

Sakong, Jung; Zentefis, Alexander K.

**Working Paper**

## Bank branch access: Evidence from geolocation data

Working Paper, No. WP 2023-15

**Provided in Cooperation with:**  
Federal Reserve Bank of Chicago

*Suggested Citation:* Sakong, Jung; Zentefis, Alexander K. (2023) : Bank branch access: Evidence from geolocation data, Working Paper, No. WP 2023-15, Federal Reserve Bank of Chicago, Chicago, IL, <https://doi.org/10.21033/wp-2023-15>

This Version is available at:  
<https://hdl.handle.net/10419/284056>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Bank Branch Access: Evidence from Geolocation Data

Jung Sakong and Alexander K. Zentefis

---

April 11, 2023

WP 2023-15

<https://doi.org/10.21033/wp-2023-15>



FEDERAL RESERVE BANK *of* CHICAGO

---

\*Working papers are not edited, and all opinions are the responsibility of the author(s). The views expressed do not necessarily reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System.

---

# Bank Branch Access: Evidence from Geolocation Data\*

Jung Sakong<sup>†</sup>

Alexander K. Zentefis<sup>‡</sup>

April 11, 2023

## Abstract

Low-income and Black households are less likely to visit bank branches than high-income and White households, despite the former two groups appearing to rely more on branches as means of bank participation. We assess whether unequal branch access can explain that disparity. We propose a measure of bank branch access based on a gravity model of consumer trips to bank branches, estimated using mobile device geolocation data. Residents have better branch access if branches are closer or have superior qualities that attract more visitors. Because the geolocation data is distorted to protect user privacy, we estimate the gravity model with a new econometric method that adapts the Method of Simulated Moments to handle high-dimensional fixed effects. We find no evidence that low-income communities lack access to bank branches and instead find that lower demand for bank branch products or services explains their lower branch use. But in Black communities, worse access explains their entire drop-off in branch use. For residents of these areas, weaker access is not from having lower quality branches, but from branches being located farther away from them. The results highlight parts of the country that would benefit the most from policies that expand access to banking.

**JEL classification:** D14, G21, J15, R20

**Keywords:** inequality, location economics, spatial analysis, banking

---

\*This paper previously circulated with the title “Bank Access Across America.” We give special thanks to Bo Honoré and Luojia Hu for suggesting parts of the econometric method we use in this paper and for their very helpful feedback. We are grateful to Tori Healey, Gen Li, Lizzie Tong, and Yi (Layla) Wang for their extraordinary research assistance. We thank Taha Ahsin, Treb Allen, Costas Arkolakis, Helen Banga, Asaf Bernstein, Mehdi Beyhaghi, Lorenzo Caliendo, Claire Célérier, Judy Chevalier, Tony Cookson, Jess Cornaggia, Doug Diamond, Jonathan Dingel, Jane Dokko, João Granja, Cecilia Fielner, Raffi Garcia, Paul Goldsmith-Pinkham, Yaming Gong, Gary Gorton, Jessie Handbury, Xuan Hung Do, Stefan Jacewitz, Kristoph Kleiner, Noura Kone, Sam Kortum, Cameron LaPoint, Simone Lenzu, Xiang Li, Runjing Lu, Yuhei Miyauchi, Luciana Orozco, Piyush Panigrahi, Karen Pence, Matthieu Picault, Roberto Robatto, Claudia Robles-Garcia, Rosa Sanchis-Guarner, Shri Santosh, Katja Siem, Fiona Scott Morton, Brad Shapiro, Kelly Shue, Mike Sinkinson, Amir Sufi, Nancy Wallace, Brian Waters; and participants at various seminars and conferences for their very helpful comments. We also thank Jill Kelly, Miriam Olivares, Yichen Yang, the Marx Science and Social Science Library at Yale, Patricia Carbajales, Pat Claflin, Mazair Fooladi Mahani, and the Clemson Center for Geospatial Technologies for their valuable assistance in geoprocessing. 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.

<sup>†</sup>Federal Reserve Bank of Chicago; 230 La Salle St, Chicago, IL 60604 (email: jung.sakong@chi.frb.org)

<sup>‡</sup>Yale School of Management; 165 Whitney Ave, New Haven, CT 06510 (email: alexander.zentefis@yale.edu), Corresponding author

# 1 Introduction

Bank branches remain vital means of bank participation in the United States, especially for low-income and Black households. But both types of households are significantly less likely to visit bank branches than high-income and White households, and they do not appear to offset their lower use of branches with greater reliance on mobile and online banking. Among policymakers and in the media, a common argument to explain these racial and income disparities in branch use is unequal branch access (Friedline and Despard 2016; Dahl and Franke 2017), which may impede some households from obtaining the full benefits to welfare of bank participation (Davidson 2018). This reasoning has motivated policy proposals to increase access to branches, such as public investments in community development banks (Ellwood and Patel 2018) and the expansion of U.S. postal banking (Baradaran 2013). But research conflicts over whether branch access differs enough by race and income to explain the disparities (Morgan, Pinkovskiy, and Yang 2016; Goodstein and Rhine 2017; Small, Akhavan, Torres, and Wang 2021). And demand-related factors, such as low cash savings or distrust in banks, provide equally reasonable explanations.<sup>1</sup>

In this paper, we attempt to make progress in this area by using newly available geolocation data from mobile devices to quantify the extent to which differences in access or in demand separately explain continued disparities in U.S. bank branch use. Quantifying these two distinct channels is the paper’s first main contribution. We find no evidence that low-income communities lack access to bank branches and instead find that lower demand for bank branch products or services explains their lower branch use. But in Black communities, worse access explains their entire drop-off in branch use. The results help inform which areas of the country to target policies that expand access to banking.

In the first part of the paper, we estimate a granular measure of bank branch access, relying on tools from spatial economics and trade. We represent consumer trips from home Census block groups to bank branches in the geolocation data with a standard gravity equation, which consists of block group  $\times$  time fixed effects, bank branch  $\times$  time fixed effects, and the distances between pairs of block groups and branches. This approach takes advantage of the network of trips that links block group residents to multiple branches. By using block group fixed effects, we compare how residents of the *same* block group visit different branches by varying amounts. The extent to which this *within* block group comparison fully absorbs the block group residents’ demand for bank branch products or services, the estimated differences between block groups in overall branch use can plausibly be

---

<sup>1</sup>Disparities in bank participation and branch visits are documented in the 2019 FDIC Survey of Household Use of Banking and Financial Services and described in more detail in Section 2 of this paper. Regarding conflicting evidence on bank branch access, Morgan et al. (2016) find that residents of majority-minority tracts are less likely to live in banking deserts (areas with no branches within 10 miles) than their counterparts in non-minority tracts. But Small et al. (2021) find that the nearest banks are considerably farther than the nearest check cashers in minority neighborhoods. Meanwhile, Goodstein and Rhine (2017) argue that the influence of branch locations on bank use is fairly modest overall. For evidence of the welfare benefits of bank participation (e.g., improved access to credit, higher subjective well being, greater liquid savings, larger wealth accumulation), see, for example, Eisfeldt (2007); Fitzpatrick (2015); Prina (2015); Agarwal, Alok, Ghosh, Ghosh, Piskorski, and Seru (2017); Melzer (2018); Célerier and Matray (2019); and Brown, Cookson, and Heimer (2019).

attributed to differential access.<sup>2</sup>

The gravity equation not only allows us to control for demand-related factors, but it also generates a local measure of bank branch access. From the gravity model, a block group's expected total number of branch goers can be split into one part due to "demand" (the block group fixed effect) multiplied by a second part due to "access." We use "demand" informally to capture all resident characteristics that contribute to visiting *any* bank branch (e.g., average wealth, income, financial sophistication, trust in banks, flexibility in time). The "access" part is an index of bank branches available to residents of the block group. Each branch in a block group's index is represented by the branch's attributes, as captured by the branch's fixed effect, and the transportation costs that block group residents must bear to travel to the branch. A branch's fixed effect proxies for its "quality," capturing all characteristics of the branch that make it a destination for residents of *any* block group in the period (e.g., the branch having better deposit or loan rates, higher staff attentiveness, or an efficient drive-through ATM).

By this index, residents have better bank branch access if branches are closer or if the nearest branches have superior attributes that attract more visitors. Unlike pure supply-side measures of access used in the banking literature, such as the number of branches per capita (Claessens 2006) or the availability of broadband Internet lines for online banking (Arnaboldi and Claeys 2008), which reveal the opportunity set available to consumers, this measure of access embodies consumers' actual choices. 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; De Sousa, Mayer, and Zignago 2012; Donaldson and Hornbeck 2016; Fajgelbaum and Gaubert 2020; and Adão, Carrillo, Costinot, Donaldson, and Pomeranz 2020).

While the geolocation data are critical to estimate the local measure of branch access, the data are subject to differential privacy methods, which try to shield the personal identifiable information of individuals from becoming public. Noise is added to the number of visitors from a block group to a branch, and these visitor counts are either truncated or censored if the number is too low. These distortions introduce non-classical measurement error into the data. An OLS regression of the gravity equation (or a Poisson pseudo-maximum-likelihood estimation) would render biased estimates and contaminate the measure of access.

To account for the differential privacy, we instead use the Method of Simulated Moments (MSM) to estimate the gravity equation. A key insight of our approach is to simulate data from the gravity model and then apply the same differential privacy algorithm to the simulated data that the data provider used to privacy-protect the real-world data. Standard MSM is straightforward to implement this way in models with no or few fixed effects (McFadden 1989). But the gravity equation has hundreds of

---

<sup>2</sup>Using block group fixed effects to control for consumer demand is akin to using firm fixed effects to absorb credit demand shocks when estimating the consequences of credit supply shocks, in the style of Khwaja and Mian (2008).

thousands of fixed effects across block groups and branches, which severely complicate the procedure. To tackle this issue in the gravity model estimation, we introduce an econometric method that adapts MSM to handle high-dimensional fixed effects.

This econometric method is the paper's second main contribution, as the approach can be implemented in other empirical settings involving high-dimensional fixed effects in MSM estimation. We run the MSM procedure month-by-month to account for dynamic branch entry and exit. The monthly point estimates of the gravity coefficient range from -1.45 to -1.26, implying that, if a representative branch is located 1% farther away from a representative block group, the expected number of residents from that block group who travel to that branch will drop by around 1.26% to 1.45% per month. This range is slightly higher in magnitude than the gravity coefficient estimate of -1.05 that [Agarwal, Jensen, and Monte \(2018b\)](#) find for the average nonfinancial, out-of-home consumption purchase, where the authors evaluate how consumer expenditures in nonfinancial sectors vary with distances from merchants.

In the second part of the paper, we use the gravity model estimates to uncover how bank branch access varies across the country. We document three geographic patterns: First, access varies substantially across regions. Residents of the eastern shores of New England and the Mid-Atlantic experience the highest access nationwide, whereas residents of the Deep South observe considerably weaker access. Second, the most pronounced differences in access are between urban and rural areas. Residents of big cities like Boston, Miami, Houston, and Minneapolis observe roughly three times better access to bank branches than residents of rural towns. Third, branch access varies greatly even across nearby block groups within the same town or city. For example, in New York City, access is highest in Midtown Manhattan and lowest in the Bronx, and even within each borough there is large variation. Likewise, around Chicago, residents of the North side experience better access than residents of the South side; and in Greater Los Angeles, the neighborhoods of Beverly Hills and the Hollywood area observe meaningfully better access than parts south of the city, such as Compton and Long Beach. A population-weighted, block-group-level regression of bank branch access on county fixed effects estimates that a sizeable 23% of the variation in access nationwide is within-county.

After examining geographic variation in branch access, we evaluate how access covaries with the household incomes and racial makeups of local communities. We compute national statistics, but we also zero in on large Metropolitan core areas, because these parts of the country contain the majority of U.S. households, and they have Black population shares that resemble the national average. Understanding bank branch access in these big cities is essential to explaining why Black households use branches less.

With controls for the racial and age population shares of block groups, we find that residents of low-income block groups both cross-country and in Metro cores have better access to bank branches than residents of high-income block groups. The better access can arise from either low-income

residents living relatively closer to branches or visiting branches of higher quality. To determine which channel explains the result, we split access into two components such that their product equals the original measure. The first component equalizes the qualities of all branches so that any variation in the first component across block groups is entirely driven by residents' different proximities to branches. The second component equalizes residents' proximities to branches so that any variation in the second component across block groups is entirely driven by residents' different weighted average qualities of branches, where closer branches are given higher weights. Performing this separation, we find that the higher access in low-income communities is entirely due to better branch proximity and not quality. We find no significant difference in the average quality of branches that residents of low-income and high-income areas experience.

Unlike residents of low-income block groups who observe relatively better bank branch access, we find that residents of block groups with high Black population shares experience *worse* access than residents living in block groups with high White population shares. This finding controls for a block group's median household income and the population shares of its residents' ages. Splitting access into the two components like before reveals that lower branch proximity is the reason behind the worse access. Residents of areas with high Black population shares experience *better* average branch quality compared to residents of areas with high White population shares. But because branches are located farther away from Black communities, those communities bear larger transportation costs to reach any branch. These costs more than offset the advantage of having the closest branches be of higher quality, leading to lower overall branch access.

In the third part of the paper, we isolate the extent to which access or demand explains the observed racial and income disparities in branch use. We start by regressing the expected total number of branch goers per block group, as estimated from the MSM, onto demographic characteristics of the block group residents. Controlling for block group racial and age shares, we find that every doubling in a block group's median household income is associated with 15.5 percent more expected branch goers per month, both nationwide and in Metro cores. This income gradient in branch use is large, as the unconditional likelihood of a resident in the geolocation data visiting a bank branch during the year is about 72 percentage points. Examining differences by race, and controlling for median household income and age shares, we find that a block group with a 100% Black population share expects roughly 5.6 percent fewer branch goers per month relative to a block group with a 100% White population share. In Metro cores, this Black-White gap in branch use is slightly smaller at 3.9 percent per month.

We next decompose the income gradient and the Black-White gap in branch use into constituent parts due to access and demand. Doing so simply requires regressing each block group's measure of access and its estimated fixed effect onto the same set of block-group-level characteristics that we use to evaluate the block group's expected total number of branch goers. By construction from the gravity equation, the estimated coefficients from the three regressions satisfy an identity along

any demographic attribute. The identity separates the elasticity of branch use with respect to the demographic attribute into two pieces: one part due to access and the other due to demand.

Performing this decomposition, we find that the +15.5 percent income gradient in branch use nationwide consists of a -7.5 percent income gradient in access and a +23.0 percent income gradient in demand. Thus, while residents of low-income block groups have relatively better access to bank branches, they exhibit a lower propensity to visit any branch, which translates overall into lower branch use compared to residents of high-income block groups. The Black-White gap in branch use nationwide has a different explanation. Black communities exhibit no robust statistical difference in their demand for branch products or services nationwide compared to White communities. And yet, residents of Black communities visit branches less, which implies that the 5.6 percent Black-White gap in branch use cross-country is entirely due to worse access.

In Metro cores, the decomposition by income reveals a similar pattern as the national estimates. Residents of low-income block groups have better access but weaker demand, which combine overall into lower branch use compared to residents of high-income block groups. Performing the decomposition by race, we find that residents of Black communities in big cities exhibit *higher* demand for branch products or services compared to residents of White communities. But this higher demand is more than offset by worse bank branch access, such that Black communities in big cities use branches less. At the end of the paper, we discuss different policy implications of our results, as the findings highlight parts of the country that would benefit the most from policies that expand access to banking.

**Related Literature.** This paper relates to several areas. The first is the 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. In a seminal paper, [Beck, Demirguc-Kunt, and Peria \(2007\)](#) develop indicators of banking sector outreach across 98 countries (e.g., the number of ATMs or loans per capita). [Beck, Demirgüç-Kunt, and Martinez Peria \(2008\)](#) measure bank access barriers (e.g., account fees or minimum account balances) across 62 countries. See also [Washington \(2006\)](#), [Blank \(2008\)](#), [Ho and Ishii \(2011\)](#), and [Goodstein and Rhine \(2017\)](#), who carefully investigate how improvements in different measures of bank access affect bank use in the United States. [Nguyen \(2019\)](#) exploits a creative instrumental variables design to study how branch closures causally affect local small firms' access to credit. [Argyle, Nadauld, and Palmer \(Forthcoming\)](#) quantify loan search frictions using the number of branches within a 20-minute drive, and they estimate the important real costs of these frictions on borrowers. [Yogo, Whitten, and Cox \(2022\)](#) examine U.S. bank and retirement account participation across an impressive swath of the population using tax records. [Jiang, Yu, and Zhang \(2023\)](#) propose a rich model of bank competition, and they find that branch closures lead to surplus losses for older depositors, and a policy that restricts closures can improve aggregate consumer surplus. [Célérier and Tak \(2023\)](#)



study racial disparities in banking through a keen historical lens, investigating the rise and fall of the Freedman’s Savings Bank in the 19th century. An advantage of this paper’s access measure is that it encapsulates the actual choices of individual consumers, rather than reflecting survey responses or being constructed from supply-side factors alone, such as the local branch density or the availability of low cost accounts. And while the causal benefits of increasing access are well established, less is known about the extent to which prevailing inequities in access sustain disparities in bank use. We attempt to make further progress in this body of work with a research approach that disentangles the effects of access and demand on the use of bank branches.

Second, the paper relates to the large literature in spatial economics on commuting flows and the arrangement of economic activity. See [Redding \(2013\)](#) and [Redding and Rossi-Hansberg \(2017\)](#) for surveys. Much of this work has focused on either firm or household location decisions (e.g., [Lucas and Rossi-Hansberg 2002](#); [Allen and Arkolakis 2014](#); [Ahlfeldt, Redding, Sturm, and Wolf 2015](#)) or agglomeration effects (e.g., [Dekle and Eaton 1999](#); [Rosenthal and Strange 2004](#)). Recently, many articles in this literature have taken advantage of geolocation data to answer economic questions. A pioneering example is [Chen and Rohla \(2018\)](#), who analyze how political partisanship affects time spent together at 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 \(2021\)](#) develop an innovative measure of segregation based on where people visit over the course of a day. See also [Büchel, Ehrlich, Puga, and Viladecans-Marsal \(2020\)](#); [Miyuchi, Nakajima, and Redding \(2021\)](#); [Kreindler and Miyuchi \(2022\)](#); and [Atkin, Chen, and Popov \(2022\)](#) for elegant analyses of social mobility and interactions using mobile device records. Many researchers have also used geolocation data to explore critical topics related to the Covid-19 pandemic (e.g., [Coven, Gupta, and Yao Forthcoming](#); [Almagro, Coven, Gupta, and Orane-Hutchinson 2021](#); [Goolsbee and Syverson 2021](#); [Couture, Dingel, Green, Handbury, and Williams 2022](#); and [Chen, Chevalier, and Long 2021](#)). We complement this literature by using micro-level observations of travel behaviors to quantify spatial patterns and transportation costs in banking.

Third, the econometric method we introduce in the paper relates to estimation procedures used in the spatial economics literature. Novel methods to handle econometric issues when estimating gravity models have been proposed before. In trade, [Eaton and Tamura \(1994\)](#); [Helpman, Melitz, and Rubinstein \(2008\)](#); and [Westerlund and Wilhelmsson \(2011\)](#) suggest clever methods to account for zero trade flows, including Tobit procedures, two-step Heckman procedures, and Poisson fixed-effects estimators. Recognizing potential biases introduced when log-linearizing a gravity equation, [Silva and Tenreyro \(2006\)](#) propose a Poisson pseudo-maximum-likelihood (PPML) procedure, which became seminal to the literature. Extensions of this estimator can handle a large number of fixed effects ([Larch, Wanner, Yotov, and Zylkin 2019](#)). Recently, [Dingel and Tintelnot \(2021\)](#) introduce a remarkably tractable spatial model and estimation procedure for “granular” environments, which

have a finite number of individuals whose idiosyncratic choices impact equilibrium outcomes. When the data in a spatial gravity model's estimation is subject to differential privacy and many fixed effects require estimation, we hope this paper's econometric method can be of use. More broadly, the paper provides a way to implement the Method of Simulated Moments when high-dimensional fixed effects require identification. Even in areas outside of spatial economics, differential privacy algorithms are masking more and more economic data sets over time, including the 2020 Census tables and American Community Survey microdata (Ruggles, Fitch, Magnuson, and Schroeder 2019), financial transactions data (Karger and Rajan 2020), and health records (Allen, Bavitz, Crosas, Gaboardi, Hay, Honaker, King, Korolova, Mironov, Phelan, Vadhan, and Wu 2020). The paper's econometric method can be helpful in estimating models of economic environments that rely on privacy-preserving datasets. Finally, the paper's conceptual measure of local access, empirically estimated using geolocation data, can be applied to other markets where physical retail locations matter, such as in grocery and healthcare markets.

**Outline.** The paper proceeds as follows. [Section 2](#) provides background information on the continued importance of bank branches in the U.S., especially for low-income and Black households. [Section 3](#) presents the paper's gravity model and measure of bank branch access. [Section 4](#) describes the geolocation data we use to estimate the gravity model. [Section 5](#) describes the econometric method of estimation. [Section 6](#) analyzes variation in branch access by geography and demographic characteristics of local areas. [Section 7](#) decomposes racial and income disparities in branch use by differences in access versus demand. [Section 8](#) discusses policy implications of the results. [Section 9](#) concludes.

## 2 Background on Bank Branches in the U.S.

Online and mobile banking over the past two decades have become major methods of household access to bank products and services (Boel and Zimmerman 2022). Uptake of Internet banking has been especially strong in developing countries (Laforet and Li 2005; Mbiti and Weil 2015; D'Andrea and Limodio 2023) and parts of Europe (D'Andrea, Pelosi, and Sette 2021; Mazet-Sonilhac 2022). Despite this advancement of digital banking, even in the U.S., physical bank branches continue to be important modes of bank access for U.S. consumers, particularly low-income and Black households.

According to the [2019 FDIC Survey of Household Use of Banking and Financial Services](#), roughly 81% of all banked and unbanked households visited a bank branch in the past 12 months, and 29.7% visited a branch 10 or more times. Traveling to a branch is the primary (i.e., most common) method of accessing bank accounts for about 21% of banked respondents. Mobile banking is more frequently cited as a primary method of use (34%) for banked households, but mobile banking is an incomplete substitute, as 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.

In addition to the FDIC Survey, the Survey of Consumer Finances (SCF) suggests that branches remain subjectively important to households in their use of banking products and services. According to the [2019 SCF](#), the locations of branches is cited most frequently as the most important reason for choosing an institution for a main checking account (43% of respondents), which is the highest cited reason by far. Despite advances in mobile and online banking over time, the proportion of respondents citing branch locations as their most important reason has remained roughly the same since 1989 (between 43% and 49%). Also, from the [2016 SCF](#), of the 84 percent of households with bank accounts who reported visiting a branch in the past year, almost all did so to use services other than just an ATM ([Anenberg, Chang, Grundl, Moore, and Windle 2018](#)).

Other survey evidence indicates several reasons why bank branches remain high in importance for consumers. Many households still prefer personal interactions for general banking services, complex transactions, and financial advice. A 2019 Deloitte survey of 17,000 banking consumers found that most respondents prefer branches over online or mobile banking when opening accounts (e.g., mortgage, wealth management, checking, credit card), and this preference for branches was uniform across generational cohorts ([Srinivas and Wadhvani 2019](#)). Branches appear to be symbols of trust for consumers, fostering brand recognition, a sense of security, and helping maintain face-to-face, personal banking relationships. A 2017 J.D. Power survey of retail banking customers found that respondents who used both digital and branch methods of access expressed greater satisfaction with their banks than respondents who used digital alone ([Hielscher 2017](#)). The inability to solve banking problems using the digital channel by itself was a key source of dissatisfaction.

Survey evidence also finds that consumers consider having accessible branches and ATMs nearby to be among the most important retail banking benefits ([Gaughan 2021; Martin 2023](#)). A 2020 survey from the financial advisory firm Novantas found that most respondents wanted their banks to maintain branches close to where they live or work, and 70% of consumers felt having a branch nearby was important ([Cocheo 2020](#)). Community Reinvestment Act disclosures also reveal that nearby branches are still vital for small business borrowers, as the share of loans made by lenders without a local presence, while increasing, remains low ([Anenberg et al. 2018](#)).

Beyond their importance to the general populace, bank branches play a meaningful role in serving low-income and minority communities. Empirical studies have found that when branches are located in these areas, borrowers living there—particularly those with limited credit histories—have greater access to credit and default less ([Ergungor 2010; Ergungor and Moulton 2011; Agarwal, Chomsisengphet, Liu, Song, and Souleles 2018a](#)). Bank branches also appear to be crucial modes of access for low-income and Black households. In Online [Appendix B](#), we analyze the 2019 FDIC survey's microdata and find that both household types indicate relying more on bank branches/ATMs and less on mobile/online banking as their most common method of bank access. Controlling for age and race in a multivariate linear

probability regression, we find that respondents in the lowest income bracket (< \$15,000) are roughly 25% more likely than those in the highest income bracket (\$75,000+) to say that bank tellers/ATMs are their primary method to access their bank accounts compared to mobile/online. Similarly, controlling for income and age, we find that Black respondents are about 6.4% more likely than White respondents to call bank tellers/ATMs their primary access method. Probit regressions tell a similar story.

And yet, despite evidently relying on branches more, both low-income and Black households visit branches less than high-income and White households, a pattern we aim to explain in our study. Controlling for age and race, we find that respondents in the highest income bracket are roughly 22% more likely to say they visited a branch in the previous year than respondents in the lowest income bracket. A substantial Black-White gap in reported branch use is also present. Controlling for income and age, we find that Black respondents are 10% less likely to answer having visited a branch in the past year than White respondents.

Finally, commercial banks themselves see physical branches as critical channels for acquiring customers, retaining them, and knowing them better (Horton 2019). Branches remain significant sources to attract bank deposits: JP Morgan Chase, for instance, saw 75% of its deposits growth in 2018 arising from customers using branches (Wathen 2018). Although the number of branches in the U.S. has been declining since 2013 (Anenberg et al. 2018), many financial institutions think of branches as salient “billboards” to consumers, and banks are reluctant to close their physical presence entirely in communities, in part because they lose information about the local economy (Nguyen 2021). To enhance the branch experience and draw customers in, several commercial banks are making large investments in redesigning their physical locations into upgraded “smart branches” that integrate with new technologies and amenities (Dallerup, Jayantilal, Konov, Legradi, and Stockmeier 2018). Just some investments are coffee bars, interactive video kiosks, AI-powered greeting robots, digital walls with product information, and private conference rooms for customer support (PNC Insights 2022).

Overall, bank branches remain integral to the financial lives of most U.S. households and a central customer touch point for commercial banks. Furthermore, we observe significant disparities in their use by both race and income that could be driven by differences in demand or in access. We turn next to providing the paper’s measure of bank branch access.

### 3 A Measure of Bank Branch Access

The paper’s measure of bank branch access originates from a standard log-linear, fixed-effects gravity equation that models consumer flows from home Census block groups to bank branches per time period. That gravity equation is:

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

The left-hand-side of Eq. (1) is the natural logarithm of the number of branch visitors traveling

from their home Census block group  $i$  to bank branch  $j$  in time period  $t$ . The right-hand-side of Eq. (1) includes four terms. The first term,  $\gamma_{it}$ , is a block group  $\times$  time fixed effect that captures all characteristics of block group  $i$ 's residents that contribute to them visiting any branch in the period. Informally, it represents block group-specific factors that influence residents' overall "demand" for branch products or services at any establishment (e.g., population, wealth, income, financial sophistication, trust in banks, flexibility in time).

The second term,  $\lambda_{jt}$ , is a branch  $\times$  time fixed effect that captures all characteristics of branch  $j$  that make it a destination for residents of any block group in the period. Informally, it represents branch-specific factors that contribute to a branch's "quality" (e.g., the branch having attractive deposit or loan rates, higher staff attentiveness, or many ATMs that avoid long customer queues).<sup>3</sup>

In the third term, the parameter  $\beta_t$  is the elasticity of visitor flows with respect to distance in the period. In many microfounded spatial models, the parameter can be interpreted as the product of residents' traveling costs and their elasticity of substitution between branches (Eaton and Kortum 2002; Ahlfeldt et al. 2015). The term  $\text{Distance}_{ij}$  is the geographic distance between block group  $i$  and branch  $j$ . In the estimation, we measure distance using the haversine formula, which accounts for the curvature of the Earth, and we compute the distances between branches and block groups' centers of population (See Footnote 19, where we also discuss driving time as a measure of "distance.") The fourth term,  $\varepsilon_{ijt}$ , is a mean-zero disturbance.<sup>4</sup>

To arrive at a measure of branch access using the gravity model, we exponentiate Eq. (1), take expectations, and sum across all bank branches available to visit in the period. Doing so gives block group  $i$ 's expected number of branch goes in the period:

$$E[V_{it}] = \exp(\gamma_{it}) \Phi_{it}. \quad (2)$$

The term  $\Phi_{it}$  is defined as:

$$\Phi_{it} \equiv \sum_{j \in B_t} \exp(\lambda_{jt}) d_{ij}^{-\beta_t}, \quad (3)$$

where  $B_t$  is the set of bank branches open in period  $t$ , and  $d_{ij}$  is the geographic distance between block group  $i$  and branch  $j$ .

Eq. (3) is the paper's granular measure of bank branch access. It summarizes information about the set of branches available to residents of a block group per time period. Given its form,  $\Phi_{it}$  can be interpreted as an attribute-adjusted branch index that is unique to each block group per period. Each branch in a block group's index is represented by (i) the "quality" of the branch's attributes in the period, as measured by its fixed effect,  $\lambda_{jt}$ , and (ii) the branch's distance,  $d_{ij}$ , away. The impact of

<sup>3</sup>Using origin and destination fixed effects to estimate gravity equations has become standard practice in the trade literature since Harrigan (1996). See also Fally (2015) for a connection between fixed-effects and structural gravity models. Because we have a panel, the cross-sectional fixed effects are time-varying.

