Financial Inclusion Across the United States*

Motohiro Yogo[†] Andrew Whitten[‡] Natalie Cox[§]

December 8, 2022

Abstract

We study retirement and bank account participation for the universe of U.S. households with a member aged 50 to 59 in the administrative tax data. In the lowest income quintile in 2019, 21 and 70 percent of households had respectively retirement and bank accounts. For the same group, 38 percent of households had access to an employer retirement plan. Geographic variation in financial participation relates to income rather than racial composition. By instrumental variables, we estimate the causal effect of access to an employer retirement plan. Universal access could increase retirement account participation by 10 percentage points in the lowest income quintile over ten years.

JEL: D14, G51

Keywords: Financial participation, Household finance, Inequality, Racial disparities, Tax policy

*For comments and discussions, we thank Edith Brashares, Alice Henriques Volz, Neviana Petkova, and John Sabelhaus. We also thank seminar participants at Dartmouth College, Federal Reserve Bank of Chicago, Federal Reserve Bank of New York, Harvard University, National University of Singapore, Peking University, Princeton University, Rutgers University, University of Michigan, U.S. Department of the Treasury, Wharton, the 2021 Office of Tax Analysis Research Conference, the 2021 NBER Conference on Innovative Data in Household Finance: Opportunities and Challenges, the 2022 Q-Group Spring Seminar, the 2022 NBER Public Economics Program Meeting, the 2022 UCLA Fink Center Conference on Financial Markets, and the 2022 Chicago Household Finance Conference. This research was conducted while Whitten was an employee of the U.S. Department of the Treasury. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily reflect the views or the official positions of the U.S. Department of the Treasury or the National Bureau of Economic Research. Any confidential taxpayer data used in this research were kept in a secured Treasury or IRS data repository, and all results have been reviewed to ensure that no confidential information is disclosed.

[†]Princeton University and NBER (myogo@princeton.edu)

[‡]U.S. Department of the Treasury (andrew.whitten@treasury.gov)

[§]Princeton University (nmcox@princeton.edu)

An inclusive society should strive for financial participation of all households, regardless of income or race. Based on survey evidence, we know that retirement and bank account participation are much lower for low-income and nonwhite households. However, survey evidence could paint an incomplete or inaccurate picture of financial participation because of small samples, a limited panel dimension, and measurement error. Big administrative data without these limitations could help us understand and hopefully improve financial participation for low-income and nonwhite households. Toward that effort, we study retirement and bank account participation for the universe of U.S. households with a member aged 50 to 59 in the 2015 to 2019 administrative tax data. These data contain virtually all tax returns and information returns and has the same population count as the census.

We define retirement accounts comprehensively to include employer retirement plans (both defined benefit and defined contribution plans) and individual retirement arrangements (IRAs). We confirm that the unconditional participation rates for retirement and bank accounts match between the administrative tax data and the Survey of Consumer Finances (SCF). This basic fact suggests that our sample and measurement assumptions mimic the survey data. We also confirm low financial participation for low-income households. In the lowest income quintile in 2019, 21 percent of households had retirement accounts, and 70 percent of households had bank accounts. For the same group, 38 percent of households had access to an employer retirement plan. The heterogeneity in financial participation conditional on income implies that low participation is not a simple matter of not having enough income to save. Therefore, we study geography and access to an employer retirement plan as potential determinants of financial participation.

The large sample allows us to tabulate financial participation at the level of ZIP Code Tabulation Areas (ZCTAs). We study whether the geographic variation in financial participation relates to average income, racial composition, or access to financial services. Retirement account participation is negatively correlated with the Hispanic and black population shares, but these correlations significantly weaken conditional on income. Bank account participation is also negatively correlated with the Hispanic and black population shares, but these correlations disappear conditional on income. Bank account participation is not correlated with bank branch density at the ZCTA level. However, we emphasize that our findings do not rule out the importance of supply factors such as banking fees (Dlugosz et al. 2021) and a more local level of spatial discrimination Sakong and Zentefis (2022). Overall, income is the primary geographic determinant of financial participation. In fact, the explanatory power of income is so high that geography appears to play little role in financial participation conditional on income.

