Bank Access Across America*

Jung Sakong† Alexander K. Zentefis‡

February 15, 2022

Abstract

We use location data from millions of mobile devices to construct a granular measure of bank access throughout the United States. The measure originates from a spatial gravity model and is a function of a local area's distance from available bank branches and branch characteristics. To overcome methods that protect user privacy in the mobile device data, we estimate the access measure using the Method of Simulated Moments. The estimated gravity coefficient used in the access measure is -0.8, which implies that the number of residents visiting a bank branch drops by 80% for every doubling in the branch's distance away from home. We document substantial variation in bank access nationwide. Rural areas experience considerably weaker access than big cities. We use the access measure to evaluate a policy of postal banking. We estimate that the policy would improve bank access the most in low income areas and areas with higher Black and Hispanic population shares.

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

*We thank Jane Dokko, Gary Gorton, Bo Honoré, Luojia Hu, Xuan Hung Do (discussant), Roberto Robatto, conference participants at the 2021 New Zealand Finance Meeting, and seminar participants at the Federal Reserve Bank of Chicago for their very helpful 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

Poor 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 researchers to rely on coarse measures, like the number of loans per capita or the 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 construct a micro-level measure of access to banking services across the United States, putting attention on bank branches. We document significant heterogeneity in bank access throughout the country and we use our measure to evaluate the extent to which a public policy of postal banking can improve access.

Our measure of bank access is derived from a spatial model of consumer travel patterns between home Census block groups and bank branches. We estimate the measure using anonymous location data from millions of mobile devices throughout the US per month from 2018-2019. We organize our analysis in three parts.

In the first part, we construct and estimate our local bank access measure. It is valuable to briefly describe the measure's derivation. The conceptual framework in Section 2 that microfounds the measure yields a gravity equation that characterizes visitor flows between residents' home Census block groups and bank branches over time:

$$\log\left(\text{No. of visitors}_{ijt}\right) = \gamma_{it} + \lambda_{jt} - \beta \log\left(\text{Distance}_{ij}\right) + \varepsilon_{ijt}. \tag{1}$$

The log number of visitors from block group i to branch j at time t is a linear function of four factors: (i) a block-group-by-time fixed effect γ_{it}, capturing all characteristics of block group i that contribute to residents visiting any bank branch at time t; (ii) a branch-by-time fixed effect λ_{jt}, capturing all characteristics of branch j that make it a destination for any block group at time t; (iii) the log distance between the block group and branch; and (iv) a mean-zero disturbance ε_{ijt}. The parameter β is the elasticity of visitor flows with respect to distance.

Taking predictions of Eq. (1), exponentiating, and summing across branches j produces the expected number of branch goers from block group i at time t, which we denote as $\hat{V}_{it}$:

$$\hat{V}_{it} = \exp\left(\hat{\gamma}_{it}\right)\hat{\Phi}_{it}. \tag{2}$$

Importantly, the term $\hat{\Phi}_{it}$ combines information across branches and is defined as

$$\hat{\Phi}_{it} \equiv \sum_{j \in B_{it}} \exp\left(\hat{\lambda}_{jt}\right) d_{ij}^{-\hat{\beta}}, \tag{3}$$

where B_{it} is the set of branches available to residents of block group i at time t, and d_{ij} is the distance between branch and block group.

Our micro-level measure of bank access is $\hat{\Phi}_{it}$. The measure is an attribute-adjusted branch index that is unique to each Census block group. Each branch in the index is represented by its characteristics, as captured by the fixed effect λ_{jt}, and its distance away from the local area. Distances are scaled by the parameter β, which, as we discuss in Section 2, can be interpreted as the product of consumers' elasticity of substitution between branches and their costs of travel. The measure is conceptually related to indices in the economic geography and trade literature that describe an exporting country's access to the importing markets of other countries (e.g., Harris, 1954, Head and Mayer, 2004,Redding and Venables, 2004, Hanson, 2005, and De Sousa, Mayer and Zignago, 2012).

By our measure, local areas have better bank access if more branches are available (higher B_{it}), the branches are closer (lower d_{ij}), or the branches feature superior attributes (higher λ_{jt}). Access is also higher if traveling costs are lower or if branches are less substitutable (lower β). Notice that a lower elasticity of substitution implies that residents find it worthwhile to travel to faraway branches despite the resistance imposed by their distance.

We compute the access measure $\hat{\Phi}_{it}$ per block group by estimating the gravity equation in Eq. (1) using cross-country, monthly foot traffic data from SafeGraph. The data provide detailed travel patterns of consumers based on the whereabouts of their mobile devices. Most importantly for us, the data include the locations of bank branches, the home Census block groups of branch visitors, and the associated number of branch visitors from each block group, which we apply directly to the gravity equation.

While this detailed micro-data on consumer trips is central to estimating our local measure of bank access, the data is subject to differential privacy methods, which try to shield the movements of individuals from becoming public. Noise is added to branch visitor counts, and the number of visitors from a block group to a branch are either omitted or censored if the count is too low. For this reason, an OLS regression of Eq. (1) would render biased estimates and distort our measure of access. To account for the differential privacy, we instead use the method of simulated moments (MSM) in the estimation. Standard MSM is straightforward to implement in models with no or few fixed effects

(McFadden, 1989). However, Eq. (1) has hundreds of thousands of fixed effects across two dimensions, which complicates the procedure. To handle such a large volume of fixed effects, we introduce a new computational routine, which we describe in Section 4. When the number of fixed effects in each dimension is large, as in our setting, the routine produces consistent estimates. The routine can easily be implemented by other researchers applying MSM to a model with voluminous fixed effects.

From the MSM procedure over the full panel of data, the MSM estimate $-\hat{\beta}_{\text{MSM}} = -0.8$ with standard error 0.056. Thus, across the country, 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 around 0.8% per month. We also run the MSM procedure month-by-month to evaluate the stability of the gravity coefficient. The monthly point estimates range from -0.525 to -0.973.

Our access measure is based on bank branches, but has the growth of online and mobile banking spoiled branches as important indicators of financial access? In Online Appendix C.1, we present national survey evidence from the "2019 FDIC Survey of Household Use of Banking and Financial Services" that suggests not. Roughly 81% of respondents answered having visited a branch in the past 12 months and about a quarter indicate that a branch is their primary means to access their accounts. Overall, bank branches remain integral to the financial lives of most US households, which means that a measure of access based on bank branches is informative.

In the second part of the article, we characterize variation in our bank access measure across the country. Nationwide, the most pronounced differences in access are between urban and rural areas. Big cities like Boston, Richmond, Miami, and Minneapolis observe substantially higher bank access than even nearby rural areas. Suburban areas that outlie large metropolitan centers also observe greater access than rural parts of the country.

There is substantial within-city variation in bank access as well. For example, in Greater Los Angeles, the neighborhoods of Beverly Hills and the Hollywood area observe substantially better access than parts south of the city, such as Compton and Long Beach. Neighborhoods in the Palos Verdes Peninsula, San Fernando, and near the Anaheim Hills observe the lowest access. Around Chicago, the North side experiences better access than the South side, and the Northwest suburbs near Lake Forest observe the highest access. Around Washington, access is highest in the District and lowest in Anacostia. Finally, around New York City, Midtown Manhattan observes the highest access—more than the Upper East or West Sides—Brooklyn and Queens experience lower access than Manhattan overall, and Staten Island observes the lowest access among the five Boroughs.

In the final part of the article, we use the access measure to evaluate a policy proposal that might improve bank access: postal banking. Such a policy would extend checking, savings, and possibly credit services to some or all Post Office branches. 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). Relaunching the Postal Savings System as a way to improve access has been proposed by members of Congress (Warren, 2014; Gillibrand, 2021; Sanders, 2021) and parts of academia (Baradaran, 2013; Johnson, 2017).

With our data, we can assess how postal banking would affect access of consumers to their (private and public) bank branches. We use our gravity estimate of β from the MSM to measure a partial policy impact of postal banking. For simplicity, we also presume that the block group fixed effects γ_{it} would remain the same, which understates the impact of postal banking because it ignores any possible introduction of new branch goers among the unbanked as a result of the policy. This kind of analysis, which keeps the gravity estimate and block group fixed effects the same 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).

Nevertheless, we exert discretion over the establishment fixed effects λ_{jt} that postal bank branches might have. A branch's fixed effect stands in for the "quality" of the location, as it captures all features that draw in visitors from any block group. In our analysis, we apply uniform fixed effects to all identified Post Office branches registered in SafeGraph, implicitly presuming that each would become a postal bank. We therefore provide closer to an upper bound on the potential postal branch choice set, as not all the postal locations we include might 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 shrinking the branch choice set.

To evaluate the policy, we consider three scenarios of Postal branch fixed effects, with each one based on a percentile of the distribution of private bank fixed effects per year-month of the sample. In the first scenario, Postal branches are "low quality," in that they share the estimated fixed effect of the 10th percentile of private banks per year-month. In the second scenario, Postal branches are "medium quality," in that they share the estimated fixed effect of the 50th percentile of private banks per year-month. Finally, in the third scenario, Postal branches are "high quality," in that they share the estimated fixed effect of the 90th percentile of private banks per year-month. Postal banks being perceived as high quality is acquainted with the notion that a government-sponsored banking institution is considered more trustworthy than private banks (Office of the USPS Inspector General, 2014; Baradaran, 2015).

With this in mind, we evaluate how bank use and access would change with postal banking under the three scenarios. Regarding use, without postal banking, the median number of residents from a typical block group nationwide who visit a private bank branch is 7.16 per month. In addition, block groups with higher median house hold income also observe more branch visitors. Under a "low quality" postal banking system, median branch use would rise to 7.67 per month (7% increase). It would rise to 9.80 per month (37% increase) under a "medium quality" postal banking system and 18.74 per month (162% increase) under a "high quality" postal banking system.

Breaking down this increase by demographic variables, we show that the income gradient on branch use would flatten by roughly 3 percentage points nationally under a "medium quality" postal banking system. In other words, postal banking would be associated with higher visitation rates nationwide, and poorer block groups would observe a 3 percentage point greater increase in bank visitation per month than richer block groups. We further estimate that postal banking would increase branch use the most among residents of block groups with higher Black or Hispanic population shares. Under a "medium quality" postal banking system, branch use would increase about 0.013 percentage points per month for every 1% increase in a block group's Black or Hispanic population share.

The increase in branch use arises from the increased access that postal banking would afford. Under a median quality postal banking system, the bank access of a typical block group nationwide would increase by roughly 2.7 percentage points for every doubling in median household income. Under a low quality postal banking system, the increase is about 0.70 percentage points, and under a high quality postal banking system, bank access increases about 6.8 percentage points. Block groups with higher Black population shares would experience a roughly 1.3 percentage point increase in access under a medium quality system, a 0.2 percentage point increase under a low quality system, and a 3.4 percentage point increase under a high quality system. Similarly, the increase in access of block groups with higher Hispanic population shares would range from 0.9-6.2 percentage points nationwide under the three scenarios.

We estimate that the positive impact on access would be even higher among block groups in big cities with higher Black or Hispanic population shares. Compared to a hypothetical block group with a 100% White population share, a hypothetical block group with a 100% Black population share would experience a rise in access between 0.7-7.7 percentage points under a postal banking system. The analogous increase in bank access for a hypothetical block group with a 100% Hispanic share would be between 1.1 and 9.1 percentage points. This last finding implies that postal banking would have the biggest impact on bank access and use in poorer block groups and block groups with higher Black or Hispanic population shares, particularly in big cities.

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, Ferguson, Gentzkow and Schmidt (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 access and use across the US.

Second, this article contributes to the vast array of work that investigates financial access, financial use, and their joint relation to 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. See also Washington (2006), Blank (2008), Ho and Ishii (2011), Goodstein and Rhine (2017), and Célerier and Matray (2019) that investigate other measures of access in the US. An advantage of our access measure is that it embodies the equilibrium choices of individual consumers, rather than just reflecting their survey responses or being constructed from pure supply-side factors, such an area having a branch nearby, the local branch density, or the availability of low cost accounts.

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 estimate spatial patterns in banking activity.

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 access and use.

Outline. The article proceeds as follows. Section 2 presents a spatial model that serves as a conceptual framework for our measure of bank access. Section 3 describes the mobile device data we use to estimate the access measure. Section 4 details the approach to estimating the measure. Section 5 analyzes variation in bank access by geography and other characteristics of local areas. Section 6 analyzes racial and income differences in branch use. Section 7 evaluates a policy of postal banking using the access measure. Section 8 concludes.

2 A Conceptual Framework for Bank Access

In this section, we provide a conceptual framework that informs our local measure of bank access. The framework describes the spatial movements of residents between their home Census block groups and bank branches. Each resident r among a continuum inhabits one block group $i \in G$. Each bank branch across the country is indexed by $j \in B_t$, and the set of branches can vary over time from store openings or closings.

In every time period, each resident chooses which single bank branch to visit so as to maximize utility. Residents may also choose not to visit a branch, either remaining home or visiting another point-of-interest. We index this outside option choice by $j = 0$. The indirect utility of resident r living in home block group i visiting branch j at time t is

$$U_{rjt} = \frac{z_{rjt}\Lambda_{jt}}{\delta_{ij}}. \tag{4}$$

The term Λ_{jt} is an index of all attributes of branch j that make it a destination for residents of any block group at time t (e.g., the branch having better deposit or loan rates, higher staff quality, or an efficient drive-through ATM). The term z_{rjt} is an idiosyncratic, unobserved error that captures individual differences in residents' personal preferences for banking at branch j (i.e., favoring Chase over Wells Fargo, relishing the branch's proximity to the children's daycare, or appreciating the building's historic architecture). Finally, the term δ_{ij} is an iceberg traveling cost that is defined as

$$\delta_{ij} = d_{ij}^{\kappa}, \tag{5}$$

where d_{ij} is the distance between home block group i and branch j, and $\kappa > 1$ controls the scale of the traveling costs.

To derive mathematically convenient functional forms for the branch choice behavior of the population, we follow McFadden (1974), Eaton and Kortum (2002), and Ahlfeldt et al. (2015) by assuming that the idiosyncratic component of utility z_{rjt} is drawn from an independent Fréchet distribution:

$$F\left(z_{rjt}\right) = e^{-H_{jt}z_{rjt}^{-\varepsilon}}, \quad V_{jt} > 0,\ \varepsilon > 1. \tag{6}$$

Here, the branch-specific parameter $H_{jt} > 0$ influences the mean of the distribution. A larger H_{jt} implies that a high utility draw for branch j is more likely among residents of any block group. The term $\varepsilon > 1$ governs the heterogeneity of idiosyncratic utility. A smaller $\varepsilon > 1$ implies that residents are more heterogeneous in their preferences for branches.[1]

Substituting the expression for U_{rjt} into the distribution of idiosyncratic tastes in Eq. (6) implies that residents of block group i at time t are presented with a distribution of utility across branches $G_{ijt}(u) = \Pr\left[U_{rjt} \leq u\right] = F\left(u\delta_{ij}/\Lambda_{jt}\right)$, or

$$G_{ijt}(u) = e^{-\left[H_{jt}\left(\frac{\Lambda_{jt}}{\delta_{ij}}\right)^{\varepsilon}\right]u^{-\varepsilon}} \tag{7}$$

We normalize the value from the outside point-of-interest $H_{0t}\Lambda_{0t}^{\varepsilon}\delta_{i0}^{-\varepsilon} = 1$. Each resident chooses a location to visit that yields the maximum utility. Hence, the distribution of utility across all possible

[1]The parameter ε plays a role like the elasticity of substitution between bank branches in a model where residents have CES preferences over bank services from all branches. A smaller ε is akin to branches being less substitutable, which implies that residents find it worthwhile to travel to a branch despite resistance imposed by the geographic barrier δ_{ij}.

locations that a resident would actually visit is

$$G_{it}(u) = \Pi_{j=0}^{J_t} G_{ijt}(u).$$

Inserting Eq. (7), one obtains the utility distribution:

$$G_{it}(u) = e^{-(1+\Phi_{it})u^{-\varepsilon}}, \tag{8}$$

where the parameter Φ_{it} of block group i's utility distribution is

$$\Phi_{it} = \sum_{j \in B_t} H_{jt} \Lambda_{jt}^{\varepsilon} d_{ij}^{-\kappa\varepsilon}. \tag{9}$$

The object Φ_{it} is the theoretical foundation of our measure of bank access. It summarizes information about the set of branches available to residents of each block group per time period. Given its form, Φ_{it} can be interpreted as an attribute-adjusted branch index that is unique to each block group i. Each branch in the index is represented by the average idiosyncratic utility $\left(H_{jt}\right)$ that residents assign to the branch's attributes $\left(\Lambda_{jt}\right)$ and the branch's distance $\left(d_{ij}\right)$ away. The impact of distance is influenced by the scale of traveling costs (κ) and the substitutability between branches (ε). The object Φ_{it} is conceptually related to what some in the economic geography and trade literature have described as an exporting country's "access" to the importing markets of other countries (e.g., Harris, 1954, Head and Mayer, 2004, Redding and Venables, 2004, Hanson, 2005, and De Sousa et al., 2012). In our environment, we treat Φ_{it} as a local measure of residents' access to banks.

A convenient property of the utility distribution is that it generates a gravity equation in visits between home block groups and bank branches. The share π_{ijt} of residents living in block group i who visit branch j at time t is

$$\pi_{ijt} = \frac{H_{jt} \Lambda_{jt}^{\varepsilon} d_{ij}^{-\kappa\varepsilon}}{1+\Phi_{it}}. \tag{10}$$

The visitor share depends on the characteristics of the branch $\left(\Lambda_{jt}\right)$, the average utility draw of the branch $\left(H_{jt}\right)$, and the "bilateral resistance" derived from the intervening distance $\left(d_{ij}\right)$. Other things equal, a resident is more likely to visit a branch if it has superior characteristics, receives a higher average idiosyncratic utility, or is closer. In the denominator, Φ_{it} plays the role of "multilateral resistance," which is a term from the trade literature (Head and Mayer, 2014). It affects residents' visitation to *all possible* branches. The probability that residents of a block group in, say, Palo Alto, visit a nearby Chase branch depends not only on the benefits of the branch and the costs of getting there, but also on the benefits and costs of visiting all other available branches.[2]

Eq. (10) offers a conceptual interpretation of the gravity equation we employ to compute our measure of bank access. We estimate the gravity equation using information on residents' visits from their home block groups to bank branches. Estimates from the gravity equation are then aggregated to the block-group level to obtain local measures of bank access. We turn next to describing the micro-level data we use in the estimation.

3 Data on Branch Visits

Estimating the gravity equation requires data on the travel patterns of residents from their homes to bank branches. We measure these travel patterns using mobile device data from SafeGraph between January 2018 and December 2019. The data are monthly and include both branch locations and information about branch visitors. The SafeGraph data is benefited by elaborate algorithms the company has developed to determine whether a mobile device visits a particular destination and to pinpoint a mobile device's home origin. A visitor is identified by a mobile device, one device is treated as one visitor, and a device must spend at least 4 minutes at an establishment to qualify as a visitor. Importantly for us, SafeGraph provides the home Census block groups of bank branch visitors and the associated number of branch visitors from each block group. The home block group of an individual visitor is not given. 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 summary.[3]

[2] We obtain the gravity relation in Eq. (10) by evaluating:

$$\pi_{ijt} = \Pr\left[u_{ijt} \geq \max\left\{u_{ijt} \;;\forall j\right\}\right] = \int_0^\infty \Pi_s \left[G_{is}(u)\right] dG_{ijt}(u)\, du.$$

[3] 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.

3.1 Primary Sample

Our primary (core) data set includes bank branches in all 50 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 2019 Summary of Deposits (SOD).

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 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.[4]

Our core sample is confined to bank branches for which SafeGraph has visitor data. Many 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 visitors to a unique place that is part of a shared space awfully difficult. To reduce false attributions, SafeGraph instead allocates visitors to the larger "parent" space, such as the encompassing mall. 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 visitors to an enclosed branch. Instead, those visitors 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.) Visitors to these enclosed branches cannot be separated from visitors to Walmart, and so, these branches are deprived of visitor data.[5]

[4]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.

[5]Regarding branch openings and closings, if a bank branch closed and SafeGraph were aware of its closure, any visitors to the building (say, if a new business opened there) would no longer be attributed to the branch. Likewise, if a branch

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 50,999 of these places have visitor data and constitute our core sample. Our core sample thus covers around 60% of bank branches in the United States.[6]

3.2 Sampling Bias

Our core sample experiences three types of sampling bias: (i) differential privacy, (ii) sample selection, and (iii) bank employees. We discuss each bias below and describe how we address it.

Differential Privacy. The first bias emerges from SafeGraph's efforts to preserve user privacy. The company applies differential privacy methods to avoid identifying people by their home locations. First, Safegraph adds Laplace noise to all positive counts of visitors to a branch from each home Census block group of the branch's visitors. Second, they round each of these block group-branch visitor counts down to the nearest integer. Third, they drop from the data all rounded visitor counts less than 2. Fourth, if a rounded visitor count equals 2 or 3, they set it to 4. These last two data adjustments render our sample subject to both truncation from below and censoring from below. Figure I presents the distribution of the observed (raw) visitor counts, which reveals both the truncation and censoring. Roughly three-quarters of the observed visitor counts equal 4. We account for SafeGraph's differential privacy methods by estimating the gravity equation using the method of simulated moments. The full procedure is detailed in Online Appendix B, and a summary of the procedure is provided in Section 4.

Sample Selection. The second bias relates to sample selection, as our data on branch visitation patterns might not be representative of the true population behavior in the US. Potential sampling bias arises from two sources: our set of branches and our set of visitors. To address potential sampling bias from missing around 40% of US branches, in Section 3.3 we compare the representation of different demographic groups in the areas covered by our core sample of branches to the areas covered by all branches in the SOD. Overall, differences in demographic characteristics between the two sets of areas are precisely estimated, but small. Regarding our sample of visitors, SafeGraph aggregates data from

opened and SafeGraph were aware of it, visitors would start being attributed to the branch. Nevertheless, if SafeGraph is unaware of a branch's opening or closing, visitors would be incorrectly attributed and count toward measurement error.

[6]Online Figure A.1 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.

around 10% of all mobile devices in the country. We calculate about 30 million unique mobile devices per month on average visiting all businesses recorded in SafeGraph, and our core sample reports 1.6 million visitors to bank branches per month on average.[7] The 2010 US Census records 217,740 Census block groups, and our core sample includes 215,686 unique visitor home block groups, implying close to complete coverage of US local home areas. In Online Appendix C.3, we compare by household income the share of households in the 2019 FDIC survey who report having visited a bank branch in the previous 12 months to the share of mobile devices in SafeGraph that visit branches in the same period. There is a strong resemblance between the two sources, as both reported branch visitor shares from the FDIC survey and observed branch visitor shares from the mobility data are increasing and concave in household income.

Nevertheless, we cannot rule out non-random sampling of mobile devices based on unobserved characteristics of visitors. As we discuss in detail in Section 5.2, we do not know the precise demographic attributes of an individual bank branch visitor, and instead, we assign attributes to visitors according to the demographic characteristics of their home Census block groups. Applying this same approach, Squire (2019) quantifies the sampling bias across the entire SafeGraph dataset. He documents that the number of devices from SafeGraph's identified home locations correlates highly at the county level with 2010 US Census numbers in terms of population counts (97%), educational attainment (99%), and household income (99%).[8]

Despite this strong alignment between the Census and SafeGaph at the county level, Thaenraj (2021) identifies around 1,000 Census block groups in the SafeGraph data that register more devices residing than there are people according to the 2019 ACS. Squire (2019) also discusses this feature of the SafeGraph panel, and he interprets these outlier Census block groups as most likely representing errors or technical limits in SafeGraph's attribution of devices to home block groups. Less extreme misattributions are also possible. But any misattribution is likely between neighboring block groups with similar characteristics because the SafeGraph representation lines up well at the county level. In Online Appendix D.2, we weight the branch visitor data by Census population counts. Doing so down-weights visitor counts from block groups with disproportionately high visitors relative to their

[7] Online Figure A.1 presents a time-series of the number of branch visitors each month over the sample period. The number of branch visitors per month rises over the sample period, starting from around 900 thousand in January 2018 and ending with 1.85 million in December 2019. The change could reflect a combination of increasing bank visitation and improving visitor coverage over time.

[8] Couture et al. (2021) analyze mobile device data from the provider PlaceIQ, and the authors find that it too is broadly representative of the general population based on assigned household attributes and movement patterns.

populations, and up-weights counts from areas with disproportionately low visitors relative to their populations, so as to make the visitor sample more representative.

