Bank Branch Access: Evidence from Geolocation Data*

Jung Sakong† Alexander K. Zentefis‡

May 15, 2024

Abstract

We investigate whether unequal access explains why low-income and Black households use bank branches less than high-income and White households, despite relying on them more. We obtain a measure of access from a gravity model of consumer trips to bank branches, estimated using mobile device geolocation data. We find no evidence that low-income communities lack access, and instead find that lower demand for branch products or services explains their lower branch use. But in Black communities, poor access explains their entire drop-off in branch use. The results spotlight areas of the country to best target policies expanding access to banking.

JEL classification: R20, D14, G21
Keywords: spatial economics, household finance, banking

*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 Bob Adams, Taha Ahsin, Treb Allen, Bronson Argyle, Costas Arkolakis, Helen Banga, Nick Barberis, Asaf Bernstein, Mehdi Beyhaghi, Vitaly Bord, Nick Buchholz, Lorenzo Caliendo, Claire Célérier, Judy Chevalier, Tony Cookson, Jess Cornaggia, Doug Diamond, Jonathan Dingel, Jane Dokko, João Granja, Cecilia Fieler, Raffi Garcia, Paul Grieco, Paul Goldsmith-Pinkham, Yaming Gong, Gary Gorton, Jessie Handbury, Sean Higgins, Xuan Hung Do, Jungbin Hwang, Stefan Jacewitz, Umang Khetan, Edward Kim, Kristoph Kleiner, Noura Kone, Sam Kortum, Cameron LaPoint, Simone Lenzu, Xiang Li, Runjing Lu, Pengfei Ma, Chuck Manski, Erik Mayer, Yuhei Miyauchi, Luciana Orozco, Christopher Palmer, Piyush Panigrahi, Karen Pence, Matthieu Picault, Roberto Robatto, Claudia Robles-Garcia, Rosa Sanchis-Guarner, Mario Samano, Shri Santosh, Katja Siem, Brad Shapiro, Kelly Shue, Mike Sinkinson, David Stillerman, Amir Sufi, Nancy Wallace, Brian Waters; and participants at various seminars and conferences for their very helpful comments. We 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. We are grateful to Bianca Battaglia for her exceptional copyediting. The views expressed in this paper are those of the authors and do not reflect those of the Federal Reserve Bank of Chicago or the Federal Reserve System. All errors are our own.

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

‡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 for low-income and Black households in the United States, but they visit them less often than high-income and White households, and do not offset their lower branch use with greater reliance on mobile and online banking. This disparity has prompted debates about unequal branch access as a central cause (Friedline and Despard 2016; Dahl and Franke 2017) that impedes certain households from obtaining the full welfare benefits of bank participation (Davidson 2018). As a remedy, policies have been proposed to increase branch access, like investing in community development banks (Ellwood and Patel 2018) and expanding U.S. postal banking (Baradaran 2013). However, research is divided on 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.[1]

In this paper, we attempt to make progress in this area 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 that low demand for branch products or services, not lack of access, drives lower use in low-income areas. But for Black communities, poor access fully explains their lower branch use. These results illuminate areas around 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 with a standard gravity equation, which consists of block group × time fixed effects, bank branch × time fixed effects, and the distances between pairs of block groups and branches. This approach takes advantage of the network of trips that link 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' overall demand for bank branch products

[1]Regarding conflicting evidence on bank branch access, Morgan et al. (2016) find minorities are less likely to live in "banking deserts" (areas with no branches within 10 miles), but Small et al. (2021) show branches are farther from minority neighborhoods than check cashers. Meanwhile, Goodstein and Rhine (2017) argue the overall influence of branch locations on branch use appears modest. For evidence of the welfare benefits of bank participation (e.g., improved access to credit, higher subjective well-being, greater liquid savings, reduced poverty, larger wealth accumulation), see, for example, Burgess and Pande (2005); Eisfeldt (2007); Fitzpatrick (2015); Prina (2015); Agarwal, Alok, Ghosh, Ghosh, Piskorski and Seru (2017); Melzer (2018); Célerier and Matray (2019); Brown, Cookson and Heimer (2019); Ji, Teng and Townsend (2023).

or services, then the estimated differences between block groups in branch use can plausibly be attributed to differential access.[2]

The gravity equation not only allows us to control for demand-related factors, but also generates a measure of bank branch access. From the gravity model, a block group's expected 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 schedules). 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 costs of traveling 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 (e.g., the branch having a modern design, higher staff attentiveness, or an efficient drive-through ATM).[3]

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 traditional supply-side measures of access from 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 only reveal consumers' opportunity set, this measure embodies consumers' actual choices.[4]

Applying this measure of access to a banking setting is the paper's second main contribution, as it can be used to examine other parts of household financial markets, such as access to credit unions, investment advisors, or payday lenders. Based on the median dwell times that visitors in the geolocation data spend at branches—ranging from less than 6 minutes to over 2.5 hours—the data appear to pick up both quick trips for straightforward transactions like depositing cash and long trips for complex transactions like small business borrowing. If lack of access deters vulnerable populations from visiting bank branches—without clear evidence of them fully compensating with alternative methods like mobile and online banking—these groups are at greater risk of bearing the welfare costs of reduced bank participation documented in the literature.

[2] 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).

[3] We correlate the estimated branch fixed effects with observable property characteristics, and the results suggest that the fixed effects are sensible proxies for branch quality: Branches with higher fixed effects have more square footage, higher property market values, higher price/sq. ft., longer store hours, and are open on weekends.

[4] The measure of access 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).

The central worry over uncovering the causal effect of access on branch visitation from the gravity equation is omitted variable bias. Namely, the distance between block groups and branches is endogenous: People may choose to live near particular kinds of bank branches, and banks may build their branches in certain areas to cater to particular kinds of clients. Hence, there may exist unobservable characteristics of residents and branches that correlate both with the distances between them and visits. To quantitatively assess the impact of possible selection on these unobservable characteristics, we add an extensive vector of controls to the gravity equation that attempt to account for endogenous location choices. The controls take inspiration from Balassa (1965)'s measure of "revealed comparative advantage," a measure also recently adapted in Paravisini, Rappoport and Schnabl (2023) for quantifying bank specialization. Overall, a sensitivity analysis reveals that the gravity parameter estimate is robust to the rich set of controls.

Beyond omitted variables, we face an additional identification challenge: The geolocation 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. Standard gravity model estimation methods, such as Poisson pseudo-maximum-likelihood (PPML), 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; Pakes and Pollard 1989). But the gravity equation has hundreds of 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 identify high-dimensional fixed effects. This econometric method is the paper's third main contribution, as the approach can be implemented in other empirical settings involving high-dimensional fixed effects in MSM estimation.

In the second part of the paper, we use the gravity model estimates to uncover how bank branch access covaries with the household incomes and racial makeups of local communities. Controlling for the racial and age population shares of block groups, we find that residents of low-income block groups have *better* branch access 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. We find that the higher access in low-income communities is entirely due to better branch proximity and not quality. Unlike those in low-income areas, residents of block groups with high Black population shares have *worse* access compared to residents of block groups with high White shares, controlling for a block group's median household income and the population shares of its residents' ages. The worse access for Black communities stems entirely from lower branch proximity and not quality. These results reconcile with the existing literature on banking deserts when recognizing that the paper's demographic comparisons of branch access are done at the granular block-group level. For example, Morgan et al. (2016) find that residents of majority-minority census tracts are less prone to live in banking deserts and argue that this finding reflects majority-minority tracts being situated mostly in cities. We find that, *within cities*, residents of predominately Black block groups live farther away from branches relative to their counterparts in predominately White block groups.

Finding that the Black-White gap in branch access is from branches being located farther away from Black communities, we attempt to shed light on some potential reasons for this disparity. To do so, we correlate the branch access gaps at the county-level with local crime indices (Nau, Sidell, Clift, Koebnick, Desai and Rohm-Young 2020) and measures of racial bias against Blacks among Whites (Xu, Nosek and Greenwald 2014; Chetty, Hendren, Jones and Porter 2020). We find that counties with higher expected risk of crime—especially robbery, murder, and assault—along with counties with higher levels of either implicit or explicit racial bias have larger Black-White gaps in 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. 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.

We next decompose the income gradient and the Black-White gap in branch use into constituent parts due to access and demand. We do so by 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, separating the elasticity of branch use 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. On the other hand, 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.

Near the end of the paper, we discuss several policy implications of the results. Of note, in a counterfactual exercise, we add post office locations to the set of available bank branches—akin to a national expansion of postal banking (Baradaran 2013; Sanders 2021; Gillibrand 2021)—under varying assumptions about the quality of postal banks (i.e., their fixed effects). In a low or medium quality system, the racial gap in access actually *widens* because post offices also tend to be located comparatively closer to White communities than Black communities, just like private banks. However, if the postal banks were of high quality, the Black-White gap in branch access shrinks, most significantly in big cities.

Related Literature. This paper relates to three 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. 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), Agarwal and Hauswald (2010), 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, as well as Bachas, Gertler, Higgins and Seira (2018), Bachas, Gertler, Higgins and Seira (2021), Agarwal, Mukherjee and Naaraayanan (2023), and Fonseca and Matray (2024), who do the same in other countries. Recent works have also meticulously studied how branch closures affect small business credit access (Nguyen 2019), how they lead to welfare losses for older depositors (Jiang, Yu and Zhang 2023), and how they became an unintended consequence of the Community Reinvestment Act (Cespedes, Jiang, Parra and Zhang 2024). Other recent papers have carefully examined how search frictions influence consumer loan access and choice (Argyle, Nadauld and Palmer 2023), how bank and retirement account participation varies across the U.S. (Yogo, Whitten and Cox 2022), and how Black depositors were afflicted by apparent access but actual fraud in the 19th century (Célérier and Tak 2023). An advantage of this paper's measure of financial access is that it encapsulates the actual choices of individual consumers, rather than reflecting survey responses or supply-side factors alone, such as the local branch density or the availability of low-cost accounts. And while the benefits of increasing access are well established, less is known about the extent to which prevailing U.S. inequities in access sustain disparities in bank use. We attempt to make 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 examined spatial differences in consumption access (e.g., Marshall and Pires 2018; Allcott, Diamond, Dubé, Handbury, Rahkovsky and Schnell 2019; Ellickson, Grieco and Khvastunov 2020; Eizenberg, Lach and Oren-Yiftach 2021). We take a closer look at spatial differences in access to financial services. Recently, several articles in the literature use geolocation data to answer economic questions. Research has looked at political partisanship (Chen and Rohla 2018), restaurant dining choices (Athey, Blei, Donnelly, Ruiz and Schmidt 2018), voting wait times (Chen, Haggag, Pope and Rohla 2019), segregation (Athey, Ferguson, Gentzkow and Schmidt 2021), firm boycotts (Hacamo 2023), social interactions (Büchel, Ehrlich, Puga and Viladecans-Marsal 2020; Miyauchi, Nakajima and Redding 2021; Kreindler and Miyauchi 2022; Atkin, Chen and Popov 2022), and the Covid-19 pandemic (Almagro, Coven, Gupta and Orane-Hutchinson 2021; Goolsbee and Syverson 2021; Chen, Chevalier and Long 2021; Couture, Dingel, Green, Handbury and Williams 2022; Coven, Gupta and Yao 2023). We complement this literature by using geolocation data to quantify spatial patterns and transportation costs in banking.

Third, the econometric method we introduce relates to estimation procedures used in the spatial economics and trade literatures. Novel methods to handle econometric issues when estimating gravity models have been proposed before, such as Tobit procedures, two-step Heckman procedures, and Poisson fixed-effects estimators to address zero flows (Eaton and Tamura 1994; Helpman, Melitz and Rubinstein 2008; Westerlund and Wilhelmsson 2011). 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, and its extensions 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. 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 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 et al. 2020). The paper's econometric method can be helpful in estimating models of economic environments that rely on privacy-preserving datasets.

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 use method (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 location of branches is most frequently cited as the most important reason for choosing an institution for a main checking account (43% of respondents). 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%). The 2016 SCF also reports that almost all of the 84 percent of households with bank accounts who visited a branch in the past year 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 Wadhwani 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). Furthermore, a 2018 Mercator Advisory Group survey found that 79% of small business owners visited a branch at least once a week and 24% visited daily (Augustine 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 F, 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. Based on raw averages, 61% of respondents in the lowest income bracket ($< \$15,000$) say that bank tellers/ATMs are their primary access method, compared to only 32% of respondents in the highest income bracket ($\$75,000+$). Similarly, 50% of Black respondents indicate that bank tellers/ATMs are their primary access method, compared to only 41% of White respondents. And when controlling for age and race in a multivariate linear probability regression, we find that respondents in the lowest income bracket are roughly 25% more likely than those in the highest income bracket 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.

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. Based on raw averages, 63% of respondents in the lowest income bracket say they visited a bank branch in the past twelve months, compared to 86% of respondents in the highest income bracket. Similarly, 69% of Black respondents say they visited a branch in the past twelve months, compared to 84% of White respondents. And when 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).

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.

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 the number of visitors from home Census block group i to bank branch j in time period t.[5] That gravity equation is:

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

Block group fixed effect. The first term in the gravity equation, γ_{it}, is a block group × time fixed effect that captures all characteristics of block group i's residents that contribute to them visiting *any* bank branch in the period. The population size, as well as residents' wealth and income, cash needs, extent of financial sophistication, amount of trust in banks, and general methods of transportation, are all encoded in the block group fixed effect. Because it embodies all block group-specific factors that promote residents to visit any bank, and is orthogonal to the supply-side characteristics of any particular branch, we refer to a block group fixed effect (albeit informally) as proxying for block group residents' "demand" for physical branch products or services. More precisely, it is the *residual* demand not met by other bank access methods—like online, mobile, and telephone banking, or visiting offsite ATMs—that residents optimally choose to satisfy in-person at a branch. The geolocation data do not reveal the precise activity that a customer partakes at a branch, but we

[5]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. Eq. (1) can be derived from a differentiated product, discrete choice model of consumers selecting branches to visit per period. Online Appendix I provides one simple model of that sort. We specify Eq. (1) in levels instead of in logs because the estimation procedure described in Section 5 is run in levels to account for the mobile device data's differential privacy distortions.

do know median dwell times spent there, which give some clues. From Table 1, the 10-90 quantile range is 6 minutes to 2.5 hours. Thus, the data likely include a wide scope of activities, such as depositing cash, obtaining a loan, or requesting an international wire transfer.

Branch fixed effect. The second term, λ_{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. The bank's brand, the efficiency and courteousness of the branch's staff, the expertise and experience of financial advisors, the hours open, the size and design of the physical building along with its furnishings, and the availability of amenities like safety deposit boxes and drive-through ATMs are all encoded in the branch fixed effect. Even the excellence or mediocrity of the bank's online and mobile offerings is embedded, since this could affect a customer's choice to visit in-person as a substitute. Because it embodies all branch-specific factors that draw in visitors from any origin, and is orthogonal to the demand-side characteristics of any particular block group, we refer to a branch fixed effect (again informally) as proxying for the bank branch's "quality." Nevertheless, a potential concern with this interpretation is that λ_{jt} conflates two opposing effects: High quality branches may attract more customers from farther away, but low quality, inefficient branches requiring multiple visits to solve issues may also attract more visitors. To better understand what drives the branch fixed effects, we study their correlation with observable property characteristics in Online Appendix D. The fixed effects indeed align with measures of quality: Branches with higher fixed effects have more square footage, higher property market values, higher price/sq. ft., longer weekday hours, and are open on weekends.[6]

Bilateral travel barrier. The third term, $\beta_t \log\left(\text{Distance}_{ij}\right)$, is the bilateral travel barrier between block group i's residents and branch j. It combines travel costs, captured by $\log\left(\text{Distance}_{ij}\right)$, with their respective elasticity, β_t, to measure the overall impact on visitor flows.[7] 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

[6]To allay concerns that the estimated branch fixed effects are mechanically higher in more populated areas, we regress them in Online Table A.5 on the population densities of the census tracts in which the branches are located and find a negative and imprecisely estimated relationship.

[7]One potential concern is that the elasticity β_t should vary by demographic attributes. Online Table A.6 presents PPML estimates of Eq. (1) using the raw geolocation data, allowing β_t to differ by income and race separately. Residents of low-income block groups and Black communities exhibit precisely estimated lower values of β_t, but the differences are relatively small. For this reason, we let β_t vary over time but keep it constant across block groups. A homogeneous β_t also lets branch proximity and branch quality determine a block group's branch access, instead of differences in residents' elasticity to distance, which may arise from some groups like low-income and Black communities relying more on branches and less on alternative banking methods like mobile or online.

distances between branches and block groups' centers of population. (See Footnote 39 in the Online Appendix for details.)[8] Finally, Distance_{ij} implicitly presumes that residents visit branches directly from home. The geolocation data we use do not reveal the exact travel paths of visitors, but only their home block groups. Hence, the branch access measure below is best thought of as capturing access around residents' homes. We know of no accessible data on the commuting or shopping patterns of U.S. residents at the block group level. But to the extent that residents of low-access areas visit branches in high-access areas where they work or shop, the access measure will be underestimated for those groups.[9]

Identification. Identification of β_t as the causal effect of distance on branch visitation hinges on unobserved determinants of actual visitor counts being orthogonal to Distance_{ij}. However, Distance_{ij} is endogenous to where people choose to live and where banks choose to build their branches. For example, existing Chase account owners may elect to live in a town because it has many nearby Chase branches; or, conversely, Chase may erect a branch with private wealth management services in an area if it anticipates that these services are attractive to residents there. The same can be said of minority depository institutions (MDIs) that are located in underserved neighborhoods precisely to provide for those communities. Unobservable characteristics of residents and branches may correlate both with the distances between them and visits, which would render Eq. (1) subject to an omitted variables problem.[10] In Online Appendix B.2, we quantitatively assess the impact of possible selection on these unobservable characteristics by adding a rich set of controls that try to proxy for the endogenous location choices of residents and branches. These controls measure the relative shares of types of visitors over others at a branch, reminiscent of Balassa (1965)'s "revealed comparative advantage." If people live near a branch that caters to them, and if that branch is located near the customers it wishes to cater to, then we should expect to see disproportionate shares of certain types of visitors over others at that branch. We follow the common approach of assessing the sensitivity of the estimated parameter of interest (the gravity

[8] An alternative to haversine distance is the road driving time between locations. In Online Appendix Table A.8, we regress the driving times between about 1 million random block groups and bank branches onto the corresponding haversine distances. Regressions are run across the entire 1 million observation sample and over subsample block groups associated with various demographic attributes. Across all samples, the regression adjusted R^2s are very high, and the variation in driving time with respect to haversine distance is stable. Haversine distance is computationally easier to calculate, and these results suggest that it a good substitute for driving time.

[9] See Relihan (2022) for a remarkable recent analysis of the online and offline shopping habits of millions of customers using JP Morgan's proprietary transactions data.

[10] Notice that if people, upon moving, *became* Chase account owners and visited Chase branches because their new neighborhood was populated with Chase branches, this effect would be attributed to the causal effect of distance.

coefficient) to the inclusion of the observed controls using OLS and PPML. Overall, the gravity coefficient estimate is robust to the rich vector of controls.[11]

Functional form. Eq. (1) imposes a log-linear relation between visitor counts and distance. Fig. 1, Panel A depicts a clearly negative and fairly linear relation between distance and visitation in the geolocation data. The relation does flatten out (i.e., becomes nonlinear) when the log number of visitors approaches 1.4, but that change corresponds to the geolocation data's privacy protections that bottom code the visitor counts at 4. When only block group × branch pairs with >4 visitor counts are plotted in Panel B, a near-linear relation retains throughout. A log-linear gravity model as in Eq. (1) is standard in the spatial economics literature, and Fig. 1 suggests that it represents the geolocation data quite well.

Bank branch access. To arrive at a measure of branch access using the gravity model, we sum across all bank branches available to visit in the period. Doing so gives block group i's expected number of visitors to any branch in the period:

$$E\left(\text{No. of visitors}_{it}\right) = \exp\left(\gamma_{it}\right)\Phi_{it}. \tag{2}$$

The term Φ_{it} is defined as:

$$\Phi_{it} \equiv \sum_{j \in B_t} \exp\left(\lambda_{jt}\right)\left(\text{Distance}_{ij}\right)^{-\beta_t}, \tag{3}$$

where B_t is the set of bank branches open in period t across the country.

Eq. (3) is the paper's 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, Φ_{it} can be interpreted as an attribute-adjusted branch index that is unique to each block group. Each branch in a block group's index is represented by (1) the "quality" of the branch's attributes in the period, as measured by its fixed effect, λ_{jt}, and (2) the branch's distance away. Local areas have better access if bank branches are relatively closer, especially branches with better attributes.[12]

[11] Paravisini et al. (2023) recently use similar Balassa (1965)-style share measures to study bank specialization in lending.

[12] The object Φ_{it} is conceptually related to what some in the economic geography and trade literature have described as an exporting country's "access" to the importing markets of other countries (e.g., Harris 1954; Head and Mayer 2004; Redding and Venables 2004; Hanson 2005; De Sousa et al. 2012; Donaldson and Hornbeck 2016; Fajgelbaum and Gaubert 2020; and Adão et al. 2020).

4 Geolocation Data on Branch Visitors

Branch visitors are based on monthly, anonymous geolocation data from mobile devices between January 2018 and December 2019. The data provider is the firm SafeGraph. We do not use the "raw" pings from individual mobile devices, but instead, 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 bank branch visitors per month by their home Census block groups (i.e., the network of consumer trips from home block groups to bank branches).

The aggregated data are benefited by SafeGraph's algorithms that pinpoint a device owner's home origin and estimate whether the owner visits a particular branch. However, a limitation is that the data do not give the demographic attributes of the mobile device owners, the starting points of their trips, their individual durations spent at a branch, nor what they do at the branch. A unique visitor in the data is identified by a mobile device, one device is treated as one visitor, and a device must spend at least 4 minutes at a branch to qualify as a visitor. Online Appendix C provides background information on the SafeGraph data and a detailed explanation of how we construct our primary sample.[13]

4.1 Primary Sample

Our primary (core) data set includes bank branches in all 50 states and the District of Columbia. To ensure that we only analyze depository institutions, 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 FDIC's 2019 Summary of Deposits (SOD). 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 were typically more accurate.[14]

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

[13]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 (SafeGraph 2018-2019).

[14]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.)

mobile device visits to particular branches for two reasons. First, in dense environments such as multi-story buildings or strip malls, SafeGraph might lack confidence about the geometric boundary of a place. To reduce false visitor 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 aggregates visitors 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.[15]

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 Online Appendix B.1, 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 (and median) of the estimated fixed effects of branches in SafeGraph within the period. Including all SOD branches strengthens the main findings on how access varies by race and income.[16]

4.2 Sampling Bias

Our core sample experiences two types of sampling biases: (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

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

[16] We focus our analysis on commercial banks in this paper and leave for follow-up work the study of access to other depository institutions like credit unions, and non-traditional financial institutions, like check cashers, pawnbrokers, and payday lenders.

locations. First, SafeGraph adds Laplace noise to all positive counts of visitors to a branch from each home Census block group of the branch's visitors. Second, they round each of these block group × branch visitor counts down to the nearest integer. Third, they drop from the data all rounded visitor counts less than 2. Fourth, if a rounded visitor count equals 2 or 3, they 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 varies 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 in Eq. (1) using an econometric method that adapts the Method of Simulated Moments to estimate high-dimensional fixed effects. Online Appendix A details the full procedure, and Section 5 provides a summary.

