics*, 2007, *85* (1), 234–266.

**_ , Asli Demirgüç-Kunt, and Maria Soledad Martinez Peria**, "Banking services for everyone? Barriers to bank access and use around the world," *World Bank Economic Review*, 2008, *22* (3), 397–430.

**_ , _ , and Patrick Honohan**, "Access to financial services: Measurement, impact, and policies," *World Bank Research Observer*, 2009, *24* (1), 119–145.

**Billings, Stephen B, David J Deming, and Jonah Rockoff**, "School segregation, educational attainment, and crime: Evidence from the end of busing in Charlotte-Mecklenburg," *Quarterly Journal of Economics*, 2014, *129* (1), 435–476.

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

**Blank, Rebecca M and Michael S Barr**, *Insufficient funds: Savings, assets, credit, and banking among low-income households*, Russell Sage Foundation, 2009.

**Card, David, Alexandre Mas, and Jesse Rothstein**, "Tipping and the dynamics of segregation," *Quarterly Journal of Economics*, 2008, *123* (1), 177–218.

**Cargill, Thomas F and Naoyuki Yoshino**, *Postal savings and fiscal investment in Japan: The PSS and the FILP*, Oxford University Press, 2003.

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

**Caskey, John P**, *Fringe banking: Check-cashing outlets, pawnshops, and the poor*, Russell Sage Foundation, 1994.

_ **and Andrew Peterson**, "Who has a bank account and who doesn't: 1977 and 1989," *Eastern Economic Journal*, 1994, *20* (1), 61–73.

**Chen, M Keith and Ryne Rohla**, "The effect of partisanship and political advertising on close family ties," *Science*, 2018, *360* (6392), 1020–1024.

_ **, Judith A Chevalier, and Elisa F Long**, "Nursing home staff networks and COVID-19," *Proceedings of the National Academy of Sciences*, 2021, *118* (1).

_ **, Kareem Haggag, Devin G Pope, and Ryne Rohla**, "Racial disparities in voting wait times: Evidence from smartphone data," *Review of Economics and Statistics*, 2019, pp. 1–27.

**Claessens, Stijn**, "Access to financial services: A review of the issues and public policy objectives," *World Bank Research Observer*, 2006, *21* (2), 207–240.

_ **and Enrico Perotti**, "Finance and inequality: Channels and evidence," *Journal of Comparative Economics*, 2007, *35* (4), 748–773.

**Clotfelter, Charles T**, "Public school segregation in metropolitan areas," *Land Economics*, 1999, pp. 487–504.

**Cook, Lisa D, Maggie EC Jones, David Rosé, and Trevon D Logan**, "The Green Books and the geography of segregation in public accommodations," March 2020. Working paper no. 26819. National Bureau of Economic Research, Cambridge, MA.

**Couture, Victor, Jonathan I Dingel, Allison E Green, Jessie Handbury, and Kevin R Williams**, "Measuring movement and social contact with smartphone data: A real-time application to COVID-19," 2021. Working paper no. 27560. National Bureau of Economic Research, Cambridge, MA.

**Coven, Joshua, Arpit Gupta, and Iris Yao**, "Urban flight seeded the Covid-19 pandemic across the United States," October 2020. Unpublished working paper. New York University, New York, NY.

**Currie, Janet and Patricia B Reagan**, "Distance to hospital and children's use of preventive care: Is being closer better, and for whom?," *Economic Inquiry*, 2003, *41* (3), 378–391.

**Cutler, David M and Edward L Glaeser**, "Are ghettos good or bad?," *Quarterly Journal of Economics*, 1997, *112* (3), 827–872.

**Davis, Donald R, Jonathan I Dingel, Joan Monras, and Eduardo Morales**, "How segregated is urban consumption?," *Journal of Political Economy*, 2019, *127* (4), 1684–1738.

**Dekle, Robert and Jonathan Eaton**, "Agglomeration and land rents: Evidence from the prefectures," *Journal of Urban Economics*, 1999, *46* (2), 200–214.

**Duncan, Otis Dudley and Beverly Duncan**, "A methodological analysis of segregation indexes," *American Sociological Review*, 1955, *20* (2), 210–217.

**Echenique, Federico and Roland G Fryer Jr.**, "A measure of segregation based on social interactions," *Quarterly Journal of Economics*, 2007, *122* (2), 441–485.

**Fogli, Alessandra and Veronica Guerrieri**, "The end of the American Dream? Inequality and segregation in US cities," August 2019. Working paper no. 26143. National Bureau of Economic Research, Cambridge, MA.

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

**Gillibrand, Kristen**, "Senators Gillibrand And Sanders, Representatives Ocasio-Cortez, Pascrell, and Kaptur call on Congress to implement Postal Banking pilot programs," Apr 2021.

**Goolsbee, Austan and Chad Syverson**, "Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020," *Journal of Public Economics*, 2021, *193*, 104311.

**Hamrick, Karen S, David Hopkins et al.**, "The time cost of access to food–Distance to the grocery store as measured in minutes," *International Journal of Time Use Research*, 2012, *9* (1), 28–58.

**Harrigan, James**, "Openness to trade in manufactures in the OECD," *Journal of International Economics*, 1996, *40* (1-2), 23–39.

**Head, Keith and Thierry Mayer**, "Gravity equations: Workhorse, toolkit, and cookbook," in "Handbook of International Economics," Vol. 4 2014, pp. 131–195.

**Hogarth, Jeanne M and Kevin H O'Donnell**, "Being accountable: A descriptive study of unbanked households in the US," in "Proceedings of the Association for Financial Counseling and Planning Education" 1997, pp. 58–67.

**__ , Christoslav E Anguelov, and Jinhook Lee**, "Who has a bank account? Exploring changes over time, 1989–2001," *Journal of Family and Economic Issues*, 2005, *26* (1), 7–30.

**Honohan, Patrick**, "Measuring microfinance access: Building on existing cross-country data," May 2005. Unpublished working paper. World Bank, Washington, DC.

**Iceland, John**, "Beyond black and white: Metropolitan residential segregation in multi-ethnic America," *Social Science Research*, 2004, *33* (2), 248–271.

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

— **and Melissa Scopilliti**, "Immigrant residential segregation in US metropolitan areas, 1990–2000," *Demography*, 2008, *45* (1), 79–94.

**Inagami, Sanae, Deborah A Cohen, Brian Karl Finch, and Steven M Asch**, "You are where you shop: Grocery store locations, weight, and neighborhoods," *American Journal of Preventive Medicine*, 2006, *31* (1), 10–17.

**Jahn, Julius, Calvin F Schmid, and Clarence Schrag**, "The measurement of ecological segregation," *American Sociological Review*, 1947, *12* (3), 293–303.

**Johnson, Randall K**, "How the United States Postal Service (USPS) Could Encourage More Local Economic Development," *Chicago Kent Law Rev.*, 2017, *92*, 593–615.

**Kain, John F**, "Housing segregation, negro employment, and metropolitan decentralization," *Quarterly Journal of Economics*, 1968, *82* (2), 175–197.

**King, Gary**, *A Solution to the Ecological Inference Problem*, Princeton University Press, 1997.

— **, Martin A Tanner, and Ori Rosen**, *Ecological Inference: New Methodological Strategies*, Cambridge University Press, 2004.

**Kreindler, Gabriel E and Yuhei Miyauchi**, "Measuring commuting and economic activity inside cities with cell phone records," February 2021. Working paper no. 28516. National Bureau of Economic Research, Cambridge, MA.

**Logan, John R, Andrew Foster, Hongwei Xu, and Wenquan Zhang**, "Income Segregation: Up or down, and for whom?," *Demography*, 2020, *57* (5), 1951–1974.

**Logan, Trevon D and John M Parman**, "Segregation and homeownership in the early twentieth century," *American Economic Review*, 2017, *107* (5), 410–14.

**Lucas, Robert E and Esteban Rossi-Hansberg**, "On the Internal Structure of Cities," *Econometrica*, 2002, *70* (4), 1445–1476.

**Macartney, Hugh and John D Singleton**, "School boards and student segregation," *Journal of Public Economics*, 2018, *164*, 165–182.

**Massey, Douglas S and Nancy A Denton**, "The dimensions of residential segregation," *Social Forces*, 1988, *67* (2), 281–315.

**Miyauchi, Yuhei, Kentaro Nakajima, and Stephen J Redding**, "Consumption access and agglomeration: Evidence from smartphone data," February 2021. Working paper no. 28497. National Bureau of Economic Research, Cambridge, MA.

**Nicholl, Jon, James West, Steve Goodacre, and Janette Turner**, "The relationship between distance to hospital and patient mortality in emergencies: An observational study," *Emergency Medicine Journal*, 2007, *24* (9), 665–668.

**Office of the USPS Inspector General**, "Providing non-bank financial services for the underserved," 2014.

**O'Hara, Maureen and David Easley**, "The Postal Savings System in the Depression," *Journal of Economic History*, 1979, *39* (3), 741–753.

**Owens, Ann, Sean F Reardon, and Christopher Jencks**, "Income segregation between schools and school districts," *American Educational Research Journal*, 2016, *53* (4), 1159–1197.

**Reardon, Sean F**, "Measures of income segregation," *Unpublished Working Paper. Stanford Center for Education Policy Analysis*, 2011.

**_ and Glenn Firebaugh**, "Measures of multigroup segregation," *Sociological methodology*, 2002, *32* (1), 33–67.

**_ and Kendra Bischoff**, "Income inequality and income segregation," *American Journal of Sociology*, 2011, *116* (4), 1092–1153.

**_ , _ , Ann Owens, and Joseph B Townsend**, "Has income segregation really increased? Bias and bias correction in sample-based segregation estimates," *Demography*, 2018, *55* (6), 2129–2160.

**Rosenthal, Stuart S and William C Strange**, "Evidence on the nature and sources of agglomeration economies," in "Handbook of Regional and Urban Economics," Vol. 4 2004, pp. 2119–2171.

**Sanders, Bernie**, "Fair banking for all," August 2021.

**Schuster, Steven Sprick, Matthew Jaremski, and Elisabeth Ruth Perlman**, "An empirical history of the US Postal Savings System," *Social Science History*, 2020, *44* (4), 667–696.

**Shaw, Christopher W**, "'Banks of the people': The life and death of the US Postal Savings System," *Journal of Social History*, 2018, *52* (1), 121–152.

**Sunstein, Cass**, *Republic.com*, Princeton University Press, February 2001.

**Theil, Henri**, *Statistical decomposition analysis: With applications in the social and administrative sciences*, Vol. 14, North-Holland Publishing Company, 1972.

**Warren, Elizabeth**, "The big benefits of postal service banking," Jul 2014.

**Wilson, William Julius**, *The truly disadvantaged: The inner city, the underclass, and public policy*, University of Chicago Press, 1987.

**Yantzi, Nicole, Mark W Rosenberg, Sharon O Burke, and Margaret B Harrison**, "The impacts of distance to hospital on families with a child with a chronic condition," *Social Science & Medicine*, 2001, *52* (12), 1777–1791.