Bank Employees. The third bias relates to bank employees versus bank customers. SafeGraph does not distinguish branch employees from branch patrons in the visitor data. All mobile devices that travel to a branch are considered "visitors." One way to approximate the number of employees is by the number of mobile device visits that dwell at the location for a long time (e.g., greater than 4 hours). The ratio of a branch's number of "long dwells" to its total visit count indicates the extent to which employees' presence skews our gravity estimates. In Online Appendix D.3, we present several ways we account for bank employees in checks for robustness.

3.3 Descriptive Statistics

Table I reports descriptive statistics from our core sample. The typical branch has 40 visitors per month on average and an interquartile range of 5 to 48 visitors. For each branch, SafeGraph provides both the median distance visitors travel to get there and the median time they spend there. On average, the median distance traveled is 5 miles, and the 90th percentile is 9 miles. The median dwell time 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 by our core sample of branches with those in the areas represented by all branches in the SOD. Demographic attributes in the table are taken from the 2019 5-year 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 in our core sample are over-represented by about 3% compared to the SOD, which coincides with the greater mobile phone 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.

4 Bank Access Estimation

The conceptual framework of Section 2 produces a gravity equation in Eq. (10) that characterizes the share π_{ijt} of consumer visitors between home block groups and bank branches. Our measure of bank access is computed by estimating a gravity equation that can be interpreted through the lens of that framework. SafeGraph's differential privacy methods bias any OLS estimation of the gravity equation. In this section, we describe our alternative estimation approach and discuss the gravity estimates that enter the measure of access.[9]

4.1 Estimation Approach

In the estimation, we express the gravity equation in levels rather than shares to account for the differential privacy methods that SafeGraph applies to the visitor data. Let V^*_{ijt} be the true number of visitors from block group i to branch j at time t that SafeGraph observes, and let R_{it} denote the number of residents of home block group i at time t. The true number of visitors can be expressed as:

$$V^*_{ijt} = \pi_{ijt} R_{it}, \tag{11}$$

where π_{ijt} is the share of block group i's residents visiting branch j at time t. Let L_{ijt} denote the Laplace noise that SafeGraph adds to V^*_{ijt} to protect user privacy. Noise is added only if SafeGraph observes a visitor (i.e., $V^*_{ijt} > 0$). The noise $L_{ijt} \sim \text{Laplace}(0, b)$, where b is the scale of the distribution, and SafeGraph informed us that $b = \frac{10}{9}$. Let V^+_{ijt} denote the number of visitors after the noise is added, giving:

$$V^+_{ijt} = V^*_{ijt} + L_{ijt}. \tag{12}$$

Let $\lfloor V^+_{ijt} \rfloor$ denote the integer floor to which SafeGraph rounds the noisy visitor count. To accommodate SafeGraph's truncation and censoring, we denote z_{ijt} as an indicator for whether a block group-branch

[9] The differential privacy distortions as they relate to the gravity equation are clearly observable in Online Figure A.5 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 SafeGraph's censored value of 4 visitors.

visitor count is present in the sample. The selection equation is

$$z_{ijt} = \begin{cases} 1 & \text{if } \lfloor V_{ijt}^{+} \rfloor \geq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Let V_{ijt} denote the number of visitors observed in the sample, subject to SafeGraph's censoring. The observation equation is

$$V_{ijt} = \max\left\{4, \lfloor V_{ijt}^{+} \rfloor\right\}. \tag{14}$$

To estimate the gravity equation, we use the method of simulated moments (MSM), relying on Eqs. (11)–(14). The MSM chooses model parameters to make simulated model moments match the data moments. We run the estimation across the full panel and separately per year-month of the sample to evaluate the stability of the estimates over time. A full description of our procedure is provided in Online Appendix B, but we provide a brief summary here.[10]

In the simulations, we specify that V_{ijt}^{*} is Poisson distributed with mean $\pi_{ijt}R_{it}$. Using the gravity equation from the conceptual framework (Eq. 10) to inform π_{ijt}, we express the true visitor count as obeying

$$V_{ijt}^{*} \sim \text{Pois}\left(\exp\left(\gamma_{it} + \lambda_{jt} - \beta \log \text{Distance}_{ij}\right)\right), \tag{15}$$

where γ_{it} is a block-group-by-year-month fixed effect that captures all characteristics of block group i that contribute to residents visiting any branch in year-month t, and λ_{jt} is a branch-by-year-month fixed effect that captures all characteristics of branch j that make it a destination for any block group in year-month t. A block group's fixed effect can be thought of as reflecting its residents' "demand" for bank branch services, and a branch's fixed effect can be thought of as reflecting the "quality" of the establishment.[11] The parameter β is the elasticity of visitor flows with respect to distance. From the perspective of the conceptual framework, β can be interpreted as the product of residents' traveling cost scale (κ) and their elasticity of substitution between branches (ε).

Technically speaking, every possible block group i and branch j pair should enter Eq. (15). But our sample of over fifty-thousand branches, over two-hundred-thousand block groups, altogether spanning twenty-four months, makes it computationally infeasible to calculate an access measure that accounts for all possible block group-branch pairs. Instead, we reduce the choice set per block group to include

[10] For textbook treatments of MSM, see Adda and Cooper (2003), Davidson and MacKinnon (2004), and Evans (2018).

[11] Using origin and destination fixed effects to estimate gravity equations has become standard practice in the trade literature since Harrigan (1996). Because we have a panel, the cross-sectional fixed effects are time-varying.

only those branches located within a 10-mile radius around the block group's center of population. A distance-based cutoff of 10 miles conforms with the definition of "bank deserts" used in the literature (Morgan, Pinkovskiy, Yang et al., 2016; Dahl and Franke, 2017), which implicitly regards branches located outside a 10-mile radius as unreasonable destinations. In Online Appendix D.1, we present robustness checks by expanding the radius to 25 miles and 60 miles.[12]

In the estimation, we simulate visitor counts according to Eq. (15) and apply the differential privacy methods represented in Eqs. (12)–(14). The MSM then uses the visitor data and the model parameters to minimize the distance between simulated model moments and data moments. With the very large number of block groups and branches in our sample, the model of visitor counts in Eq. (15) requires thousands of fixed effects be estimated. There are simply too many parameters to identify from the MSM minimization problem alone. Instead, we adopt an iterative routine to identify the fixed effects $\{\gamma_{it}, \lambda_{jt}$ and let the minimization problem identify β. Given an estimate of β, the estimates of the fixed effects repeatedly update until the differences in the average simulated visitor counts and average observed visitor counts of each block group i and each branch j per year-month t become sufficiently small. After the fixed effects converge per estimate of β, the MSM minimization problem then selects the optimal β estimate that satisfies the moment conditions. When the number of fixed effects in each dimension is large, as in our setting, the routine produces consistent estimates.

4.2 Gravity Estimates

From the MSM procedure over the full panel of data, the MSM estimate $-\hat{\beta}_{\text{MSM}} = -0.8$ with standard error 0.056. Thus, across the country, 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 around 0.8% per month. Figure II presents the MSM estimates from the month-by-month estimation, along with 95% confidence intervals. The monthly point estimates range from -0.525 to -0.973.

Figure III presents histograms of the estimated Census block group and and bank branch fixed effects across all months of the sample period. A block group's fixed effect can be interpreted as

[12] Online Figure A.3 illustrates the CDF of visitor counts by the log miles between the block group and visited branch. Seventy-five percent of visitors travel to a branch within 10 miles, and ninety-five percent travel within 60 miles. Survey evidence also suggests that banking is quite local. Studying the Survey of Consumer Finances, Amel, Kennickell and Moore (2008) find that, in both 1992 and 2004, 75% of U.S. households had a checking account at a branch located within just five miles of their place of residence or work.

the average log number of residents from that block group who visit any branch in the year-month, controlling for branch fixed effects and distance. The bulk of the distribution of block group fixed effects range from exponentiated values around 0.1 to 50. Similarly, a branch's fixed effect can be interpreted as the branch's average log number of visitors in the year-month, controlling for visitors' block group fixed effects and distance. Most of the mass is within a range of exponentated values between 0.01 and 21.

In Figure I, we compare the distributions of observed "raw" visitor counts to simulated "true" visitor counts according to the month-by-month MSM estimation of the Poisson model in Eq. (15). The comparison reveals how well the MSM uncovers the visitor data before it is affected by SafeGraph's differential privacy methods. The simulated counts distribution includes all positive draws from all simulations across every year-month in the sample period. Among observed visitor counts, 74.5% equal 4 due to SafeGraph's truncation and censoring, whereas among simulated "true" visitor counts, 72.5% equal 4 or less. The MSM does a reasonable job spreading out the mass of visitors into the left tail of the distribution that is lost in the observed data. The two distributions also line up fairly well at values exceeding 4. For instance, around 4.3% of observed visitor counts equal 15 or more, whereas 5.1% of simulated visitor counts do.

Figure IV directly compares the observed number of visitors from each Census block group to their expected counterparts from the simulation. Specifically, it presents a scatter plot of the log observed number of branch goers from each block group (i.e., $\log V_{it} \equiv \log \sum_j V_{ijt}$, where V_{ijt} is given in Eq. 14) versus the log expected number of branch goers from each block group based on the month-by-month MSM estimates (i.e., $\log \hat{V}^*_{it} \equiv \log \sum_j \hat{V}^*_{ijt}$, where $\hat{V}^*_{ijt}$ is the predicted mean of V^*_{ijt} in Eq. 15 and β is time-varying). If SafeGraph applied no differential privacy methods to their data, all dots in the figure would line up neatly on the red 45° line. The single caveat is that the expected number of visitors might not be whole numbers, whereas the observed number of visitors must be. The censoring levels off the observed visitor counts at 1.4, which corresponds to 4 visitors. The truncation causes the observed visitor counts to enter below the expected visitor counts, and the gap between observed and expected counts is largest for block groups with few branch goers, which are areas where the truncation has the largest impact. The gap shrinks as the number of branch goers from a block group increases, and in block groups with many bank goers, the observed and expected number of visitors nearly match. This implies that the MSM generates estimates that fit the data well in regions least affected by the differential privacy distortions, which one would hope for.

SafeGraph's differential privacy methods biases any OLS estimation of the gravity equation and prompts an alternative estimation method like the MSM. Computing the OLS estimates is still useful, though, to informally assess the magnitude of the bias. To this end, Online Table A.6 presents estimates from OLS regressions of the gravity equation in Eq. (1) using the observed visitor counts. We include a specification in which racial shares are also interacted with distance as independent variables. We run two sets of regressions: those including all block group-branch pairs and those that limit pairs with more than 4 branch visitors (which avoid SafeGraph's censoring).

When all block group-branch pairs are included, the OLS estimate $-\hat{\beta}_{\text{OLS}} = -0.053$, which is an order of magnitude below the MSM estimate. When racial shares are interacted with distance, $-\hat{\beta}_{\text{OLS}}$ remains largely unchanged at -0.056. The coefficients on the racial interaction terms are small, which suggests that the relation between distance and visitor flows is universal across racial groups. But the coefficient on the interaction term with the Black share is precisely estimated and positive, which is consistent with residents from block groups with larger Black population shares having a lower elasticity of substitution across branches.

When the sample is limited to block group-branch pairs with greater than 4 branch visitors, the estimate $-\hat{\beta}_{\text{OLS}}$ nationwide increases to -0.283, which still falls short of the MSM estimate. Online Figure A.5, 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. When racial shares are added as interaction terms, the OLS estimate of $-\hat{\beta}_{\text{OLS}} = -0.258$. The sign on the interaction term with the Black population share is still positive, though no longer precisely estimated. Overall, Online Table A.6 reveals the downward bias that SafeGraph's differential privacy methods introduce to an OLS estimation of the gravity equation, and it stresses the need for the alternative MSM procedure.

5 Spatial Variation in Bank Access

We move now to examining variation in bank access throughout the United States. Substituting the gravity equation estimates into the theoretical measure of access in Eq. (9), our empirical measure of block group i's bank access in year-month t is

$$\hat{\Phi}_{it} \equiv \sum_{j \in B_{it}} \exp\left(\hat{\lambda}_{jt}\right) d_{ij}^{-\hat{\beta}}, \tag{16}$$

where B_{it} is the set of branches in the period within a 10-mile radius of the block group.

The relation between bank access and branch visitor counts from the Poisson model of Eq. (15) bestows economic meaning to the magnitude of $\hat{\Phi}_{it}$. From that equation, the expected total number of residents of block group i who visit any branch in year-month t is

$$\hat{V}^{*}_{it} = \exp\left(\hat{\gamma}_{it}\right)\hat{\Phi}_{it}. \tag{17}$$

In the MSM estimation, we normalize the block group fixed effects $\exp\left(\hat{\gamma}_{it}\right)$ to have mean one across block groups within the year-month. This normalization implies that the mean of $\hat{\Phi}_{it}$ is the expected number of branch goers per month from the average block group.

We characterize the spatial heterogeneity in bank access by evaluating it over different parts of the US and by measuring its association with demographic characteristics of block group residents.[13,14]

5.1 The Geography of Bank Access

Here we describe the variation in bank access nationwide and in different pockets of the country. Figure V illustrates a dot density map of bank access by Census block groups cross-country. Each dot is positioned at a block group's center of population. Access estimates are calculated month-by-month per block group, and the figure presents weighted monthly averages, where each month's weight is its share of the block group's total branch visitors over the sample period. The map is constructed by grouping block groups into deciles and shading the dots so that higher-ordered colors in the rainbow gradient (i.e., indigo and violet) imply higher bank access values and lower-ordered colors (i.e., red and orange) imply lower access values. Block groups where no resident was recorded in SafeGraph as having visited a branch in the sample period and block groups having no bank branch within a 10-mile radius are shaded white.

Across the country, the most pronounced differences in access are between urban and rural areas. Big cities like Boston, Richmond, Miami, and Minneapolis observe substantially higher bank access than even nearby rural areas. Suburban areas that outlie large metropolitan centers also observe greater access than rural parts of the country.

[13] In Online Appendix E, we study how the actual distances residents travel to their branches vary by demographic attributes. Residents of poorer block groups travel shorter distances on average, but residents of block groups with higher Black population shares travel farther.

[14] In Online Appendix F, we examine the extent to which different groups choose different *menus* of branches. We do so by estimating measures of segregation among bank branch visitors. Big cities observe the highest segregation of branch goers by race and income, and the urbanized Northeast is more segregated than the rural South.

There is substantial within-city variation in bank access as well. Figure VI zeroes in on major Metropolitan Statistical Areas of four big cities: Los Angeles, Chicago, Washington, DC, and New York City. In Greater Los Angeles, Beverly Hills and the Hollywood area observe substantially better access than parts south of the city, such as Compton and Long Beach. Neighborhoods in the Palos Verdes Peninsula, San Fernando, and near the Anaheim Hills observe the lowest access. Around Chicago, the North side experiences better access than the South side, and the Northwest suburbs near Lake Forest observe the highest access. Around Washington, access is highest in the District and lowest in Anacostia. Finally, around New York City, Midtown Manhattan observes the highest access—more than the Upper East or West Sides—Brooklyn and Queens experience lower access than Manhattan overall, and Staten Island observes the lowest access among the five Boroughs.

We turn next to discussing how bank access varies with the demographic attributes of local areas. Before we get there, we describe next how to interpret regressions of bank access (and later branch use) on demographic attributes when our data on visitors is anonymous.

5.2 Assigning Demographic Attributes to Individual Visitors

A goal of our study is not only to develop a micro-level measure of bank access, but also to apply that measure to explain differences in branch use by demographic characteristics, such as race and income. But by 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 the 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 number of residents who visit a bank branch increases in the median household income of their 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 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 when explaining differential patterns of branch access and use.

Even so, benefiting from block-group-level information does not mean that we escape from the ecological inference problem. Focusing on household income, Online Figure A.2, 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 our log access measure is regressed on block-group-level racial population shares (with the White population shares omitted) and a control for the log number of devices residing in the block group. 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 x% poorer access."

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.2, 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 $100x\%$ poorer access, 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-level 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 experiences $100x\%$ poorer access than a White resident."

5.3 Bank Access and Demographic Attributes

We next examine how bank access at Census block groups nationwide covaries with the demographic attributes of area residents. Because a central focus of our study is answering how bank access and use 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 between Black and White residents. 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.[15]

Online Table A.1 presents household counts and Black shares throughout the US and within each 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.[16] 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 $\log \hat{\Phi}_{it}$ on demographic attributes of block group

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

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

residents. We compute $\hat{\Phi}_{it}$ using estimates from the month-by-month MSM procedure such that $\hat{\beta}$ varies over time. Observations in the regressions 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. 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 all specifications, we control for the natural logarithm of the number of mobile devices residing in the block group, which is unaffected by SafeGraph's differential privacy methods. Because we normalize the block group fixed effects $\exp(\hat{\gamma}_{it})$ to have mean 1, the coefficients from the table can be interpreted as the percent change in the expected number of branch goers per month, holding constant the block group fixed effects.

Our access measure embodies information about the costs of travel through the elasticity parameter β, which we take to be constant nationwide for simplicity. But travel costs indeed 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. 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.

Column (1) conditions the access regressions on household income and race. Block groups with higher median household income observe lower bank access. Holding constant the block group fixed effects, a doubling in median household income is associated with about 15% fewer expected branch goers per month. The reason behind the negative relation between income and bank access is that residents of richer block groups tend to live farther away from bank branches. In unreported regressions, we find that every doubling in median household income is associated with the nearest branch being located roughly 7% farther away. This greater remoteness is relative to residents nationwide living 1.98 miles away from their nearest branch on average and 1.07 miles away at the median. In Metropolitan core areas, residents are on average 1.21 miles away from their nearest branch, while the median distance away from the nearest branch in Metro cores is 0.91 miles.

The relation between bank access and a block group's Black population share nationwide is positive, but not precisely estimated. Controlling for age in column (2) still preserves the negative relation between income and access, though the magnitude is cut by roughly half. Also controlling for age, block groups with larger Black population shares observe higher access, which is now precisely estimated. Extrapolation of the coefficient implies that a Black resident observes about 11% better access than a White resident. Column (3) restricts the sample to block groups in Metropolitan core areas. There, the coefficient on income is negative (-0.168), and the coefficient on the Black share is negative and precisely estimated. In big cities, a doubling of median household income is associated with about 16.8% weaker access, and extrapolation of the coefficient implies that a Black resident experiences roughly 6% poorer access than a White resident. Once controlling for age in column (4), the coefficient on income remains negative, but the coefficient on the Black population share is positive and precisely estimated.

In columns (5)-(8), we regress the estimated block group fixed effects $\hat{\gamma}_{it}$ from Eq. (15) on demographic attributes of the block groups, both cross-country and in Metro cores. The block group fixed effects capture all local characteristics that contribute to residents visiting any bank branch. These characteristics include both geographic features of the area (e.g., quality of transit networks, presence of geological barriers) and traits of its residents (e.g., average wealth, income, financial sophistication, flexibility in time, car ownership rates, lack of mobile or online access) that altogether influence "demand" for bank branches.

In all columns, richer block groups observe higher fixed effects, which implies that residents of these areas have a greater propensity to visit any bank branch. A doubling in median household income is associated with a 30-47% increase in the number of branch goers per month, holding fixed the measure of access. With controls for age and income, block groups with higher Black population shares, across the country in column (6) and in Metro cores in column (8), observe lower fixed effects, implying a weaker tendency to visit branches. Extrapolations of the coefficient suggest that a block group with a 100% Black share observes about 12% fewer branch visitors per month nationwide compared to a comparable block group with a 100% White share, and roughly 5% fewer branch goers per month in Metro cores.

Overall, results from Table 5.3 show that (i) across the country and in Metro cores, residents of higher income block groups observe lower bank access but are associated with a greater propensity to visit branches, and (ii) there is no robust association between a block group's bank access and its Black population share, but residents of block groups with higher Black shares are associated with a lower propensity to visit any branch.

6 Bank Branch Use

We transition now to estimating the statistical relation between visitor's demographic attributes and their branch use. Across all block groups and year-months over the sample period, the median number of expected block group residents who visit a branch per year-month $\left(\text{i.e., Med}\left(\hat{V}_{it}^{*}\right)\right)$ is 7.16. Table IV presents weighted OLS regressions of bank branch visitation by demographic attributes nationwide and in Metro core areas. Like in the previous section, observations are at the level of a home Census block group per year-month over the 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.

In columns (1)-(5), the dependent variable is the natural logarithm of the expected number of branch goers from each block group based on the month-by-month MSM estimates. That is, $\log \hat{V}_{it}^{*} \equiv \log \sum_{j} \hat{V}_{ijt}^{*}$, where $\hat{V}_{ijt}^{*}$ is the predicted mean of V_{ijt}^{*} in Eq. (15) and β is time-varying. Column (1) reports coefficients on household income and race. The coefficients in the column define a "bank use gradient" as a function of demographic attributes. Higher income block groups have more expected bank branch goers per month (about 21.7% more for every doubling in median household income). Block groups with larger Black population shares have more expected branch goers compared to block groups with larger White population shares. A 1% increase in the Black share is associated with 0.067% more expected bank visitors per month.

Column (2) adds year-month, county, and RUCA fixed effects to the income-racial/ethnic specification of column (1). The positive relation between income and visitation strengthens (from 21.7% to 28.7%), and the coefficient on the Black population share is virtually unchanged (from 0.067 to 0.07). An important factor that might drive branch visits are differences in financial savvy or technical sophistication from differences in age or education (Caskey and Peterson, 1994; Caskey, 1994; Hogarth and O'Donnell, 1997; Hogarth, Anguelov and Lee, 2005; Blank and Barr, 2009; Rhine and Greene, 2013). Controlling for income in columns (1) and (2) already proxies for the permanent component of human capital, but in column (3), we add age shares. The coefficient on income is largely unchanged (0.232), but the coefficient on the Black population share is now negative and not precisely estimated. 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 (about 6.7% more). This finding is consistent with relatively younger, more technologically proficient residents opting for online and mobile banking over visiting branches; and older, less technologically savvy cohorts relying on face-to-face interactions with bankers and tellers over mobile and online banking.

A natural question might be whether residents of poorer block groups or block groups with higher Black population shares substitute branch visitation with greater use of alternative bank methods, such as online or mobile banking. Responses from the FDIC survey suggest not. In Online Appendix C.2, we present evidence from the survey suggesting that lower-income and Black households do not make greater use of online or mobile banking. In fact, both lower income and Black households indicate relying on mobile/online banking less and bank branches more as their most common access method.

Moving away from nationwide estimates of branch use to local ones, we next focus on Metro core areas in column (4). Regressing the log expected number of visitors on income and the racial/ethnic categories with year-month and county fixed effects produces a positive coefficient on income roughly the same as in column (2), which used all block groups. The coefficient on the Black population share increases slightly from 0.07 to 0.097. Column (5) again focuses on Metro core areas, but adds age shares. Here, the magnitude of the Black population share coefficient is no longer precisely estimated.

Finally, columns (6) and (7) use the natural logarithm of the *observed* visitor counts as dependent variables for comparison. Column (6) looks nationwide and column (7) limits observations to Metro cores. Cross-country, the sensitivity of observed visitor counts with respect to income in column (6) is slightly higher (0.244) than the sensitivity of expected visitor counts (0.232) in column (3) that uses the same specification. Compared to column (3), the coefficient on the Black population share is negative and precisely estimated, implying a Black-White gap in branch use of 4.7%. Extrapolations of the coefficient suggest that among Black residents, we expect 4.7% fewer to visit a branch in a month compared to White residents. Focusing on Metro cores in column (7), one can observe that the income gradient is roughly unchanged compared to the same specification in column (5) using expected visitor counts. The Black-White gap in Metro cores is about 2.5%.

The results from the previous section demonstrated that richer block groups observed weaker access but higher "demand" for branch services in the form of larger block group fixed effects. This section reveals that the latter channel dominates the former so as to lead to higher branch use overall among residents of richer block groups.

7 Postal Banking

Having argued that bank access varies substantially across the country, we turn next to evaluating a policy proposal that might improve access for 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). Only limited financial services still remain at some Post Offices, such as domestic and international money orders and wire transfers. 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 and gravity model estimates, we can 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 both access to and use of banking services at branches. From Eq. (17), the expected number of residents in a block group who visit a branch per year-month under a banking system that includes both postal and private banks is affected by five components: (i) the block group's fixed effect γ_{it}, (ii) the fixed effects of both postal and private bank branches λ_{jt}, (iii) the distances between the block group and branches d_{ij}, (iv) the gravity parameter β, and (iv) the set of both postal and private branches available to the block group B_{it}. Our evaluation of a postal banking policy requires assumptions for each component.