<sup>4</sup>Eq. (1) can be derived from a differentiated product, discrete choice model of consumers selecting branches to visit per period. Online Appendix F provides one simple model of that sort, though not the only possible one.

distance is influenced by the gravity coefficient  $\beta_t$ . The term  $d_{ij}^{-\beta_t}$  is the inverse of the transportation costs that block group  $i$ 's residents must bear to reach branch  $j$ . While all branches across the country in the period are available to all residents to visit, the costs of traveling to those branches vary by block group. Local areas have better access if bank branches are relatively closer, especially branches with better attributes.

The object  $\Phi_{it}$  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](#); [De Sousa et al. 2012](#); [Donaldson and Hornbeck 2016](#); [Fajgelbaum and Gaubert 2020](#); and [Adão et al. 2020](#)). In our environment, we treat  $\Phi_{it}$  as a local measure of residents' access to bank branches.

The gravity model implies that residents visit bank branches directly from home. In the geolocation data, we will know visitors' home block groups, which will let us infer crucial demographic information from anonymous mobile devices. But households might not visit a branch directly from home, instead, perhaps, dropping by during lunchtime at work or in the middle of other errands. Unfortunately, the data we will use in the estimation does not reveal the exact travel paths of visitors. The access measure is therefore best thought of as capturing branch access around residents' home block groups.

We estimate the parameters of the gravity model in [Eq. \(1\)](#) and then aggregate estimates to the block-group level to obtain local measures of bank branch access. We turn next to describing the geolocation data we use in the estimation.

## 4 Geolocation Data on Branch Visitors

Branch visitors are based on geolocation data from mobile devices between January 2018 and December 2019. The data provider is the firm [SafeGraph](#). The data are monthly and include both branch locations and information about branch visitors. We do not use the "raw" pings from individual mobile devices, but rather, we use SafeGraph's aggregated geolocation data that try to protect user privacy. Rather than reporting the physical whereabouts of an individual device through time, this aggregated data report the home Census block groups of branch visitors and the associated number of visitors from each block group per month. In essence, the data provide the network of consumer trips from home block groups to bank branches each month.

The aggregated data are benefited by elaborate algorithms that SafeGraph has developed to accurately estimate whether a mobile device visits a particular destination and to pinpoint a mobile device's home origin, using the device's reported pings over time. However, the data do not give the demographic attributes of the mobile device owners, nor their home addresses or starting points of their trips, nor their duration spent at a branch, nor what they do at the branch.

A visitor in the SafeGraph data 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. Online [Appendix C](#) provides background information on the SafeGraph data and a detailed explanation of the way we construct our primary sample. Here, we give a summary.<sup>5</sup>

## 4.1 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 the SOD. We therefore include all Wells Fargo and SunTrust Bank branch locations in SafeGraph. We identify the physical locations of bank branches from SafeGraph’s geographic coordinates, and not from the SOD’s, as we found that SafeGraph’s coordinates typically were more accurate.<sup>6</sup>

Our core sample is confined to bank branches for which SafeGraph has visitor data. Many bank locations 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 shopping malls, SafeGraph might not be confident about the geometric boundary of a place. Not knowing the boundary makes it awfully difficult to attribute visitors to a unique place that is part of a shared space. 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 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.<sup>7</sup>

---

<sup>5</sup>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](#).

<sup>6</sup>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.)

<sup>7</sup>Regarding branch openings and closings, if a bank branch closed and SafeGraph were aware of its closure, any visitors

The SOD registers 86,374 bank branch locations as of 2019. While SafeGraph can account for 71,468 branches according to our core sample definition (83% coverage), only 51,369 of these places have visitor data and constitute our core sample. Our core sample thus covers around 60% of bank branches in the United States. Online Fig. A.1 presents a time-series of the number of branches per month in our core sample. Per month, the number of recorded branches is fairly stable and averages around 38,000. In Section 6.4, we perform a robustness check of our main findings on access by including all 2019 SOD bank branches. For branches in the SOD but not in SafeGraph, we use their geographic distances from block groups, and we impute their estimated fixed effects with the national average of the estimated fixed effects of branches in SafeGraph within the period. We also try the national median. Including all SOD branches confirms our main findings on how access varies by race and income.<sup>8</sup>

## 4.2 Sampling Bias

Our core sample experiences two types of sampling bias: (i) differential privacy and (ii) sample selection. 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  $\times$  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 raise it to 4. These last two data adjustments render our sample subject to both truncation from below and censoring from below, leading to non-classical measurement error. Fig. 2 presents the distribution of the observed (raw) visitor counts in black, which reveals both the truncation and censoring. Roughly 84% of the observed visitor counts equal 4, which implies a substantial amount of data distortion. The distortion also appears to vary by demographic attributes of residents. For example, in block groups with predominately Black residents (80%+), about 88% of visitor counts equal 4, whereas in the remaining block groups, about 83% equal 4. We account for SafeGraph’s differential privacy methods by estimating the gravity equation of Eq. (1) using an econometric method that adapts the Method of Simulated Moments to handle the estimation of high-dimensional fixed effects. Appendix A details the full procedure, and Section 5 provides a summary.

---

to the building (say, if a new business opened there) would no longer be attributed to the branch. Likewise, if a branch opened and SafeGraph were aware of it, 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.

<sup>8</sup>We focus our analysis on depository institutions in this paper and leave for follow-up work the study of access to non-depository institutions, like credit unions, and non-traditional financial institutions, like check cashers and payday lenders.



**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 U.S. 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 U.S. branches, in [Section 4.3](#) 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. In addition, we conduct a robustness check on access in [Section 6.4](#), where we include all branches in the 2019 SOD when evaluating how branch access varies across the country. As mentioned previously, our statistical findings on bank branch access are corroborated with the additional branches.

Regarding our sample of visitors, SafeGraph aggregates data from 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.<sup>9</sup> The 2010 U.S. 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 U.S. local home areas.

Nevertheless, we cannot rule out non-random sampling of mobile devices based on unobserved characteristics of visitors. As we discuss in more detail in [Online Appendix D](#), we do not know the precise demographic attributes of an individual bank branch visitor, and instead, we infer attributes of visitors according to the demographic characteristics of their home Census block groups. That is, we face the problem of ecological inference ([King 1997](#); [King, Tanner, and Rosen 2004](#)). The 2019 FDIC Survey reports smartphone ownership rates by household characteristics. Overall, 85.4% of respondents own smartphones, with Black respondents reporting mildly lower rates of ownership (81.5%) compared to White respondents (85.4%). Ownership rates decline to 66.4% among those aged 65+, 63.3% for those earning less than \$15,000 per year, and 75.6% for residents living outside Metropolitan areas. Smartphone ownership rates are also lower among the unbanked (63.7%) compared to the banked (86.6%). We likely under sample these groups with lower mobile device ownership rates.

Although the geolocation data have these sampling limitations, they are among the few sources of information on observed patterns of regular consumer travel, patterns that can be linked to important demographic data. As long as a block group is represented in the data—and nearly all U.S. block groups are in our sample—and the findings are framed at the block-group and not the individual level, the under sampling of certain households based on differential smartphone ownership is less of an issue. In fact, lower smartphone ownership rates among low-income and Black households

---

<sup>9</sup>Online [Fig. A.1](#) presents a time-series of the number of branch visitors each month over the sample period. The number of visitors 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.

reasonably make our estimates of branch use, branch access, and branch demand by income and race more conservative. The FDIC survey evidence shows that our two demographic groups of interest (low-income and Black households) visit branches less. Because these two groups are possibly under sampled in the geolocation data, the extent we find that they also visit branches less in that data is reasonably an underestimate.

To assess how well the mobile device visitation patterns align with representative survey evidence on branch visits, in Online [Appendix B.3](#), we compare 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. The comparison is by reported household income of the respondent and median household income of the mobile device’s home block group. 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 geolocation data are increasing and concave in household income.

Looking at the entire SafeGraph sample, [Squire \(2019\)](#) quantifies the sampling bias in the company’s data. He documents that the number of devices from SafeGraph’s identified home locations correlates highly at the county level with 2010 U.S. Census numbers in terms of population counts (97%), inferred educational attainment (99%), and inferred household income (99%).<sup>10</sup>

Despite this strong alignment between the Census and SafeGraph at the county level, [Thaenraj \(2021\)](#) identifies around 1,000 Census block groups in the SafeGraph data that register more devices residing there than the number of people living there according to the Census. [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 demographics because the SafeGraph representation lines up well at the county level.

For robustness, we weight block-group level regressions of branch access, branch use, and branch demand in [Sections 6 to 7](#) by the 2019 5-year American Community Survey (ACS) block group population counts. Population weighting down-weights block groups with disproportionately high mobile devices relative to their Census populations, and up-weights block groups with disproportionately low mobile devices relative to their populations, so as to make the visitor sample more representative.

### 4.3 Descriptive Statistics

Online [Table A.1](#) reports descriptive statistics of our core sample. The typical branch has 40 unique visitors per month on average, but there is wide dispersion across branches, as the standard deviation

---

<sup>10</sup>[Couture et al. \(2022\)](#) 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.

of visitors is over twice as high at 94. 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, but the standard deviation is 16 miles. The median dwell time is 49 minutes on average, but for half the branches in the sample, the median dwell time is 9 minutes or less. 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.

Online [Table A.2](#) compares demographic characteristics of residents living in the geographic areas covered by our core sample of bank branches with those in the areas covered by the full set of branches in the SOD. Demographic attributes in the table are taken from the 2019 5-year ACS and are averaged at the Census Bureau’s zip code tabulation area (ZCTA). In ZCTAs having branches in the SOD, the fraction of White households is 80.5%, which aligns closely with the 79.9% share of White households in ZCTAs having branches in 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 greater smartphone ownership rates in urban over rural areas. The differences in demographic attributes between the two samples are precisely estimated, but overall, the economic magnitudes of the differences are small relative to the mean values across areas.

## 5 Gravity Model Estimation

SafeGraph’s differential privacy methods bias any OLS or PPML estimation of the gravity model in [Eq. \(1\)](#). In this section, we describe our alternative econometric method to estimate the parameters of the gravity model. The method adapts the Method of Simulated Moments (MSM) to handle the estimation of high-dimensional fixed effects. The full details of the procedure are in [Appendix A](#); here, we provide a summary and present the model estimates.

### 5.1 Econometric Method

The goal of the econometric method is to uncover the parameters of the “true” distribution of branch visitors (based on the gravity model) from the distorted visitor data. Let  $V_{ijt}^*$  be the true number of visitors from block group  $i$  to branch  $j$  in year-month  $t$  that SafeGraph observes. 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:

$$V_{ijt}^+ = V_{ijt}^* + L_{ijt}. \quad (4)$$

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  $\times$  branch visitor count is present in the sample. The selection equation is

$$z_{ijt} = \begin{cases} 1 & \text{if } \lfloor V_{ijt}^+ \rfloor \geq 2, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

Let  $V_{ijt}$  denote the visitor count observed in the geolocation data, subject to SafeGraph’s censoring. The observation equation is

$$V_{ijt} = \max\{4, \lfloor V_{ijt}^+ \rfloor\}, \quad (6)$$

The econometric method involves simulating “true” visitor counts,  $V_{ijt}^*$ , from a presumed data generating process that follows the gravity model, manipulating the simulated data according to Eqs. (4) to (6), and then choosing the gravity model parameters to make selected moments of the manipulated simulated data match moments of the SafeGraph geolocation data as closely as possible. We run the estimation separately per year-month of the sample to account for branch openings and closings and to evaluate the stability of the estimates over time. The steps of the procedure follow.<sup>11</sup>

**Specify the data generating process.** In the simulations, we specify  $V_{ijt}^*$  as Poisson distributed. We presume a Poisson model for the visitor count data rather than an alternative distribution such as Negative binomial because it is parsimonious and interpretations of the fixed effects and the gravity coefficient are straightforward (Cameron and Trivedi 2013). Moreover, estimators based on the Poisson likelihood function have been ubiquitous to gravity model estimation since Silva and Tenreyro (2006).

Using the gravity model in Eq. (1), we express the true visitor count as following

$$V_{ijt}^* \sim \text{Pois}\left(\exp\left(\gamma_{it} + \lambda_{jt} - \beta_t \log \text{Distance}_{ij}\right)\right). \quad (7)$$

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 (see Footnote 19).

Eq. (7) imposes a log-linear relation between visitor counts and distance. Fig. 1 suggests that this relation represents the geolocation data fairly well. Panel A 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, with the fixed effects removed. A clear negative relation between distance and visitation is visible, and that relation is nearly linear. The relation does flatten out when the log number of visitors

<sup>11</sup>For textbook treatments of MSM, see Adda and Cooper (2003), Davidson and MacKinnon (2004), and Evans (2018).

approaches 1.4, but that change corresponds to SafeGraph’s censored value of 4 visitors. When only block group  $\times$  branch pairs exhibiting greater than 4 visitor counts are plotted in Panel B, the fairly linear relation retains throughout.

While the observed branch visitor patterns reasonably follow a gravity relation in distance, we do not claim that the relation is causal. It is true that the structural model of Eq. (7) implicitly assumes that the distances between branches and block groups is exogenous with respect to the parameters  $\{\gamma_{it}, \lambda_{jt}, \beta_t\}$ . This exogeneity assumption implies that knowledge of the distribution of distances is not required for inference of these parameters, and hence, we can validly condition the distribution of visitor counts on distance. But the distribution of distances is really the distribution of where people choose to live within block groups and where banks choose to build their branches. We treat these two choices as exogenous not because we believe they satisfy this classification in reality, but because the additional joint modeling of both residential choices and branch location choices is outside the scope of our investigation. We tackle the much narrower goal of explaining the discrete choices of branch visitors, taking as given their home locations and the branch locations in each year-month. We also estimate the gravity equation month-by-month, and at that short-term frequency, it is reasonable to assume that residential and branch location decisions are fairly fixed. Finally, Eq. (7) might omit other block group  $\times$  branch variables that influence branch visits (e.g., block-group targeted advertising or special block-group loan rates based on soft information). The extent to which these omitted variables have no predictive value for  $V_{ijt}^*$  after conditioning on distance, the parameter estimates are consistent.

**Sample block group  $\times$  branch pairs.** Technically speaking, every possible block group  $i$  and branch  $j$  pair should enter Eq. (7). But our data of over fifty-thousand branches, over two-hundred-thousand block groups, altogether spanning twenty-four months, makes it computationally impractical to have the billions of block group  $\times$  branch pairs enter the gravity model estimation. We instead include only stratified sampled pairs.

We implement the stratified sampling in the following manner. In each year-month, block group  $\times$  branch pairs in the SafeGraph data register either positive (and  $\geq 4$ ) or missing observed visitor counts. If a block group  $\times$  branch pair has a positive visitor count, then we know that residents of the block group visited the branch in the period, and we sample this block group  $\times$  branch pair in our simulation with probability 1. If a block group  $\times$  branch pair has a missing visitor count in the year-month, then either residents of the block group did not visit the branch in the period, or the visitor count was left out of the data from SafeGraph’s differential privacy methods. In each year-month, we sample from this *alternative* set of missing block group  $\times$  branch pairs such that (i) every pair in the alternative set has the same probability of being sampled, and (ii) each block group and each branch is part of at least one block group  $\times$  branch pair in the alternative set. The second condition ensures that each block group and branch is represented in the stratified sampling. We set the sampling probability

to  $1/2000$ , which implies that, on average, the randomly sampled alternative set of block group  $\times$  branch pairs represents slightly higher than a 0.05% sample size of all possible block group  $\times$  branch pairs with missing visitor counts. A larger sample size of 0.1% did not alter the estimation results.

Because we use stratified sampling, we must apply probability weights to any variable measured at the block group  $\times$  branch level, such as visitor counts or pairwise distances, so as to rebalance the data and make it represent the target population as closely as possible. Following standard practice, we use probability weights that are the reciprocal of the likelihood of being sampled (subject to rounding).<sup>12</sup>

Rather than using a stratified sampling of branches, we could have restricted the set of branches per block group using a distance-based cutoff (e.g., including in residents' choice sets only the branches within a 10-mile radius of their block group). But doing so would contaminate the comparison of block group fixed effects and bank branch access between areas. For example, if a block group had twice as many branches within a 10-mile radius than another block group, the former's access measure would be twice as high as the latter, all else equal. But actual branch visitation might be similar between the two block groups. If so, the estimated fixed effect of the block group with more branches would mechanically lower to equalize visitation. The stratified sampling implies that *all* branches in the sample per period are commonly available to *all* U.S. residents. In addition, including all branches in the choice sets of residents will let us uniquely pin down the branch fixed effects, which we discuss momentarily.

**Simulate the visitor counts.** We simulate visitor counts according to the Poisson model of Eq. (7) and apply the differential privacy methods expressed in Eqs. (4) to (6). The simulation process differs between the block group  $\times$  branch pairs that are sampled with probability 1 and those sampled with probability  $1/2000$ . For the pairs sampled with probability 1, we draw Poisson random variables with distinct means given in Eq. (7). We then (i) add a Laplace draw to all non-zero "true" visitor counts to form a "noisy" block group  $\times$  branch visitor count, (ii) round down each noisy visitor count 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.

For the block group  $\times$  branch pairs sampled with probability  $1/2000$ , we do not draw random variables in the simulation. If these visitor counts were randomly drawn, they would have disproportionate impact on any computed moments because of the high probability weights that would multiply them. Noise from the simulation would be amplified and make the estimation unstable. In other words, there is a peso problem. Rather than randomly drawing these pairs of visitor counts, we construct their implied empirical probability distribution given the parameter estimates. If an infinite number

---

<sup>12</sup>The stratified sampling is similar in spirit to the choice reduction procedure in McFadden (1977) and Davis, Dingel, Monras, and Morales (2019), where the choice sets of consumers include the selections actually chosen (in our case, branches visited) plus a random subset of all other alternatives (other branches that residents could have visited). We say "similar in spirit" because McFadden (1977) and Davis et al. (2019) also sample from a set of available alternatives, but they estimate conditional logit models, unlike here.

of observations were in fact simulated for these pairs, their distribution would coincide with this constructed empirical distribution.<sup>13</sup>

Because the Laplace noise is added after the Poisson draw, this empirical distribution is a truncated and censored Laplace distribution whose mean is the realization of the Poisson draw. We construct 7 components of this empirical distribution: (i)-(ii) the probability that the visitor count equals 0, as well as exceeds 0; (iii)-(iv) the probability that the visitor count equals 4, as well as exceeds 4; (v) the expected visitor count; and (vi)-(vii) the expected natural logarithm of visitor counts, conditional on visitor counts exceeding 0, as well as exceeding 4.

**Iterate the fixed effects until convergence.** The MSM 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. (7) requires the estimation of hundreds of thousands of fixed effects. Estimating all these parameters from the MSM minimization problem alone would be computationally impractical. Instead, we adopt an iterative routine to identify the fixed effects  $\{\gamma_{it}, \lambda_{jt}\}$  and let the minimization problem identify  $\beta_t$ . First, given an estimate of  $\beta_t$  and estimates of the block group fixed effects,  $\{\gamma_{it}\}$ , we take advantage of another data field in SafeGraph: a branch’s total number of visitors. This field is unaffected by SafeGraph’s differential privacy methods. Because the stratified sampling implicitly presumes that visitors can arrive from any block group, we can uniquely identify each branch’s fixed effect from an “adding up” condition. Specifically, we sum the means of Eq. (7) across block groups, for each branch, which gives the model’s *predicted* total number of visitors to the branch in the year-month. We then set that sum equal to the branch’s *observed* total number of visitors in SafeGraph, and then invert the equation to extract the branch’s fixed effect. Then, given estimates of  $\beta_t$  and the branch fixed effects,  $\{\lambda_{jt}\}$ , from the inversions, we repeatedly update the block group fixed effects,  $\{\gamma_{it}\}$ , until the differences in the average simulated visitor counts and average observed visitor counts of each block group  $i$  across all branches per year-month  $t$  become sufficiently small. Notice that, with every update to the block group fixed effects,  $\{\gamma_{it}\}$ , the branch fixed effects,  $\{\lambda_{jt}\}$ , also update from the inversions. When the number of fixed effects in each dimension is large, as in our setting, this routine produces consistent estimates.<sup>14</sup>

**Construct the MSM estimator.** After both dimensions of fixed effects are identified per estimate of  $\beta_t$ , the MSM minimization problem then selects the optimal  $\beta_t$  estimate that minimizes the weighted

---

<sup>13</sup>Notice that we cannot apply this approach to the set of block group  $\times$  branch pairs sampled with probability 1 because each pair in that set is drawn from a distinct distribution, due, in part, to the block group- and branch-specific fixed effects. For the pairs sampled with probability 1, we simulate draws. However, the pairs in the alternative set that are sampled with probability  $1/2000$  are meant to represent the remaining population of block group  $\times$  branch pairs, which are very high in number. One stratified sampled observation from the alternative set is meant to represent 2,000 observations from the same distribution. We need only construct the empirical distribution that these sampled pairs represent.

<sup>14</sup>The iterative process we use to identify the fixed effects is similar in spirit to the “zig-zag” algorithm, or Gauss-Seidel method, that is commonly used to identify high-dimensional fixed effects in linear models (Guimaraes and Portugal 2010).

sum of squared errors (expressed in percentage points) between the simulated model moments and data moments. We use 6 unconditional moments that describe important parts of the distribution of visitor counts: (i)-(ii) the fractions of visitor counts equaling 0 and equaling 4; (iii)-(iv) the average log distances, when visitor counts equal 0 and equal 4; and (v)-(vi) the OLS coefficients from regressing log visitor counts onto their associated log distances, when visitor counts equal 0 and equal 4. The model moments include both the simulated draws from the block group  $\times$  branch pairs that are sampled with probability 1 and components of the empirical distribution that represent the alternative set of pairs that are sampled with probability  $1/2000$ .

## 5.2 Gravity Estimates

**Fig. 2** compares the distribution of observed “raw” visitor counts (in black) to simulated “true” visitor counts (in blue) according to the month-by-month MSM estimation of the Poisson model in **Eq. (7)**. The simulated visitor counts include all positive draws from all simulations across every year-month in the sample period. The numbers of 0 visitor counts in both distributions are very large and are omitted for clarity. The black distribution reveals the effects of the differential privacy on the raw visitor counts, having a large mass at 4. The MSM does a reasonable job spreading out the mass of visitors into the lower portion of the distribution that is lost in the observed data. The two distributions line up fairly well at the right tail, where we have censored the visitor counts at 10 in the figure for clarity. The “true” visitor distribution in blue obeys our assumed Poisson structure, which may not coincide with the true data generating process of visitor counts known only to SafeGraph. Nevertheless, as with standard MSM, potential misspecification of the simulating distribution does not interfere with the consistency of the estimates (**McFadden 1989**). Also displayed in the figure is the distribution of simulated “manipulated” visitor counts in red, which is the distribution of the “true” visitor counts after they are manipulated by the differential privacy methods in **Eqs. (4) to (6)**.

**Fig. 3** compares the observed number of visitors from each Census block group to their expected (i.e., predicted) counterparts from the simulation. It presents a binned scatter plot of the log observed number of branch goers from each block group versus the log expected number of branch goers from the block group based on the MSM estimates. If SafeGraph applied no differential privacy methods to their geolocation 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 log 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. In block groups with many branch goers, the observed and expected number of visitors nearly match. This implies that the MSM generates estimates that fit the



geolocation data well in regions least affected by the differential privacy distortions, which one would hope for.

Fig. 4, Panel A presents the gravity coefficient estimates through time, along with 95% confidence intervals. The monthly point estimates of the gravity coefficient range from about -1.45 to -1.26, and they are fairly stable month-to-month. Thus, across the country, if a representative branch were located 1% farther away from a representative block group, the number of residents from that block group who travel to that branch would drop by around 1.26-1.45% per month. As for comparison, Agarwal et al. (2018b) estimate a gravity model of consumer expenditures in nonfinancial sectors. They find a gravity coefficient of -1.05 for the average out-of-home purchase, but they document significant heterogeneity across sectors, with Food Stores, for example, observing an estimate of -0.85; and Health Services, -0.33.

Fig. 4, Panels B and C present histograms of the estimated Census block group and bank branch fixed effects across all months of the sample period. A block group's fixed effect can be interpreted as the average log number of residents from that block group who visit any branch in the year-month, controlling for branch fixed effects and transportation costs. The bulk of the distribution of block group fixed effects range from exponentiated values around 0.01 to 20. 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 their transportation costs. Most of the mass is within a range of exponentiated values between 0.01 and 30. In an unreported regression, roughly 77% of the variation in a branch's fixed effect over time can be explained by the branch itself, suggesting that branch quality is fairly stable over time.

SafeGraph's differential privacy methods bias traditional methods of estimating the gravity model and prompts an alternative econometric method like the MSM. But computing estimates from the traditional methods is still useful to informally assess the magnitude of the bias. To this end, Online Table A.6 presents gravity coefficient estimates and standard errors from the MSM estimation, along with estimates and standard errors from OLS and PPML estimations. The PPML and OLS estimates are computed on the observed, "raw" visitor counts. We run each estimation approach per year-month of the sample. PPML and OLS estimations are also run over the full sample panel period (January 2018 - December 2019). In addition, we run the OLS estimation on block group  $\times$  branch pairs with more than 4 visitors (which avoid SafeGraph's censoring). PPML and OLS standard errors are two-way clustered by both Census block groups and bank branches.

The gravity coefficient estimates from the MSM range from -1.45 to -1.26. The estimates from OLS range from -0.062 to -0.038, roughly twenty to thirty times smaller in magnitude. The OLS estimate over the full panel is -0.053, still an order of magnitude below the MSM estimates. When the sample is limited to block group  $\times$  branch pairs with greater than 4 visitors, the OLS estimates rise in magnitude, ranging from -0.33 to -0.27, which is still roughly four to five times smaller in magnitude than the

MSM estimates. Computed over all block group  $\times$  branch pairs, the PPML estimates register higher magnitudes than the OLS ones, ranging in values from -0.108 to -0.066. But they still are roughly ten to twenty times smaller in magnitude than the MSM estimates. The PPML gravity coefficient estimate over the full panel is -0.091.

Overall, Online [Table A.6](#) reveals the downward bias that SafeGraph’s differential privacy methods introduce to traditional methods of estimating the gravity equation, and it stresses the need for the alternative MSM procedure.

## 6 Bank Branch Access in the United States

With the gravity model estimates, we next construct local measures of bank branch access, and we study variation in access throughout the country. The estimated counterpart of [Eq. \(3\)](#) is the empirical measure of branch access for residents of block group  $i$  in year-month  $t$ :

$$\hat{\Phi}_{it} \equiv \sum_{j \in B_t} \exp(\hat{\lambda}_{jt}) d_{ij}^{-\hat{\beta}_t}, \quad (8)$$

Because all block group and branch pairs are represented in the estimation from the stratified sampling, every branch has an estimated fixed effect, and thus, each home block group’s measure of access includes all branches nationwide in the period.

The relation between branch access and branch visitor counts from the Poisson model of [Eq. \(7\)](#) bestows economic meaning to the magnitude of  $\hat{\Phi}_{it}$ . Because the access measure aggregates across all branches, it can be interpreted as a block group’s expected total number of branch goers per month, when the block group’s fixed effect is set to zero. The block group fixed effects control for block group populations, so residents of high-population areas do not mechanically have better access. We characterize bank branch access by evaluating its spatial heterogeneity over different parts of the U.S. and by measuring its association with demographic characteristics of block group residents.

### 6.1 Geography of Branch Access

[Fig. 5](#) illustrates a dot density map of bank branch access by Census block group across the country. Each dot is positioned at a block group’s center of population. We compute the access estimates 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 observed branch visitors over the sample period. We construct the map by grouping block groups into deciles and shading the dots so that higher ordered colors in the rainbow gradient (indigo and violet) imply higher bank branch access, and lower ordered colors (red and orange) imply lower access. Block groups where no resident was recorded in SafeGraph as having visited a branch in the sample period are shaded white.

Throughout the country, three patterns are apparent. First, access varies substantially across regions. The eastern shores of New England and the Mid Atlantic, along with the upper Midwest experience the highest access nationwide. The Appalachian states, such as parts of West Virginia, Tennessee, and Kentucky, observe relatively better access than the Deep South, such as the southern parts of Mississippi and Alabama. A population-weighted, block-group-level regression of branch access on Census region fixed effects suggests that 74.4% of the variation in access nationwide is within the four Census regions.

Second, the most pronounced differences in access are between urban and rural areas. Big cities like Boston, Richmond, Miami, Philadelphia, Houston, and Minneapolis observe substantially higher bank branch access than even nearby suburban areas. Suburban areas that outlie large metropolitan centers observe greater access than rural parts of the country. Online [Fig. A.2](#) compares branch access by Rural-Urban Commuting Areas (RUCAs), as classified by the U.S. 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. We categorize each block group into its RUCA code, and we present in the figure the population-weighted averages of bank branch access per RUCA. Access declines monotonically as one transitions from Metropolitan core to Metropolitan suburb, micropolitan suburb, and rural areas. Branch access in the typical rural area is about two-thirds less than access in a typical Metropolitan core.