The panel dimension allows us to estimate the causal effect of access to an employer retirement plan on retirement account participation. We start with the sample of households who do not have access to an employer retirement plan in 2010. The identifying assumption is that workers do not sort into employers in 2010 based on expectations of whether their employer would eventually offer a retirement plan. Some workers subsequently gain access if their employer introduces a new retirement plan or they switch to an employer that offers a retirement plan. The intent-to-treat instrument is the counterfactual access to an employer retirement plan if the worker had remained with the same employer since 2010. This instrument addresses the endogeneity problem that arises from workers switching employers to gain access to a retirement plan. On the extensive margin, access to an employer retirement plan increases retirement account participation by 17 percentage points in the lowest income quintile. On the intensive margin, each additional year of access increases retirement account participation by 3 percentage points.

These estimates are consistent with the evidence that a nudge (i.e., easy access to a retirement account through an employer) could have a large impact on retirement account participation (Madrian and Shea 2001; Chetty et al. 2014). Moreover, they suggest that universal access to an employer retirement plan is a policy intervention that could boost retirement account participation for low- and middle-income households. Starting with Oregon in 2017, ten states now have mandates requiring most employers to enroll all workers in a state-sponsored retirement savings program if they do not already offer a retirement plan. Because these mandates are relatively recent, we cannot yet evaluate their long-run impact on retirement account participation. However, we could estimate the counterfactual retirement account participation in 2019 if universal access to an employer retirement plan had already been in place in 2010. Over ten years, retirement account participation would increase by 10 percentage points in the lowest income quintile and 8 percentage points in the second income quintile.

Policymakers encourage retirement savings through a variety of tax incentives for both employers and workers. Eligible employers can claim tax credits for the cost of setting up retirement plans. Workers can deduct retirement contributions from taxable income and earn tax-deferred returns. Tax incentives affect even low-income workers, who face low marginal tax rates and may not pay capital gains taxes, because they could claim a Saver's Credit of up to 50 percent of retirement contributions. However, Ramnath (2013) finds that the Saver's Credit has a limited causal effect on retirement contributions. Based on our findings, tax incentives for employers to offer retirement plans may be more effective than those for workers to save in retirement accounts.

The administrative tax data allow us to measure the extensive margin of whether a household has a retirement or bank account but not the intensive margin of the account balance. The extensive margin is important from the perspective of life-cycle saving and could be a welfare metric under some assumptions. By not participating in retirement accounts, households are forgoing valuable tax savings. In particular, the Saver's Credit is like a matching contribution of 50 percent (up to $2,000 for joint filers) by the federal government. In the presence of behavioral biases, participation in a retirement account with automatic contributions or a bank account with electronic deposit of wages could get households in the habit of saving (Mullainathan and Shafir 2009). Consistent with this hypothesis, Célerier and Matray (2019) find that bank account participation increases wealth accumulation and durable good purchases. Finally, households without retirement, bank, or brokerage accounts do not own any risky financial assets. With smooth preferences and no fixed costs, portfolio theory predicts that at least some equity exposure is optimal.

Our study relates to the literature on wealth inequality. Wealth inequality is worse than income inequality in the United States (Bricker et al. 2020). According to the 2019 SCF, the income share was 15 percent for the bottom half of households and 19 percent for the top one percent. In comparison, the wealth share was 2 percent for the bottom half of households and 33 percent for the top one percent. Therefore, the bottom half of households earns 19 percent of income but owns only 3 percent of wealth in the subpopulation that excludes the top one percent. Because retirement and bank accounts are the most important means of accumulating financial wealth, low participation at the bottom of the income distribution is important for a better understanding wealth inequality.

The remainder of the paper is organized as follows. Section I describes how we construct the sample of households and measures of income and financial participation from the administrative tax data. Section II summarizes financial participation by income. In Section III, we study the geographic determinants of financial participation. In Section IV, we estimate the causal effect of access to an employer retirement plan on retirement account participation. We also estimate the potential impact of universal access to employer retirement plans. Section V concludes.

I. Data Construction

We describe how we construct the sample of households and measures of income and financial participation from the administrative tax data.