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: the set of branches and the set of visitors. To address potential sampling bias from missing around 40% of U.S. branches, Online Table A.1 compares 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 economically small compared to the mean values across areas. In addition, we conduct a robustness check on access in Online Appendix B.1, where we include all branches in the 2019 SOD, and the paper's main findings on access are strengthened.

Regarding the sample of visitors, our core sample includes 215,686 unique visitor home Census block groups, and the 2010 U.S. Census records 217,740 block groups, implying close to complete coverage of U.S. local home areas. SafeGraph aggregates data from around 10% of all mobile devices in the country, and we calculate about 30 million unique mobile devices visiting all establishments recorded in SafeGraph, and 1.6 million visiting bank branches per month on average. Online Fig. A.1 presents a time-series of the number of branch visitors over the sample period. Several other researchers have documented that geolocation data from mobile devices is broadly representative of the general population (Squire 2019; Chen and Pope 2020; Athey et al. 2021; Couture et al.

2022). To gauge our sample's representativeness, in Online Appendix F.3, we compare the share of households in the 2019 FDIC survey who reported visiting a bank branch in the previous 12 months to the share of mobile devices in SafeGraph that visited branches in the same period, with the comparison done by household income. There is a strong resemblance between the two sources, as both reported and observed branch visitor shares are increasing and concave in household income.

Nevertheless, we cannot rule out non-random sampling of mobile devices based on unobserved characteristics of visitors. As we discuss in Online Appendix G, 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, thus facing an ecological inference problem (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 slightly 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 63.7% among the unbanked. We likely under sample these groups with lower mobile device ownership rates.

Even so, lower smartphone ownership rates among low-income households, Black households, and the unbanked 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 to which we find that they also visit branches less in our sample is reasonably an underestimate. Furthermore, 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.

Finally, SafeGraph may misattribute a mobile device to a home block group. To account for possible misattribution, we weight block-group level regressions of branch access, branch demand, and branch use in Sections 6 to 7 by the 2019 5-year American Community Survey (ACS) block group population counts, thereby down-weighting block groups with disproportionately high mobile devices and up-weighting block groups with disproportionately low mobile devices.[17]

[17]Indeed, Thaenraj (2021) identifies around 1,000 Census block groups in the SafeGraph data that register more

4.3 Descriptive Statistics

Table 1 reports descriptive statistics of our core sample. On average, the typical branch has 40 unique visitors per month, but there is wide dispersion across branches, as the standard deviation of visitors is over twice as high at 94.[18] The median distance traveled to a branch is 5 miles on average, 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. Hence the sample likely includes quick trips for straightforward transactions (e.g., ATM cash withdrawals, check cashing, and making change for a $100 bill), as well as long trips for complex financial transactions (e.g., opening a new account, applying for a home mortgage or small business loan, receiving personalized retirement planning). With the median dwell time on average being 49 minutes, the distribution of trips may be skewed towards more of these complex transactions.[19]

5 Gravity Model Estimation

SafeGraph's differential privacy methods bias any OLS or PPML estimation of Eq. (1). In this section, we summarize our alternative econometric method, which adapts the Method of Simulated Moments (MSM) to identify high-dimensional fixed effects. Online Appendix A has the full details.

5.1 Econometric Method

The method's goal 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, which 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,

devices residing there than the number of people living there according to the Census. 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 (Squire 2019).

[18] While the number of unique visitors per branch may appear low, it is important to keep in mind that SafeGraph includes roughly 10% of smartphones in its sample, and so, the population average number of unique visitors per branch may be closer to 400 per month (about 20 per business day). Also from the table, on average, a typical branch has about 67 visits per month, making the ratio of total visits to unique visitors about 1.7. This ratio is close to the average number of distinct branches (2.3) that people from a typical block group visit per month.

[19] A caveat is that we cannot distinguish branch employees from customers in the sample of visitors, and the presence of the former may raise the median dwell time. Nevertheless, the number of employees relative to customers in the sample is likely to be low and have a small impact.

giving:

$$V^{+}_{ijt} = V^{*}_{ijt} + L_{ijt}. \tag{4}$$

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

$$z_{ijt} = \begin{cases} 1 & \text{if } \lfloor V^{+}_{ijt} \rfloor \geq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

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

$$V_{ijt} = \max\left\{4, \lfloor V^{+}_{ijt} \rfloor\right\}, \tag{6}$$

The econometric method (i) simulates "true" visitor counts, V^{*}_{ijt}, from a presumed data-generating process that follows the gravity model, (ii) manipulates the simulated data according to Eqs. (4) to (6), and then (iii) chooses the gravity model parameters to minimize the distance between selected moments of the (manipulated) simulated data and moments of the SafeGraph geolocation data. 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.[20]

Specify the data-generating process. In the simulations, we specify V^{*}_{ijt} as Poisson distributed. We presume a Poisson model instead of 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). Using the gravity model in Eq. (1), we express the true visitor count as following

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

Sample block group × branch pairs. The sample of over fifty-thousand branches and over two-hundred-thousand block groups, altogether spanning twenty-four months, makes it computationally impractical to include all the billions of block group × branch pairs in Eq. (7). We instead include

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

only stratified sampled pairs. If a block group × 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 × branch pair in our simulation with probability 1. If a block group × 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 × branch pairs such that (i) every pair in the alternative set has the same probability of being sampled, and (ii) 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 × branch pairs represents slightly higher than a 0.05% sample size of all possible block group × branch pairs with missing visitor counts. A larger sample size of 0.1% did not alter the estimation results. We apply probability weights to block group × branch level variables to rebalance the stratified sampled data, and the weights are the reciprocal of the sampling likelihood, following standard practice. The stratified sampling implies that *all* branches in the sample per period are commonly available to *all* households.[21,22]

Simulate the visitor counts. We simulate visitor counts per Eq. (7) and apply the differential privacy methods expressed in Eqs. (4) to (6). The simulation process differs between the block group × 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 from Eq. (7) and then apply Eqs. (4) to (6). For the block group × 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 multiply them. Noise from the simulation would be amplified and make the estimation unstable. Rather than randomly drawing these pairs of visitor counts, we construct

[21] 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.

[22] Rather than using a stratified sample, 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 groups). 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 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. Including all branches in the choice sets of residents will also let us uniquely pin down the branch fixed effects, which we discuss momentarily.

their implied empirical probability distribution given the parameter estimates. If an infinite number of observations were in fact simulated for these pairs, their distribution would coincide with this constructed empirical distribution. 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.[23]

Iterate the fixed effects until convergence. Estimating the hundreds of thousands of fixed effects in Eq. (7) using the MSM minimization problem alone would be computationally impractical. Instead, we adopt an iterative routine to identify the fixed effects $\left\{\gamma_{it}, \lambda_{jt}\right\}$ and let the minimization problem identify β_t. First, given an estimate of β_t and estimates of the block group fixed effects, $\{\gamma_{it}\}$, we uniquely pin down each branch's fixed effect, λ_{jt}, by utilizing another data field in SafeGraph that is unaffected by differential privacy: a branch's total number of visitors. Because the stratified sampling implicitly presumes that visitors can arrive from any block group, we can identify each branch's fixed effect from an "adding up" condition. Namely, 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 we invert the equation to extract the branch's fixed effect. Second, given estimates of β_t and the branch fixed effects, $\left\{\lambda_{jt}\right\}$, 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. When the number of fixed effects in each dimension is large, as in our setting, this routine produces consistent estimates.[24]

[23] We cannot apply this approach to the set of block group × 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 × 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.

[24] With every update to the block group fixed effects, $\{\gamma_{it}\}$, the branch fixed effects, $\left\{\lambda_{jt}\right\}$, also update from the inversions. The iterative process we use 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).

Construct the MSM estimator. After both dimensions of fixed effects are identified per estimate of β_t, the MSM minimization problem then selects the optimal β_t estimate that minimizes the weighted 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 × 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). The simulated visitor counts include all positive draws from all simulations across every year-month in the sample period. 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 "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; Pakes and Pollard 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. SafeGraph's data censoring levels off the log observed visitor counts at 1.4, which corresponds to 4 visitors. The company's data 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, such that 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 that are computed using the MSM standard errors. 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. In comparison, Agarwal, Jensen and Monte (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. Fig. 4, Panels B and C present histograms of the estimated block group and bank branch fixed effects across all months of the sample period. Roughly 35% (77%) of the variation in a block group's (branch's) fixed effect over time can be explained by the block group (branch) itself. This suggests that block group demand tends to vary over time, but branch quality is fairly stable.[25]

Online Table A.7 presents gravity coefficient estimates and standard errors from the MSM estimation, along with estimates and standard errors from traditional OLS and PPML estimations using the "raw" visitor counts, both per year-month and over the full sample period. Whereas the MSM gravity coefficient estimates range from -1.45 to -1.26, the OLS estimates range from -0.062 to -0.038, roughly twenty to thirty times smaller in magnitude. The PPML estimates register higher magnitudes than the OLS ones, ranging in values from -0.108 to -0.066, but they are still roughly ten to twenty times smaller in magnitude than the MSM estimates. Overall, the table reveals the downward bias that SafeGraph's differential privacy methods introduce to traditional methods of estimating the gravity equation, stressing the need for the alternative econometric method.

[25]These values are R's from panel regressions of γ_{it} on i fixed effects, and separately, λ_{jt} on j fixed effects.

6 Bank Branch Access in the United States

With the gravity model estimates, we next compute the empirical counterpart of Eq. (3), which is the branch access for residents of block group i in year-month t:

$$\hat{\Phi}_{it} \equiv \sum_{j \in B_t} \exp\left(\hat{\lambda}_{jt}\right)\left(\text{Distance}_{ij}\right)^{-\hat{\beta}_t}. \tag{8}$$

The magnitude of $\hat{\Phi}_{it}$ has economic meaning as a block group's expected total number of branch goers per month, when branch demand is equalized across block groups (i.e., block group fixed effects are normalized 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 geographic heterogeneity over the U.S. and by measuring its association with the 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 U.S., where each dot is positioned at a block group's center of population. Throughout the country, three patterns are apparent. First, access varies substantially across regions. The eastern shores of New England, the Mid Atlantic, and the upper Midwest experience the highest access nationwide, whereas the Deep South observes the lowest access. A population-weighted, block-group-level regression 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. Online Fig. A.2 shows that access declines monotonically as one transitions from Metropolitan core to Metropolitan suburb, micropolitan suburb, and rural areas, such that access in the typical rural area is about two-thirds less than access in a typical Metropolitan core.[26] Third, even within a local area, branch access varies significantly. A population-weighted, block-group-level regression 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, access is highest for residents of Midtown Manhattan and lowest for those living in the Bronx, and within each borough there is large variation. In Greater

[26]The U.S. Department of Agriculture's Economic Research Service provides Rural-Urban Commuting Area codes that 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.

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 the Palos Verdes Peninsula, despite that latter neighborhood also 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. 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 match the decile break points of the distribution of branch access nationwide. Finally, in Houston, access is highest for residents living downtown, and declines as one moves farther away from the city.[27]

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 investigate Metropolitan core areas, which exhibit Black population shares close to the national average. Studying access in these big cities is essential to explaining why Black households use branches less.[28]

Table 2 presents weighted OLS regressions of $\log \hat{\Phi}_{it}$ on demographic attributes of block group residents. 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 are non-Hispanic Asian, non-Hispanic Black, non-Hispanic White, non-Hispanic Other Races, and Hispanic. To make the notion of "distance" as comparable as possible across different types of areas (urban, rural, and suburban), we add Rural-Urban Commuting Area (RUCA) fixed effects to our national-level specifications. 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.

[27] A common measure of bank access is an area's density of branches. In Online Table A.16, 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 the paper's measure of bank branch access captures more information than just branch density. 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.

[28] From the 2019 5-year ACS, the national Black share is 12%. The Rural-Urban 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.

Column (1) conditions the branch access regressions on median household income and population racial shares. Residents of block groups with low median household income experience *better* access. If a block group's median household income doubles, its access drops by roughly 11%, a negative income gradient. However, Black communities have *worse* access: Residents of a block group with a 100% Black population share observe 8.2% poorer access than residents of a comparable block group with a 100% White population share. Because differences in financial savvy or technical sophistication from age differences might influence branch visits (Caskey and Peterson 1994; Hogarth, Anguelov and Lee 2005; Rhine and Greene 2013), we add age shares in column (2). Controlling for age 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 Black communities still have worse access, on the order of 5.3%.

Column (3) restricts the sample to block groups in Metropolitan core areas. The negative coefficients on income and the Black share sharpen. 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).[29]

6.3 Decomposing Access: Branch Proximity and Distance-Adjusted Average Quality

The measure of bank branch access combines information about the quality of branches available and the cost of traveling to them. We next evaluate how each component contributes to differential access. One may wonder, for instance, whether Black communities have lower branch access because there are no branches near them or because their nearest branches are lower quality.

Bank branch access in Eq. (8) can be rewritten as

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

[29]The results in Table 2 reconcile with the earlier literature on banking deserts, such as Morgan et al. (2016), when recognizing that here, the demographic comparisons of branch access are done at the very local block-group level within a county and within a type of area (e.g., city, suburb, rural town). For example, Morgan et al. (2016) find that residents of majority-minority census tracts are less prone to live in banking deserts and argue that this finding reflects majority-minority tracts being situated mostly in cities. Here, we find that, *within cities*, residents of predominately Black block groups live farther away from branches relative to their counterparts in predominately White block groups.

where the weights $\pi_{ijt} \equiv \frac{d_{ij}^{-\hat{\beta}_t}}{\sum_{k \in B_t} d^{-\hat{\beta}_t}}$. Eq. (9) separates access into two components. The first is *branch proximity*, measured as the sum of the inverse of the transportation costs for residents to reach bank branches across the country. The closer people are to bank branches, the higher their branch proximity. It can be interpreted as residents' hypothetical access if branch quality were equalized across all branches (and normalized to one). The second component is *distance-adjusted average 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 2 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 distance-adjusted average 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 higher quality branches.

Focusing on the racial differences, we find that Black communities experience significantly lower branch proximity, 14.3% lower nationwide and 16.8% lower in Metro cores. However, both nationwide and in Metro cores, members of Black communities—after having made longer-distance trips to branches—experience higher distance-adjusted average quality (9.0% higher cross-country and 10.4% higher in big cities). Yet this higher distance-adjusted average quality falls short of the lower branch proximity. Overall then, reduced proximity to branches is the central explanation for the lower branch access in Black communities.[30]

[30]It is important to stress that the results in columns (7)-(8) of Table 2 do not imply that the branches *near* Black communities have higher average quality. Indeed, Online Table A.5 shows that the correlation between a branch's estimated fixed effect and the Black racial share of the block group where the branch is located is negative and imprecisely estimated. Rather, the results in Table 2 reveal that residents of Black communities travel farther to reach branches of higher average quality.

6.4 Correlates of Black-White Gaps in Branch Access

The previous section established that the reason why Black communities lack branch access is because branches are located relatively farther away from them. Fully explaining why these racial gaps in access exist is beyond the scope of this paper. But to shed some light on them, we run the specification in column (2) of Table 2 county-by-county, so that the estimated coefficients on the Black population share are county-specific Black-White gaps in branch access. We then study the correlation of these racial gaps with county-level measures of neighborhood crime and racial bias against Blacks among Whites.

Neighborhood Crime. We first examine whether the Black-White gap in branch access is associated with local indices of crime, with the full analysis described in Online Appendix E.1. The indices are from the database CrimeRisk, are generated by Applied Geographic Solutions (AGS), and are distributed by Esri.[31] CrimeRisk provides indices for several categories of personal and property crime, where a score of 100 implies that a neighborhood's expected risk for that crime matches the national average. We focus on the five indices that Nau et al. (2020) validated using LAPD crime rates: robbery, murder, assault, motor-vehicle theft, and personal crime (which includes the first three categories and rape). Online Table A.9 presents the associations between neighborhood crime risk and estimated Black-White gaps in branch access across counties. In all cases, counties with higher expected crime risk have larger Black-White gaps in branch access. For example, in counties having a 10% higher expected risk of robbery, the access gap is higher by 0.02. By comparison, the Black-White access gap nationwide is 0.053 (column 2 of Table 2). One explanation for this positive correlation is that higher local crime raises the cost of operating a branch due to greater investments in security, making banks reluctant to enter. But another possibility is that other factors drive both higher expected crime risk and the relative dearth of bank branches in these areas.

Racial bias. We next investigate whether the Black-White gap in branch access is associated with racial bias, where the full details are provided in Online Appendix E.2.[32] We consider a measure of *implicit* racial bias from the Implicit Association Test (IAT) in Greenwald, McGhee and Schwartz

[31] The primary source of CrimeRisk is the FBI Uniform Crime Reports (UCR), which compile crime statistics from 18,000 law enforcement agencies across the U.S., mainly in the largest cities, counties, and metro areas. An advantage of the CrimeRisk database is that it covers all block groups in the country, albeit it relies on a predictive model to estimate local crime index values.

[32] Prior work has studied the adverse consequences of racial bias, for example, on Black boys in school (Simpson and Erickson 1983; Chavous, Rivas-Drake, Smalls, Griffin and Cogburn 2008), how racial bias is associated with intergenerational mobility (Chetty et al. 2020), and how it changes after Black electoral victories (Sakong 2023).

(1998)—which measures the difference in participants matching positive and negative words with Black versus White faces—and a measure of *explicit* bias that measures the difference in participants' answers to whether they "feel warmer toward" Whites versus Blacks. A value of zero for either measure represents no racial bias against Blacks, and higher levels imply greater racial bias. We obtain data on the two measures for non-Hispanic White test participants at the county level from Project Implicit (Xu et al. 2014). Because of potential selection bias from test participation being voluntary, we also create "adjusted" racial bias measures that are the residuals from projecting the two measures on respondent demographics and test variables (i.e., month, hour, weekday, and order of test). Online Table A.10 and Online Table A.11 present the associations between implicit and explicit racial bias, respectively, and estimated Black-White gaps in branch access across counties. In all cases, counties with higher racial bias against Blacks among Whites have larger Black-White gaps in branch access. For instance, in counties with a one standard deviation higher level of implicit racial bias against Blacks, the access gap is higher by 0.18.[33] Again by comparison, the Black-White access gap nationwide is 0.053. This positive correlation has several potential explanations. One is that local bank managers are reluctant to recommend new branch sites (or eager to close underperforming branches) in these areas because of racial bias. An alternative explanation is that racial bias adversely affects the incomes, employment levels, or wealth of local residents, making it less profitable for branches to enter. A third possibility is that racial bias is associated with other underlying factors that contribute to these correlations.

7 Bank Branch Use: Access versus Demand

The gravity model isolates how the unequal spatial distribution of bank branches explains disparities in household branch use. Since the comparison is across branches for the *same* block group, block group-specific demand is absorbed by the block group fixed effects. As an added benefit, the model generates a simple decomposition of branch visitor counts into two parts. Summing the log predicted means of Eq. (7) across branches (i.e., $\log \hat{V}^*_{it} \equiv \log \sum_j \hat{V}^*_{ijt}$) delivers the expected total number of residents of block group i who visit any branch in year-month t:

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

[33]The cross-county standard deviation of implicit racial bias is 0.037.

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 were equalized. This decomposition thus separates the parts of branch use that are due to demand versus access.

One may 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

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

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

$$\hat{\gamma}_{it} = X_i \theta_\gamma + \varepsilon_{\gamma,it}, \tag{12}$$

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

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

$$\hat{\theta}_{V,x} \equiv \hat{\theta}_{\gamma,x} + \hat{\theta}_{\Phi,x}, \tag{14}$$

which separates the elasticity of branch use with respect to the demographic attribute into 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 and differential demand.

7.1 Bank Branch Use

Table 3 presents weighted OLS regressions of bank branch visitors and branch demand by demographic attributes, both nationwide and in Metro cores. In columns (1)-(4), the dependent variable is a block group's log expected number of branch visitors, $\log \hat{V}^*_{it}$. 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). However, Black communities have fewer expected branch goers compared to White communities (4.1% fewer branch goers per month moving from a block group with a 100% Black share to a comparable block group with a 100% White share). These observed income and racial differences in branch use are consistent with the reported FDIC survey evidence. 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 population share is more sharply negative (-5.6%). Columns (3) and (4) focus on Metro core areas, and 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 Black-White gap in branch use drops slightly to 3.9% in big cities.

7.2 Bank Branch Demand

In columns (5)-(8) of Table 3, the dependent variable is a block group's fixed effect $\hat{\gamma}_{it}$. 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 access were equalized across the country and normalized to one, this range would correspond to the decline in the block group's expected number of branch goers per month.[34,35]

Turning to racial differences in branch demand, we find that Black communities observe higher fixed effects than White communities (4.1% higher) in column (5) with just an income control, but when age controls are added in column (6), the sign of the coefficient on the Black share turns

[34] 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 or services overall. This relation contrasts with the small and negative loading of branch access on the log number of devices in Table 2.

[35] 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 respondent's income 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.

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, Black communities do observe precisely estimated higher demand without age controls (8.9% higher in column 7), and the coefficient stays precisely estimated once age controls are added (2.5% higher in column 8).

7.3 Decomposing Branch Use

We next decompose the variation in branch use into parts due to differences in access versus demand by inserting the estimates from the branch access regressions of Table 2 and the estimates from the branch use and demand regressions of Table 3 into the identity of Eq. (14).

Nationwide, the income gradient in branch use from column (2) of Table 3 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 2. 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 leads to their lower overall branch use. When it comes to racial disparities, column (2) of Table 3 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 3) is +15.5%, which can be separated into an income gradient in demand of +24.2% (column 8 in Table 3) and an income gradient in access of -8.7% (column 4 in Table 2). 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, producing a 3.9% Black-White gap in branch use.

8 Discussion of Policy Implications

Many researchers and policymakers have proposed programs to increase bank access by enlarging the physical presence of branches (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). The paper's results indicate these policies would have their largest impact in Black communities, particularly those in big cities.

In Online Appendix H, we investigate U.S. postal banking through a counterfactual exercise. In September 2021, the U.S. Postal Service launched a pilot program in four post office locations that allowed customers to cash payroll and business checks as gift cards (Heckman 2022). To study how a national expansion of the policy would affect branch access, we re-estimate each block group's access after adding all post office locations registered in SafeGraph to the set of available branches. We maintain the previously estimated time-series gravity coefficients $\hat{\beta}_t$, the estimated fixed effects for private banks $\hat{\lambda}_{jt}$, and the estimated block group fixed effects $\hat{\gamma}_{it}$. In the counterfactual, we explore three distinct scenarios for the fixed effects of postal banks. Each scenario has postal branches featuring a different quality based on a percentile of the distribution of private branch fixed effects in each year-month. The scenarios are "low-quality" (postal branches have the 10^{th} percentile of private branch fixed effects), "medium-quality" (50^{th} percentile), and "high-quality" (90^{th} percentile). This analysis, similar to a "partial trade impact" evaluation in the trade literature dealing with policy changes in trade costs like tariffs (Head and Mayer 2014), does not account for general equilibrium effects resulting from the policy change.