Third, even within a local area, branch access varies significantly. A population-weighted, block-group-level regression of access on county fixed effects estimates that a sizeable 23% of cross block-group variance in access nationwide is within-county. [Fig. 6](#) zeroes in on the four largest cities in the U.S. by population: New York City, Los Angeles, Chicago, and Houston. Around New York City, bank branch access is highest for residents of Midtown Manhattan and lowest for those living in the Bronx, and within each borough there is large variation. But relatively speaking, access is high overall for residents of New York City, such that the numbers in the city's lowest access neighborhoods are comparable to those in the highest access neighborhoods of Los Angeles. In Greater Los Angeles, residents of Beverly Hills and the Hollywood area observe substantially better access than residents living south of the city, such as in Compton and Long Beach. Residents of neighborhoods in the Palos Verdes Peninsula, San Fernando, and near the Anaheim Hills observe the lowest access. Residents of the Palos Verdes Peninsula experience weaker access despite the neighborhood being relatively affluent. Around Chicago, residents of the North side experience better access than those living in the South side, an area with a high Black population share. Residents of the Northwest suburbs near Lake Forest observe the highest access outside the Loop, which also corresponds to neighborhoods with high White population shares. Variation in bank branch access is so large around Chicago, that the decile break points of the distribution of access in Cook County, where the city is located, nearly matches

the decile break points of the distribution of branch access nationwide. Finally, in Houston, access is highest for residents living downtown, and access declines more and more, almost uniformly, as one moves farther away from the city into such areas as Sugar Land, Brookside Village, and Humble. The radial decline in access moving away from the downtown area mirrors the general pattern observed nationally in the neighborhoods surrounding big cities.<sup>15</sup>

The heterogeneity in bank branch access displayed in these four big cities hint that access is correlated with the income and racial composition of residents in local areas. We turn next to formally evaluating how access varies with those attributes.

## 6.2 Branch Access by Income and Race

Because a central focus of our study is understanding how bank branch access differs between Black and White households, we look at national comparisons, but we also zero in on parts of the country that present Black population shares close to the national average. From the 2019 5-year ACS, the national Black share is 12%. The commuting areas with Black population shares closest to this national number are Metropolitan area core (Metro core), having a 15% Black share, and Micropolitan area core (micro core), having a 9% Black share. Although Metro and micro cores share similar racial shares, Metro cores vastly outnumber micro cores in household counts (99.5 million vs. 8.5 million), and Metro cores capture roughly 72% of the 138.9 million total households in the U.S. For this reason, we supplement our national estimates of branch access with local estimates in Metro cores.

**Table 1** presents weighted OLS regressions of  $\log \hat{\Phi}_{it}$  on demographic attributes of block group residents. Observations in the regressions are at the level of a home Census block group per year-month of the core sample period, they are weighted by 2019 5-year ACS block-group population counts, and standard errors are clustered at the block-group level. Independent variables are population-based shares from the 2019 5-year ACS and the log number of mobile devices residing in the block group within the period, which is unaffected by SafeGraph’s differential privacy methods. The number of devices is included to interpret access on a per capita basis. 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. The coefficients from the table can be interpreted as the percent change in a block group’s expected number of branch goers per month, assuming the block group fixed effect is held constant at zero.

The branch access measure embodies information about the per-mile cost of travel through the gravity coefficient  $\beta_t$ , which we take to be constant nationwide per year-month for simplicity. But

---

<sup>15</sup>A common measure of bank access is an area’s density of branches. In Online **Table A.9**, we regress the branch access measure on branch density at both the county and census tract levels. At these geographies, between 44-62% of the variation in access is explained by density when state fixed effects are included, suggesting that our measure of bank branch access captures more information than just supply-side factors. In addition, the coefficient in the regression changes sign from positive to negative as the landscape enlarges from census tracts to counties, which reveals the importance of measuring access over as narrow a terrain as possible.

commuting costs indeed vary substantially throughout the diverse U.S. 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 alone 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. In addition to the RUCA fixed effects, all specifications include year-month and county fixed effects. Thus, one can interpret the specifications as comparing the bank branch access of residents living within the same county, at the same time period, within, say, a small town, but in different block groups.

Column (1) conditions the branch access regressions on median household income and population racial shares. Residents of block groups with lower median household income experience higher access. A doubling in a block group’s median household income is associated with its residents experiencing about 11.0% weaker access. Residents of block groups with high Black population shares also have weaker access, on the order of 8.2%. Extrapolating from the coefficients on the racial shares, we thus estimate that a hypothetical block group with a 100% Black population share would observe roughly 8.2% fewer branch goers per month than a comparable block group with a 100% White population share, holding constant the block group fixed effects at zero.

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). Including income in column (1) already proxies for the permanent component of human capital, but in column (2), we add age shares. Controlling for age in column (2) still preserves the negative relation between income and access, though the magnitude is cut from -11.0% to -7.6%. Also controlling for age, we observe that residents of block groups with large Black population shares also have worse access. Extrapolation of the coefficient implies that residents of a hypothetical block group with a 100% Black population share observe about 5.3% poorer access than residents of a comparable block group with a 100% White population share.

Column (3) restricts the sample to block groups in Metropolitan core areas. (Here, because we limit the sample of block groups to belong to one type of RUCA, we exclude the RUCA fixed effects.) In this column, the negative coefficients on income and the Black share are sharper. In big cities, a doubling of a block group’s median household income is associated with its residents observing about 12.6% weaker access, and the Black-White gap in access is 10.7%. Controlling for age shares in column (4), we find that the coefficients on income and the Black population share remain negative, though smaller in magnitude (-8.7% income gradient and 6.4% Black-White gap).

### 6.3 Explaining Access: Branch Proximity and Branch Quality

The measure of bank branch access combines information about the “quality” of branches available to residents and the cost of traveling to those branches. We next evaluate how each component contributes to differential access across block groups. Bank branch access in Eq. (8) can be rewritten as

$$\hat{\Phi}_{it} \equiv \underbrace{\left( \sum_{j \in B_t} d_{ij}^{-\beta_t} \right)}_{\text{Branch Proximity}} \times \underbrace{\left( \sum_{j \in B_t} \pi_{ijt} \exp(\hat{\lambda}_{jt}) \right)}_{\text{Branch Quality}}, \quad (9)$$

where the weights  $\pi_{ijt} \equiv \frac{d_{ij}^{-\beta_t}}{\sum_{j \in B_t} d_{ij}^{-\beta_t}}$ . Eq. (9) separates access into two components. The first is branch proximity, measured as the sum of the inverse of the transportation costs that residents would bear to reach bank branches across the country. The closer that residents are to bank branches, the higher their branch proximity. The component can be interpreted as residents’ hypothetical access if branch quality were equalized across all branches (and normalized to one). The second component is branch quality, measured as the weighted average quality of branches that residents experience, where closer branches are assigned higher weight. It can be interpreted as residents’ hypothetical access if branch proximity were equalized across block groups (and normalized to one).

Columns (5)-(8) of Table 1 present the coefficients of weighted OLS regressions of these two access components on block-group median household income, racial shares, and age shares. Columns (5) and (7) provide nationwide estimates, whereas columns (6) and (8) focus on Metro cores. Nationwide, a doubling in a block group’s median household income is associated with a 7.6% decline in its residents’ proximity to all bank branches. In Metro cores, the drop in proximity is higher at 8.6%. Looking at branch quality in columns (7)-(8), we find no statistical difference in the average quality of branches that residents experience in low-income versus high-income communities, both nationwide and in big cities. Put together, the decomposition reveals that the higher bank branch access for residents of low-income block groups is entirely driven by their greater proximity to bank branches and not from experiencing a higher quality of branches.

Focusing on the racial differences in the two access components, we find that residents of block groups with high Black population shares experience significantly lower branch proximity. Nationwide, residents of a block group with a 100% Black share observe 14.3% lower branch proximity compared to residents of a block group with a 100% White share. In Metro cores, the drop in proximity to branches for Black communities is 16.8%. However, both nationwide and in Metro cores, residents of block groups with high Black population shares experience better average branch quality (9.0% higher quality cross-country and 10.4% higher quality in big cities). This result implies that reduced proximity to branches, rather than lower quality branches, explains the lower branch access in Black communities. Put differently, the branches nearest Black communities tend to be of higher average quality, but those

branches are located relatively farther away from those communities, and hence, are costlier to reach.<sup>16</sup>

Overall, results from Sections 6.2 to 6.3 show that (1) across the country and in Metro cores, residents of low-income block groups have better branch access, and we find no meaningful difference in their average quality of branches compared to high-income communities; and (2) residents of areas with high Black population shares have worse branch access, and this worse access is driven by bank branches being located relatively farther away from Black communities, rather than these communities experiencing lower average quality branches.

#### 6.4 Robustness Check: Access With All SOD Branches

In our main analysis, we estimate the gravity model in Eq. (7) using distances and visitor counts from block groups to branches in SafeGraph for which we have visitor data. We then use Eq. (8) as an empirical measure of bank branch access per block group. Our sample covers virtually all U.S. Census block groups, but not all U.S. branches. Here, we re-examine branch access by demographic attributes when all branches in the 2019 SOD are included.

A block group’s estimated bank branch access per period in Eq. (8) consists of four components: (i) the set of branches available to all residents  $B_t$ , (ii) the distances between the block group and branches  $\{d_{ij}\}_{v,j}$ , (iii) the estimated gravity coefficient  $\hat{\beta}_t$ , and (iv) the estimated fixed effects of the branches  $\{\hat{\lambda}_{jt}\}$ . For component (i), we use the set of branches in SafeGraph already included in the main analysis plus the set of branches in the 2019 SOD that are missing from SafeGraph. For component (ii), we use the haversine distances used before for the branches in the SafeGraph sample plus the haversine distances between the block groups’ centers of populations and the addresses of the SOD branches.

In many microfounded gravity models, component (iii), the estimated gravity coefficient  $\hat{\beta}_t$ , can be interpreted as the product of consumer’s traveling costs and elasticity of substitution between branches per period. We need to make an assumption for this component. It is reasonable to presume that, had we been able to estimate the gravity model of Eq. (7) with all bank branches in the 2019 SOD, households’ per unit traveling costs would not have been different than the one implicitly embedded in our earlier estimated  $\hat{\beta}_t$ . But the elasticity of substitution between branches might have been different. Adjusting for this possible change would be challenging. For simplicity, we use each month’s estimated value of  $\hat{\beta}_t$  from the main analysis.

Finally, the estimated branch fixed effects  $\{\hat{\lambda}_{jt}\}$  of component (iv) require two assumptions: values for the estimated fixed effects of the branches in SafeGraph and values for the estimated fixed effects of the branches in the SOD but not in SafeGraph. First, for the SafeGraph branches, had we estimated the gravity model on geolocation data involving all bank branches, the estimated fixed effects of those branches in SafeGraph could have been different from the values we estimated using data only on our

---

<sup>16</sup>If the coefficients on the Black population share in columns (7) and (8) registered the opposite sign, bank branch access would have been doubly bad for Black communities. Not only would branches have been located farther away from them, but their nearest branches would be of lower average quality.

core sample of branches. Rather than speculating the direction of the change, for simplicity we instead apply their estimated fixed effects from the main analysis. For the branches in the 2019 SOD but not in SafeGraph, we assume that their estimated fixed effects equal the national average of the estimated fixed effects of the SafeGraph branches within the year-month. We also try the national median.<sup>17</sup>

Online Table A.8 repeats the branch access OLS regressions of Table 1, but now with all branches from the 2019 SOD included in each block group’s measure of access. For the estimated fixed effects of the branches in the SOD but missing from SafeGraph, columns (1)-(4) use the national mean of the estimated branch fixed effects of the branches in SafeGraph per year-month, whereas columns (5)-(8) use the national median per year-month.

Using either the mean or median produces the same conclusion: Adding all the SOD branches reinforces the earlier statistical findings on branch access by race and income. The income gradient nationwide in column (2) of Table 1 was -7.6%. In column (2) of Online Table A.8, the income gradient sharpens in magnitude to -8.4%. In Metro cores, the negative income gradient sharpens from -8.7% to -9.3%. The Black-White gap in access nationwide widens when all SOD branches are included. In column (2) of Table 1, the Black-White gap in access was 5.3%. In column (2) of Online Table A.8, the Black-White gap is 15.6%. In Metro cores, the Black-White gap in access widens from 6.4% to 17.8%.

## 7 Bank Branch Use: Access versus Demand

The gravity model takes advantage of the network of consumer flows linking block group residents to multiple branches. By using block group fixed effects, it isolates how the unequal spatial distribution of bank branches explains disparities in household use of those branches, by controlling for block group differences in demand for branch products or services. Since the comparison is across branches for the *same* block group, block group-specific demand is absorbed by the block group fixed effect.

As an added benefit, the gravity model in Eq. (7) generates a simple decomposition of branch visitor counts into two parts. From that equation, the natural logarithm of the predicted total number of residents of block group  $i$  who visit any branch in year-month  $t$  is

$$\log \hat{V}_{it}^* = \hat{\gamma}_{it} + \log \hat{\Phi}_{it}. \quad (10)$$

Eq. (10) separates a block group’s predicted total number of branch goers into (i) the number of visitors if branch access were equalized across block groups, plus (ii) the number of visitors if branch demand instead were equalized. This decomposition thus separates the parts of branch use that are due to demand versus access. Notice that Eq. (10) is the log estimated counterpart of Eq. (2).

---

<sup>17</sup>In this section, we only examine how access would change if all bank branches in the 2019 SOD were included. Based on Eq. (2), one could further study how expected branch visitation would change per block group, but doing so would require more assumptions on how the block group fixed effects  $\{\gamma_{it}\}$  would change had all branches been included. With more bank branches available, we may have observed more branch goers generally per block group, which would raise a block group’s estimated fixed effect. Rather than further speculating the change in the fixed effect per block group when all SOD branches are included, we instead examine only how branch access might change with the additional branches.



We can also evaluate the extent to which *variation* in branch use across population groups can be attributed to differences in each component. An OLS regression of  $\log \hat{V}_{it}^*$  on a vector  $X_i$  of resident characteristics at the block-group level gives

$$\log \hat{V}_{it}^* = X_i \theta_V + \varepsilon_{V,it}. \quad (11)$$

Similar regressions of both the estimated block group fixed effects and the access measure on  $X_i$  produce

$$\hat{\gamma}_{it} = X_i \theta_\gamma + \varepsilon_{\gamma,it}, \quad (12)$$

$$\log \hat{\Phi}_{it} = X_i \theta_\Phi + \varepsilon_{\Phi,it}. \quad (13)$$

Along any demographic attribute  $x$ , the estimated coefficients from Eqs. (11) to (13) satisfy the identity:

$$\hat{\theta}_{V,x} \equiv \hat{\theta}_{\gamma,x} + \hat{\theta}_{\Phi,x}, \quad (14)$$

which separates the elasticity of branch use with respect to the demographic attribute into two constituent parts: one part due to demand for branch products or services,  $\hat{\theta}_{\gamma,x}$ , and the other part due to bank branch access,  $\hat{\theta}_{\Phi,x}$ .

We begin the empirical analysis of this section by running regressions of the sort in Eq. (11) and Eq. (12). Regressions of the sort in Eq. (13) were presented earlier in Section 6.2. We then use the estimates to decompose disparities in branch use into parts explained by differential access to and differential demand for branch products or services.

## 7.1 Bank Branch Use

Table 2 presents weighted OLS regressions of bank branch visitors by demographic attributes, both nationwide and in Metro cores. Just as in the access regressions of Table 1, observations are at the level of home Census block groups per year-month over the sample period; they are weighted by 2019 5-year ACS block-group population counts; standard errors are clustered at the block-group level; and year-month, county, and RUCA fixed effects are included. The natural log number of devices is included in each specification to interpret the coefficients on a per capita basis.

In columns (1)-(4), the dependent variable 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. (7). Column (1) reports coefficients on median household income and racial shares. High-income block groups have more expected bank branch goers per month (about 18.6% more for every doubling in the block group's median household income). Block groups with high Black population shares have fewer expected branch goers compared to block groups with high White population shares. A hypothetical block group with a 100% Black share would have roughly 4.1% fewer branch goers per month relative to a comparable block group with a 100% White share. Controlling for block group age shares in column (2), we find that the coefficient on income drops slightly (15.5%), but

the coefficient on the Black share is more sharply negative (-5.6%). Extrapolations of the coefficient suggest that a block group with a 100% Black population share would expect 5.6% fewer branch visitors per month relative to a comparable block group with a 100% White share.

Moving away from nationwide estimates, we next focus on Metro core areas in columns (3) and (4). Having controls for age shares in column (4), we find that the positive income gradient in branch use in big cities remains the same as the cross-country estimate (15.5%). The magnitude of the coefficient on the Black population share is -3.9%, which is slightly narrower than the national estimate. Thus, in big cities, we detect a Black-White gap in branch use of roughly 3.9% per month.

## 7.2 Bank Branch Demand

In columns (5)-(8) of [Table 2](#), we report coefficients from weighted OLS regressions of the estimated block group fixed effects,  $\{\hat{\gamma}_{it}\}$ , on demographic attributes of block group residents, both cross-country and in Metro cores. In all columns, low-income block groups observe lower fixed effects, which implies that residents of these areas have a lower propensity to visit any bank branch. A 1% decline in a block group's median household income is associated with a roughly 0.23-0.32% drop in its residents' "demand" for branch products or services. If the block group's access were held fixed and normalized to one, this range would correspond with the decline in the block group's expected number of branch goers per month.<sup>18</sup>

We cannot say with certainty why residents of low-income areas exhibit lower demand for bank branch products or services. But evidence from the [2019 FDIC Survey](#) provides some potential explanations. Among unbanked households, the top 5 reasons cited for not having a bank account are (i) not having enough money to meet minimum balance requirements (48.9% of respondents), (ii) not trusting banks (36.3%), (iii) avoiding a bank gives more privacy (36.0%), (iv) bank account fees are too high (34.2%), and (v) fees are too unpredictable (31.3%). The extent to which these reasons correlate with a respondents' income, they can be explanations for the lower demand among residents of low-income block groups. But our sample also includes banked, and quite likely, underbanked residents, the latter of whom have bank accounts but still rely on alternative financial services like payday loans. Low-income banked and underbanked residents might exhibit lower demand for bank branch products or services for the same reasons as unbanked residents. But they might also have less demand for the kinds of premium services that require visiting a branch, such as storing valuables in a safety deposit box, using notary services, or consulting with a banker about more complex financial issues like wealth management or small business banking. Finally, nonbank financial institutions like check cashers might target their services and advertising to low-income customers, shifting their demand away from banks.

---

<sup>18</sup>Notice that the block group fixed effects load positively and strongly on the block group log number of devices, confirming intuition that higher population areas would exhibit higher "demand" for branch products and services overall. This relation contrasts with the small and negative loading of branch access on the log number of devices in [Table 1](#).

Turning to racial differences in bank branch demand, we find that block groups with high Black population shares observe larger fixed effects than block groups with high White shares, with controls for income alone in column (5). Extrapolations of the coefficient suggest that a block group with a 100% Black share would observe about 4.1% more branch visitors per month compared to a comparable block group with a 100% White share, if bank branch access were equalized across the country (and normalized to one). But when age controls are added, the sign of the coefficient on the Black share turns negative and is no longer precisely estimated. The change suggests that demand for branch products or services nationwide between Black and White communities is not robustly different. In Metro cores, block groups with high Black population shares do observe precisely estimated higher fixed effects without age controls (8.9% higher demand in column 7), and the coefficient stays precisely estimated once age controls are added (2.5% higher demand in column 8).

Overall, results from [Sections 7.1 to 7.2](#) show that (1) across the country and in Metro cores, residents of low-income block groups use branches less than residents of high-income block groups, but residents of these low-income block groups exhibit significantly lower demand for branches via lower block group fixed effects, and (2) residents of areas with high Black population shares also use branches less cross-country and in big cities than residents of comparable block groups with high White population shares. But residents of Black communities nationwide show no robust difference in their demand for branch products or services, and they exhibit slightly higher demand in big cities.

### 7.3 Decomposing Branch Use

Looking at different retail markets, researchers have studied how weaker access to vital establishments, such as grocery stores and hospitals, depresses the use of goods at those locations and hurts welfare ([Yantzi, Rosenberg, Burke, and Harrison 2001](#); [Inagami, Cohen, Finch, and Asch 2006](#); [Nicholl, West, Goodacre, and Turner 2007](#)). Furthermore, access to these places varies by socioeconomic status ([Currie and Reagan 2003](#); [Hamrick, Hopkins et al. 2012](#)). See also [Allcott, Diamond, Dubé, Handbury, Rahkovsky, and Schnell \(2019\)](#) for an excellent analysis that separates out demand and supply explanations for the inequalities in nutrition observed by income. When it comes to banking markets, we can decompose the variation in branch use into parts due to differences in access versus demand. We do so by inserting the estimates from the branch access regressions of [Table 1](#) and the estimates from the branch use and demand regressions of [Table 2](#) into the identity of [Eq. \(14\)](#).

Nationwide, the income gradient in branch use from column (2) of [Table 2](#) is +15.5%. This income gradient consists of a demand gradient of +23.0% from column (6) of the table and an access gradient of -7.5% in column (2) of [Table 1](#). The decomposition thus shows that the lower demand for branch products or services among residents of low-income block groups dominates their higher access and is associated with their lower overall branch use. When it comes to racial disparities, column (2) of [Table 2](#) reveals a Black-White gap in branch use of 5.6%. This gap consists of 0.3% lower demand and 5.3%

lower access. But the difference in demand is imprecisely estimated, implying that the Black-White gap in branch use is entirely due to a Black-White gap in branch access rather than in branch demand.

In Metro cores, the income gradient in branch use (column 4 in [Table 2](#)) is +15.5%, which can be separated into an income gradient in demand of +24.2% (column 8 in [Table 2](#)) and an income gradient in access of -8.7% (column 4 in [Table 1](#)). As was the case nationally, residents of low-income block groups experience higher access but lower demand, with the latter eclipsing the former enough so that overall branch use is lower in low-income communities. As for racial differences in big cities, residents of areas with high Black population shares exhibit slightly higher demand (an elasticity of +2.5%) compared to residents of areas with high White population shares, but residents of Black communities experience significantly weaker access (an elasticity of -6.4%). Thus, in big cities, the lack of access in Black communities is powerful enough to surpass the higher demand for branch products or services, indicating a 3.9% Black-White gap in branch use.

To summarize, we find that low-income communities have better bank branch access, both nationally and in big cities, but their demand for branch products or services is lower compared to high-income communities. For these low-income communities, a lack of access cannot explain their lower branch use. Black communities, on the other hand, have worse access compared to White communities both cross-country and in big cities. Black communities exhibit no robust difference nationwide in demand for branch products or services compared to White communities and mildly stronger demand in big cities. But their relative lack of access is substantial enough to overcome their higher demand, revealing a Black-White gap in branch use in big cities as well.

## 8 Discussion of Policy Implications

Many researchers and policymakers have proposed programs to increase bank access by enlarging the physical presence of branches, aiming to remedy disparities in bank participation ([Dahl and Franke 2017](#); [Davidson 2018](#)). Example proposals are investing in community development banks, which are certified commercial banks that principally serve minority and low-income communities ([Ellwood and Patel 2018](#)); and expanding U.S. postal banking, which would add checking, savings, and possibly credit services to some or all U.S. Post Office branches ([Baradaran 2013](#)). Our results suggest that these policies would have their largest impact on bank access in Black communities, particularly those in big cities.

In [Online Appendix E](#), we examine one of these policies—U.S. postal banking—investigating how its expansion might affect both access to and the use of bank branches. In September 2021, the U.S. Postal Service (USPS) launched a pilot program in four post office locations that allowed customers to cash payroll and business checks in the form of gift cards ([Heckman 2022](#)). Adding more financial services and expanding the program to all Post Office locations would require congressional legislation.

To study the potential expansion of the policy, we re-estimate each block group’s bank branch access after adding all Post Office locations registered in SafeGraph to the set of available branches. We use the time-series estimates of the gravity coefficient  $\hat{\beta}_t$  from before; we leave unchanged each private bank’s estimated fixed effect,  $\hat{\lambda}_{jt}$ ; and we retain the same estimate of each block group fixed effect,  $\hat{\gamma}_{it}$ . This kind of analysis, which ignores the general equilibrium effects under the policy change, is a “partial policy” evaluation of postal banking, 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).

Whereas private banks in the analysis retain their original estimated fixed effects, we consider three different scenarios of fixed effects for the USPS locations. In each scenario, all postal locations share the same fixed effect in a year-month for simplicity. In the first scenario, postal branches are “low quality,” in that they all share the estimated fixed effect of the 10<sup>th</sup> percentile of private banks per year-month. In the second scenario, postal branches are “medium quality,” having the estimated fixed effect of the 50<sup>th</sup> percentile of private banks per year-month. In the third scenario, postal branches are “high quality,” having the estimated fixed effect of the 90<sup>th</sup> percentile of private banks per year-month.

In the low- and medium-quality scenario, we find that an expansion of the postal banking system would *widen* the Black-White gap in access, both nationwide and in big cities. Extending banking services to Post Office locations would mechanically increase bank branch access for everyone, but it would enhance access relatively more for residents of predominately White areas than for residents of predominately Black areas. The racial gap in access widens because Post Offices also tend to be located comparatively closer to White communities than Black communities, just like private bank branches. Only in the case of a high quality postal banking system would the Black-White gap in access shrink. Nationwide, it would shrink from 5.3% to 4.8% per month, and in Metro cores, it would shrink from 6.4% to 5.3% per month. Because block group fixed effects retain their same values from before, any change in the racial gap in access would equal the relative amounts that branch use would change between groups. By raising access for everyone, postal banking would be associated with increased branch use across all demographic groups, but the increase would be relatively smaller in Black communities than in White communities. The exception is a high quality system, which would shrink the Black-White gap in branch use by 0.5 percentage points cross-country and 1.1 percentage points in Metro cores per month.

Our partial policy evaluation ignores any changes to block group residents’ demand for banking products or services that postal banking might bring about, which is a notable shortcoming of the analysis. Distaste for private banks might partly explain the low estimated block group fixed effects of some communities. In the [2019 Survey of Consumer Finances](#), 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 unbanked respondents selecting this option as their reason has increased steadily over time (from 15% in 1989 to 22.9% in 2019). Proponents of postal banking argue that the unbanked will perceive

Post Office bank branches as more trustworthy than private banks, making the unbanked more likely to utilize Post Office locations (Office of the USPS Inspector General 2014; Baradaran 2015). If this is true, postal banking could improve bank participation by raising demand for bank branch products or services. Our main analysis suggests that low-income communities would benefit the most from this potential outcome of the policy, as residents of these areas observed the lowest estimated block group fixed effects.

An expansion of postal banking could also trigger a competitive reaction from private banks, leading to an endogenous change in their branch fixed effects, something we also do not account for. Private banks might develop different fee and service structures to attract unbanked and underbanked residents. If this were to happen, a postal banking policy could indirectly increase access and consumer welfare by inducing private banks to improve their branch quality in response to competitive pressures. But why do private banks not already cater more to the unbanked and underbanked? We find that Black communities exhibit slightly higher demand for branch products or services in big cities, but they experience significantly weaker access there. Why do private banks not enter these areas to a greater extent? It is beyond the paper's scope to answer this question, but it is important to stress that we measure branch demand from branch visitation, which might not perfectly correlate with bank profits. Private banks might not enter because doing so is unprofitable. If that were the case, another policy that could remedy weak branch access would be allocating tax credits to banks that establish new branches in Black communities, or subsidizing community development banks to further expand there.

The policy responses discussed so far relate to changes in the locations or quality of bank branches. But another policy that could improve bank access is enhancing broadband Internet connectivity to reduce the costs of mobile or online banking. The two groups we identify making the least use of bank branches (low-income and Black households), simultaneously appear to rely on branches *more* as their primary method of banking, rather than mobile or online. Weak access to broadband could be a reason why. Black adults are about 9 percentage points less likely to have home broadband than White adults (71% vs. 80%), and low income families earning less than \$30,000 are 13 percentage points less likely to have broadband at home than families earning more than \$75,000 (86% vs. 99%) (Pew Research Center 2021). Programs to expand broadband connectivity in low-income and Black communities could encourage residents to make greater use of mobile or online banking as a substitute for visiting a physical branch.

Expanding broadband could also raise consumer access to Fintech firms, which are non-depository institutions offering financial services entirely online. A purported goal of Fintech is increasing the financial inclusion of groups whom traditional banks have historically underserved. But evidence so far on Fintech's success at achieving that goal is mixed. Erel and Liebersohn (2022) find that Fintech lenders expanded credit from the Paycheck Protection Program to small business owners located in

areas with more low-income and minority residents. But Fuster, Plosser, Schnabl, and Vickery (2019) find that Fintech borrowers of purchase mortgages tend to have higher incomes and are less likely to be minorities. In addition, Friedline, Narahariseti, and Weaver (2020) and Friedline and Chen (2021) find that low-income communities of color have the lowest Fintech adoption rates. More time is needed to observe the full potential of Fintech on closing disparities in financial access. But in every wave of the Survey of Consumer Finances since 1989, the location of a bank’s branches is cited as the most important reason by far for choosing an institution for a main checking account (43.1% of respondents). This persistent stated preference suggests that programs targeting the geographic distribution of retail locations where consumers receive their financial services will remain important for public policy.

## 9 Conclusion

We use anonymous geolocation data from mobile devices to develop a local measure of bank branch access. The measure is derived from a spatial gravity model and is an expression of a local area’s distance from surrounding bank branches and branch quality. Both the model and the rich data of consumer travel patterns allow us to implement a research design that quantifies the extent to which differences in access or in demand explain the continued racial and income disparities in household use of bank branches.

To overcome distortions in the geolocation data that safeguard user privacy, we estimate the gravity model using a new econometric method that adapts the Method of Simulated Moments to handle the estimation of hundreds of thousands of fixed effects. We hope the method can be of use in other applications relying on big data that are potentially contaminated by differential privacy methods.

Controlling for block-group racial and age shares, we find that access is better for residents of low-income communities, which implies that inadequate branch access cannot explain their lower use of bank branches. Instead we find that lower demand for bank branch products or services explains their lower branch use. In contrast, residents of areas with high Black population shares experience weaker access, principally because branches are located farther away from them. The relative lack of branch access for Black communities in big cities is so significant that it eclipses their slightly higher demand for branch products or services. The consequence is their overall lower use of bank branches.