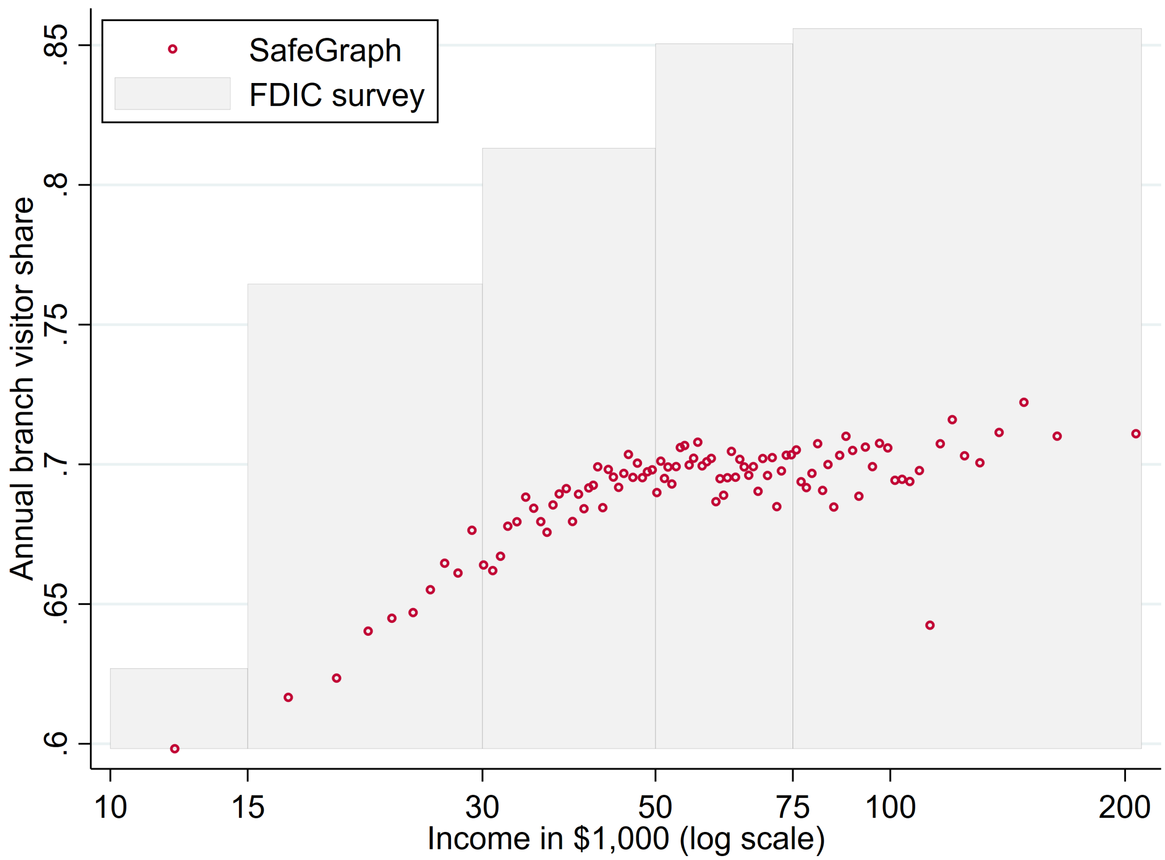
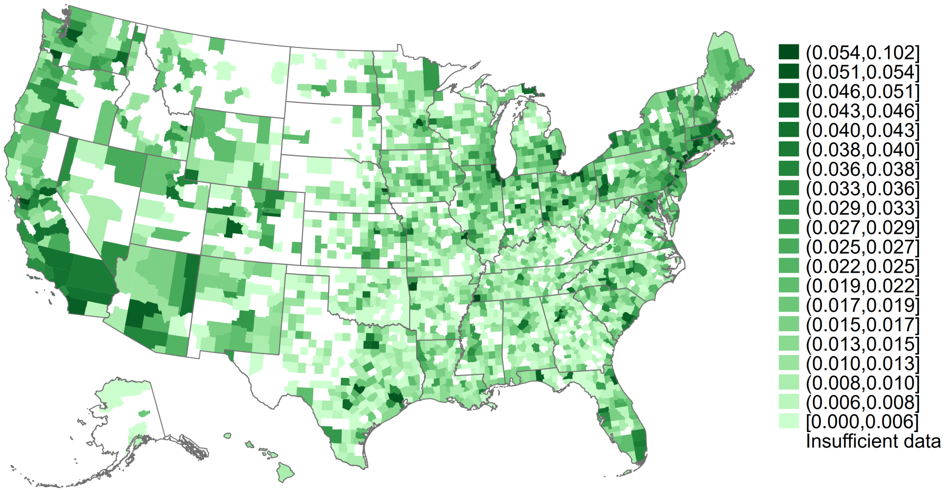



**FIGURE I**

BANK BRANCH VISITOR SHARE BY INCOME (FDIC SURVEY & SAFEGRAPH)

*Notes.* The figure presents a binned scatter plot of the shares of residents that visit bank branches according to household income, comparing survey responses to actual visitors. Survey responses are from the "2019 FDIC Survey of Household Use of Banking and Financial Services", conducted in June 2019. Both banked and unbanked respondents are included. Actual branch visitor shares are from our core SafeGraph sample between July 2018 and June 2019; i.e., only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) with visitor data whose brands are also listed in the 2019 vestige of the FDIC's Summary of Deposits. The survey responses (represented as grey bars) are the shares of households in the five income categories of the survey that acknowledged visiting a bank branch within the past 12 months. The width of a bar corresponds to the income range of its category, except for the first income category (<$15,000) and the last category (>$75,000), where we extend the width of the bars to the nearest thousand dollars that also includes the reaches of the SafeGraph data. The corresponding SafeGraph values are the annual shares of mobile devices recorded in SafeGraph that visit a bank branch over the same 12-month period. To compute these annual shares of branch visitors, we first divide a month's total branch visitors by the total recorded mobile devices in the month for each home Census block group. This ratio gives an estimate of the probability that a device from each home block group visits a bank branch at least once during the month. Let this estimated branch visitor probability for block group $j$ in month $t$ be denoted $p_{j,t}$. Not every block group has a visitor probability each month, so, let $k_j$ denote the number of months for which block group $j$ has observations. The annual branch visitor share $s_j$ for block group $j$ is computed as $s_j = 1 - \prod_{t=1}^{12/k_j} (1 - p_{j,t})^{12/k_j}$. A binned scatter plot of these calculated annual visitor shares by household income overlays the bars from the survey responses. Household income is measured as median household income from the 2019 5-year American Community Survey. To construct this binned scatter plot, we divide the horizontal axis into 100 equal-sized (percentile) bins and plot the mean annual share of visitors to a bank branch versus the mean household income within each bin.

**(A) Income Segregation by County**




**(B) Racial Segregation by County**

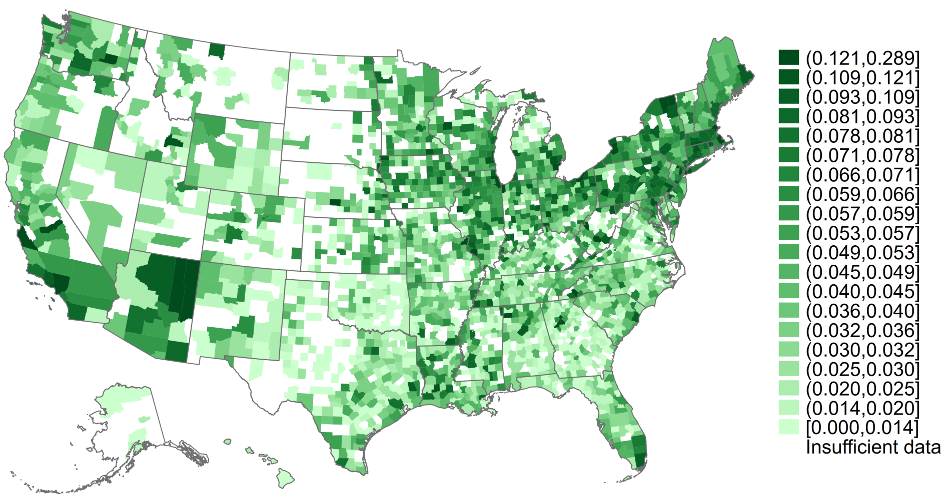



**FIGURE II**

GEOGRAPHY OF BANK BRANCH SEGREGATION

*Notes.* The figure presents heatmaps of income and racial segregation at US 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 2019 vestige of the FDIC's Summary of Deposits. The income entropy segregation index values portrayed in Panel A are estimates of Eq. (7), made using the procedure described in Reardon (2011). The racial entropy segregation index values portrayed in Panel B are estimates of Eq. (5). 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.

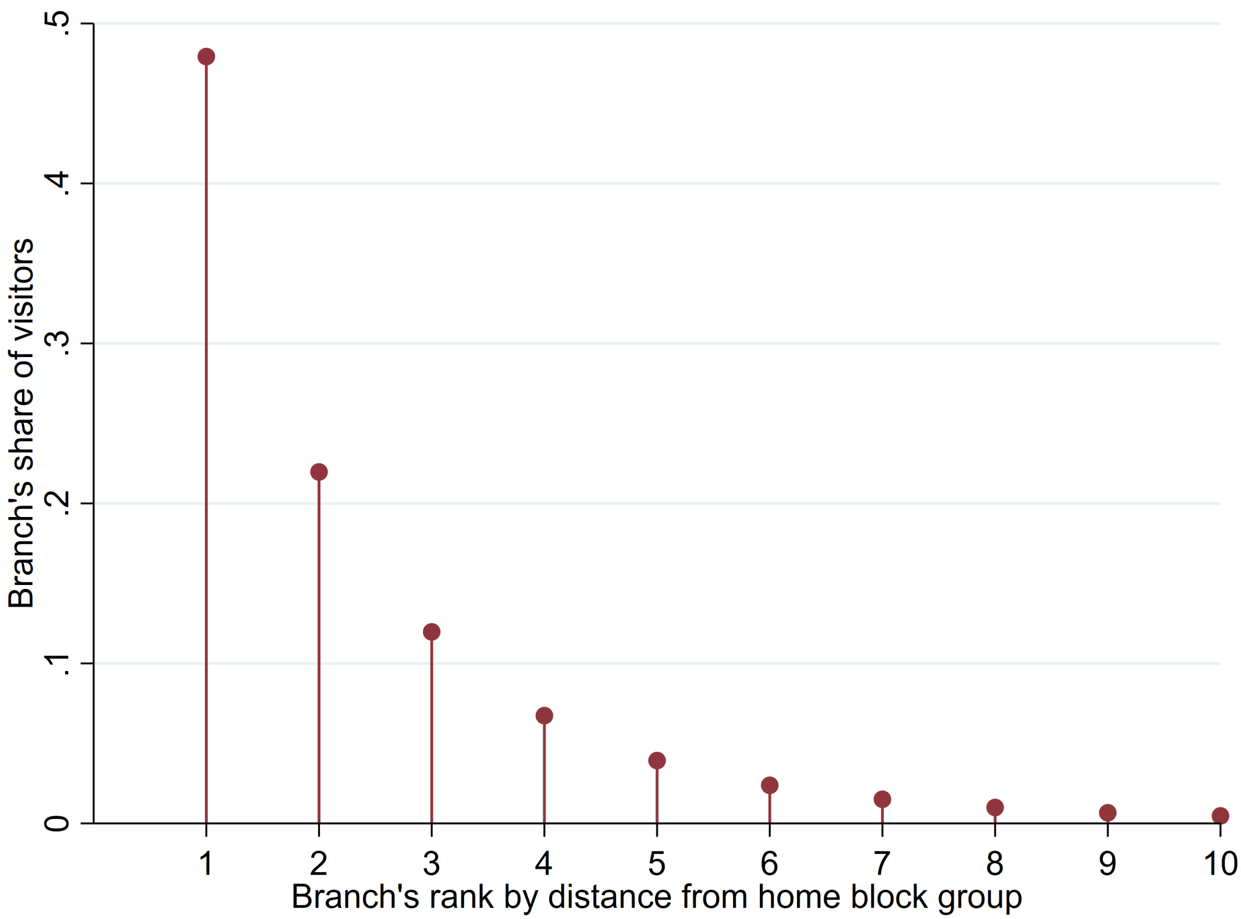



**FIGURE III**

SHARE OF VISITORS BY BANK BRANCH'S RANKED DISTANCE FROM HOME

*Notes.* The figure presents the shares of visitors from a Census block group that travel to a bank branch according to the branch's ranked distance from home. Visitor information is from our core SafeGraph sample; 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 2019 vestige of the FDIC's Summary of Deposits. To construct this distribution of shares, we start from the perspective of a single Census block group. For this block group, we compute the distances between its population-weighted center and the latitude-longitude points of all branches visited by the block group's residents. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance in miles (see Footnote 17). We then rank visited branches by their distances from the block group's center. (We use an integer rank starting from one instead of a percentile rank.) We repeat this exercise for all block groups that are home to branch visitors in our core sample. We then sum across all block groups and months the number of visitors to each rank and divide each sum by the total number of visitors to all branches throughout all months. The empirical distribution presented in the figure is each ranked branch's share of visitors.

