

Discussion of
**Bank Branch Access:
Evidence from Geolocation Data**

Jung Sakong and Alexander K. Zantefis

Bronson Argyle¹

UNC, June 2023

¹Brigham Young University

What's the Basic Idea?

- Use proprietary geolocation micro data to analyze the variation in bank access (quality $\times$ proximity) and demand by income and race.
 - Block group-level geographic data from SafeGraph: privacy protection schemes from SafeGraph make the analysis more tricky (truncation and censoring), so
 - Develop a Method of Simulated Moments to deal with the privacy-related data obfuscation.

What's the Basic Idea?

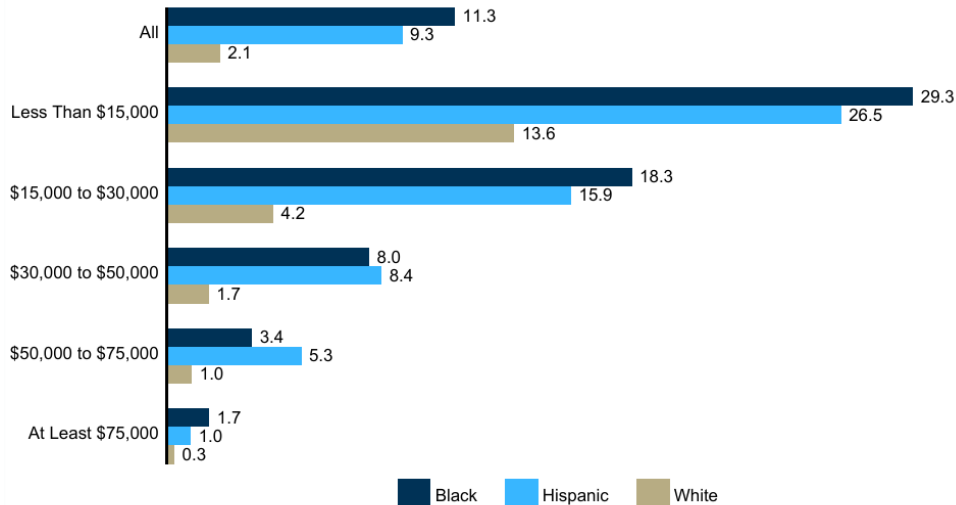
- Use proprietary geolocation micro data to analyze the variation in bank access (quality $\times$ proximity) and demand by income and race.
 - Block group-level geographic data from SafeGraph: privacy protection schemes from SafeGraph make the analysis more tricky (truncation and censoring), so
 - Develop a Method of Simulated Moments to deal with the privacy-related data obfuscation.
- Low-income and predominantly-black census block groups (CBG) are less likely to visit bank branches.
 - Low-Income CBGs: better access and no difference in branch quality; **less demand**.
 - Predominantly black CBGs: **worse access**, not because of quality but **because of proximity**; if anything, higher demand.

Comments Outline

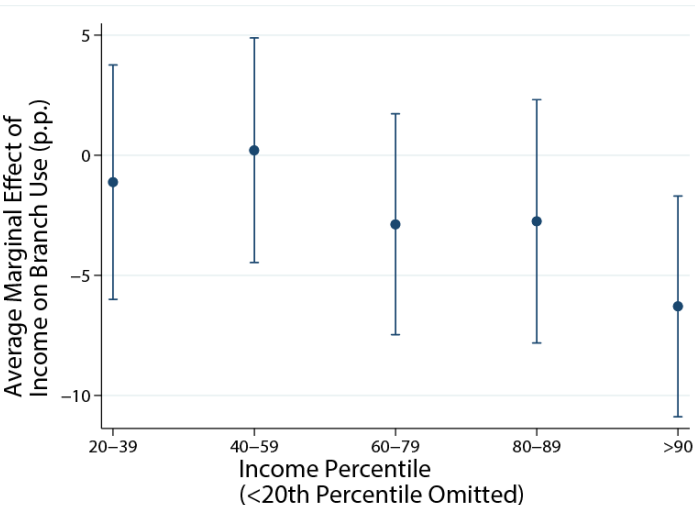
- I found the paper fun to read and very engaging. I think it's an important topic.
- Outline for my thoughts:
 - Contribution & Background
 - Data
 - Interpretation & Policy Implications
 - Smaller issues/suggestions