## References

- ADÃO, R., P. CARRILLO, A. COSTINOT, D. DONALDSON, AND D. POMERANZ (2020): “Imports, Exports, and Earnings Inequality: Measures of Exposure and Estimates of Incidence,” *Quarterly Journal of Economics*.
- ADDA, J. AND R. COOPER (2003): *Dynamic Economics: Quantitative Methods and Applications*, MIT Press.
- AGARWAL, S., S. ALOK, P. GHOSH, S. GHOSH, T. PISKORSKI, AND A. SERU (2017): “Banking the unbanked: What do 255 million new bank accounts reveal about financial access?” Working paper. Columbia University, New York, NY.

- AGARWAL, S., S. CHOMSISENGPHET, C. LIU, C. SONG, AND N. S. SOULELES (2018a): "Benefits of relationship banking: Evidence from consumer credit markets," *Journal of Monetary Economics*, 96, 16–32.
- AGARWAL, S., J. B. JENSEN, AND F. MONTE (2018b): "The geography of consumption," Working paper. Georgetown University, Washington, D.C.
- AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.
- ALLCOTT, H., R. DIAMOND, J.-P. DUBÉ, J. HANDBURY, I. RAHKOVSKY, AND M. SCHNELL (2019): "Food deserts and the causes of nutritional inequality," *Quarterly Journal of Economics*, 134, 1793–1844.
- ALLEN, J., C. BAVITZ, M. CROSAS, M. GABOARDI, M. HAY, J. HONAKER, G. KING, A. KOROLOVA, I. MIRONOV, M. PHELAN, S. VADHAN, AND A. WU (2020): "The OpenDP White Paper," Working Paper.
- ALLEN, T. AND C. ARKOLAKIS (2014): "Trade and the Topography of the Spatial Economy," *Quarterly Journal of Economics*, 129, 1085–1140.
- ALMAGRO, M., J. COVEN, A. GUPTA, AND A. ORANE-HUTCHINSON (2021): "Racial disparities in frontline workers and housing crowding during COVID-19: Evidence from geolocation data," Working paper. New York University, New York, NY.
- ANENBERG, E., A. C. CHANG, S. GRUNDL, K. B. MOORE, AND R. WINDLE (2018): "The Branch Puzzle: Why Are there Still Bank Branches?" *FEDS Notes*.
- ARELLANO, M. AND S. BOND (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *Review of Economic Studies*, 58, 277–297.
- ARGYLE, B., T. D. NADAULD, AND C. PALMER (Forthcoming): "Real effects of search frictions in consumer credit markets," *Review of Financial Studies*.
- ARNABOLDI, F. AND P. CLAEYS (2008): "Internet banking in Europe: a comparative analysis," *Research Institute of Applied Economics*, 8, 1–28.
- ATHEY, S., D. BLEI, R. DONNELLY, F. RUIZ, AND T. SCHMIDT (2018): "Estimating heterogeneous consumer preferences for restaurants and travel time using mobile location data," in *AEA Papers and Proceedings*, vol. 108, 64–67.
- ATHEY, S., B. FERGUSON, M. GENTZKOW, AND T. SCHMIDT (2021): "Estimating experienced racial segregation in US cities using large-scale GPS data," *Proceedings of the National Academy of Sciences*, 118, 1–9.
- ATKIN, D., M. K. CHEN, AND A. POPOV (2022): "The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley," Working paper no. 30147. National Bureau of Economic Research, Cambridge, MA.
- BARADARAN, M. (2013): "It's time for postal banking," *Harvard Law Review Forum*, 127, 165–175.
- (2015): *How the other half banks: Exclusion, exploitation, and the threat to democracy*, Harvard University Press.
- BECK, T., A. DEMIRGÜÇ-KUNT, AND M. S. MARTINEZ PERIA (2008): "Banking services for everyone? Barriers to bank access and use around the world," *World Bank Economic Review*, 22, 397–430.
- BECK, T., A. DEMIRGUC-KUNT, AND M. S. M. PERIA (2007): "Reaching out: Access to and use of banking services across countries," *Journal of Financial Economics*, 85, 234–266.
- BLANK, R. M. (2008): "Access to financial services, savings, and assets among the poor," *National Poverty Center Policy Brief* (13).
- BLANK, R. M. AND M. S. BARR (2009): *Insufficient funds: Savings, assets, credit, and banking among low-income households*, Russell Sage Foundation.
- BOEL, P. AND P. ZIMMERMAN (2022): "Unbanked in America: A Review of the Literature," *Economic Commentary*.
- BROWN, J. R., J. A. COOKSON, AND R. Z. HEIMER (2019): "Growing up without finance," *Journal of Financial Economics*, 134, 591–616.
- BÜCHEL, K., M. V. EHRLICH, D. PUGA, AND E. VILADECANS-MARSAL (2020): "Calling from the outside: The role of networks in residential mobility," *Journal of Urban Economics*, 119, 103277.
- CAMERON, A. C. AND P. K. TRIVEDI (2013): *Regression analysis of count data*, vol. 53, Cambridge University Press.
- CASKEY, J. P. (1994): *Fringe banking: Check-cashing outlets, pawnshops, and the poor*, Russell Sage Foundation.
- CASKEY, J. P. AND A. PETERSON (1994): "Who has a bank account and who doesn't: 1977 and 1989," *Eastern Economic Journal*, 20, 61–73.

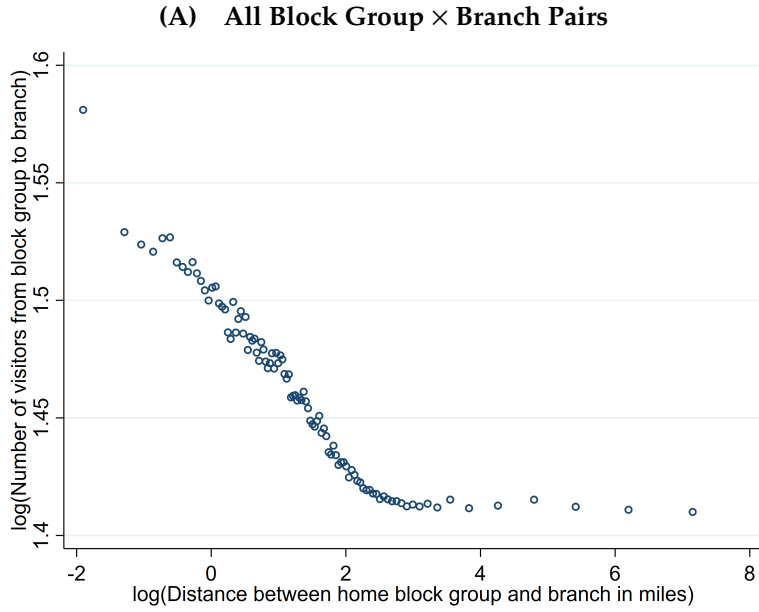


- CÉLERIER, C. AND A. MATRAY (2019): "Bank-branch supply, financial inclusion, and wealth accumulation," *Review of Financial Studies*, 32, 4767–4809.
- CÉLERIER, C. AND P. TAK (2023): "Finance, Advertising, and Fraud: The Rise and Fall of the Freedman's Savings Bank," Working Paper. University of Toronto, Toronto, CA.
- CHEN, M. K., J. A. CHEVALIER, AND E. F. LONG (2021): "Nursing home staff networks and COVID-19," *Proceedings of the National Academy of Sciences*, 118.
- CHEN, M. K., K. HAGGAG, D. G. POPE, AND R. ROHLA (2019): "Racial disparities in voting wait times: Evidence from smartphone data," *Review of Economics and Statistics*, 1–27.
- CHEN, M. K. AND R. ROHLA (2018): "The effect of partisanship and political advertising on close family ties," *Science*, 360, 1020–1024.
- CLAESSENS, S. (2006): "Access to financial services: A review of the issues and public policy objectives," *World Bank Research Observer*, 21, 207–240.
- CLAESSENS, S. AND E. PEROTTI (2007): "Finance and inequality: Channels and evidence," *Journal of Comparative Economics*, 35, 748–773.
- COCHEO, S. (2020): "Rethinking Branch Networks Without Killing Sales or Jeopardizing Growth," *The Financial Brand*.
- COUTURE, V., J. I. DINGEL, A. GREEN, J. HANDBURY, AND K. R. WILLIAMS (2022): "JUE Insight: Measuring movement and social contact with smartphone data: a real-time application to COVID-19," *Journal of Urban Economics*, 127, 1–9.
- COVEN, J., A. GUPTA, AND I. YAO (Forthcoming): "Urban flight seeded the Covid-19 pandemic across the United States," *Journal of Urban Economics: Insights, Forthcoming*.
- CURRIE, J. AND P. B. REAGAN (2003): "Distance to hospital and children's use of preventive care: Is being closer better, and for whom?" *Economic Inquiry*, 41, 378–391.
- DAHL, D. AND M. FRANKE (2017): "'Banking deserts' become a concern as branches dry up," *Population*, 12, 5–9.
- DALLERUP, K., S. JAYANTILAL, G. KONOVA, A. LEGRADI, AND H.-M. STOCKMEIER (2018): "A bank branch for the digital age," *McKinsey Company*.
- DAVIDSON, C. (2018): "Lack of access to financial services impedes economic mobility," *Economy Matters: Federal Reserve Bank of Atlanta*.
- DAVIDSON, R. AND J. G. MACKINNON (2004): *Econometric Theory and Methods*, vol. 5, Oxford University Press New York.
- DAVIS, D. R., J. I. DINGEL, J. MONRAS, AND E. MORALES (2019): "How segregated is urban consumption?" *Journal of Political Economy*, 127, 1684–1738.
- DE SOUSA, J., T. MAYER, AND S. ZIGNAGO (2012): "Market access in global and regional trade," *Regional Science and Urban Economics*, 42, 1037–1052.
- DEKLE, R. AND J. EATON (1999): "Agglomeration and land rents: Evidence from the prefectures," *Journal of Urban Economics*, 46, 200–214.
- DINGEL, J. I. AND F. TINTELNOT (2021): "Spatial economics for granular settings," Working paper. The University of Chicago, Chicago, IL.
- DONALDSON, D. AND R. HORNBECK (2016): "Railroads and American economic growth: A "market access" approach," *Quarterly Journal of Economics*, 131, 799–858.
- D'ANDREA, A. AND N. LIMODIO (2023): "High-Speed Internet, Financial Technology, and Banking," *Management Science*.
- D'ANDREA, A., M. PELOSI, AND E. SETTE (2021): "Broadband and Bank Intermediation," Working Paper. Bocconi University, Milan, Italy.
- EATON, J. AND S. KORTUM (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.
- EATON, J. AND A. TAMURA (1994): "Bilateralism and regionalism in Japanese and U.S. trade and direct foreign investment patterns," *Journal of the Japanese and International Economies*, 8, 478–510.
- EISFELDT, A. L. (2007): "Smoothing with liquid and illiquid assets," *Journal of Monetary Economics*, 54, 1572–1586.
- ELLWOOD, D. T. AND N. G. PATEL (2018): "Restoring the American Dream," *U.S. Partnership on Mobility from Poverty*.
- EREL, I. AND J. LIEBERSOHN (2022): "Can FinTech reduce disparities in access to finance? Evidence from the Paycheck Protection Program," *Journal of Financial Economics*, 146, 90–118.

- ERGUNGOR, O. E. (2010): "Bank branch presence and access to credit in low-to moderate-income neighborhoods," *Journal of Money, Credit and Banking*, 42, 1321–1349.
- ERGUNGOR, O. E. AND S. MOULTON (2011): "Do bank branches matter anymore?" *Economic Commentary*.
- EVANS, J. W. (2018): "Simulated Method of Moments (SMM) Estimation," *QuantEcon Notes*.
- FAJGELBAUM, P. D. AND C. GAUBERT (2020): "Optimal spatial policies, geography, and sorting," *The Quarterly Journal of Economics*, 135, 959–1036.
- FALLY, T. (2015): "Structural gravity and fixed effects," *Journal of International Economics*, 97, 76–85.
- FITZPATRICK, K. (2015): "The effect of bank account ownership on credit and consumption: Evidence from the UK," *Southern Economic Journal*, 82, 55–80.
- FRIEDLINE, T. AND Z. CHEN (2021): "Digital redlining and the Fintech marketplace: evidence from US zip codes," *Journal of Consumer Affairs*, 55, 366–388.
- FRIEDLINE, T. AND M. DESPARD (2016): "Life in a banking desert," *The Atlantic*.
- FRIEDLINE, T., S. NARAHARISSETTI, AND A. WEAVER (2020): "Digital redlining: Poor rural communities' access to fintech and implications for financial inclusion," *Journal of Poverty*, 24, 517–541.
- FUSTER, A., M. PLOSSER, P. SCHNABL, AND J. VICKERY (2019): "The role of technology in mortgage lending," *Review of Financial Studies*, 32, 1854–1899.
- GAUGHAN, M. (2021): "Accessing a branch is 'highly important' to banking customers," *Insider Intelligence*.
- GOODSTEIN, R. M. AND S. L. RHINE (2017): "The effects of bank and nonbank provider locations on household use of financial transaction services," *Journal of Banking & Finance*, 78, 91–107.
- GOOLSBEE, A. AND C. SYVERSON (2021): "Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020," *Journal of Public Economics*, 193, 104311.
- GUIMARAES, P. AND P. PORTUGAL (2010): "A simple feasible procedure to fit models with high-dimensional fixed effects," *The Stata Journal*, 10, 628–649.
- HAMRICK, K. S., D. HOPKINS, ET AL. (2012): "The time cost of access to food—Distance to the grocery store as measured in minutes," *International Journal of Time Use Research*, 9, 28–58.
- HANSON, G. H. (2005): "Market potential, increasing returns and geographic concentration," *Journal of International Economics*, 67, 1–24.
- HARRIGAN, J. (1996): "Openness to trade in manufactures in the OECD," *Journal of International Economics*, 40, 23–39.
- HARRIS, C. D. (1954): "The market as a factor in the localization of industry in the United States," *Annals of the Association of American Geographers*, 44, 315–348.
- HEAD, K. AND T. MAYER (2004): "Market potential and the location of Japanese investment in the European Union," *Review of Economics and Statistics*, 86, 959–972.
- (2014): "Gravity equations: Workhorse, toolkit, and cookbook," in *Handbook of International Economics*, vol. 4, 131–195.
- HECKMAN, J. (2022): "USPS continues postal banking pilot, despite House Republicans' objections," *Federal News Network*.
- HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): "Estimating trade flows: Trading partners and trading volumes," *Quarterly Journal of Economics*, 123, 441–487.
- HIELSCHER, J. (2017): "Bank customers still want their branches," *Herald-Tribune*.
- HO, K. AND J. ISHII (2011): "Location and competition in retail banking," *International Journal of Industrial Organization*, 29, 537–546.
- HOGARTH, J. M., C. E. ANGUELOV, AND J. LEE (2005): "Who has a bank account? Exploring changes over time, 1989–2001," *Journal of Family and Economic Issues*, 26, 7–30.
- HOGARTH, J. M. AND K. H. O'DONNELL (1997): "Being accountable: A descriptive study of unbanked households in the US," in *Proceedings of the Association for Financial Counseling and Planning Education*, 58–67.
- HORTON, D. (2019): "The Top 3 Reasons why Banks Branches are still important," *LinkedIn*.
- HWANG, J. AND Y. SUN (2018): "Should we go one step further? An accurate comparison of one-step and two-step procedures in a generalized method of moments framework," *Journal of Econometrics*, 207, 381–405.
- INAGAMI, S., D. A. COHEN, B. K. FINCH, AND S. M. ASCH (2006): "You are where you shop: Grocery store locations, weight, and neighborhoods," *American Journal of Preventive Medicine*, 31, 10–17.

- JIANG, E. X., G. Y. YU, AND J. ZHANG (2023): "Bank Competition amid Digital Disruption: Implications for Financial Inclusion," Working paper. University of Southern California, Los Angeles, CA.
- KARGER, E. AND A. RAJAN (2020): "Heterogeneity in the marginal propensity to consume: Evidence from Covid-19 stimulus payments," Working paper. Federal Reserve Bank of Chicago, Chicago, IL.
- KHWAJA, A. I. AND A. MIAN (2008): "Tracing the impact of bank liquidity shocks: Evidence from an emerging market," *American Economic Review*, 98, 1413–42.
- KREINDLER, G. E. AND Y. MIYAUCHI (2022): "Measuring commuting and economic activity inside cities with cell phone records," *Review of Economics and Statistics*, 1–48.
- LAFORET, S. AND X. LI (2005): "Consumers' attitudes towards online and mobile banking in China," *International journal of bank marketing*.
- LARCH, M., J. WANNER, Y. V. YOTOV, AND T. ZYLKIN (2019): "Currency unions and trade: A PPML re-assessment with high-dimensional fixed effects," *Oxford Bulletin of Economics and Statistics*, 81, 487–510.
- LUCAS, R. E. AND E. ROSSI-HANSBERG (2002): "On the internal structure of cities," *Econometrica*, 70, 1445–1476.
- MARTIN, D. (2023): "The more things change, the more branches still matter," *American Banker*.
- MAZET-SONILHAC, C. (2022): "Information Frictions in Credit Markets," Working Paper. Bocconi University, Milan, Italy.
- MBITI, I. AND D. N. WEIL (2015): "Mobile banking: The impact of M-Pesa in Kenya," in *African successes, Volume III: Modernization and development*, University of Chicago Press, 247–293.
- McFADDEN, D. (1977): "Modelling the choice of residential location," *Cowles Foundation Discussion Paper*, no. 477.
- (1989): "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica*, 995–1026.
- MELZER, B. T. (2018): "Spillovers from costly credit," *Review of Financial Studies*, 31, 3568–3594.
- MIYAUCHI, Y., K. NAKAJIMA, AND S. J. REDDING (2021): "The economics of spatial mobility: Theory and evidence using smartphone data," Working paper no. 28497. National Bureau of Economic Research, Cambridge, MA.
- MORGAN, D. P., M. L. PINKOVSKIY, AND B. YANG (2016): "Banking deserts, branch closings, and soft information," Working Paper. Federal Reserve Bank of New York. New York, NY.
- NGUYEN, H.-L. Q. (2019): "Are credit markets still local? Evidence from bank branch closings," *American Economic Journal: Applied Economics*, 11, 1–32.
- NGUYEN, J. (2021): "Why is there still a need for physical bank branches?" *Marketplace*.
- NICHOLL, J., J. WEST, S. GOODACRE, AND J. TURNER (2007): "The relationship between distance to hospital and patient mortality in emergencies: An observational study," *Emergency Medicine Journal*, 24, 665–668.
- OFFICE OF THE USPS INSPECTOR GENERAL (2014): "Providing non-bank financial services for the underserved," *USPS Office of the Inspector General, Working Paper*, Washington, DC.
- PEW RESEARCH CENTER (2021): "Internet/Broadband Fact Sheet," Online.
- PNC INSIGHTS (2022): "PNC Solution Centers Reflect Evolution of Branch Banking," .
- PRINA, S. (2015): "Banking the poor via savings accounts: Evidence from a field experiment," *Journal of Development Economics*, 115, 16–31.
- REDDING, S. AND A. J. VENABLES (2004): "Economic geography and international inequality," *Journal of International Economics*, 62, 53–82.
- REDDING, S. J. (2013): "Economic geography: A review of the theoretical and empirical literature," *Palgrave Handbook of International Trade*, 497–531.
- REDDING, S. J. AND E. ROSSI-HANSBERG (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9, 21–58.
- RHINE, S. L. AND W. H. GREENE (2013): "Factors that contribute to becoming unbanked," *Journal of Consumer Affairs*, 47, 27–45.
- ROSENTHAL, S. S. AND W. C. STRANGE (2004): "Evidence on the nature and sources of agglomeration economies," in *Handbook of Regional and Urban Economics*, vol. 4, 2119–2171.
- RUGGLES, S., C. FITCH, D. MAGNUSON, AND J. SCHROEDER (2019): "Differential privacy and census data: Implications for social and economic research," in *AEA Papers and Proceedings*, vol. 109, 403–08.
- SILVA, J. S. AND S. TENREYRO (2006): "The log of gravity," *Review of Economics and Statistics*, 88, 641–658.

- SMALL, M. L., A. AKHAVAN, M. TORRES, AND Q. WANG (2021): "Banks, alternative institutions and the spatial-temporal ecology of racial inequality in US cities," *Nature Human Behaviour*, 5, 1622–1628.
- SQUIRE, R. F. (2019): "Quantifying sampling bias in SafeGraph Patterns," *SafeGraph Blog*.
- SRINIVAS, V. AND R. WADHWANI (2019): "Recognizing the value of bank branches in a digital world: Findings from the global banking survey," *Deloitte Insights*.
- THAENRAJ, P. (2021): "Identifying CBG Sinks," *SafeGraph Blog*.
- WASHINGTON, E. (2006): "The impact of banking and fringe banking regulation on the number of unbanked Americans," *Journal of Human Resources*, 41, 106–137.
- WATHEN, J. (2018): "Believe It or Not, Bank Branches Still Matter," *The Motley Fool*.
- WESTERLUND, J. AND F. WILHELMSSON (2011): "Estimating the gravity model without gravity using panel data," *Applied Economics*, 43, 641–649.
- YANTZI, N., M. W. ROSENBERG, S. O. BURKE, AND M. B. HARRISON (2001): "The impacts of distance to hospital on families with a child with a chronic condition," *Social Science & Medicine*, 52, 1777–1791.
- YOGO, M., A. WHITTEN, AND N. COX (2022): "Financial Inclusion Across the United States," Working paper. Princeton University, Princeton, NJ.



**(B) Block Group  $\times$  Branch Pairs with  $>4$  Visitors, with Fixed Effects**

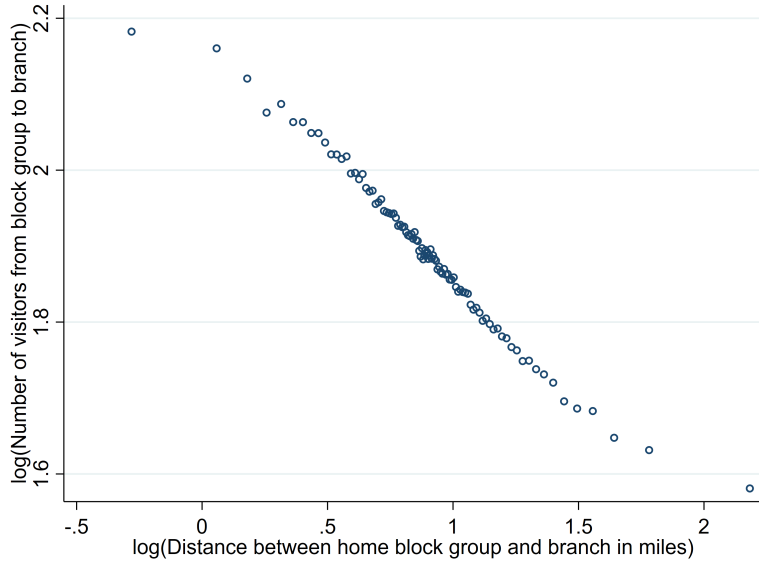


FIGURE 1

NUMBER OF VISITORS FROM BLOCK GROUPS TO BANK BRANCHES BY DISTANCE

The figure presents binned scatter plots of the log number of visitors from home Census 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 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. Distance is computed from the population-weighted center of a block group to a branch. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance (see [Footnote 19](#)). Panel A presents the observed (raw) geolocation data and includes all block group  $\times$  branch pairs, including those with visitor counts of 2 or 3 that SafeGraph rounds up to 4. Panel B only includes block group  $\times$  bank branch pairs with greater than 4 visitors. In that panel, the log numbers of visitors are residualized by block group  $\times$  year-month fixed effects and branch  $\times$  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.

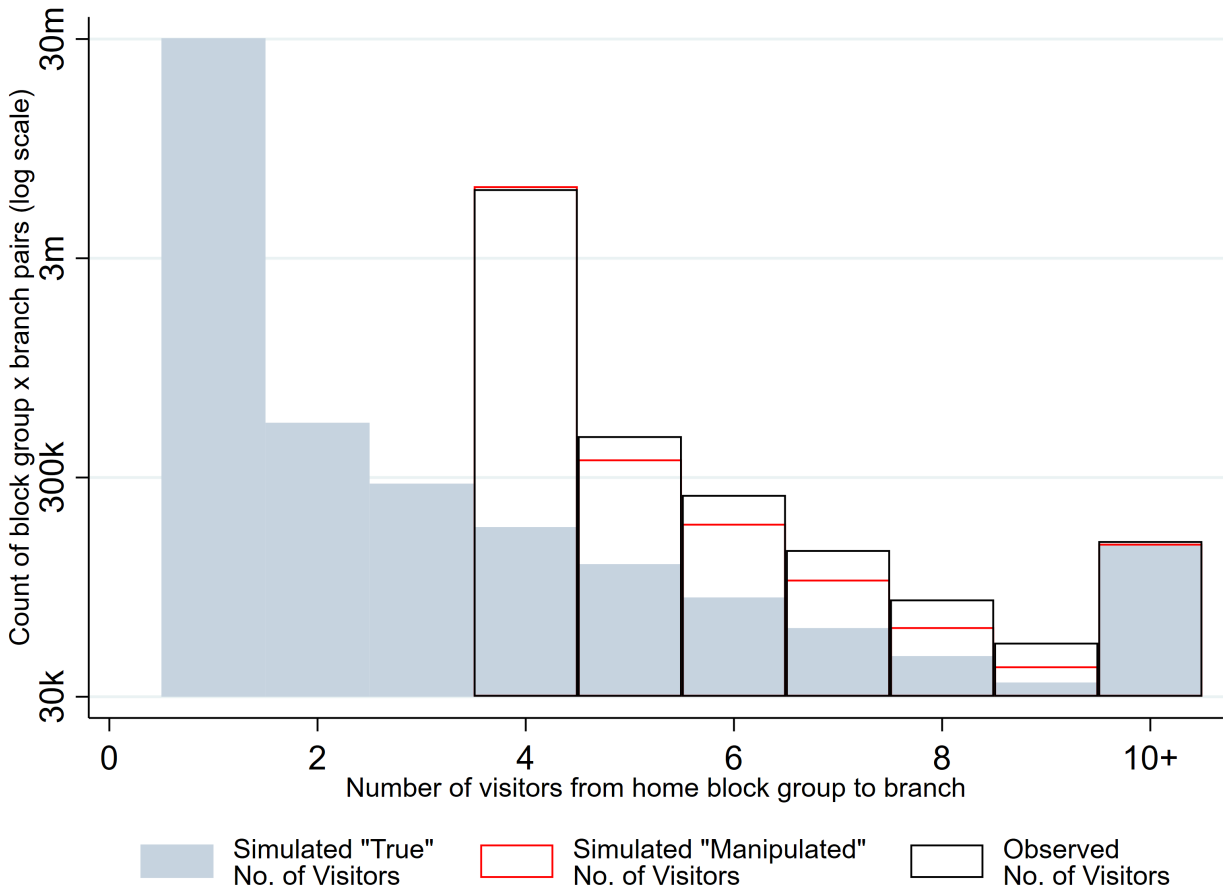


FIGURE 2  
DISTRIBUTIONS OF VISITOR COUNTS

The figure presents distributions of observed visitor counts, simulated “true” visitor counts, and simulated “manipulated” visitor counts from visitors’ home Census block groups to bank branches. Observed visitor counts, denoted  $V_{ijt}$  from Eq. (6), are the raw geolocation 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. Simulated “true” visitor counts, denoted  $V_{ijt}^*$  from Eq. (7), are draws from the underlying “true” distribution of visitors, which we assume to be Poisson. Simulated “manipulated” visitor counts are the “true” visitor counts after being manipulated via differential privacy methods presented in Eqs. (4) to (6). The simulated values are computed from the month-by-month Method of Simulated Moments estimation described in Section 5, with full details of the method in Appendix A. The distribution of simulated visitor counts includes all positive draws from all simulations across every year-month in the sample period. To enhance the depictions of the distributions, we censor them at 10 visitors. That is, the number of block group  $\times$  branch pairs with visitor counts exceeding 10 is assigned to 10+ visitors in the figure.