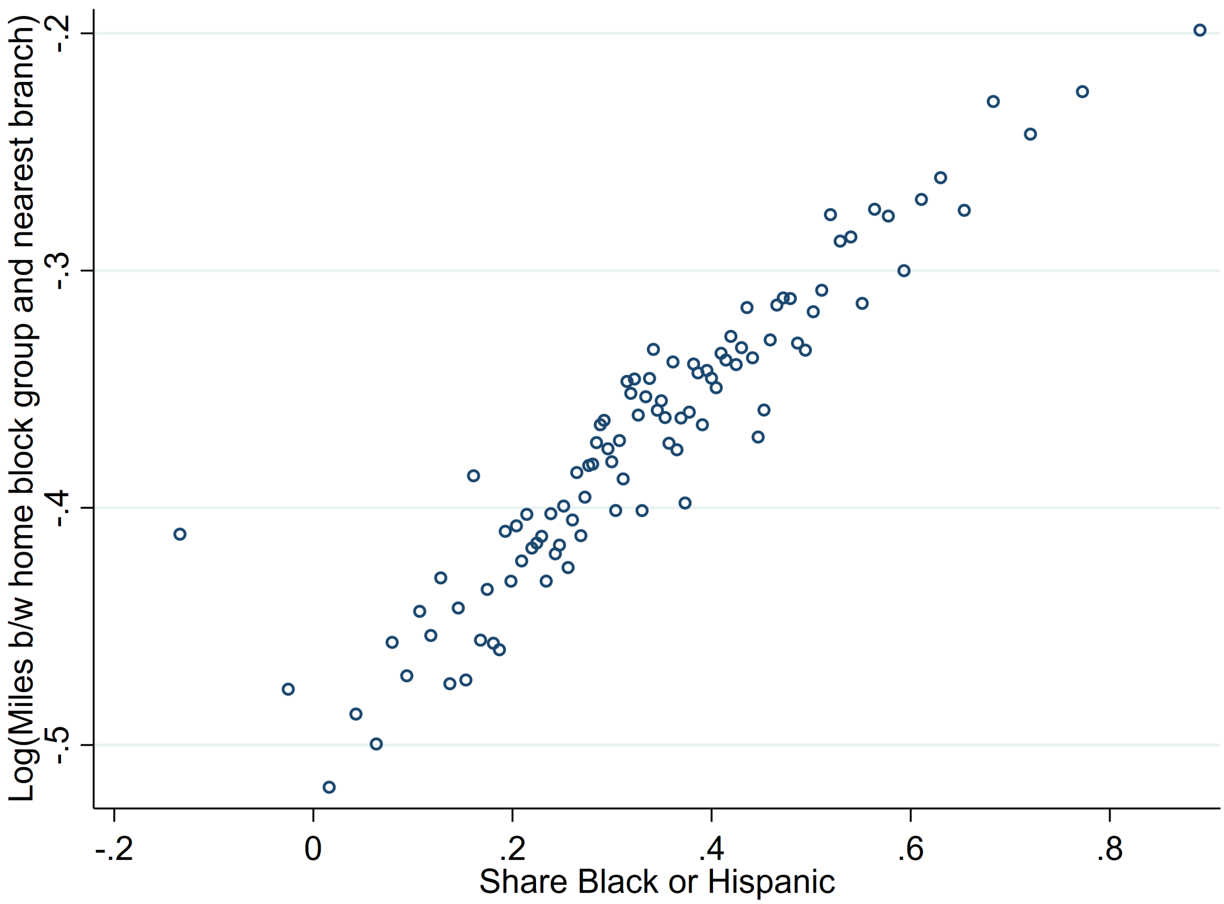
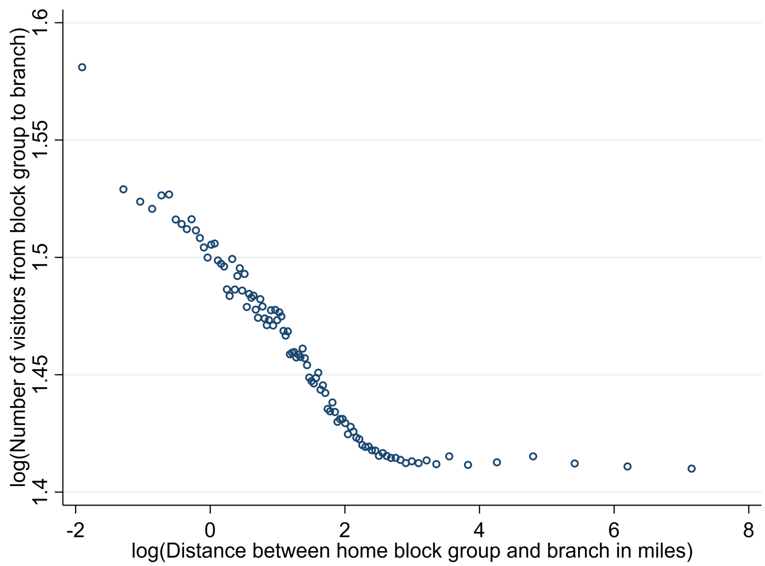



**FIGURE IV**

DIST. TO NEAREST BANK BRANCH IN METRO CORES BY BLACK OR HISPANIC SHARE

*Notes.* The figure presents a binned scatter plot the log distance in miles between home block groups in Metropolitan area cores and their nearest bank branches according to block groups' population shares of Black or Hispanic residents. All branch locations in Metro cores that present in the 2018 or 2019 vestiges of the FDIC's Summary of Deposits are included in the calculations. Distance is computed from the population-weighted center of a block group to the nearest branch. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance (see Footnote 17). Population shares of Black or Hispanic residents are from the 5-year American Community Survey. Log distances are residualized by county fixed effects, year-month fixed effects, population shares of Asian and Other Races, age shares, and log median household income. Black or Hispanic population shares are residualized by the same set of variables. To construct the binned scatter plot, we divide the residualized Black or Hispanic shares into 100 equal-sized (percentile) bins. We then calculate the mean of the residualized log distances and the mean of the residualized share of Black or Hispanic residents within each bin. Finally, we add back the unconditional mean of the log distances and the unconditional mean of the Black or Hispanic shares to re-scale values. These two latter objects are then plotted.

**(A) All Block Group-Branch Pairs**




**(B) Block Group-Branch Pairs with >4 Visitors, with Fixed Effects**

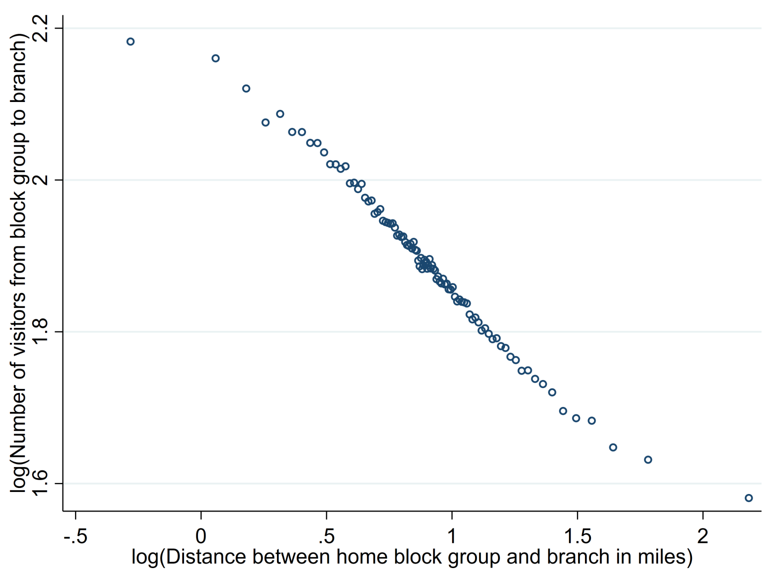



FIGURE V

NUMBER OF VISITORS FROM BLOCK GROUP TO BANK BRANCH BY DISTANCE

*Notes.* The figure presents binned scatter plots of the log number of visitors from home block groups to bank branches according to the log mile distance between the block groups and branches. Visitor information is from our core SafeGraph sample. Distance is computed from the population-weighted center of a block group to the branch. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance (see Footnote 17). Panel A presents the raw data and includes all block group-branch pairs, including those with visitor counts <4 that are bottom-coded to 4 by SafeGraph. Panel B only includes block group-branch pairs with greater than 4 visitors. In that panel, the log numbers of visitors are residualized by block group-year-month fixed effects and branch-year-month fixed effects. The log distances are residualized by the same set of fixed effects. To construct the binned scatter plots, we divide the x-axis values into 100 equal-sized (percentile) bins. We then calculate the mean of the y-axis values and the mean of the x-axis values within each bin. In addition, for Panel B we add back the unconditional mean of the log numbers of visitors and the unconditional mean of the log distances to re-scale values. These two objects are plotted.

**TABLE I**
DESCRIPTIVE STATISTICS - CORE SAFEGRAPH SAMPLE

| | Mean | Std. Dev | P10 | P25 | P50 | P75 | P90 | N |
|---|---|---|---|---|---|---|---|---|
| No. of Visits | 67 | 180 | 6 | 14 | 35 | 78 | 147 | 919,076 |
| No. of Visitors | 40 | 94 | 5 | 10 | 23 | 48 | 90 | 919,076 |
| Med. Dist. from Home (mi) | 5 | 16 | 2 | 3 | 4 | 6 | 9 | 822,569 |
| Med. Dwell Time (min) | 49 | 102 | 6 | 7 | 9 | 30 | 152 | 919,076 |
| Device Type - iOS | 52% | | | | | | | 19,238,792 |
| Device Type - Android | 46% | | | | | | | 17,207,356 |

*Notes.* The table reports descriptive statistics of key variables related to bank branch visits. All values are based on our core sample of branch locations, which consists of 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 2019 vestige of the FDIC's Summary of Deposits. Data are monthly, at the branch level, and range from January 2018 - December 2019. *No. of Visits* is the total number of visits to a typical bank branch in a month. *No. of Visitors* is the total number of visitors (i.e., mobile devices) to a typical branch in a month. *Med. Dist. from Home (mi)* is the median distance in miles that visitors travel to a branch from home (among visitors whose home is identified). *Med. Dwell Time (min)* is the median amount of time in minutes that visitors stay at a branch. *Device Type* is the fraction of total branch visitors using Google Android vs. Apple iOS mobile devices. The number of observations *N* used in the first four rows is the total number of branch-year-months. The number of observations used in the last two rows is the total number of mobile devices with device-type information over the core sample period.