In the low- and medium-quality scenarios, expanding postal banking *worsens* the Black-White gap in branch access. This is because post offices are typically closer to White communities, just like private banks, and adding banking services to post offices would benefit them more. Only a high-quality postal banking system would narrow the Black-White gap, reducing it by 0.5 percentage points per month nationally (5.3% to 4.8%) and 1.1 percentage points (6.4% to 5.3%) in Metro core areas. Because block group fixed effects retain their same values from before, any change in the racial gap in access would equal the change in branch use between groups.

The partial policy analysis ignores changes in demand for branch products or services that postal banking could spur, a limitation of the exercise since distaste for private banks may explain the low estimated block-group fixed effects of low-income communities.[36] Postal banking proponents argue that the unbanked perceive post office branches as more trustworthy than private banks (Office of the USPS Inspector General 2014; Baradaran 2015). If this is the case, the policy could improve bank participation by raising demand for branch products or services, which would benefit low-income communities the most based on the paper's main analysis. Postal banking could also prompt a competitive reaction from private banks to attract the unbanked and underbanked, leading to an endogenous change in their branch fixed effects, something we also omit.

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. It is beyond the paper's scope to fully explain why private banks do not enter more into these areas. The neighborhood crime and racial bias correlations in Section 6.4 may give some clues, 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 so, another policy that could remedy weak branch access would be allocating tax credits to banks that establish new branches in Black communities, or subsidizing both community development banks and minority depository institutions (MDIs) to further expand there (Hurtado and Sakong 2022).

Beyond expanding physical branches, 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 branches (low-income and Black households) simultaneously rely on branches *more* as their primary method of banking, rather than mobile or online, and weak access to broadband could be a reason why.[37] Expanding broadband could also raise consumer access to Fintech firms that offer services entirely online. While a purported goal of Fintech is increasing financial inclusion, evidence so far on achieving that goal is mixed, and more time is needed to fully evaluate Fintech's impact on closing disparities in financial access.[38] Even so, in every wave of the Survey of Consumer Finances since 1989, the location of a bank's branches is

[36] 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).

[37] 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).

[38] Erel and Liebersohn (2022) find that Fintech lenders expanded credit from the Paycheck Protection Program to small business owners in areas with more low-income and minority residents. But Fuster, Plosser, Schnabl and Vickery (2019) find that Fintech borrowers of purchase mortgages have higher incomes and are less likely to be minorities. In addition, Friedline, Naraharisetti and Weaver (2020) and Friedline and Chen (2021) find that low-income communities of color have the lowest Fintech adoption rates.

cited as the most important reason by far for choosing an institution for a main checking account. This persistent stated preference suggests that programs targeting the geographic distribution of bank branch locations will remain important for public policy.

9 Conclusion

We use mobile device geolocation data to develop a local bank branch access measure from a spatial gravity model. The model and consumer travel patterns enable a research design to quantify whether differences in access or in demand drive racial and income disparities in branch use. To overcome distortions in the geolocation data that protect user privacy, we estimate the gravity model with a new econometric method that adapts the Method of Simulated Moments to identify high-dimensional fixed effects. This approach may aid other applications involving privacy-protected big data.

Controlling for block-group racial and age shares, we find that access is better for residents of low-income communities, and lower demand for branch products or services explains their lower branch use. In contrast, Black communities have weaker access due to branches being located relatively far away from them. In big cities, the lack of access for Black communities is so significant that it eclipses their higher demand for branch products or services, leading to their overall lower branch use. These results inform where to target policies that increase access to banking across the country.

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. and R. Hauswald (2010) "Distance and private information in lending," *The Review of Financial Studies*, 23 (7), 2757–2788.

Agarwal, S., J. B. Jensen, and F. Monte (2018b) "The geography of consumption," May, Working paper. Georgetown University, Washington, D.C.

Agarwal, S., A. Mukherjee, and S. L. Naaraayanan (2023) "Roads and loans," *The Review of Financial Studies*, 36 (4), 1508–1547.

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 (4), 1793–1844.

Allen, J., C. Bavitz, M. Crosas et al. (2020) "The OpenDP White Paper," May, https://projects.iq.harvard.edu/files/opendifferentialprivacy/files/opendp_white_paper_11may2020.pdf, Working Paper.

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," February, 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*, https://www.federalreserve.gov/econres/notes/feds-notes/why-are-there-still-bank-branches-20180820.html.

Argyle, B., T. Nadauld, and C. Palmer (2023) "Real effects of search frictions in consumer credit markets," *Review of Financial Studies*, 36 (7), 2685–2720.

Arnaboldi, F. and P. Claeys (2008) "Internet banking in Europe: a comparative analysis," *Research Institute of Applied Economics*, 8 (11), 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*, 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 (46), 1–9.

Atkin, D., M. K. Chen, and A. Popov (2022) "The returns to face-to-face interactions: Knowledge spillovers in Silicon Valley," june, Working paper no. 30147. National Bureau of Economic Research, Cambridge, MA.

Augustine, K. (2018) "Small Business Banking: A Captive Audience," *Mercator Advisory Group Summary Report*, https://javelinstrategy.com/research/small-business-banking-captive-audience.

Bachas, P., P. Gertler, S. Higgins, and E. Seira (2018) "Digital financial services go a long way: Transaction costs and financial inclusion," in *AEA Papers and Proceedings*, 108, 444–448, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

——— (2021) "How debit cards enable the poor to save more," *The Journal of finance*, 76 (4), 1913–1957.

Balassa, B. (1965) "Trade liberalisation and "revealed" comparative advantage 1," *The manchester school*, 33 (2), 99–123.

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 (3), 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 (1), 234–266.

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

Boel, P. and P. Zimmerman (2022) "Unbanked in America: A Review of the Literature," *Economic Commentary* (2022-07).

Brown, J. R., J. A. Cookson, and R. Z. Heimer (2019) "Growing up without finance," *Journal of Financial Economics*, 134 (3), 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.

Burgess, R. and R. Pande (2005) "Do rural banks matter? Evidence from the Indian social banking experiment," *American economic review*, 95 (3), 780–795.

Cameron, A. C. and P. K. Trivedi (2013) *Regression analysis of count data*, 53: Cambridge University Press.

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

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

Célérier, C. and P. Tak (2023) "Finance, Advertising, and Fraud: The Rise and Fall of the Freedman's Savings Bank," February, Working Paper. University of Toronto, Toronto, CA.

Cespedes, J., E. X. Jiang, C. Parra, and J. Zhang (2024) "Branching Out Inequality: The Impact of Credit Equality Policies," January, Working paper.

Chavous, T. M., D. Rivas-Drake, C. Smalls, T. Griffin, and C. Cogburn (2008) "Gender matters, too: the influences of school racial discrimination and racial identity on academic engagement outcomes among African American adolescents.," *Developmental psychology*, 44 (3), 637.

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 (1).

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 D. G. Pope (2020) "Geographic mobility in America: Evidence from cell phone data," Working Paper.

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

Chetty, R., N. Hendren, M. R. Jones, and S. R. Porter (2020) "Race and Economic Opportunity in the United States: an Intergenerational Perspective*," *Quarterly Journal of Economics*, 135 (2), 711–783, 10.1093/qje/qjz042.

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

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

Cocheo, S. (2020) "Rethinking Branch Networks Without Killing Sales or Jeopardizing Growth," *The Financial Brand*, https://thefinancialbrand.com/news/banking-branch-transformation/covid-19-pandemic-coronavirus-branch-network-digital-online-channel-video-zoom-meeting-1035

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, https://doi.org/10.1016/j.jue.2021.103328.

Coven, J., A. Gupta, and I. Yao (2023) "JUE Insight: Urban flight seeded the COVID-19 pandemic across the United States," *Journal of Urban Economics*, 133, 103489, https://doi.org/10.1016/j.jue.2022.103489, Special Issue: JUE Insight Shorter Papers.

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

Dallerup, K., S. Jayantilal, G. Konov, A. Legradi, and H.-M. Stockmeier (2018) "A bank branch for the digital age," *McKinsey Company*, https://www.mckinsey.com/industries/financial-services/our-insights/a-bank-branch-for-the-digital-age.

Davidson, C. (2018) "Lack of access to financial services impedes economic mobility," *Economy Matters: Federal Reserve Bank of Atlanta*, https://www.atlantafed.org/economy-matters/community-and-economic-development/2018/10/16/lack-of-access-to-financial-services-impedes-economic-mobility#:~:text=Putsimply,lackofaccess,apaymentandcreditrecord.

Davidson, R. and J. G. MacKinnon (2004) *Econometric Theory and Methods*, 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 (4), 1684–1738.

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

DINGEL, J. I. AND F. TINTELNOT (2021) "Spatial economics for granular settings," January, 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 (2), 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," November, Working Paper. Bocconi University, Milan, Italy.

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 (4), 478–510.

EISFELDT, A. L. (2007) "Smoothing with liquid and illiquid assets," *Journal of Monetary Economics*, 54 (6), 1572–1586.

EIZENBERG, A., S. LACH, AND M. OREN-YIFTACH (2021) "Retail prices in a city," *American Economic Journal: Economic Policy*, 13 (2), 175–206.

ELLICKSON, P. B., P. L. GRIECO, AND O. KHVASTUNOV (2020) "Measuring competition in spatial retail," *The RAND Journal of Economics*, 51 (1), 189–232.

ELLWOOD, D. T. AND N. G. PATEL (2018) "Restoring the American Dream," *U.S. Partnership on Mobility from Poverty*, https://www.mobilitypartnership.org/restoring-american-dream.

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 (1), 90–118.

ERGUNGOR, O. E. AND S. MOULTON (2011) "Do bank branches matter anymore?" *Economic Commentary* (2011-13).

ERGUNGOR, O. E. (2010) "Bank branch presence and access to credit in low-to moderate-income neighborhoods," *Journal of Money, Credit and Banking*, 42 (7), 1321–1349.

EVANS, J. W. (2018) "Simulated Method of Moments (SMM) Estimation," *QuantEcon Notes*, https://notes.quantecon.org/submission/5b3db2ceb9eab00015b89f93.

FAJGELBAUM, P. D. AND C. GAUBERT (2020) "Optimal spatial policies, geography, and sorting," *The Quarterly Journal of Economics*, 135 (2), 959–1036.

FALLY, T. (2015) "Structural gravity and fixed effects," *Journal of International Economics*, 97 (1), 76–85.

FITZPATRICK, K. (2015) "The effect of bank account ownership on credit and consumption: Evidence from the UK," *Southern Economic Journal*, 82 (1), 55–80.

FONSECA, J. AND A. MATRAY (2024) "Financial Inclusion, Economic Development, and Inequality: Evidence from Brazil," January, Working paper.

FRIEDLINE, T. AND Z. CHEN (2021) "Digital redlining and the Fintech marketplace: evidence from US zip codes," *Journal of Consumer Affairs*, 55 (2), 366–388.

FRIEDLINE, T. AND M. DESPARD (2016) "Life in a banking desert," *The Atlantic*, https://www.theatlantic.com/business/archive/2016/03/banking-desert-ny-fed/473436/.

FRIEDLINE, T., S. NARAHARISETTI, AND A. WEAVER (2020) "Digital redlining: Poor rural communities' access to fintech and implications for financial inclusion," *Journal of Poverty*, 24 (5-6), 517–541.

FUSTER, A., M. PLOSSER, P. SCHNABL, AND J. VICKERY (2019) "The role of technology in mortgage lending," *Review of Financial Studies*, 32 (5), 1854–1899.

GAUGHAN, M. (2021) "Accessing a branch is 'highly important' to banking customers," *Insider Intelligence*, https://www.insiderintelligence.com/content/accessing-branch-highly-important-banking-customers.

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*, https://www.gillibrand.senate.gov/news/press/release/senators-gillibrand-and-sanders-representatives-ocasio-cortez-pascrell-and-kaptur-call-on-c

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.

Greenwald, A. G., D. E. McGhee, and J. L. Schwartz (1998) "Measuring individual differences in implicit cognition: the implicit association test.," *Journal of personality and social psychology*, 74 (6), 1464.

Guimaraes, P. and P. Portugal (2010) "A simple feasible procedure to fit models with high-dimensional fixed effects," *The Stata Journal*, 10 (4), 628–649.

Hacamo, I. (2023) "Racial Prejudice in the Workplace and Firm Boycotts," Working paper.

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

Harrigan, J. (1996) "Openness to trade in manufactures in the OECD," *Journal of International Economics*, 40 (1-2), 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 (4), 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 (4), 959–972.

——— (2014) "Gravity equations: Workhorse, toolkit, and cookbook," in *Handbook of International Economics*, 4, 131–195.

Heckman, J. (2022) "USPS continues postal banking pilot, despite House Republicans' objections," *Federal News Network*, https://federalnewsnetwork.com/agency-oversight/2022/04/usps-continues-postal-banking-pilot-despite-house-republicans-objections/.

Helpman, E., M. Melitz, and Y. Rubinstein (2008) "Estimating trade flows: Trading partners and trading volumes," *Quarterly Journal of Economics*, 123 (2), 441–487.

Hielscher, J. (2017) "Bank customers still want their branches," *Herald-Tribune*, https://www.heraldtribune.com/story/business/briefs/2017/05/15/bank-customers-still-want-their-branches/21043877007/.

Ho, K. and J. Ishii (2011) "Location and competition in retail banking," *International Journal of Industrial Organization*, 29 (5), 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 (1), 7–30.

Horton, D. (2019) "The Top 3 Reasons why Banks Branches are still important," *LinkedIn*, https://www.linkedin.com/pulse/top-3-reasons-why-banks-branches-still-important-david-horton/.

Hurtado, A. and J. Sakong (2022) "The effect of minority bank ownership on minority credit," Working paper.

Ji, Y., S. Teng, and R. M. Townsend (2023) "Dynamic Bank Expansion: Spatial Growth, Financial Access, and Inequality," *Journal of Political Economy, Forthcoming*.

Jiang, E. X., G. Y. Yu, and J. Zhang (2023) "Bank Competition amid Digital Disruption: Implications for Financial Inclusion," February, 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 (4), 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 (3), 487–510.

Marshall, G. and T. Pires (2018) "Measuring the impact of travel costs on grocery shopping," *The Economic Journal*, 128 (614), 2538–2557.

Martin, D. (2023) "The more things change, the more branches still matter," *American Banker*, https://www.americanbanker.com/opinion/the-more-things-change-the-more-branches-still-matter.

Mazet-Sonilhac, C. (2022) "Information Frictions in Credit Markets," September, 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*, 247–293: University of Chicago Press.

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 (9), 3568–3594.

Miyauchi, Y., K. Nakajima, and S. J. Redding (2021) "The economics of spatial mobility: Theory and evidence using smartphone data," February, 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.

Nau, C., M. Sidell, K. Clift, C. Koebnick, J. Desai, and D. Rohm-Young (2020) "A commercially available crime index may be a reliable alternative to actual census-tract crime in an urban area," *Preventive medicine reports*, 17, 100996.

Nguyen, H.-L. Q. (2019) "Are credit markets still local? Evidence from bank branch closings," *American Economic Journal: Applied Economics*, 11 (1), 1–32.

Nguyen, J. (2021) "Why is there still a need for physical bank branches?" *Marketplace*, https://www.marketplace.org/2021/11/18/why-is-there-still-a-need-for-physical-bank-branches/.

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*.

Pakes, A. and D. Pollard (1989) "Simulation and the asymptotics of optimization estimators," *Econometrica: Journal of the Econometric Society*, 1027–1057.

Paravisini, D., V. Rappoport, and P. Schnabl (2023) "Specialization in Bank Lending: Evidence from Exporting Firms," *The Journal of Finance*, 78 (4), 2049–2085.

Pew Research Center (2021) "Internet/Broadband Fact Sheet," Online, April, https://www.pewresearch.org/internet/fact-sheet/internet-broadband/?menuItem=480dace1-fd73-4f03-ad88-eae66e1f4217.

Prina, S. (2015) "Banking the poor via savings accounts: Evidence from a field experiment," *Journal of Development Economics*, 115, 16–31.

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.

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

Relihan, L. E. (2022) "Is online retail killing coffee shops? Estimating the Winners and Losers of Online Retail using Customer Transaction Microdata," Working Paper.

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

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*, 109, 403–08.

SafeGraph (2018-2019) "Places and Patterns Data," https://docs.safegraph.com/docs/welcome, accessed January 7, 2021.

Sakong, J. (2023) "Effect of Black Electoral Victories on Racial Bias and Economic Gaps," July, Working paper.

Sanders, B. (2021) "Fair banking for all," *Bernie Sanders Official Website*, https://berniesanders.com/issues/fair-banking-for-all/.

Silva, J. S. and S. Tenreyro (2006) "The log of gravity," *Review of Economics and Statistics*, 88 (4), 641–658.

Simpson, A. W. and M. T. Erickson (1983) "Teachers' verbal and nonverbal communication patterns as a function of teacher race, student gender, and student race," *American Educational Research Journal*, 20 (2), 183–198.

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 (12), 1622–1628.

Squire, R. F. (2019) "Quantifying sampling bias in SafeGraph Patterns," *SafeGraph Blog*, https://colab.research.google.com/drive/1u15afRytJMsizySFqA2EPlXSh3KTmNTQ#sandboxMode=true&scrollTo=xsNNli6GTN6s.

Srinivas, V. and R. Wadhwani (2019) "Recognizing the value of bank branches in a digital world: Findings from the global banking survey," *Deloitte Insights*, https://www2.deloitte.com/content/dam/insights/us/articles/4999_Global-banking-survey/DI_Bank-branches-digital-world.pdf.

Thaenraj, P. (2021) "Identifying CBG Sinks," *SafeGraph Blog*, https://colab.research.google.com/drive/17-cpOxXN7PFUjdEHf329fmbyirmeuza6?usp=sharing#scrollTo=RRqtUyPBg40H.

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

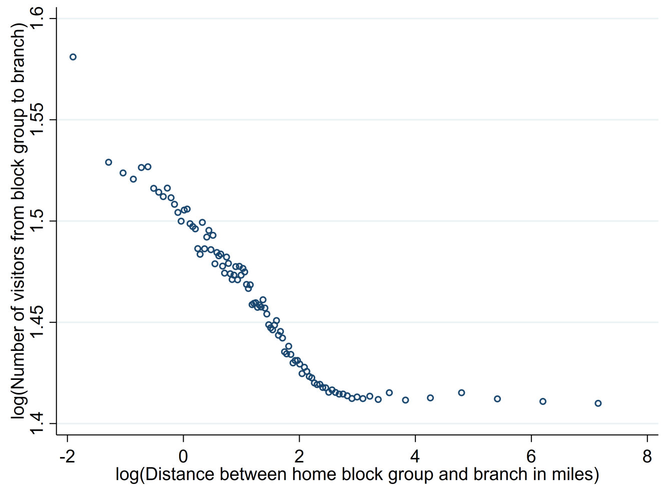
Wathen, J. (2018) "Believe It or Not, Bank Branches Still Matter," *The Motley Fool*, https://www.fool.com/investing/2018/03/04/believe-it-or-not-bank-branches-still-matter.aspx.

Westerlund, J. and F. Wilhelmsson (2011) "Estimating the gravity model without gravity using panel data," *Applied Economics*, 43 (6), 641–649.

Xu, K., B. Nosek, and A. Greenwald (2014) "Psychology data from the race implicit association test on the project implicit demo website," *Journal of Open Psychology Data*, 2 (1).

Yogo, M., A. Whitten, and N. Cox (2022) "Financial Inclusion Across the United States," April, Working paper. Princeton University, Princeton, NJ.

(A) All Block Group × Branch Pairs



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

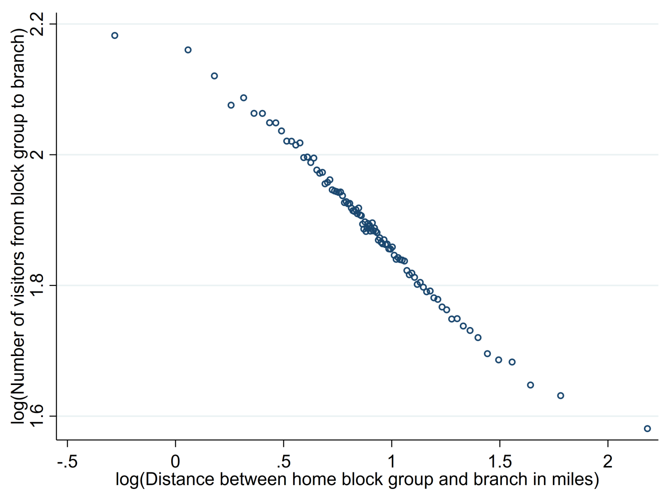


Figure 1

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. Panel A presents the observed (raw) geolocation data and includes all block group × branch pairs, including those with visitor counts of 2 or 3 that SafeGraph rounds up to 4. Panel B only includes block group × bank branch pairs with greater than 4 visitors. In that panel, the log numbers of visitors are residualized by block group × year-month fixed effects and branch × year-month fixed effects. The log distances are residualized by the same set of fixed effects. To construct the binned scatter plots, we divide the x-axis values into 100 equal-sized (percentile) bins. We then calculate the mean of the y-axis values and the mean of the x-axis values within each bin. In addition, for Panel B we add back the unconditional mean of the log numbers of visitors and the unconditional mean of the log distances to re-scale values.

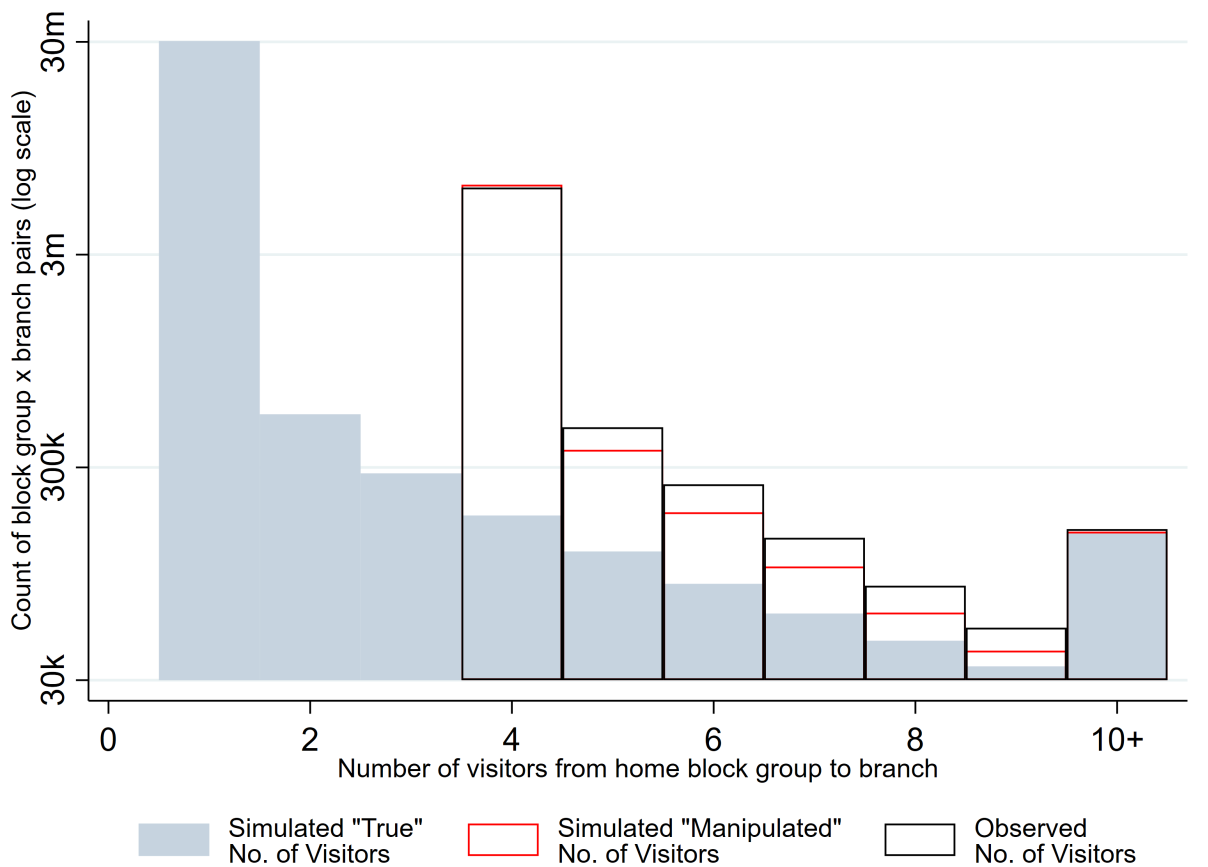


FIGURE 2
DISTRIBUTIONS OF VISITOR COUNTS

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 "maniputed" 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 Online 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 omit the large number of 0 visitor counts and censor the distributions at 10 visitors. That is, the number of block group × branch pairs with visitor counts exceeding 10 is assigned to 10+ visitors in the figure.