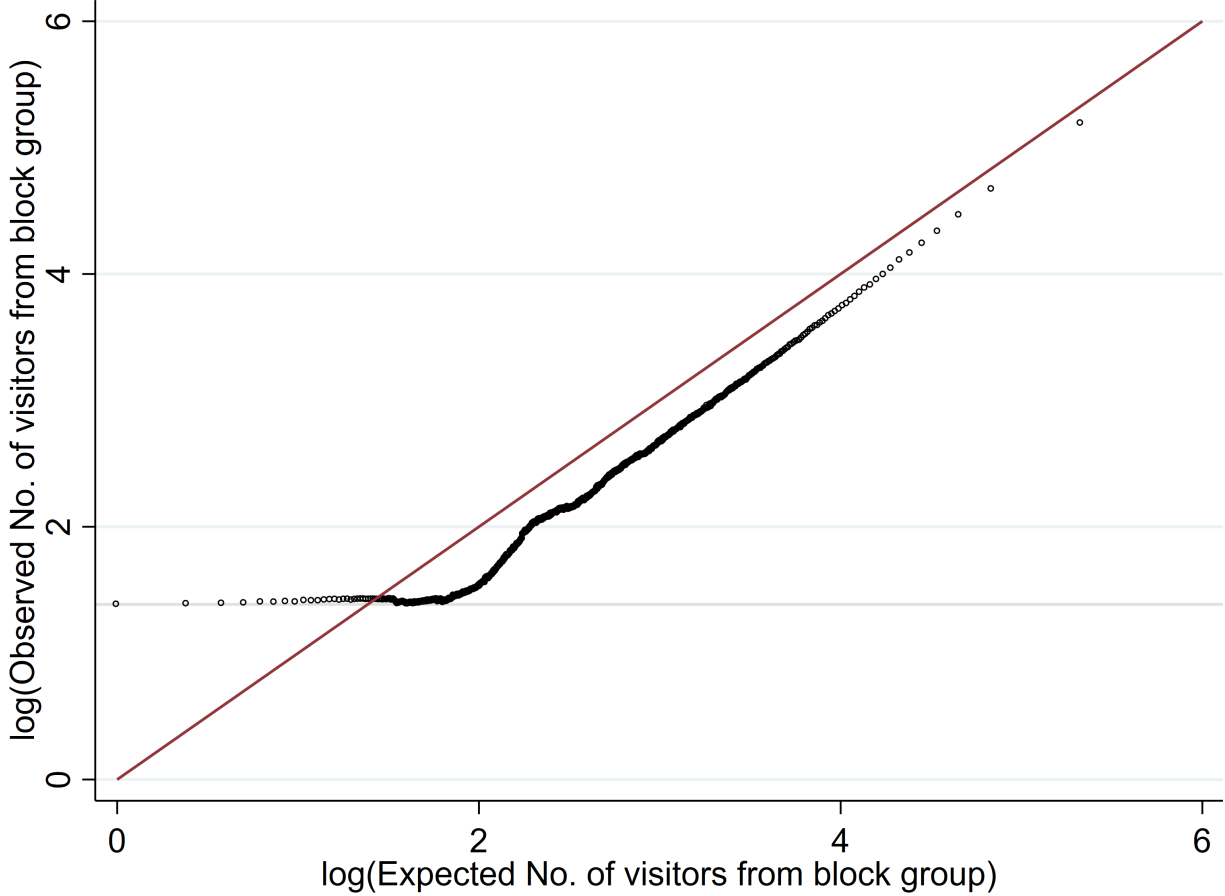
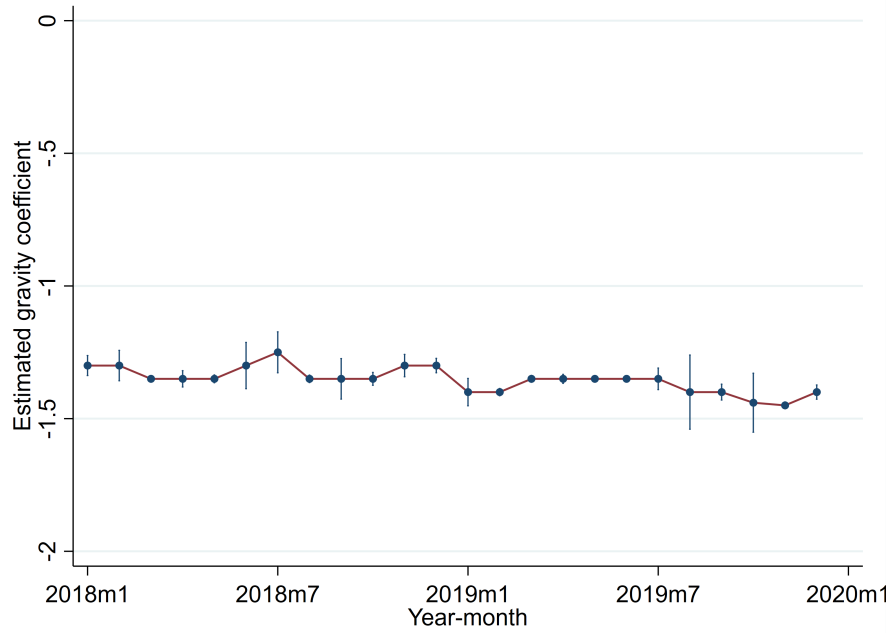


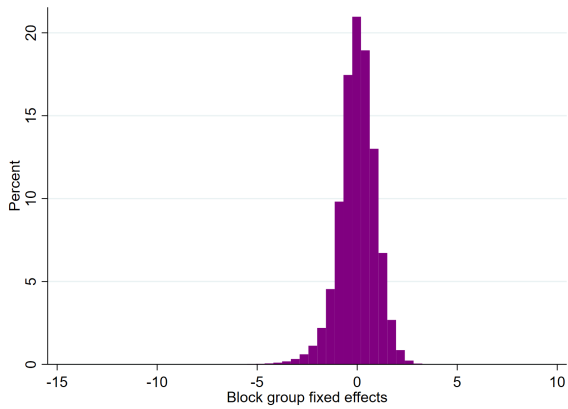
FIGURE 3  
OBSERVED VS. EXPECTED BRANCH VISITORS PER CENSUS BLOCK GROUP

The figure presents a binned scatter plot of the log observed number of branch visitors from each Census block group (i.e.,  $\log V_{it} \equiv \log \sum_j V_{ijt}$ , where  $V_{ijt}$  is given in Eq. (6)) versus the log expected number of branch visitors from each block group based on the month-by-month Method of Simulated Moments (MSM) estimates (i.e.,  $\log \hat{V}_{it}^a \equiv \hat{\gamma}_{it} + \log \hat{\Phi}_{it}^a$ , where the access measure  $\hat{\Phi}_{it}^a \equiv \sum_{j \in B_{it}} \omega_j^i \exp(\hat{\lambda}_{jt}) d_{ij}^{-\hat{\beta}_i}$  reflects the branch probability weights used in the stratified sampling and defined in Eq. (A.30)). 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 Section 5, with full details in Appendix A. To construct the binned scatter plot, we divide the x-axis values into 1,000 equal-sized bins. We then calculate the mean of the y-axis values and the mean of the x-axis values within each bin.

(A) Time Series of  $-\hat{\beta}_{t,MSM}$



(B) Census Block Group Fixed Effects



(C) Bank Branch Fixed Effects

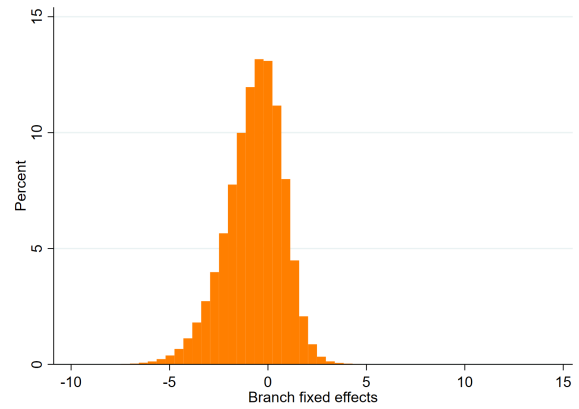


FIGURE 4

METHOD OF SIMULATED MOMENTS PARAMETER ESTIMATES

The figure presents the parameter estimates from the month-by-month Method of Simulated Moments (MSM) estimation of the visitor count gravity relation in Eq. (7). Panel A illustrates the monthly time series of the  $-\hat{\beta}_{t,MSM}$  gravity coefficient estimates, along with 95% confidence intervals. Panel B presents a histogram of the estimated Census block group fixed effects,  $\{\hat{\gamma}_{it}^{\infty}\}$ , and Panel C presents a histogram of the estimated bank branch fixed effects,  $\{\hat{\lambda}_{jt}^{\infty}\}$ . In each histogram, the fixed effects are grouped into 50 equally-sized bins, and the estimated fixed effects for all months in the sample period are presented. A summary of the MSM estimation is provided in Section 5, with full details in Appendix A.



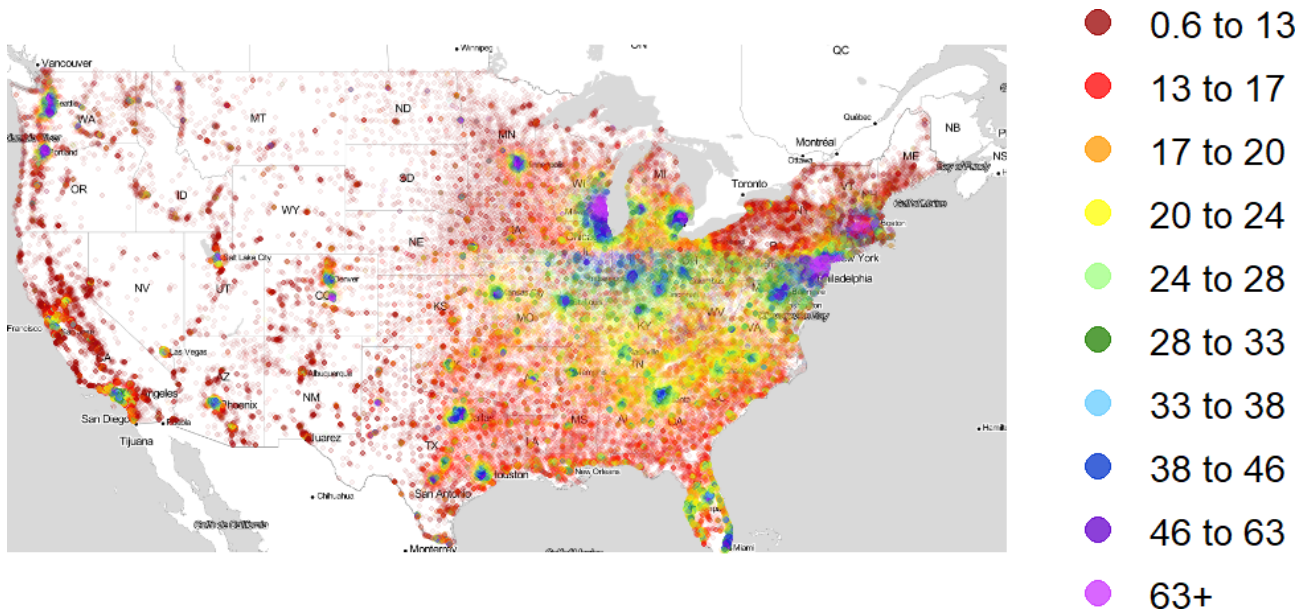
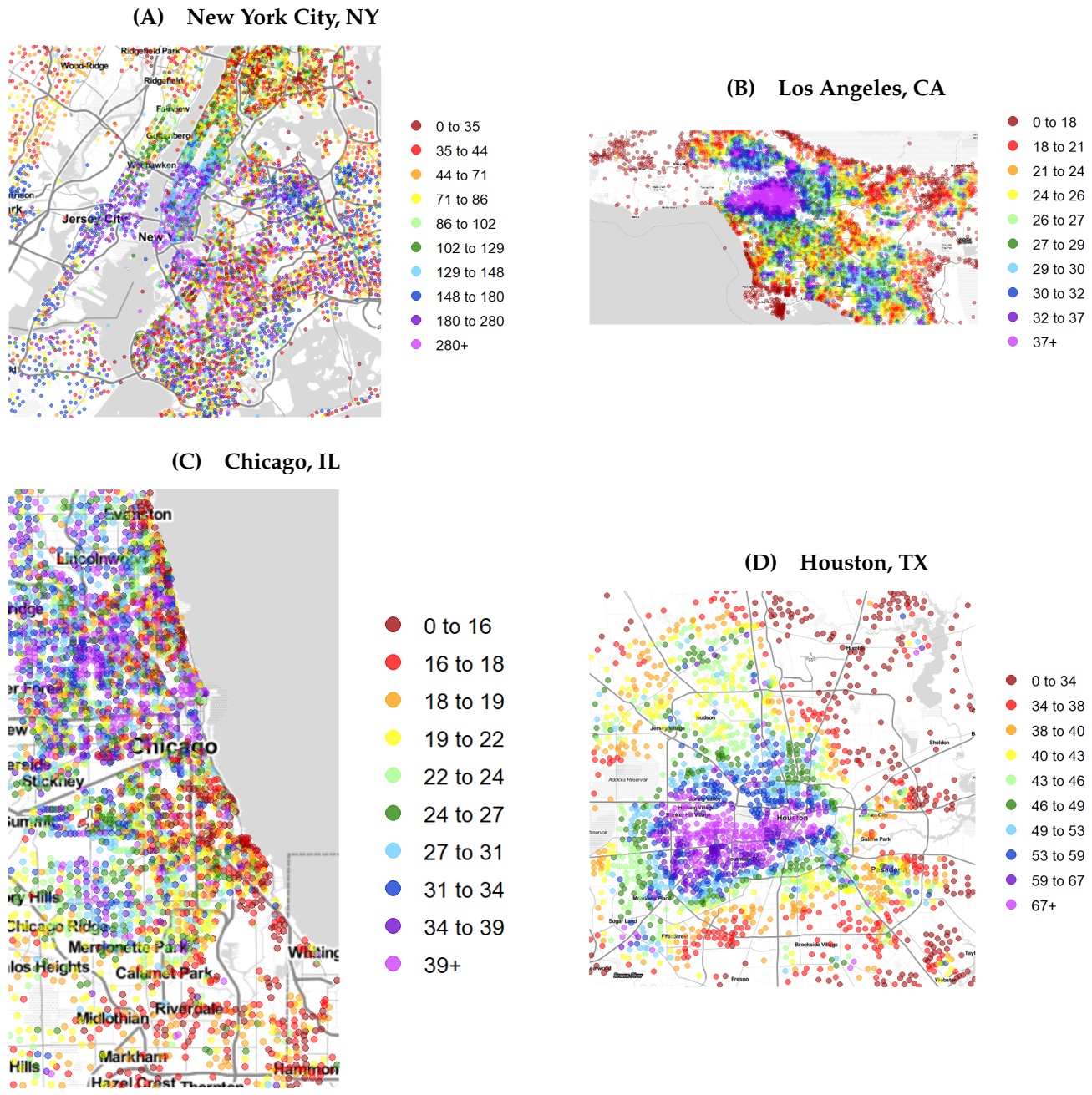


FIGURE 5  
BANK BRANCH ACCESS NATIONWIDE

The figure illustrates a dot density map of bank branch 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. Branch access estimates are calculated from Eq. (8) and are based on the Method of Simulated Moments estimation described in Section 5, with full details in Appendix A. Bank branch 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 observed 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 (indigo and violet) imply higher branch access values and lower-ordered colors (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 are shaded white.



**FIGURE 6**  
**BANK BRANCH ACCESS IN THE FOUR LARGEST U.S. CITIES**

The figure illustrates dot density maps of bank branch access by Census block groups in the four largest U.S. cities by population as of the 2020 Census. 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. Branch access estimates are calculated from Eq. (8) and are based on the Method of Simulated Moments estimation described in Section 5, with full details in Appendix A. Bank branch 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 observed branch visitors over the core sample period (January 2018 - December 2019). Each panel’s map is constructed by grouping block groups within the panel into deciles and shading the dots so that higher-ordered colors in the rainbow gradient (indigo and violet) imply higher access values and lower-ordered colors (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 are shaded white.

TABLE 1  
BANK BRANCH ACCESS BY DEMOGRAPHIC ATTRIBUTES

Dep. var.:	log(Bank branch access of block groups)				log(Branch proximity)		log(Branch quality)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	-0.110 (0.003)	-0.076 (0.003)	-0.126 (0.003)	-0.087 (0.003)	-0.076 (0.003)	-0.086 (0.003)	0.001 (0.001)	-0.001 (0.002)
Black	-0.082 (0.005)	-0.053 (0.005)	-0.107 (0.006)	-0.064 (0.006)	-0.143 (0.006)	-0.168 (0.006)	0.090 (0.003)	0.104 (0.003)
Asian	0.470 (0.014)	0.438 (0.013)	0.429 (0.014)	0.398 (0.013)	0.422 (0.012)	0.375 (0.013)	0.016 (0.007)	0.023 (0.007)
Other	0.023 (0.023)	0.020 (0.023)	0.078 (0.033)	0.073 (0.033)	0.006 (0.022)	0.064 (0.031)	0.015 (0.011)	0.009 (0.015)
Hispanic	0.046 (0.007)	0.081 (0.007)	0.021 (0.007)	0.071 (0.008)	-0.003 (0.007)	-0.030 (0.008)	0.085 (0.004)	0.101 (0.004)
Age <15		-0.721 (0.017)		-0.814 (0.020)	-0.757 (0.017)	-0.855 (0.019)	0.036 (0.010)	0.041 (0.011)
Age 35-54		-0.238 (0.017)		-0.204 (0.020)	-0.253 (0.017)	-0.209 (0.020)	0.015 (0.009)	0.005 (0.011)
Age 55-64		-0.551 (0.019)		-0.573 (0.022)	-0.675 (0.019)	-0.691 (0.023)	0.124 (0.010)	0.117 (0.012)
Age 65+		-0.245 (0.013)		-0.262 (0.015)	-0.256 (0.013)	-0.277 (0.015)	0.010 (0.007)	0.014 (0.008)
log(No. of devices)	-0.050 (0.002)	-0.053 (0.002)	-0.057 (0.002)	-0.059 (0.002)	-0.063 (0.002)	-0.072 (0.002)	0.010 (0.001)	0.012 (0.001)
Observations	2,549,020	2,549,020	1,847,252	1,847,252	2,549,020	1,847,252	2,549,020	1,847,252
Adjusted $R^2$	0.704	0.708	0.634	0.640	0.820	0.774	0.684	0.646
Sample	Core	Core	MC	MC	Core	MC	Core	MC
Year-month FE	○	○	○	○	○	○	○	○
County FE	○	○	○	○	○	○	○	○
RUCA FE	○	○			○		○	

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 bank branch 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. The log number of devices is SafeGraph's record of the number of mobile devices residing in the block group in the year-month. In columns (1)-(4), the dependent variable is the natural logarithm of the estimated bank branch access measure,  $\log \hat{\Phi}_{it}$ , from Eq. (8). In columns (5)-(6), the dependent variable is the natural logarithm of the "branch proximity" component in the decomposition of  $\hat{\Phi}_{it}$  in Eq. (9), whereas in columns (7)-(8), the dependent variable is the natural logarithm of the "average branch quality" component in that decomposition. All dependent variables are computed from the month-by-month Method of Simulated Moments estimation described in Section 5, with full details in Appendix A. 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 and age range 15-34.

TABLE 2  
BANK BRANCH USE AND BLOCK GROUP FIXED EFFECTS BY DEMOGRAPHIC ATTRIBUTES

Dep. var.:	log(Expected no. of visitors)				Block group fixed effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	0.186 (0.004)	0.155 (0.004)	0.190 (0.004)	0.155 (0.004)	0.296 (0.004)	0.230 (0.004)	0.315 (0.005)	0.242 (0.005)
Black	-0.041 (0.009)	-0.056 (0.009)	-0.017 (0.010)	-0.039 (0.010)	0.041 (0.010)	-0.003 (0.010)	0.089 (0.011)	0.025 (0.011)
Asian	0.238 (0.020)	0.241 (0.020)	0.227 (0.020)	0.224 (0.019)	-0.232 (0.023)	-0.197 (0.022)	-0.202 (0.023)	-0.174 (0.022)
Other	0.034 (0.028)	0.041 (0.029)	0.009 (0.038)	0.016 (0.039)	0.010 (0.033)	0.020 (0.033)	-0.068 (0.046)	-0.058 (0.046)
Hispanic	0.008 (0.009)	-0.016 (0.010)	0.021 (0.010)	-0.012 (0.011)	-0.038 (0.010)	-0.097 (0.011)	-0.000 (0.011)	-0.083 (0.012)
Age <15		0.341 (0.022)		0.343 (0.025)		1.061 (0.026)		1.157 (0.030)
Age 35-54		0.486 (0.023)		0.529 (0.026)		0.724 (0.025)		0.732 (0.029)
Age 55-64		0.101 (0.025)		0.053 (0.030)		0.651 (0.029)		0.626 (0.034)
Age 65+		0.248 (0.018)		0.253 (0.020)		0.494 (0.021)		0.515 (0.023)
log(No. of devices)	0.606 (0.004)	0.606 (0.004)	0.601 (0.005)	0.601 (0.005)	0.656 (0.005)	0.659 (0.005)	0.658 (0.005)	0.660 (0.005)
Observations	2,549,020	2,549,020	1,847,252	1,847,252	2,549,020	2,549,020	1,847,252	1,847,252
Adjusted R <sup>2</sup>	0.380	0.381	0.339	0.340	0.310	0.314	0.309	0.313
Sample	Core	Core	MC	MC	Core	Core	MC	MC
Year-month FE	○	○	○	○	○	○	○	○
County FE	○	○	○	○	○	○	○	○
RUCA FE	○	○			○	○		

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 bank branch 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 log number of devices is SafeGraph's record of the number of mobile devices residing in the block group in the year-month. The dependent variable in columns (1)-(4) is the natural logarithm of the expected number of branch goes 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. (7). The dependent variable in columns (5)-(8) is the estimated block group fixed effects,  $\{\hat{\gamma}_{it}\}$ , from the gravity relation in Eq. (7). The estimation method is described in Section 5, with full details in Appendix A. Columns (1), (2), (5), and (6) include all block groups for which we have branch visitor data, whereas columns (3), (4), (7), and (8) restrict the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitan area core). The omitted demographic groups are non-Hispanic Whites and age range 15-34.

# Appendix

## A Econometric Method

Here, we layout the steps of the econometric method we use to estimate the parameters of the fixed effects gravity model of Eq. (1) using the privacy-protected geolocation data from mobile devices. The approach adapts the Method of Simulated Moments (MSM) to handle a large volume of fixed effects. A key insight of the approach is to simulate data from the gravity model and then apply the same differential privacy algorithm to the simulated data that the data provider used to privacy-protect the geolocation data. We run the method separately per year-month of our sample period (January 2018 - December 2019).

### A.1 Specify the DGP for visitors

The data generating process (DGP) we simulate is the number of visitors from block groups to branches through time. We assume that the true number of visitors from block group  $i$  to branch  $j$  in year-month  $t$ , denoted  $V_{ijt}^*$ , is Poisson distributed. Using the gravity model from Eq. (1), we express the true visitor count as obeying

$$V_{ijt}^* \sim \text{Pois}\left(\exp\left(\gamma_{it} + \lambda_{jt} - \beta_t \log \text{Distance}_{ij}\right)\right). \quad (\text{A.15})$$

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.<sup>19</sup>

To account for the differential privacy algorithm in the simulation, we 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:

$$V_{ijt}^+ = V_{ijt}^* + L_{ijt}. \quad (\text{A.16})$$

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  $\times$  branch visitor count is present in the sample. The selection equation is

$$z_{ijt} = \begin{cases} 1 & \text{if } \lfloor V_{ijt}^+ \rfloor \geq 2, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.17})$$

Let  $V_{ijt}$  denote the visitor count observed in the sample, subject to SafeGraph's censoring. The observation equation is

$$V_{ijt} = \max\{4, \lfloor V_{ijt}^+ \rfloor\}, \quad (\text{A.18})$$

In the simulation, we implement Eqs. (A.15) to (A.18).

<sup>19</sup> 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  $h = \text{hav}(lat_1 - lat_2) + \cos(lat_1) \cos(lat_2) \text{hav}(long_2 - long_1)$ . 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 travel is the road driving time between locations. Even so, the haversine formula is simple, fairly accurate, and convenient to compute. In Online Appendix Table A.7, we regress the driving times between about 1 million random block groups and bank branches onto the corresponding haversine distances. Driving times are computed using the Origin-Destination Cost Matrix of ArcGIS Pro under the default settings. Regressions are run across the entire 1 million sample and over parts of the sample associated with various demographic attributes, such as including only block groups with Black population shares exceeding 80% from the 2019 5-yr. American Community Survey. Across the samples, the regressions produce very high  $R^2$ , ranging from 0.972 to 0.993. Haversine distance is computationally easier to calculate, and these regression results suggest that it correlates highly with driving time.

## A.2 Sample block group $\times$ branch pairs

Technically speaking, every possible block group  $i$  and branch  $j$  pair should enter Eq. (A.15). But our data of over fifty-thousand branches, over two-hundred-thousand block groups, altogether spanning twenty-four months, makes it computationally impractical to have the billions of possible block group  $\times$  branch pairs enter the MSM estimation. Instead, we sample pairs using stratified sampling.

In each year-month, block group  $\times$  branch pairs in the SafeGraph data register either positive (and  $\geq 4$ ) or missing observed visitor counts. If a block group  $\times$  branch pair has a positive visitor count, then we know that residents of the block group visited the branch in the period, and we sample this block group  $\times$  branch pair in our simulation with probability 1. If a block group  $\times$  branch pair has a missing visitor count in the year-month, then either residents of the block group did not visit the branch in the period, or the visitor count was left out of the data from SafeGraph’s differential privacy methods. In each year-month, we sample from this *alternative* set of missing block group  $\times$  branch pairs such that (i) every pair in the alternative set has the same probability of being sampled, and (ii) each block group and each branch is part of at least one block group  $\times$  branch pair in the alternative set. The second condition ensures that each block group and branch is represented in the stratified sampling. We set the sampling probability to  $1/2000$ , which implies that, on average, the randomly sampled alternative set of block group  $\times$  branch pairs represents slightly higher than a 0.05% sample size of all possible block group  $\times$  branch pairs with missing visitor counts.

To establish notation for the stratified sampling of block group  $\times$  branch pairs, we let  $n_t$  denote the stratified sample of block group  $\times$  branch pairs in year-month  $t$ . This set is the union of the set of pairs with positive observed visitor counts that are sampled with probability 1, denoted  $n_t^1$ , and the alternative set of pairs with missing observed visitor counts that are sampled with probability  $1/2000$ , denoted  $n_t^0$ . Let  $N_{it}^0$  denote the *population* of the block group  $\times$  branch pairs with missing observed visitor counts that are associated with block group  $i$ . (We use lowercase notation for the stratified sample of pairs and uppercase notation for the population of pairs.)

We implement the stratified sampling in the following manner to satisfy conditions (i) and (ii) above. To satisfy (ii), we pick 1 pair randomly from  $N_{it}^0$  for each block group  $i$ . Notice that we could have chosen more than one pair per block group  $i$  to satisfy (ii), but choosing just one reduces the estimation time. Next, to satisfy (i), we draw a uniform random variable  $u_{ijt} \sim U[0, 1]$  for each pair in  $N_{it}^0$ . We include the pair in the sample if  $u_{ijt} \leq \frac{p - \frac{m}{|N_{it}^0|}}{1 - \frac{m}{|N_{it}^0|}}$ , where  $p = 1 - \sqrt{1 - 1/M}$ , and  $1/M$  is our target sampling probability of  $1/2000$ , and  $|\cdot|$  is cardinality of a set. We loop this procedure through each block group  $i$ . We then repeat the process for each branch  $j$  (i.e., draw a uniform random variable for each pair again in  $N_{it}^0$ , but looping through all branches).

Notice that we rely on the uniform random variable draw falling short of a threshold to determine whether a block group  $\times$  branch pair is sampled because the number of pairs in each set  $N_{it}^0$  is discrete, but we want the sampling probability to be the same across all block group  $\times$  branch pairs with missing visitor counts. The probability of a pair being sampled when looping through block groups is the union of the initial 1 random pair choice satisfying condition (ii) and the threshold condition on the uniform draw. That probability is the following:

$$\frac{1}{|N_{it}^0|} + \frac{p - \frac{1}{|N_{it}^0|}}{1 - \frac{1}{|N_{it}^0|}} - \frac{1}{|N_{it}^0|} \frac{p - \frac{1}{|N_{it}^0|}}{1 - \frac{1}{|N_{it}^0|}}, \quad (\text{A.19})$$

which is simply the probability of the union of the two independent events, where we have used the relation  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  for independent events  $A$  and  $B$ . Some algebra reveals that Eq. (A.19) equals  $p$ . Because we repeat the process across all branches, the probability of a block group  $\times$  branch pair being sampled either from the loop through block groups or the loop through branches is

$$p + p - p^2 = \frac{1}{M},$$

which matches our target sampling probability, as desired.

The stratified sampling requires that we apply probability weights to any variable measured at the block group  $\times$  branch level, such as visitor counts or pairwise distances, so as to rebalance the data and make it represent the target population as closely as possible. We assign probability weights equaling 1 to the sampled pairs in the set  $n_t^1$  because these pairs were sampled with probability 1. We assign probability weights denoted  $\omega_t$  to the sampled pairs in the set  $n_t^0$ . These probability weights satisfy:

$$\omega_t |n_t^0| + 1 |n_t^1| = \text{Total no. of block groups in year-month } t \times \text{Total number of branches in year-month } t. \quad (\text{A.20})$$

Rearranging Eq. (A.20) shows that the probability weight  $\omega_t$  per year-month is the number of *population* pairs with missing observed visitor counts divided by the number of *sampled* pairs with missing observed visitor counts. Following standard practice, we have the probability weights equal the reciprocal of the likelihood of being sampled ( $M = 2000$ ), but they can deviate slightly from  $M$  by chance because of the random sampling.

### A.3 Initialize the fixed effects routine

In each year-month of the sample period, the MSM uses the visitor data  $v$  and the model parameters  $\psi \equiv \{\beta_t, \gamma_{it}, \lambda_{jt}\}$  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. (A.15) requires hundreds of thousands of fixed effects to be estimated. Estimating all these parameters from the MSM minimization problem alone would be computationally impractical. Instead, we adopt an iterative routine to identify the fixed effects  $\{\gamma_{it}, \lambda_{jt}\}$  and let the minimization problem identify  $\beta_t$ . Holding fixed an estimate of  $\beta_t$  and given initial estimates of the fixed effects, the routine updates the fixed effects estimates until they converge. After the fixed effects converge per estimate of  $\beta_t$  in the year-month, the MSM minimization problem then chooses the optimal  $\beta_t$  estimate that satisfies the moment conditions in the year-month. We initialize the fixed effects routine with guessed estimates  $\hat{\gamma}_{it}^0 = \hat{\lambda}_{jt}^0 = 1$  for all  $i$  and  $j$  and  $t$ .

### A.4 Simulate visitor counts

We run  $S = 10$  simulations of the visitor counts per block group  $\times$  branch pair. The  $S$  simulations are run per year-month of the sample. We differentially simulate visitor counts from the two sets of sampled block group  $\times$  pairs,  $n_t^0$  and  $n_t^1$ , because of their different probability weights.