**TABLE II**

DEMOGRAPHIC ATTRIBUTES OF CORE SAFEGRAPH SAMPLE VS. FDIC SOD

| | Core Sample $\hat{\mu}_1$ $\hat{\sigma}_1$ | SOD $\hat{\mu}_2$ $\hat{\sigma}_2$ | Diff $\hat{\mu}_1 - \hat{\mu}_2$ (se) | | Core Sample $\hat{\mu}_1$ $\hat{\sigma}_1$ | SOD $\hat{\mu}_2$ $\hat{\sigma}_2$ | Diff $\hat{\mu}_1 - \hat{\mu}_2$ (se) |
|---|---|---|---|---|---|---|---|
| N branch | 51,369 | 86,374 | -35,005 | < HS | 0.104 | 0.104 | -0.000 |
| | | | | | 0.075 | 0.075 | (0.000) |
| White | 0.799 | 0.805 | -0.006 | HS degree | 0.260 | 0.263 | -0.003 |
| | 0.184 | 0.183 | (0.001) | | 0.103 | 0.106 | (0.001) |
| Black | 0.103 | 0.095 | 0.008 | Some college | 0.199 | 0.198 | 0.001 |
| | 0.153 | 0.146 | (0.001) | | 0.052 | 0.054 | (0.000) |
| Asian | 0.046 | 0.047 | -0.001 | College degree | 0.297 | 0.295 | 0.002 |
| | 0.076 | 0.082 | (0.000) | | 0.090 | 0.091 | (0.001) |
| Hispanic | 0.109 | 0.106 | 0.003 | > College | 0.140 | 0.140 | 0.000 |
| | 0.155 | 0.156 | (0.001) | | 0.098 | 0.100 | (0.001) |
| Homeowner | 0.645 | 0.643 | 0.002 | Unemp rate | 0.050 | 0.050 | 0.000 |
| | 0.168 | 0.173 | (0.001) | | 0.026 | 0.028 | (0.000) |
| Age 15-34 | 0.188 | 0.190 | -0.001 | HH income ($) | 70,338 | 69,802 | 536 |
| | 0.089 | 0.093 | (0.001) | | 29,657 | 30,052 | (166) |
| Age 35-54 | 0.350 | 0.347 | 0.004 | In poverty | 0.127 | 0.129 | -0.002 |
| | 0.069 | 0.069 | (0.000) | | 0.082 | 0.084 | (0.000) |
| Age 55-64 | 0.195 | 0.195 | -0.000 | Urban | 0.815 | 0.787 | 0.028 |
| | 0.038 | 0.039 | (0.000) | | 0.295 | 0.319 | (0.002) |
| Age 65+ | 0.267 | 0.269 | -0.002 | Home value ($) | 286,881 | 290,086 | -3,205 |
| | 0.086 | 0.085 | (0.000) | | 243,368 | 254,504 | (1,383) |

*Notes.* The table compares demographic characteristics of geographic areas represented in our core sample of bank branches with geographic areas represented by all bank branches in the 2019 vestige of the FDIC's Summary of Deposits (SOD). Our core sample consists only of 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 2019 vestige of the FDIC's Summary of Deposits. Demographic characteristics in the table are taken from the 2019 5-year American Community Survey and are averaged at the level of the Census Bureau's zip code tabulation areas (ZCTA). The only non-demographic variable is N branch, which is the number of branches in the core sample and the SOD. In columns (1), (2), (4), and (5), the first row of each demographic attribute is its sample mean across ZCTAs, whereas the second row is the sample standard deviation. In columns (3) and (6), the first row is the difference in sample means between the core sample and the SOD, whereas the second row is the the heteroskedasticity-robust standard error of the estimated difference between the two sample means.

## TABLE III
BANK BRANCH VISITATION BY DEMOGRAPHIC ATTRIBUTES

| Dep. var.: | Visitor Ratio×100 | | | | | | log(No. of Visitors) | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log(Income) | 0.290 | 1.018 | 0.622 | 0.909 | 1.195 | 0.792 | 0.244 | 0.254 |
| | (0.048) | (0.044) | (0.045) | (0.054) | (0.050) | (0.052) | (0.006) | (0.007) |
| Black | -1.696 | -0.803 | -0.536 | -0.735 | -0.426 | -0.257 | -0.057 | -0.025 |
| | (0.102) | (0.110) | (0.111) | (0.111) | (0.121) | (0.123) | (0.014) | (0.016) |
| Asian | -4.830 | -0.528 | 0.625 | 1.102 | -0.157 | 0.874 | 0.375 | 0.406 |
| | (0.214) | (0.192) | (0.192) | (0.198) | (0.200) | (0.201) | (0.025) | (0.026) |
| Other | -8.666 | -1.695 | 0.095 | -0.002 | -2.094 | -0.076 | -0.029 | -0.016 |
| | (0.363) | (0.351) | (0.355) | (0.354) | (0.451) | (0.461) | (0.042) | (0.056) |
| Hispanic | -1.515 | -0.712 | -0.069 | -0.333 | -0.383 | 0.151 | -0.100 | -0.092 |
| | (0.104) | (0.130) | (0.137) | (0.163) | (0.144) | (0.153) | (0.020) | (0.022) |
| Age <15 | | | 5.506 | 5.153 | | 5.794 | 0.934 | 1.000 |
| | | | (0.296) | (0.296) | | (0.342) | (0.036) | (0.042) |
| Age 35-54 | | | 4.925 | 4.497 | | 4.579 | 0.613 | 0.635 |
| | | | (0.280) | (0.283) | | (0.333) | (0.032) | (0.039) |
| Age 55-64 | | | 7.653 | 6.990 | | 7.261 | -0.002 | -0.110 |
| | | | (0.321) | (0.329) | | (0.395) | (0.037) | (0.045) |
| Age 65+ | | | 6.161 | 6.079 | | 6.429 | 0.198 | 0.179 |
| | | | (0.239) | (0.240) | | (0.277) | (0.027) | (0.030) |
| HS degree | | | | 0.973 | | | | |
| | | | | (0.301) | | | | |
| Some college | | | | 0.780 | | | | |
| | | | | (0.295) | | | | |
| College degree | | | | -0.198 | | | | |
| | | | | (0.263) | | | | |
| > College | | | | -1.356 | | | | |
| | | | | (0.310) | | | | |
| log(No. of Devices) | | | | | | | 0.527 | 0.511 |
| | | | | | | | (0.008) | (0.009) |
| Constant | 7.775 | | | | | | | |
| | (0.532) | | | | | | | |
| Observations | 3,134,728 | 3,134,720 | 3,134,720 | 3,134,663 | 2,246,239 | 2,246,239 | 3,134,720 | 2,246,239 |
| Adjusted $R^2$ | 0.012 | 0.186 | 0.193 | 0.194 | 0.174 | 0.182 | 0.538 | 0.536 |
| Sample | Core | Core | Core | Core | MC | MC | Core | MC |
| County FE | | O | O | O | O | O | O | O |
| Year-month FE | | O | O | O | O | O | O | O |

*Notes.* Each column reports coefficients from a multivariate, weighted OLS regression with standard errors clustered at the Census-block-group level reported in parentheses. One observation is a block group per month per year in the sample period from January 2018 - December 2019. Observations are weighted by the number of mobile devices residing in the block groups in the year-months. Dependent variable observations are based on our core sample of branch locations, which consists of 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 2019 vestige of the FDIC's Summary of Deposits. Demographic independent variable observations are population-based decimal shares from the 2019 5-year American Community Survey. Income is household income. The dependent variable in columns (1)-(6) is defined as follows. Let $n_{ijt}$ denote the number of devices residing in block group $i$ that visit branch $j$ in month $t$. Let $h_{it}$ denote the number of devices residing in block group $i$ in month $t$. The dependent variable is then $(\Sigma_j n_{ijt}/h_{it}) \times 100$. The dependent variable in columns (7) and (8) is the natural logarithm of $\sum_j n_{ijt}$. The variable *log(No. of Visitors)* in columns (7) and (8) is the natural logarithm of $h_{it}$. Columns (1)-(4) and column (7) include all block groups for which we have branch visitor data, whereas columns (5), (6), and (8) restrict the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitan area core). The omitted demographic groups are non-Hispanic Whites, age range 15-34, and education less than High School degree.

**TABLE IV**
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) |

*Notes.* 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 2019 vestige of the FDIC's Summary of Deposits. The dissimilarity index in this paper is an estimate of Eq. 1, as described in Section 3.2.1. The two groups in the dissimilarity index computation are Black and non-Black. The racial entropy index is an estimate of Eq. (5), as described in Section 3.2.2. The four racial groups used in computing the racial entropy index are Hispanics, non-Hispanic White, non-Hispanic Blacks, and others. The income entropy index is an estimate of Eq. (7), as described in Section 3.2.3. 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 V
DISTANCE FROM NEAREST BANK BRANCH BY DEMOGRAPHIC ATTRIBUTES

| Dep. var.: | log(Distance b/w home block group and nearest branch) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log(Income) | 0.069 | 0.425 | 0.309 | 0.559 | 0.441 | 0.325 | 0.364 | 0.365 |
| | (0.008) | (0.007) | (0.008) | (0.010) | (0.008) | (0.008) | (0.008) | (0.009) |
| Black | -0.602 | 0.392 | 0.272 | 0.095 | 0.520 | 0.354 | 0.355 | 0.422 |
| | (0.016) | (0.017) | (0.017) | (0.018) | (0.018) | (0.019) | (0.018) | (0.019) |
| Asian | -2.414 | -0.766 | -0.672 | -0.496 | -0.662 | -0.607 | -0.597 | -0.546 |
| | (0.044) | (0.034) | (0.034) | (0.033) | (0.035) | (0.034) | (0.034) | (0.034) |
| Other | -0.158 | -0.348 | -0.367 | -0.451 | -0.389 | -0.488 | -0.260 | -0.462 |
| | (0.076) | (0.063) | (0.063) | (0.062) | (0.078) | (0.078) | (0.064) | (0.077) |
| Hispanic | -0.792 | 0.102 | -0.031 | -0.535 | 0.224 | 0.019 | 0.054 | 0.089 |
| | (0.015) | (0.019) | (0.020) | (0.023) | (0.020) | (0.022) | (0.021) | (0.022) |
| Age <15 | | | 1.668 | 1.343 | | 1.903 | 1.929 | 2.279 |
| | | | (0.053) | (0.051) | | (0.059) | (0.057) | (0.063) |
| Age 35-54 | | | 0.737 | 0.388 | | 0.523 | 0.851 | 0.613 |
| | | | (0.053) | (0.052) | | (0.060) | (0.055) | (0.062) |
| Age 55-64 | | | 1.815 | 1.293 | | 1.417 | 2.123 | 1.677 |
| | | | (0.058) | (0.057) | | (0.067) | (0.060) | (0.069) |
| Age 65+ | | | 0.261 | 0.117 | | 0.259 | 0.438 | 0.461 |
| | | | (0.042) | (0.041) | | (0.047) | (0.043) | (0.048) |
| HS degree | | | | -0.166 | | | | |
| | | | | (0.044) | | | | |
| Some college | | | | -0.606 | | | | |
| | | | | (0.046) | | | | |
| College degree | | | | -1.100 | | | | |
| | | | | (0.042) | | | | |
| > College | | | | -1.566 | | | | |
| | | | | (0.052) | | | | |
| Constant | -0.297 | | | | | | | |
| | (0.094) | | | | | | | |
| Observations | 3,134,728 | 3,134,720 | 3,134,720 | 3,134,663 | 2,246,239 | 2,246,239 | 3,134,720 | 2,246,239 |
| Adjusted $R^2$ | 0.094 | 0.426 | 0.440 | 0.454 | 0.241 | 0.265 | 0.434 | 0.318 |
| Sample | Core | Core | Core | Core | Core-MC | Core-MC | SOD | SOD-MC |
| Year-month FE | | O | O | O | O | O | O | O |
| County FE | | O | O | O | O | O | O | O |
| RUCA FE | | O | O | O | | | O | |