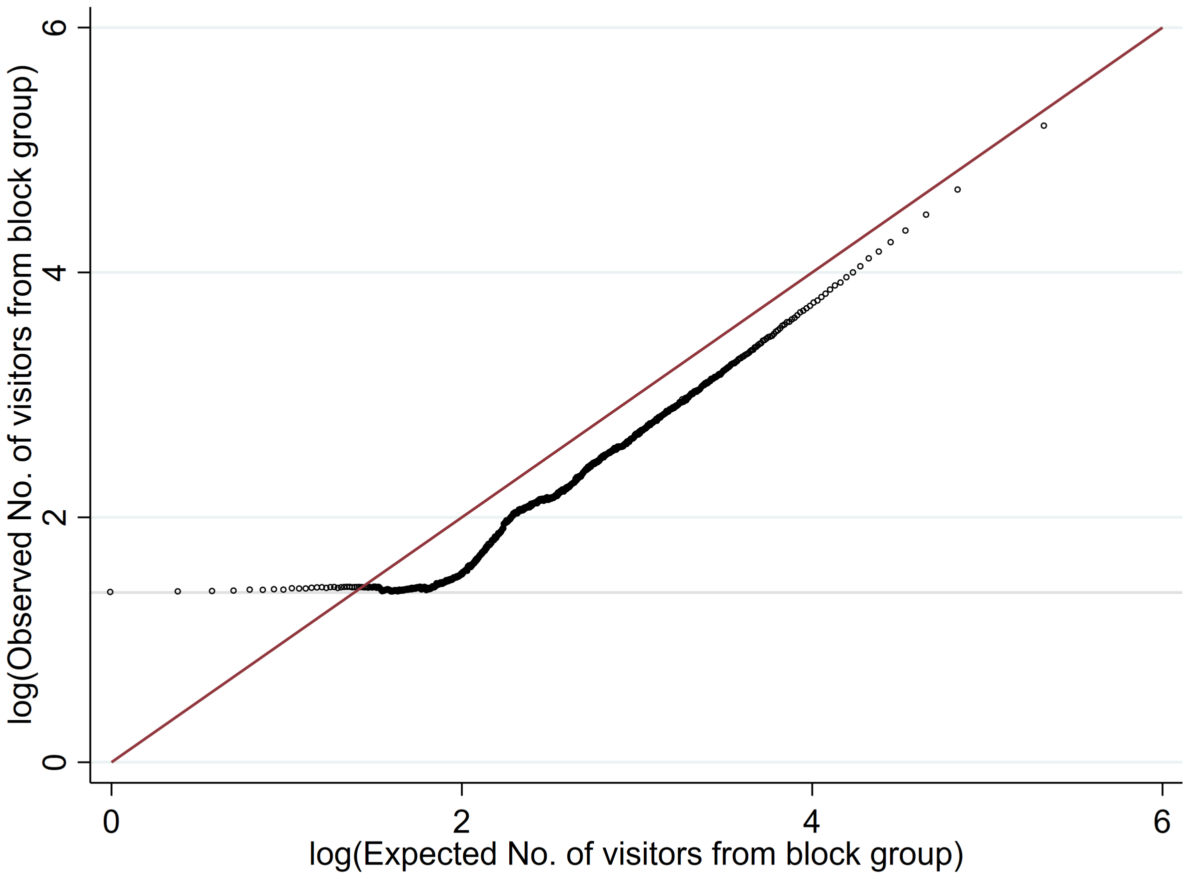
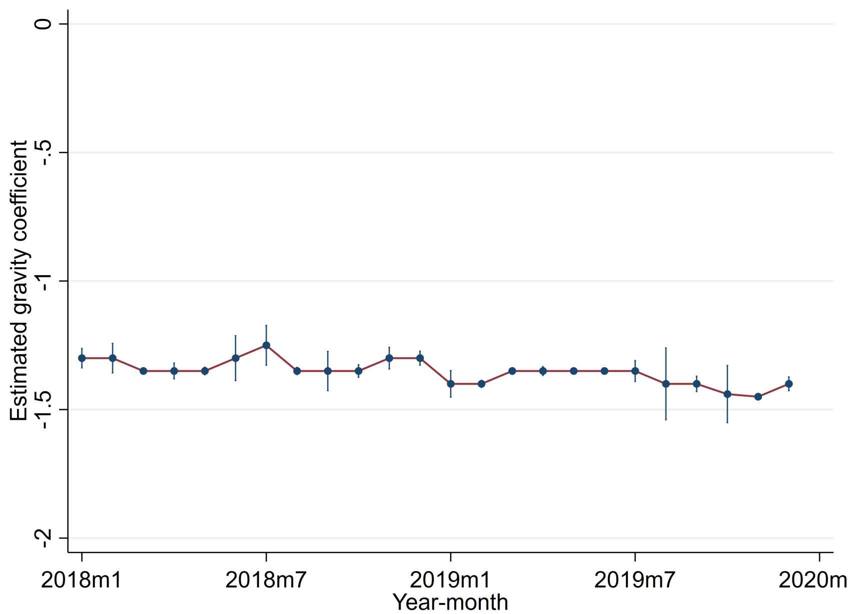


FIGURE 3

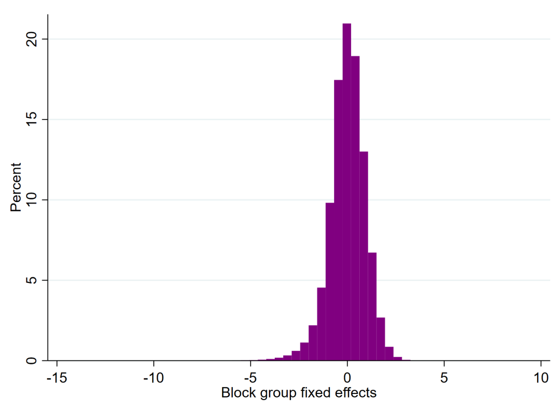
OBSERVED VS. EXPECTED BRANCH VISITORS PER CENSUS BLOCK GROUP

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., $\log \hat{V}^a_{it} \equiv \hat{\gamma}_{it} + \log \hat{\Phi}^a_{it}$, where the access measure $\hat{\Phi}^a_{it} \equiv \sum_{j \in b_{it}} \omega^i_t \exp\left(\hat{\lambda}_{jt}\right) d_{ij}^{-\hat{\beta}_t}$ reflects the branch probability weights used in the stratified sampling that defined in Eq. (A.30)). The observed and expected number of visitors range over the full sample period from January 2018 to December 2019. 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 Online 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 $-\hat{\beta}_{t,\text{MSM}}$



(B) Census Block Group Fixed Effects



(C) Bank Branch Fixed Effects

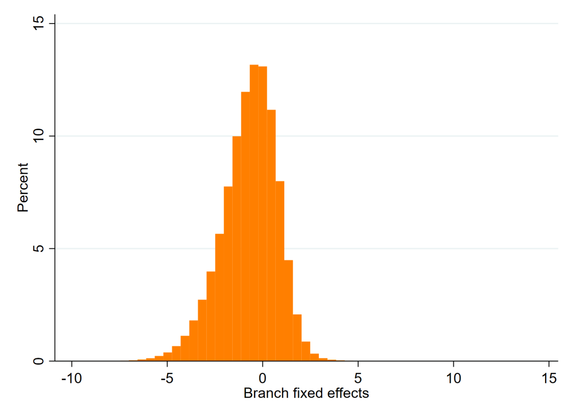


Figure 4
Method of Simulated Moments Parameter Estimates

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 $-\hat{\beta}_{t,\text{MSM}}$ gravity coefficient estimates, along with 95% confidence intervals, which are computed using the MSM standard errors in Eq. (A.65). Panel B presents a histogram of the estimated Census block group fixed effects, $\left\{\hat{\gamma}_{it}^{\infty}\right\}$, and Panel C presents a histogram of the estimated bank branch fixed effects, $\left\{\hat{\lambda}_{jt}^{\infty}\right\}$. 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 Online Appendix A.

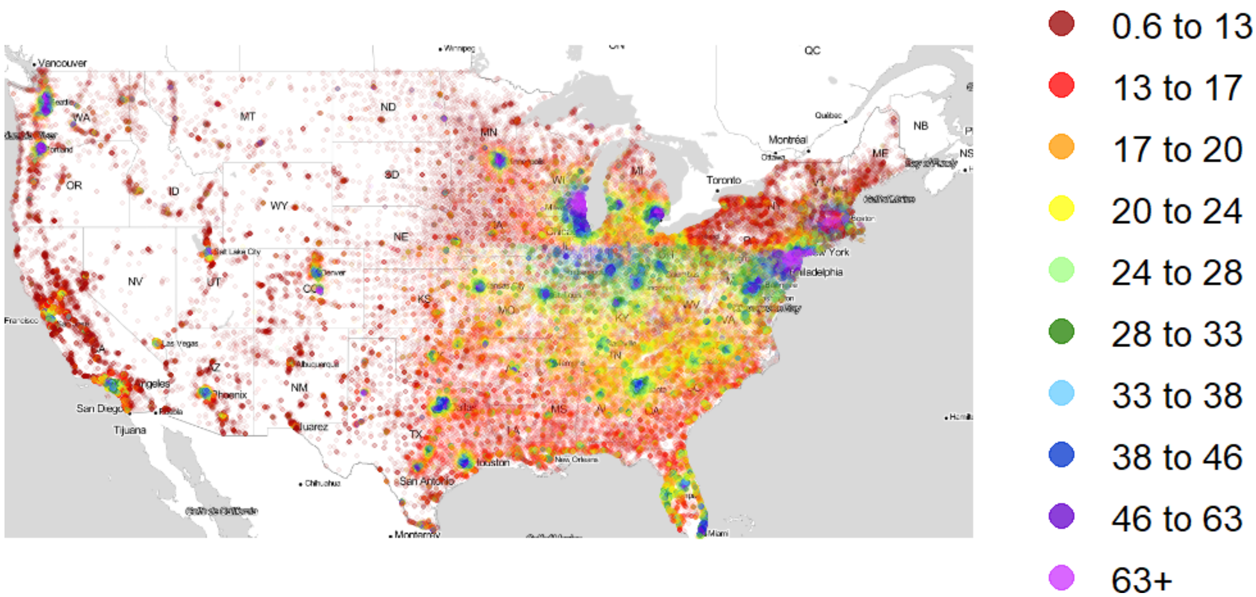


Figure 5
Bank Branch Access Nationwide

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 Online 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.

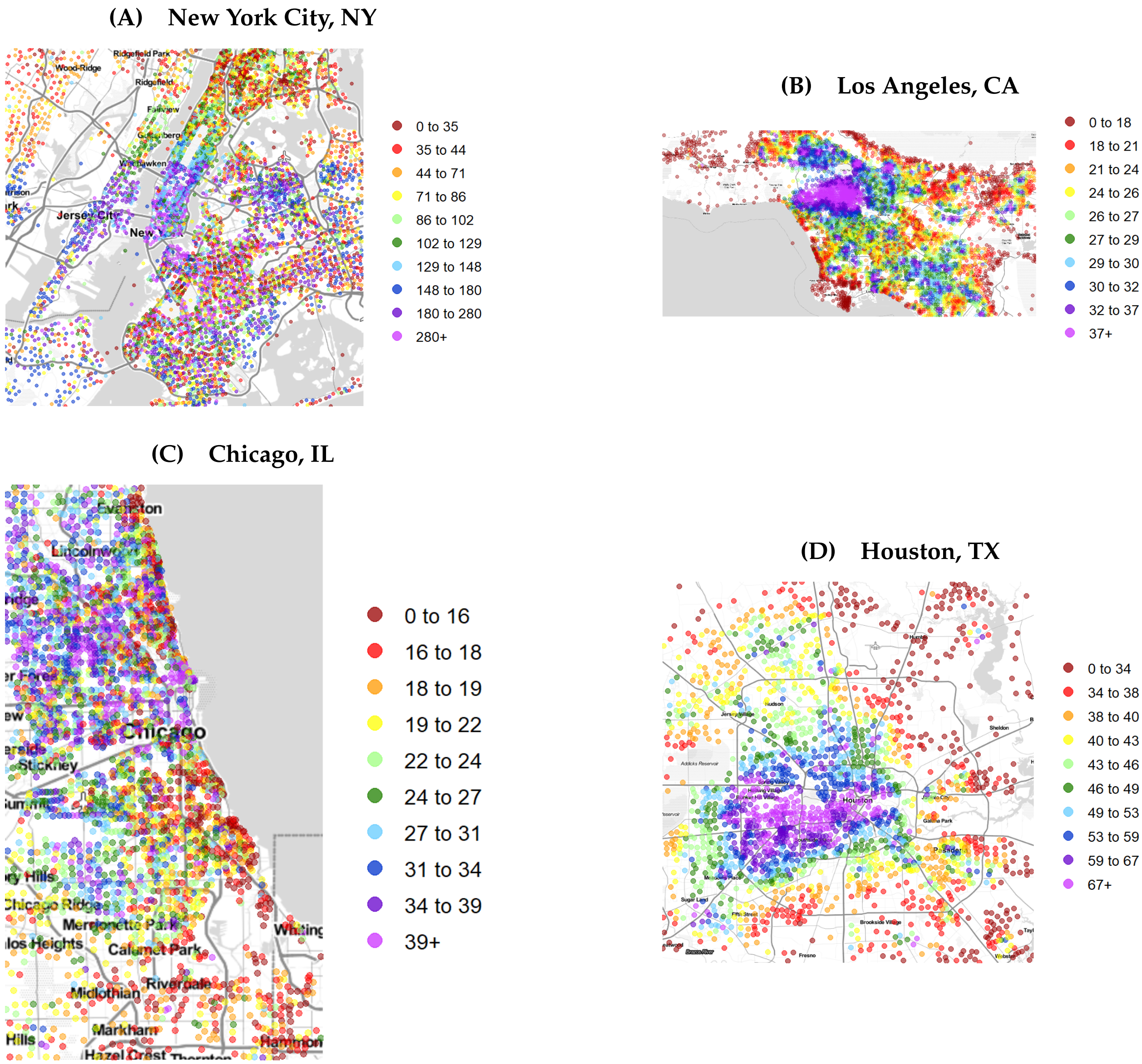


Figure 6
Bank Branch Access in the Four Largest U.S. Cities

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 Online 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
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 2
Bank Branch Access by Demographic Attributes

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

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 "distance-adjusted average 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 Online 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 3
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	0.155	0.190	0.155	0.296	0.230	0.315	0.242
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Black	-0.041	-0.056	-0.017	-0.039	0.041	-0.003	0.089	0.025
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
Asian	0.238	0.241	0.227	0.224	-0.232	-0.197	-0.202	-0.174
	(0.020)	(0.020)	(0.020)	(0.019)	(0.023)	(0.022)	(0.023)	(0.022)
Other	0.034	0.041	0.009	0.016	0.010	0.020	-0.068	-0.058
	(0.028)	(0.029)	(0.038)	(0.039)	(0.033)	(0.033)	(0.046)	(0.046)
Hispanic	0.008	-0.016	0.021	-0.012	-0.038	-0.097	-0.000	-0.083
	(0.009)	(0.010)	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)	(0.012)
Age <15		0.341		0.343		1.061		1.157
		(0.022)		(0.025)		(0.026)		(0.030)
Age 35-54		0.486		0.529		0.724		0.732
		(0.023)		(0.026)		(0.025)		(0.029)
Age 55-64		0.101		0.053		0.651		0.626
		(0.025)		(0.030)		(0.029)		(0.034)
Age 65+		0.248		0.253		0.494		0.515
		(0.018)		(0.020)		(0.021)		(0.023)
log(No. of devices)	0.606	0.606	0.601	0.601	0.656	0.659	0.658	0.660
	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(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^2	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	O	O	O	O	O	O	O	O
County FE	O	O	O	O	O	O	O	O
RUCA FE	O	O			O	O		

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 goers from each block group based on the month-by-month Method of Simulated Moments estimates; i.e., $\log \hat{V}^*_{it} \equiv \log \sum_j \hat{V}^*_{ijt}$, where $\hat{V}^*_{ijt}$ is the predicted mean of V^*_{ijt} in Eq. (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 Online 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.

For Online Publication: 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. The approach adapts the Method of Simulated Moments (MSM) to identify high-dimensional 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

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

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.[39]

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:

$$V^+_{ijt} = V^*_{ijt} + L_{ijt}. \tag{A.16}$$

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

$$z_{ijt} = \begin{cases} 1 & \text{if } \lfloor V^+_{ijt} \rfloor \geq 2, \\ 0 & \text{otherwise.} \end{cases} \tag{A.17}$$

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

$$V_{ijt} = \max\left\{4, \lfloor V^+_{ijt} \rfloor\right\}, \tag{A.18}$$

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

A.2 Sample block group × 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 and over two-hundred-thousand block groups, altogether spanning twenty-four months, makes it computationally impractical to have the billions of possible block group × branch pairs enter the MSM estimation. Instead, we sample pairs using stratified sampling.

[39] The haversine distance between two latitude-longitude coordinates $(lat_1, long_1)$ and $(lat_2, long_2)$ is $2r \arcsin\left(\sqrt{h}\right)$, where r is the Earth's radius and $h = \text{hav}(lat_1 - lat_2) + \cos(lat_1)\cos(lat_2)\,\text{hav}(long_2 - long_2)$. The haversine function $\text{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right)$. The centers of population are computed using population counts from the 2010 Census and are found here: 2010 Census Centers of Population. 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. Even so, the haversine formula is simple, fairly accurate, and convenient to compute.

In each year-month, block group × branch pairs in the SafeGraph data register either positive (and ≥ 4) or missing observed visitor counts. If a block group × 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 × branch pair in our simulation with probability 1. If a block group × 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 × 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 × 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 × branch pairs represents slightly higher than a 0.05% sample size of all possible block group × branch pairs with missing visitor counts.

To establish notation for the stratified sampling of block group × branch pairs, we let n_t denote the stratified sample of block group × 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 × 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 × 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 × 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:

$$\frac{1}{|N_{it}^0|} + \frac{p - \frac{1}{|N_{it}^0|}}{1 - \frac{1}{|N_{it}^0|}} - \frac{1}{|N_{it}^0|} \frac{p - \frac{1}{|N_{it}^0|}}{1 - \frac{1}{|N_{it}^0|}}, \tag{A.19}$$

which is simply the probability of the union of the two independent events, where we have used the relation $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$ 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 × 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 × 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 ω_t to the sampled pairs in the set n_t^0. These probability weights satisfy:

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

Rearranging Eq. (A.20) shows that the probability weight ω_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 $\psi \equiv \left\{\beta_t, \gamma_{it}, \lambda_{jt}\right\}$ 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 $\left\{\gamma_{it}, \lambda_{jt}\right\}$ and let the minimization problem identify β_t. Holding fixed an estimate of β_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 β_t in the year-month, the MSM minimization problem then chooses the optimal β_t estimate that satisfies the moment conditions in the year-month. We initialize the fixed effects routine with guessed estimates $\hat{\gamma}^0_{it} = \hat{\lambda}^0_{jt} = 1$ 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^0_t and n^1_t, because of their different probability weights.

Consider first the set n^1_t of pairs with positive observed visitor counts that were sampled with probability 1. Per year-month, we begin the simulation by drawing $|n^1_t| \times S$ Laplace random variables having mean zero and scale 10/9, and we draw $|n^1_t| \times S$ independent Uniform random variables over the unit interval. We draw these random variables only once at the beginning of each year-month's run so that the MSM does not have the underlying sample change for every guess of the model parameters. Given an estimate $\hat{\beta}_t$ of the gravity coefficient and the initial guessed estimates $\left\{\hat{\gamma}^0_{it}, \hat{\lambda}^0_{jt}\right\}$ of the fixed effects, we then apply the inverse Poisson CDF to transform the Uniform random variables into Poisson random variables with distinct means given in Eq. (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. Simulated visitor counts are 0 if either the true visitor count (from the Poisson draw) is 0 or the noisy visitor count (from the Poisson draw plus the Laplace draw) falls below 2. This way, simulated visitor counts that equal 0 arise in the same two ways as would 0 visitor counts in the observed SafeGraph data. Let $\tilde{v} = \{\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_S\}$ be the S simulated visitor counts in year-month t, where we have excluded a t subscript to simplify notation.

Consider next the set n^0_t of block group $\times$ branch pairs with missing visitor counts that were sampled with probability 1/2000. If an extra $|n^0_t| \times S$ 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^0_t, we construct their implied empirical probability distribution according to the parameter estimate of ψ in each iteration. If an infinite number of visitor counts from the pairs in n^0_t were in fact simulated, their distribution would coincide with this constructed empirical distribution. Notice that we cannot apply this approach to the set n^1_t 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^0_t 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.

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 Poission 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 distribution of visitor counts in Eq. (A.15). Namely,

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

Finally, let the probability that the Poisson distribution draws a visitor count of k, given its estimated mean $\hat{\mu}_{ijt}$ be denoted $p\left(k, \hat{\mu}_{ijt}\right)$. Notice that the parameters of the empirical distribution update with every iteration of the estimated fixed effects and guess of β_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 $K \to \infty$ and $L \to \infty$, they would hold exactly. The 7 components of the empirical distribution we compute are:

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

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

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:**

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

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:**

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

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 Lapalace 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 $4.\overrightarrow{9}$ would round down to 4.