Consider first the set  $n_t^1$  of pairs with positive observed visitor counts that were sampled with probability 1. Per year-month, we begin the simulation by drawing  $|n_t^1| \times S$  Laplace random variables having mean zero and scale  $10/9$ , and we draw  $|n_t^1| \times S$  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. Given an estimate  $\hat{\beta}_t$  of the gravity coefficient and the initial guessed estimates  $\{\hat{\gamma}_{it}^0, \hat{\lambda}_{jt}^0\}$  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. (A.15).

Each Poisson draw is a "true" block group  $\times$  branch visitor count. 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  $\times$  branch visitor count, (ii) round each noisy visitor count down to the nearest integer, (iii) set to 0 all noisy visitor counts below 2, and (iv) replace all noisy visitor counts that equal 2 or 3 with 4 (see Eqs. (A.15) to (A.18)). 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  $\tilde{v} = \{\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_S\}$  be the  $S$  simulated visitor counts in year-month  $t$ , where we have excluded a  $t$  subscript to simplify notation.

Consider next the set  $n_t^0$  of block group  $\times$  branch pairs with missing visitor counts that were sampled with probability  $1/2000$ . If an extra  $|n_t^0| \times S$  pairs of visitor counts were simulated in the same manner described in the previous two paragraphs, those simulated visitor counts would have disproportionate impact on any computed moments because of the high probability weights that would multiply them. Noise from the simulation would be amplified and make the estimation unstable. Rather than simulating visitor counts for the block group  $\times$  branch pairs in  $n_t^0$ , we construct their implied empirical probability distribution according to the parameter estimate of  $\psi$  in each iteration. If an infinite number of visitor counts from the pairs in  $n_t^0$  were in fact simulated, their distribution would coincide with this constructed empirical distribution. Notice that we cannot apply this approach to the set  $n_t^1$  of sampled block group  $\times$  branch pairs because each pair in that set is drawn from a distinct distribution, due, in part, to the block group- and branch-specific fixed effects. For those pairs, we simulate draws. However, the sampled pairs in the set  $n_t^0$  are meant to represent the remaining block group  $\times$  branch pairs in the population with missing observed visitor counts, which are very high in number. One stratified sampled observation is meant to represent 2,000 observations from the same distribution. We construct the empirical distribution that these sampled pairs represent.

That empirical distribution is a truncated and censored version of a Poisson probability mass function per Eq. (A.15) plus a Laplace distribution with mean 0 and scale  $10/9$ . Because the Laplace noise is added after the Poisson random variable is drawn, this empirical distribution can be thought of as a truncated and censored Laplace distribution whose parameter location (mean) is the realization of the Poisson draw. With this in mind, let  $G(y, k)$  be the CDF of a Laplace

distribution with mean  $k$  and scale  $10/9$ . And let  $\hat{\mu}_{ijt}$  denote the estimated mean of the Poisson visitor count in Eq. (A.15). Namely,

$$\hat{\mu}_{ijt} \equiv \exp\left(\hat{\gamma}_{it} + \hat{\lambda}_{jt} - \hat{\beta}_t \log \text{Distance}_{ij}\right). \quad (\text{A.21})$$

Finally, let the probability that the Poisson distribution draws a visitor count of  $k$ , given its estimated mean  $\hat{\mu}_{ijt}$  be denoted  $p(k, \hat{\mu}_{ijt})$ . Notice that the parameters of the empirical distribution update with every iteration of the estimated fixed effects and guess of  $\beta_t$ .

We construct 7 components of the empirical distribution that we use in the moments of the estimation. Because both the Laplace and Poisson distributions have infinite support, we must insert an upper bound to both supports when constructing the empirical distribution. We bound the Poisson support at  $K = 20$  and the Laplace support at  $L = 30$ . The upper bounds imply that the 7 components of the empirical distribution hold approximately. As  $K \rightarrow \infty$  and  $L \rightarrow \infty$ , they would hold exactly. The 7 components of the empirical distribution we compute are:

**1. Probability that the visitor count equals 0:**

$$\Pr\left(\tilde{V}_{ijt} = 0 | \hat{\mu}_{ijt}\right) \approx p(0, \hat{\mu}_{ijt}) + \sum_{k=1}^K p(k, \hat{\mu}_{ijt}) \times G(2, k). \quad (\text{A.22})$$

The probability that a simulated visitor count is zero equals the probability that the Poisson draw equals zero, represented by the first term in Eq. (A.22), plus the cumulative probability that the Poisson draw has a positive value but the Laplace draw reduces that positive value to the lower bound of 0. That cumulative probability is represented by the second term in Eq. (A.22). In that term, the Laplace draw has mean  $k$  to adjust for different possible positive draws of the Poisson. Moreover, the CDF value of the Laplace distribution given that mean,  $G(2, k)$ , is positioned at 2 because SafeGraph truncates any visitor count below 2. Thus, the second term is the cumulative probability that the simulated visitor count falls below 2 after the Laplace noise is added to a positive Poisson draw. The Laplace probability multiplies the Poisson probability because the two draws are independent. Notice that no Laplace piece enters the first term because SafeGraph adds Laplace noise only to positive observed visitor counts.

**2. Probability that the visitor count exceeds 0:**

$$\Pr\left(\tilde{V}_{ijt} > 0 | \hat{\mu}_{ijt}\right) \approx \sum_{k=1}^K p(k, \hat{\mu}_{ijt}) \times (1 - G(2, k)). \quad (\text{A.23})$$

This probability is simply the complement of the previous one. Because the visitor count exceeds 0 in this scenario, Laplace noise is always added to the Poisson draw, and hence, the ‘‘survival function’’ of the Laplace, given by  $1 - G(2, k)$ , multiplies each Poisson probability. The survival value is the probability that the visitor count avoids truncation.

**3. Probability that the visitor count equals 4:**

$$\Pr\left(\tilde{V}_{ijt} = 4 | \hat{\mu}_{ijt}\right) \approx \sum_{k=1}^K p(k, \hat{\mu}_{ijt}) \times (G(5, k) - G(2, k)). \quad (\text{A.24})$$

The probability that the visitor count equals 4 is the probability that the Poisson draw lands at or above 1 visitor count times the probability that the Laplace draw pushes the visitor count to a value in the interval between 2 and 4 inclusive (i.e., the censoring region). Because SafeGraph rounds visitor counts down to the nearest integer, the probability that the Laplace draw carries the visitor count into the censored region is  $G(5, k) - G(2, k)$ . For example, a Poisson draw plus a Laplace draw that equaled  $4.\overline{9}$  would round down to 4.

**4. Probability that the visitor count exceeds 4:**

$$\Pr\left(\tilde{V}_{ijt} > 4 | \hat{\mu}_{ijt}\right) \approx \sum_{k=1}^K p(k, \hat{\mu}_{ijt}) \times (1 - G(5, k)). \quad (\text{A.25})$$

This probability is simply the complement of the previous one. The survival function of the Laplace above 4, given by  $1 - G(5, k)$ , multiplies each Poisson probability. The survival value is the probability that the visitor count avoids censoring.



### 5. Expected visitor count:

$$\mathbb{E}(\tilde{V}_{ijt}|\hat{\mu}_{ijt}) \approx \sum_{k=1}^K p(k, \hat{\mu}_{ijt}) \left[ 4 \times \{G(5, k) - G(2, k)\} + \sum_{l=5}^L l \times \{G(l+1, k) - G(l, k)\} \right]. \quad (\text{A.26})$$

The formula for the mean visitor count is broken up into two parts. Both parts are multiplied by the probability,  $p(k, \hat{\mu}_{ijt})$ , that the Poisson draw lands at or above 1 visitor count so that the observation enters the support of the empirical distribution. The first part is the probability that the Laplace draw pushes the visitor count to a value in the interval between 2 and 4 inclusive (the censoring region) multiplied by 4 visitors. The second part is the probability that the Laplace draw pushes the visitor count to a value of 5 or higher, multiplied by that value. Because SafeGraph rounds visitor counts down to the nearest integer, the probability of each value in this second part is the CDF of the Laplace distribution at 1 above that value less the CDF at the value, given by  $G(l+1, k) - G(l, k)$ .

### 6. Expected log visitor count, conditional on the visitor count exceeding 0:

$$\mathbb{E}(\log \tilde{V}_{ijt} | \tilde{V}_{ijt} > 0, \hat{\mu}_{ijt}) \approx \frac{\sum_{k=1}^K p(k, \hat{\mu}_{ijt}) \left[ \log 4 \times \{G(5, k) - G(2, k)\} + \sum_{l=5}^L \{\log l \times (G(l+1, k) - G(l, k))\} \right]}{\Pr(\tilde{V}_{ijt} > 0 | \hat{\mu}_{ijt})}. \quad (\text{A.27})$$

The formula for the mean of the natural logarithm of the visitor count is very similar to that of the mean of the visitor count from Eq. (A.26). The only adjustments are that the natural logarithm is taken as needed and that the mean is re-weighted to account for the positive visitor count requirement. That re-weighting is exhibited via the division by  $\Pr(\tilde{V}_{ijt} > 0 | \hat{\mu}_{ijt})$ , defined in Eq. (A.23), which is the way to compute the mean of a truncated random variable.

### 7. Expected log visitor count, conditional on the visitor count exceeding 4:

$$\mathbb{E}(\log \tilde{V}_{ijt} | \tilde{V}_{ijt} > 4, \hat{\mu}_{ijt}) \approx \frac{\sum_{k=1}^K p(k, \hat{\mu}_{ijt}) \left[ \sum_{l=5}^L \{\log l \times (G(l+1, k) - G(l, k))\} \right]}{\Pr(\tilde{V}_{ijt} > 4 | \hat{\mu}_{ijt})}. \quad (\text{A.28})$$

This conditional mean is even simpler to compute than the one in Eq. (A.27). The formula consists of just the second component in the numerator of Eq. (A.27), and the re-weighting in the denominator is the probability of the visitor count exceeding 4, given in Eq. (A.25).

## A.5 Iterate the fixed effects until convergence

Under a fixed estimate  $\hat{\beta}_t$ , the next step is to iterate the estimated fixed effects until they converge. Because the fixed effects are measured at the block group or branch level, and not the block group  $\times$  branch level like the visitor counts, we need two other sets of probability weights for the fixed effects estimation due to the stratified sampling. The block group and branch weights are defined similarly as the block group  $\times$  branch weights in Eq. (A.20), but they are measured from the perspective of a block group or branch.

Notice that the stratified sample of block group  $\times$  branch pairs also creates a stratified sample of block groups and branches *separately*. With this in mind, we let  $b_{it}$  denote the stratified sample of branches for block group  $i$  in year-month  $t$ . This set is the union of the set of branches from the pairs sampled with probability 1, denoted  $b_{it}^1$ , and the set of branches from the pairs sampled with probability  $1/2000$ , denoted  $b_{it}^0$ . Likewise, let  $h_{jt}$  denote the stratified sample of home block groups for branch  $j$  in year-month  $t$ . This set is the union of the set of block groups from the pairs sampled with probability 1, denoted  $h_{jt}^1$ , and the set of block groups from the pairs sampled with probability  $1/2000$ , denoted  $h_{jt}^0$ . The block groups in  $h_{jt}^1$  and branches in  $b_{it}^1$  have probability weights equal to 1. The block groups in  $h_{jt}^0$  have probability weights denoted  $\omega_{jt}^j$ , and the branches in  $b_{it}^0$  have probability weights denoted  $\omega_{it}^i$ . These probability weights are defined as:

$$\omega_{it}^i | h_{jt}^0 | + 1 | h_{jt}^1 | = \text{Total no. of block groups in year-month } t, \quad \forall (i, j) \in n_t, \quad (\text{A.29})$$

$$\omega_{it}^i | b_{it}^0 | + 1 | b_{it}^1 | = \text{Total no. of branches in year-month } t, \quad \forall (i, j) \in n_t. \quad (\text{A.30})$$

We use the block group- and branch-specific probability weights from Eqs. (A.29) to (A.30) only in the fixed effects iteration routine. We iterate the estimated fixed effects sequentially. We begin with the estimated branch fixed effects  $\{\hat{\lambda}_{jt}\}$ , while holding constant the estimated block group fixed effects  $\{\hat{\gamma}_{it}^0 = 1, \forall i \text{ and } \forall t\}$ .

To estimate the branch fixed effects, we take advantage of another data field in SafeGraph: a branch's total number of visitors. The SafeGraph name for this field is RAW\_VISITOR\_COUNT. Unlike the number of visitors from a block group to the branch, a branch's total number of visitors is unaffected by SafeGraph's differential privacy methods. Because we presume that block group residents can visit any branch cross-country in the year-month, we can take advantage of a branch's total visitors to uniquely pin down the estimate of the branch's fixed effect. Let  $V_{jt}^T$  denote branch  $j$ 's total visitors in year-month  $t$ .

The iteration process for estimating the branch fixed effects is as follows. Suppose we are on the  $k$ -th iteration. From Eq. (A.15), the expected number of visitors to branch  $j$  from block group  $i$  in year-month  $t$  based on the  $k$ -th iteration estimates of the fixed effects is

$$\hat{V}_{ijt}^k = \exp(\hat{\lambda}_{jt}^k) \exp(\hat{\gamma}_{it}^k) d_{ij}^{-\hat{\beta}_t}. \quad (\text{A.31})$$

Summing across block groups, and adjusting for the probability weights defined in Eq. (A.29), we obtain a branch's expected total visitor count:

$$\hat{V}_{jt}^k = \exp(\hat{\lambda}_{jt}^k) \sum_{i \in h_{jt}} \omega_t^i \exp(\hat{\gamma}_{it}^k) d_{ij}^{-\hat{\beta}_t}. \quad (\text{A.32})$$

Given  $\hat{\beta}_t$  and the  $k$ -th iteration of the estimated block group fixed effects,  $\{\hat{\gamma}_{it}^k\}$ , we determine the  $k$ -th iteration of each branch's estimated fixed effect,  $\hat{\lambda}_{jt}^k$ , by solving for the value that equates the branch's expected total visitor count,  $\hat{V}_{jt}^k$  from Eq. (A.32), with the branch's observed total visitor count,  $V_{jt}^T$ . Mathematically speaking, the branch's fixed effect estimate satisfies:

$$\hat{\lambda}_{jt}^k = \log V_{jt}^T - \log \sum_{i \in h_{jt}} \omega_t^i \exp(\hat{\gamma}_{it}^k) d_{ij}^{-\hat{\beta}_t}. \quad (\text{A.33})$$

Per iteration, Eq. (A.33) pins down each branch's estimated fixed effect as a function of the estimated block group fixed effects (and the estimate of  $\beta_t$ ). The estimated block group fixed effects will iterate until they converge, and by Eq. (A.33), once the estimated block group fixed effects converge, so too do the estimated branch fixed effects, given an estimate of  $\beta_t$ .

The iteration process for estimating the block group fixed effects is as follows. Suppose we are on the  $k$ -th iteration. For each block group  $i$  in the year-month, we divide the average observed visitor counts  $V_{ijt}$  across the branches in set  $b_{it}$ , by the average simulated visitor counts across all branches in set  $b_{it}$  and all simulations  $S$ . With this in mind, we let the average observed visitor count of block group  $i$  be

$$\bar{V}_{it} = \frac{1}{|b_{it}|} \sum_{j \in b_{it}} V_{ijt}. \quad (\text{A.34})$$

Let the simulated visitor counts from simulation  $s$  in iteration  $k$  be denoted  $\tilde{V}_{ijt}^k(s)$ . The average simulated visitor count of block group  $i$  in simulation  $s$  is

$$\bar{\tilde{V}}_{it}^k(s) = \frac{\sum_{j \in b_{it}^1} \tilde{V}_{ijt}^k(s) + \sum_{j \in b_{it}^0} \omega_t^j \mathbb{E}(\tilde{V}_{ijt}^k | \hat{\mu}_{ijt})}{\sum_{j \in b_{it}^1} 1 + \sum_{j \in b_{it}^0} \omega_t^j}, \quad (\text{A.35})$$

where  $\mathbb{E}(\tilde{V}_{ijt}^k | \hat{\mu}_{ijt})$  is provided in Eq. (A.26). Because the calculation is at the block-group level, the probability weights we use are from the block-group perspective, and they either equal 1 or satisfy Eq. (A.30). Averaging across simulations delivers the mean simulated visitor count of block group  $i$  as

$$\bar{\tilde{V}}_{it}^k = \frac{1}{S} \sum_s \bar{\tilde{V}}_{it}^k(s). \quad (\text{A.36})$$

The ratio of block group  $i$ 's average observed visitor count to average simulated visitor count is thus:

$$\chi_{it}^k = \frac{\bar{V}_{it}}{\bar{\tilde{V}}_{it}^k} \quad (\text{A.37})$$

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

$$\hat{\gamma}_{it}^{k+1} = \hat{\gamma}_{it}^k \times (\lambda_{it}^k)^g, \quad (\text{A.38})$$

where  $g$  is a modifying term to avoid oscillating estimates, and we set its value to 0.5. Notice that if block group  $i$ ’s average simulated visitor count is higher than its average observed visitor count in the data, then  $\lambda_{it}^k < 1$ , and the block group’s estimated fixed effect is revised downward.

After each update of the estimated block group fixed effects, we re-transform the  $|n_t^1| \times S$  Uniform random variables into Poisson random variables using (i) the estimate  $\hat{\beta}_t$ ; (ii) the updated block group fixed effect estimates,  $\{\hat{\gamma}_{it}^{k+1}\}$ ; and (iii) the updated branch fixed effect estimates,  $\{\hat{\lambda}_{jt}^{k+1}\}$ , based on Eq. (A.33). We then apply differential privacy methods to the “updated” simulated data. The process iterates until the estimated block group fixed effects converge.<sup>20</sup>

While the estimated fixed effects are updated using *ratios* of the averages between observed and simulated values, we found that the estimates converged faster under a convergence criterion that uses *differences* in the averages instead. We define convergence as the squared change between iterations in the mean squared difference between average observed and simulated visitor counts of a block group 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 of each block group  $i$ , 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 block groups with branch goes to more branches (higher  $|b_{it}|$ ). Mathematically, the convergence condition is

$$\left[ \frac{1}{|n_t|} \sum_i |b_{it}| (\bar{V}_{it}^{k+1} - \bar{V}_{it})^2 - \frac{1}{|n_t|} \sum_i |b_{it}| (\bar{V}_{it}^k - \bar{V}_{it})^2 \right]^2 < \varepsilon \quad (\text{A.39})$$

for small  $\varepsilon$ , which we set to  $1e^{-9}$ .

After the condition in Eq. (A.39) is met, we have converged fixed effects estimates, denoted  $\{\hat{\gamma}_{it}^\infty\}$  and  $\{\hat{\lambda}_{jt}^\infty\}$ , for a given estimated  $\hat{\beta}_t$ . The final piece of the estimation is to select the optimal  $\hat{\beta}_t$  that minimizes the distance between simulated and data moments in the year-month.

## A.6 Select the moments

To identify  $\beta_t$ , we choose 6 unconditional moments of the distribution of visitor counts. We select moments that describe important parts of the distribution. The moments are computed per year-month across all block groups and branches. Denote the vector of the data moments in the year-month as  $m(v)$ , and denote as  $m(\tilde{v}_s|\psi)$  the analogous vector of simulated moments from simulation  $s$ .

Recall that  $n_t$  is the set of stratified sampled block group  $\times$  branch pairs in year-month  $t$ . The set is the union of the set of pairs in  $n_t^1$  that were sampled with probability 1 and the set of pairs in  $n_t^0$  that were sampled with probability  $1/2000$ . Recall also that  $\omega_t$  are the probability weights assigned to the pairs in the set  $n_t^0$ , given in Eq. (A.20). Both the data and simulated moments only include block group  $\times$  branch pairs from the stratified sample. The 6 data and simulated moments are:

### 1. Percent of visitor counts equal to 0:

$$m_1(v) \equiv \frac{\sum_{(i,j) \in n_t^1} \mathbb{1}(V_{ijt} = 0) + \sum_{(i,j) \in n_t^0} \mathbb{1}(V_{ijt} = 0) \omega_t}{\sum_{(i,j) \in n_t^1} 1 + \sum_{(i,j) \in n_t^0} \omega_t}, \quad (\text{A.40})$$

$$m_1(\tilde{v}|\psi) \equiv \frac{\sum_{(i,j) \in n_t^1} \mathbb{1}(\tilde{V}_{ijt} = 0) + \sum_{(i,j) \in n_t^0} \Pr(\tilde{V}_{ijt} = 0 | \hat{\mu}_{ijt}) \omega_t}{\sum_{(i,j) \in n_t^1} 1 + \sum_{(i,j) \in n_t^0} \omega_t}, \quad (\text{A.41})$$

where  $\mathbb{1}(\cdot)$  stands for the indicator function and  $\Pr(\tilde{V}_{ijt} = 0 | \hat{\mu}_{ijt})$  is from Eq. (A.22). The data moment  $m_1(v)$  is straightforward, separating pairs in the two sampled sets,  $n_t^0$  and  $n_t^1$ , and applying the different probability

<sup>20</sup>The iterative process we use to identify the fixed effects is similar in spirit to the “zig-zag” algorithm, or Gauss-Seidel method, that is commonly used to identify high-dimensional fixed effects in linear models (Guimaraes and Portugal 2010).

weights. The simulated moment  $m_1(\tilde{v}|\psi)$  adds the fraction of the simulated visitor counts from the sampled set  $n_t^1$  equaling 0 to the probability of the visitor counts from the sampled set  $n_t^0$  equaling 0, adjusted by the probability weights.

**2. Percent of visitor counts equal to 4:**

$$m_2(v) \equiv \frac{\sum_{(i,j) \in n_t^1} \mathbb{1}(V_{ijt} = 4) + \sum_{(i,j) \in n_t^0} \mathbb{1}(V_{ijt} = 4) \omega_t}{\sum_{(i,j) \in n_t^1} 1 + \sum_{(i,j) \in n_t^0} \omega_t}, \quad (\text{A.42})$$

$$m_2(\tilde{v}|\psi) \equiv \frac{\sum_{(i,j) \in n_t^1} \mathbb{1}(\tilde{V}_{ijt} = 4) + \sum_{(i,j) \in n_t^0} \Pr(\tilde{V}_{ijt} = 4|\hat{\mu}_{ijt}) \omega_t}{\sum_{(i,j) \in n_t^1} 1 + \sum_{(i,j) \in n_t^0} \omega_t}, \quad (\text{A.43})$$

where  $\Pr(\tilde{V}_{ijt} = 4|\hat{\mu}_{ijt})$  is from Eq. (A.24).

**3. Average log distance, in cases where  $V_{ijt}, \tilde{V}_{ijt} = 0$ :**

$$m_3(v) \equiv \frac{\sum_{(i,j) \in n_t^1} \mathbb{1}(V_{ijt} = 0) \log d_{ij} + \sum_{(i,j) \in n_t^0} \mathbb{1}(V_{ijt} = 0) \omega_t \log d_{ij}}{\sum_{(i,j) \in n_t^1} 1 + \sum_{(i,j) \in n_t^0} \omega_t}, \quad (\text{A.44})$$

$$m_3(\tilde{v}|\psi) \equiv \frac{\sum_{(i,j) \in n_t^1} \mathbb{1}(\tilde{V}_{ijt} = 0) \log d_{ij} + \sum_{(i,j) \in n_t^0} \Pr(\tilde{V}_{ijt} = 0|\hat{\mu}_{ijt}) \omega_t \log d_{ij}}{\sum_{(i,j) \in n_t^1} 1 + \sum_{(i,j) \in n_t^0} \omega_t}. \quad (\text{A.45})$$

**4. Average log distance, in cases where  $V_{ijt}, \tilde{V}_{ijt} = 4$ :**

$$m_4(v) \equiv \frac{\sum_{(i,j) \in n_t^1} \mathbb{1}(V_{ijt} = 4) \log d_{ij} + \sum_{(i,j) \in n_t^0} \mathbb{1}(V_{ijt} = 4) \omega_t \log d_{ij}}{\sum_{(i,j) \in n_t^1} 1 + \sum_{(i,j) \in n_t^0} \omega_t}, \quad (\text{A.46})$$

$$m_4(\tilde{v}|\psi) \equiv \frac{\sum_{(i,j) \in n_t^1} \mathbb{1}(\tilde{V}_{ijt} = 4) \log d_{ij} + \sum_{(i,j) \in n_t^0} \Pr(\tilde{V}_{ijt} = 4|\hat{\mu}_{ijt}) \omega_t \log d_{ij}}{\sum_{(i,j) \in n_t^1} 1 + \sum_{(i,j) \in n_t^0} \omega_t}. \quad (\text{A.47})$$

**5. OLS coefficient from regressing log visitor counts onto their associated log distances, in cases where  $V_{ijt}, \tilde{V}_{ijt} > 0$ :**

First, using the observed data, we define the regression's dependent and independent variables, respectively, as

$$y_{ijt} = \langle \log V_{ijt} \rangle_{(i,j) \in n_t^1}, \quad (\text{A.48})$$

$$X_{ijt} = \left[ \langle 1 \rangle_{(i,j) \in n_t^1}, \langle \log d_{ij} \rangle_{(i,j) \in n_t^1} \right]. \quad (\text{A.49})$$

Here,  $\langle \cdot \rangle_{(i,j) \in n_t^1}$  denotes a vector with length equaling the number of elements in the set  $n_t^1$ . The dependent variable  $y_{ijt}$  consists of a vector of log visitor counts, whereas the independent variables are a vector of ones and a vector of log distances. With these variables established, the data moment is

$$m_5(v) \equiv \text{Second element of } (X'_{ijt} X_{ijt})^{-1} (X'_{ijt} y_{ijt}) \quad (\text{A.50})$$

Notice that, because the data moment reflects only positive observed visitor counts from the set  $n_t^1$  of sampled block group  $\times$  branch pairs, the probability weights all equal 1 and do not appear in the data moment.

The corresponding simulated moment uses a weighted least squares (WLS) coefficient because the probability weights do not all equal 1. With this in mind, we define the observation weights of the WLS as

$$\tilde{\eta}_{ijt} \equiv \left[ \begin{array}{c} \langle 1 \rangle_{(i,j) \in n_t^1: \tilde{V}_{ijt} > 0} \\ \langle \omega_t \Pr(\tilde{V}_{ijt} > 0|\hat{\mu}_{ijt}) \rangle_{(i,j) \in n_t^0} \end{array} \right], \quad (\text{A.51})$$

where  $\Pr(\tilde{V}_{ijt} > 0 | \hat{\mu}_{ijt})$  is from Eq. (A.23). The observation weights consist of (1) a vector of ones with length equaling the number of block group  $\times$  branch pairs in  $n_t^1$  that also have positive simulated visitor counts, and (2) a vector of weighted probabilities that the simulated visitor counts from the pairs in the sampled set  $n_t^0$  exceed 0.

The dependent variable in the WLS is defined as

$$\tilde{y}_{ijt} \equiv \sqrt{\tilde{\eta}_{ijt}} \odot \begin{bmatrix} \langle \log \tilde{V}_{ijt} \rangle_{(i,j) \in n_t^1: \tilde{V}_{ijt} > 0} \\ \langle \mathbb{E}(\log \tilde{V}_{ijt} | \tilde{V}_{ijt} > 0, \hat{\mu}_{ijt}) \rangle_{(i,j) \in n_t^0} \end{bmatrix}, \quad (\text{A.52})$$

where  $\odot$  is the element-wise product and  $\mathbb{E}(\log \tilde{V}_{ijt} | \tilde{V}_{ijt} > 0, \hat{\mu}_{ijt})$  is from Eq. (A.27). The dependent variable consists of (1) a weighted vector of log simulated visitor counts with length equaling the number of block group  $\times$  branch pairs in  $n_t^1$  that also have positive simulated visitor counts, and (2) a weighted vector of mean log simulated visitor counts from the pairs in the sampled set  $n_t^0$ , conditional on the simulated visitor counts exceeding 0.

The independent variable in the WLS is defined as

$$\tilde{X}_{ijt} \equiv \begin{bmatrix} \sqrt{\tilde{\eta}_{ijt}}, & \sqrt{\tilde{\eta}_{ijt}} \odot \begin{bmatrix} \langle \log d_{ij} \rangle_{(i,j) \in n_t^1: \tilde{V}_{ijt} > 0} \\ \langle \log d_{ij} \rangle_{(i,j) \in n_t^0} \end{bmatrix} \end{bmatrix}. \quad (\text{A.53})$$

The independent variable consists of (1) the square root of the weights from Eq. (A.51), and (2) the element-wise product of the square root of the weights and log distances.