Components (iii) and (v) are the least controversial. For the set of branches B_{it}, we include all private bank branches per year-month within a block group's 10-mile radius like before, but now also include all Post Offices within that radius as well. 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 10-mile postal branch choice set, as not all the postal locations we include might 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 shrinking the branch choice set. For component (iii), we measure distances between block groups and branches d_{ij} in the same manner as before using the haversine formula between locations and the population-weighted centers of block groups.

Component (iv) requires an assumption about how the elasticity of branch visitor flows with respect to distance β might change under a postal banking system. Per Section 2, β can be interpreted as the product of consumer's traveling costs and elasticity of substitution between branches. It is reasonable to presume that postal banking will not affect per unit traveling costs. But the elasticity of substitution between postal and private branches might easily differ. One clear reason is that postal banks enable economies of scope that permit residents to spread out fixed costs of travel in a way that private banks cannot, for a person can access financial services at a postal bank when dropping off mail. For simplicity, we assume that all bank branches, postal and private, share the same β per year-month, as estimated in the month-by-month MSM procedure from before, which implicitly presumes a common elasticity of substitution across banking institutions.

The introduction of a postal banking system would reasonably affect component (i), a block group's fixed effect γ_{it}, which captures all attributes of the block group that influences "demand" for any branch's services. The clearest change is postal banking inducing account ownership among the unbanked. If the policy had such an effect, residents of the block group who were once non branch goers would likely become new visitors, which would raise the block group's fixed effect and imply greater expected branch use. Rather than speculating the change in the fixed effect per block group from a postal banking policy, we instead fix them at their estimated values from before. Doing so means that their impact on branch use in the policy evaluation will likely be underestimated.

Finally, the branch fixed effects λ_{jt} of component (ii) is the most challenging to manage. Undoubtedly, the private bank fixed effects would change under a postal banking system. Residents might substitute away from a private bank toward a postal bank, which would reduce the average visitor count of the private bank and cut into its fixed effect. Alternatively, private banks would almost surely respond endogenously to the new competition from postal banks, perhaps with new price promotions or investments in staff or infrastructure, so as to lift their branches' perceived "quality," which would increase the fixed effects. For simplicity, we assume away any changes in private bank fixed effects, and instead apply their estimated fixed effects from before. By presuming both unchanged block group and private bank fixed effects, our approach is a partial impact assessment of a postal banking policy that does not account for the general equilibrium effects on consumer and producer behavior 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.

Not only must we assume estimated values of fixed effects for private banks under a postal banking system, we must also assign fixed effects to the new postal banks. Here, we consider a set of possible fixed effects to produce a range of estimates on both bank use and access under a postal banking policy. We first assume that all postal banks per year-month in the sample share the same fixed effect. This assumption is simple, but restrictive, because it ignores potential variation in postal bank quality cross-country. Second, we assign three fixed effects to postal banks based on different parts of the distribution of private bank fixed effects per year-month: the 10th percentile, 50th percentile, and 90th percentile. The first assignment implicitly assumes that the “quality” of postal banks would be that of the bottom 10 percent of private banks per year-month. We call this a “low quality” postal banking system. Similarly, the 50th percentile assumes that the typical postal bank would have the quality of the median private bank per year-month (i.e., a “medium quality” system), and the 90th percentile assumes that postal banks are perceived to have the same quality as the top 10 percent of private banks per year-month (i.e., a “high quality” system).[17]

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).

Without postal banking, the median number of residents from a typical block group nationwide who visit a private bank branch was 7.16 per month. Median branch use would rise to 7.67 per month (7% increase) under a “low quality” postal banking system, 9.80 per month (37% increase) under a “medium quality” postal banking system, and 18.74 per month (162% increase) under a “high quality” postal banking system.

To measure the extent to which bank access would change under postal banking, we re-run the

[17]Postal banks being perceived as high quality is acquainted with the notion that a government-sponsored banking institution is considered more trustworthy than private banks (Office of the USPS Inspector General, 2014; Baradaran, 2015). In fact, among respondents to the 2019 Survey of Consumer Finances (SCF), the choice “do not like dealing with banks” is the second-most cited reason for families not having a checking account, and the fraction of respondents selecting this option as their reason has increased steadily over time. The 2019 SCF tables can be found here.

access regressions from Table III, but in computing Φ_{it} per Census block group, we now include the locations of all Post Office branches that are registered in SafeGraph within a 10-mile radius. To make the policy evaluation comparable to our analysis earlier that considered only private banks, we only include block groups that have a private bank branch located within a 10-mile radius in the year-month. This implies that block groups which would only have a new postal bank within 10 miles are ignored from the analysis.

The results are in Table V. The coefficient on income in the nationwide estimates in column (2) under a medium quality system is -0.103. This value contrasts to the corresponding coefficient on income in Table III of -0.076. Hence, a postal banking system of medium quality steepens the negative income gradient in bank access by 2.7 percentage points, which implies that access will improve relatively more for residents of poorer block groups than richer ones. Under a low quality postal banking system, the income gradient still sharpens, but by only 0.7 percentage points, and under a high quality system, the improvement in access rises to 6.8 percentage points. The coefficient of Black population share in column (2) is 0.079, which contrasts with a coefficient of 0.047 in III, implying a 1.3 percentage point increase in access for block groups with higher Black population shares under a medium quality postal banking system. The range in the access improvement from the low to high quality postal banking system is 0.2-3.4 percentage points for block groups with higher Black population shares. Similarly, a hypothetical block group that had a 100% Hispanic population share would observe between a 0.9-6.2 percentage point increase in access from a postal banking system, compared to a hypothetical block group with a 100% White population share.

Zeroing in on Metro cores, we find that the improvement in access would be even large for block groups with higher Black or Hispanic population shares. For the Black population share, the coefficients imply a 0.7-7.7 percentage point increase in access compared to the coefficient in column (4) in Table III. Similarly, the range of access improvement for Hispanic residents is between 1.1-9.1 percentage points.

Overall, we find that a postal banking policy would have the biggest impact on bank access and use in poorer block groups and block groups with higher Black or Hispanic population shares, particularly in big cities.

8 Conclusion

We use anonymous location data from millions of mobile devices to develop a new micro-level measure of bank access. Our measure is derived from a spatial gravity model and is a function of a local area's distance from surrounding bank branches and branch characteristics. To overcome distortions in the mobility data that safeguard user privacy, we estimate the access measure using the Method of Simulated Moments and a novel computational routine that can handle hundreds of thousands of fixed effects.

We document significant geographic variation in access. Nationwide, the most pronounced differences in access are between urban and rural areas. There is substantial within-city variation in bank access as well, with different parts of Los Angeles, Washington, DC, Chicago, and New York City observed vastly different access estimates, for example.

Finally, we use the access measure to evaluate a policy of postal banking. Depending on the assumed quality of a postal banking system, we estimate that the policy would improve bank access between 0.7-6.8 percentage points in low income areas and 0.2-6.2 percentage points in areas with higher Black and Hispanic population shares. We estimate that a postal banking system's effect on access by racial shares would be even large in big cities.

References

Adda, Jerome and Russell Cooper, *Dynamic Economics: Quantitative methods and applications*, MIT press, 2003.

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.

Amel, Dean F, Arthur B Kennickell, and Kevin B Moore, "Banking market definition: evidence from the survey of consumer finances," *Federal Reserve Board Finance and Economics Discussion Series*, 2008.

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, "Access to financial services, savings, and assets among the poor," *National Poverty Center Policy Brief (13)*, 2008.

_ **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.

Célerier, Claire and Adrien Matray, "Bank-branch supply, financial inclusion, and wealth accumulation," *The Review of Financial Studies*, 2019, *32* (12), 4767–4809.

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.

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

Dahl, Drew and Michelle Franke, ""Banking deserts" become a concern as branches dry up," *Population*, 2017, *12* (21.2), 5–9.

Davidson, Russell and James G MacKinnon, *Econometric Theory and Methods*, Vol. 5, Oxford University Press New York, 2004.

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.

Eaton, Jonathan and Samuel Kortum, "Technology, geography, and trade," *Econometrica*, 2002, *70* (5), 1741–1779.

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

Evans, James W, "Simulated Method of Moments (SMM) Estimation," Jul 2018.

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.

Goodstein, Ryan M and Sherrie LW Rhine, "The effects of bank and nonbank provider locations on household use of financial transaction services," *Journal of Banking & Finance*, 2017, *78*, 91–107.

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

Hanson, Gordon H, "Market potential, increasing returns and geographic concentration," *Journal of International Economics*, 2005, *67* (1), 1–24.

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

Harris, Chauncy D, "The market as a factor in the localization of industry in the United States," *Annals of the Association of American geographers*, 1954, *44* (4), 315–348.

Head, Keith and Thierry Mayer, "Market potential and the location of Japanese investment in the European Union," *Review of Economics and Statistics*, 2004, *86* (4), 959–972.

_ and _ , "Gravity equations: Workhorse, toolkit, and cookbook," in "Handbook of International Economics," Vol. 4 2014, pp. 131–195.

Ho, Katherine and Joy Ishii, "Location and competition in retail banking," *International Journal of Industrial Organization*, 2011, *29* (5), 537–546.

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.

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.

McFadden, Daniel, "The measurement of urban travel demand," *Journal of Public Economics*, 1974, *3* (4), 303–328.

_ , "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 1989, pp. 995–1026.

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.

Morgan, Donald P, Maxim L Pinkovskiy, Bryan Yang et al., "Banking deserts, branch closings, and soft information," 2016. Unpublished Working Paper. Federal Reserve Bank of New York. New York, NY.

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.

Redding, Stephen and Anthony J Venables, "Economic geography and international inequality," *Journal of International Economics*, 2004, *62* (1), 53–82.

Rhine, Sherrie LW and William H Greene, "Factors that contribute to becoming unbanked," *Journal of Consumer Affairs*, 2013, *47* (1), 27–45.

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.

Sousa, José De, Thierry Mayer, and Soledad Zignago, "Market access in global and regional trade," *Regional Science and Urban Economics*, 2012, *42* (6), 1037–1052.

Squire, Ryan Fox, "Quantifying sampling bias in SafeGraph Patterns," *SafeGraph Blog*, October 2019.

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

Thaenraj, Pranav, "Identifying CBG Sinks," *SafeGraph Blog*, 2021.

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.

Washington, Ebonya, "The impact of banking and fringe banking regulation on the number of unbanked Americans," *Journal of Human Resources*, 2006, *41* (1), 106–137.

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

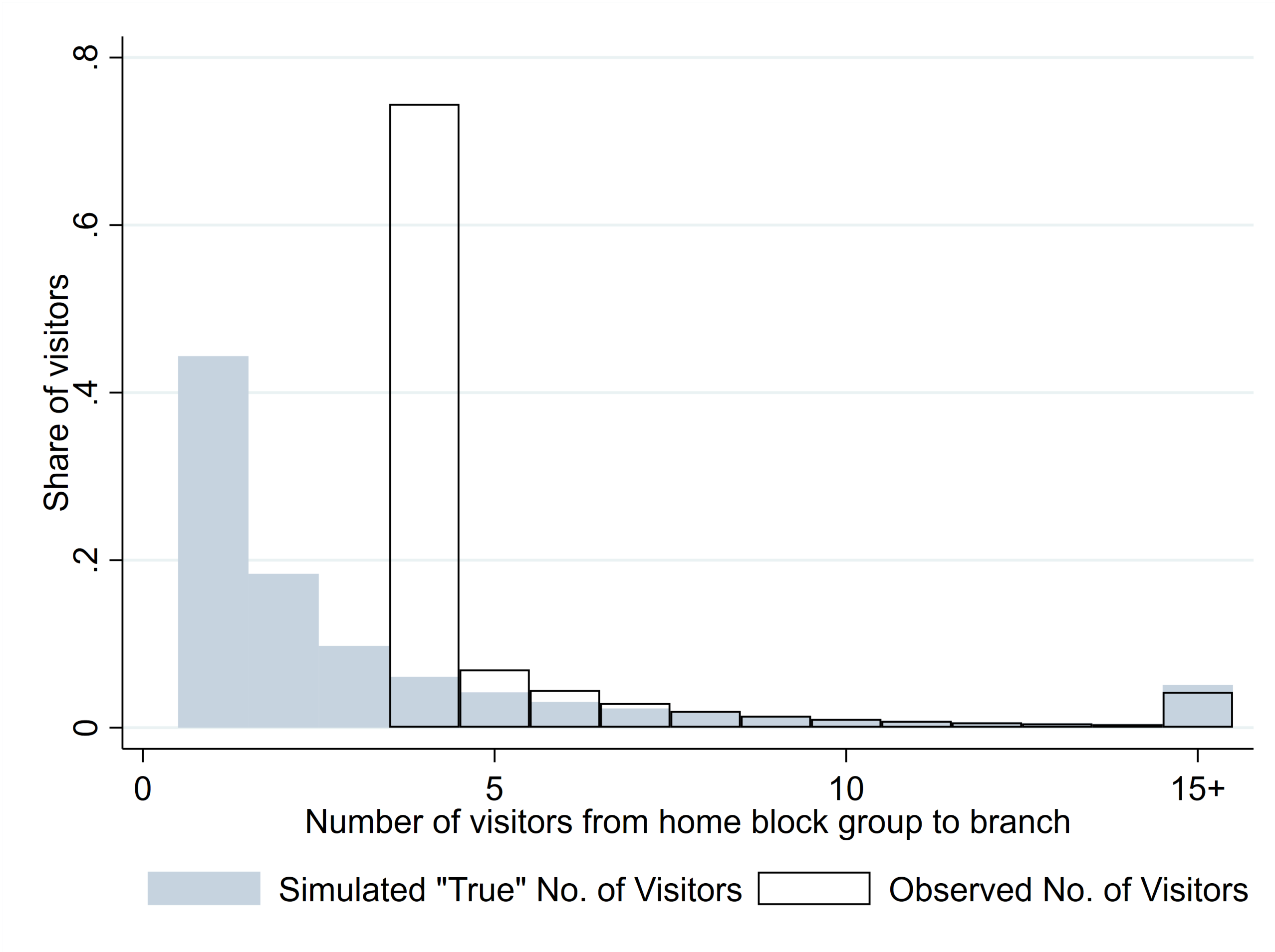


FIGURE I

DISTRIBUTIONS OF OBSERVED AND SIMULATED "TRUE" VISITOR COUNTS

Notes. The figure presents distributions of observed visitor counts and simulated "true" visitor counts from visitors' home Census block groups to bank branches. Observed visitor counts, denoted V_{ijt} from Eq. (14), are the raw data from our core SafeGraph sample ranging from January 2018 to December 2019. The core sample includes only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits. Among the observed visitor counts, 74.5% equal 4, 4.3% equal 15 or more, and 0.47% equal 100 or more. Simulated "true" visitor counts, denoted V^*_{ijt} from Eq. (15), are draws from the underlying "true" distribution of visitors, which we assume to be Poisson. The simulated values are computed from the month-by-month Method of Simulated Moments estimation described in Online Appendix B. The distribution of simulated visitor counts includes all positive draws from all simulations across every year-month in the sample. Among the simulated visitor counts, 72.5% are less than 4, 6.1% equal 4, 5.1% equal 15 or more, and 0.98% equal 100 or more. To enhance their depictions, we censor the distributions at 15 visitors. That is, the share of visitor counts equaling or exceeding 15 is assigned to 15+ visitors in the figure.

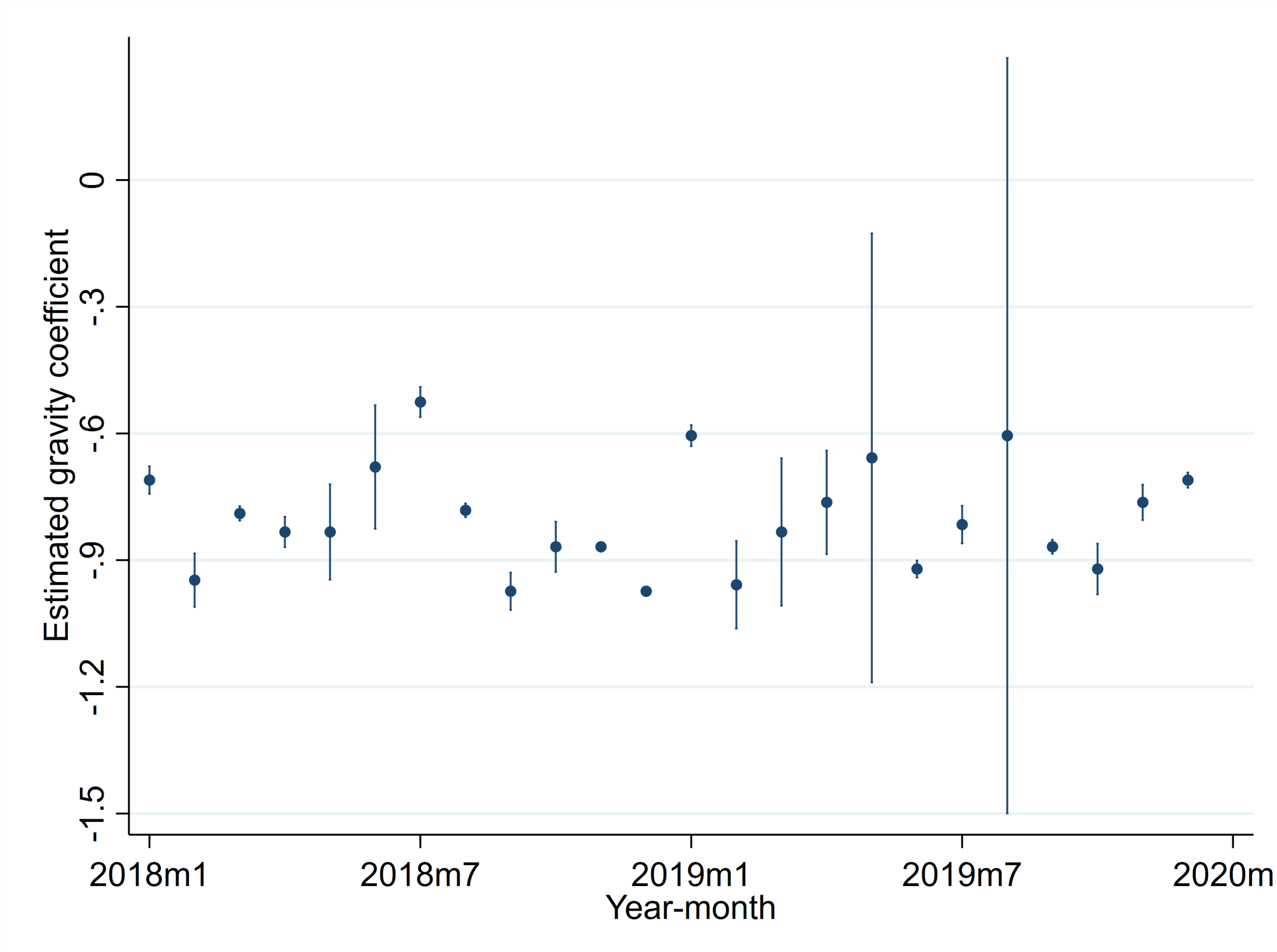
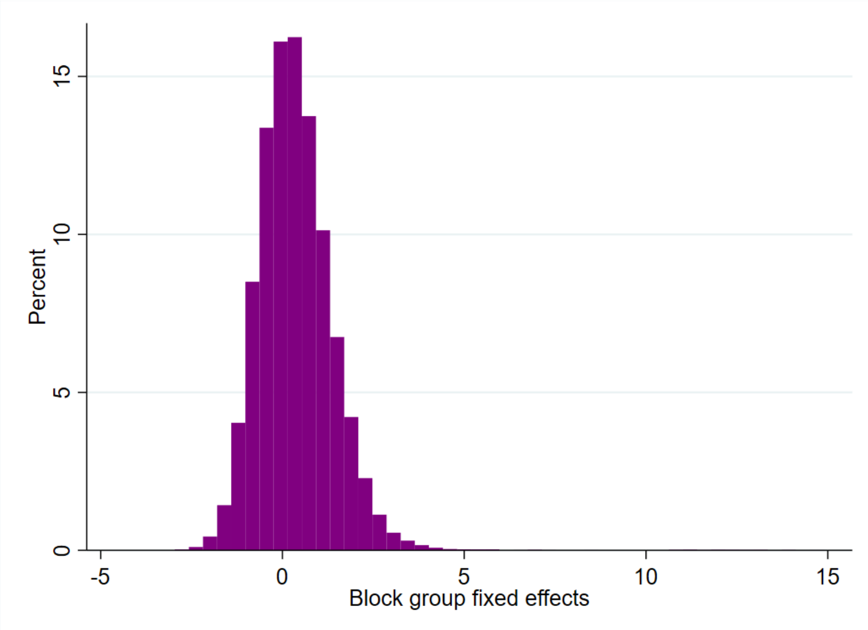


FIGURE II

TIME SERIES OF $-\hat{\beta}_{MSM}$ GRAVITY COEFFICIENT ESTIMATES

Notes. The figure presents the monthly time series of $-\hat{\beta}_{\text{MSM}}$ gravity coefficient estimates from the month-by-month Method of Simulated Moments (MSM) estimation, along with 95% confidence intervals. The parameter β is the elasticity of visitor flows with respect to distance, as measured in the gravity model of Eq. (15). The step-by-step details of the MSM estimation are provided in Online Appendix B.

(A) Census Block Groups



(B) Bank Branches

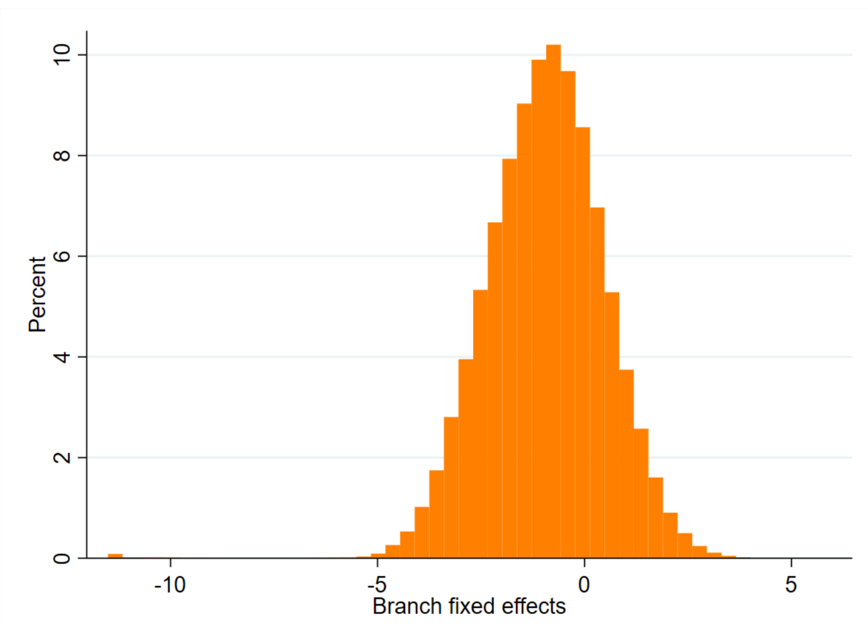


FIGURE III
HISTOGRAMS OF ESTIMATED CENSUS BLOCK GROUP AND BANK BRANCH FIXED EFFECTS

Notes. The figure presents histograms of the estimated Census block group and bank branch fixed effects from the Method of Simulated Moments (MSM) estimation of the visitor count gravity relation in Eq. (15). In each histogram, the fixed effects are grouped into 50 equally-sized bins. The histograms present the estimated fixed effects for all months in the sample period under the month-by-month MSM estimation. The steps of the MSM procedure are in Online Appendix B. From the notation in that appendix, Panel A presents a histogram of $\hat{\gamma}_{it}^{\infty}$, whereas Panel B presents a histogram of $\hat{\lambda}_{jt}^{\infty}$.