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

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

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:**

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

The formula for the mean visitor count is broken up into two parts. Both parts are multiplied by the probability, $p\left(k, \hat{\mu}_{ijt}\right)$, 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:**

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

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 $\Pr\left(\tilde{V}_{ijt} > 0 | \hat{\mu}_{ijt}\right)$, 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:**

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

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 × 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 × 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 × 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 ω_t^j, and the branches in b_{it}^0 have probability weights denoted ω_t^i. These probability weights are defined as:

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

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

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 $\left\{\hat{\lambda}_{jt}\right\}$, while holding constant the estimated block group fixed effects $\{\hat{\gamma}_{it}\}$ at $\hat{\gamma}_{it}^0 = 1$, $\forall i$ 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

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

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

$$\hat{V}_{jt}^{k} = \exp\left(\hat{\lambda}_{jt}^{k}\right)\left(\sum_{i\in h_{jt}^{1}}\exp\left(\hat{\gamma}_{it}^{k}\right)\text{Distance}_{ij}^{-\hat{\beta}_t} + \sum_{i\in h_{jt}^{0}}\omega_t^{j}\exp\left(\hat{\gamma}_{it}^{k}\right)\text{Distance}_{ij}^{-\hat{\beta}_t}\right) \tag{A.32}$$

Given $\hat{\beta}_t$ and the k-th iteration of the estimated block group fixed effects, $\left\{\hat{\gamma}_{it}^{k}\right\}$, 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:

$$\hat{\lambda}_{jt}^{k} = \log V_{jt}^{T} - \log\left(\sum_{i\in h_{jt}^{1}}\exp\left(\hat{\gamma}_{it}^{k}\right)\text{Distance}_{ij}^{-\hat{\beta}_t} + \sum_{i\in h_{jt}^{0}}\omega_t^{j}\exp\left(\hat{\gamma}_{it}^{k}\right)\text{Distance}_{ij}^{-\hat{\beta}}\right). \tag{A.33}$$

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 β_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 β_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

$$\overline{V}_{it} = \frac{1}{|b_{it}|}\sum_{j\in b_{it}} V_{ijt}. \tag{A.34}$$

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

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

where $\mathbb{E}\left(\tilde{V}_{ijt}^{k}|\hat{\mu}_{ijt}\right)$ 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

$$\overline{\tilde{V}}_{it}^{k} = \frac{1}{S}\sum_{s}\overline{\tilde{V}}_{it}^{k}(s). \tag{A.36}$$

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

$$\chi_{it}^k = \frac{\overline{V}_{it}}{\overline{\tilde{V}}_{it}^k} \tag{A.37}$$

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:

$$\hat{\gamma}_{it}^{k+1} = \hat{\gamma}_{it}^k \times \left(\chi_{it}^k\right)^g , \tag{A.38}$$

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 $\chi_{it}^k < 1$, 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, $\left\{\hat{\gamma}_{it}^{k+1}\right\}$; and (iii) the updated branch fixed effect estimates, $\left\{\hat{\lambda}_{jt}^{k+1}\right\}$, 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.[40]

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 goers to more branches (higher $|b_{it}|$). Mathematically, the convergence condition is

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

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

After the condition in Eq. (A.39) is met, we have converged fixed effects estimates, denoted $\left\{\hat{\gamma}_{it}^{\infty}\right\}$ and $\left\{\hat{\lambda}_{jt}^{\infty}\right\}$, 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 β_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 × 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 ω_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 × branch pairs from the stratified sample. The 6 data and simulated moments are:

[40] 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).

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

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

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

where $\mathbb{I}(\cdot)$ stands for the indicator function and $\Pr\left(\tilde{V}_{ijt}=0|\hat{\mu}_{ijt}\right)$ 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 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:**

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

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

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

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

$$m_3(v) \equiv \frac{\sum_{(i,j)\in n_t^1} \mathbb{I}\left(V_{ijt}=0\right)\log \text{Distance}_{ij} + \sum_{(i,j)\in n_t^0} \mathbb{I}\left(V_{ijt}=0\right)\omega_t \log \text{Distance}_{ij}}{\sum_{(i,j)\in n_t^1} 1 + \sum_{(i,j)\in n_t^0} \omega_t}, \tag{A.44}$$

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

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

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

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

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

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

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

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

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

Notice that because the data moment reflects only positive observed visitor counts from the set n_t^1 of sampled block group × 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

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

where $\Pr\left(\tilde{V}_{ijt} > 0|\hat{\mu}_{ijt}\right)$ is from Eq. (A.23). The observation weights consist of (1) a vector of ones with length equaling the number of block group × 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

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

where $\odot$ is the element-wise product and $\mathbb{E}\left(\log \tilde{V}_{ijt}|\tilde{V}_{ijt} > 0, \hat{\mu}_{ijt}\right)$ 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 × 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

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

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

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

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

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

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

The data moment is then

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

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

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

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

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

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

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

With these terms established, we set the simulated moment as

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

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

$$\overline{m}\left(\tilde{v}|\psi\right) = \frac{1}{S}\sum_{S} m\left(\tilde{v}_s|\psi\right). \tag{A.62}$$

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 compute the error $e_r\left(\tilde{v}, v|\psi\right)$ per moment r, which is the percent difference between a data moment and its corresponding model moment:

$$e_r\left(\tilde{v}, v|\psi\right) \equiv \frac{\overline{m}_r\left(\tilde{v}|\psi\right) - m_r\left(v\right)}{m_r\left(v\right)}, \quad \forall r \in \{1, \ldots, 6\}. \tag{A.63}$$

Let $e\left(\tilde{v}, v|\psi\right)$ denote the vector of moment errors. The MSM estimator is then

$$\hat{\beta}_{t,\text{MSM}} = \underset{\beta_t}{\text{argmin}}\; e\left(\tilde{v}, v|\psi\right)' W e\left(\tilde{v}, v|\psi\right), \tag{A.64}$$

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

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 of the MSM estimator $\hat{\beta}_{t,\text{MSM}}$ as

$$\widehat{\text{Var}}\left(\hat{\beta}_{t,\text{MSM}}\right) = \left(1 + \frac{1}{S}\right)\left(\hat{J}'\hat{J}\right)^{-1}\hat{J}'\hat{\Sigma}\hat{J}\left(\hat{J}'\hat{J}\right)^{-1}, \tag{A.65}$$

where $\hat{J} \equiv \frac{\partial e(\tilde{v},v|\psi)}{\partial \beta_t}$ is the estimated derivative of the vector of moment errors, evaluated at $\hat{\beta}_{t,\text{MSM}}$, and $\hat{\Sigma} \equiv e(\tilde{v},v|\psi)\, e(\tilde{v},v|\psi)'$ is a consistent estimate of the covariance matrix of moment errors. We calculate the derivatives numerically by taking a central difference around $\hat{\beta}_{t,\text{MSM}}$.

B Robustness

In this section, we conduct robustness checks on the paper's main empirical conclusions. In Appendix B.1, we measure branch access when all SOD branches are included instead of only those in SafeGraph, and in Appendix B.2, we evaluate the robustness to omitted variable bias due to the endogenous location choices of residents and bank branches.

B.1 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 $\left\{\text{Distance}_{ij}\right\}_{\forall j}$, (iii) the estimated gravity coefficient $\hat{\beta}_t$, and (iv) the estimated fixed effects of the branches $\left\{\hat{\lambda}_{jt}\right\}$. 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 $\left\{\hat{\lambda}_{jt}\right\}$ 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 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.[41]

Online Table A.12 repeats the branch access OLS regressions of Table 2, 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

[41] 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.

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 2 was -7.6%. In column (2) of Online Table A.12, 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 2, the Black-White gap in access was 5.3%. In column (2) of Online Table A.12, the Black-White gap is 15.6%. In Metro cores, the Black-White gap in access widens from 6.4% to 17.8%.

B.2 Omitted Variable Bias

The primary concern over whether β_t in the gravity model of Eq. (1) represents the causal effect of distance on branch visitation is potential omitted variable bias. Namely, Distance_{ijt} is endogenous to the location choices of households and banks. People might choose to live in neighborhoods that have bank branches they prefer to visit, and banks might build branches in neighborhoods whose residents would find those branches attractive. Unobservable characteristics of residents and branches may correlate both with the distances between them and visits, which would render Eq. (1) subject to an omitted variables problem.

To make this potential omitted variables problem more concrete, consider the following thought experiment. Suppose that all branches in high-income areas catered to high-income customers, and all branches in low-income areas catered to low-income customers. Now randomly swap residents of low-income neighborhoods with residents of high-income neighborhoods. There is an omitted variables problem if both types of the randomly reallocated residents then visited branches *less*, holding constant all block group and branch fixed effects. Similarly from the branch perspective, randomly swap branches between low-income and high-income neighborhoods. There is an omitted variables problem if visitor flows to those randomly replanted branches were then lower, all else equal. The potential for households' *bank brand-specific loyalty* and/or branches' *customer-specific specialization* (as part of households' and branches' location decisions) confounds the causal interpretation of β_t.[42]

To evaluate the robustness to this possible omitted variable bias, we follow the common approach of assessing the sensitivity of the estimated parameter of interest (the gravity coefficient) to the inclusion of observed controls. In our context, such controls should proxy for the true omitted variables that relate to the endogenous location choices of residents and branches. We assess the sensitivity by running OLS and PPML on Eq. (1) using the raw geolocation data.

What controls do we consider? If people live near a branch that caters to them and if that branch is located near the customers it wishes to cater to, then we should expect to see disproportionate shares of certain types of visitors over others at that branch. This logic motivates adding controls that measure the relative shares of types of visitors over others at a branch, reminiscent of Balassa (1965)'s "revealed comparative advantage." Because visitors differ by several characteristics, we add relative shares along three observable categories: two from the branch perspective and one from the block group perspective. From the branch perspective, we consider block group median household income and the five racial categories of visitors. From the block group perspective, we consider visitors' bank brand choice. The last characteristic is meant to capture the notion that some residents may visit certain branches over others irrespective of distance because of brand loyalty.[43,44]

We define the relative shares as follows, first from the branch perspective. The number of visitors from block group i with characteristic c who visit branch j in year-month t is V^c_{ijt}. The branch's total number of visitors with characteristic c over the sample period is $V^c_{jt} = \sum_t \sum_i V^c_{ijt}$. The branch's total number of visitors over the sample period

[42]Notice that if residents visited a branch because that branch was nearby attractive amenities like cafes or parks, that choice would be encoded in the branch fixed effect. Likewise, if a bank chose to build a branch in a high foot-traffic area, that high foot traffic is likely a consequence of the area's features, such as being near public transportation or a large retail shopping center. That choice too would be encoded in the branch fixed effect.

[43]Recently, Paravisini et al. (2023) adopt a relative shares measure similar to Balassa (1965) in their study of bank specialization in lending. In their microfoundation for the measure, banks with unobservable advantages in lending toward certain activities have a disproportionate share of their loans funding those activities. Likewise, firms engaged in certain activities disproportionately fund those activities with credit from specialized banks.

[44]If a branch located in an area because that area has amenities that are attractive to particular subpopulations, then we should also expect to see a disproportionate share of these types of customers visiting that branch.

is $V_j = \sum_t \sum_c V^c_{jt}$. We start by computing the share of branch j's visitors who have characteristic c:

$$\tilde{S}^c_j = \frac{V^c_j}{V_j} = \frac{\sum_t \sum_i V^c_{ijt}}{\sum_t \sum_c \sum_i V^c_{ijt}}. \tag{A.66}$$

We then compare branch j's share to the visitor-c share of all branches in the country. Using the terminology of Balassa (1965), we obtain branch j's "revealed comparative advantage" (RCA) in serving visitors with characteristic c:

$$RCA^c_j = \frac{\tilde{S}^c_j}{\tilde{S}^c}, \tag{A.67}$$

where $\tilde{S}^c$ is the share of all branch visitors across the country over the sample period who have characteristic c:

$$\tilde{S}^c = \frac{\sum_t \sum_j V^c_{jt}}{V} = \frac{\sum_t \sum_j \sum_i V^c_{ijt}}{\sum_t \sum_j \sum_c \sum_i V^c_{ijt}}. \tag{A.68}$$

Unfortunately, we cannot construct the ideal measure RCA^c_j because the anonymous geolocation data makes V^c_{ijt} unobservable. Instead, we proxy for RCA^c_j with an observable counterpart that uses block group visitor counts in the period, V_{ijt}, and block group population information about characteristic c from the 2019 5-year ACS. When c refers to racial categories, we define the proxy:

$$S^{\text{race } c}_j \equiv \frac{\sum_t \sum_i V_{ijt} X^{\text{race } c}_i}{\sum_t \sum_i V_{ijt}}, \tag{A.69}$$

where $X^{\text{race } c}_i$ is block group i's population share of residents with race c. The racial category proxy measures branch j's number of visitors, weighted by their block group racial shares, as a fraction of all of branch j's visitors.

Because block-group level income is measured as median household income, we have no categories to work with. Instead, we define the income-related shares as:

$$S^{\text{income}}_j \equiv \frac{\sum_t \sum_i V_{ijt} X^{\text{income}}_i}{\sum_t \sum_i V_{ijt}}, \tag{A.70}$$

where X^{income}_i is block group i's log median household income. The income proxy measures branch j's number of visitors, weighted by their block group median household incomes, as a fraction of all of branch j's visitors.

From the block group perspective, we compute proxies for residents' bank brand choice. Specifically, we define

$$S^{\text{brand } c}_i \equiv \frac{\sum_t \sum_j V_{ijt} X^{\text{brand } c}_j}{\sum_t \sum_j V_{ijt}}, \tag{A.71}$$

where $X^{\text{brand } c}_j$ is a dummy equaling 1 if branch j belongs to brand c, and 0 otherwise. The brand proxy measures block group i's number of branch goers, weighted by whether they visit brand c, as a fraction of all branch goers in block group i. We compute $S^{\text{brand } c}_i$ for the top 5 bank brands by number of branches (JPMorgan Chase, Bank of America, Wells Fargo, PNC, and U.S. Bank).

Notice that the share proxies in Eqs. (A.69) to (A.71) do not adjust for the shares of all branches (or block groups). Thus, to gauge a branch's or block group's relative comparative advantage, we also rely on the distribution of shares across branches (or block groups). To include as controls parts of the distribution of shares, we use dummies for the quartile of the characteristic-c distribution that a branch (or block group) belongs to.

Online Table A.13 presents the means and percentiles of the distributions of the various shares. The mean visitor shares by race are consistent with the Census population shares, but there is meaningful dispersion in the distributions across branches. For example, a typical (i.e., mean) branch has about 10% of its visitors from predominately Black areas, and the U.S. Black population share is 13.6%. But there are some branches (the 95$^{\text{th}}$ percentile) with roughly 35% of their visitors from Black communities. In fact, if we isolate branches belonging to minority depository institutions (MDIs), we see an even larger share of visitors coming from Black communities.[45] Consistent with intuition and the

[45] According to the FDIC Statement of Policy Regarding Minority Depository Institutions, an MDI "may be a federal insured depository institution for which (1) 51 percent or more of the voting stock is owned by minority individuals; or (2) a majority of the

motivation for these Balassa-style controls, we observe a disproportionate share of visitors to MDI branches from predominately minority communities. The typical Black MDI branch has 52% of its visitors from Black communities, and Black MDI branches at the 95$^{\text{th}}$ percentile have 77%. Similarly, a typical Asian MDI branch has disproportionate shares of visitors from Asian communities (26%), and a typical Hispanic MDI has disproportionate shares of visitors from Hispanic communities (70%). Regarding the brand loyalty shares in the table, the distributions for Wells Fargo, Bank of America, and JPMorgan Chase compare similarly, whereas the distributions for U.S. Bank and PNC are roughly the same. A typical block group has about 11% of its branch goers visiting a Chase branch, for instance, but block groups at the 95$^{\text{th}}$ percentile have 43% of its branch goers visiting a Chase.

In our empirical strategy, we add the share measures to the gravity model of Eq. (1). To avoid potential spurious correlation between visitors from block group i to branch j (i.e., V_{ijt}) and the share measure of branch j, we leave out i's visitor count when constructing a branch's share:

$$S_{(-i)j}^{\text{race } c} \equiv \frac{\sum_t \sum_{k \neq i}^{I} V_{kjt} X_k^{\text{race } c}}{\sum_t \sum_{k \neq i}^{I} V_{kjt}}, \tag{A.72}$$

$$S_{(-i)j}^{\text{income}} \equiv \frac{\sum_t \sum_{k \neq i}^{I} V_{kjt} X_k^{\text{income}}}{\sum_t \sum_{k \neq i}^{I} V_{kjt}}. \tag{A.73}$$

Similarly, we leave out branch j's visitor counts when constructing a block group's brand share:

$$S_{i(-j)}^{\text{brand } c} \equiv \frac{\sum_t \sum_{k \neq j}^{J} V_{ikt} X_k^{\text{brand } c}}{\sum_t \sum_{k \neq j}^{J} V_{ikt}}. \tag{A.74}$$

With these share measures, we employ two specifications. The first uses the level of shares:

$$\begin{aligned}
\log\left(\text{No. of Visitors}_{ijt}\right) = \gamma_{it} + \lambda_{jt} - \beta \log\left(\text{Distance}_{ij}\right) \\
+ \sum_c \alpha_1^c X_i^{\text{race } c} \times S_{(-i)j}^{\text{race } c} \\
+ \alpha_2 X_i^{\text{income}} \times S_{(-i)j}^{\text{income}} \\
+ \sum_c \alpha_3^c X_j^{\text{brand } c} \times S_{i(-j)}^{\text{brand } c} + \varepsilon_{ijt}.
\end{aligned} \tag{A.75}$$

Notice that the branch share variables, $S_{(-i)j}^{\text{race } c}$ and $S_{(-i)j}^{\text{income}}$, interact with block group i characteristics because the concern over omitted variables is a block group × branch-specific component. Likewise, the block group share variable $S_{i(-j)}^{\text{brand } c}$ interacts with the branch j characteristic.

Our second specification utilizes dummies for a branch's (and block group's) quartile in the distribution of shares, which also allows for a nonlinear relation between visitation and shares:

$$\begin{aligned}
\log\left(\text{No. of Visitors}_{ijt}\right) = \gamma_{it} + \lambda_{jt} - \beta \log\left(\text{Distance}_{ij}\right) \\
+ \sum_c \sum_{q=2}^{4} \alpha_{1q}^c X_i^{\text{race } c} \times D\left(S_{(-i)j}^{\text{race } c} \in Q_q^{\text{race } c}\right) \\
+ \sum_{q=2}^{4} \alpha_{2q} X_i^{\text{income}} \times D\left(S_{(-i)j}^{\text{income}} \in Q_q^{\text{income}}\right) \\
+ \sum_c \sum_{q=2}^{4} \alpha_{3q}^c X_j^{\text{brand } c} \times D\left(S_{i(-j)}^{\text{brand } c} \in Q_q^{\text{brand } c}\right) + \varepsilon_{ijt},
\end{aligned} \tag{A.76}$$

where, for example, $D\left(S_{(-i)j}^{\text{race } c} \in Q_q^{\text{race } c}\right)$ is a dummy that equals 1 if the branch's share is in quartile $q = 1, \ldots, 4$ of the

board of directors is minority and the community that the institution serves is predominantly minority."

distribution of $S^{\text{race } c}_{(-i)j}$ for each racial category c over the full sample. The bottom quartiles of the distributions are omitted in Eq. (A.76), which implies that the coefficients on the interaction terms capture the elasticity for branches (or block groups) with q-level shares relative to those with bottom quartile shares. We try versions of Eqs. (A.75) to (A.76) using OLS and PPML over the full panel that includes each share variable separately and together with standard errors two-way clustered by Census block groups and bank branches.

Online Table A.14 and Online Table A.15 present the results. Throughout all specifications, the gravity coefficient estimate stays virtually unchanged despite meaningful increases in the R^2s. Across all PPML estimations, the coefficient remains around -0.09, and across all OLS estimations, it remains about -0.053. Overall, the gravity coefficient estimate is robust to the rich vector of controls proxying for endogenous location choice.

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 spend money and/or time."[46] 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.[47] 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.[48]

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

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

[48] Full details of the algorithm are found here: Home Identification Algorithm.

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).[49] 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.

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

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

& 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 Correlates of Branch Fixed Effects

The estimated branch fixed effects proxy for all attributes of branches that make them destinations for residents of any block group, controlling for residents' distances away and residents' demand for branch products or services, as proxied by the block group fixed effects. To provide some suggestive evidence on what might influence the branch fixed effects we estimate, we explore their correlation with observable branch characteristics, in particular, branch property values, square footage, and hours/days open. Online Appendix D.1 presents detailed definitions of the characteristics. Descriptive statistics of all characteristics are in Online Table A.17.

Branch property values and square footage are from CoreLogic, a popular data provider of real estate information. CoreLogic data are from 2021 and are matched by latitude-longitude. Branch hours and days open are from SafeGraph itself.

Online Fig. A.5 reports bivariate OLS regressions of the estimated branch fixed effects on the branch characteristics. We compute the correlations using the panel of branch fixed effects. Roughly 77% of the variation over time in a typical branch's fixed effect can be explained by the branch's identity. Standard errors are clustered at the branch level. In each regression, the estimated branch fixed effects and characteristics have been normalized to have mean zero and standard deviation one. The figure displays simple correlations and need not reflect causal effects, but the results follow intuitive patterns.

Branches with higher fixed effects have higher property market values. This positive correlation may reflect several factors about the building, its condition, the profitability of the branch, and the attractiveness of the surrounding location that draws in visitors. Larger branches (i.e., those with higher square footage) also have higher fixed effects. Branches with higher fixed effects also have higher price/sq. ft. Finally, branches with higher fixed effects are open for longer hours during Monday through Friday, and they are also open on weekends, particularly on Saturday. It is not surprising that larger branches, those with longer open hours, and those open on the weekend attract more visitors, all else equal. These characteristics should be encoded in a branch's fixed effect, and they are also arguably measures of quality.[50]

D.1 Definitions of Branch Characteristics

The following are the definitions of the branch characteristics we correlate with the estimated branch fixed effects. Names of data fields are provided in parentheses.

- *Property Value* - natural logarithm of the total market value of a parcel's land and improvement values as provided by county or local taxing/assessment authority (market_total_value)
- *Square Footage* - natural logarithm of the building square footage that reflects the most accurate available for use in assessments/comparables (universal_building_sqft)
- *Price/Sq. Ft.* - natural logarithm of the ratio of property value to square footage

[50] We considered correlating the branch fixed effects with measures of local competition—such as the number of competing entrants or exiters within a certain radius—to assess whether competition is associated with a branch's estimated "quality." But we decided against it for three reasons. First, a branch's fixed effect tends to vary little over time despite the entry or exit of other branches in the sample, which may suggest a small effect. Second, from the results above, the branch fixed effects already appear to correlate in the cross-section with observable measures of quality. Third, the effects of competition on a branch's fixed effect depend on the time it takes the market to reach equilibrium, which is hard to judge empirically. To see why, consider the following two hypothetical scenarios: (i) one Bank of America branch is located 1 mile from a block group, and (ii) one Bank of America and one Wells Fargo are located 1 mile from the block group. In the short run when the total number of visitors is held fixed, the Bank of America branch's estimated fixed effect in the second scenario may register lower than in the first scenario because the same number of branch goers may split between the two branches. Hence, competition may make it appear as if branch quality is lower. However, in the longer run, competition may induce the Bank of America and Wells Fargo branches to offer promotions or add amenities or services that would draw more customers in on the extensive margin, raising the estimated fixed effects of both locations. Just as importantly, even if local competition may lower a branch's estimated fixed effect in the short run, *branch access*, which is the paper's focus, need not lower because a new branch is now available to block groups.