A. Administrative Tax Data

We use the administrative tax data of the Internal Revenue Service, which contain tax returns (Form 1040) and information returns for tax years 1999 to 2019. The data are digitized since 1999 and are not fully available before then. The relevant information returns are Forms W-2 (wage and tax statement), 1099-INT (interest income), 1099-DIV (dividends and distributions), 1099-R (distributions from pensions, annuities, retirement or profit-sharing plans, IRAs, insurance contracts, etc.), 1099-MISC (miscellaneous income), 1099-G (certain government payments), SSA-1099 (Social Security benefit statement), 1099-B (proceeds from broker and barter exchange transactions), 5498 (IRA contribution information), and 1095 (health insurance marketplace statement, health coverage, or employer-provided health insurance offer and coverage).

Sample

We sample all individuals aged 50 to 59 from 2015 to 2019, who have either a tax return or an information return and have a ZIP Code within the U.S. states or Washington, DC. We focus on ages 50 to 59 because it is the most relevant part of the life cycle for retirement saving. As we show in Appendix A, the sample includes over 43 million individuals, which is essentially the same population count as the resident population aged 50 to 59 in the census. The sample starts in 2015 to coincide with the start of Form 1095, which is necessary to achieve the same population count as the census (Lurie and Pearce 2019).

We use a crosswalk file to map the ZIP Codes to ZIP Code Tabulation Areas (ZCTAs).[1] The Census Bureau constructed ZCTAs by assigning census blocks to approximately 32,000 geographic areas. In most cases, the ZCTA assigned to a census block is the same as its ZIP Code. However, they could be different if a census block contains multiple ZIP Codes. For some of our specifications, we use commuting zone fixed effects. The Census Bureau defines 709 commuting zones by aggregating counties according to local labor markets.

For each sampled individual, we also obtain her tax data for the previous nine years. So for a sampled individual in 2015, we obtain her tax data for 2006 to 2015. We also sample the spouses of sampled individuals (regardless of age), defined as a current joint filer on Form 1040, a previous joint filer in the last ten years who currently has the same household identifier (i.e., the same address), or a previous joint filer in the last ten years who is currently on the same Form 1095.[2] The latter two criteria ensure that we do not break up households that stop filing taxes. We intentionally do not link individuals who have the same household identifier if they have not filed taxes together. Our goal is to measure joint access to financial accounts, and it is unclear to what extent financial accounts are shared among non-spousal

[1]When the ZIP Code is not available on Form 1040, we use the ZIP Code from information returns prioritized in the order listed above. If the ZIP Code is still not available, we use the most commonly reported ZIP Code on all other information returns.

[2]Larrimore et al. (2021) constructed the household identifiers based on a textual analysis of the addresses on tax returns and information returns.

household members (e.g., parents or children living at the same address).

Income

We construct pre-tax household income following Larrimore et al. (2021). For tax filers who have a Form 1040, we start with total income (line 7b on the 2019 Form 1040), which includes wages and salaries, pass-through business income (including self-employment income), taxable interest, dividends, realized capital gains, taxable private retirement income, taxable Social Security benefits, rents, royalties, unemployment compensation, and alimony. We adjust total income by adding tax-exempt interest, subtracting realized capital gains, replacing taxable private retirement income with gross private retirement income (excluding rollovers) on Form 1099-R, and replacing taxable Social Security benefits with total Social Security benefits on Form SSA-1099 (including disability insurance). Finally, we truncate pre-tax household income at zero to limit the impact of business losses.

For nonfilers who do not have a Form 1040, we again follow Larrimore et al. (2021). Pre-tax individual income is the sum of wages and salaries on Form W-2, interest income on Form 1099-INT, dividends on Form 1099-DIV, gross private retirement income (excluding rollovers) on Form 1099-R, total Social Security benefits on Form SSA-1099, unemployment benefits on From 1099-G, and 30 percent of income on Form 1099-MISC (assuming 70 percent for offsetting expenses). For nonfilers who form households through a common household identifer or Form 1095, pre-tax household income is the sum of the pre-tax individual incomes.

We adjust income to 2016 dollars using the consumer price index for all urban consumers. We then define usual household income as the moving average of inflation-adjusted household income over a five-year history (i.e., the current year and the previous four). Usual household income is meant to capture permanent income that smoothes out transitory shocks.