2021 FDIC National Survey of Unbanked and UnderBanked Households

Figure ES.2 Unbanked Rates by Income Level and Race and Ethnicity, 2021 (Percent)

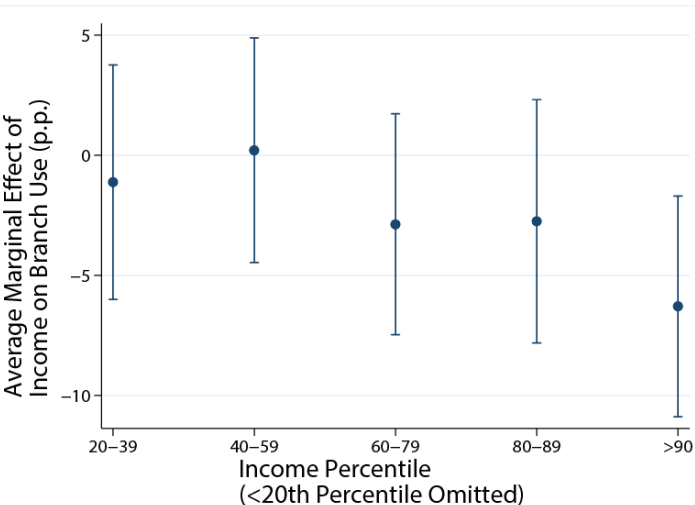


2016 Survey of Consumer Finances



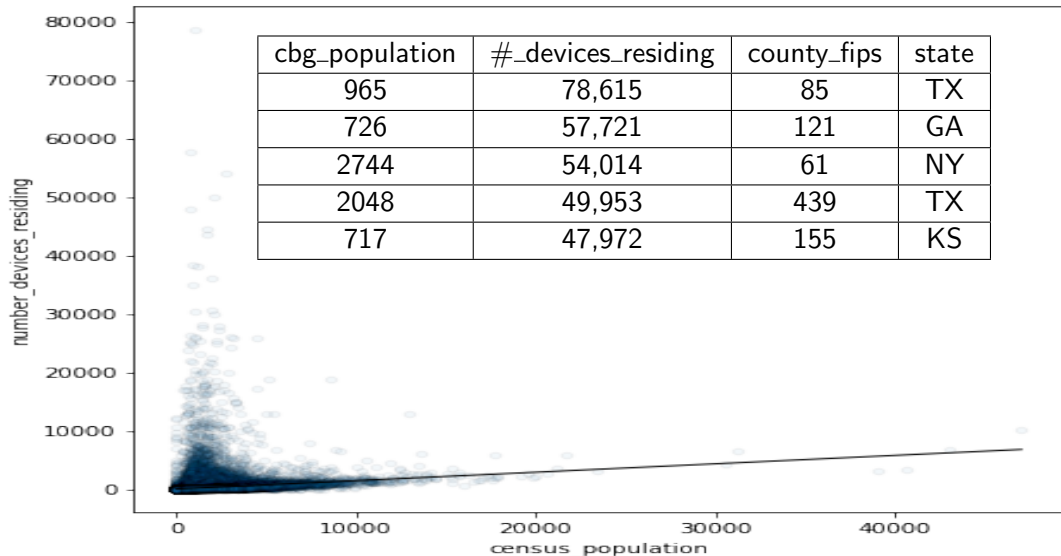
- How to square the findings in this paper regarding the relationship between income and branch use and the 2016 SCF?
 - SCF: Higher income $\Rightarrow$ less visits to bank branches
 - SZ: Higher income $\Rightarrow$ more visits to bank branches

2016 Survey of Consumer Finances



- How to square the findings in this paper regarding the relationship between income and branch use and the 2016 SCF?
 - SCF: Higher income $\Rightarrow$ less visits to bank branches
 - SZ: Higher income $\Rightarrow$ more visits to bank branches
- Also, higher income is less likely to use a bank teller/ATM (Table A.4), but more likely to visit a branch (Table 2).
- With such great data, this difference is probably indicating interesting economics.

SafeGraph Census Block Group (CBG) Bias



Representativeness 1/2

- The block groups in SafeGraph data seem representative (Table A.2), but a very significant 2.8 pp difference in Urban vs Rural.