With these terms established, we set the simulated moment as

$$m_5(\tilde{v} | \psi) \equiv \text{Second element of } (\tilde{X}'_{ijt} \tilde{X}_{ijt})^{-1} (\tilde{X}'_{ijt} \tilde{y}_{ijt}). \quad (\text{A.54})$$

## 6. OLS coefficient from regressing log visitor counts onto their associated log distances, where $V_{ijt}, \tilde{V}_{ijt} > 4$ :

The sixth data moment is similar to the fifth data moment, except that it conditions on the visitor count exceeding 4 rather than 0. Specifically, let

$$q_{ijt} = \langle \log V_{ijt} \rangle_{(i,j) \in n_t^1: V_{ijt} > 4} \quad (\text{A.55})$$

$$Z_{ijt} = \begin{bmatrix} \langle 1 \rangle_{(i,j) \in n_t^1: V_{ijt} > 4}, & \langle \log d_{ij} \rangle_{(i,j) \in n_t^1: V_{ijt} > 4} \end{bmatrix}. \quad (\text{A.56})$$

The data moment is then

$$m_6(v) \equiv \text{Second element of } (Z'_{ijt} Z_{ijt})^{-1} (Z'_{ijt} q_{ijt}). \quad (\text{A.57})$$

The sixth simulated moment is also similar to the fifth simulated moment, just now conditioning on  $\tilde{V}_{ijt} > 4$ . Thus, let the WLS observation weights be

$$\tilde{\xi}_{ijt} \equiv \begin{bmatrix} \langle 1 \rangle_{(i,j) \in n_t^1: \tilde{V}_{ijt} > 4} \\ \langle \omega_t \Pr(\tilde{V}_{ijt} > 4 | \hat{\mu}_{ijt}) \rangle_{(i,j) \in n_t^0} \end{bmatrix}, \quad (\text{A.58})$$

where  $\Pr(\tilde{V}_{ijt} > 4 | \hat{\mu}_{ijt})$  is from Eq. (A.25). The dependent variable in the WLS is defined as

$$\tilde{q}_{ijt} \equiv \sqrt{\tilde{\xi}_{ijt}} \odot \begin{bmatrix} \langle \log \tilde{V}_{ijt} \rangle_{(i,j) \in n_t^1: \tilde{V}_{ijt} > 4} \\ \langle \mathbb{E}(\log \tilde{V}_{ijt} | \tilde{V}_{ijt} > 4, \hat{\mu}_{ijt}) \rangle_{(i,j) \in n_t^0} \end{bmatrix}, \quad (\text{A.59})$$

where  $\mathbb{E}(\log \tilde{V}_{ijt} | \tilde{V}_{ijt} > 4, \hat{\mu}_{ijt})$  is from Eq. (A.28). Likewise, the independent variable in the WLS is defined as

$$\tilde{Z}_{ijt} \equiv \begin{bmatrix} \sqrt{\tilde{\xi}_{ijt}}, & \sqrt{\tilde{\xi}_{ijt}} \odot \begin{bmatrix} \langle \log d_{ij} \rangle_{(i,j) \in n_t^1: \tilde{V}_{ijt} > 4} \\ \langle \log d_{ij} \rangle_{(i,j) \in n_t^0} \end{bmatrix} \end{bmatrix}. \quad (\text{A.60})$$

With these terms established, we set the simulated moment as

$$m_6(\tilde{v}|\psi) \equiv \text{Second element of } \left(\tilde{Z}'_{ijt}\tilde{Z}_{ijt}\right)^{-1} \left(\tilde{Z}'_{ijt}\tilde{q}_{ijt}\right). \quad (\text{A.61})$$

In the procedure, we take the mean of the simulated moments by averaging values across the  $S$  simulations. Let  $\hat{m}(\tilde{v}|\psi)$  be the estimate of the model moments from the  $S$  simulations:

$$\hat{m}(\tilde{v}|\psi) = \frac{1}{S} \sum_S m(\tilde{v}_s|\psi). \quad (\text{A.62})$$

The final step of the MSM procedure is to find the estimated  $\hat{\beta}_t$  that minimizes the distance between the data moments and simulated model moments.

## A.7 Construct the MSM estimator

The MSM estimator  $\hat{\beta}_{t,\text{MSM}}$  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:

$$e(\tilde{v}, v|\psi) \equiv \frac{\hat{m}(\tilde{v}|\psi) - m(v)}{m(v)}. \quad (\text{A.63})$$

The MSM estimator is

$$\hat{\beta}_{t,\text{MSM}} = \underset{\hat{\beta}_t}{\text{argmin}} e(\tilde{v}, v|\psi)' W e(\tilde{v}, v|\psi), \quad (\text{A.64})$$

where  $W$  is a  $6 \times 6$  weighting matrix that controls how each moment is weighted in the minimization problem. Notice that each candidate  $\hat{\beta}_t$  in Eq. (A.64) is associated with a different set of converged fixed effects estimates  $\{\hat{\gamma}_{it}^\infty, \hat{\lambda}_{jt}^\infty\}$ .

We use the identity matrix  $I$  for the weighting matrix  $W$ . We also implemented a two-step procedure to select an optimal weighting matrix  $W$ , but that approach produced unstable estimates. This is not surprising, given evidence in the literature of the underperformance of the two-step procedure when there is uncertainty in the estimation of the weighting matrix (Arellano and Bond 1991; Hwang and Sun 2018).

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

$$\widehat{\text{Var}}(\hat{\beta}_{t,\text{MSM}}) = \left(1 + \frac{1}{S}\right) \left[ \frac{\partial \hat{m}(\tilde{v}|\psi)'}{\partial \hat{\beta}_t} W^{-1} \frac{\partial \hat{m}(\tilde{v}|\psi)}{\partial \hat{\beta}_t} \right]^{-1}, \quad (\text{A.65})$$

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

Online Appendix to  
**Bank Branch Access:  
Evidence from Geolocation Data**

Jung Sakong

Alexander K. Zentefis

April 11, 2023

## B 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 U.S. Census Bureau’s Current Population Survey, which covers a representative sample of households in the U.S. each month. The FDIC survey queries both banked and unbanked households, and the 2019 survey collected responses from almost 33,000 households. In [Online Appendix B.1](#), we discuss survey findings about bank branch use; in [Online Appendix B.2](#), we analyze differences by demographic characteristics in the primary methods that banked respondents use to access their bank accounts; and in [Online Appendix B.3](#), we compare reported branch visitor shares according to household income from the survey to observed shares from the SafeGraph geolocation data.

### B.1 Bank Branch Use

Visiting bank branches remains a common and popular bank access method. In the survey, 80.9% of all respondents (banked and unbanked) answered having visited a bank branch in the past 12 months, and 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 for banked households (31.4%). But 81.2% of respondents who cite mobile banking as their primary method also say they visited a branch over the past year and about 1 in 5 in this group visited a branch ten or more times.

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.3](#), we report coefficients from multivariate linear probability regressions of survey responses on self-reported demographic characteristics. The survey reveals a positive income gradient in reported branch use. Banked and unbanked respondents are included. Controlling for age and race, we find that respondents in the highest income bracket (\$75,000+) are roughly 22% more likely to say they visited 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, we find that 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 the survey responses.

### B.2 Primary Bank Access Methods by Household Characteristics

The FDIC survey provides 6 choices for banked respondents to select as their primary method of banking: Bank Teller, ATM/Kiosk, Online Banking, Mobile Banking, Telephone Banking, and Other. Across all respondents, the first four choices dominate as primary access methods. 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 banking into one category, as those are the two major alternatives to visiting a branch.

The survey responses show that low-income and Black households do not appear to make up their lesser branch use with greater use of online or mobile banking. In [Online Table A.4](#), we report coefficients from multivariate linear probability regressions of stated primary access methods on self-reported demographic characteristics. Controlling for age and race, we find that respondents in the lowest income bracket are roughly 31% less likely than those in the highest income bracket to say that mobile or online banking is their primary method to access their bank accounts. Controlling for income and age, we find that Black respondents are about 6.6% less likely than White respondents to call mobile or online banking their primary access method. Analogous estimates from Probit regressions in [Online Table A.5](#) document similar differences by income and race. Overall, the survey evidence reveals that banked low income and Black households respond as relying on mobile/online banking less and bank branches/ATMs more as their primary access methods.

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

[Online Fig. A.4](#) presents a binned scatter plot of the share of bank branch visitors by household income from the SafeGraph observed (raw) data. 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 American Community Survey (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 Online Fig. A.4, 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 respondents (among both banked and unbanked) that visited 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.<sup>1</sup>

Comparing the FDIC’s survey responses on branch visits to the SafeGraph data is imperfect. The survey responses measure whether a respondent visited any U.S. 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 Online Fig. A.4. The same person visiting the same branch multiple times in the year-month would count as one visitor, but the same person traveling to multiple branches in the same year-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 observed branch visitor shares from the mobile device data are increasing and concave in household income. Around 63% of respondents with household income less than \$15,000 say they visited a branch over the past year, whereas 86% of those with income \$75,000 and above reported having visited. From the geolocation data, we see that the observed visitor share is 59% for block groups with median household income around \$12,000 and 71% for block groups with median household income 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 U.S. bank branches (and their visitors). Another contributing explanation is that SafeGraph entirely misses 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 low-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 observed visitor shares at the bottom of the income distribution. Another possibility, though, is that SafeGraph regularly misses branch visitors from high income block groups, which would understate the observed visitor shares at the top of the income distribution and compress the slope.

## C Core Sample Construction

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

### C.1 SafeGraph Geolocation Data

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

---

<sup>1</sup>To compute this annual share of branch visitors, we first divide the total branch visitors in each Census block group by the total recorded mobile devices residing in the block group per year-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  $i$  in year-month  $t$  be denoted  $p_{i,t}$ . Not every block group has a visitor probability each month, so let  $k_i$  denote the number of months for which block group  $i$  has branch visitors. The annual branch visitor share  $s_i$  for block group  $i$  is  $s_i = 1 - \prod_{t=1}^{12/k_i} (1 - p_{i,t})^{12/k_i}$ . After computing each block group’s annual branch visitor share, we categorize block groups by median household income from the 2019 5-year ACS.

spend money and/or time.”<sup>2</sup> 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 establishment 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 establishment opened, and when SafeGraph began tracking information about the establishment. 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.<sup>3</sup> In this dataset, 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 that visitors traveled to reach the POI, 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 home Census block groups of visitors, and the number of visitors from each of those home block groups. To protect user privacy, SafeGraph employs differential privacy methods to the visitor home block group data. First, it adds Laplace noise to each block group’s visitor count (when it observes at least one visitor from the block group). Second, after the noise is added, SafeGraph rounds the visitor counts down to their nearest integers. Third, SafeGraph then truncates the rounded visitor counts by only reporting data from block groups with at least two visitors. Fourth, home block groups with only two, three, or four visitors are reported as having four visitors.

SafeGraph determines a visitor’s home Census block group using an algorithm. A brief description of that algorithm is as follows. 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 the device’s 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 identified. Finally, SafeGraph computes a confidence score for each device’s calculated home block group. Only high-confidence home locations are included; otherwise, the device’s home location is classified as unknown.<sup>4</sup>

## C.2 FDIC Summary of Deposits

To construct our core sample, we rely on branch information from the Federal Deposit Insurance Corporation (FDIC). Branch data are from the FDIC’s 2019 Summary of Deposits (SOD).<sup>5</sup> 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.

## C.3 Construction Process

Our core sample can be thought of as consisting of two components: (i) a set of locations and (ii) consumer movement to those locations. We call these two components “places” and “visitors.” In our case, the places and visitors are specific to bank branches. SafeGraph is our only source of visitor data, and so, we rely on it exclusively. The visitors data field we use that contains the home Census block groups of the visitors to a branch is VISITOR\_HOME\_CBGs. As we describe in the text, this data field is subject to SafeGraph’s differential privacy.

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.

---

<sup>2</sup>See the [SafeGraph Places Manual](#) and [Data Guide](#) for more details.

<sup>3</sup>See the [SafeGraph Privacy Policy](#) for more details.

<sup>4</sup>Full details of the algorithm are found here: [Home Identification Algorithm](#).

<sup>5</sup>FDIC SOD data are located here: [SOD](#).

### C.3.1 Placekey

Placekey is a free, standardized identifier of 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 consider an example Placekey, take 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](#).

### C.3.2 Choosing the Set of Places

Both the 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 the 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 visitor 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 visitor 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, which we define next.

Our **core sample** includes only SafeGraph places with brands that are included in the SOD and for which we have visitor geolocation data from SafeGraph. 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 union of the SOD fields namefull and namehcr to identify a bank’s brand. In SafeGraph, we use the field LOCATION\_NAME to identify a bank brand name. 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.

## D Assigning Demographic Attributes to Individual Visitors

Our goal is not only to develop a local measure of bank branch access, but also to apply that measure to explain differences in branch use by race and income. But we face a limitation when using anonymous mobile device data: 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 identified 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 et al. 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 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. 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. Online Fig. A.3, 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 50<sup>th</sup> 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 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 the characteristics of block group residents. 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  $x\%$  weaker 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 Fig. A.3, 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  $100x\%$  weaker 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  $100x\%$  weaker access than a White resident.”

## E Postal Banking

The geolocation data and gravity model allow us to study a policy proposal that might improve access for branch goers. In particular, we examine postal banking. A Postal Savings System existed in the United States beginning in 1911, but Congress phased it out in 1966 (O’Hara and Easley 1979; Shaw 2018). The system was promoted to reach the unbanked, and non-farming immigrant populations initially used it for short-term savings and as a partial substitute for private banks (Schuster, Jaremski, and Perlman 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 assess how a Postal Banking System—which would extend checking, savings, and possibly credit services to some or all U.S. Post Office branches—might affect both access to and use of banking products or services at branches. From Eq. (2) and Eq. (3), the expected number of block group residents 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  $\gamma_{it}$ , (ii) the fixed effects of both postal and private bank branches  $\lambda_{jt}$ , (iii) the distances between the block group and branches  $d_{ij}$ , (iv) the gravity parameters  $\beta_t$ , and (v) the set of both postal and private branches available to all residents  $B_t$ . Our evaluation of a postal banking policy requires an assumption for each component.

Components (iii) and (v) are the least controversial. For the set of branches,  $B_t$ , we include all private bank branches per year-month in our core sample like before, but now we also include all Post Office branches 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 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 might change under a postal banking system. Per [Section 3](#),  $\beta_t$  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, as a person can access financial services at a postal bank when dropping off mail. For simplicity, we assume that the gravity model that governs visitor flows to all bank branches, both postal and private, has the same  $\beta_t$  per year-month, as estimated in the month-by-month MSM procedure from before, which implicitly presumes a common elasticity of substitution across institutions.

The introduction of a postal banking system would reasonably affect component (i), a block group’s fixed effect  $\gamma_{it}$ , which captures all attributes of the block group’s residents that influence demand for any branch’s products or services. The clearest change is postal banking encouraging bank 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 situate 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  $\{\lambda_{jt}\}$  of component (ii) is also 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, say, in trade costs.

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 branch access and use 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 local variation in postal bank quality cross-country. Second, we assign three estimated fixed effects to postal banks based on different parts of the distribution of estimated private bank fixed effects per year-month: the 10<sup>th</sup> percentile, 50<sup>th</sup> percentile, and 90<sup>th</sup> 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 50<sup>th</sup> percentile assumes that the typical postal bank would have the quality of the median private bank per year-month (a “medium quality” system), and the 90<sup>th</sup> percentile assumes that postal banks would be perceived as having the same quality as the top 10 percent of private banks per year-month (a “high quality” system).

To measure the extent to which bank branch access would change under postal banking, we re-run the access regressions from [Table 1](#), but in computing  $\Phi_{it}$  per Census block group, we now include the locations of all Post Office branches that are registered in SafeGraph within a block group’s set of branches. To make the policy evaluation comparable to our analysis earlier that considered only private banks, we only include block groups whose residents visited a private bank branch in the year-month.

The results are in [Online Table A.10](#). Under a medium quality system in column (4), the nationwide estimate of the coefficient on medium household income with controls for block group racial shares and age shares is -8.0%. This value

contrasts to the corresponding coefficient on medium household income in [Table 1](#) of -7.6%. Hence, a postal banking system of medium quality steepens the negative income gradient in branch access by roughly 0.4 percentage points, which implies that access would improve relatively more for residents of low-income block groups than high-income block groups. Under a low-quality postal banking system, the income gradient flattens slightly by 0.2 percentage points, and under a high quality system, the relative improvement in access for low-income block groups rises to 1.1 percentage points.

The coefficient on the Black population share in column (4) of [Online Table A.10](#) is -5.8%, which is slightly higher than the coefficient in [Table 1](#). Nationwide, then, a medium-quality postal banking system would improve access for both Black and White communities, but improve it by relatively more in White than Black communities. The racial gap in access widens because Post Offices also tend to be located comparatively closer to White communities than Black communities, just like private bank branches. Only a high-quality postal banking system would shrink the Black-White gap in access nationwide, but only by 0.5 percentage points (from 5.3% to 4.8%).

Zeroing in on Metro cores, we find that access would relatively improve for residents of low-income block groups between 0.2 to 0.6 percentage points under a medium- and high-quality system, respectively, which is a smaller range than the national estimates. But for Black communities in big cities, access would improve under only a high-quality postal system, by 1.1 percentage points. Under a low- and medium-quality system, the Black-White gap in access would widen in big cities by 0.7 and 0.3 percentage points, respectively.

Overall, we find that a postal banking policy would have the largest effects on branch access and (branch use) for residents of low-income block groups nationwide. But it generally would widen the Black-White gap in access, raising access relatively more for White communities than for Black communities. If the goal of the policy were to shrink the Black-White gap in access and use, the postal banking system would need to be of high quality, and its largest impact would be in big cities. Again, we caveat these findings with the acknowledgement that our investigation embedded several simplifying assumptions and was conducted in partial equilibrium.

## F A Simple Model of Bank Branch Choice

A continuum of residents choose destinations to visit from their home Census block groups per time period. Each resident  $r$  lives in one block group  $i \in G$ . Bank branches are located across the country, and each branch is indexed by  $j \in B_t$ , where the set of branches can vary over time from store openings and closings.

In every period, a 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  $i$  and visiting branch  $j$  at time  $t$  is

$$U_{rjt} = \frac{z_{rjt} \Lambda_{jt}}{\delta_{ij}}. \quad (\text{B.66})$$

The term  $\Lambda_{jt}$  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 attractive deposit or loan rates, higher customer service quality, or a wider variety of products). The term  $z_{rjt}$  is an idiosyncratic, unobserved error that captures individual differences in residents' personal preferences for banking at branch  $j$  (e.g., 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  $\delta_{ij}$  is an iceberg traveling cost that is defined as

$$\delta_{ij} = d_{ij}^\kappa, \quad (\text{B.67})$$

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:

$$F(z_{rjt}) = e^{-H_{jt} z_{rjt}^\varepsilon}, \quad H_{jt} > 0, \quad \varepsilon > 1. \quad (\text{B.68})$$

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  $\varepsilon$  implies that residents are more heterogeneous in their preferences for branches.<sup>6</sup>

<sup>6</sup>The parameter  $\varepsilon$  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  $\varepsilon$  is akin to branches being less substitutable, which implies that residents find it worthwhile to travel to a branch despite the resistance imposed by the geographic barrier  $\delta_{ij}$ .

Substituting the expression for  $U_{rjt}$  into the distribution of idiosyncratic tastes in Eq. (B.68), one can observe that residents of block group  $i$  at time  $t$  are presented with a distribution of utility across branches,  $G_{ijt}(u) = \Pr[U_{rjt} \leq u] = F(u\delta_{ij}/\Lambda_{jt})$ , or

$$G_{ijt}(u) = e^{-\left[H_{jt}\left(\frac{\Lambda_{jt}}{\delta_{ij}}\right)^\epsilon\right]u^{-\epsilon}}. \quad (\text{B.69})$$

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

$$G_{it}(u) = \prod_{j=0}^{B_t} G_{ijt}(u). \quad (\text{B.70})$$

Inserting Eq. (B.69) into Eq. (B.70), one obtains the utility distribution:

$$G_{it}(u) = e^{-(1+\Psi_{it})u^{-\epsilon}}, \quad (\text{B.71})$$

where the parameter  $\Psi_{it}$  of block group  $i$ 's utility distribution is

$$\Psi_{it} = \sum_{j \in B_t} H_{jt}\Lambda_{jt}^\epsilon d_{ij}^{-\kappa\epsilon}. \quad (\text{B.72})$$

The utility distribution generates a gravity equation in visits between home block groups and bank branches. The share  $\pi_{ijt}$  of residents living in block group  $i$  who visit branch  $j$  at time  $t$  is

$$\pi_{ijt} = \frac{H_{jt}\Lambda_{jt}^\epsilon d_{ij}^{-\kappa\epsilon}}{1 + \Psi_{it}}. \quad (\text{B.73})$$

The visitor share depends on the characteristics of the branch ( $\Lambda_{jt}$ ), the average utility draw of the branch ( $H_{jt}$ ), and the ‘‘bilateral resistance’’ derived from the intervening transportation costs ( $d_{ij}^{-\kappa\epsilon}$ ). Other things equal, a resident is more likely to visit a branch if it has superior attributes, delivers higher average idiosyncratic utility, or is less costly to reach. In the denominator,  $\Psi_{it}$  plays the role of ‘‘multilateral resistance,’’ which 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.<sup>7</sup>

## Online Appendix References

- AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): ‘‘The economics of density: Evidence from the Berlin Wall,’’ *Econometrica*, 83, 2127–2189.
- BARADARAN, M. (2013): ‘‘It’s time for postal banking,’’ *Harvard Law Review Forum*, 127, 165–175.
- EATON, J. AND S. KORTUM (2002): ‘‘Technology, geography, and trade,’’ *Econometrica*, 70, 1741–1779.
- GILLIBRAND, K. (2021): ‘‘Senators Gillibrand And Sanders, Representatives Ocasio-Cortez, Pascrell, and Kaptur call on Congress to implement Postal Banking pilot programs,’’ *Kirsten Gillibrand | U.S. Senator for New York*.
- HEAD, K. AND T. MAYER (2014): ‘‘Gravity equations: Workhorse, toolkit, and cookbook,’’ in *Handbook of International Economics*, vol. 4, 131–195.
- JOHNSON, R. K. (2017): ‘‘How the United States Postal Service (USPS) could encourage more local economic development,’’ *Chicago Kent Law Rev.*, 92, 593–615.
- KING, G. (1997): *A Solution to the Ecological Inference Problem*, Princeton University Press.
- KING, G., M. A. TANNER, AND O. ROSEN (2004): *Ecological Inference: New Methodological Strategies*, Cambridge University Press.
- McFADDEN, D. (1974): ‘‘The measurement of urban travel demand,’’ *Journal of Public Economics*, 3, 303–328.

<sup>7</sup>We obtain the gravity relation in Eq. (B.73) by evaluating:

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

- O'HARA, M. AND D. EASLEY (1979): "The postal savings system in the depression," *Journal of Economic History*, 39, 741–753.
- SANDERS, B. (2021): "Fair banking for all," *Bernie Sanders Official Website*.
- SCHUSTER, S. S., M. JAREMSKI, AND E. R. PERLMAN (2020): "An empirical history of the US Postal Savings System," *Social Science History*, 44, 667–696.
- SHAW, C. W. (2018): "'Banks of the people': The life and death of the US Postal Savings System," *Journal of Social History*, 52, 121–152.
- WARREN, E. (2014): "The big benefits of postal service banking," *U.S. News*.



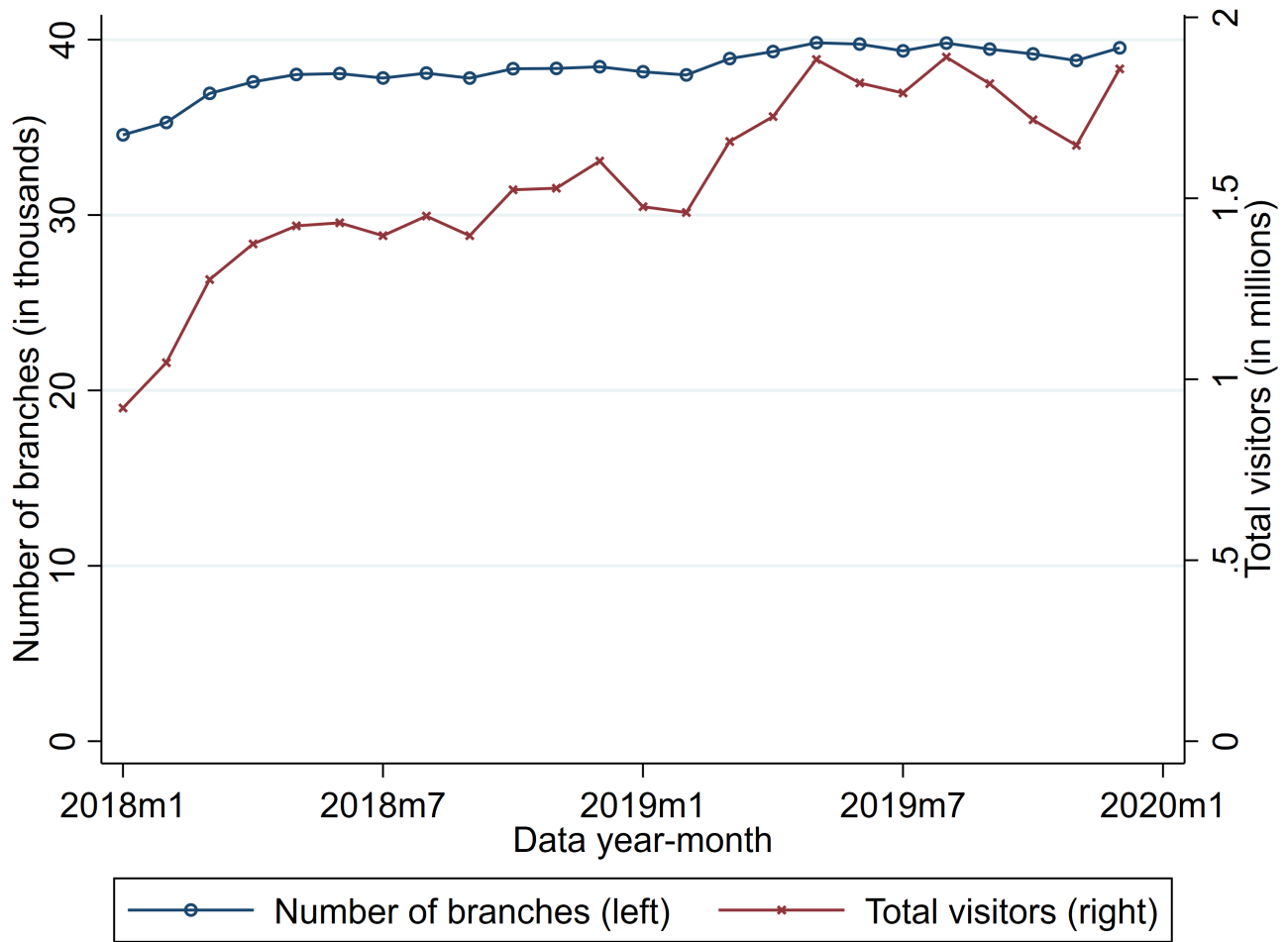


FIGURE A.1  
NUMBER OF BANK BRANCHES AND BRANCH VISITORS - CORE SAMPLE

The figure presents the number of bank branches and number of branch visitors each year-month in our core sample. The core sample of geolocation data 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.

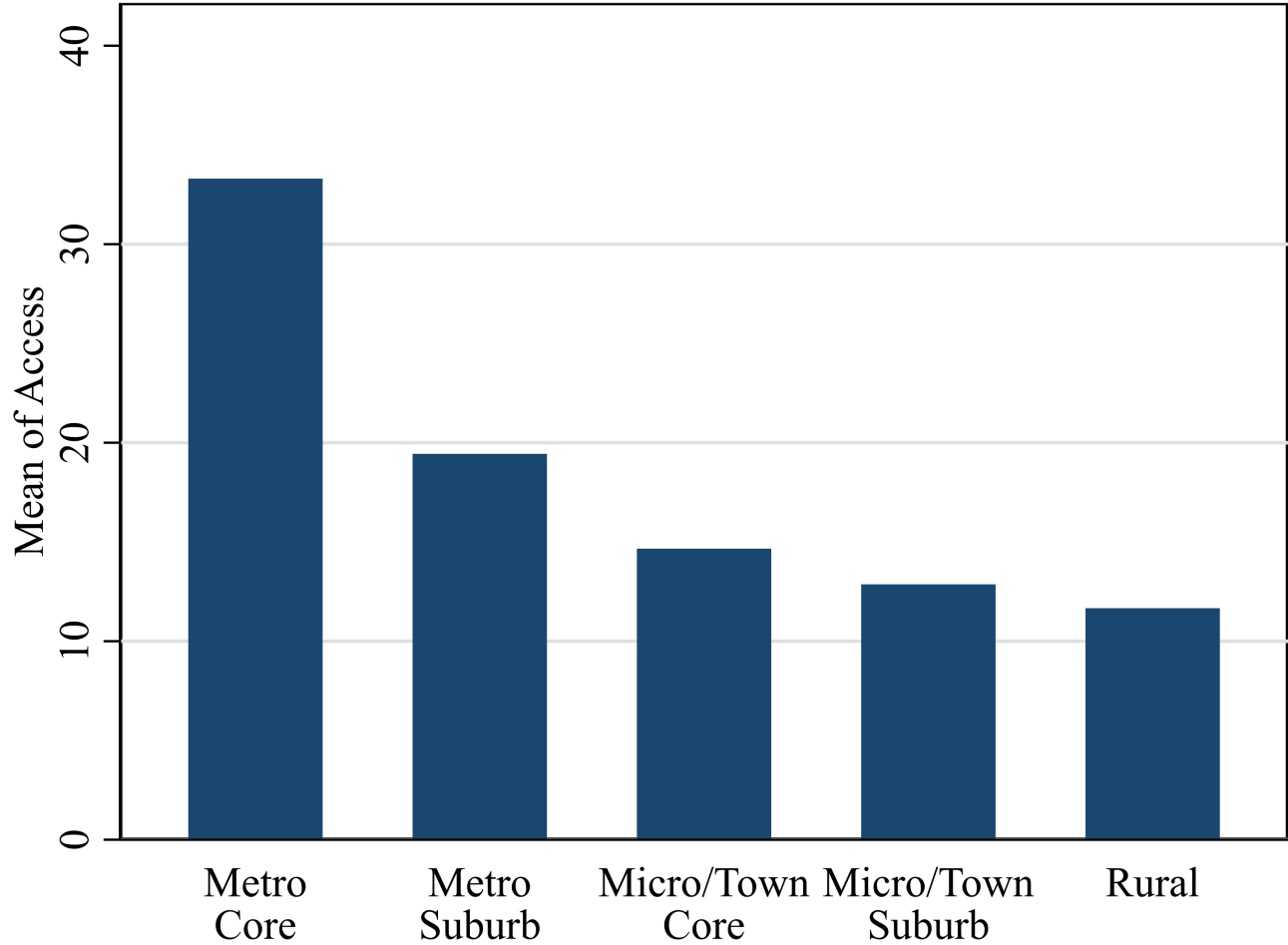
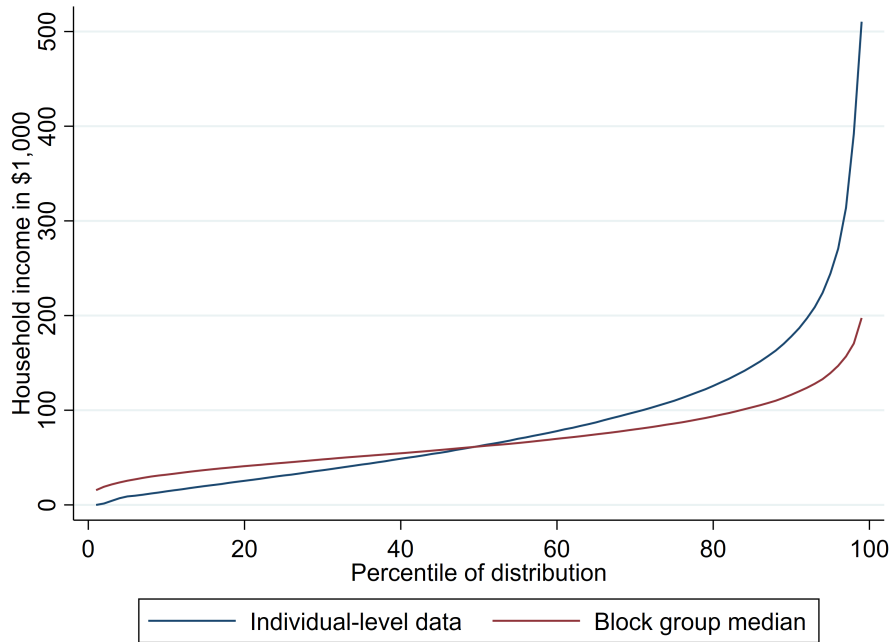


FIGURE A.2  
BANK BRANCH ACCESS BY RURAL-URBAN COMMUTING AREA

The figure presents weighted average measures of bank branch access by primary Rural-Urban Commuting Areas (RUCA). We compute each block group’s monthly access measure according to the Method of Simulated Moments estimation described in Section 5, with full details of the method in Appendix A. The monthly estimates are then averaged over time per block group, 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). Block groups are then assigned to one of the 5 displayed RUCA categories, and each category’s access value is the population-weighted average of the access measures of all block groups belonging to that category. Population shares are from the 2019 5-year American Community Survey (ACS). *Metro Core* includes RUCA 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.