- *Weekday Open Hours* - number of hours a branch is open from Monday-Friday (open_hours). For the minority of branches with conflicting hours listed (e.g., "Mon": [["7:00", "21:00"], ["8:30", "17:00"], ["9:00", "16:00"]]), the narrowest window of open hours is used.
- *Open Saturday* - dummy equal to 1 if a branch is open on Saturday (open_hours)
- *Open Sunday* - dummy equal to 1 if a branch is open on Sunday (open_hours)

E Correlates of Black-White Gaps in Access, Demand, and Visitors

Here, we provide full details of the analysis in Section 6.4 of the text. For Black-White gaps in branch access, we run the specification in column (2) of Table 2 county-by-county and gather the estimated coefficients on the Black population share. For Black-White gaps in demand and expected visitors, we run the specifications in column (2) and column (6), respectively, of Table 3 county-by-county, and we collect the estimated coefficients on the Black population share. The set of coefficients are county-specific Black-White gaps in branch access, demand, and expected visitors. We then study the correlation of these racial gaps with county-level measures of neighborhood crime and racial bias against Blacks among Whites.

E.1 Crime Indices

Crime indices are from the 2022 vintage of the CrimeRisk database that is generated by Applied Geographic Solutions and distributed by Esri. The primary source of CrimeRisk is the FBI Uniform Crime Reports (UCR), which compile crime statistics from 18,000 law enforcement agencies across the U.S. Crime in America is known to exhibit extreme spatial variation (Harries 2006), and the UCR database is limited to only the most populous parts of the country. AGS constructs the CrimeRisk indices by starting with the UCR data over the past seven years, adds crime reports from a handful of large cities, standardizes the data across jurisdictions, and then implements a statistical predictive model to produce crime index values for all U.S. block groups. AGS states that its model includes Census socioeconomic characteristics, but no data are used related to local residents' ethnicity, race, ancestry, or language spoken at home.

CrimeRisk produces indices for seven different categories of crime, and each category is modeled separately. It also provides aggregate indices for personal, property, and total crime, where each aggregate index equally weights its constituent categories, which accords with the reporting procedures used in the UCR. Block-group CrimeRisk scores are indexed to the national level, which has a score of 100. A block group with a score of 100 implies that its expected risk for that crime is close to the national average, whereas a score of 200, for instance, implies an expected doubling in the risk of that crime.[51] Accurately predicting crime rates at the very local level based on data from larger jurisdictions in the UCR and Census information is quite challenging, and no study has validated the CrimeRisk indices across the entire country. Nau et al. (2020) evaluates the validity of the indices by comparing them to crime rate data from the Los Angeles Police Department (LAPD). They find that five indices correlate fairly well and predict LAPD crime rates at the Census tract level: (1) robbery, (2) murder, (3) assault, (4) motor vehicle theft, and (5) personal crime, where (5) is an unweighted average of (1)-(3) and the index for rape. With this finding in mind, we focus on these five indices in our analysis, but we also consider the total crime index, which combines all the CrimeRisk categories across personal and property crimes (i.e., robbery, murder, assault, rape, larceny, burglary, motor vehicle theft).

The analysis regresses estimated Black-White gaps in branch access across counties (i.e., the county-specific loadings on the Black population share from the specification in column 2 of Table 2) onto the CrimeRisk indices, where we divide the indices by 100. Counties with less than 20 Census block groups with estimated access measures over the sample period are dropped. Observations are weighted by county population counts from the 2019 5-year ACS. We also consider how the crime indices are associated with the Black-White gaps in branch demand (i.e., the county-specific loadings on the Black population share from the specification in column 6 of Table 3) and in expected branch visitors (i.e., the county-specific loadings on the Black population share from the specification in column 2 of Table 3).

Online Table A.9 has the results. In all cases, areas with higher levels of expected crime risk have larger estimated Black-White gaps in branch access. All relations are precisely estimated except for motor vehicle theft. As examples, in counties having a 10% higher expected risk of robbery, the access gap is higher by 0.02, and in counties with a 10% higher expected risk of murder, the access gap is higher by 0.015. By comparison, the Black-White access gap

[51] https://appliedgeographic.com/2020/07/faq-crimerisk/.

nationwide is 0.053 (column 2 of Table 2). Associations between expected crime risk and the Black-White gaps in branch demand and in expected branch visitors are imprecisely estimated.

E.2 Racial Bias

Data on racial bias are from the Project Implicit Database (Xu et al. 2014). We use both implicit and explicit measures of bias. The *implicit* measure is based on the Implicit Association Test (IAT) from Greenwald et al. (1998), administered through an online response module. Respondents are shown pictures of faces (Black or White) and words (associated with good or bad). They use the same set of buttons on the keyboard to classify the faces into Black or White categories and words into good or bad categories. IAT is based on the premise that if a respondent has a stronger association in mind between being White and being good, the classification exercise will take longer when they have to use the same button to classify a face as being Black and a word as being good. The implicit bias measure we use, the *D* score, is the difference in time it takes to classify when Black faces and good words are paired together versus when Black faces and bad words are paired together.[52]

The *explicit* measure of racial bias is based on Project Implicit's "thermology" questions. Respondents are asked whether they "feel warmer toward" White Americans and whether they "feel warmer toward" Black Americans, and respond on a 0-to-10 scale to each question. We subtract the latter from the former to form the explicit bias measure so that a higher value means the respondent feels warmer toward White Americans than toward Black Americans. This thermology-based explicit measure has an advantage over the directly elicited preference, because by asking each preference independently and then taking the difference, it parallels the symmetric nature of the IAT.

Data for both measures span 2003-2017, with roughly 250,000 tests completed each year. We only use test results from non-Hispanic White respondents. A value of zero for either measure represents no racial bias against Blacks, and higher levels imply greater racial bias. Because the online tests are voluntary, the data are subject to potential self-selection bias. To help address potential selection, we also construct "adjusted" versions of both racial bias measures, which are the residuals from projecting respondents' racial bias measures on respondent age, race, gender, education, and test variables (i.e., the month, hour, weekday, and order of test). County-level measures are simple averages per month of the raw and adjusted measures among non-Hispanic White respondents residing in the county.

The analysis regresses estimated Black-White gaps in branch access across counties (i.e., the county-specific loadings on the Black population share from the specification in column 2 of Table 2) onto the measures of racial bias. Counties with less than 20 Census block groups with estimated access measures over the sample period are dropped. Observations are weighted by county population counts from the 2019 5-year ACS. We also consider how racial bias is associated with the Black-White gaps in branch demand (i.e., the county-specific loadings on the Black population share from the specification in column 6 of Table 3) and in expected branch visitors (i.e., the county-specific loadings on the Black population share from the specification in column 2 of Table 3).

Online Table A.10 has the results with the implicit bias measure (raw and adjusted), whereas Online Table A.11 has the results with the explicit bias measure (raw and adjusted). The raw (adjusted) measures as independent variables are in the odd (even) columns of both tables. In all cases, areas with higher levels of implicit or explicit racial bias against Blacks among Whites have larger, precisely estimated Black-White gaps in branch access. For IAT, the mean raw score cross-county is 0.4 and standard deviation is 0.037. Based on the coefficient in column (1) of Online Table A.10, this implies that in counties with a one standard deviation higher level of implicit racial bias against Blacks, the access gap is higher by 0.18. By comparison, the Black-White access gap nationwide is 0.053 (column 2 of Table 2). For the thermology-based explicit measure, the mean raw value cross-county is 0.75 and standard deviation is 0.226. Based on the coefficient in column (1) of Online Table A.11, this implies that in counties with a one standard deviation higher level of explicit racial bias against Blacks, the access gap is higher by 0.20. Associations between racial bias and the Black-White gaps in branch demand and in expected branch visitors are imprecisely estimated.

[52] Recent meta-studies find that IAT is a fairly good predictor of racial discrimination (Oswald, Mitchell, Blanton, Jaccard and Tetlock 2013). The IAT has also been used frequently in economics and other social sciences to measure racial bias (e.g., Reuben, Sapienza and Zingales 2014; Lowes, Nunn, Robinson and Weigel 2015, 2017; Carlana 2019; Chetty et al. 2020). Nonetheless, concerns have been raised over the validity of implicit measures of racial bias, for example with test-retest reliability (Blanton and Jaccard 2008; Schimmack 2021). And IAT is known to better predict interracial behavior and the average racial bias of a group than a single person's racial bias (Greenwald, Poehlman, Uhlmann and Banaji 2009). For these reasons, we complement the implicit measure with an explicit one.

F 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 F.1, we discuss survey findings about bank branch use; in Online Appendix F.2, we analyze differences by demographic characteristics in the primary methods that banked respondents use to access their bank accounts; and in Online Appendix F.3, we compare reported branch visitor shares according to household income from the survey to observed shares from the SafeGraph geolocation data.

F.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.2, 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. 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.

F.2 Primary Bank Access Methods by Respondent Demographics

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.3, 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.4 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.

F.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.[53]

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 household 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.

G 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 geolocation 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).

[53]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.

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 50th percentile and below. This close alignment of the two distributions over these percentiles suggests that individual-level behavior based on income can be inferred quite accurately from the grouped data over this income range. As the percentiles get farther above the median, however, the gap between the two distributions grows substantially. Individual-level household income at the top percentiles is over twice as large as block-group-level median household income. This divergence is unsurprising, as calculating the median household income naturally compresses the distribution across block groups.

When faced with an ecological inference problem, how 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."

H 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 asses 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 γ_{it}, (ii) the fixed effects of both postal and private bank branches λ_{jt}, (iii) the distances between the block group and branches Distance_{ij}, (iv) the gravity parameters β_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 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. In many microfounded gravity models, β_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 β_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 γ_{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 $\left\{\lambda_{jt}\right\}$ 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^{th} percentile, 50^{th} percentile, and 90^{th} 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^{th} 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^{th} 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 2, but in computing Φ_{it} per Census block group, we now include the locations of all post office branches that are registered in SafeGraph within a 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.18. 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 2 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.18 is -5.8%, which is slightly higher than the coefficient in Table 2. 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.

I 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

$$U_{rjt} = \frac{z_{rjt}\Lambda_{jt}}{\delta_{ij}}. \tag{A.77}$$

The term Λ_{jt} is an index of all attributes of branch j that make it a destination for residents of any block group at time t. 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 δ_{ij} is an iceberg traveling cost that is defined as

$$\delta_{ij} = d_{ij}^{\kappa}, \tag{A.78}$$

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, Redding, Sturm and Wolf (2015) by assuming that the idiosyncratic component of utility, z_{rjt}, is drawn from an independent Fréchet distribution:

$$F\left(z_{rjt}\right) = e^{-H_{jt} z_{rjt}^{-\varepsilon}}, \quad H_{jt} > 0,\ \varepsilon > 1. \tag{A.79}$$

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 ε implies that residents are more heterogeneous in their preferences for branches.[54]

Substituting the expression for U_{rjt} into the distribution of idiosyncratic tastes in Eq. (A.79), one can observe that residents of block group i at time t are presented with a distribution of utility across branches, $G_{ijt}(u) = \Pr\left[U_{rjt} \leq u\right] = F\left(u\delta_{ij}/\Lambda_{jt}\right)$, or

$$G_{ijt}(u) = e^{-\left[H_{jt}\left(\frac{\Lambda_{jt}}{\delta_{ij}}\right)^{\varepsilon}\right]u^{-\varepsilon}}. \tag{A.80}$$

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

$$G_{it}(u) = \Pi_{j=0}^{B_t} G_{ijt}(u). \tag{A.81}$$

Inserting Eq. (A.80) into Eq. (A.81), one obtains the utility distribution:

$$G_{it}(u) = e^{-(1+\Psi_{it})u^{-\varepsilon}}, \tag{A.82}$$

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

$$\Psi_{it} = \sum_{j\in B_t} H_{jt}\Lambda_{jt}^{\varepsilon} d_{ij}^{-\kappa\varepsilon}. \tag{A.83}$$

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

$$\pi_{ijt} = \frac{H_{jt}\Lambda_{jt}^{\varepsilon} d_{ij}^{-\kappa\varepsilon}}{1+\Psi_{it}}. \tag{A.84}$$

The visitor share depends on the characteristics of the branch $\left(\Lambda_{jt}\right)$, the average utility draw of the branch $\left(H_{jt}\right)$, and the "bilateral resistance" derived from the intervening transportation costs $\left(d_{ij}^{-\kappa\varepsilon}\right)$. 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, Ψ_{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.[55]

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 (6), 2127–2189.

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 (2), 277–297.

Balassa, B. (1965) "Trade liberalisation and "revealed" comparative advantage 1," *The manchester school*, 33 (2), 99–123.

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

Blanton, H. and J. Jaccard (2008) "Unconscious racism: A concept in pursuit of a measure," *Annu. Rev. Sociol*, 34, 277–297.

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

[55]We obtain the gravity relation in Eq. (A.84) by evaluating:

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

Carlana, M. (2019) "Implicit stereotypes: Evidence from teachers' gender bias," *Quarterly Journal of Economics*, 134 (3), 1163–1224.

Chetty, R., N. Hendren, M. R. Jones, and S. R. Porter (2020) "Race and Economic Opportunity in the United States: an Intergenerational Perspective*," *Quarterly Journal of Economics*, 135 (2), 711–783, 10.1093/qje/qjz042.

Eaton, J. and S. Kortum (2002) "Technology, geography, and trade," *Econometrica*, 70 (5), 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*, https://www.gillibrand.senate.gov/news/press/release/senators-gillibrand-and-sanders-representatives-ocasio-cortez-pascrell-and-kaptur-call-on-congress-to-im

Greenwald, A. G., D. E. McGhee, and J. L. Schwartz (1998) "Measuring individual differences in implicit cognition: the implicit association test.," *Journal of personality and social psychology*, 74 (6), 1464.

Greenwald, A. G., T. A. Poehlman, E. L. Uhlmann, and M. R. Banaji (2009) "Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity.," *Journal of personality and social psychology*, 97 (1), 17.

Guimaraes, P. and P. Portugal (2010) "A simple feasible procedure to fit models with high-dimensional fixed effects," *The Stata Journal*, 10 (4), 628–649.

Harries, K. (2006) "Extreme spatial variations in crime density in Baltimore County, MD," *Geoforum*, 37 (3), 404–416.

Head, K. and T. Mayer (2014) "Gravity equations: Workhorse, toolkit, and cookbook," in *Handbook of International Economics*, 4, 131–195.

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 (2), 381–405.

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.

Lowes, S., N. Nunn, J. A. Robinson, and J. Weigel (2015) "Understanding ethnic identity in africa: Evidence from the implicit association test (iat)," *American Economic Review*, 105 (5), 340–345.

Lowes, S., N. Nunn, J. A. Robinson, and J. L. Weigel (2017) "The evolution of culture and institutions: Evidence from the Kuba Kingdom," *Econometrica*, 85 (4), 1065–1091.

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

Nau, C., M. Sidell, K. Clift, C. Koebnick, J. Desai, and D. Rohm-Young (2020) "A commercially available crime index may be a reliable alternative to actual census-tract crime in an urban area," *Preventive medicine reports*, 17, 100996.

O'Hara, M. and D. Easley (1979) "The postal savings system in the depression," *Journal of Economic History*, 39 (3), 741–753.

Oswald, F. L., G. Mitchell, H. Blanton, J. Jaccard, and P. E. Tetlock (2013) "Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies.," *Journal of personality and social psychology*, 105 (2), 171.

Paravisini, D., V. Rappoport, and P. Schnabl (2023) "Specialization in Bank Lending: Evidence from Exporting Firms," *The Journal of Finance*, 78 (4), 2049–2085.

Reuben, E., P. Sapienza, and L. Zingales (2014) "How stereotypes impair women's careers in science," *Proceedings of the National Academy of Sciences*, 111 (12), 4403–4408.

Sanders, B. (2021) "Fair banking for all," *Bernie Sanders Official Website*, https://berniesanders.com/issues/fair-banking-for-all/.

Schimmack, U. (2021) "The Implicit Association Test: A method in search of a construct," *Perspectives on Psychological Science*, 16 (2), 396–414.

Schuster, S. S., M. Jaremski, and E. R. Perlman (2020) "An empirical history of the US Postal Savings System," *Social Science History*, 44 (4), 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 (1), 121–152.

Silva, J. S. and S. Tenreyro (2006) "The log of gravity," *Review of Economics and Statistics*, 88 (4), 641–658.

Warren, E. (2014) "The big benefits of postal service banking," *U.S. News*, https://www.warren.senate.gov/newsroom/op-eds/2014/07/07/us-news-op-ed-the-big-benefits-of-postal-service-banking-1.

Xu, K., B. Nosek, and A. Greenwald (2014) "Psychology data from the race implicit association test on the project implicit demo website," *Journal of Open Psychology Data*, 2 (1).

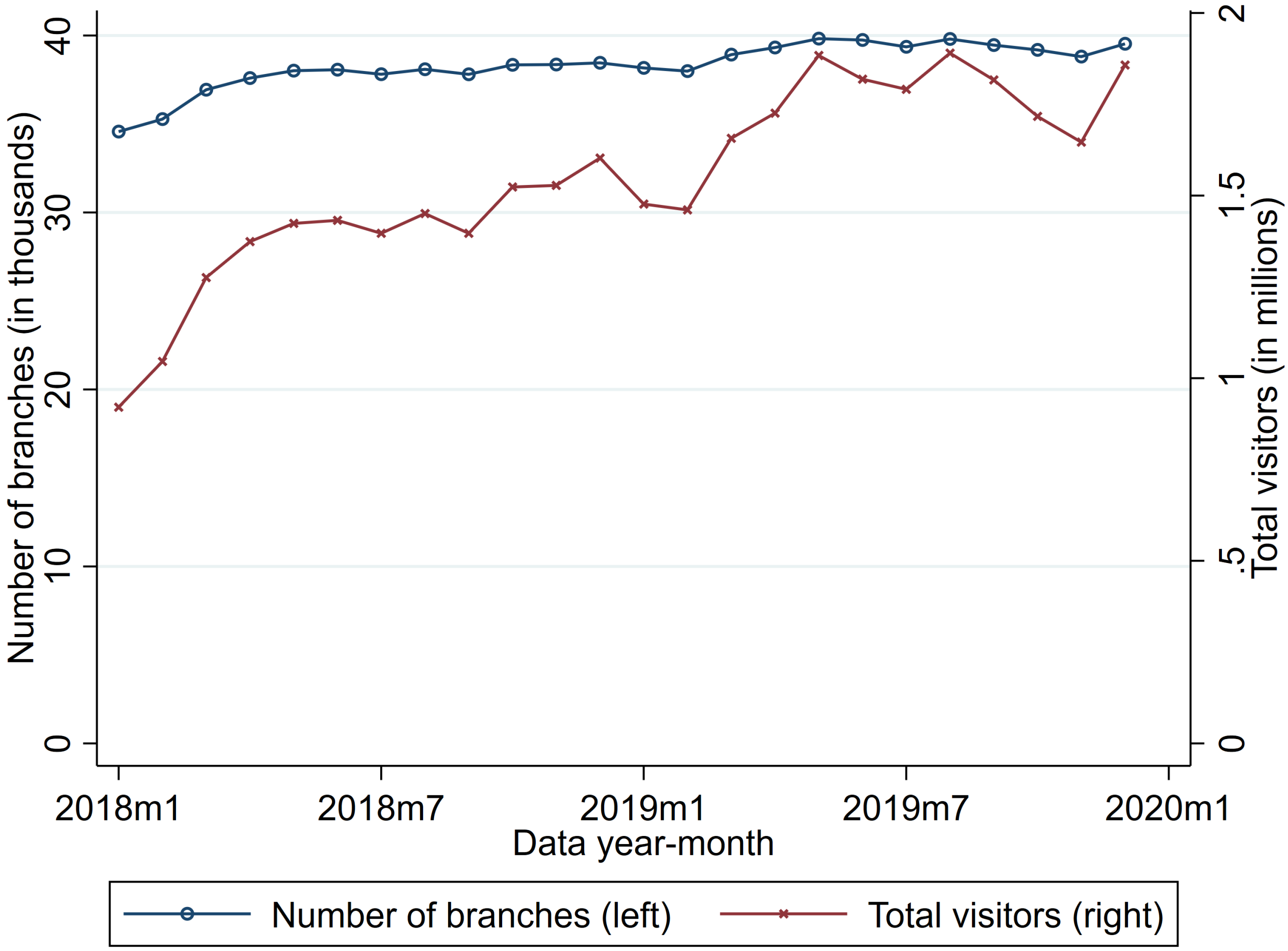


FIGURE A.1
TOTAL BANK BRANCHES AND BRANCH VISITORS - CORE SAMPLE

The figure presents the total number of bank branches and 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.

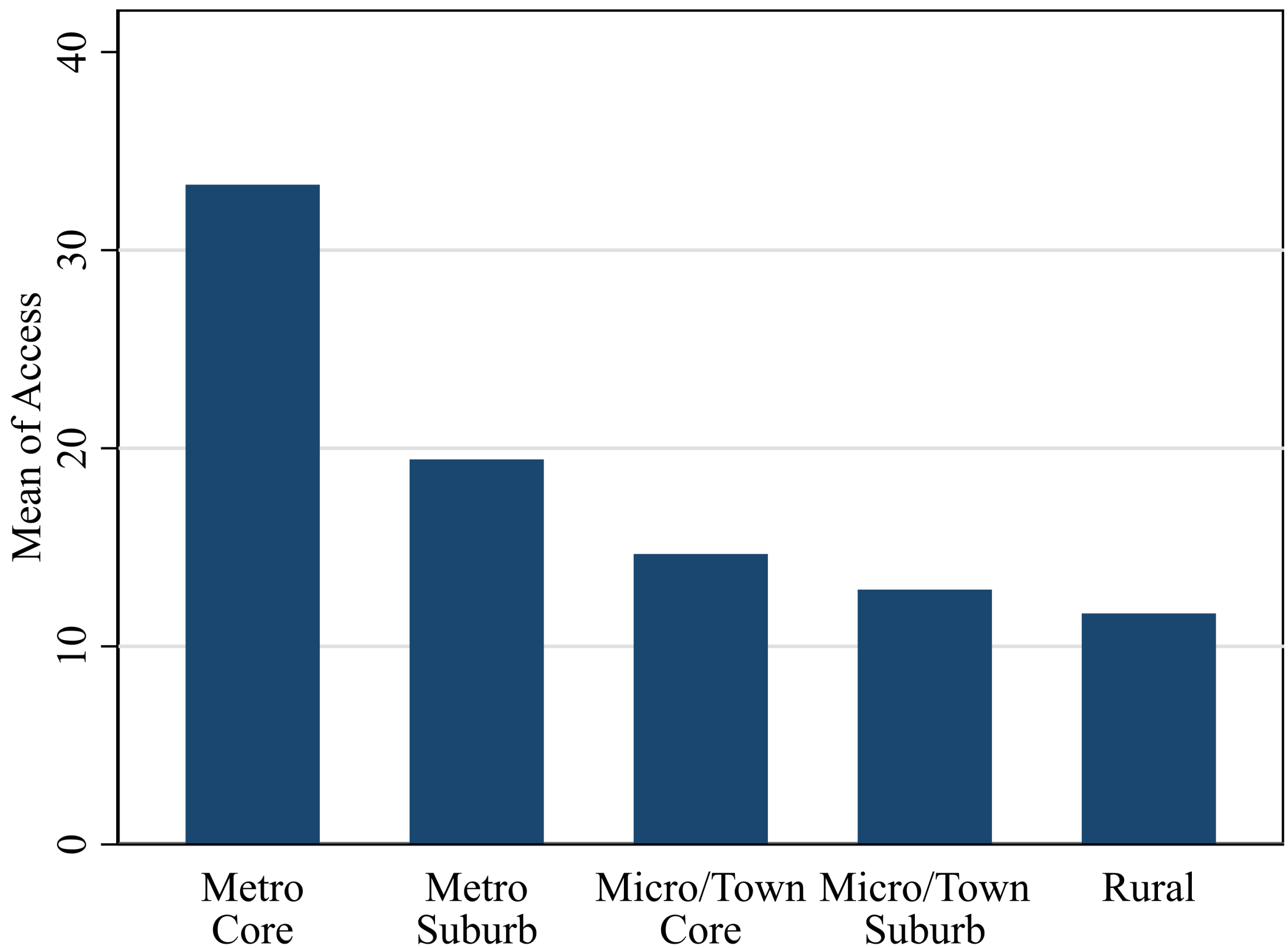


Figure A.2

Bank Branch Access by Rural-Urban Commuting Area

The figure presents weighted average measures of bank branch access by primary Rural-Urban Commuting Area (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 Online 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.