Representativeness 1/2

- The block groups in SafeGraph data seem representative (Table A.2), but a very significant 2.8 pp difference in Urban vs Rural.
- Two layers of representativeness: SG CBGs vs population CBGs, and smartphone users in a CBG vs. CBG population.
 - Are smartphone users representative of the block group population behavior? This seems likely to not hold in low-income/rural block groups.

Representativeness 1/2

- The block groups in SafeGraph data seem representative (Table A.2), but a very significant 2.8 pp difference in Urban vs Rural.
- Two layers of representativeness: SG CBGs vs population CBGs, and smartphone users in a CBG vs. CBG population.
 - Are smartphone users representative of the block group population behavior? This seems likely to not hold in low-income/rural block groups.
- How should I think about access to fintechs and credit unions and other unmodeled financial service providers?
 - I think SafeGraph is likely to have data on these Points of Interest (POIs) as well. Why not use it?

Representativeness 2/2

- The ACS block group population numbers are 5-year averages. Using the averages to reweight is likely to introduce a bias in areas with volatile population numbers (little baby robustness check: what if we used the 2018 numbers for reweighting?).

Representativeness 2/2

- The ACS block group population numbers are 5-year averages. Using the averages to reweight is likely to introduce a bias in areas with volatile population numbers (little baby robustness check: what if we used the 2018 numbers for reweighting?). Ideally, we would use block group population numbers for the year of the data, but the 1-year data only cover counties and places with populations $> 65,000$.

Representativeness 2/2

- The ACS block group population numbers are 5-year averages. Using the averages to reweight is likely to introduce a bias in areas with volatile population numbers (little baby robustness check: what if we used the 2018 numbers for reweighting?). Ideally, we would use block group population numbers for the year of the data, but the 1-year data only cover counties and places with populations $> 65,000$. This is problematic given that these are the exact areas that are likely to have an unrepresentative population number in the 5-year average data, e.g. Vineyard, UT grew by 18.5% between 2018 and 2019.

Representativeness 2/2

- The ACS block group population numbers are 5-year averages. Using the averages to reweight is likely to introduce a bias in areas with volatile population numbers (little baby robustness check: what if we used the 2018 numbers for reweighting?). *Ideally, we would use block group population numbers for the year of the data, but the 1-year data only cover counties and places with populations > 65,000.* This is problematic given that these are the exact areas that are likely to have an unrepresentative population number in the 5-year average data, e.g. Vineyard, UT grew by 18.5% between 2018 and 2019.
 - Proposal: (in addition to the tables in the paper currently) could use the 1-year data (but can only use CBGs with >65k).

Representativeness and Estimation

- 84% of the block-level bank visitations counts are equal to 4 (truncation and censoring).
- The 75th percentile of branch-level bank visitation is equal to 48 unique visitors per month (the median is 23).

Representativeness and Estimation

- 84% of the block-level bank visitations counts are equal to 4 (truncation and censoring).
- The 75th percentile of branch-level bank visitation is equal to 48 unique visitors per month (the median is 23).
 - How do these numbers compare to actual visitation (even for a sample of branches)? This would speak to how representative smartphone users are.

Representativeness and Estimation

- 84% of the block-level bank visitations counts are equal to 4 (truncation and censoring).
- The 75th percentile of branch-level bank visitation is equal to 48 unique visitors per month (the median is 23).
 - How do these numbers compare to actual visitation (even for a sample of branches)? This would speak to how representative smartphone users are.
- With so few (<100) visitors/month, the normal approximation of the Poisson process is likely **inconsistent**. Consider using the exact method using the χ^2 - distribution to get the standard errors right. (implementation in scipy: `scipy.special.gammaincinv`)

Representativeness and Estimation

- 84% of the block-level bank visitations counts are equal to 4 (truncation and censoring).
- The 75th percentile of branch-level bank visitation is equal to 48 unique visitors per month (the median is 23).
 - How do these numbers compare to actual visitation (even for a sample of branches)? This would speak to how representative smartphone users are.
- With so few (<100) visitors/month, the normal approximation of the Poisson process is likely **inconsistent**. Consider using the exact method using the χ^2 - distribution to get the standard errors right. (implementation in scipy: `scipy.special.gammaincinv`)
- Instead of the post-hoc stratified re-weighting, consider a full-blown Heckman correction to deal with the multi-dimensional sampling bias.