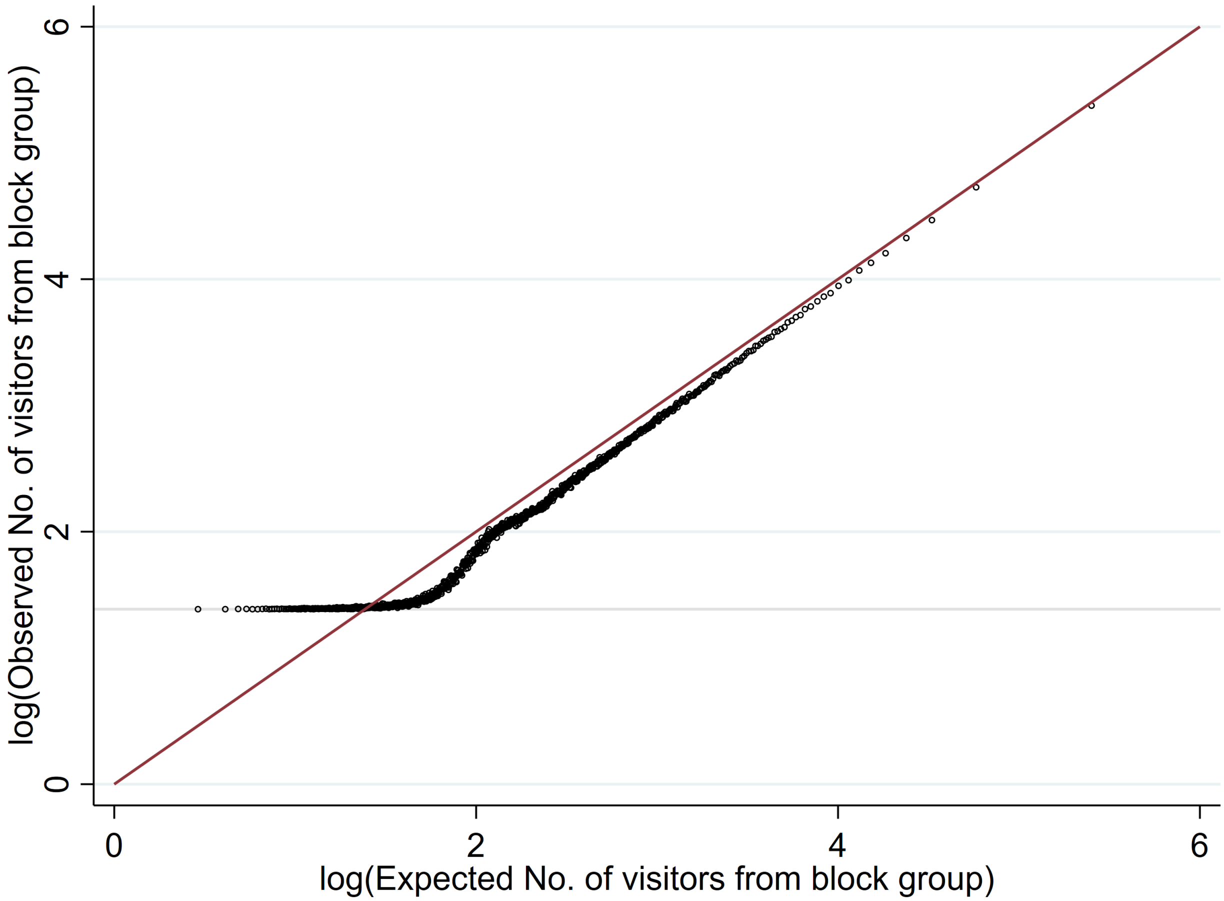


FIGURE IV

OBSERVED VS. EXPECTED BRANCH VISITORS PER CENSUS BLOCK GROUP

Notes. The figure presents a scatter plot of the log observed number of branch visitors from each Census block group (i.e., $V_{it} \equiv \sum_j V_{ijt}$, where V_{ijt} is given in Eq. 14) versus the log expected number of branch visitors from each block group based on the month-by-month MSM estimates (i.e., $\hat{V}^*_{it} \equiv \sum_j \hat{V}^*_{ijt}$, where $\hat{V}^*_{ijt}$ is the predicted mean of V^*_{ijt} in Eq. 15 and β is time-varying). The observed and expected number of visitors range over the full sample period from January 2018 to December 2019. Each dot represents a Census block group in a year-month. The red solid line is a 45° line and the light grey solid line cuts the y-axis at 1.4, which corresponds to SafeGraph's censoring at 4 visitor counts. The steps of the MSM procedure that generate the expected number of branch goers are in Online Appendix B.

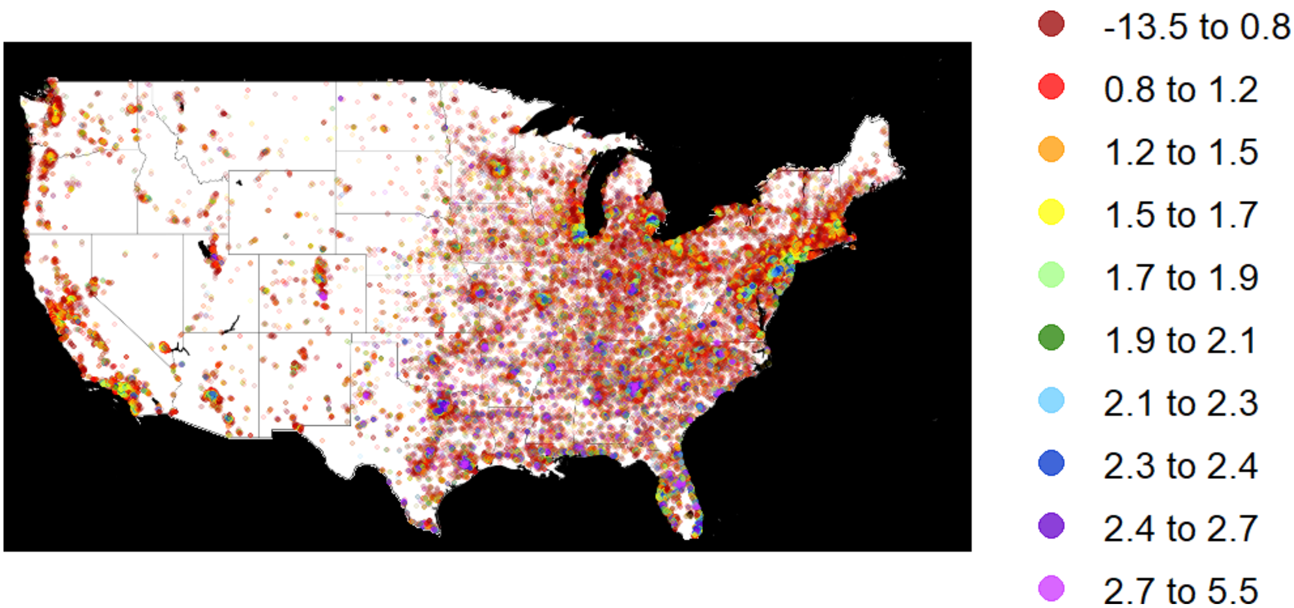


FIGURE V
BANK ACCESS NATIONWIDE

Notes. The figure illustrates a dot density map of bank access by Census block groups nationwide. The figure is based on our core SafeGraph sample of bank branches; i.e., only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits. Each dot is positioned at a block group's center of population. Bank access estimates are calculated from Eq. (16) and are based on the Method of Simulated Moments estimation described in Online Appendix B. Access estimates are calculated month-by-month per block group, and the figure presents weighted monthly averages, where each month's weight is its share of the block group's total branch visitors over the core sample period (January 2018 - December 2019). The map is constructed by grouping block groups into deciles and shading the dots so that higher-ordered colors in the rainbow gradient (i.e., indigo and violet) imply higher bank access values and lower-ordered colors (i.e., red and orange) imply lower access values. Block groups where no resident was recorded in SafeGraph as having visited a branch in the sample period and block groups having no bank branch within a 10-mile radius are shaded white.

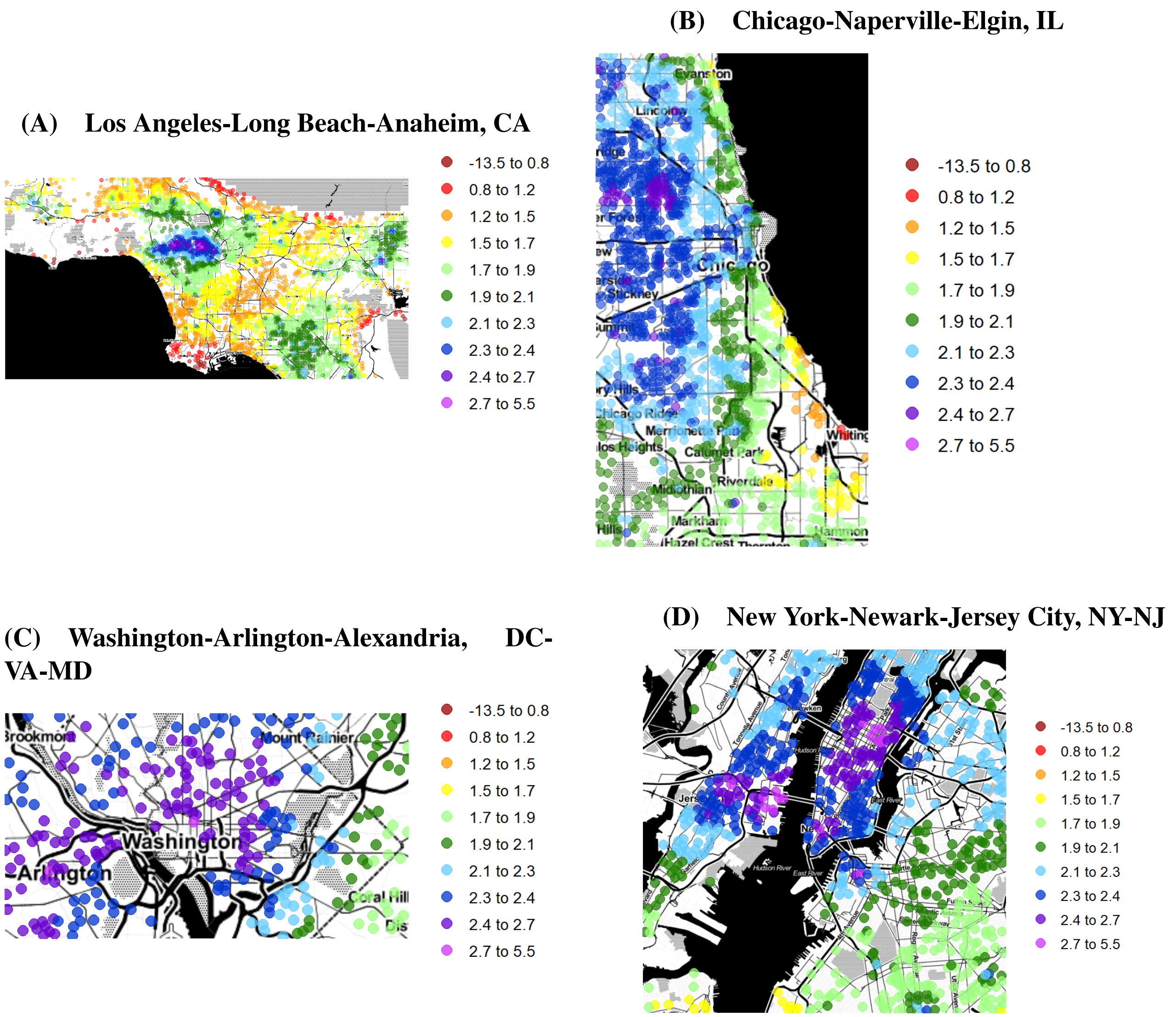


FIGURE VI

BANK ACCESS BY METROPOLITAN STATISTICAL AREAS

Notes. The figure illustrates dot density maps of bank access by Census block groups across four Metropolitan Statistical Areas (MSAs). The figure is based on our core SafeGraph sample of bank branches; i.e., only businesses in SafeGraph with NAICS codes equal to 522110 (Commercial Banking), 522120 (Savings Institutions), or 551111 (Offices of Bank Holding Companies) for which we have visitor data and whose brands are also listed in the FDIC's 2019 Summary of Deposits. Each dot in a panel is positioned at a block group's center of population. Bank access estimates are calculated from Eq. (16) and are based on the Method of Simulated Moments estimation described in Online Appendix B. Access estimates are calculated month-by-month per block group, and the panels present weighted monthly averages, where each month's weight is its share of the block group's total branch visitors over the core sample period (January 2018 - December 2019). Each panel's map is constructed by grouping block groups within the MSA into deciles and shading the dots so that higher-ordered colors in the rainbow gradient (i.e., indigo and violet) imply higher bank access values and lower-ordered colors (i.e., red and orange) imply lower access values. Block groups where no resident was recorded in SafeGraph as having visited a branch in the sample period and block groups having no bank branch within a 10-mile radius are shaded white.

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 FDIC's 2019 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 FDIC's 2019 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 SOD. 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 ACCESS AND BLOCK GROUP FIXED EFFECTS BY DEMOGRAPHIC ATTRIBUTES

Dep. var.:	log(Bank access of block groups)				Block group fixed effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	-0.151	-0.076	-0.168	-0.098	0.438	0.309	0.466	0.332
	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)	(0.008)	(0.009)	(0.009)
Black	0.015	0.110	-0.061	0.047	0.055	-0.118	0.158	-0.048
	(0.011)	(0.012)	(0.011)	(0.012)	(0.019)	(0.019)	(0.020)	(0.021)
Asian	0.870	0.851	0.790	0.780	-0.352	-0.349	-0.264	-0.283
	(0.022)	(0.022)	(0.021)	(0.021)	(0.033)	(0.032)	(0.034)	(0.032)
Other	0.360	0.434	0.324	0.422	-0.242	-0.408	-0.182	-0.403
	(0.046)	(0.046)	(0.048)	(0.049)	(0.064)	(0.064)	(0.076)	(0.076)
Hispanic	0.162	0.283	0.108	0.249	-0.050	-0.294	0.025	-0.265
	(0.013)	(0.014)	(0.013)	(0.014)	(0.023)	(0.025)	(0.026)	(0.027)
Age <15		-1.039		-1.133		1.902		2.063
		(0.030)		(0.030)		(0.053)		(0.060)
Age 35-54		-0.493		-0.361		1.058		0.986
		(0.035)		(0.037)		(0.046)		(0.052)
Age 55-64		-0.897		-0.686		0.830		0.577
		(0.038)		(0.039)		(0.052)		(0.059)
Age 65+		-0.070		-0.117		0.361		0.385
		(0.026)		(0.026)		(0.037)		(0.042)
log(No. of Devices)	-0.034	-0.036	-0.049	-0.049	0.473	0.471	0.481	0.476
	(0.004)	(0.004)	(0.004)	(0.004)	(0.011)	(0.010)	(0.012)	(0.011)
Observations	2,669,220	2,669,220	2,033,884	2,033,884	2,669,220	2,669,220	2,033,884	2,033,884
Adjusted R^2	0.498	0.504	0.518	0.531	0.378	0.392	0.374	0.395
Sample	Core	Core	MC	MC	Core	Core	MC	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. Block groups where no resident was recorded in SafeGraph as having visited a branch in the year-month and block groups having no bank branch within a 10-mile radius in the year-month are dropped. Observations are weighted by the number of mobile devices residing in the block groups in the year-months. All columns 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 FDIC's 2019 Summary of Deposits (SOD). Demographic independent variable observations are population-based decimal shares from the 2019 5-year American Community Survey. Income is median household income. In columns (1)-(4), the dependent variable is the log estimated bank access measure $\log \hat{\Phi}_{it}$ from Eq. (16), whereas in columns (5)-(8), the dependent variable is the estimated block group fixed effects $\hat{\gamma}_{it}$ from the gravity relation in Eq. (15). Both dependent variables are computed from the month-by-month Method of Simulated Moments estimation described in Online Appendix B. 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 (Metropolitian 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 USE BY DEMOGRAPHIC ATTRIBUTES

Dep. var.:	log(Expected no. of visitors)					log(Observed no. of visitors)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Income)	0.217	0.287	0.232	0.298	0.234	0.244	0.254
	(0.007)	(0.007)	(0.007)	(0.008)	(0.008)	(0.006)	(0.007)
Black	0.067	0.070	-0.008	0.097	-0.001	-0.047	-0.025
	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)	(0.014)	(0.016)
Asian	-0.099	0.518	0.502	0.526	0.497	0.393	0.406
	(0.035)	(0.029)	(0.028)	(0.029)	(0.029)	(0.025)	(0.026)
Other	-0.782	0.118	0.026	0.141	0.019	-0.015	-0.016
	(0.055)	(0.050)	(0.050)	(0.061)	(0.062)	(0.042)	(0.056)
Hispanic	0.072	0.112	-0.011	0.133	-0.016	-0.092	-0.092
	(0.017)	(0.019)	(0.021)	(0.021)	(0.023)	(0.020)	(0.022)
Age <15			0.863		0.930	0.921	1.000
			(0.047)		(0.053)	(0.036)	(0.042)
Age 35-54			0.565		0.625	0.594	0.635
			(0.038)		(0.044)	(0.033)	(0.039)
Age 55-64			-0.067		-0.110	-0.034	-0.110
			(0.040)		(0.047)	(0.037)	(0.045)
Age 65+			0.291		0.268	0.200	0.179
			(0.031)		(0.034)	(0.027)	(0.030)
log(No. of Devices)	0.525	0.439	0.435	0.432	0.426	0.527	0.511
	(0.016)	(0.009)	(0.009)	(0.010)	(0.010)	(0.008)	(0.009)
Constant	-2.675						
	(0.090)						
Observations	2,669,246	2,669,220	2,669,220	2,033,884	2,033,884	3,134,720	2,246,239
Adjusted R^2	0.315	0.454	0.459	0.452	0.459	0.538	0.536
Sample	Core	Core	Core	MC	MC	Core	MC
Year-month FE		O	O	O	O	O	O
County FE		O	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 year-month in the sample period from January 2018 - December 2019. Block groups where no resident was recorded in SafeGraph as having visited a branch in the year-month and block groups having no bank branch within a 10-mile radius in the year-month are dropped. 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 FDIC's 2019 Summary of Deposits. Demographic independent variable observations are population-based decimal shares from the 2019 5-year American Community Survey. Income is median household income. The dependent variable in columns (1)-(6) is the natural logarithm of the expected number of branch goers from each block group based on the month-by-month Method of Simulated Moments estimates; i.e., $\log \hat{V}^*_{it} \equiv \log \sum_j \hat{V}^*_{ijt}$, where $\hat{V}^*_{ijt}$ is the predicted mean of V^*_{ijt} in Eq. (15) and β is time-varying. The dependent variable in columns (7) and (8) is the natural logarithm of the observed number of branch goers from each Census block group; i.e., $\log V_{it} \equiv \log \sum_j V_{ijt}$, where V_{ijt} is given in Eq. (14). 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 V
BANK ACCESS BY DEMOGRAPHIC ATTRIBUTES UNDER POSTAL BANKING

Dep. var.:	log(Bank Access of block groups)							
USPS branch quality:	Median				Low		High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	-0.177	-0.103	-0.187	-0.116	-0.083	-0.102	-0.144	-0.148
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
Black	0.032	0.123	-0.028	0.079	0.112	0.054	0.144	0.124
	(0.011)	(0.011)	(0.011)	(0.011)	(0.012)	(0.012)	(0.010)	(0.011)
Asian	0.776	0.753	0.707	0.694	0.827	0.759	0.621	0.575
	(0.019)	(0.019)	(0.019)	(0.019)	(0.021)	(0.020)	(0.018)	(0.019)
Other	0.300	0.363	0.293	0.385	0.416	0.416	0.255	0.314
	(0.041)	(0.041)	(0.044)	(0.044)	(0.044)	(0.047)	(0.038)	(0.042)
Hispanic	0.196	0.312	0.151	0.290	0.292	0.260	0.345	0.341
	(0.012)	(0.013)	(0.012)	(0.014)	(0.014)	(0.014)	(0.013)	(0.013)
Age <15		-1.058		-1.160	-1.041	-1.138	-1.100	-1.214
		(0.028)		(0.030)	(0.029)	(0.030)	(0.028)	(0.030)
Age 35-54		-0.485		-0.371	-0.489	-0.362	-0.483	-0.399
		(0.029)		(0.031)	(0.033)	(0.035)	(0.026)	(0.028)
Age 55-64		-0.872		-0.679	-0.888	-0.683	-0.843	-0.681
		(0.034)		(0.036)	(0.037)	(0.038)	(0.032)	(0.035)
Age 65+		-0.120		-0.152	-0.081	-0.124	-0.206	-0.216
		(0.023)		(0.024)	(0.025)	(0.025)	(0.021)	(0.022)
log(No. of Devices)	-0.041	-0.043	-0.051	-0.051	-0.038	-0.049	-0.056	-0.060
	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)
Observations	2,669,220	2,669,220	2,033,884	2,033,884	2,669,220	2,033,884	2,669,220	2,033,884
Adjusted R^2	0.602	0.611	0.535	0.551	0.577	0.537	0.692	0.658
Sample	Core	Core	MC	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. Block groups where no resident was recorded in SafeGraph as having visited a private bank branch in the year-month and block groups having no private bank branch within a 10-mile radius in the year-month are dropped. Observations are weighted by the number of mobile devices residing in the block groups in the year-months. All columns use our core sample of private 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 FDIC's 2019 Summary of Deposits plus businesses in SafeGraph with NAICS codes equal to 491110 (Postal Services) for which we have visitor data. Demographic independent variable observations are population-based decimal shares from the 2019 5-year American Community Survey. Income is median household income. In all columns, the dependent variable is the log estimated bank access measure $\log \hat{\Phi}_{it}$ from Eq. (16) that includes both private bank branches and Post Office branches. The dependent variable is computed from the month-by-month Method of Simulated Moments estimation described in Online Appendix B. In columns (1)-(4), we assign to each Post Office location per year-month an establishment fixed effect λ_{jt} equal to the 50th percentile of the distribution of private bank fixed effects in the year-month. In columns (5) and (6), we assign the 10th percentile and in columns (7) and (8), we assign the 90th percentile. Columns (1), (2), (5), and (7) include all block groups for which we have visitor data, whereas columns (3), (4), (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 VI
BANK BRANCH USE BY DEMOGRAPHIC ATTRIBUTES UNDER POSTAL BANKING

Dep. var.:	log(Expected no. of visitors)							
USPS branch quality:	Median				Low		High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	0.261	0.205	0.279	0.216	0.225	0.230	0.165	0.184
	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)
Black	0.087	0.005	0.130	0.031	-0.006	0.007	0.026	0.077
	(0.014)	(0.014)	(0.015)	(0.016)	(0.015)	(0.016)	(0.014)	(0.015)
Asian	0.424	0.404	0.443	0.411	0.477	0.475	0.272	0.292
	(0.027)	(0.027)	(0.028)	(0.027)	(0.028)	(0.028)	(0.026)	(0.027)
Other	0.058	-0.045	0.110	-0.018	0.008	0.013	-0.153	-0.089
	(0.048)	(0.049)	(0.060)	(0.060)	(0.050)	(0.061)	(0.049)	(0.059)
Hispanic	0.146	0.018	0.176	0.025	-0.002	-0.005	0.051	0.076
	(0.019)	(0.020)	(0.020)	(0.022)	(0.021)	(0.022)	(0.020)	(0.021)
Age <15		0.844		0.903	0.861	0.925	0.802	0.849
		(0.045)		(0.051)	(0.046)	(0.053)	(0.043)	(0.049)
Age 35-54		0.573		0.615	0.569	0.624	0.575	0.587
		(0.036)		(0.042)	(0.037)	(0.043)	(0.036)	(0.041)
Age 55-64		-0.042		-0.102	-0.058	-0.106	-0.013	-0.104
		(0.039)		(0.045)	(0.040)	(0.046)	(0.039)	(0.045)
Age 65+		0.241		0.233	0.280	0.261	0.155	0.169
		(0.030)		(0.033)	(0.031)	(0.034)	(0.029)	(0.033)
log(No. of Devices)	0.432	0.427	0.430	0.424	0.432	0.426	0.414	0.415
	(0.009)	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)
Observations	2,669,220	2,669,220	2,033,884	2,033,884	2,669,220	2,033,884	2,669,220	2,033,884
Adjusted R^2	0.349	0.354	0.398	0.405	0.402	0.440	0.351	0.418
Sample	Core	Core	MC	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 year-month in the sample period from January 2018 - December 2019. Block groups where no resident was recorded in SafeGraph as having visited a private bank branch in the year-month and block groups having no private bank branch within a 10-mile radius in the year-month are dropped. 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 private 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 FDIC's 2019 Summary of Deposits plus businesses in SafeGraph with NAICS codes equal to 491110 (Postal Services) for which we have visitor data. Demographic independent variable observations are population-based decimal shares from the 2019 5-year American Community Survey. Income is median household income. The dependent variable is the natural logarithm of the expected number of branch goers from each block group based on the month-by-month Method of Simulated Moments estimates; i.e., $\log \hat{V}^*_{it} \equiv \log \sum_j \hat{V}^*_{ijt}$, where $\hat{V}^*_{ijt}$ is the predicted mean of V^*_{ijt} in Eq. (15) and β is time-varying. In columns (1)-(4), we assign to each Post office location per year-month an establishment fixed effect λ_{jt} equal to the 50th percentile of the distribution of private bank fixed effects in the year-month. In columns (5) and (6), we assign the 10th percentile and in columns (7) and (8), we assign the 90th percentile. Columns (1), (2), (5), and (7) include all block groups for which we have branch visitor data, whereas columns (3), (4), (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.