**(A) Household Income**



**(B) Black Population Share**

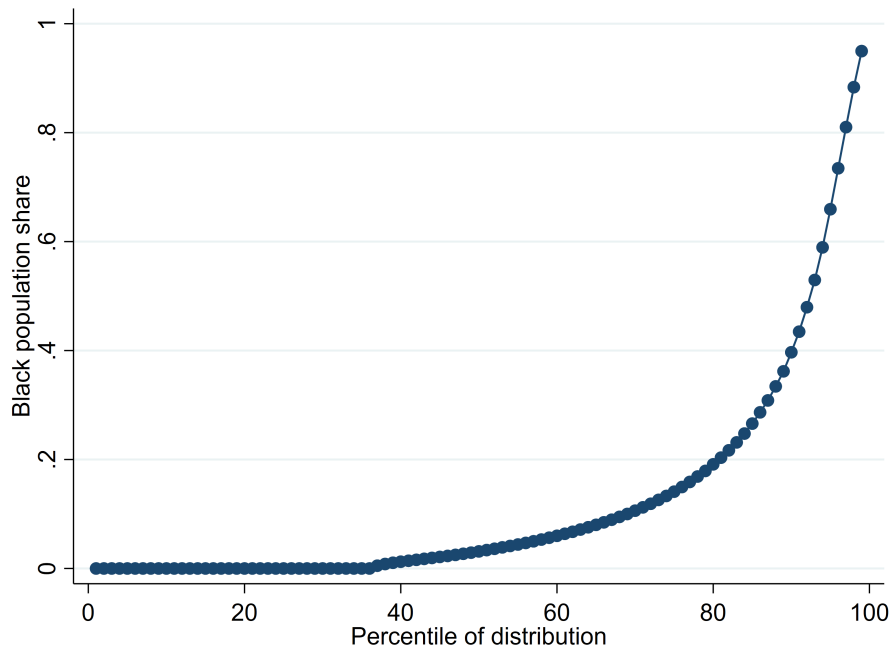


FIGURE A.3

DISTRIBUTIONS OF DEMOGRAPHIC ATTRIBUTES

The figure presents the percentiles of the distributions of U.S. household income and Black population shares. Panel A gives the percentiles of the individual-level household income distribution and the distribution of median household income at the level of Census block groups. Panel B gives the percentiles of the distribution of Black population shares across all Census block groups. 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.

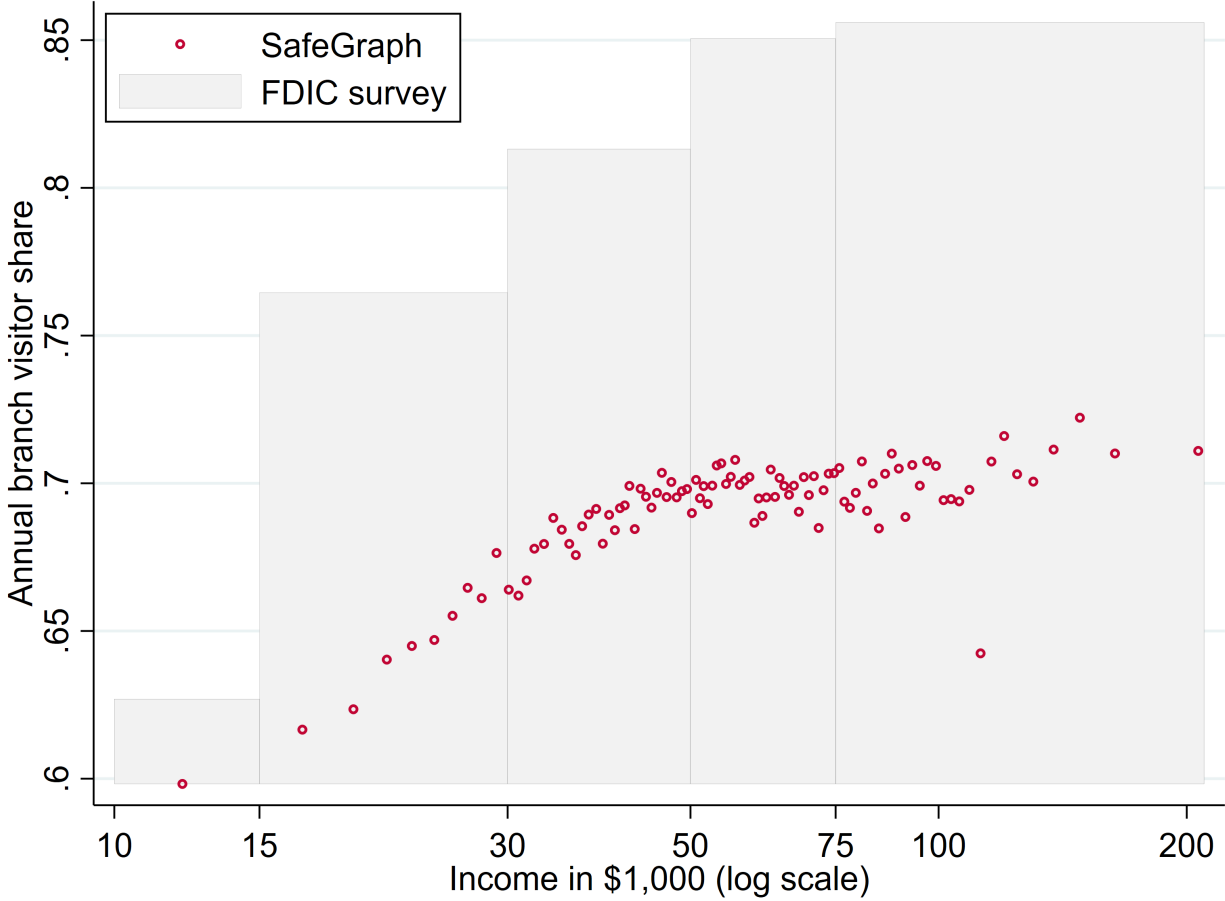


FIGURE A.4  
BANK BRANCH VISITOR SHARE BY INCOME - FDIC SURVEY & SAFEGRAPH

The figure presents a binned scatter plot of the shares of residents who visit bank branches according to household income, comparing FDIC survey responses to observed visitors in SafeGraph. 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. Observed branch visitor shares are based on our core SafeGraph sample of branch locations 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 reported 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 observed shares of branch visitors, we first divide a Census block group’s total branch visitors by its total residing mobile devices in each year-month of the period. 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 of block group  $i$  in year-month  $t$  be denoted  $p_{i,t}$ . Not every block group has a visitor probability each month, so let  $k_i$  denote the number of months for which block group  $i$  has observations. The annual branch visitor share  $s_i$  for block group  $i$  is  $s_i = 1 - \prod_{t=1}^{12/k_i} (1 - p_{i,t})^{12/k_i}$ . 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.

TABLE A.1  
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

The table reports descriptive statistics of key variables related to bank branch visitation. All values are based on our core sample of geolocation data, 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 their 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 A.2  
 DEMOGRAPHIC ATTRIBUTES OF RESIDENTS IN AREAS REPRESENTED IN CORE SAFEGRAPH SAMPLE vs. FDIC SOD

	Core Sample		SOD		Diff	
	$\hat{\mu}_1$	$\hat{\sigma}_1$	$\hat{\mu}_2$	$\hat{\sigma}_2$	$\hat{\mu}_1 - \hat{\mu}_2$ (se)	$\hat{\sigma}_1 - \hat{\sigma}_2$ (se)
N branch	51,369		86,374		-35,005	-0.000 (0.000)
White	0.799		0.805		-0.006 (0.001)	-0.003 (0.001)
Black	0.103		0.095		0.008 (0.001)	0.001 (0.000)
Asian	0.046		0.047		-0.001 (0.000)	0.002 (0.001)
Hispanic	0.109		0.106		0.003 (0.001)	0.000 (0.001)
Homeowner	0.645		0.643		0.002 (0.001)	0.000 (0.000)
Age 15-34	0.188		0.190		-0.001 (0.001)	536 (166)
Age 35-54	0.350		0.347		0.004 (0.000)	-0.002 (0.000)
Age 55-64	0.195		0.195		-0.000 (0.000)	0.028 (0.002)
Age 65+	0.267		0.269		-0.002 (0.000)	-3,205 (1,383)
< HS	0.104		0.104		0.000 (0.000)	0.000 (0.000)
HS degree	0.075		0.075		0.000 (0.000)	0.000 (0.000)
Some college	0.260		0.263		-0.003 (0.001)	0.001 (0.000)
College degree	0.103		0.106		-0.003 (0.001)	0.000 (0.000)
> College	0.199		0.198		0.001 (0.000)	0.000 (0.001)
Unemp rate	0.052		0.054		-0.002 (0.000)	0.000 (0.000)
HH income (\$)	0.297		0.295		0.002 (0.000)	0.000 (0.000)
In poverty	0.090		0.091		-0.001 (0.000)	0.000 (0.000)
Urban	0.140		0.140		0.000 (0.000)	0.028 (0.002)
Home value (\$)	0.098		0.100		-0.002 (0.000)	-3,205 (1,383)
	0.050		0.050		0.000 (0.000)	0.000 (0.000)
	0.026		0.028		-0.002 (0.000)	0.000 (0.000)
	70,338		69,802		536 (166)	0.028 (0.002)
	29,657		30,052		-356 (166)	0.028 (0.002)
	0.127		0.129		-0.002 (0.000)	0.000 (0.000)
	0.082		0.084		-0.002 (0.000)	0.000 (0.000)
	0.815		0.787		0.028 (0.000)	0.000 (0.000)
	0.295		0.319		-0.024 (0.000)	0.000 (0.000)
	286,881		290,086		-3,205 (1,383)	0.000 (0.000)
	243,368		254,504		-11,136 (1,383)	0.000 (0.000)

The table compares demographic characteristics of residents of geographic areas represented in our core sample of bank branches with characteristics of residents of geographic areas represented 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). *White* includes both Hispanic and non-Hispanic White racial shares. 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 is the difference in sample means between the core sample and the SOD, 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 heteroskedasticity-robust standard error of the estimated difference between the two sample means.

TABLE A.3  
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 (0.012)	0.127 (0.012)		0.362 (0.034)	0.363 (0.035)
\$30,000 to \$50,000		0.178 (0.011)	0.183 (0.011)		0.527 (0.033)	0.552 (0.034)
\$50,000 to \$75,000		0.206 (0.011)	0.214 (0.011)		0.636 (0.035)	0.673 (0.035)
At least \$75,000		0.207 (0.010)	0.218 (0.010)		0.643 (0.030)	0.693 (0.031)
Black	-0.144 (0.009)	-0.111 (0.009)	-0.100 (0.009)	-0.476 (0.028)	-0.370 (0.028)	-0.331 (0.028)
Hispanic	-0.121 (0.009)	-0.101 (0.009)	-0.084 (0.009)	-0.409 (0.028)	-0.345 (0.028)	-0.285 (0.029)
Asian	-0.072 (0.013)	-0.074 (0.013)	-0.060 (0.013)	-0.259 (0.042)	-0.274 (0.042)	-0.225 (0.042)
Other	-0.077 (0.023)	-0.056 (0.023)	-0.048 (0.022)	-0.274 (0.074)	-0.203 (0.075)	-0.176 (0.075)
Age 35-54			0.016 (0.008)			0.048 (0.027)
Age 55-64			0.064 (0.008)			0.236 (0.031)
Age 65+			0.074 (0.008)			0.275 (0.028)
Constant	0.836 (0.003)	0.660 (0.010)	0.612 (0.012)	0.977 (0.011)	0.457 (0.027)	0.283 (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

Each column reports coefficients from a weighted binary regression with heteroskedasticity-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.4

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

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

Each column reports coefficients from a weighted binary OLS regression with heteroskedasticity-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 respondents who 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

## SURVEY REPORTED BANK ACCOUNT PRIMARY ACCESS METHOD BY HOUSEHOLD CHARACTERISTICS - PROBIT MODEL

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

Each column reports coefficients from a weighted binary Probit regression with heteroskedasticity-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 respondents who 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.6  
COMPARING GRAVITY EQUATION ESTIMATION METHODS

Year	Month	MSM		PPML		OLS		OLS where $\geq 4$	
		$\beta$	s.e.	$\beta$	s.e.	$\beta$	s.e.	$\beta$	s.e.
2018	1	-1.26	(0.035)	-0.066	(0.003)	-0.038	(0.001)	-0.331	(0.030)
	2	-1.31	(0.227)	-0.072	(0.004)	-0.042	(0.001)	-0.319	(0.023)
	3	-1.32	(0.019)	-0.076	(0.003)	-0.046	(0.001)	-0.295	(0.018)
	4	-1.33	(0.033)	-0.073	(0.002)	-0.045	(0.001)	-0.287	(0.016)
	5	-1.32	(0.011)	-0.075	(0.003)	-0.045	(0.001)	-0.297	(0.017)
	6	-1.30	(0.007)	-0.072	(0.002)	-0.045	(0.001)	-0.288	(0.017)
	7	-1.27	(0.043)	-0.069	(0.002)	-0.043	(0.001)	-0.278	(0.018)
	8	-1.29	(0.053)	-0.079	(0.003)	-0.047	(0.001)	-0.317	(0.018)
	9	-1.34	(0.304)	-0.082	(0.002)	-0.049	(0.001)	-0.340	(0.022)
	10	-1.37	(0.090)	-0.086	(0.003)	-0.051	(0.001)	-0.303	(0.016)
	11	-1.31	(0.032)	-0.086	(0.003)	-0.051	(0.001)	-0.293	(0.014)
	12	-1.31	(0.035)	-0.091	(0.003)	-0.053	(0.001)	-0.269	(0.014)
2019	1	-1.40	(0.018)	-0.089	(0.003)	-0.053	(0.001)	-0.300	(0.014)
	2	-1.43	(0.030)	-0.089	(0.002)	-0.053	(0.001)	-0.286	(0.015)
	3	-1.37	(0.035)	-0.096	(0.003)	-0.056	(0.001)	-0.279	(0.014)
	4	-1.39	(0.016)	-0.098	(0.003)	-0.056	(0.001)	-0.268	(0.012)
	5	-1.40	(0.023)	-0.106	(0.003)	-0.061	(0.001)	-0.258	(0.010)
	6	-1.38	(0.177)	-0.096	(0.002)	-0.057	(0.001)	-0.274	(0.010)
	7	-1.35	(0.106)	-0.095	(0.003)	-0.056	(0.001)	-0.261	(0.011)
	8	-1.40	(0.039)	-0.103	(0.003)	-0.061	(0.001)	-0.270	(0.010)
	9	-1.41	(0.034)	-0.108	(0.003)	-0.060	(0.001)	-0.290	(0.011)
	10	-1.45	(0.031)	-0.102	(0.003)	-0.059	(0.001)	-0.291	(0.012)
	11	-1.43	(0.015)	-0.099	(0.003)	-0.058	(0.001)	-0.290	(0.013)
	12	-1.41	(0.033)	-0.105	(0.003)	-0.062	(0.001)	-0.285	(0.010)
Panel				-0.091	(0.002)	-0.053	(0.001)	-0.283	(0.008)

The table reports estimates and standard errors of the gravity coefficient  $\beta_t$  from the fixed-effects gravity model in Eq. (1):

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

Columns (3) and (4) present estimates from the Method of Simulated Moments estimation described in Section 5, with full details in Appendix A. Columns (5) and (6) present estimates from an unweighted Poisson pseudo-maximum-likelihood (PPML) estimation as in Silva and Tenreyro (2006), run using `ppmlhdfe` in Stata. Columns (7)-(10) present estimates from an unweighted OLS regression. The PPML and OLS estimations use the raw number of visitors from home Census block groups to bank branches based on our core sample of geolocation data, 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. Columns (9) and (10) restrict the sample to visitor counts of at least 4, which circumvent SafeGraph's truncation and censoring. The MSM, PPML, and OLS gravity coefficient estimates are calculated month-by-month over the sample period (January 2018 - December 2019). PPML and OLS estimates are also calculated over the full sample panel. Standard errors of the MSM estimates are described in Appendix A. Standard errors of the PPML and OLS estimates are two-way clustered by both Census block groups and bank branches.

TABLE A.7  
DRIVING TIME VERSUS HAVERSINE DISTANCE

Dep. var.:	Driving time b/w block group and visited branch								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Haversine distance b/w block group and visited branch	0.641 (0.000)	0.634 (0.000)	0.631 (0.000)	0.632 (0.000)	0.649 (0.000)	0.652 (0.000)	0.634 (0.000)	0.647 (0.000)	0.632 (0.000)
Constant	57.610 (0.177)	60.684 (0.150)	58.863 (0.494)	64.234 (0.156)	51.796 (0.319)	51.427 (0.289)	58.749 (0.195)	50.702 (0.303)	67.206 (0.171)
Observations	995,000	725,000	35,000	498,000	497,000	498,000	497,000	508,000	487,000
Adjusted $R^2$	0.982	0.991	0.993	0.992	0.972	0.973	0.990	0.975	0.990
Sample	Core	MC	Core	Core	Core	Core	Core	Core	Core
Black > 0.8			O						
Black $\geq$ Med. Black				O					
Black < Med. Black					O				
White $\geq$ Med. White						O			
White < Med. White							O		
log(Income) $\geq$ Med. log(Income)								O	
log(Income) < Med. log(Income)									O

Each column reports coefficients from a univariate, weighted OLS regression with heteroskedasticity-robust standard errors reported in parentheses. One observation is a block group  $\times$  branch pair from our core sample of Census block groups and bank branches, where the branches consist 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). Observations are weighted by block-group population counts from the 2019 5-year American Community Survey (ACS). Dependent variable observations are the driving times from the population-weighted centers of block groups to branches, where driving times are computed using the Origin-Destination Cost Matrix of ArcGIS Pro under the default settings. Centers of population are from the 2010 Census. Independent variable observations are the corresponding haversine distances between block groups and branches. 995,000 block group  $\times$  branch pairs were drawn randomly. Column (1) includes the entire random sample of block group  $\times$  branch pairs. Column (2) restricts the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitan area core). Column (3) restricts the sample to block groups with Black population shares exceeding 80%. Column (4) restricts the sample to block groups with Black population shares at or exceeding the median Black population share across all block groups in the entire random sample. Column (5) restricts the sample to block groups with Black population shares below the median Black population share across all block groups in the entire random sample. Column (6) restricts the sample to block groups with White population shares at or exceeding the median White population share across all block groups in the entire random sample. Column (7) restricts the sample to block groups with White population shares below the median White population share across all block groups in the entire random sample. Column (8) restricts the sample to block groups with the natural logarithm of median household income at or exceeding the median of the natural logarithm of median household income across all block groups in the entire random sample. Column (9) restricts the sample to block groups with the natural logarithm of median household income below the median of the natural logarithm of median household income across all block groups in the entire random sample. Racial shares and median household income are from the 2019 5-year ACS.

TABLE A.8  
BANK BRANCH ACCESS BY DEMOGRAPHIC ATTRIBUTES - ALL SOD BRANCHES

Dep. var.:	log(Bank branch access of block groups)								
	Imputed branch quality:	Mean branch FE in year-month				Median branch FE in year-month			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	-0.131 (0.003)	-0.084 (0.003)	-0.147 (0.003)	-0.093 (0.003)	-0.124 (0.003)	-0.081 (0.003)	-0.140 (0.003)	-0.091 (0.003)	
Black	-0.187 (0.005)	-0.156 (0.005)	-0.226 (0.006)	-0.178 (0.006)	-0.148 (0.005)	-0.118 (0.005)	-0.181 (0.006)	-0.135 (0.006)	
Asian	0.504 (0.016)	0.453 (0.016)	0.454 (0.017)	0.404 (0.017)	0.492 (0.015)	0.447 (0.015)	0.445 (0.015)	0.402 (0.015)	
Other	-0.012 (0.022)	-0.039 (0.022)	0.049 (0.032)	0.012 (0.031)	0.003 (0.022)	-0.015 (0.022)	0.061 (0.032)	0.035 (0.031)	
Hispanic	-0.045 (0.007)	-0.012 (0.007)	-0.087 (0.007)	-0.037 (0.008)	-0.010 (0.007)	0.023 (0.007)	-0.045 (0.007)	0.004 (0.008)	
Age <15		-0.962 (0.019)		-1.089 (0.022)		-0.869 (0.018)		-0.984 (0.021)	
Age 35-54		-0.350 (0.018)		-0.310 (0.022)		-0.307 (0.017)		-0.270 (0.021)	
Age 55-64		-0.799 (0.019)		-0.835 (0.023)		-0.705 (0.018)		-0.736 (0.022)	
Age 65+		-0.383 (0.014)		-0.418 (0.016)		-0.332 (0.013)		-0.360 (0.015)	
log(No. of devices)	-0.057 (0.002)	-0.063 (0.002)	-0.065 (0.002)	-0.072 (0.002)	-0.054 (0.002)	-0.058 (0.002)	-0.061 (0.002)	-0.066 (0.002)	
Observations	2,549,020	2,549,020	1,847,252	1,847,252	2,549,020	2,549,020	1,847,252	1,847,252	
Adjusted $R^2$	0.891	0.896	0.874	0.881	0.748	0.754	0.694	0.703	
Sample	Core	Core	MC	MC	Core	Core	MC	MC	
Year-month FE	○	○	○	○	○	○	○	○	
County FE	○	○	○	○	○	○	○	○	
RUCA FE	○	○			○	○			

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 bank branch in the year-month are dropped. Observations are weighted by block-group population counts from the 2019 5-year American Community Survey (ACS). In all columns, the dependent variable is the natural logarithm of the estimated bank branch access measure,  $\log \hat{\Phi}_{it}$ , from Eq. (8). All columns use the complete set of branches in the FDIC's 2019 Summary of Deposits (SOD). Branches in the SOD that are also in SafeGraph have their fixed effects estimated month-by-month from the Method of Simulated moments procedure described in Section 5, with full details in Appendix A. Branches in the SOD that are not in SafeGraph have their estimated fixed effects imputed per period with the national mean (columns 1-4) or median (columns 5-8) of the estimated fixed effects of the branches in SafeGraph. The set of branches in SafeGraph is 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 SOD. Demographic independent variable observations are population-based decimal shares from the 2019 5-year ACS. Income is median household income. The log number of devices is SafeGraph's record of the number of mobile devices residing in the block group in the year-month. Columns (1), (2), (5), and (6) include all block groups for which we have branch visitor data, whereas columns (3), (4), (7), and (8) restrict the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitan area core). The omitted demographic groups are non-Hispanic Whites and age range 15-34.

TABLE A.9  
BRANCH ACCESS VERSUS BRANCH DENSITY

Dep. var.:	Census tract weighted average of log(bank branch access)			County weighted average of log(bank branch access)	
	(1)	(2)	(3)	(4)	(5)
Branch density	3.612 (2.661)	4.002 (2.476)	2.787 (1.860)	-421.996 (29.371)	-455.377 (39.185)
Constant	3.780 (0.006)			3.306 (0.019)	
Observations	28,862	28,862	28,312	3,106	3,105
Adjusted $R^2$	0.000	0.441	0.836	0.046	0.620
State FE		O			O
County FE			O		

Each column reports coefficients from a univariate, unweighted OLS regression with heteroskedasticity-robust standard errors reported in parentheses. One observation is a Census tract (columns 1-3) or county (columns 4-5). Dependent variable observations are the weighted average of the natural logarithm of block-group-level estimated bank branch access measures through time, aggregated to either the Census tract or county. The natural logarithm of the estimated access measure per block group per year-month is  $\log \hat{\Phi}_{it}$  from Eq. (8). In computing the weighted average over time per block group, we take each month's weight as its share of the block group's total branch visitors over the core sample period (January 2018 - December 2019). Access measures are then aggregated to the Census tract or county level by weighting each block group's access measure by its share of population in either the Census tract or county to which it belongs. Independent variable observations are the branch densities of either Census tracts or counties, which are calculated as the number of branches in the area according to the 2019 FDIC Summary of Deposits (SOD) divided by the population of the area. Population counts are from the 2019 5-year American Community Survey (ACS). Columns (2) and (5) include state fixed effects and column (3) includes county fixed effects.

TABLE A.10  
BANK BRANCH ACCESS BY DEMOGRAPHIC ATTRIBUTES UNDER POSTAL BANKING

Dep. var.:	log(Bank branch access of block groups)							
	USPS branch quality:		Median				High	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Income)	-0.074 (0.003)	-0.085 (0.003)	-0.112 (0.002)	-0.080 (0.003)	-0.124 (0.003)	-0.089 (0.003)	-0.087 (0.002)	-0.093 (0.003)
Black	-0.059 (0.005)	-0.071 (0.006)	-0.082 (0.005)	-0.058 (0.005)	-0.104 (0.005)	-0.067 (0.006)	-0.048 (0.004)	-0.053 (0.005)
Asian	0.420 (0.013)	0.380 (0.013)	0.412 (0.012)	0.378 (0.012)	0.370 (0.012)	0.340 (0.012)	0.292 (0.010)	0.258 (0.011)
Other	0.022 (0.023)	0.077 (0.032)	0.021 (0.022)	0.009 (0.021)	0.074 (0.030)	0.064 (0.030)	-0.017 (0.019)	0.033 (0.027)
Hispanic	0.066 (0.007)	0.055 (0.008)	0.040 (0.006)	0.066 (0.007)	0.014 (0.007)	0.056 (0.007)	0.072 (0.006)	0.064 (0.006)
Age <15	-0.697 (0.017)	-0.786 (0.019)		-0.663 (0.016)		-0.752 (0.018)	-0.581 (0.014)	-0.666 (0.016)
Age 35-54	-0.243 (0.017)	-0.204 (0.020)		-0.230 (0.016)		-0.190 (0.019)	-0.198 (0.014)	-0.159 (0.016)
Age 55-64	-0.560 (0.018)	-0.571 (0.022)		-0.540 (0.017)		-0.539 (0.021)	-0.482 (0.015)	-0.466 (0.018)
Age 65+	-0.239 (0.013)	-0.253 (0.015)		-0.245 (0.012)		-0.256 (0.014)	-0.244 (0.011)	-0.252 (0.012)
log(No. of devices)	-0.051 (0.002)	-0.057 (0.002)	-0.048 (0.002)	-0.051 (0.002)	-0.053 (0.002)	-0.056 (0.002)	-0.051 (0.002)	-0.053 (0.002)
Observations	2,549,020	1,847,252	2,549,020	2,549,020	1,847,252	1,847,252	2,549,020	1,847,252
Adjusted $R^2$	0.666	0.578	0.671	0.675	0.592	0.599	0.715	0.675
Sample	Core	MC	Core	Core	MC	MC	Core	MC
Year-month FE	○	○	○	○	○	○	○	○
County FE	○	○	○	○	○	○	○	○
RUCA FE	○		○	○			○	

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 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. The log number of devices is SafeGraph's record of the number of mobile devices residing in the block group in the year-month. In all columns, the dependent variable is the log estimated bank branch access measure  $\log \hat{\Phi}_{it}$  from Eq. (8) 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 Section 5, with full details in Appendix A. In columns (1) and (2), we assign to each Post Office location per year-month an estimated establishment fixed effect  $\hat{\lambda}_{jt}$  equal to the 10<sup>th</sup> percentile of the distribution of estimated private bank fixed effects in the year-month. In columns (3)-(6), we assign the 50<sup>th</sup> percentile; and in columns (7) and (8), we assign the 90<sup>th</sup> percentile. Columns (1), (3), (4), and (7) include all block groups for which we have visitor data, whereas columns (2), (5), (6), and (8) restrict the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitan area core). The omitted demographic groups are non-Hispanic Whites and age range 15-34.