*Notes.* Each column reports coefficients from a multivariate, weighted OLS regression with standard errors clustered at the Census-block-group level reported in parentheses. One observation is a block group per month per year in the sample period from January 2018 - December 2019. Observations are weighted by the number of mobile devices residing in the block groups in the year-months. Columns (1)-(6) use our core sample of branch locations, which consists of 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 2019 vestige of the FDIC's Summary of Deposits (SOD). Columns (7) and (8) use all branch locations presented in the 2018 or 2019 vestiges of the SOD. Demographic independent variable observations are population-based decimal shares from the 2019 5-year American Community Survey. Income is household income. The dependent variable is the log distance from the population-weighted center of a block group to the nearest branch. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance in miles (see Footnote 17). Columns (1)-(4) and column (7) include all block groups for which we have branch visitor data, whereas columns (5), (6), and (8) restrict the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitian area core). The omitted demographic groups are non-Hispanic Whites, age range 15-34, and education less than High School degree.

## TABLE VI
### AVG. DIST. TRAVELED TO BANK BRANCHES BY DEMOGRAPHIC ATTRIBUTES

| Dep. var.: | Weighted average log(Distance b/w home block group and visited branches) | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| log(Income) | 0.058 | 0.171 | 0.182 | 0.261 | 0.169 | 0.188 |
| | (0.007) | (0.009) | (0.008) | (0.011) | (0.010) | (0.010) |
| Black | -0.272 | 0.091 | 0.033 | -0.020 | 0.135 | 0.063 |
| | (0.017) | (0.019) | (0.019) | (0.019) | (0.021) | (0.021) |
| Asian | -1.303 | -0.630 | -0.730 | -0.709 | -0.575 | -0.703 |
| | (0.033) | (0.033) | (0.034) | (0.034) | (0.034) | (0.034) |
| Other | 0.180 | -0.093 | -0.297 | -0.319 | -0.013 | -0.321 |
| | (0.059) | (0.063) | (0.065) | (0.065) | (0.080) | (0.083) |
| Hispanic | -0.438 | -0.194 | -0.292 | -0.488 | -0.148 | -0.278 |
| | (0.014) | (0.021) | (0.022) | (0.024) | (0.023) | (0.024) |
| Age <15 | | | -0.355 | -0.462 | | -0.344 |
| | | | (0.062) | (0.062) | | (0.072) |
| Age 35-54 | | | -0.353 | -0.462 | | -0.458 |
| | | | (0.054) | (0.054) | | (0.062) |
| Age 55-64 | | | -0.104 | -0.268 | | -0.431 |
| | | | (0.057) | (0.058) | | (0.067) |
| Age 65+ | | | -0.839 | -0.903 | | -0.924 |
| | | | (0.046) | (0.046) | | (0.052) |
| HS degree | | | | -0.164 | | |
| | | | | (0.049) | | |
| Some college | | | | -0.376 | | |
| | | | | (0.048) | | |
| College degree | | | | -0.517 | | |
| | | | | (0.046) | | |
| > College | | | | -0.473 | | |
| | | | | (0.054) | | |
| Constant | 1.049 | | | | | |
| | (0.081) | | | | | |
| Observations | 3,134,728 | 3,134,720 | 3,134,720 | 3,134,663 | 2,246,239 | 2,246,239 |
| Adjusted $R^2$ | 0.018 | 0.122 | 0.124 | 0.125 | 0.044 | 0.047 |
| Sample | Core | Core | Core | Core | MC | MC |
| Year-month FE | | O | O | O | O | O |
| County FE | | O | O | O | O | O |
| RUCA FE | | O | O | O | | |

*Notes.* Each column reports coefficients from a multivariate, weighted OLS regression with standard errors clustered at the Census-block-group level reported in parentheses. One observation is a block group per month per year in the sample period from January 2018 - December 2019. Observations are weighted by the number of mobile devices residing in the block groups in the year-months. Dependent variable observations are based on our core sample of branch locations, which consists of 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 2019 vestige of the FDIC's Summary of Deposits (SOD). Demographic independent variable observations are population-based decimal shares from the 2019 5-year American Community Survey. Income is household income. The dependent variable is the weighted average log distance from the population-weighted center of a block group to all branches visited by residents of that block group. Each branch's weight is its share of visitors from the block group. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance in miles (see Footnote 17). Columns (1)-(4) include all block groups for which we have branch visitor data, whereas columns (5) and (6) restrict the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitan area core). The omitted demographic groups are non-Hispanic Whites, age range 15-34, and education less than High School degree.

**TABLE VII**

GRAVITY EQUATIONS

| Dep. var.: | log(No. of visitors from block group $i$ to branch $j$ in year-month $t$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| $\log(\text{Distance}_{ijt})$ | -0.053<br>(0.001) | -0.056<br>(0.001) | -0.051<br>(0.001) | -0.050<br>(0.001) | -0.283<br>(0.007) | -0.258<br>(0.009) | -0.311<br>(0.009) | -0.284<br>(0.013) |
| $\log(\text{Distance}_{ijt})$ $\times$ Black | | 0.014<br>(0.002) | | 0.004<br>(0.002) | | 0.026<br>(0.029) | | 0.014<br>(0.036) |
| $\log(\text{Distance}_{ijt})$ $\times$ Asian | | 0.013<br>(0.005) | | -0.004<br>(0.005) | | -0.434<br>(0.092) | | -0.400<br>(0.101) |
| $\log(\text{Distance}_{ijt})$ $\times$ Other | | 0.030<br>(0.008) | | 0.037<br>(0.010) | | 0.016<br>(0.119) | | 0.035<br>(0.167) |
| $\log(\text{Distance}_{ijt})$ $\times$ Hispanic | | -0.005<br>(0.002) | | -0.012<br>(0.002) | | -0.025<br>(0.020) | | -0.019<br>(0.024) |
| Observations | 5,627,180 | 5,625,696 | 4,210,214 | 4,209,361 | 276,624 | 276,598 | 198,054 | 198,034 |
| Adjusted $R^2$ | 0.104 | 0.105 | 0.088 | 0.088 | 0.381 | 0.383 | 0.402 | 0.404 |
| Sample | Core | Core | MC | MC | Core | Core | MC | MC |
| >4 only | | | | | O | O | O | O |
| Fixed Effects | O | O | O | O | O | O | O | O |

*Notes.* Each column reports coefficients from an unweighted OLS regression with standard errors clustered at the Census-block-group level reported in parentheses. The regressions estimate visitor flows from block group $i$ to branch $j$ in year-month $t$ according to the gravity equation:

$$\log(\text{No. of visitors}_{ijt}) = \gamma_{it} + \lambda_{jt} + \beta \log(\text{Distance}_{ijt}) + \varepsilon_{ijt},$$

where $\gamma_{it}$ is a block-group by year-month fixed effect, and $\lambda_{jt}$ is a branch by year-month fixed effect. Dependent variable observations are based on our core sample of branch locations, which consists of 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 2019 vestige of the FDIC's Summary of Deposits. Independent variable observations are the log distances from the population-weighted center of block groups to visited bank branches (odd columns) and the log distances interacted with population-based racial shares from the 2019 5-year American Community Survey (even columns). Centers of population are from the 2010 Census, and we use the haversine formula to compute distance in miles (see Footnote 17). Columns (1), (2), (5), and (6) include all block groups for which we have branch visitor data, whereas columns (3), (4), (7), and (8) restrict the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitan area core).

**TABLE VIII**
DIST. FROM NEAREST BANK OR USPS BRANCH BY DEMOGRAPHIC ATTRIBUTES

| Dep. var.: | log(Distance b/w home block group and nearest branch) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| With USPS? | No | Yes | No | Yes | No | Yes | No | Yes |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| log(Income) | 0.425 | 0.506 | 0.309 | 0.399 | 0.441 | 0.494 | 0.325 | 0.385 |
| | (0.007) | (0.007) | (0.008) | (0.007) | (0.008) | (0.008) | (0.008) | (0.008) |
| Black | 0.392 | 0.379 | 0.272 | 0.285 | 0.520 | 0.474 | 0.354 | 0.324 |
| | (0.017) | (0.017) | (0.017) | (0.017) | (0.018) | (0.018) | (0.019) | (0.018) |
| Asian | -0.766 | -0.581 | -0.672 | -0.468 | -0.662 | -0.507 | -0.607 | -0.440 |
| | (0.034) | (0.032) | (0.034) | (0.031) | (0.035) | (0.033) | (0.034) | (0.032) |
| Other | -0.348 | -0.355 | -0.367 | -0.322 | -0.389 | -0.391 | -0.488 | -0.455 |
| | (0.063) | (0.061) | (0.063) | (0.061) | (0.078) | (0.075) | (0.078) | (0.075) |
| Hispanic | 0.102 | 0.108 | -0.031 | 0.015 | 0.224 | 0.203 | 0.019 | 0.023 |
| | (0.019) | (0.018) | (0.020) | (0.019) | (0.020) | (0.019) | (0.022) | (0.020) |
| Age <15 | | | 1.668 | 1.580 | | | 1.903 | 1.857 |
| | | | (0.053) | (0.052) | | | (0.059) | (0.058) |
| Age 35-54 | | | 0.737 | 0.699 | | | 0.523 | 0.488 |
| | | | (0.053) | (0.047) | | | (0.060) | (0.053) |
| Age 55-64 | | | 1.815 | 1.779 | | | 1.417 | 1.349 |
| | | | (0.058) | (0.054) | | | (0.067) | (0.062) |
| Age 65+ | | | 0.261 | 0.383 | | | 0.259 | 0.351 |
| | | | (0.042) | (0.039) | | | (0.047) | (0.044) |
| Observations | 3,134,720 | 3,134,720 | 3,134,720 | 3,134,720 | 2,246,239 | 2,246,239 | 2,246,239 | 2,246,239 |
| Adjusted $R^2$ | 0.426 | 0.361 | 0.440 | 0.376 | 0.241 | 0.266 | 0.265 | 0.289 |
| Sample | Core | Core | Core | Core | Core-MC | Core-MC | Core-MC | Core-MC |
| Year-month FE | O | O | O | O | O | O | O | O |
| County FE | O | O | O | O | O | O | O | O |
| RUCA FE | O | O | O | O | | | | |

*Notes.* Each column reports coefficients from a multivariate, weighted OLS regression with standard errors clustered at the Census-block-group level reported in parentheses. One observation is a block group per month per year in the sample period from January 2018 - December 2019. Observations are weighted by the number of mobile devices residing in the block groups in the year-months. Odd columns use our core sample of bank branch locations, which consists of 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 2019 vestige of the FDIC's Summary of Deposits. Even columns add businesses in SafeGraph with NAICS codes equal to 491110 (Postal Services). Demographic independent variable observations are population-based decimal shares from the 2019 5-year American Community Survey. Income is household income. The dependent variable is the log distance from the population-weighted center of a block group to the nearest branch. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance in miles (see Footnote 17). Columns (1)-(4) include all block groups for which we have branch visitor data, whereas columns (5)-(8) restrict the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitan area core). The omitted demographic groups are non-Hispanic Whites, age range 15-34, and education less than High School degree.