Online Appendix to

Bank Access Across America[*]

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

February 15, 2022

[*]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

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. To protect user privacy, SafeGraph employs several masking methods to the visitor home block group variable. First, it adds Laplace noise to its values. Second, after the noise is added, Safegraph truncates the variable by only reporting data from block groups that observe at least two devices. Third, home block groups with only two, three, or four devices are reported as having four devices.

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

[1] See the SafeGraph Places Manual and Data Guide for more details.

[2] See the SafeGraph Privacy Policy for more details.

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]

A.2 FDIC Summary of Deposits

To construct our mobility dataset, we rely on branch information from the Federal Deposit Insurance Corporation (FDIC). Branch data are from the FDIC's 2019 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 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. Additional technical information about Placekey can be found in their white paper located here: Placekey White Paper.

[3]Full details of the algorithm are found here: Home Identification Algorithm.
[4]FDIC SOD data are located here: SOD.

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.

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 visits 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 and for which we have visitor information. 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 Method of Simulated Moments

This section presents the steps used in the method of simulated moments (MSM) to estimate the parameters of the gravity equation that enter our measure of local bank access. The MSM is run both separately per year-month of our sample and across the entire panel.

B.1 Step 1: Specify the DGP for visitors

The data generating process (DGP) we seek to simulate is the number of visitors between block groups and branches through time. To account for measurement error in the SafeGraph data and unobservable factors that influence visits, we assume that the true visitor count V_{ijt}^{*} is Poisson distributed with mean $\pi_{ijt}R_{it}$, where R_{it} is the number of residents of block group i and π_{ijt} is the share of residents living in block group i who visit branch j in year-month t. Using the gravity equation from the conceptual framework (Eq. 10) to inform π_{ijt}, we express the true visitor count as obeying Eq. (15). We measure distance in miles between branches and the population-weighted center of visitors' home block groups. We use the haversine formula to calculate distance, which accounts for the curvature of the Earth.[5]

B.2 Step 2: Reduce the branch choice set

Technically speaking, every branch in our sample is in the choice set of each resident and ought to enter a block group's measure of bank access. A resident of a block group in, say, New York City, could travel to a First Republic Bank branch in San Francisco three thousand miles away. But the chances are low, and a sample of over fifty thousand branches, over two hundred thousand block groups, altogether spanning twenty-four months, makes it computationally infeasible to calculate an access measure that accounts for all possible block group-branch pairs.

Instead, we reduce the choice set per block group to include only those branches located within a 10-mile radius around the block group's center of population. Any method that reduces the choice set of branches available to residents must be uniform across block groups to avoid potentially introducing spurious correlation between the local access measure and characteristics of the block group. We choose a distance-based cutoff of 10 miles because that radius has been commonly used in defining "bank deserts" (Morgan et al., 2016; Dahl and Franke, 2017). That is, any area without a branch within 10 miles is a bank desert. By choosing this definition, the literature implicitly regards branches located outside a 10-mile radius as unreasonable destinations. Online Appendix D.1 expands the choice set for robustness.

If residents of block group i visit a branch outside their 10-mile radius, we drop those block group-branch visitor counts from the data. If residents do not visit a branch within their 10-mile radius, it means that either no residents truly visited that branch or the true visitor count was hidden by SafeGraph's differential privacy. In these cases, we input the observed visitor count as 0, which accommodates either scenario. Let B_{it} denote the set of branches in the 10-mile radius of block group i in year-month t, including those with inputted 0 visitor counts.

B.3 Step 3: Initialize the fixed effects routine

The MSM uses the visitor data v and the model parameters $\psi \equiv \{\beta, \gamma_{it}, \lambda_{jt}\}$ to minimize the distance between simulated model moments and data moments. With the very large number of block groups and branches in our sample, the model of visitor counts in Eq. (15) requires thousands of fixed effects to be estimated. There are simply too many parameters to identify from the MSM minimization problem alone. Instead, we adopt an iterative routine to identify the fixed effects $\{\gamma_{it}, \lambda_{jt}\}$ and let the minimization problem identify β. Holding fixed an estimate of β and given

[5] The centers of population are computed using population counts from the 2010 Census and are found here: 2010 Census Centers of Population. 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.

initial estimates of the fixed effects, the routine updates the fixed effects estimates until they converge. After the fixed effects converge per estimate of β, the MSM minimization problem then chooses the optimal β estimate that satisfies the moment conditions. In the month-by-month estimation, we estimate ψ per year-month of the sample, which means we run the MSM 24 times and produce 24 estimates of ψ. In the full panel estimation, the MSM is run once and a single estimate of ψ is obtained. In either case, we initialize the fixed effects routine with guessed estimates $\hat{\gamma}_{it}^{0} = \hat{\lambda}_{jt}^{0} = 1$ for all i and j.

B.4 Step 4: Simulate visitor counts

We run $S = 10$ simulations of the number of visitors per block group-branch pair. In the panel estimation, S simulations are run across all 24 year-months of the sample, and in the month-by-month estimation, S simulations are run per year-month of the sample. The simulation procedure is identical whether month-by-month or across the panel. The only difference is the number of visitor counts drawn. Using notation for the month-by-month estimation, let N_t denote the number of block group-branch pairs with zero or positive visitor counts in year-month t of the data. Per year-month, we begin the simulation by drawing $N_t \times S$ Laplace random variables having mean zero and scale $10/9$ and $N_t \times S$ independent Uniform random variables over the unit interval. We draw these random variables only once at the beginning of each year-month's run so that the MSM does not have the underlying sample change for every guess of the model parameters. Under an estimate $\hat{\beta}$ of the elasticity and the initial guessed estimates $\left\{\hat{\gamma}_{it}^{0}, \hat{\lambda}_{jt}^{0}\right\}$ of the fixed effects, we then apply the inverse Poisson CDF to transform the Uniform random variables into Poisson random variables with distinct means given in Eq. (15). Each Poisson draw is a "true" block group-branch visitor count. Then, to replicate SafeGraph's differential privacy methods in the simulations, we (i) add a Laplace draw to all non-zero true visitor counts to form a "noisy" block group-branch visitor count, (ii) round each noisy visitor count down to the nearest integer, (iii) make 0 all noisy visitor counts below 2, and (iv) replace all noisy visitor counts that equal 2 or 3 with 4. Simulated visitor counts are 0 if either the true visitor count (from the Poisson draw) is 0 or the noisy visitor count (from the Poisson draw plus the Laplace draw) falls below 2. This way, simulated visitor counts that equal 0 arise in the same two ways as would 0 visitor counts in the observed SafeGraph data. Let $\tilde{v} = \{\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_S\}$ be the S simulated visitor counts.

B.5 Step 5: Iterate the fixed effects until convergence

Under a fixed estimate $\hat{\beta}$, the next step is to iterate the estimated fixed effects until they converge. The iterative procedure is the same in the month-by-month and full panel estimation, save for a minor change to the convergence criterion, which we describe below. We iterate the fixed effects sequentially, beginning with λ_{jt}, while holding fixed $\hat{\gamma}_{it}$ at $\hat{\gamma}_{it}^{0} = 1$. Let the total number of block groups with visitors (zero or positive) to branch j in year-month t be denoted H_{jt}. That is, branch j has zero or positive visitors from H_{jt} block groups in year-month t. Recall that the total number of branches within a 10-mile radius of block group i is B_{it}, and the total number of block group-branch pairs in year-month t with zero or positive visitor counts is N_t. Notice that, by definition, $\sum_i B_{it} = N_t$ and $\sum_j H_{jt} = N_t$.

The iteration routine is as follows. Suppose we are on the k-th iteration. The estimates of the branch fixed effects are $\hat{\lambda}_{jt}^{k}$. For each branch j, we divide the average *observed* visitor counts V_{ijt} across the H_{jt} block groups with zero or positive visitors to branch j by the average *simulated* visitor counts across the H_{jt} block groups and all simulations S. Let the average observed visitor counts for branch j be

$$\overline{V}_{jt} = \frac{1}{H_{jt}} \sum_{i} V_{ijt}. \tag{18}$$

Let the simulated visitor counts from simulation s in iteration k be denoted $\tilde{V}_{ijt}^{k}(s)$. Let the average simulated visitor counts for branch j then be

$$\overline{\tilde{V}}_{jt}^{k} = \frac{1}{S} \sum_{s} \left(\frac{1}{H_{jt}} \sum_{i} \tilde{V}_{ijt}^{k}(s) \right) \tag{19}$$

The ratio of average observed visitor counts to average simulated visitor counts is thus:

$$\xi_{jt}^{k} = \frac{\overline{V}_{jt}}{\overline{\tilde{V}}_{jt}^{k}} \tag{20}$$

We take ratios of averages rather than differences of averages because the fixed effects in the visitor count model in Eq. (15) are exponentiated. These branch-level ratios then multiplicatively update each branch's estimated fixed effect:

$$\hat{\lambda}_{jt}^{k+1} = \hat{\lambda}_{jt}^{k} \times \left(\xi_{jt}^{k} \right)^{h}, \tag{21}$$

where h is a modifying term to avoid oscillating estimates, and we set its value to 0.5. Notice that if the average simulated visitor count for branch j is higher than the average observed count in the data, $\xi_{jt}^{k} < 1$, and the branch's estimated fixed effect is revised downward.

After each update of the branch fixed effects estimates, we re-transform the $N_t \times S$ Uniform random variables into Poisson random variables, but this time inserting the updated branch fixed effects estimates $\hat{\lambda}_{jt}^{k+1}$. The estimate $\hat{\beta}$ and the block group fixed effects estimates $\hat{\gamma}_{it}^{0}$ are kept the same. We then apply the differential privacy methods to these "updated" simulated data, and then we revise the branch fixed effects estimates again using the same procedure above until the estimates converge.

While the fixed effects are updated using ratios of averages between observed and simulated visitor counts, we found that convergence occurred faster under a criterion using the differences in the averages instead. We define convergence as the squared change between iterations in the average squared difference between observed and simulated visitor counts across all branches being sufficiently small. The criterion is similar in spirit to a GMM minimization problem in which the moments are the difference in means between the observed and simulated visitor counts for each branch j, using an identity weighting matrix. Minimization is reached when the change in the GMM objective function becomes sufficiently small. In the calculation of the average squared difference, we assign more weight to branches with visitors from more block groups (higher H_{jt}). Mathematically, the convergence condition is

$$\left[\frac{1}{N_t} \sum_{j} H_{jt} \left(\overline{\tilde{V}}_{jt}^{k+1} - \overline{V}_{jt} \right)^{2} - \frac{1}{N_t} \sum_{j} H_{jt} \left(\overline{\tilde{V}}_{jt}^{k} - \overline{V}_{jt} \right)^{2} \right]^{2} < \varepsilon, \tag{22}$$

for small ε, which we set to $1e^{-5}$. In the full panel estimation, the summation in Eq. (22) is over all j branches and all t year-months, and rather than N_t in the denominator, we use the total number of block group-branch pairs across all 24 months in the sample (i.e., $N = \sum_t N_t$).

Once the branch fixed effects converge to $\hat{\lambda}_{jt}^{\infty}$, we then iterate the block group fixed effects in a similar fashion. Because the block group and branch fixed effects enter linearly in Eq. (15), one set of fixed effects needs to be normalized for identification. Otherwise, a higher estimate of λ_{jt} can be undone by a lower estimate of γ_{it}. We normalize the block group fixed effects by presuming they have a geometric mean of 1 when averaged across all visitor counts N_t.

Suppose we are on the k-th iteration of the block group fixed effects. For each block group in the year-month, we divide the average observed visitor counts V_{ijt} across the B_{it} branches within a 10-mile radius by the average simulated visitor counts across simulations. With this in mind, let the average observed visitor counts for block group i be

$$\overline{V}_{it} = \frac{1}{B_{it}} \sum_j V_{ijt}, \tag{23}$$

and let the average simulated visitor counts for block group i be

$$\overline{\tilde{V}}_{it}^{k} = \frac{1}{S} \sum_s \quad \frac{1}{B_{it}} \sum_j \tilde{V}_{ijt}^{k}(s) \Bigg) \Bigg(\tag{24}$$

The ratio of average observed visitor counts to average simulated visitor counts is thus:

$$\chi_{it}^{k} = \frac{\overline{V}_{it}}{\overline{\tilde{V}}_{it}^{k}} \tag{25}$$

Each block-group-level ratio updates the block group's estimated fixed effect using the same modifier:

$$\hat{\gamma}_{it}^{k+1} = \hat{\gamma}_{it}^{k} \times \left(\chi_{it}^{k}\right)^{h}. \tag{26}$$

With every update of the block group fixed effects, the Uniform random variables are re-transformed into Poisson random variables using the fixed estimate $\hat{\beta}$, the converged branch fixed effects estimates $\hat{\lambda}_{jt}^{\infty}$ and the updated block group fixed effects $\hat{\gamma}_{it}^{k+1}$. Differential privacy methods are applied to the updated simulated data, and the process iterates until the branch fixed effects estimates converge according to a similar criterion, putting more weight on block groups with more branches within a 10-mile radius:

$$\left[\frac{1}{N_t} \sum_i B_{it} \left(\overline{\tilde{V}}_{it}^{k+1} - \overline{V}_{it}\right)^2 - \frac{1}{N_t} \sum_i B_{it} \left(\overline{\tilde{V}}_{it}^{k} - \overline{V}_{it}\right)^2\right]^2 < \varepsilon \tag{27}$$

Again, in the full panel estimation, convergence is defined over all i and all t and over all block group-branch pairs N. After convergence, we have fixed effects estimates $\hat{\gamma}_{it}^{\infty}$ and $\hat{\lambda}_{jt}^{\infty}$ for a given estimate of β. The final piece of the estimation is to select the optimal β estimate that minimizes the distance between simulated and data moments.

B.6 Step 6: Select the moments

To identify β, we choose 8 unconditional moments of the distribution of visitor counts. In the month-by-month estimation, the moments are computed per year-month across all block groups and branches, where each block group's choice set of branches is set to B_{it}. In the panel estimation, the moments are computed across all year-months as well. Here we present the moments using notation from the month-by-month estimation. Again, the total number of zero or positive visitor counts in year-month t is N_t. We select moments that describe important parts of the distribution of these visitor counts. Denote the vector of these data moments as $m(v)$. These moments are

1. Mean of visitor counts:
$$m_1(v) \equiv \frac{1}{N_t}\sum_{N_t} V_{ijt} \tag{28}$$

2. Variance of visitor counts:
$$m_2(v) \equiv \frac{1}{N_t}\sum_{N_t} \left(V_{ijt} - m_1(v)\right)^2 \tag{29}$$

3. Percent of visitor counts equal to 0:
$$m_3(v) \equiv \frac{1}{N_t}\sum_{N_t} \mathbb{I}\left(V_{ijt} = 0\right) \tag{30}$$

4. Percent of visitor counts equal to 4:
$$m_4(v) \equiv \frac{1}{N_t}\sum_{N_t} \mathbb{I}\left(V_{ijt} = 4\right) \tag{31}$$

5. OLS coefficient from regressing log visitor counts onto their associated log distances, in cases where $V_{ijt} > 0$:
$$m_5(v) \equiv \left[\left(X'X\right)^{-1}(X)' \log V_{ijt}\right] \mathbb{I}\left(V_{ijt} > 0\right), \tag{32}$$
where the matrix $X = \left[\overrightarrow{1},\ \log d_{ij}\right]$ consists of a vector of ones and the vector of log distances.

6. OLS coefficient from regressing log visitor counts onto their associated log distances, where $V_{ijt} > 4$:
$$m_6(v) \equiv \left[\left(X'X\right)^{-1}(X)' \log V_{ijt}\right] \mathbb{I}\left(V_{ijt} > 4\right). \tag{33}$$

7. Average log distance, in cases where $V_{ijt} = 0$:
$$m_7(v) \equiv \frac{1}{N_t}\sum_{N_t} \log d_{ij}\, \mathbb{I}\left(V_{ijt} = 0\right) \tag{34}$$

8. Average log distance, in cases where $V_{ijt} = 4$:
$$m_8(v) \equiv \frac{1}{N_t}\sum_{N_t} \log d_{ij}\, \mathbb{I}\left(V_{ijt} = 4\right) \tag{35}$$

With the data moments in hand, we then calculate analogous moments from the simulated data and take average values across the S simulations. Denote $m(\tilde{v}_s|\psi)$ as the vector of simulated moments from simulation s and let $\hat{m}(\tilde{v}|\psi)$ be the estimate of the model moments from the S simulations:

$$\hat{m}(\tilde{v}|\psi) = \frac{1}{S}\sum_{S} m(\tilde{v}_s|\psi). \tag{36}$$

The final step is to find the estimate of β that minimizes the distance between the data moments and simulated model moments.

B.7 Step 7: Construct the MSM estimator

The MSM estimator $\hat{\beta}_{\text{MSM}}$ minimizes the weighted sum of squared errors between the simulated model moments and data moments. So that all errors are expressed in the same units and the minimization problem is scaled properly, we use the error function $e(\tilde{v}, v|\psi)$, which is the percent difference between the two vectors of moments:

$$e(\tilde{v}, v|\psi) \equiv \frac{\hat{m}(\tilde{v}|\psi) - m(v)}{m(v)}. \tag{37}$$

In this case, the MSM estimator is

$$\hat{\beta}_{MSM} = \underset{\beta}{\operatorname{argmin}}\; e(\tilde{v}, v|\psi)' W e(\tilde{v}, v|\psi), \tag{38}$$

where W is an 8×8 weighting matrix that controls how each moment is weighted in the minimization problem. Notice that each candidate β in Problem (38) is associated with a different set of converged fixed effects estimates $\left\{\hat{\gamma}_{it}^{\infty}, \hat{\lambda}_{jt}^{\infty}\right\}$.

We use a two-step procedure to select an optimal weighting matrix W. In the first step, we assign the identity matrix I. We then use the vector of moment errors from the MSM estimator of the first step, denoted $e\left(\tilde{v}, v|\hat{\beta}_{MSM(1)}; \hat{\gamma}_{it}^{\infty}, \hat{\lambda}_{jt}^{\infty}\right)$, to obtain a new estimate of the variance-covariance matrix of the moment errors:

$$\hat{\Omega} = e\left(\tilde{v}, v|\hat{\beta}_{MSM(1)}; \hat{\gamma}_{it}^{\infty}, \hat{\lambda}_{jt}^{\infty}\right) e\left(\tilde{v}, v|\hat{\beta}_{MSM(1)}; \hat{\gamma}^{\infty}, \hat{\lambda}_{jt}^{\infty}\right)'. \tag{39}$$

In the second step, we re-esimate the MSM estimator in Problem (38) by setting $W = \hat{\Omega}^{-1}$.

Under this weighting matrix, one can derive the variance-covariance matrix of the MSM estimator $\hat{\beta}_{MSM}$ as

$$\widehat{\text{Var}}\left(\hat{\beta}_{MSM}\right) = \left(1 + \frac{1}{S}\right)\left[\frac{\partial \hat{m}(\tilde{v}|\psi)'}{\partial \beta}\Omega^{-1}\frac{\partial \hat{m}(\tilde{v}|\psi)}{\partial \beta}\right]^{-1}, \tag{40}$$

where $\frac{\partial \hat{m}(\tilde{v}|\psi)}{\partial \beta}$ is the derivative of the vector of simulated moments, evaluated at $\hat{\beta}_{MSM}$. We calculate the derivatives numerically by taking a central difference around $\hat{\beta}_{MSM}$.

C FDIC Survey Analysis

In this section, we analyze survey evidence from the "2019 FDIC Survey of Household Use of Banking and Financial Services." The FDIC fields the survey every two years in June as 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. In Section C.1, we discuss survey findings about bank branch use; in Section C.2, we analyze differences by demographic characteristics in the primary methods that banked respondents use to access their bank accounts; and in Section C.3, we compare reported branch visitor shares according to household income from the survey to observed shares from the mobility data.

C.1 Bank Branch Use

Here, we discuss the survey findings about bank branch use among all households as well as by household characteristics. Overall, visiting branches remain a common and popular bank access method. In the survey, 80.9% of all respondents answered having visited a bank branch in the past 12 months, and just over 29.7% reported having visited a branch 10 or more times. Traveling to a branch is the primary (i.e., most common) method of accessing bank accounts 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% stated visiting a branch over the past year and about 1 in 5 in that group visited ten or more times.[6]

Household responses to the survey imply significant demographic differences in the likelihood of visiting a branch over the previous 12 months. In Online Table A.2, we report coefficients from multivariate linear probability regressions of survey responses on self-reported household characteristics. The survey reveals a positive income gradient in reported branch use. Controlling for age and race, respondents in the highest income bracket ($75,000+$) are roughly 22% more likely to reply visiting a branch in the previous year than respondents in the lowest income bracket ($< \$15,000$). A substantial Black-White gap in reported branch use is also present. Controlling for income and age, Black respondents are 10% less likely to report having visited a branch than White respondents. Probit regressions, also presented in the table, provide similar estimates of the racial and income differences in branch use based on survey evidence.

C.2 Primary Bank Access Methods by Household Characteristics

Here, we analyze how the primary method of accessing bank accounts varies by household characteristics. In the survey, the FDIC provides 6 choices for banked respondents to choose as their primary method: Bank Teller, ATM/Kiosk, Online Banking, Mobile Banking, Telephone Banking, and Other. Across all respondents, the first four choices dominate as primary access methods

[6]The 2019 Survey of Consumer Finances (SCF) also suggests that branches remain subjectively important to households in their use of financial services. The locations of branches is cited most frequently (43% of respondents) as the most important reason for choosing an institution for their main checking account, which is the most cited reason by far. Despite advances in mobile and online banking, the proportion of respondents citing branch locations as their most important reason has remained about the same since 1989 (between 43% and 49%). The 2019 SCF tables can be found here.

(xx). We therefore focus on these methods. Because ATMs and kiosks are commonly, though not exclusively, located at bank branches, we combine Bank Teller and ATM/Kiosk into one category that we treat as "visiting a bank branch." We also combine online and mobile into one category, as those are the two major alternatives to visiting a branch.

The survey responses suggest that lower-income and Black households do not make up their lesser branch use with greater use of online or mobile banking. In Online Table A.3, we report coefficients from multivariate linear probability regressions of stated primary access methods on self-reported household characteristics. Controlling for age and race, respondents in the lowest income bracket are roughly 31% less likely than those in the highest income bracket to say that mobile or online bank is their primary method to access their bank accounts. Lower income households are also about 25% more likely to say that bank tellers or ATMs are their primary access method. Similarly, controlling for income and age, Black respondents are about 6.6% less likely than White respondents to call mobile or online banking their primary access method and 6.4% more likely to say that bank tellers or ATMs are their primary method. Analogous estimates from Probit regressions in Online Table A.4 document similar differences by income and race. Overall, the survey evidence reveals that both lower income and Black households respond as depending on mobile/online banking less and bank branches more as their primary access methods.

C.3 Branch Visitor Shares by Household Income: FDIC Survey vs. SafeGraph

In this section, we compare branch visitor shares by household income observed in SafeGraph with branch visitor shares reported in the FDIC's survey. Figure A.6 presents a binned scatter plot of the share of bank branch visitors by household income from SafeGraph. 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.

Behind the binned scatter plot in Figure A.6, 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.[7]

The comparison of the FDIC's survey responses to the visitation patterns observed in SafeGraph

[7] 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.

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 A.6. 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 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.