Presenting Magnitudes

- How should I benchmark the effects?
 - A 1% increase in CBG-level income is associated with a 0.15% increase in visitors. Given a mean visitor count of 40 (How should this be scaled?), increasing the income by 10% is associated with another 0.6 people at the average branch.
 - A 10 pp increase in the share of Black individuals (mean share is 0.103) in a CBG is associated with a reduction in visitors by -0.56%, so 2 less people at the average branch.

Presenting Magnitudes

- How should I benchmark the effects?
 - A 1% increase in CBG-level income is associated with a 0.15% increase in visitors. Given a mean visitor count of 40 (How should this be scaled?), increasing the income by 10% is associated with another 0.6 people at the average branch.
 - A 10 pp increase in the share of Black individuals (mean share is 0.103) in a CBG is associated with a reduction in visitors by -0.56%, so 2 less people at the average branch.
- Standardizing these coefficients may be a more intuitive way to present these results. (IMO):
 - dos d'enveloppe: A one standard deviation increase in income (Black share) is associated with (ish) 2.5 more (3.2 less) visitors at the average branch. (again, scaling issue).

Measurement of Quality

- Quality is measured as a weighted average of branch fixed effects from a visitations per month regression. But is this really quality? ~~I think I might just need a little more hand-holding here.~~ Yes, I think it is the right measure of the revealed preferences of customers.
 - But can I conclude that if $\lambda_{jt} > \lambda_{j^*t}$ that the quality of branch j is higher than j^* ?
 - What is the distribution of this measure through time and space and does this seem sensible?

Policy?

- Given the endogeneity of household and branch location, there is no causal identification here. I think the channel of these effects must be pinned down to really push for policy implications (what is done in the paper is still a very useful and necessary project).

Policy?

- Given the endogeneity of household and branch location, there is no causal identification here. I think the channel of these effects must be pinned down to really push for policy implications (what is done in the paper is still a very useful and necessary project).
- For example: what if low-income/black households are more likely to use mobile banking?
 - how should I think about mobile banking in general (given that these data are conditional upon having a smartphone)?

Table ES.2 Primary Method of Bank Account Access, 2017–2021

Banked Households That Accessed Their Account in the Past 12 Months, Row Percent

Year	Bank Teller	ATM/Kiosk	Telephone Banking	Online Banking	Mobile Banking	Other	Bank Teller or ATM/Kiosk
2017	24.8	19.5	2.9	37.0	15.1	0.7	44.3
2019	21.0	19.5	2.4	22.8	34.0	0.3	40.5
2021	14.9	16.0	2.9	22.0	43.5	0.7	31.0

(2021 FDIC National Survey of Unbanked and UnderBanked Households)

Policy?

- Given the endogeneity of household and branch location, there is no causal identification here. I think the channel of these effects must be pinned down to really push for policy implications (what is done in the paper is still a very useful and necessary project).
- For example: what if low-income/black households are more likely to use mobile banking?
 - how should I think about mobile banking in general (given that these data are conditional upon having a smartphone)?

Table ES.2 Primary Method of Bank Account Access, 2017–2021

Banked Households That Accessed Their Account in the Past 12 Months, Row Percent

Year	Bank Teller	ATM/Kiosk	Telephone Banking	Online Banking	Mobile Banking	Other	Bank Teller or ATM/Kiosk
2017	24.8	19.5	2.9	37.0	15.1	0.7	44.3
2019	21.0	19.5	2.4	22.8	34.0	0.3	40.5
2021	14.9	16.0	2.9	22.0	43.5	0.7	31.0

(2021 FDIC National Survey of Unbanked and UnderBanked Households)

- For policy, what really matters is the elasticity of usage with respect to proximity and quality.**