Online Appendix to

# Banking Across America: Distance and Branch Use*

Jung Sakong† Alexander K. Zentefis‡

September 1, 2021

*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)

# A Mobility Dataset Construction

Here, we supply background information on SafeGraph data and a detailed explanation of how we construct our core mobile device sample.

## A.1 SafeGraph

We use two of SafeGraph's primary datasets: Core Places and Patterns. Both datasets have information on millions of points-of-interest (POIs) in the United States, which SafeGraph defines as "specific location[s] where consumers can spend money and/or time."[1] Locations such as restaurants, grocery stores, parks, museums and hospitals are included, but not residential homes or apartment buildings.

The Core Places dataset provides the location name (e.g., Salinas Valley Ford Lincoln), brand (e.g., Ford), six-digit NAICS code, latitude and longitude coordinates, address, phone number, hours open, when the location opened, and when SafeGraph began tracking the location. SafeGraph describes creating this dataset using thousands of diverse sources. We use the January 2021 version of the Core Places dataset, which was the most up-to-date and accurate as of the time of our analysis.

The Patterns dataset contains information on visitors to different locations. A visitor is identified via his or her mobile device, and one device is treated as one visitor. SafeGraph collects this information from third-party mobile application developers. Through these mobile applications, SafeGraph gathers a device's advertisement identifier, the latitude and longitude coordinates of the device at a designated time, and the horizontal accuracy of the geographic coordinates.[2] In this data set, SafeGraph aggregates the visitor data and provides several bits of information, including the number of visits and unique visitors to a POI during a specified date range, the median distance from home traveled by visitors, the median dwell time spent at the POI, and the number of visitors using Apple's iOS or Google's Android operating system. The Patterns dataset is backfilled to reflect the Core Places from the January 2021 version.

Most importantly for us, the Patterns dataset contains the Census block groups of visitors in an aggregated form. Specifically, it includes the number of visitors to a POI whose home is in a Census block group.

Using an algorithm, SafeGraph determines a visitor's home location at the level of a Census block group. Briefly, the algorithm starts by clustering GPS signals from a device during the nighttime hours between 6pm - 7am local time. The Census block group with the most clusters is recorded as the device's potential home location for the day. SafeGraph reviews the previous six weeks of daily home locations and identifies the most frequent one as the device's home Census block group. This home location applies for the device over the next thirty days, at which point the home location is updated. New devices that appear in the panel require at least five days of data before they are eligible to have their home locations configured. Finally, SafeGraph computes a confidence score for each device's calculated home block group. Only high confidence home locations are included; otherwise, the home is classified as unknown.[3]

[1] See the SafeGraph Places Manual and Data Guide for more details.
[2] See the SafeGraph Privacy Policy for more details.
[3] Full details of the algorithm are found here: Home Identification Algorithm.

## A.2 FDIC Summary of Deposits

To construct our mobility dataset, we rely on branch location information from the Federal Deposit Insurance Corporation (FDIC). Bank location data are from the 2019 vestige of the FDIC's Summary of Deposits (SOD).[4] We rely on the SOD to confirm that branch locations we use from SafeGraph belong to actual depository institutions, instead of other financial institutions that SafeGraph might mistakenly label as a "bank," but do not take deposits, such as an investment advisory firm.

## A.3 Mobility Dataset Construction

Our mobile device data set can be thought of as consisting of two components: (i) a set of locations and (ii) consumer movement to those locations. We call those two components "places" and "visits." In our case, the places and visits are specific to bank branches. SafeGraph is our only source of visits data, and so, we rely on it exclusively.

Places data, on the other hand, are available in both SafeGraph and the SOD. Before we detail how we make use of both sources, we first need to introduce *placekey*, which is a crucial way we identify a place.

### A.3.1 Placekey

Placekey is a free, standardized identifier for physical locations. It supplants a location's address and latitude-longitude geocode with a unique identifier. Using this identifier overcomes the challenge of linking locations by addresses that are spelled differently (e.g., 1215 Third Street, Suite 10 vs. 1215 3rd St., #10) or by latitude-longitude geocodes that differ slightly but refer to the same place.

A business's placekey consists of two parts (called "What" and "Where"), and it is written as What@Where. The What component encodes an address and a point-of-interest. The point-of-interest piece adjusts if a new business opens at the same address of a previous business that closed. For example, if a bank branch closed, but its building converted into a bakery, the two businesses would share the same address, but different points-of-interest, and therefore, they would be assigned different placekeys.

The Where component consists of a unique character sequence. It encodes a hexagonal region on the surface of the Earth based on the latitude and longitude of the business. The hexagon contains the centroid of the business, and the Where component is the full encoding of the hexagon. To make Placekey concrete, consider the Chase branch at 1190 S. Elmhurst Rd. in Mount Prospect, IL 60056. This branch's placekey is 223-222@5sb-8gg-jn5.[5]

### A.3.2 Choosing the Set of Places

Both SOD and SafeGraph have bank branch locations. SafeGraph locations are already identified by their placekeys. We generate placekeys for the SOD locations using Placekey's free API.

[4] FDIC SOD data are located here: SOD.

[5] Additional technical information about Placekey can be found in their white paper located here: Placekey White Paper.

To construct an accurate and comprehensive set of places, we take advantage of place information in SafeGraph and the SOD. The *quality* of SafeGraph places is higher than those in the SOD. Often, an address in SOD has an invalid placekey, and a Google Maps search confirms that no physical place exists at that address. (The place's absence is not due to a branch closing.) A higher quality set of places from SafeGraph should come at little surprise, as the success of the company's business relies in part on providing highly accurate place information.

On the other hand, the *quantity* of places is higher in the SOD than in SafeGraph. In SafeGraph, bank branches are classified by their 6 digit NAICS codes (522110 for Commercial Banking, 522120 for Savings Institutions, and 551111 for Offices of Bank Holding Companies). The number of places in SafeGraph under these categories is less than the number of branches in the SOD.

So that we can link places information to patterns information, all places we analyze must be included in SafeGraph. For example, a branch in the SOD that is not part of SafeGraph whatsoever has no visits information to study. But we can use place information from the SOD to choose the set of places from SafeGraph that balances quality and quantity. Doing so constructs our core sample of branches, which we define next.

Our **core sample** of branches includes only SafeGraph places with brands that are included in the SOD. In the SOD, the field `CERT` identifies a unique banking institution. We rely on this field to select the list of unique banks, and we use the the field `LOCATION_NAME` to label a bank brand name in SafeGraph. For example, Wells Fargo & Company and SunTrust Banks, Inc. are two bank brands with locations in the SOD. All Wells Fargo and SunTrust Bank places in SafeGraph would be included, and their locations would be identified by SafeGraph's placekeys for them. All SOD locations (and their placekeys) are ignored, as they tended to be less reliable that SafeGraph's.

# B Income Segregation Computational Steps

This section presents the steps to compute the income entropy segregation indices of Section 3.2.3 in the text. 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

$$IS = 2\ln 2 \int_0^1 E(p) H(p)\, dp, \tag{9}$$

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

$$E(p) = p \ln\left(\frac{1}{p}\right) + (1-p)\ln\left(\frac{1}{1-p}\right). \tag{10}$$

## B.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.

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.

## B.2 Step 1: Calculate $E(p_k) \equiv E_k$ for 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 US, though the same logic applies for the county analysis we present in the text. We start by dropping all home block groups that have zero population according to the ACS.

Suppose the country has $N$ branches. 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 explain in text for estimating the share of all branch visitors in the country who are part of a particular race group. (See Section 3.2.2.) There, we used the notation $\pi_s$ 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 equation (10), the entropy for this percentile is

$$E(p_k) \equiv E_k = p_k \ln\left(\frac{1}{p_k}\right) + (1-p_k)\ln\left(\frac{1}{1-p_k}\right). \tag{11}$$

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

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

Here, we perform the same calculation for entropy, 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 Section 3.2.2.) 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).

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

The entropy index aggregates information across branches in the country. It is calculated 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. Online Figure A.3 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 equation (9) looks like. At least in this month, among branch visitors in the US, income segregation is seen to monotonically increase.

## B.5 Step 4: Estimate the function $H(p)$ in equation (9)

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 Online Figure A.3, 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 equation (11). 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 1/2, where $H_k$ is weighted most.

The choice of polynomial order is at the discretion of the researcher, and 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 plotted the 95% confidence intervals around each point estimate. (Obtaining the standard error of the estimate is described below). The plot is provided in Online Figure A.4. 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, again, 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 $V$.

## B.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 $\delta_m$ can be evaluated as

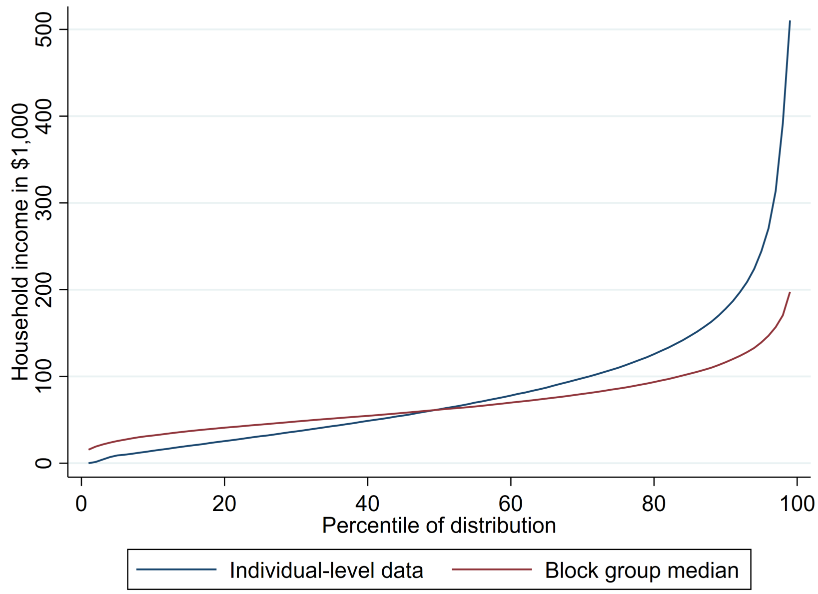
$$\delta_m = \frac{2}{(2+m)^2} + 2\sum_{n=0}^{m} \frac{(-1)^{m-n}\binom{m}{n}}{(m-n+2)^2}, \tag{12}$$

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 5.