We construct five income groups based on the distribution of usual household income over the entire sample. We refer to the income quintiles as the lowest quintile (0–20 percentiles), the second quintile (20–40 percentiles), the third quintile (40–60 percentiles), the fourth quintile (60–80 percentiles), and the highest quintile (80–100 percentiles).

Financial Participation

We measure bank account participation based on electronic funds transfer for payment of taxes or receipt of refunds on Form 1040. According to the instructions for Form 1040, about 80 percent of tax filers who receive refunds do so by direct deposit. Moreover, the name on the tax filing must match the name on the bank account, which rules out tax filers receiving refunds in a bank account that they do not own. We also measure bank account participation based on taxable (box 1) or tax-exempt (box 8) interest on Form 1099-INT.[3]

The resulting variable for bank account participation could have gaps in the panel dimension if an individual does not file taxes or receive a Form 1099-INT in a given year. Therefore, we use a nine-year lookback to improve our measure of bank account participation. For example, we measure bank account participation in 2015 if the criteria for having a bank account are satisfied in any year between 2006 and 2015. Thus, the definition of bank account participation in 2015 is having an account in the last ten years, even if that account is closed as of 2015. From a practical perspective, we do not expect individuals to switch from participant to non-participant at high frequency. From an economic perspective, an individual who has ever had a bank account is different from one who has never had a bank account.

We define retirement account participation comprehensively to include employer retirement plans and IRAs. We measure participation in an employer retirement plan if retirement plan (box 13) is checked on Form W-2. This covers all employer retirement plans including defined benefit and defined contribution plans. We measure participation in an IRA based on the presence of Form 5498, which is annually filed with the Internal Revenue Service even when no contributions are made. We also measure retirement account participation based on a retirement distribution on Form 1099-R. For some of our analyses, we use the distribution codes on Form 1099-R to distinguish an employer retirement plan versus an IRA.

The resulting variable for retirement account participation could have gaps in the panel dimension if an individual does not receive a Form W-2 or a Form 1099-R in a given year. Therefore, we use a nine-year lookback to improve our measure of retirement account participation. We observe that a retirement account is closed when total distribution (box 2b) is checked on Form 1099-R. Thus, the definition of retirement account participation in 2015 is having an account in the last ten years that is still open as of 2015.

We measure access to an employer retirement plan if retirement plan (box 13) is checked on *any* Form W-2 issued by their employer in a given year. We search through all Forms W-2 (not just sampled individuals) to construct this variable. Furthermore, we require that the income received from the employer is greater than the federal minimum wage times 1,000 hours to infer that the individual is eligible for retirement benefits. Using a nine-year lookback, we define access to an employer retirement plan as access through any employer in the last ten years.

[3] Form 1099-INT has incomplete coverage because it is required only for accounts with at least $10 of annual interest. Thus, we measure bank account participation primarily through electronic funds transfer as part of a tax filing.

We define indicator variables for married households by taking the maximum over the two indicator variables for bank account participation, retirement account participation, and access to an employer retirement plan. That is, a household has a bank account if either spouse has a bank account.

During the initial research design, we considered other measures of financial participation. We could define financial participation more broadly to include mutual funds and brokerage accounts, based on Forms 1099-DIV and 1099-R. However, we have verified that virtually all households that have these accounts already have a bank account. Participation in stocks and equity mutual funds would have been interesting from the perspective that all households should participate under smooth preferences and no fixed costs. However, the administrative tax data do not contain any information about stocks and equity mutual funds in retirement accounts. We could measure mortgage participation based on Form 1098. However, the administrative tax data do not tell us whether households do not have a mortgage because they do not need one or have been denied. In a related paper, Lurie and Pearce (2019) use the administrative tax data to study health insurance coverage.

B. Survey of Consumer Finances

To benchmark our findings to survey evidence, we use the 2016 and 2019 SCF. We restrict the sample to households with a respondent aged 50 to 59 in these years. The data contain the respondent's race, which is grouped into white, Hispanic, black, and other nonwhite (including Asian). Usual income in the SCF is a self-reported measure of permanent income that smoothes out transitory shocks. It is broader than our measure of usual income in the administrative tax data by including realized capital gains, food stamps, and other government support. We construct retirement and bank account participation, following the same definitions that we use for the administrative tax data.