Mechanism

Quality

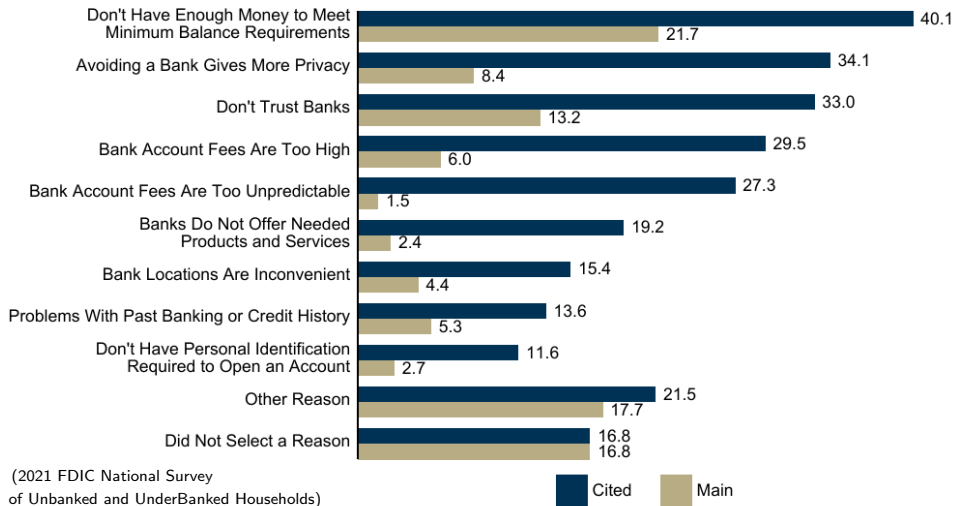
- ~~What about looking at branch usage around the introduction of mobile banking apps for specific banks?~~
 - This is likely to be outside of the time series (2018-2019).
- What about parent-bank level change in branch hours? Is there latent demand?
- What about introduction of new products (e.g. CDs, HELOCs) by the parent-bank?

Proximity

- What about Nguyen (2019)-style instrument exploiting branch closure following parent bank merge activity?
- Completion of public transit projects (definitely random).

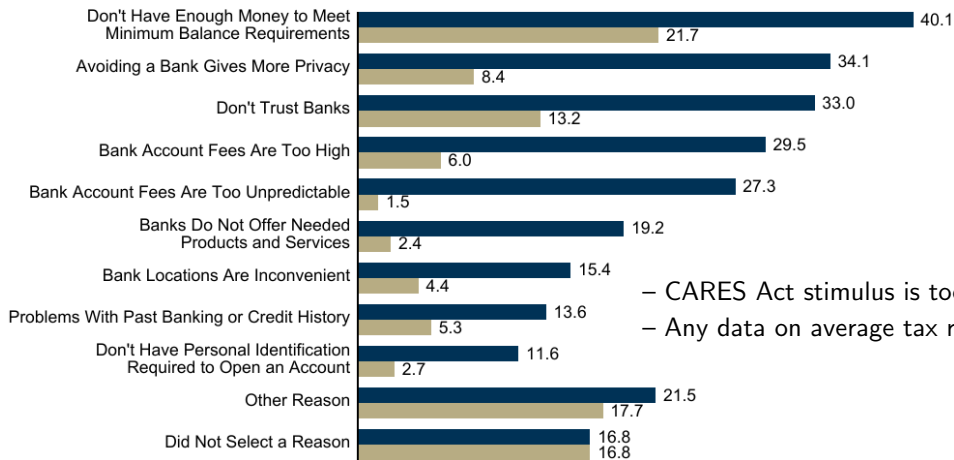
Mechanism: Survey Responses

Figure ES.3 Unbanked Households' Reasons for Not Having a Bank Account, 2021 (Percent)



Mechanism: Survey Responses

Figure ES.3 Unbanked Households' Reasons for Not Having a Bank Account, 2021 (Percent)



– CARES Act stimulus is too late.

– Any data on average tax refund by CBG?

Incidentals?

- Please add significance stars to the tables.
- Other Races?
 - Higher CBG-level Asian share is associated with more visitors AND lower block demand (coefficients are much larger than for Black share).
- Look at other interesting dimensions of the open census data (education, commuting time, etc).
- Are there Block FEs in Table 1 and Table 2 (I was confused by the title vs. the description).
- Assuming non-included branches all have same FE (at various points in the distribution) 😞
 - Alternative: use FE of nearest branch and assume similar quality. This isn't a great assumption either, but it's better than assuming they're all the same, I think. 😞
- Robustness: for just the main specification: do it with drive time and calculating all Branch and CBG pairs (feasible?).
- Cut the postal banking extension and broadband discussion.

Conclusion

- This is a really cool paper that goes a long way in establishing bank access disparities along dimensions of race and income.
- A little tightening on some of the methods
- The real grand prize is establishing channels through which these equilibrium outcomes materialize, but this paper (importantly) establishes the set of stylized facts that set the stage.