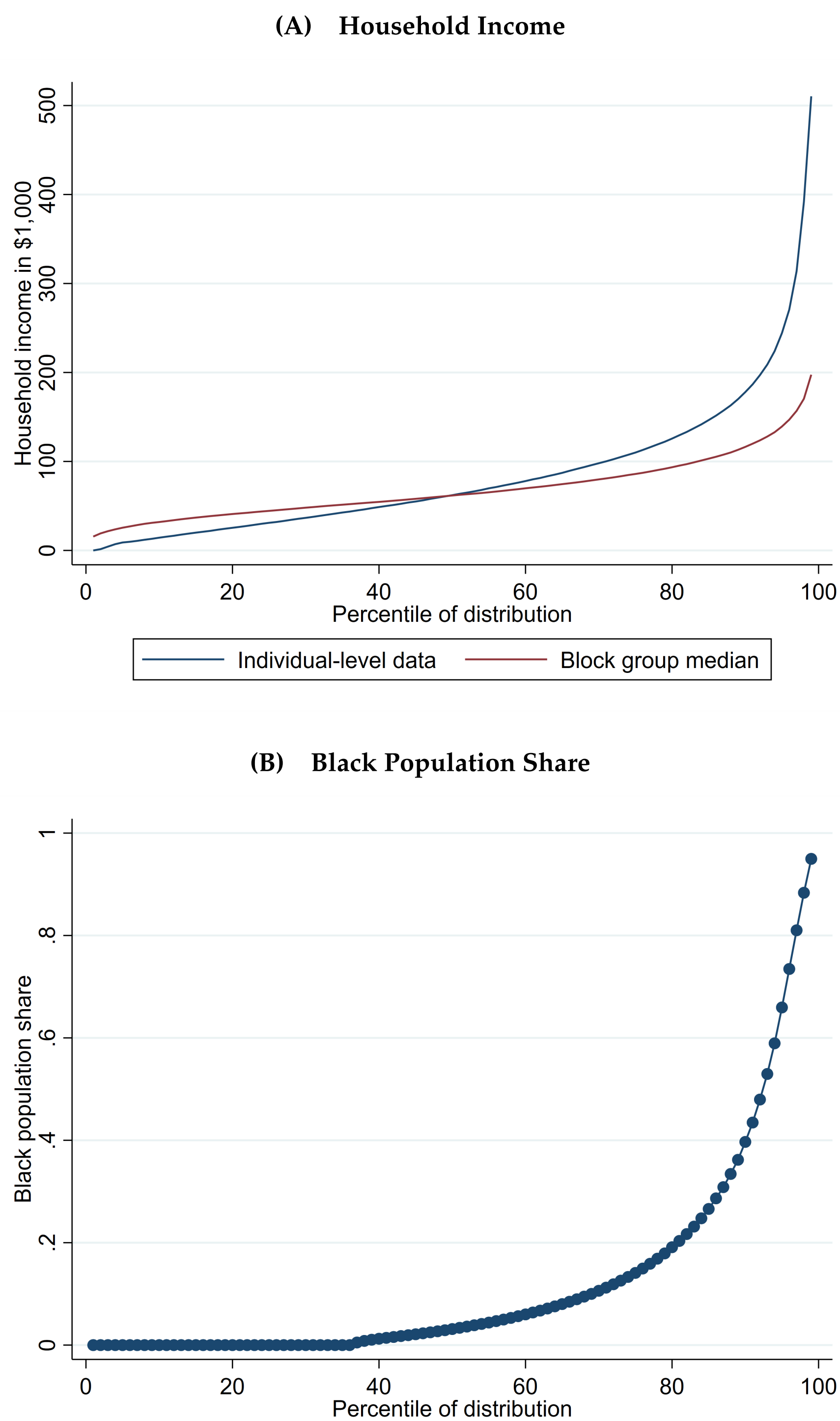


Figure A.3
Distributions of Demographic Attributes

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.

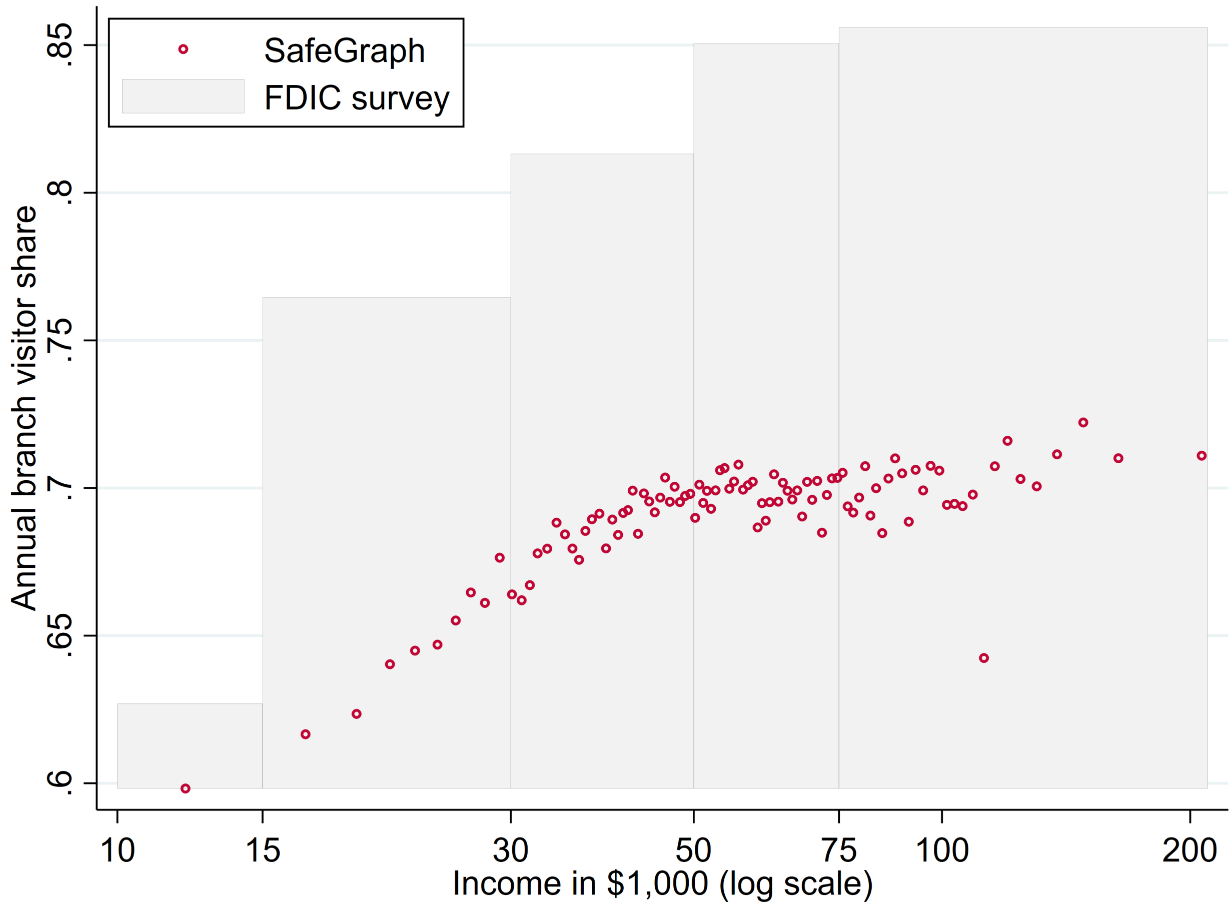


FIGURE A.4

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.

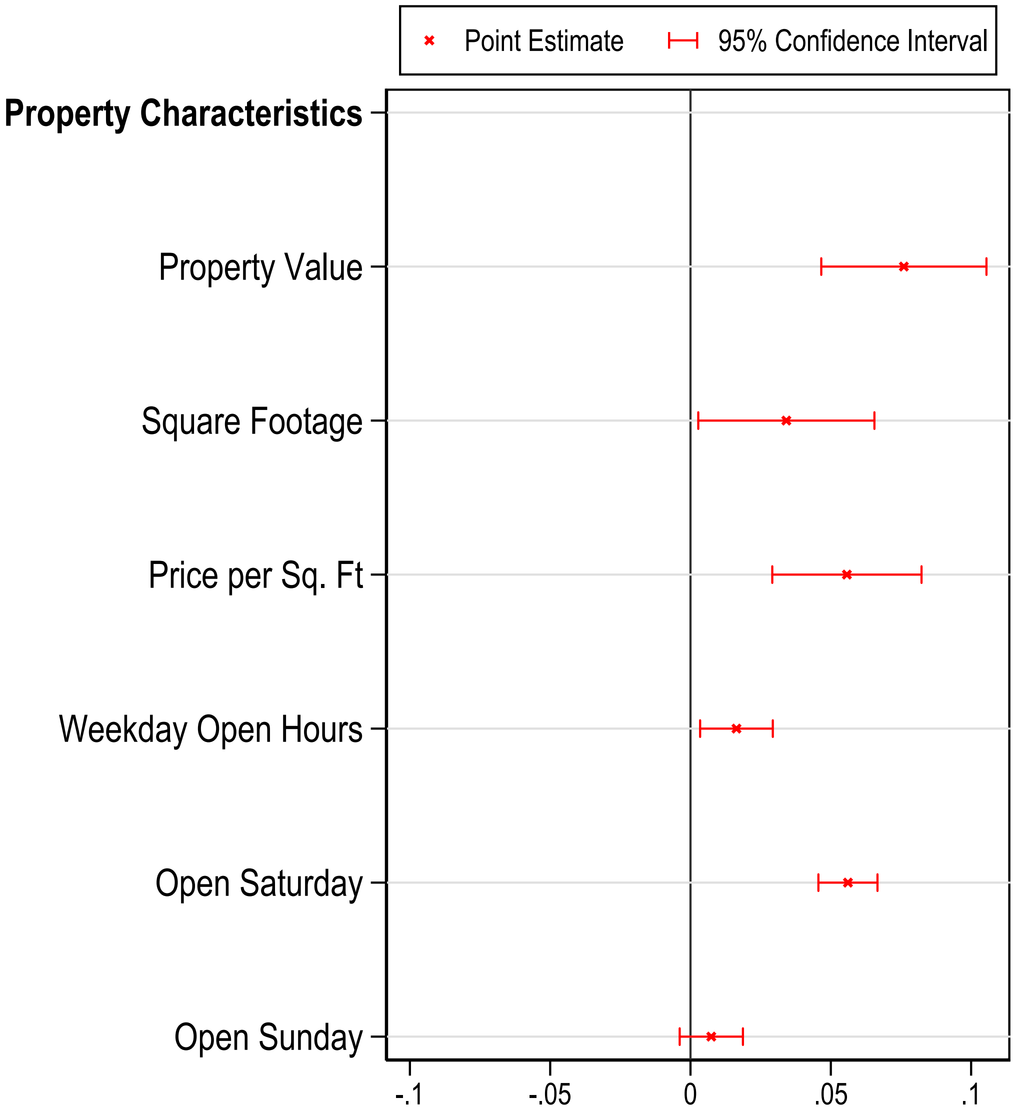
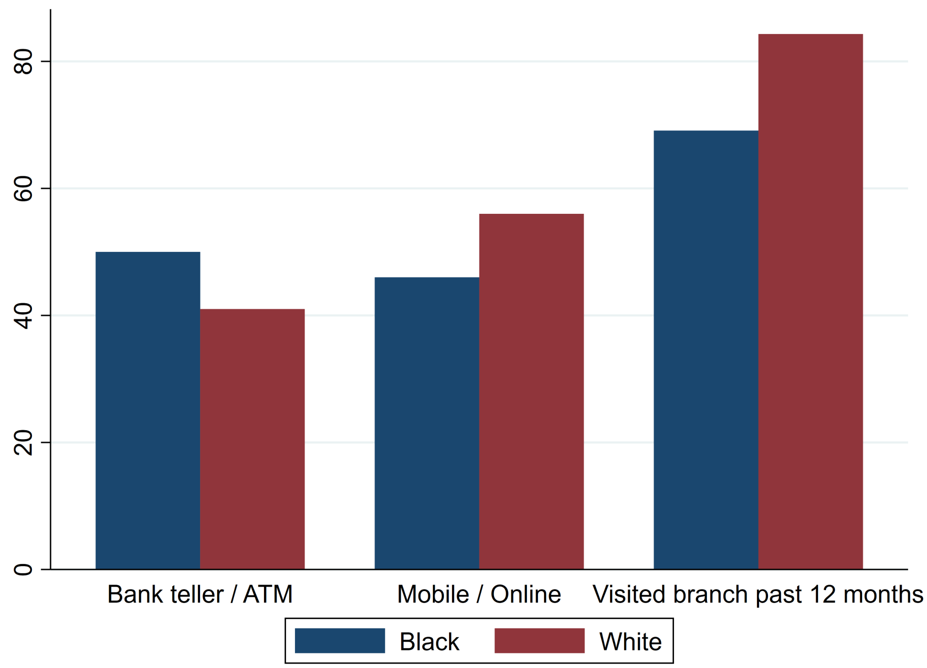


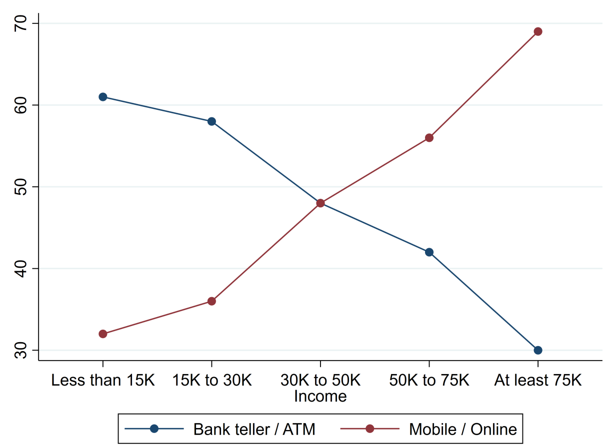
FIGURE A.5
CORRELATIONS OF BRANCH FIXED EFFECTS WITH PROPERTY CHARACTERISTICS

The figure illustrates unweighted bivariate OLS regression results of z-scores of the estimated panel of branch fixed effects, $\{\hat{\lambda}_{jt}\}$, on z-scores of branch property characteristics. Branch fixed effects are estimated from the Method of Simulated Moments, described in Online Appendix A. Online Appendix D.1 provides detailed definitions of the characteristics. The x's in the figure mark coefficient point estimates, whereas the bands are 95% confidence intervals, and standard errors are clustered at the branch level.

(A) Responses by Race: Primary Banking Methods & Branch Visitation



(B) Responses by Income: Primary Banking Methods



(C) Responses by Income: Branch Visitation

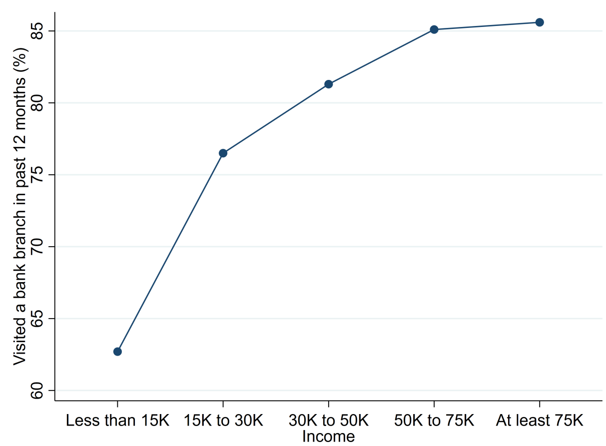


FIGURE A.6

FDIC SURVEY RESPONSES ON PRIMARY BANKING METHODS AND BRANCH VISITATION

The figure presents raw averages of survey responses about primary banking methods and branch visitation by race and income from the 2019 FDIC Survey of Household Use of Banking and Financial Services, conducted in June 2019. Primary banking method is the average of respondents' answers to the most common way they accessed their bank accounts, among banked respondents who accessed their accounts in the past twelve months. Branch visitation is the average of respondents' answers to whether they have visited a bank branch in the past twelve months, among both banked and unbanked respondents. Panel A illustrates average responses among Black and White respondents. Panel B presents average responses by respondent income for primary banking method. Panel C presents average responses by respondent income for branch visitation.

Table A.1
Demographic Attributes of Residents in Areas Represented in Core SafeGraph Sample vs. FDIC SOD

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

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 of each demographic attribute is its sample mean across ZCTAs, whereas the second row is the sample standard deviation. In columns (3) and (6), the first row is the difference in sample means between the core sample and the SOD, whereas the second row is the heteroskedasticity-robust standard error of the estimated difference between the two sample means.

Table A.2
Survey Reported Branch Visit Shares by Household Characteristics

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

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 whether respondents have visited a bank branch in the past twelve months. Demographic independent variable observations are self-reported characteristics of respondents. Income is household income. Coefficients in columns (1)-(3) are from linear probability models estimated using OLS. Coefficients in columns (4)-(6) are from Probit regressions. Omitted demographic categories are household income less than \$15,000, non-Hispanic Whites, and age range 15-34.

Table A.3

Survey Reported Bank Account Primary Access Method by Household Characteristics - Linear Probability Model

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

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.4

Survey Reported Bank Account Primary Access Method by Household Characteristics - Probit Model

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

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.5
Estimated Branch Fixed Effects by Population Density and Local Demographics

Dep. var.:	Branch fixed effects (standardized)	
	(1)	(2)
log(Population density)	-0.004	
	(0.004)	
Black		-0.009
		(0.025)
Asian		-0.195
		(0.056)
Other		0.996
		(0.097)
Hispanic		0.044
		(0.024)
log(Income)		-0.091
		(0.010)
Constant	0.025	0.970
	(0.025)	(0.105)
Observations	918,310	896,692
R^2	0.024	0.029
Year-month FE	O	O
SE cluster	Tract	CBG

Each column reports coefficients from an unweighted OLS regression with standard errors clustered at the Census-tract level (column 1) or Census-block-group level (column 2) reported in parentheses. One observation is a branch per month per year in the sample period from January 2018 - December 2019. 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 densities of the Census tracts where branches are located (column 1) and population-based decimal shares of the Census block groups where branches are located (column 2), recorded in the 2019 5-year American Community Survey (ACS). Income is median household income. The dependent variable is z-scores of the estimated panel of branch fixed effects, $\{\hat{\lambda}_{jt}\}$. The estimation method is described in Section 5, with full details in Online Appendix A. The omitted racial group is non-Hispanic Whites.

Table A.6
Heterogeneous Gravity Coefficients by Race and Income

Dep. var.:	log(No. of visitors$_{ijt}$)	
	(1)	(2)
log(Distance$_{ij}$)	-0.091	-0.088
	(0.003)	(0.002)
log(Distance$_{ij}$) × Black	0.021	
	(0.006)	
log(Distance$_{ij}$) × Asian	-0.013	
	(0.019)	
log(Distance$_{ij}$) × Other	0.066	
	(0.020)	
log(Distance$_{ij}$) × Hispanic	-0.018	
	(0.009)	
log(Distance$_{ij}$) × High Income		-0.018
		(0.003)
log(Distance$_{ij}$) × Low Income		0.015
		(0.002)
Observations	5,625,696	5,549,547

The table reports estimates and standard errors of the gravity coefficients $\beta_{1,t}$ and $\beta_{2,t}$ from the fixed-effects gravity equation:

$$\log\left(\text{No. of visitors}_{ijt}\right) = \gamma_{it} + \lambda_{jt} - \beta_{1,t}\log\left(\text{Distance}_{ij}\right) - \beta_{2,t}\log\left(\text{Distance}_{ij}\right) \times P_i + \varepsilon_{ijt}.$$

Estimates are from an unweighted Poisson pseudo-maximum-likelihood (PPML) estimation, as in Silva and Tenreyro (2006), run using `ppmlhdfe` in Stata. Standard errors are two-way clustered by Census block groups and bank branches. Dependent variable observations are 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. Independent variable observations are the log distances from the population-weighted center of block groups to visited bank branches and the log distances interacted with population-based racial shares (column 1) and interacted with dummies for high-income and low-income block groups (column 2), based on the 2019 5-year American Community Survey. Racial shares and income dummies are denoted P_i in the above equation. High-income is the third tercile of the distribution of block groups' median household incomes, whereas low-income is the first tercile. Omitted demographic groups are non-Hispanic Whites and the second tercile of median household income. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance in miles (see Footnote 39).

Table A.7
Comparison Of Gravity Equation Estimation Methods

Year	Month	MSM β	MSM s.e.	PPML β	PPML s.e.	OLS β	OLS s.e.	OLS where > 4 β	OLS where > 4 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 β_t from the fixed-effects gravity model in Eq. (1):

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

Independent variable observations are the log distances from the population-weighted center of block groups to visited bank branches, where centers of population are from the 2010 Census, and we use the haversine formula to compute distance in miles (see Footnote 39). Columns (3) and (4) present estimates from the Method of Simulated Moments estimation described in Section 5, with full details in Online 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. Dependent variable observations in the PPML and OLS estimations are 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 greater than 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 computed using Eq. (A.65). Standard errors of the PPML and OLS estimates are two-way clustered by both Census block groups and bank branches.