II. Facts About Financial Participation

We summarize financial participation by income in the administrative tax data. We confirm that the unconditional participation rates for retirement and bank accounts match between the administrative tax data and the SCF. We also summarize the determinants of financial participation in the SCF, which help us interpret the geographic analysis in Section III.

A. *Financial Participation by Income*

Table 1 reports retirement account participation for households with a member aged 50 to 59 in the administrative tax data and the SCF. The overall retirement account participation matches between the administrative tax data and the SCF. In 2016, 70 percent of households with a member aged 50 to 59 had retirement accounts in the administrative tax data, which is close to 72 percent in the SCF. In 2019, 69 percent of households with a member aged 50 to 59 had retirement accounts in both the administrative tax data and the SCF.

Retirement account participation in the administrative tax data has a slightly steeper income gradient than that in the SCF. Thus, retirement account participation for low-income households in the administrative tax data is lower than that in the SCF. In the lowest income quintile in 2019, 21 percent of households had retirement accounts in the administrative tax data, compared with 24 percent in the SCF. In the second income quintile in 2019, 54 percent of households had retirement accounts in the administrative tax data, compared with 63 percent in the SCF.

Table 4 reports bank account participation for households with a member aged 50 to 59 in the administrative tax data and the SCF. The overall bank account participation matches between the administrative tax data and the SCF. In 2016, 93 percent of households with a member aged 50 to 59 had bank accounts in both the administrative tax data and the SCF. In 2019, 92 percent of households with a member aged 50 to 59 had bank accounts in the administrative tax data, which is close to 95 percent in the SCF.

Bank account participation in the administrative tax data has a steeper income gradient than that in the SCF. Thus, bank account participation for low-income households in the administrative tax data is lower than that in the SCF. In 2016, 71 percent of households in the lowest income quintile had bank accounts in the administrative tax data, compared with 76 percent in the SCF. In the lowest income quintile in 2019, 70 percent of households had bank accounts in the administrative tax data, compared with 80 percent in the SCF.

The administrative tax data has the same population count as the census, and the SCF is based on a random sample of the census population. Thus, the fact that the unconditional participation rates match between the administrative tax data and the SCF suggests that our measurement assumptions mimic the survey data. Several factors could explain why financial participation in the administrative tax data has a steeper income gradient than that in the SCF. As we discussed in Section I, the SCF uses a broader definition of income that includes realized capital gains, food stamps, and other government support. Because the SCF is a survey, both financial participation and income are subject to measurement error or imputation error when households misreport or refuse to answer survey questions. Any measurement error in income attenuates the true relation between financial participation and income, which could explain the flatter gradient in the SCF.

B. *Additional Facts About Retirement Accounts*

Table 2 breaks down retirement account participation into employer retirement plans versus IRAs for households with a member aged 50 to 59 in 2019. In the lowest income quintile, 11 percent of households have only an employer retirement plan, 5 percent have only an IRA, and 4 percent have both. Thus, low-income households have retirement accounts primarily through their employer. Higher-income households are more likely to have both an employer retirement plan and an IRA. In the highest income quintile, 24 percent of households have only an employer retirement plan, 5 percent have only an IRA, and 68 percent have both.[4]

Table 3 reports access to an employer retirement plan for households with a member aged 50 to 59 in the administrative tax data. In the lowest income quintile in 2019, only 38 percent of households had access to an employer retirement plan. Access to an employer retirement plan increases to 80 percent of households in the second income quintile and 93 percent of households in the third income quintile. Although our time series is short, access to an employer retirement plan appears to decline in the lowest income quintile from 42 percent of households in 2015 to 38 percent of households in 2019.

C. *Determinants of Financial Participation in the SCF*

To help us interpret the geographical analysis in Section III, we summarize the determinants of financial participation in the SCF (Hogarth et al. 2005). Table 5 reports regressions of retirement or bank account participation on household characteristics for households with a member aged 50 to 59 in the 2016 and 2019 SCF. The regression is a linear probability model, where the dependent variable is an indicator variable that is one for participants. We include year fixed effects to estimate the cross-sectional relation between financial participation and household characteristics.