D Robustness Checks

In this section, we conduct several checks to evaluate the robustness of key results. Section D.1 extends the 10-mile radius of branches considered available to residents in the estimation of access. Section D.3 adjusts for estimates of the number of branch employees that potentially encompass branch visitor counts. Section D.2 weights branch visitor counts in the OLS regressions of Sections 5 and **??** by Census population counts to adjust for potential over- or under-sampling from some block groups over others.

D.1 Radius of Branch Choice Set

So that the computation of our bank access measure is feasible, we reduced the choice set of branches per Census block group that are considered "available" to its residents to visit. Our core analysis shrunk the set of available branches to those located within a 10-mile radius of each block group's center of population.

Online Figure A.3 illustrates the CDF of visitor counts by the log miles between the block group and visited branch. Seventy-five percent of visitors travel to a branch within 10 miles, and ninety-five percent travel within 60 miles. The figure reveals some outliers, though, as the 99th percentile is 680 miles traveled and the maximum travel distance observed is 5,088 miles. Such extreme distances likely reflect cases in which a resident journeyed to part of the country far from home on a trip, such as leaving Maine for Alaska or Hawaii, and incidentally visited a bank branch.

Here, we extend the 10-mile radius cutoff to verify the robustness of our access measure to including more branches per block group. Based on the distribution of visitor counts in Online Figure A.3, we compute the access measure under a radius of 25 miles, which captures 90% of all visitor counts, 60 miles (95%), and 150 miles (97%).

D.2 Block Group Population Counts

As we discuss in the text, the SafeGraph panel is broadly representative of the general population at the county level based on several demographic dimensions. At a more granular level, however, SafeGraph sometimes errs in assigning mobile devices to the correct Census home block group. In the most extreme cases, SafeGraph reports some Census block groups as having more devices residing than there are people. These block groups are called "CBG sinks," and SafeGraph identifies about 1,000 of these cases among the roughly 220,000 Census block groups in the US. Seventy percent of these cases have a Census population count of zero. SafeGraph's analysis of CBG sinks is described here. In this section, we weight the visitor data by population counts according to the 2019 ACS. Block groups with disproportionately high visitors relative to their populations are down-weighted. The population weighting also attempts to make the visitor counts more representative.

D.3 Bank Employees

SafeGraph does not distinguish branch employees from branch patrons in the visitor data. All mobile devices that travel to a branch are considered "visitors." One way to approximate the number of employees is by the number of mobile visits that dwell at the location for a long time (e.g., greater than 4 hours). The ratio of a branch's number of "long dwells" to its total visit count indicates the extent to which employees' presence skews our results.

Online Figure A.7 provides a histogram of this ratio across all branches and months in our core sample, and Online Table A.5 presents statistics of the distribution. Among all branch-year-months in our core sample, 32.8% have no long dwells. These observations likely represent cases where SafeGraph only registered the mobile devices of bank customers (missing bank employees entirely). The median ratio of long-dwell visits to total branch visits is 5.95% and the mean ratio is 13.5%. A high ratio could imply that few actual customers visited the branch relative to employees or that

SafeGraph did not record many customer devices. Some branches observe a ratio of 1, which likely represents cases in which SafeGraph only registered employees (missing bank customers entirely).

E Distance Traveled to Branches

In this section, we examine how actual distances that residents journey to branches vary by demographic attributes. Analysis of the actual distance traveled is less affected by SafeGraph's differential privacy than the gravity analysis, as we rely only on the set of branches that residents visit rather than the count of visitors to each branch. The analysis is affected by SafeGraph's visitor truncation—as we miss the branches that are left out the data from having too few visitors from a block group—but the actual distance traveled is unaffected by the censoring of visitor counts.

Actual distance traveled is a person's equilibrium choice. It combines both supply factors (e.g., available bank brands in the area, the distances to nearby branches, and the characteristics of those 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.

To evaluate which surrounding branches that residents typically visit, 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 A.4 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.

Finding that residents predominately visit their nearest branches, we turn next to estimating the distances that branch visitors actually travel. Table A.11 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 **??**. 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 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.

F Bank Branch Segregation

In this section, we examine 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.

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.[8] 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, Dingel, Monras and Morales, 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 A.12 presents several segregation measures at the national level. Our three main segregation measures are (i) racial dissimilarity, (ii) racial entropy, and (iii) income entropy.

[8] 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).

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

$$\text{Dissimilarity Index} = \frac{1}{2}\sum_{i=1}^{N} \frac{\text{Black}_i}{\text{Black}_{\text{total}}} - \frac{\text{Non-Black}_i}{\text{Non-Black}_{\text{total}}} \ , \tag{41}$$

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

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

We evaluate the racial dissimilarity index in Eq. (41) 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 π_j denote the Black population share of block group j. The estimate

$$\widehat{\text{Black}}_i = \sum_j n_{j,i}\pi_j. \tag{42}$$

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

$$\hat{\Pi}_i = \sum_j \left(\frac{n_{j,i}}{N_i}\right)\pi_j. \tag{43}$$

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

$$\widehat{\text{Black}}_{\text{total}} = \sum_i N_i\hat{\Pi}_i. \tag{44}$$

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 A.12 is a simple average over the core sample period. The monthly estimates are quite stable, and they are provided in Online Table A.14.

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.

F.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 π_s denote the fraction of total bank branch visitors in the country who belong to group s. The entropy of the groups of branch visitors across the country is $E = \sum \pi_s \ln\left(\frac{1}{\pi_s}\right)$. Similarly, the entropy of the groups of visitors to bank branch i is $E_i = \sum \pi_{s,i} \ln\left(\frac{1}{\pi_{s,i}}\right)$, where $\pi_{s,i}$ is the fraction of branch i's visitors who belong to group s.[9]

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

$$\text{Entropy Index} = \sum_{i=1}^{N} \frac{\text{visitors}_i}{\text{visitors}_{\text{total}}} \left(1 - \frac{E_i}{E}\right), \tag{45}$$

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

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

The only terms in Eq. (45) that require estimation are the fractions of branch visitors belonging to a group, both for individual branches $(\pi_{s,i})$ and across the country (π_s). We estimate $\pi_{s,i}$ in an identical fashion as $\hat{\Pi}_i$ in Eq. (43) 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.[10] Each group has its own estimate, denoted $\hat{\Pi}_{s,i}$. The estimate for π_s is computed similarly as Eq. (44) 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

$$\hat{\Pi}_s = \sum_i \left(\frac{N_i}{N}\right) \hat{\Pi}_i. \tag{46}$$

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.14.) 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,

[9]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 $0 \cdot \ln\left(\frac{1}{0}\right) \neq \lim_{\pi \to 0}\left(\pi \ln\left(\frac{1}{\pi}\right)\right) = 0$. In addition, it clearly is assumed that some racial/ethnic heterogeneity exists among branch visitors in the country so that $E \neq 0$.

[10]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.

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.

F.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. (45) 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

$$\text{Income Segregation Index} = 2\ln(2) \int_0^1 E(p) H(p)\, dp. \tag{47}$$

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. (47) 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 G.

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

F.4 Geography of Bank Branch Segregation

In this section, we draw attention to spatial variation in bank branch segregation. We focus on both the racial and income entropy segregation measures, and we compute them at the county level in the same manner described in Sections F.2 and F.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.[11]

Figure A.8 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.[12]

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 Table A.13 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.[13] 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,

[11] 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%.

[12] 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.

[13] 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%.

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.11 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.12 compares segregation values by RUCA classifications. Presented are coefficients from county-level OLS regressions of the income and racial segregation estimates on county population shares that reside in each area type. Both racial and income bank segregation increases the most when transitioning into a Metropolitan core, with the change more than doubling the effects from switching into a Metropolitan suburb, a Micropolitan/Small town core, or a Rural area.

G Income Segregation Computational Steps

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

$$IS = 2\ln 2 \int_0^1 E(p) H(p) \, dp, \tag{48}$$

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

$$E(p) = p \ln\left(\frac{1}{p}\right) + (1-p)\ln\left(\frac{1}{1-p}\right) \tag{49}$$

G.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 $k \in \{1, \dots, 16\}$. 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.

G.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 F.2.) There, we used the notation π_s for the share belonging to race group s. Here, we use p_k for the share of visitors at or below the right point of a particular ACS household income range.

Using equation (49), the entropy for this percentile is

$$E(p_k) \equiv E_k = p_k \ln\left(\frac{1}{p_k}\right) + (1 - p_k) \ln\left(\frac{1}{1 - p_k}\right) \tag{50}$$

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

G.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 F.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).

G.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.9 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 (48) looks like. At least in this month, among branch visitors in the US, income segregation is seen to monotonically increase.

G.5 Step 4: Estimate the function $H(p)$ in equation (48)

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.9, 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 (50). 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.10. 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.

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

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

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

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

$$\delta_m = \frac{2}{(2+m)^2} + 2 \sum_{n=0}^{m} \frac{(-1)^{m-n} \binom{m}{n}}{(m-n+2)^2} \Bigg(\tag{51}$$

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

The 5 values for δ_m that we require are $\left(1, \frac{1}{2}, \frac{11}{36}, \frac{5}{24}, \frac{137}{900}\right)$. The measure of uncertainty about the estimated income segregation is $\text{Var}\left(\hat{IS}\right) = \Delta' V \Delta$, which we use to compute the 95% confidence intervals in Online Figure A.10.

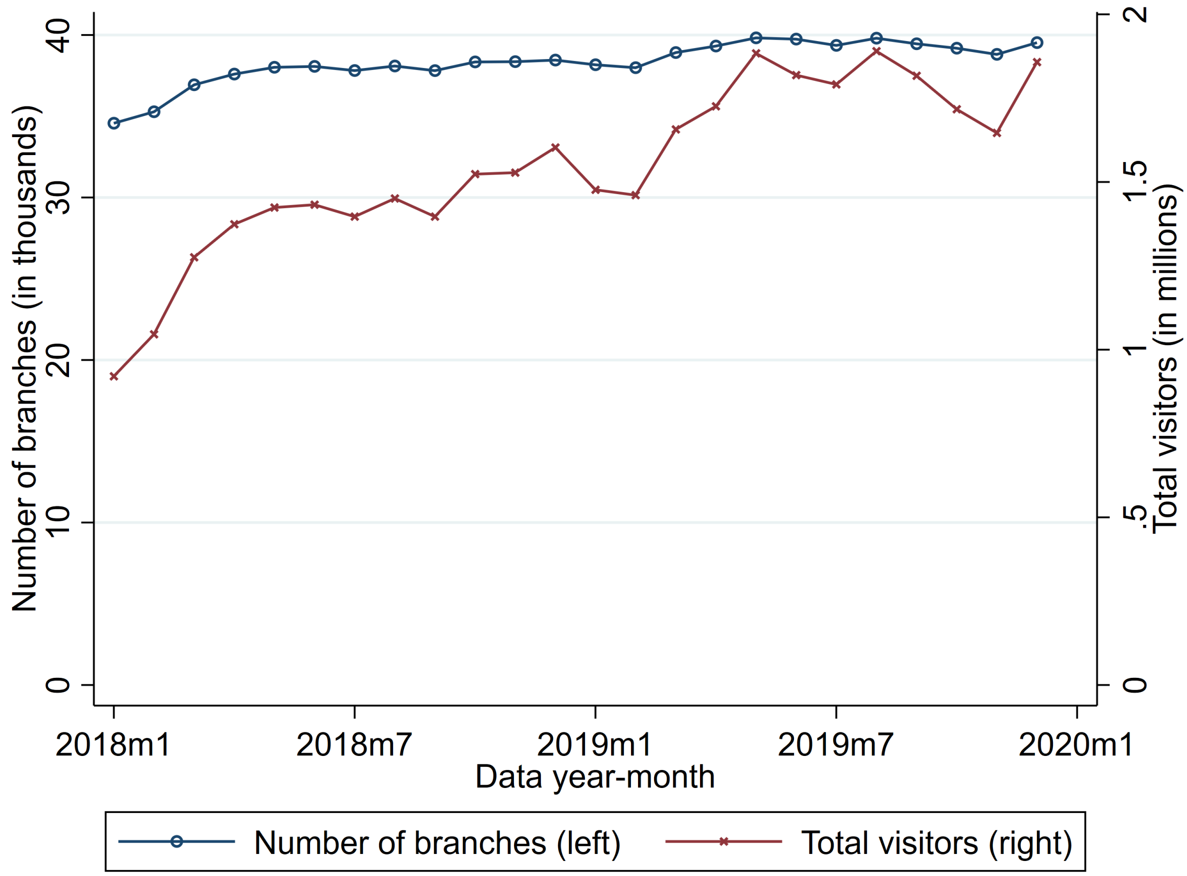
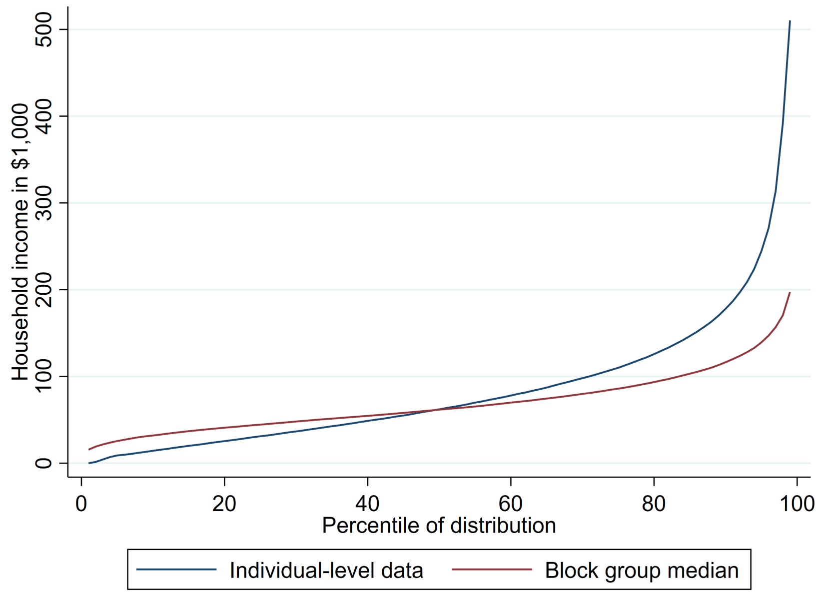


FIGURE A.1

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 FDIC's 2019 Summary of Deposits.

(A) Household Income



(B) Black Population Share

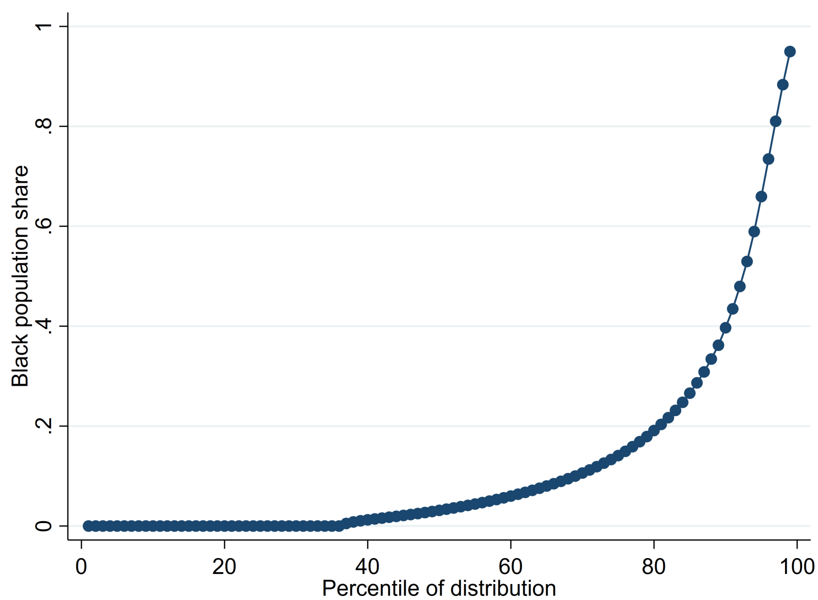


FIGURE A.2

DISTRIBUTIONS 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.

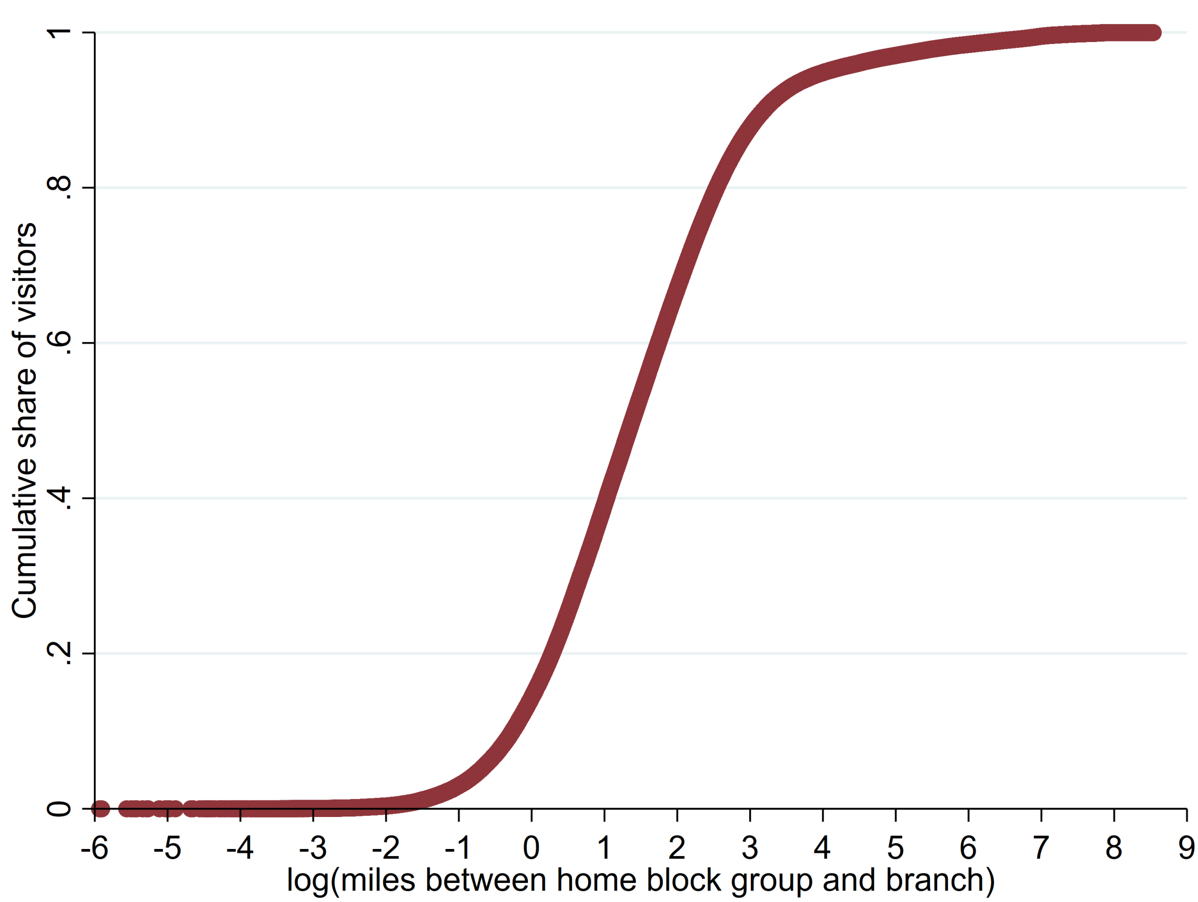


FIGURE A.3

CUMULATIVE SHARE OF BRANCH VISITORS BY LOG MILES TRAVELED FROM HOME

Notes. The figure presents the cumulative distribution of branch visitors according to their distance traveled in log miles 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 FDIC's 2019 Summary of Deposits. To construct this cumulative distribution, 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 5). 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 who travel within a particular distance and divide each sum by the total number of visitors to all branches throughout all months. We span the range of distances traveled by visitors in the data. The empirical distribution presented in the figure is the cumulative share of branch visitors per log mile traveled.

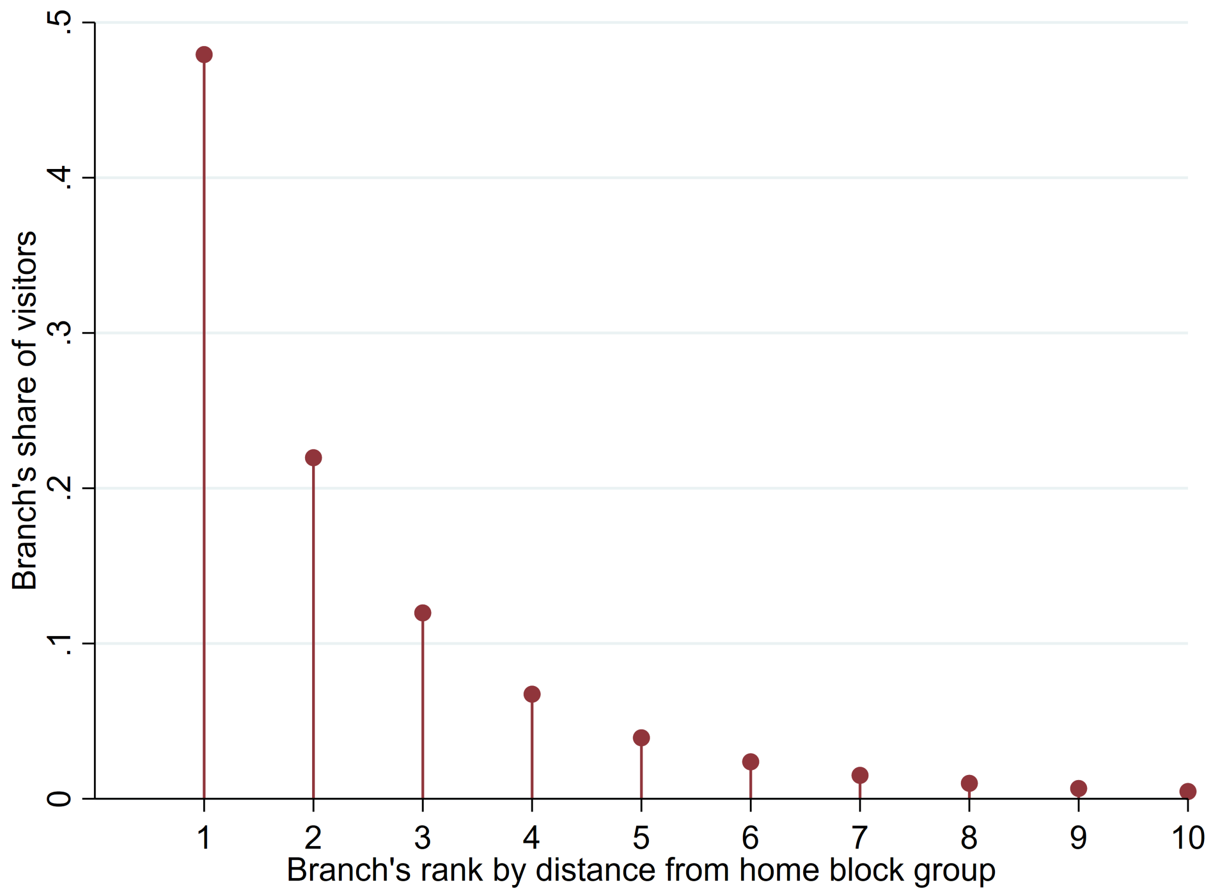
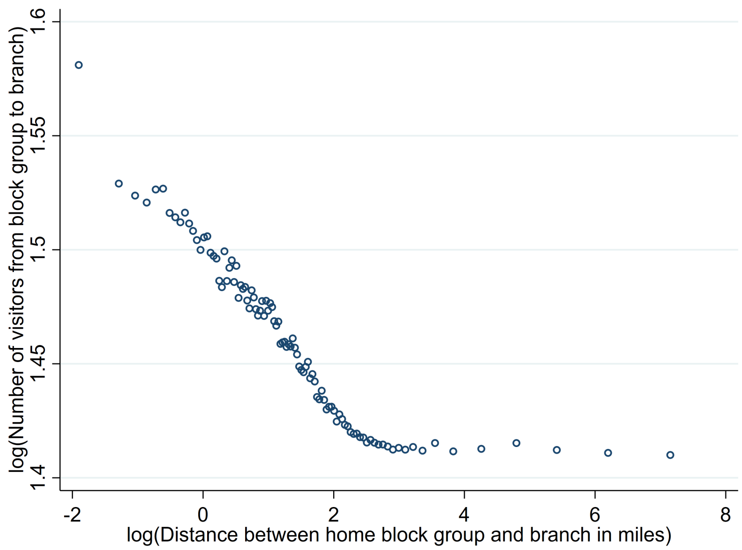


FIGURE A.4

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

Notes. The figure presents the shares of visitors who travel to bank branches according to a 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 FDIC's 2019 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 5). 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.