The 5 values for $\delta_m$ 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 $\mathrm{Var}\left(\hat{IS}\right) = \Delta' V \Delta$, which we use to compute the 95% confidence intervals in Online Figure A.4.

# C Supplemental Figures and Tables

**(A) Household Income**




**(B) Black Population Share**

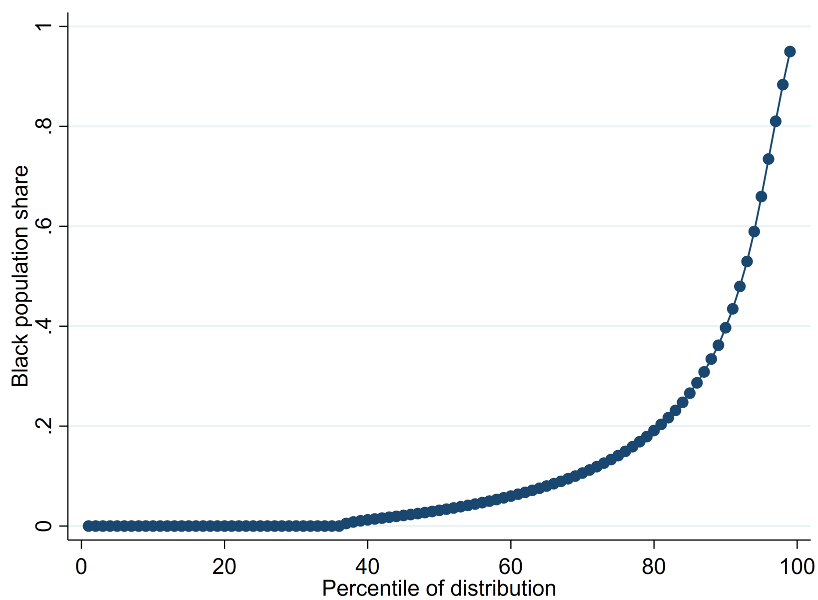



**FIGURE A.1**

DISTRIBUTION OF BLOCK-GROUP DEMOGRAPHIC ATTRIBUTES

*Notes.* The figure presents the percentiles of the distributions for US household income and black population shares. Panel A gives the percentiles for the individual-level household income distribution and the distribution of median household income at the level of Census block groups. Panel B gives black population shares by block group. Data are from the 5-year American Community Survey. The individual-level data was accessed through IPUMS and represents a 5% random sample of the population.

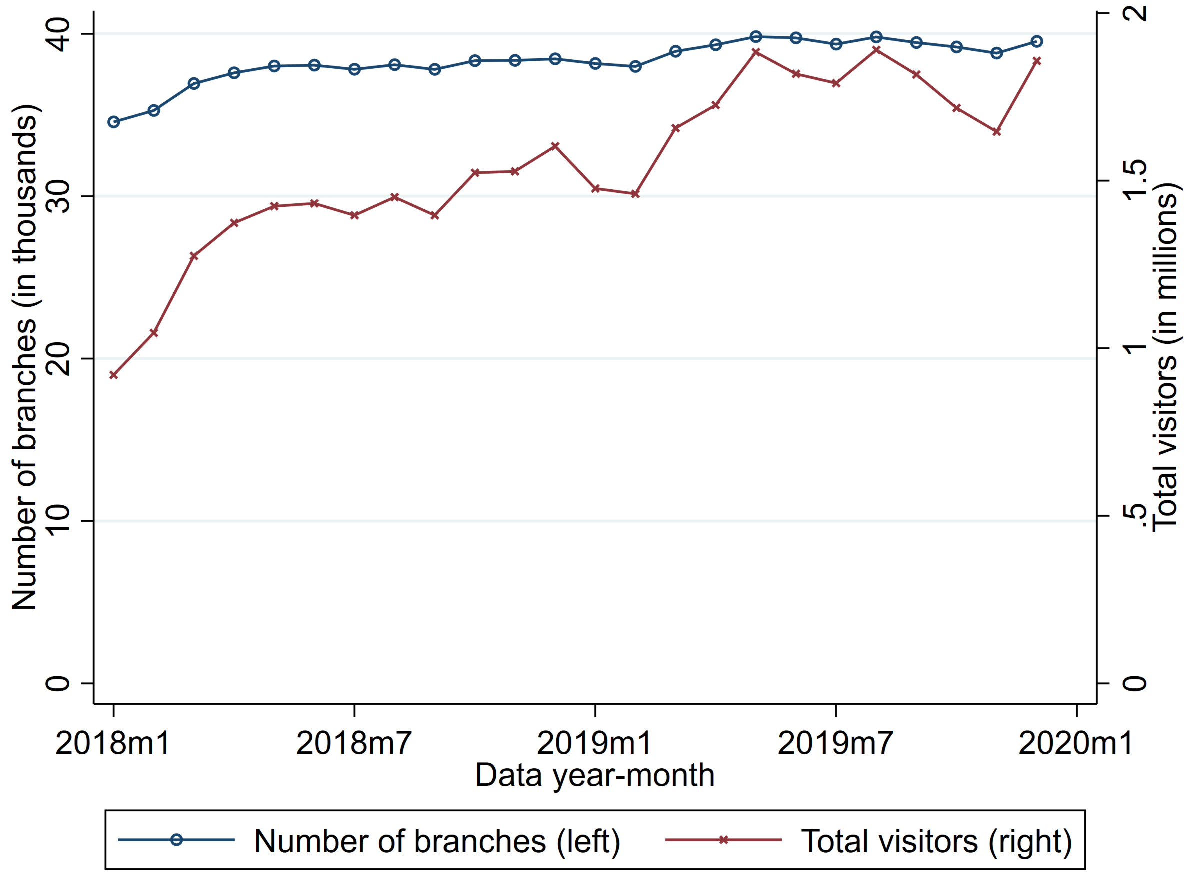



**FIGURE A.2**

NUMBER OF BANK BRANCHES AND BRANCH VISITORS - CORE SAMPLE

*Notes.* The figure presents the number of bank branches and number of branch visitors each year-month in 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 2019 vestige of the FDIC's Summary of Deposits.

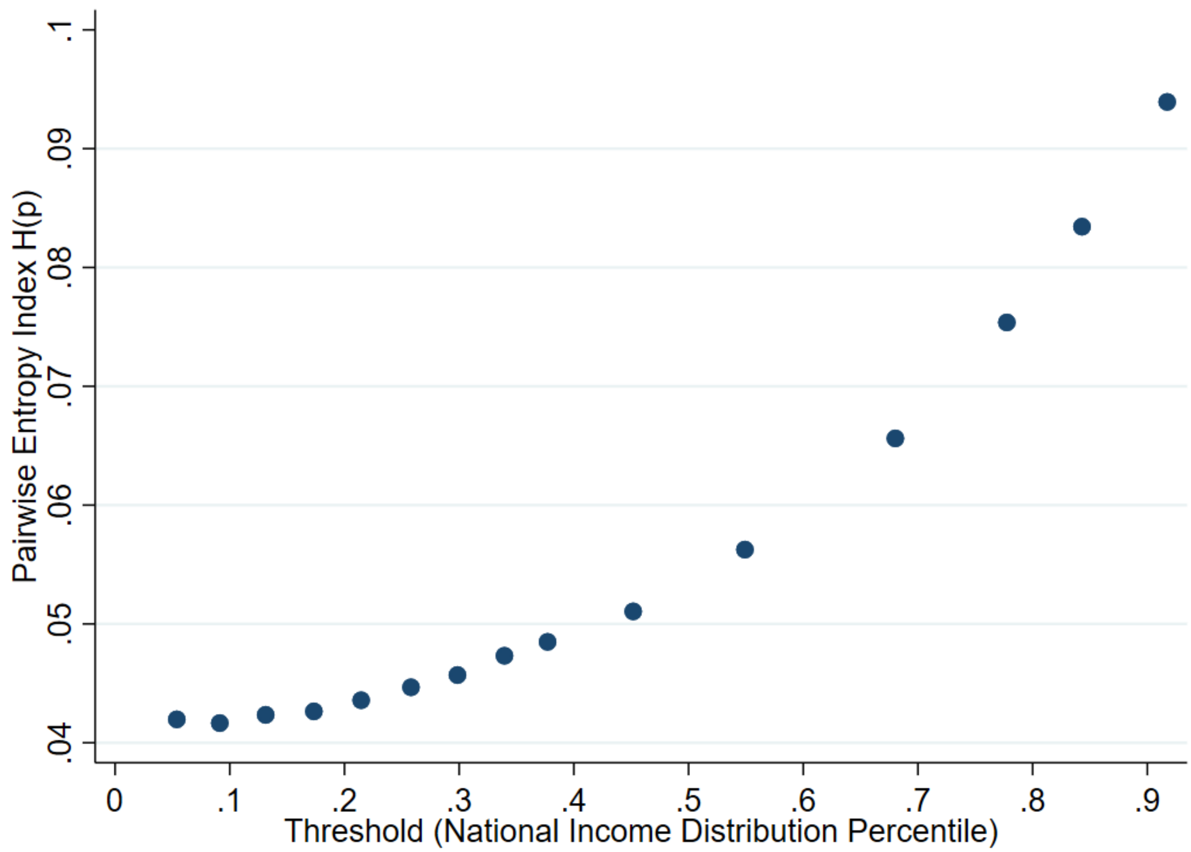



**FIGURE A.3**

PAIRWISE INCOME SEGREGATION PROFILES - SEPT. 2019

*Notes.* 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 2019 vestige of the FDIC's Summary of Deposits. The pairwise income segregation profiles are the 15 values of $H_k$, calculated using the steps described in Online Appendix B. 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.