TABLE A.8
DRIVING TIME VERSUS HAVERSINE DISTANCE

Dep. var.:	log(Driving time b/w block group and visited branch)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
log(Haversine distance b/w block group and visited branch)	0.895	0.890	0.885	0.893	0.896	0.901	0.891	0.890	0.902
	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.396	0.430	0.450	0.402	0.395	0.356	0.423	0.442	0.342
	(0.001)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	995,000	725,000	35,000	498,000	497,000	498,000	497,000	508,000	487,000
Adjusted R^2	0.988	0.989	0.987	0.989	0.987	0.986	0.989	0.987	0.988
Sample	Core	MC	Core	Core	Core	Core	Core	Core	Core
Black > 0.8			O						
Black ≥ Med. Black				O					
Black < Med. Black					O				
White ≥ Med. White						O			
White < Med. White							O		
log(Income) ≥ 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 × 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 natural log 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 natural log haversine distances between block groups and branches. 995,000 block group × branch pairs were drawn randomly. Column (1) includes the entire random sample of block group × branch pairs. Column (2) restricts the sample to block groups with Rural-Urban Commuting Areas (RUCA) codes equaling 1 (Metropolitian 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.9
Black-White Gaps and Neighborhood Crime Risk

(a) Robbery

Dep. var.:	Black-White Gap in		
	Access	Demand	Expected visitors
	(1)	(2)	(3)
Robbery	-0.207	0.106	-0.101
	(0.045)	(0.068)	(0.071)
Constant	0.447	-0.244	0.203
	(0.074)	(0.126)	(0.130)
Observations	1,531	1,531	1,531
R^2	0.009	0.001	0.000

(b) Assault

Dep. var.:	Black-White Gap in		
	Access	Demand	Expected visitors
	(1)	(2)	(3)
Assault	-0.136	0.009	-0.127
	(0.064)	(0.082)	(0.065)
Constant	0.365	-0.141	0.224
	(0.086)	(0.128)	(0.115)
Observations	1,531	1,531	1,531
R^2	0.002	0.000	0.000

(c) Personal Crime

Dep. var.:	Black-White Gap in		
	Access	Demand	Expected visitors
	(1)	(2)	(3)
Personal crime	-0.207	0.050	-0.157
	(0.060)	(0.086)	(0.082)
Constant	0.438	-0.183	0.256
	(0.084)	(0.139)	(0.137)
Observations	1,531	1,531	1,531
R^2	0.004	0.000	0.000

(d) Murder

Dep. var.:	Black-White Gap in		
	Access	Demand	Expected visitors
	(1)	(2)	(3)
Murder	-0.152	0.101	-0.051
	(0.038)	(0.042)	(0.034)
Constant	0.385	-0.236	0.149
	(0.070)	(0.097)	(0.087)
Observations	1,531	1,531	1,531
R^2	0.006	0.001	0.000

(e) Motor Vehicle Theft

Dep. var.:	Black-White Gap in		
	Access	Demand	Expected visitors
	(1)	(2)	(3)
Motor Vehicle Theft	-0.050	-0.005	-0.055
	(0.056)	(0.078)	(0.067)
Constant	0.280	-0.127	0.153
	(0.068)	(0.104)	(0.104)
Observations	1,531	1,531	1,531
R^2	0.000	0.000	0.000

(f) Total Crime

Dep. var.:	Black-White Gap in		
	Access	Demand	Expected visitors
	(1)	(2)	(3)
Total crime	-0.174	-0.185	-0.359
	(0.088)	(0.183)	(0.195)
Constant	0.406	0.056	0.462
	(0.102)	(0.222)	(0.250)
Observations	1,531	1,531	1,531
R^2	0.001	0.000	0.001

Each column reports coefficients from a univariate, weighted OLS regression with heteroskedasticity-robust standard errors reported in parentheses. One observation is a county. Counties with less than 20 Census block groups with estimated branch access measures over the sample period are dropped. Observations are weighted by county population counts from the 2019 5-year American Community Survey (ACS). Independent variable observations are crime risk indices at the county-level, obtained from the CrimeRisk database. In columns (1)-(2), the dependent variable is the estimated Black-White gap in branch access across counties (i.e., the county-specific loading on the Black population share from the specification in column 2 of Table 2). In columns (3)-(4), the dependent variable is the estimated Black-White gap in branch demand (i.e., the county-specific loading on the Black population share from the specification in column 6 of Table 3). In columns (5)-(6), the dependent variable is the estimated Black-White gap in expected branch visitors (i.e., the county-specific loading on the Black population share from the specification in column 2 of Table 3).

Table A.10
Black-White Gaps and Implicit Racial Bias

Dep. var.:	Black-White Gap in Branch Access		Black-White Gap in Branch Demand		Black-White Gap in Expected Branch Visitors	
	(1)	(2)	(3)	(4)	(5)	(6)
IAT score for Whites	-4.816	-4.898	-4.358	-2.841	-9.175	-7.739
	(2.301)	(2.176)	(7.108)	(5.926)	(8.655)	(7.052)
Constant	2.121	0.612	1.580	0.090	3.701	0.702
	(0.929)	(0.199)	(2.837)	(0.511)	(3.453)	(0.606)
Observations	1,531	1,531	1,531	1,531	1,531	1,531
R^2	0.007	0.006	0.001	0.000	0.005	0.003
Adj. Racial Bias		O		O		O

Each column reports coefficients from a univariate, weighted OLS regression with heteroskedasticity-robust standard errors reported in parentheses. One observation is a county. Counties with less than 20 Census block groups with estimated branch access measures over the sample period are dropped. Observations are weighted by county population counts from the 2019 5-year American Community Survey (ACS). Independent variable observations are measures of implicit racial bias against Blacks based on mean scores on implicit association tests (IATs) for non-Hispanic White participants by county, obtained from the Project Implicit Database. Raw test scores are used in odd columns, whereas adjusted test scores are used in even columns. Adjusted test scores are the residuals from projecting raw scores on respondent age, race, gender, education, and test variables (the month, hour, weekday, and order of test). In columns (1)-(2), the dependent variable is the estimated Black-White gap in branch access across counties (i.e., the county-specific loading on the Black population share from the specification in column 2 of Table 2). In columns (3)-(4), the dependent variable is the estimated Black-White gap in branch demand (i.e., the county-specific loading on the Black population share from the specification in column 6 of Table 3). In columns (5)-(6), the dependent variable is the estimated Black-White gap in expected branch visitors (i.e., the county-specific loading on the Black population share from the specification in column 2 of Table 3).

Table A.11
Black-White Gaps and Explicit Racial Bias

Dep. var.:	Black-White Gap in Branch Access		Black-White Gap in Branch Demand		Black-White Gap in Expected Branch Visitors	
	(1)	(2)	(3)	(4)	(5)	(6)
Thermology score for Whites	-0.867	-0.932	0.434	0.361	-0.433	-0.571
	(0.301)	(0.311)	(0.390)	(0.405)	(0.323)	(0.339)
Constant	0.872	0.656	-0.454	-0.298	0.418	0.358
	(0.260)	(0.178)	(0.338)	(0.233)	(0.292)	(0.207)
Observations	1,531	1,531	1,531	1,531	1,531	1,531
R^2	0.009	0.009	0.001	0.000	0.000	0.001
Adj. Racial Bias		O		O		O

Each column reports coefficients from a univariate, weighted OLS regression with heteroskedasticity-robust standard errors reported in parentheses. One observation is a county. Counties with less than 20 Census block groups with estimated branch access measures over the sample period are dropped. Observations are weighted by county population counts from the 2019 5-year American Community Survey (ACS). Independent variable observations are measures of explicit racial bias against Blacks based on mean scores on Project Implicit's "thermology" questions for non-Hispanic White participants by county, obtained from the Project Implicit Database. Respondents are asked whether they "feel warmer toward" White Americans and whether they "feel warmer toward" Black Americans, and respond on a 0-to-10 scale to each question. We subtract the latter from the former to form the explicit bias measure so that a higher value means the respondent feels warmer toward White Americans than toward Black Americans. Raw measures are used in odd columns, whereas adjusted measures are used in even columns. Adjusted measures are the residuals from projecting raw measures on respondent age, race, gender, education, and test variables (the month, hour, weekday, and order of test). In columns (1)-(2), the dependent variable is the estimated Black-White gap in branch access across counties (i.e., the county-specific loading on the Black population share from the specification in column 2 of Table 2). In columns (3)-(4), the dependent variable is the estimated Black-White gap in branch demand (i.e., the county-specific loading on the Black population share from the specification in column 6 of Table 3). In columns (5)-(6), the dependent variable is the estimated Black-White gap in expected branch visitors (i.e., the county-specific loading on the Black population share from the specification in column 2 of Table 3).

Table A.12

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	-0.084	-0.147	-0.093	-0.124	-0.081	-0.140	-0.091
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Black	-0.187	-0.156	-0.226	-0.178	-0.148	-0.118	-0.181	-0.135
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
Asian	0.504	0.453	0.454	0.404	0.492	0.447	0.445	0.402
	(0.016)	(0.016)	(0.017)	(0.017)	(0.015)	(0.015)	(0.015)	(0.015)
Other	-0.012	-0.039	0.049	0.012	0.003	-0.015	0.061	0.035
	(0.022)	(0.022)	(0.032)	(0.031)	(0.022)	(0.022)	(0.032)	(0.031)
Hispanic	-0.045	-0.012	-0.087	-0.037	-0.010	0.023	-0.045	0.004
	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)
Age <15		-0.962		-1.089		-0.869		-0.984
		(0.019)		(0.022)		(0.018)		(0.021)
Age 35-54		-0.350		-0.310		-0.307		-0.270
		(0.018)		(0.022)		(0.017)		(0.021)
Age 55-64		-0.799		-0.835		-0.705		-0.736
		(0.019)		(0.023)		(0.018)		(0.022)
Age 65+		-0.383		-0.418		-0.332		-0.360
		(0.014)		(0.016)		(0.013)		(0.015)
log(No. of devices)	-0.057	-0.063	-0.065	-0.072	-0.054	-0.058	-0.061	-0.066
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(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	O	O	O	O	O	O	O	O
County FE	O	O	O	O	O	O	O	O
RUCA FE	O	O			O	O		

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 Online 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 (Metropolitian area core). The omitted demographic groups are non-Hispanic Whites and age range 15-34.

Table A.13
Balassa-style Shares Descriptive Statistics

			Percentiles				
	N	Mean	5th	25th	50th	75th	95th
Branch customer specialization:							
Share Asian	48,813	0.05	0.00	0.01	0.02	0.05	0.17
Share Asian (MDI)	133	0.26	0.07	0.15	0.25	0.36	0.49
Share Black	48,813	0.10	0.00	0.02	0.06	0.13	0.35
Share Black (MDI)	27	0.52	0.18	0.35	0.54	0.71	0.77
Share Hispanic	48,813	0.14	0.01	0.04	0.08	0.18	0.47
Share Hispanic (MDI)	150	0.70	0.29	0.50	0.79	0.89	0.93
Share other	48,813	0.03	0.01	0.02	0.03	0.03	0.06
Share White	48,813	0.68	0.25	0.56	0.73	0.85	0.94
Log income	48,813	10.93	10.09	10.75	11.00	11.24	11.58
Block group brand loyalty:							
Wells Fargo	212,475	0.11	0.00	0.00	0.03	0.18	0.44
BoA	212,475	0.11	0.00	0.00	0.04	0.18	0.43
Chase	212,475	0.11	0.00	0.00	0.02	0.17	0.45
US Bank	212,475	0.04	0.00	0.00	0.00	0.00	0.24
PNC	212,475	0.04	0.00	0.00	0.00	0.01	0.22

The table reports descriptive statistics of the Balassa (1965)-style share variables added as controls to the gravity equation in Eq. (1), as described in Online Appendix B.2. The top half of the table presents shares relating to branches' customer-specific specialization, whereas the bottom-half presents shares relating to block group residents' brand-specific loyalty. In the top half, the racial shares are defined in Eq. (A.72), whereas the income shares are defined in Eq. (A.73). Income is median household income. MDI stands for Minority Depository Institutions, whose definition is here. We use the 2019 list of MDIs. Rows with the "(MDI)" label include only branches belonging to such institutions whose owners are also of that race. For example, "Share Black (MDI)" includes only Black-owned MDI branches. The other rows in the top half include both MDI and non-MDI branches in the core sample. In the bottom half, brand-specific loyalty shares are for the top 5 brands by number of branches and defined in Eq. (A.74).

Table A.14
Gravity Equation with Balassa-style Level Shares

	PPML				OLS			
Dep. var.:	No. of visitors$_{ijt}$				log(No. of visitors$_{ijt}$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Distance$_{ij}$)	-0.090	-0.088	-0.088	-0.087	-0.053	-0.053	-0.053	-0.052
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Spec$_j$ × White		0.152		0.156		0.074		0.078
		(0.064)		(0.064)		(0.024)		(0.023)
Spec$_j$ × Black		0.028		0.016		0.065		0.058
		(0.060)		(0.060)		(0.023)		(0.023)
Spec$_j$ × Asian		-0.934		-0.959		-0.309		-0.323
		(0.318)		(0.317)		(0.093)		(0.093)
Spec$_j$ × Hispanic		-0.611		-0.552		-0.221		-0.186
		(0.206)		(0.203)		(0.125)		(0.123)
Spec$_j$ × Other		-0.244		-0.258		-0.128		-0.136
		(0.062)		(0.061)		(0.025)		(0.025)
Spec$_j$ × log(Income)		-0.215		-0.214		-0.127		-0.127
		(0.019)		(0.019)		(0.009)		(0.009)
BL$_i$ × Wells Fargo			-0.156	-0.154			-0.121	-0.121
			(0.037)	(0.036)			(0.017)	(0.017)
BL$_i$ × BoA			-0.314	-0.309			-0.172	-0.173
			(0.053)	(0.052)			(0.025)	(0.024)
BL$_i$ × Chase			-0.268	-0.257			-0.164	-0.160
			(0.046)	(0.044)			(0.020)	(0.019)
BL$_i$ × US Bank			-0.118	-0.116			-0.069	-0.068
			(0.060)	(0.061)			(0.033)	(0.033)
BL$_i$ × PNC			-0.037	-0.036			-0.022	-0.020
			(0.067)	(0.073)			(0.028)	(0.029)
Constant	1.721	27.933	1.734	27.868				
	(0.003)	(2.340)	(0.003)	(2.338)				
Observations	5,549,547	5,549,547	5,549,547	5,549,547	5,549,547	5,549,547	5,549,547	5,549,547
R^2	0.714	0.724	0.715	0.725	0.477	0.484	0.478	0.485

The table reports estimates and standard errors of the fixed-effects gravity equation augmented with Balassa-style share controls, presented in Eq. (A.75). Columns (1) to (4) present estimates from an unweighted Poisson pseudo-maximum-likelihood (PPML) estimation, as in Silva and Tenreyro (2006), run using `ppmlhdfe` in Stata. Columns (5) to (8) present estimates from an unweighted OLS regression. Standard errors are two-way clustered by Census block groups and bank branches. Dependent variable observations are the raw number of visitors (columns 1 to 4) and the log raw number of visitors (columns 5 to 8) 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. Independent variable observations are (i) the log distances from the population-weighted center of block groups to visited bank branches, (ii) branch-level specialization share variables (Spec$_j$) given in Eqs. (A.72) to (A.73) interacted with block group racial shares and log median household incomes from the 2019 5-year ACS, and (iii) block group-level brand loyalty shares (BL$_i$) for the top 5 bank brands by number of branches given in Eq. (A.74) interacted with dummies equaling 1 if the branch visited belongs to the particular brand and 0 otherwise. The PPML R^2s are computed using the method described here. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance in miles (see Footnote 39).

Table A.15
Gravity Equation with Balassa-style Quartile Shares

	PPML				OLS			
Dep. var.:	No. of visitors$_{ijt}$				log(No. of visitors$_{ijt}$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(Distance$_{ij}$)	-0.090	-0.091	-0.089	-0.090	-0.053	-0.054	-0.053	-0.054
	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Spec$_j$ Q2 × White		0.014		0.013		0.003		0.003
		(0.024)		(0.024)		(0.008)		(0.008)
Spec$_j$ Q3 × White		0.013		0.012		0.003		0.003
		(0.028)		(0.028)		(0.010)		(0.010)
Spec$_j$ Q4 × White		-0.057		-0.057		-0.032		-0.032
		(0.035)		(0.035)		(0.015)		(0.015)
Spec$_j$ Q2 × Black		-0.145		-0.141		-0.066		-0.064
		(0.032)		(0.032)		(0.017)		(0.017)
Spec$_j$ Q3 × Black		-0.105		-0.103		-0.051		-0.050
		(0.035)		(0.035)		(0.018)		(0.018)
Spec$_j$ Q4 × Black		-0.119		-0.118		-0.041		-0.040
		(0.042)		(0.042)		(0.019)		(0.019)
Spec$_j$ Q2 × Asian		0.154		0.163		0.099		0.105
		(0.084)		(0.084)		(0.035)		(0.036)
Spec$_j$ Q3 × Asian		0.132		0.144		0.093		0.100
		(0.080)		(0.080)		(0.035)		(0.035)
Spec$_j$ Q4 × Asian		0.050		0.060		0.043		0.049
		(0.076)		(0.076)		(0.034)		(0.034)
Spec$_j$ Q2 × Hispanic		0.105		0.110		0.037		0.038
		(0.155)		(0.155)		(0.051)		(0.051)
Spec$_j$ Q3 × Hispanic		-0.033		-0.028		-0.008		-0.006
		(0.153)		(0.152)		(0.054)		(0.054)
Spec$_j$ Q4 × Hispanic		-0.095		-0.087		-0.050		-0.047
		(0.141)		(0.141)		(0.054)		(0.054)

TABLE A.15 (CONTINUED)

Dep. var.:	PPML				OLS			
	No. of visitors_{ijt}				$\log(\text{No. of visitors}_{ijt})$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Spec_j Q2 × Other		-0.121		-0.116		-0.064		-0.061
		(0.029)		(0.029)		(0.015)		(0.015)
Spec_j Q3 × Other		-0.266		-0.258		-0.154		-0.149
		(0.039)		(0.039)		(0.017)		(0.016)
Spec_j Q4 × Other		-0.258		-0.253		-0.163		-0.160
		(0.042)		(0.042)		(0.018)		(0.017)
Spec_j Q2 × log(Income)		-0.005		-0.005		-0.003		-0.003
		(0.002)		(0.002)		(0.001)		(0.001)
Spec_j Q3 × log(Income)		-0.017		-0.017		-0.010		-0.010
		(0.003)		(0.003)		(0.001)		(0.001)
Spec_j Q4 × log(Income)		-0.028		-0.028		-0.016		-0.016
		(0.004)		(0.004)		(0.002)		(0.002)
BL_i Q2 × Wells Fargo			-0.018	-0.017			-0.005	-0.004
			(0.008)	(0.008)			(0.005)	(0.004)
BL_i Q3 × Wells Fargo			0.025	0.027			0.011	0.012
			(0.008)	(0.008)			(0.004)	(0.004)
BL_i Q4 × Wells Fargo			-0.020	-0.019			-0.018	-0.017
			(0.009)	(0.009)			(0.004)	(0.004)
BL_i Q2 × BoA			0.002	0.003			0.005	0.005
			(0.011)	(0.011)			(0.005)	(0.005)
BL_i Q3 × BoA			0.024	0.024			0.016	0.017
			(0.010)	(0.010)			(0.004)	(0.004)
BL_i Q4 × BoA			-0.044	-0.043			-0.023	-0.022
			(0.012)	(0.011)			(0.005)	(0.005)
BL_i Q2 × Chase			-0.028	-0.028			-0.009	-0.009
			(0.010)	(0.010)			(0.006)	(0.006)
BL_i Q3 × Chase			0.024	0.025			0.013	0.013
			(0.008)	(0.008)			(0.004)	(0.004)
BL_i Q4 × Chase			-0.035	-0.033			-0.019	-0.018

Table A.15 (CONTINUED)

Dep. var.:	PPML				OLS			
	No. of visitors$_{ijt}$				log(No. of visitors$_{ijt}$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			(0.012)	(0.012)			(0.005)	(0.005)
BL$_i$ Q2 $\times$ US Bank			0.000	0.000			0.000	0.000
			(.)	(.)			(.)	(.)
BL$_i$ Q3 $\times$ US Bank			-0.264	-0.264			-0.114	-0.114
			(0.102)	(0.102)			(0.029)	(0.029)
BL$_i$ Q4 $\times$ US Bank			0.046	0.047			0.007	0.007
			(0.029)	(0.029)			(0.016)	(0.016)
BL$_i$ Q2 $\times$ PNC			0.000	0.000			0.000	0.000
			(.)	(.)			(.)	(.)
BL$_i$ Q3 $\times$ PNC			-0.050	-0.049			-0.027	-0.027
			(0.010)	(0.010)			(0.005)	(0.005)
BL$_i$ Q4 $\times$ PNC			-0.022	-0.021			-0.013	-0.012
			(0.006)	(0.006)			(0.004)	(0.004)
Constant	1.721	1.928	1.720	1.925				
	(0.003)	(0.026)	(0.003)	(0.026)				
Observations	5,549,547	5,549,547	5,549,547	5,549,547	5,549,547	5,549,547	5,549,547	5,549,547
R^2	0.714	0.716	0.715	0.717	0.477	0.478	0.478	0.479

The table reports estimates and standard errors of the fixed-effects gravity equation augmented with dummies for the quartiles of the distributions of Balassa-style share controls, presented in Eq. (A.76). Columns (1) to (4) present estimates from an unweighted Poisson pseudo-maximum-likelihood (PPML) estimation, as in Silva and Tenreyro (2006), run using `ppmlhdfe` in Stata. Columns (5) to (8) present estimates from an unweighted OLS regression. Standard errors are two-way clustered by Census block groups and bank branches. Dependent variable observations are the raw number of visitors (columns 1 to 4) and the log raw number of visitors (columns 5 to 8) 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. Independent variable observations are (i) the log distances from the population-weighted center of block groups to visited bank branches, (ii) dummies for the quartiles of branch-level specialization share variables (Spec$_j$) given in Eqs. (A.72) to (A.73) interacted with block group racial shares and log median household incomes from the 2019 5-year ACS, and (iii) dummies for the quartiles of block group-level brand loyalty shares (BL$_i$) for the top 5 bank brands by number of branches given in Eq. (A.74) interacted with dummies equaling 1 if the branch visited belongs to the particular brand, and 0 otherwise. The quartile dummies equal 1 if the branch's (or block group's) share is in the particular quartile of the distribution of the relevant share variable, and 0 otherwise. The bottom quartiles of the distributions are omitted. The PPML R^2s are computed using the method described here. Centers of population are from the 2010 Census, and we use the haversine formula to compute distance in miles (see Footnote 39).

Table A.16
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	4.002	2.787	-421.996	-455.377
	(2.661)	(2.476)	(1.860)	(29.371)	(39.185)
Constant	3.780			3.306	
	(0.006)			(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.17
Descriptive Statistics of Branch Characteristics

Variable Name	Mean	Std. Dev.	N
Property Value ($1M)	1.60	4.69	66,533
Square Footage (1K)	8.36	41.71	66,533
Price per Sq. Ft. ($)	269.81	288.91	66,533
Weekday Open Hours	40.40	8.68	487,897
Open Saturday (%)	61.86	48.57	487,897
Open Sunday (%)	3.87	19.30	487,897

The table reports descriptive statistics of branch-level property value and square footage variables from CoreLogic and branch days/hours open from SafeGraph. Variable definitions are provided in Online Appendix D.1. For property value, square footage, and price/sq. ft., we include branches that appear in both CoreLogic and our core sample of branch locations. Our core sample 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. The number *N* of observations per variable is the number of branch × year-month observations used in the bivariate OLS regressions of the estimated branch fixed effects on that variable, whose results are in Online Fig. A.5.

Table A.18
Bank Branch Access by Demographic Attributes under Postal Banking

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

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 Online 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^{th} percentile of the distribution of estimated private bank fixed effects in the year-month. In columns (3)-(6), we assign the 50^{th} percentile; and in columns (7) and (8), we assign the 90^{th} 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 (Metropolitian area core). The omitted demographic groups are non-Hispanic Whites and age range 15-34.