(A) All Block Group-Branch Pairs



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

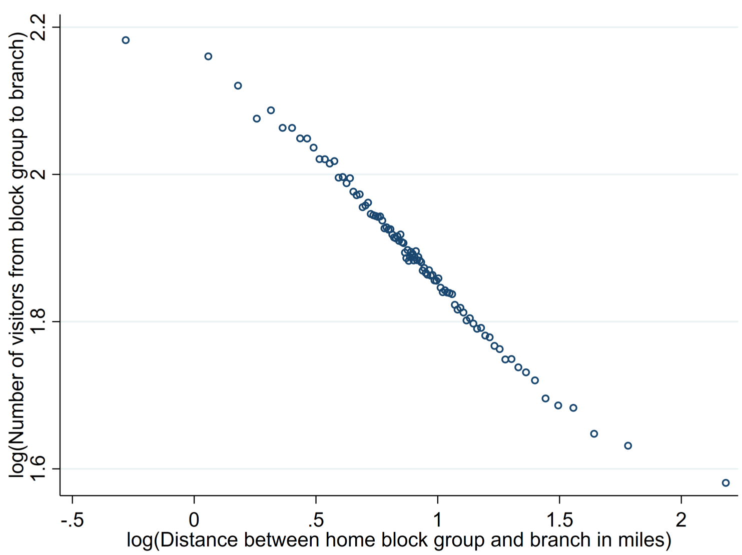


FIGURE A.5

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 5). Panel A presents the observed (raw) data and includes all block group-branch pairs, including those with visitor counts <4 that are censored 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.

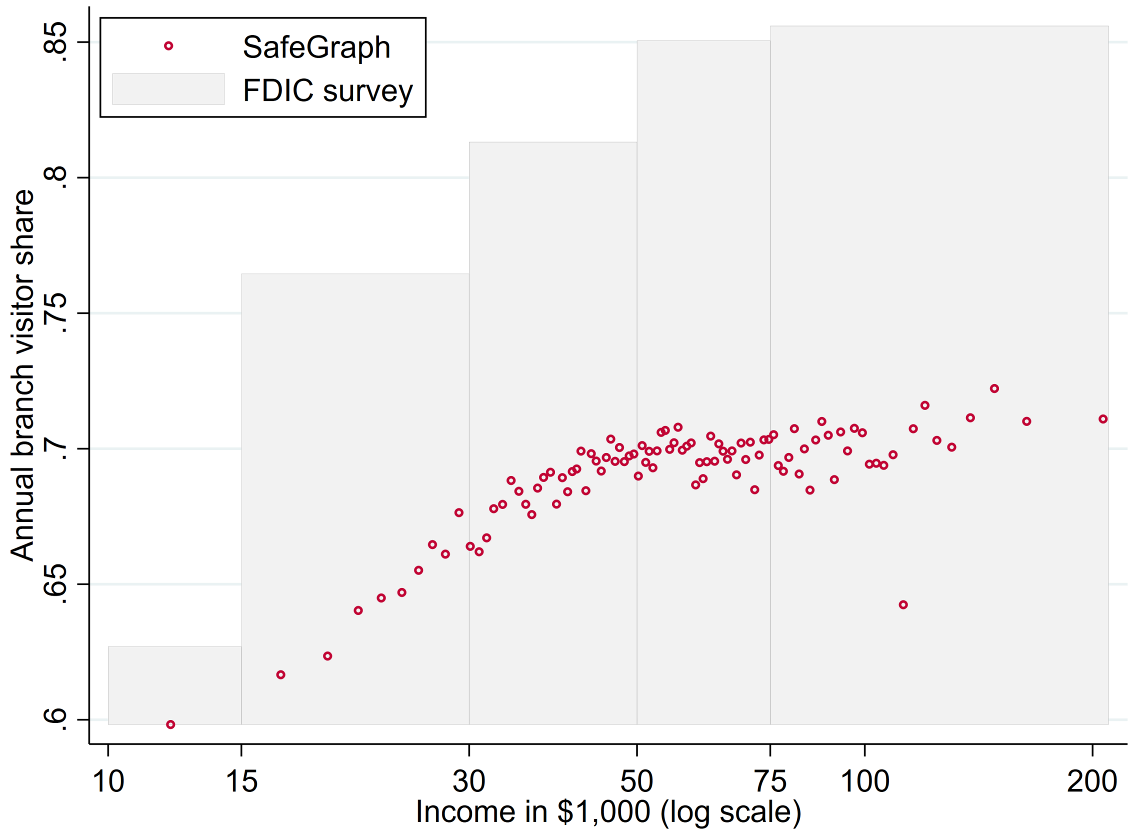


FIGURE A.6

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

Notes. The figure presents a binned scatter plot of the shares of residents who 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 FDIC's 2019 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.

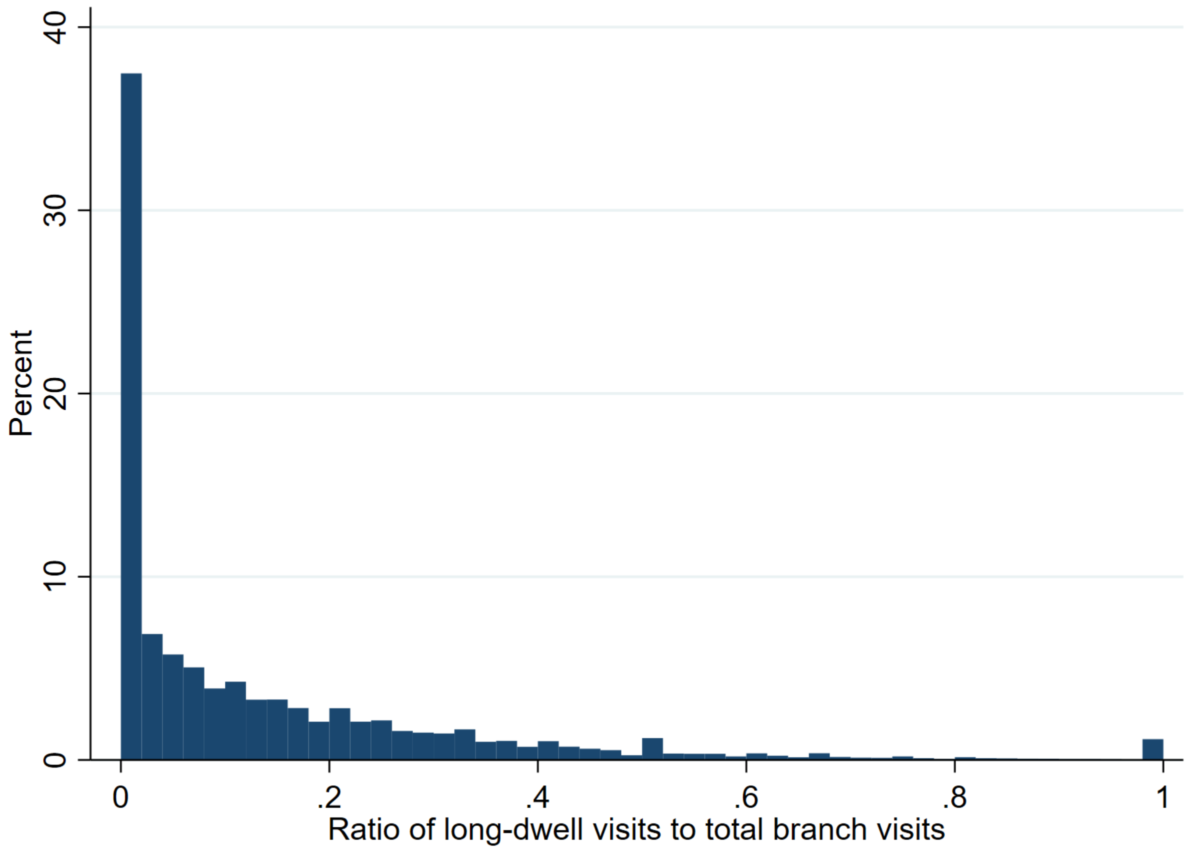
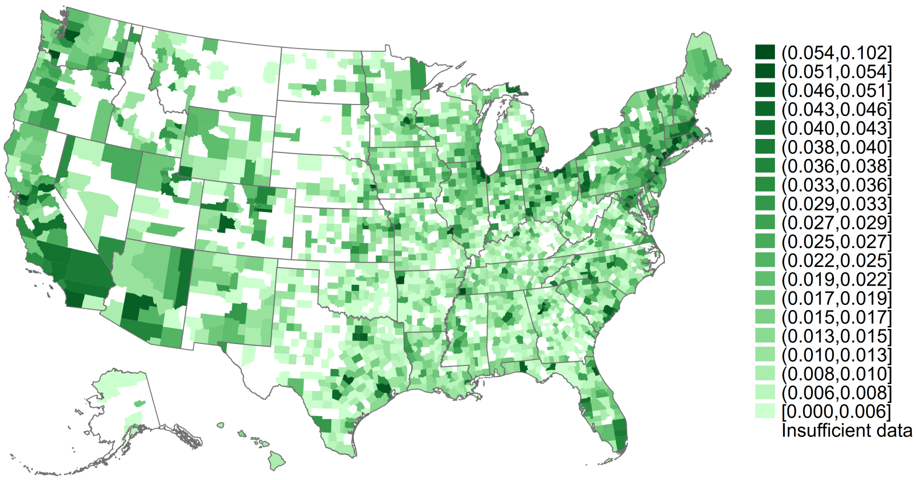


FIGURE A.7

HISTOGRAM OF RATIO OF LONG-DWELL VISITS TO TOTAL BRANCH VISITS

Notes. The figure presents a histogram of the ratio of long-dwell visits to total branch visits across all branch-year-months in our core sample. The core sample of branch locations 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 FDIC's 2019 Summary of Deposits. *Long-dwell visits* are mobile device visits with recorded SafeGraph dwell times at a branch exceeding 240 minutes (4 hours).

(A) Income Segregation by County



(B) Racial Segregation by County

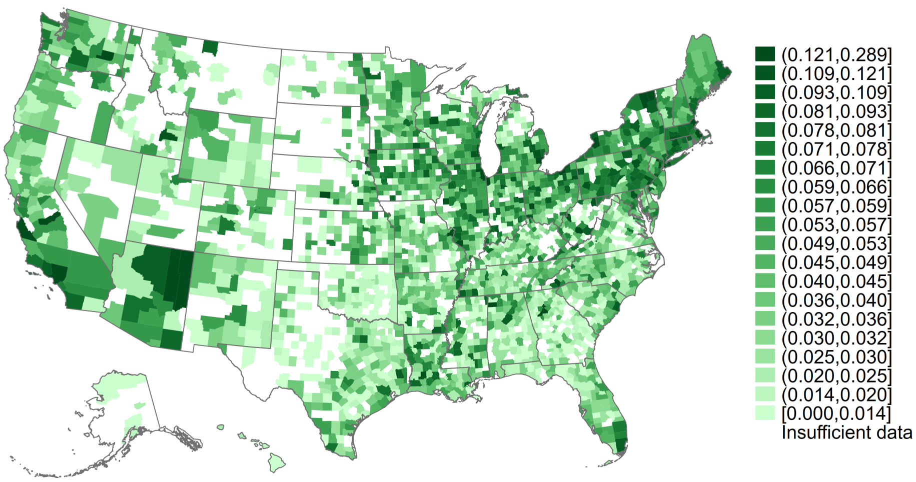


FIGURE A.8
GEOGRAPHY OF BANK BRANCH SEGREGATION

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 FDIC's 2019 Summary of Deposits. The income entropy segregation index values portrayed in Panel A are estimates of Eq. (47), made using the procedure described in Reardon (2011). The racial entropy segregation index values portrayed in Panel B are estimates of Eq. (45). 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.

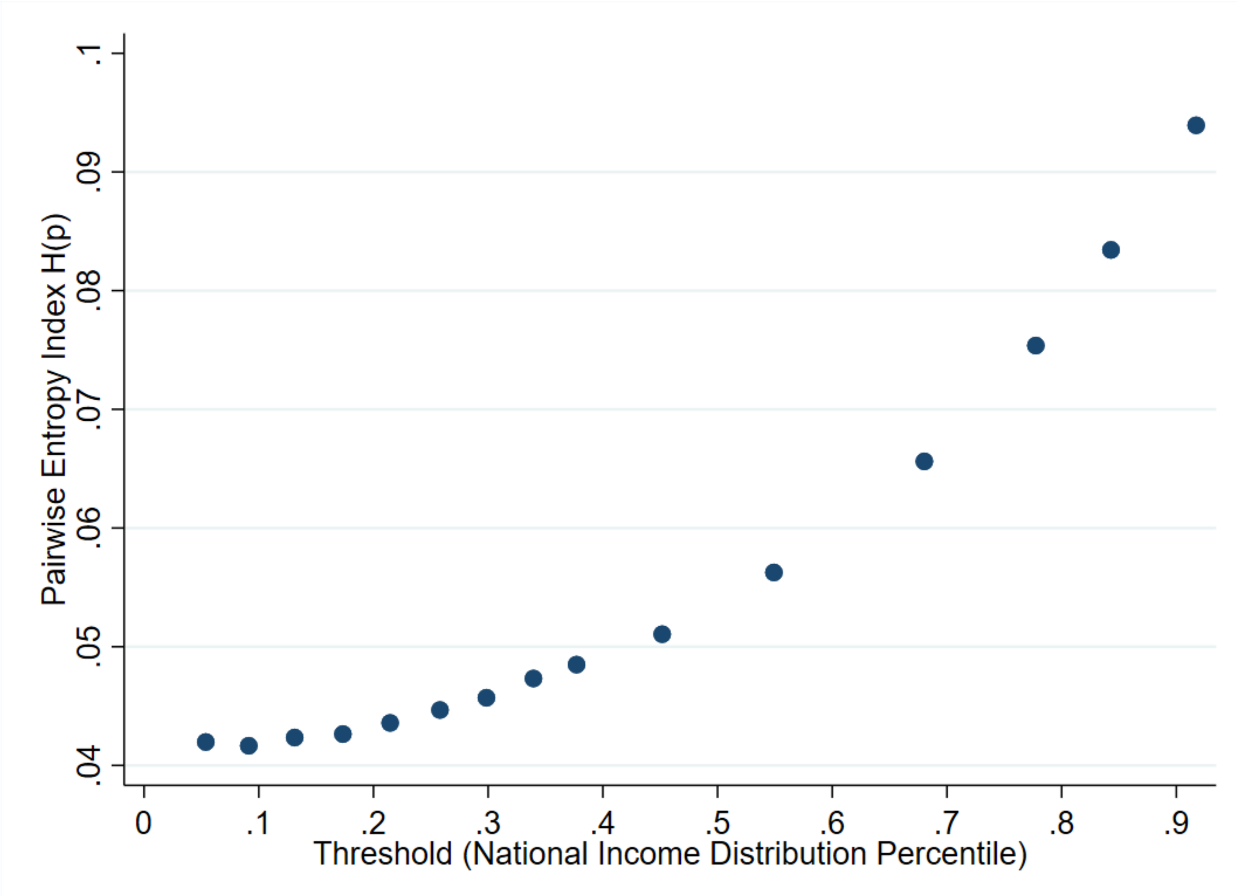


FIGURE A.9

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

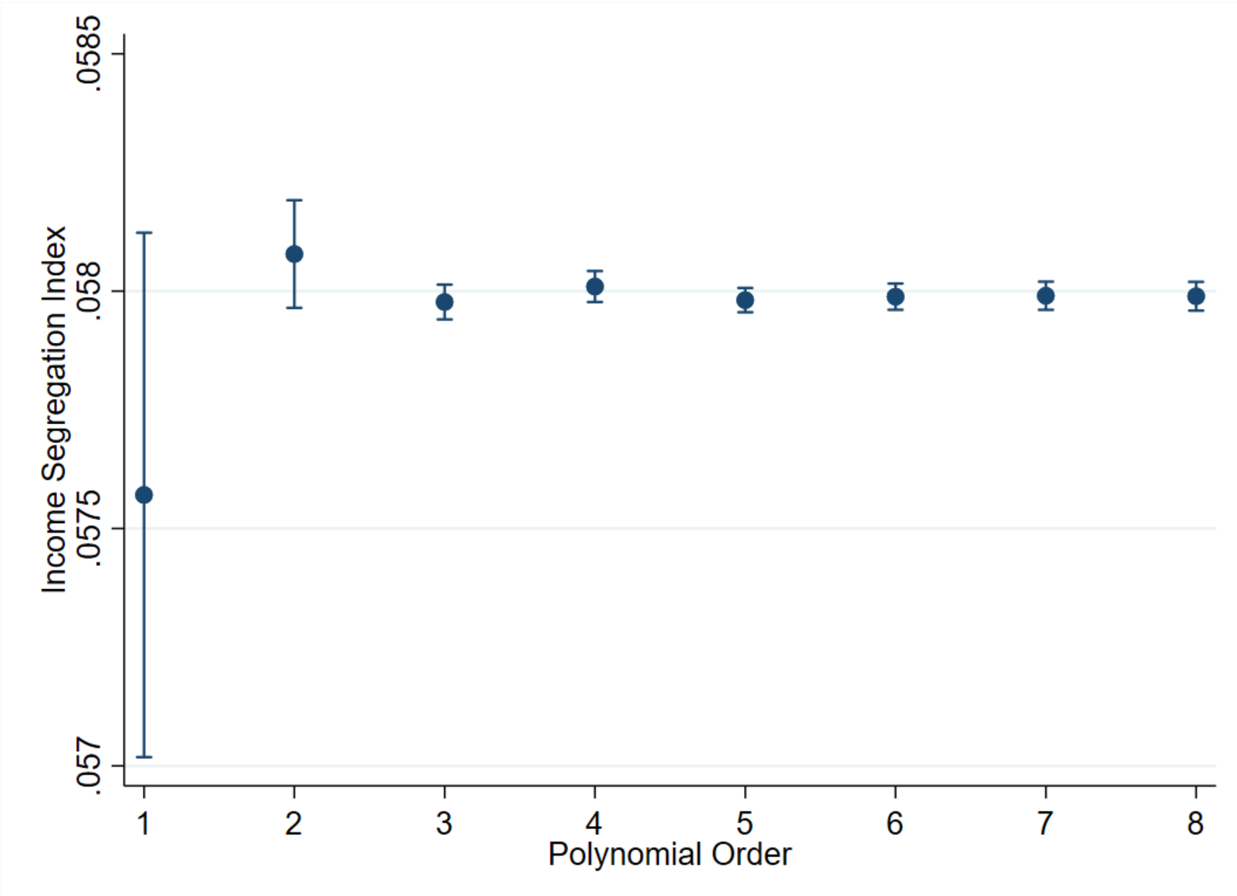
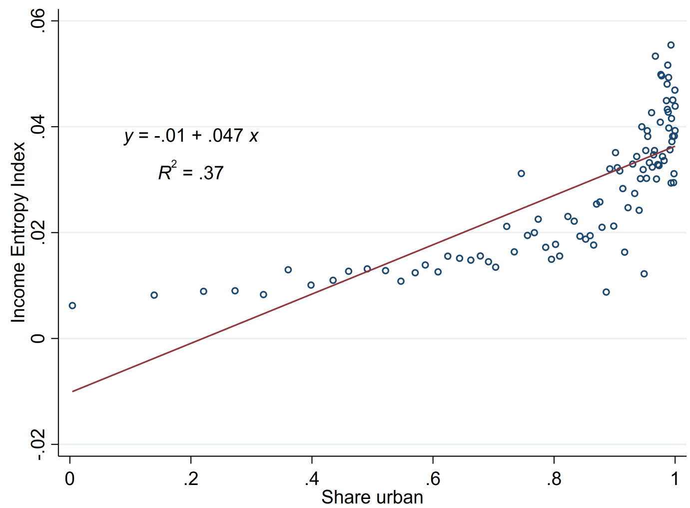


FIGURE A.10

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 FDIC's 2019 Summary of Deposits. The polynomial orders stand for the orders of the polynomials that fit the 15 values of pairwise income segregation H_k, which themselves are calculated using the steps described in Online Appendix G. 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

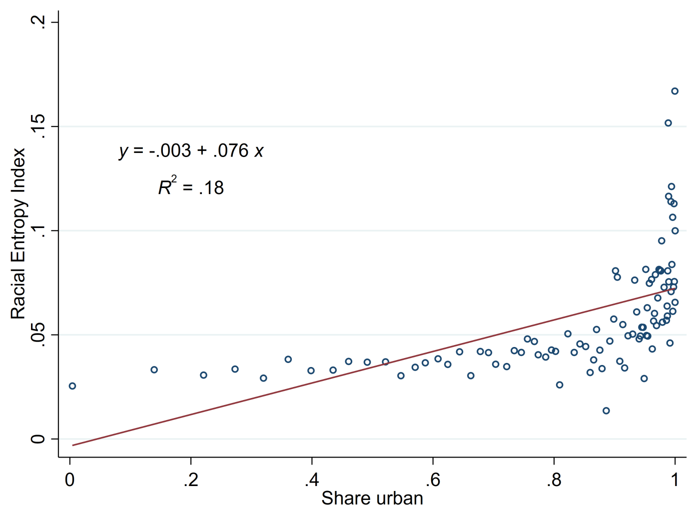


FIGURE A.11

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 FDIC's 2019 Summary of Deposits. The income entropy index values are estimates of Eq. (47). The racial entropy index values are estimates of Eq. (45). 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.