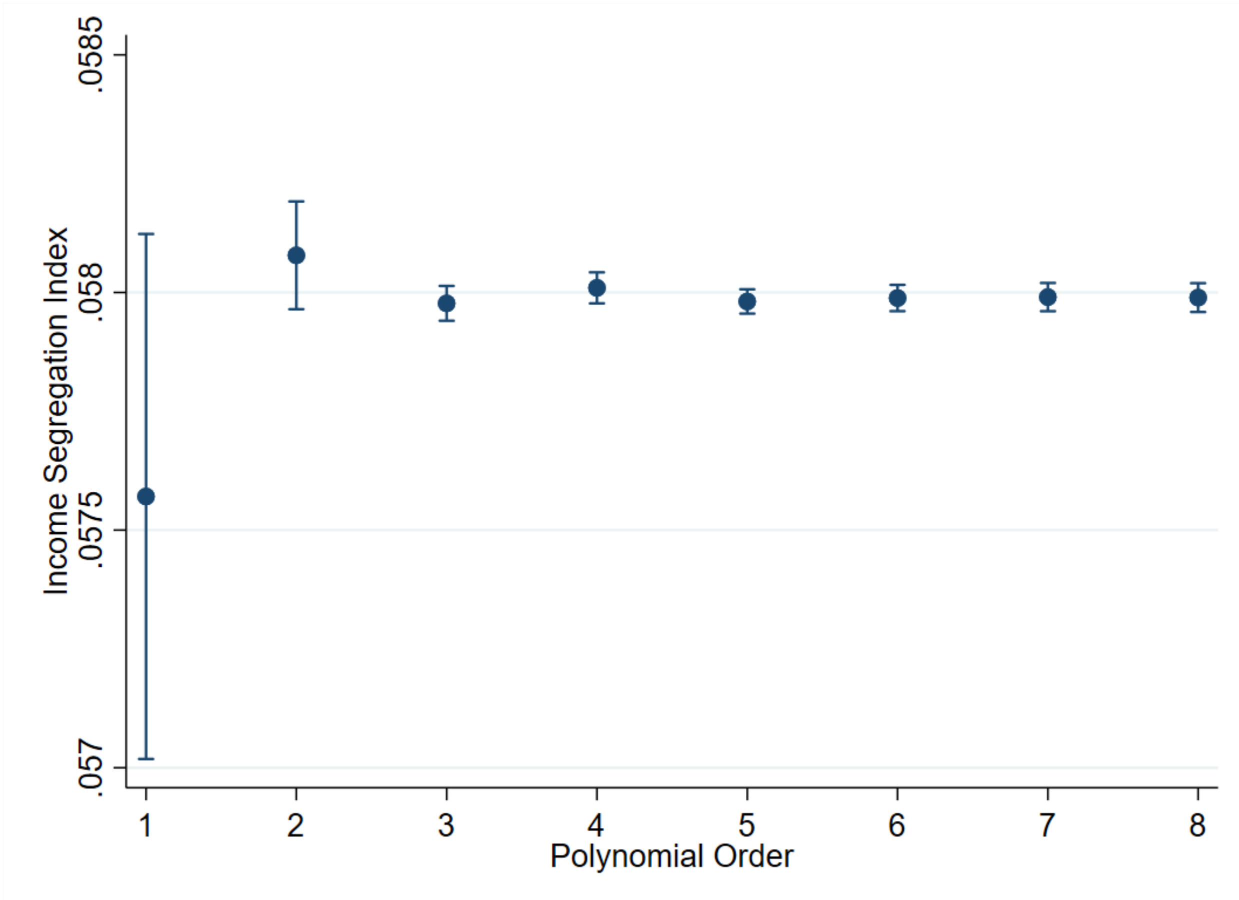
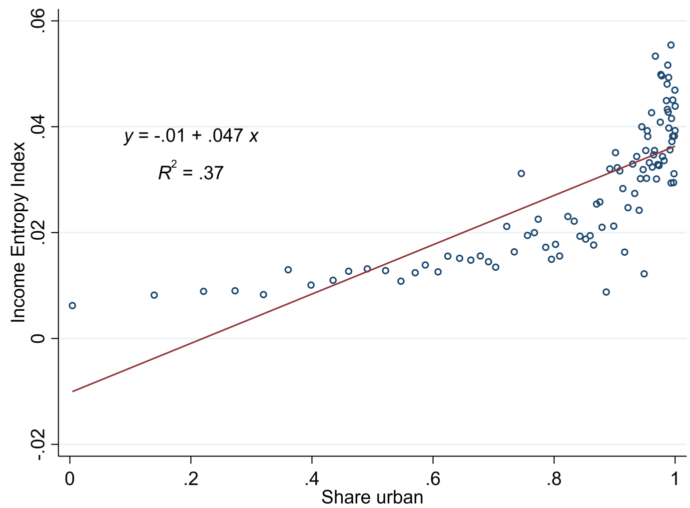



**FIGURE A.4**

ESTIMATED INCOME SEGREGATION BY POLYNOMIAL ORDER - SEPT. 2019

*Notes.* 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 2019 vestige of the FDIC's 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 Online Appendix B. The method for computing the standard errors for the income segregation estimates are also described in that online appendix.

(A) Income Segregation




(B) Racial Segregation

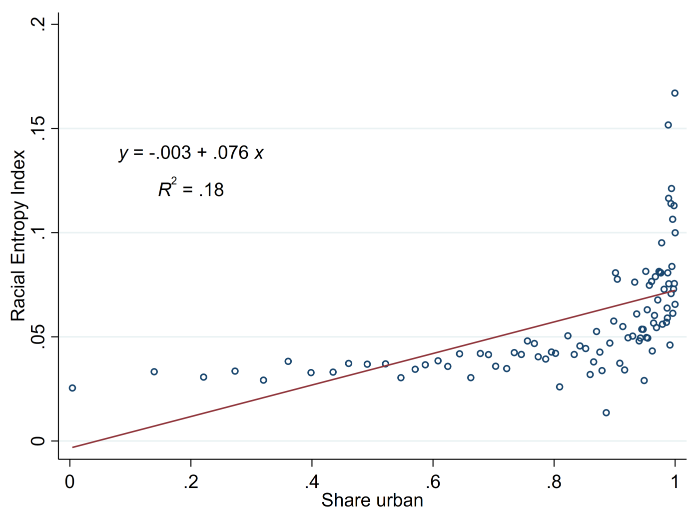



**FIGURE A.5**

BANK BRANCH SEGREGATION BY COUNTY'S URBAN SHARE

*Notes.* 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 2019 vestige of the FDIC's Summary of Deposits. The income entropy index values are estimates of Eq. (7). The racial entropy index values are estimates of Eq. (5). 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.

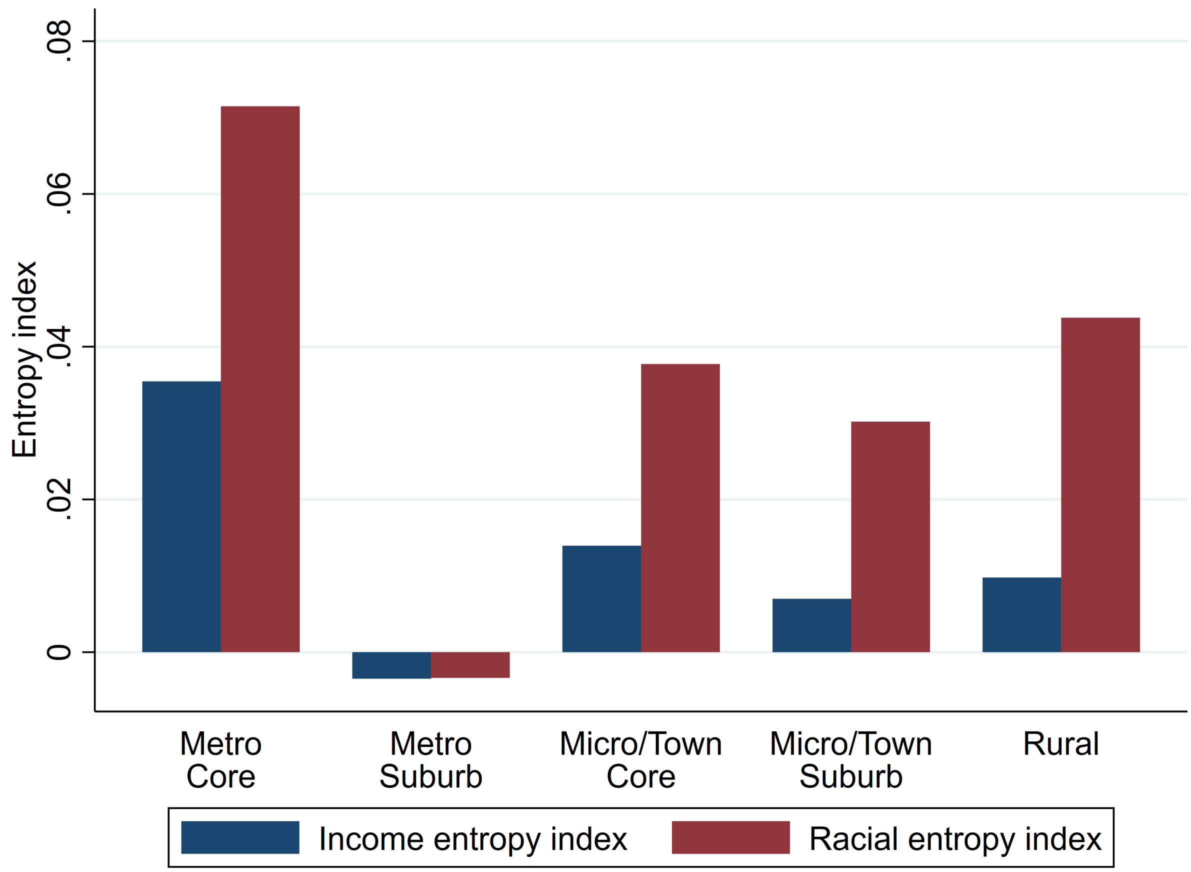



**FIGURE A.6**

BANK BRANCH SEGREGATION BY RUCA CLASSIFICATION

*Notes.* 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 code. *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. (7). The racial entropy index values are estimates of Eq. (5). 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 A.1**
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 |

**TABLE A.1** (CONTINUED)

| | Income Segregation | | | | | Racial Segregation | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **County** | **State** | **# Visitors** | **Value** | | **County** | **State** | **# Visitors** | **Value** |
| 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 |
| 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 |

*Notes.* The table reports the top-50 US 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 2019 vestige of the FDIC's Summary of Deposits). Branches are assigned to counties based on their locations in SafeGraph. Segregation estimates are calculated according to the methods described in Section 3.2. 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 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 A.2**

RACIAL SHARES OF THE PRIMARY RURAL-URBAN COMMUTING AREAS

| RUCA Code | Area type | N households | Black Share | Hispanic Share |
|---|---|---|---|---|
| 1 | Metropolitan area core | 99,473,952 | 0.15 | 0.17 |
| 2 | Metropolitan area high commuting | 13,270,243 | 0.06 | 0.06 |
| 3 | Metropolitan area low commuting | 1,262,793 | 0.06 | 0.06 |
| 4 | Micropolitan area core | 8,504,001 | 0.09 | 0.11 |
| 5 | Micropolitan high commuting | 3,682,427 | 0.06 | 0.03 |
| 6 | Micropolitan low commuting | 774,586 | 0.07 | 0.03 |
| 7 | Small town core | 4,356,721 | 0.09 | 0.08 |
| 8 | Small town high commuting | 1,439,308 | 0.06 | 0.03 |
| 9 | Small town low commuting | 625,530 | 0.07 | 0.03 |
| 10 | Rural areas | 5,549,527 | 0.03 | 0.04 |
| 99 | Not coded | 977 | 0.72 | 0.10 |
| | Total | 138,940,064 | 0.12 | 0.14 |

*Notes.* The table reports the number of households and shares of black and Hispanic households for the various Rural-Urban Commuting Areas (RUCA) in the US. RUCAs classify areas by their urban/rural status and their commuting relationships with other areas using Census measures of population density, levels of urbanization, and daily home-to-work commuting. Codes are provided for each Census tract and ZIP code by the US Department of Agriculture Economic Research Service, and the data are available here: RUCA classification. The values in the table reflect the area classifications from the 2019 update to the RUCA codes that are themselves based on the 2010 decennial US Census. Household counts and racial/ethnic shares come from the 2019 5-year American Community Survey.

**TABLE A.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 |

*Notes.* 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 2019 vestige of the FDIC's Summary of Deposits. The dissimilarity index in this paper is an estimate of Eq. 1, as described in Section 3.2.1. The two racial groups in the dissimilarity index computation are Black and non-Black. The racial entropy index is an estimate of Eq. (5), as described in Section 3.2.2. 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. (7), as described in Section 3.2.3. 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.