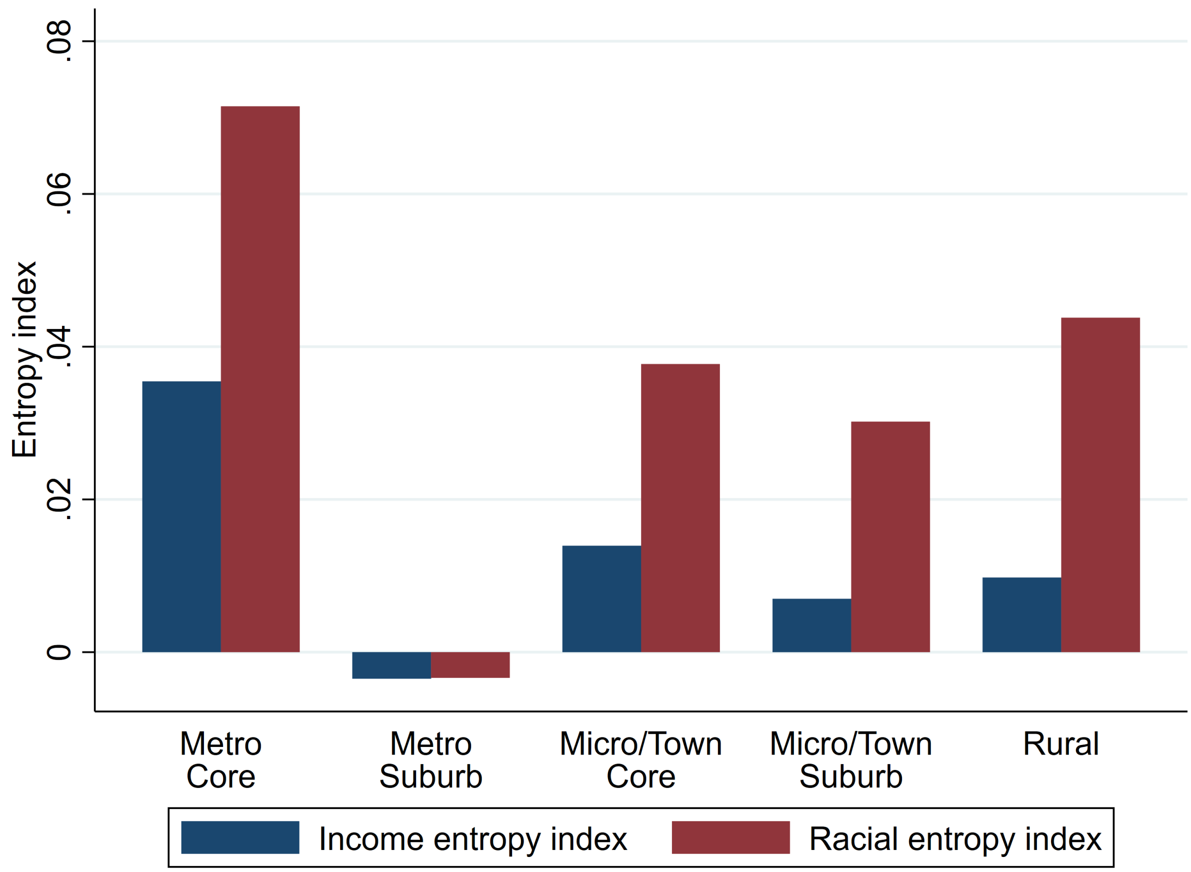


FIGURE A.12
BANK BRANCH SEGREGATION BY RUCA CLASSIFICATION

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. (47). The racial entropy index values are estimates of Eq. (45). 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
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.2
Survey Reported Branch Visit Shares by Household Characteristics

Dep. var.:	Visited a Bank Branch in the Past 12 months (Y=1, N= 0)					
Model:	OLS			Probit		
	(1)	(2)	(3)	(4)	(5)	(6)
$15,000 to $30,000		0.128	0.127		0.362	0.363
		(0.012)	(0.012)		(0.034)	(0.035)
$30,000 to $50,000		0.178	0.183		0.527	0.552
		(0.011)	(0.011)		(0.033)	(0.034)
$50,000 to $75,000		0.206	0.214		0.636	0.673
		(0.011)	(0.011)		(0.035)	(0.035)
At least $75,000		0.207	0.218		0.643	0.693
		(0.010)	(0.010)		(0.030)	(0.031)
Black	-0.144	-0.111	-0.100	-0.476	-0.370	-0.331
	(0.009)	(0.009)	(0.009)	(0.028)	(0.028)	(0.028)
Hispanic	-0.121	-0.101	-0.084	-0.409	-0.345	-0.285
	(0.009)	(0.009)	(0.009)	(0.028)	(0.028)	(0.029)
Asian	-0.072	-0.074	-0.060	-0.259	-0.274	-0.225
	(0.013)	(0.013)	(0.013)	(0.042)	(0.042)	(0.042)
Other	-0.077	-0.056	-0.048	-0.274	-0.203	-0.176
	(0.023)	(0.023)	(0.022)	(0.074)	(0.075)	(0.075)
Age 35-54			0.016			0.048
			(0.008)			(0.027)
Age 55-64			0.064			0.236
			(0.008)			(0.031)
Age 65+			0.074			0.275
			(0.008)			(0.028)
Constant	0.836	0.660	0.612	0.977	0.457	0.283
	(0.003)	(0.010)	(0.012)	(0.011)	(0.027)	(0.034)
Observations	32,904	32,904	32,904	32,904	32,904	32,904
Adjusted R^2	0.021	0.045	0.051			
Pseudo R^2				0.020	0.041	0.047

Notes. Each column reports coefficients from a weighted binary regression with robust standard errors reported in parentheses. Observations are survey responses from the "2019 FDIC Survey of Household Use of Banking and Financial Services," conducted in June 2019. Both banked and unbanked respondents are included. Observations are weighted using sample weights provided in the survey data. Dependent variable observations are binary indicators for "Yes" or "No" responses to the survey question: "Have you visited a bank branch in the past twelve months?" Demographic independent variable observations are self-reported characteristics of respondents. Income is household income. Coefficients in columns (1)-(3) are from linear probability models estimated using OLS. Coefficients in columns (4)-(6) are from Probit regressions. Omitted demographic categories are household income less than $15,000, non-Hispanic Whites, and age range 15-34.

TABLE A.3

SURVEY REPORTED BANK ACCOUNT PRIMARY ACCESS METHOD BY HOUSEHOLD CHARACTERISTICS: LINEAR PROBABILITY MODEL

Dep. var.:	Binary Indicator for Primary Method Used to Access Bank Accounts					
Access Method:	Bank Teller or ATM/Kiosk			Mobile or Online		
	(1)	(2)	(3)	(4)	(5)	(6)
\$15,000 to \$30,000		-0.032	-0.040		0.052	0.061
		(0.015)	(0.014)		(0.014)	(0.013)
\$30,000 to \$50,000		-0.130	-0.108		0.169	0.144
		(0.014)	(0.013)		(0.014)	(0.013)
\$50,000 to \$75,000		-0.186	-0.150		0.235	0.195
		(0.014)	(0.013)		(0.014)	(0.013)
At least \$75,000		-0.302	-0.252		0.364	0.308
		(0.013)	(0.012)		(0.012)	(0.012)
Black	0.068	0.018	0.064	-0.074	-0.015	-0.066
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)
Hispanic	0.066	0.025	0.096	-0.060	-0.013	-0.091
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Asian	-0.061	-0.045	0.013	0.077	0.058	-0.005
	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.013)
Other	0.057	0.025	0.063	-0.060	-0.023	-0.064
	(0.029)	(0.029)	(0.028)	(0.029)	(0.029)	(0.028)
Age 35-54			0.113			-0.121
			(0.008)			(0.009)
Age 55-64			0.244			-0.265
			(0.010)			(0.010)
Age 65+			0.361			-0.397
			(0.009)			(0.009)
Constant	0.391	0.589	0.363	0.581	0.337	0.585
	(0.004)	(0.012)	(0.013)	(0.004)	(0.012)	(0.013)
Observations	30,425	30,425	30,425	30,425	30,425	30,425
Adjusted R^2	0.005	0.053	0.121	0.005	0.070	0.152

Notes. Each column reports coefficients from a weighted binary OLS regression with robust standard errors reported in parentheses. Observations are survey responses from the "2019 FDIC Survey of Household Use of Banking and Financial Services," conducted in June 2019. Responses are from banked households. Observations are weighted using sample weights provided in the survey data. Dependent variable observations are binary indicators for the primary (i.e., most common) method used to access bank accounts among banked households that accessed their account in the past 12 months. Demographic independent variable observations are self-reported characteristics of respondents. Income is household income. The dependent variable in columns (1)-(3) equals 1 if the primary method is "Bank Teller" or "ATM/Kiosk," and 0 otherwise. The dependent variable in columns (4)-(6) equals 1 if the primary method is "Mobile Banking" or "Online Banking," and 0 otherwise. Omitted demographic categories are household income less than \$15,000, non-Hispanic Whites, and age range 15-34.

TABLE A.4
SURVEY REPORTED BANK ACCOUNT PRIMARY ACCESS METHOD BY HOUSEHOLD CHARACTERISTICS: PROBIT MODEL

Dep. var.:	Binary Indicator for Primary Method Used to Access Bank Accounts					
Access Method:	Bank Teller or ATM/Kiosk			Mobile or Online		
	(1)	(2)	(3)	(4)	(5)	(6)
\$15,000 to \$30,000		-0.082	-0.112		0.140	0.184
		(0.038)	(0.039)		(0.039)	(0.040)
\$30,000 to \$50,000		-0.330	-0.292		0.437	0.414
		(0.036)	(0.037)		(0.036)	(0.037)
\$50,000 to \$75,000		-0.470	-0.404		0.604	0.550
		(0.036)	(0.037)		(0.036)	(0.037)
At least \$75,000		-0.789	-0.699		0.950	0.872
		(0.033)	(0.034)		(0.034)	(0.035)
Black	0.173	0.047	0.182	-0.187	-0.041	-0.192
	(0.028)	(0.029)	(0.030)	(0.028)	(0.029)	(0.030)
Hispanic	0.169	0.066	0.274	-0.152	-0.033	-0.262
	(0.028)	(0.028)	(0.029)	(0.028)	(0.028)	(0.030)
Asian	-0.164	-0.128	0.033	0.202	0.164	-0.011
	(0.039)	(0.039)	(0.040)	(0.039)	(0.040)	(0.040)
Other	0.147	0.068	0.185	-0.151	-0.061	-0.193
	(0.073)	(0.075)	(0.079)	(0.072)	(0.075)	(0.080)
Age 35-54			0.338			-0.361
			(0.027)			(0.027)
Age 55-64			0.696			-0.752
			(0.029)			(0.029)
Age 65+			0.998			-1.104
			(0.027)			(0.028)
Constant	-0.276	0.226	-0.398	0.204	-0.423	0.236
	(0.010)	(0.031)	(0.038)	(0.010)	(0.032)	(0.038)
Observations	30,425	30,425	30,425	30,425	30,425	30,425
Pseudo R^2	0.004	0.040	0.094	0.004	0.052	0.117

Notes. Each column reports coefficients from a weighted binary Probit regression with robust standard errors reported in parentheses. Observations are survey responses from the "2019 FDIC Survey of Household Use of Banking and Financial Services," conducted in June 2019. Responses are from banked households. Observations are weighted using sample weights provided in the survey data. Dependent variable observations are binary indicators for the primary (i.e., most common) method used to access bank accounts among banked households that accessed their account in the past 12 months. Demographic independent variable observations are self-reported characteristics of respondents. Income is household income. The dependent variable in columns (1)-(3) equals 1 if the primary method is "Bank Teller" or "ATM/Kiosk," and 0 otherwise. The dependent variable in columns (4)-(6) equals 1 if the primary method is "Mobile Banking" or "Online Banking," and 0 otherwise. Omitted demographic categories are household income less than \$15,000, non-Hispanic Whites, and age range 15-34.

TABLE A.5
DESCRIPTIVE STATISTICS: RATIO OF LONG-DWELL VISITS TO TOTAL BRANCH VISITS

	Equals Zero	Mean	Std. Dev	P10	P25	P50	P75	P90	N
$\frac{\text{No. Long Dwells}}{\text{No. Visits}}$	0.328	0.135	0.190	0	0	0.0595	0.200	0.376	919,077

Notes. The table reports descriptive statistics of the ratio of long dwells-to-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 FDIC's 2019 Summary of Deposits. *Long Dwells* are mobile device visits with recorded SafeGraph dwell times at a branch exceeding 240 minutes (4 hours). The number of long dwells at a branch within a month divided by the branch's total number of visits in the month is an estimate of the fraction of bank employees at the branch. This ratio is computed for each branch per month over the sample period (January 2018 - December 2019). The number of observations N is the total number of branch-year-months used in generating the statistics.

TABLE A.6
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	-0.056	-0.051	-0.050	-0.283	-0.258	-0.311	-0.284
	(0.001)	(0.001)	(0.001)	(0.001)	(0.007)	(0.009)	(0.009)	(0.013)
$\log(\text{Distance}_{ijt})$		0.014		0.004		0.026		0.014
$\times$ Black		(0.002)		(0.002)		(0.029)		(0.036)
$\log(\text{Distance}_{ijt})$		0.013		-0.004		-0.434		-0.400
$\times$ Asian		(0.005)		(0.005)		(0.092)		(0.101)
$\log(\text{Distance}_{ijt})$		0.030		0.037		0.016		0.035
$\times$ Other		(0.008)		(0.010)		(0.119)		(0.167)
$\log(\text{Distance}_{ijt})$		-0.005		-0.012		-0.025		-0.019
$\times$ Hispanic		(0.002)		(0.002)		(0.020)		(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 γ_{it} is a block-group by year-month fixed effect, and λ_{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 FDIC's 2019 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 5). 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 (Metropolitian area core).

TABLE A.7
BANK ACCESS AND BLOCK GROUP FIXED EFFECTS BY DEMOGRAPHIC ATTRIBUTES - POP. WEIGHTED

Dep. var.:	log(Bank access of block groups)				Block group fixed effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	-0.129	-0.065	-0.137	-0.080	0.313	0.215	0.326	0.229
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Black	0.023	0.081	-0.043	0.019	-0.004	-0.071	0.076	-0.001
	(0.010)	(0.011)	(0.010)	(0.010)	(0.012)	(0.012)	(0.012)	(0.012)
Asian	0.808	0.781	0.755	0.727	-0.490	-0.438	-0.445	-0.396
	(0.020)	(0.020)	(0.019)	(0.019)	(0.021)	(0.021)	(0.021)	(0.020)
Other	0.386	0.408	0.391	0.411	-0.345	-0.325	-0.374	-0.343
	(0.041)	(0.042)	(0.044)	(0.045)	(0.042)	(0.042)	(0.047)	(0.047)
Hispanic	0.163	0.232	0.123	0.197	-0.100	-0.180	-0.047	-0.140
	(0.012)	(0.013)	(0.011)	(0.013)	(0.012)	(0.013)	(0.012)	(0.013)
Age <15		-0.878		-0.959		1.431		1.547
		(0.026)		(0.026)		(0.029)		(0.031)
Age 35-54		-0.513		-0.363		1.006		0.914
		(0.037)		(0.040)		(0.031)		(0.033)
Age 55-64		-0.991		-0.812		1.183		1.019
		(0.036)		(0.036)		(0.034)		(0.036)
Age 65+		-0.132		-0.216		0.561		0.666
		(0.025)		(0.025)		(0.025)		(0.026)
log(No. of Devices)	-0.072	-0.075	-0.100	-0.102	0.630	0.635	0.669	0.673
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.005)
Observations	2,669,220	2,669,220	2,033,884	2,033,884	2,669,220	2,669,220	2,033,884	2,033,884
Adjusted R^2	0.483	0.488	0.518	0.527	0.395	0.404	0.429	0.442
Sample	Core	Core	MC	MC	Core	Core	MC	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. Block groups where no resident was recorded in SafeGraph as having visited a branch in the year-month and block groups having no bank branch within a 10-mile radius in the year-month are dropped. Observations are weighted by block-group population counts from the 2019 5-year American Community Survey (ACS). All columns 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 FDIC's 2019 Summary of Deposits (SOD). Demographic independent variable observations are population-based decimal shares from the 2019 5-year ACS. Income is median household income. In columns (1)-(4), the dependent variable is the log estimated bank access measure $\log \hat{\Phi}_{it}$ from Eq. (16), whereas in columns (5)-(8), the dependent variable is the estimated block group fixed effects $\hat{\gamma}_{it}$ from the gravity relation in Eq. (15). Both dependent variables are computed from the month-by-month Method of Simulated Moments estimation described in Online Appendix B. 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 (Metropolitian area core). The omitted demographic groups are non-Hispanic Whites, age range 15-34, and education less than High School degree.

TABLE A.8
BANK BRANCH USE BY DEMOGRAPHIC ATTRIBUTES - POP. WEIGHTED

Dep. var.:	log(Expected no. of visitors)					log(Observed no. of visitors)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Income)	0.126	0.185	0.150	0.189	0.149	0.161	0.167
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.003)	(0.004)
Black	0.128	0.019	0.010	0.033	0.018	-0.025	-0.002
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Asian	-0.022	0.317	0.343	0.309	0.332	0.236	0.235
	(0.022)	(0.019)	(0.019)	(0.019)	(0.019)	(0.016)	(0.016)
Other	-0.685	0.041	0.083	0.018	0.068	0.061	0.031
	(0.034)	(0.029)	(0.029)	(0.035)	(0.036)	(0.023)	(0.031)
Hispanic	0.077	0.063	0.052	0.076	0.057	-0.016	-0.006
	(0.009)	(0.010)	(0.011)	(0.010)	(0.011)	(0.009)	(0.010)
Age <15			0.552		0.589	0.631	0.678
			(0.023)		(0.026)	(0.020)	(0.023)
Age 35-54			0.494		0.551	0.551	0.588
			(0.029)		(0.034)	(0.021)	(0.025)
Age 55-64			0.192		0.207	0.258	0.256
			(0.027)		(0.031)	(0.022)	(0.026)
Age 65+			0.429		0.450	0.385	0.417
			(0.020)		(0.022)	(0.017)	(0.019)
log(No. of Devices)	0.630	0.558	0.561	0.569	0.572	0.660	0.671
	(0.007)	(0.005)	(0.005)	(0.006)	(0.006)	(0.004)	(0.005)
Constant	-2.151						
	(0.050)						
Observations	2,669,246	2,669,220	2,669,220	2,033,884	2,033,884	3,134,720	2,246,239
Adjusted R^2	0.375	0.481	0.484	0.487	0.490	0.563	0.574
Sample	Core	Core	Core	MC	MC	Core	MC
Year-month FE		O	O	O	O	O	O
County FE		O	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 year-month in the sample period from January 2018 - December 2019. Block groups where no resident was recorded in SafeGraph as having visited a branch in the year-month and block groups having no bank branch within a 10-mile radius in the year-month are dropped. Observations are weighted by block-group population counts from the 2019 5-year American Community Survey (ACS). 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 FDIC's 2019 Summary of Deposits. Demographic independent variable observations are population-based decimal shares from the 2019 5-year ACS. Income is median household income. The dependent variable in columns (1)-(6) is the natural logarithm of the expected number of branch goers from each block group based on the month-by-month Method of Simulated Moments estimates; i.e., $\log \hat{V}^*_{it} \equiv \log \sum_j \hat{V}^*_{ijt}$, where $\hat{V}^*_{ijt}$ is the predicted mean of V^*_{ijt} in Eq. (15) and β is time-varying. The dependent variable in columns (7) and (8) is the natural logarithm of the observed number of branch goers from each Census block group; i.e., $\log V_{it} \equiv \log \sum_j V_{ijt}$, where V_{ijt} is given in Eq. (14). 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 A.9

BANK ACCESS AND BLOCK GROUP FIXED EFFECTS BY DEMOGRAPHIC ATTRIBUTES UNDER POSTAL BANKING - POP. WEIGHTED

Dep. var.:	log(Bank Access of block groups)							
USPS branch quality:	Median				Low		High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	-0.151	-0.089	-0.154	-0.097	-0.071	-0.084	-0.123	-0.124
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Black	0.043	0.096	-0.007	0.052	0.083	0.027	0.123	0.103
	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)	(0.009)
Asian	0.734	0.701	0.684	0.654	0.761	0.709	0.603	0.561
	(0.017)	(0.017)	(0.016)	(0.016)	(0.019)	(0.018)	(0.016)	(0.016)
Other	0.325	0.333	0.351	0.363	0.390	0.402	0.224	0.284
	(0.035)	(0.036)	(0.039)	(0.039)	(0.040)	(0.043)	(0.032)	(0.036)
Hispanic	0.203	0.263	0.169	0.239	0.242	0.208	0.305	0.296
	(0.010)	(0.011)	(0.010)	(0.011)	(0.013)	(0.012)	(0.010)	(0.011)
Age <15		-0.885		-0.976	-0.876	-0.962	-0.907	-1.007
		(0.023)		(0.024)	(0.025)	(0.025)	(0.022)	(0.023)
Age 35-54		-0.502		-0.374	-0.508	-0.366	-0.495	-0.397
		(0.028)		(0.030)	(0.034)	(0.037)	(0.023)	(0.024)
Age 55-64		-0.967		-0.805	-0.983	-0.809	-0.935	-0.802
		(0.029)		(0.031)	(0.033)	(0.034)	(0.026)	(0.028)
Age 65+		-0.185		-0.250	-0.144	-0.223	-0.269	-0.309
		(0.021)		(0.021)	(0.023)	(0.023)	(0.018)	(0.019)
log(No. of Devices)	-0.085	-0.088	-0.106	-0.108	-0.079	-0.103	-0.110	-0.124
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Observations	2,669,220	2,669,220	2,033,884	2,033,884	2,669,220	2,033,884	2,669,220	2,033,884
Adjusted R^2	0.602	0.608	0.544	0.556	0.572	0.541	0.689	0.659
Sample	Core	Core	MC	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. Block groups where no resident was recorded in SafeGraph as having visited a private bank branch in the year-month and block groups having no private bank branch within a 10-mile radius in the year-month are dropped. Observations are weighted by block-group population counts from the 2019 5-year American Community Survey (ACS). All columns use our core sample of private 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 FDIC's 2019 Summary of Deposits plus businesses in SafeGraph with NAICS codes equal to 491110 (Postal Services) for which we have visitor data. Demographic independent variable observations are population-based decimal shares from the 2019 5-year ACS. Income is median household income. In all columns, the dependent variable is the log estimated bank access measure $\log \hat{\Phi}_{it}$ from Eq. (16) that includes both private bank branches and Post Office branches. The dependent variable is computed from the month-by-month Method of Simulated Moments estimation described in Online Appendix B. In columns (1)-(4), we assign to each Post Office location per year-month an establishment fixed effect λ_{jt} equal to the 50th percentile of the distribution of private bank fixed effects in the year-month. In columns (5) and (6), we assign the 10th percentile and in columns (7) and (8), we assign the 90th percentile. Columns (1), (2), (5), and (7) include all block groups for which we have visitor data, whereas columns (3), (4), (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 A.10

BANK BRANCH USE BY DEMOGRAPHIC ATTRIBUTES - POP. WEIGHED

Dep. var.:	log(Expected no. of visitors)							
USPS branch quality:	Median				Low		High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	0.162	0.126	0.172	0.132	0.144	0.145	0.092	0.105
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Black	0.039	0.024	0.068	0.051	0.012	0.025	0.051	0.101
	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Asian	0.243	0.263	0.239	0.258	0.323	0.313	0.164	0.165
	(0.018)	(0.018)	(0.018)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Other	-0.020	0.008	-0.023	0.021	0.065	0.060	-0.101	-0.058
	(0.029)	(0.029)	(0.034)	(0.034)	(0.029)	(0.035)	(0.032)	(0.036)
Hispanic	0.103	0.083	0.121	0.099	0.061	0.068	0.125	0.156
	(0.009)	(0.010)	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)	(0.011)
Age <15		0.546		0.571	0.555	0.585	0.524	0.541
		(0.023)		(0.025)	(0.023)	(0.026)	(0.023)	(0.025)
Age 35-54		0.504		0.540	0.499	0.548	0.512	0.517
		(0.024)		(0.027)	(0.027)	(0.031)	(0.024)	(0.026)
Age 55-64		0.216		0.214	0.200	0.210	0.248	0.218
		(0.025)		(0.028)	(0.026)	(0.030)	(0.027)	(0.029)
Age 65+		0.376		0.415	0.417	0.442	0.292	0.357
		(0.019)		(0.020)	(0.020)	(0.022)	(0.019)	(0.021)
log(No. of Devices)	0.546	0.547	0.563	0.565	0.556	0.570	0.525	0.549
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
Observations	2,669,220	2,669,220	2,033,884	2,033,884	2,669,220	2,033,884	2,669,220	2,033,884
Adjusted R^2	0.357	0.359	0.433	0.437	0.412	0.470	0.355	0.447
Sample	Core	Core	MC	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 year-month in the sample period from January 2018 - December 2019. Block groups where no resident was recorded in SafeGraph as having visited a private bank branch in the year-month and block groups having no private bank branch within a 10-mile radius in the year-month are dropped. Observations are weighted by block-group population counts from the 2019 5-year American Community Survey (ACS). Dependent variable observations are based on our core sample of private 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 FDIC's 2019 Summary of Deposits plus businesses in SafeGraph with NAICS codes equal to 491110 (Postal Services) for which we have visitor data. Demographic independent variable observations are population-based decimal shares from the 2019 5-year ACS. Income is median household income. The dependent variable is the natural logarithm of the expected number of branch goers from each block group based on the month-by-month Method of Simulated Moments estimates; i.e., $\log \hat{V}^*_{it} \equiv \log \sum_j \hat{V}^*_{ijt}$, where $\hat{V}^*_{ijt}$ is the predicted mean of V^*_{ijt} in Eq. (15) and β is time-varying. In columns (1)-(4), we assign to each Post office location per year-month an establishment fixed effect λ_{jt} equal to the 50th percentile of the distribution of private bank fixed effects in the year-month. In columns (5) and (6), we assign the 10th percentile and in columns (7) and (8), we assign the 90th percentile. Columns (1), (2), (5), and (7) include all block groups for which we have branch visitor data, whereas columns (3), (4), (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 A.11

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 FDIC's 2019 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 5). 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 (Metropolitian area core). The omitted demographic groups are non-Hispanic Whites, age range 15-34, and education less than High School degree.

Table A.12
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 FDIC's 2019 Summary of Deposits. The dissimilarity index in this paper is an estimate of Eq. 41, as described in Section F.1. The two groups in the dissimilarity index computation are Black and non-Black. The racial entropy index is an estimate of Eq. (45), as described in Section F.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. (47), as described in Section F.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 A.13

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.13 (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 FDIC's 2019 Summary of Deposits). Branches are assigned to counties based on their locations in SafeGraph. Segregation estimates are calculated according to the methods described in Section F. 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.14
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 FDIC's 2019 Summary of Deposits. The dissimilarity index in this paper is an estimate of Eq. 41, as described in Section F.1. The two racial groups in the dissimilarity index computation are Black and non-Black. The racial entropy index is an estimate of Eq. (45), as described in Section F.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. (47), as described in Section F.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.