In column 1 of Table 5, race is a significant determinant of retirement account participation. Households with Hispanic and black respondents are respectively 34 and 21 percentage points less likely to have a retirement account than households with white respondents. Households with other nonwhite respondents are 6 percentage points less likely to have a retirement account than households with white respondents. Controlling for log income in column 2, the coefficients for race decrease in magnitude but remain statistically significant.

[4] Some IRAs could have been funded by rollovers from employer retirement plans. Therefore, Table 2 could understate the importance of employer retirement plans if some households have IRAs that were entirely funded by rollovers from employer retirement plans.

Households with Hispanic and black respondents are respectively 21 and 8 percentage points less likely to have a retirement account than households with white respondents, conditional on income. Households with other nonwhite respondents are 5 percentage points less likely to have a retirement account than households with white respondents, conditional on income.

Columns 3 and 4 of Table 5 report analogous regressions for bank account participation. In column 3, race is a significant determinant of bank account participation. Households with Hispanic and black respondents are respectively 11 and 12 percentage points less likely to have a bank account than households with white respondents. Households with other nonwhite respondents are 5 percentage points less likely to have a bank account than households with white respondents. Households with other nonwhite respondents are 5 percentage points less likely to have a bank account than households with white respondents. Controlling for log income in column 4, the coefficients for race decrease in magnitude but remain statistically significant. Households with Hispanic and black respondents are respectively 6 and 8 percentage points less likely to have a bank account than households with white respondents, conditional on income. Households with other nonwhite respondents are 5 percentage points less likely to have a bank account than households with white respondents, conditional on income. Hayashi and Minhas (2018) also find that Hispanic and black households are less likely to have a bank account conditional on income in the 2015 FDIC Survey of Unbanked and Underbanked Households.

The fact that race is an important determinant of financial participation, even after controlling for income, raises two hypotheses. The first hypothesis is racial discrimination. This includes both current discrimination and the lingering effects of historical discrimination, such as a mistrust of financial institutions that could persist as a cultural norm. A second hypothesis is spatial discrimination based on the location of bank branches. The cost of accessing banking services may be high for nonwhite household if there are no bank branches nearby. These hypotheses motivate the geographic analysis in the next section.

III. Geography of Financial Participation

We first construct maps to highlight the variation in financial participation across ZCTAs. We then examine whether the variation in financial participation across ZCTAs relates to average income, racial composition, or access to financial services. Because we do not observe race in the administrative tax data, we must be careful in interpreting the relation between financial participation and racial composition at the ZCTA level. We conclude this section with a discussion of how our findings for the geographic determinants of financial participation relate to Table 5 for the SCF.

A. Maps of Financial Participation

Figure 1 is a map of retirement account participation in the lowest income quintile in 2019. We focus on the lowest income quintile to highlight the variation in financial participation across ZCTAs. The colors range from yellow (40–100 percent participation) to red (0–10 percent participation). The shade depends on the population aged 50 to 59, where a darker shade represents a more populous ZCTA. For example, a dark shade of red indicates a populous ZCTA with low bank account participation. Figure 2 is a similar map of bank account participation in the lowest income quintile in 2019. The colors range from yellow (90–100 percent participation) to red (0–60 percent participation).

In Figures 2 and 1, the geographic variation in financial participation is not a simple matter of the north versus the south or the coasts versus the inland. Within geographic areas smaller than states, red areas of low participation are mixed with yellow areas of high participation. We examine whether the variation in financial participation across ZCTAs relates to average income, racial composition, or access to financial services.

B. Geographic Determinants of Financial Participation

Retirement Accounts

We estimate panel regressions of ZCTA-level retirement account participation on geographic characteristics with year fixed effects over the sample period of 2015 to 2019. We weight the observations by the census-derived household count in each cell (see Appendix A). In column 1 of Table 6, race is a significant determinant of retirement account participation. A percentage point increase in the Hispanic and black population shares relates respectively to a 37 and 29 basis point decrease in retirement account participation. A percentage point increase in the other nonwhite population share relates to a 29 basis point decrease in retirement account participation.

In column 2 of Table 6, we control for average log income, which is standardized. A statistically significant coefficient of 0.10 means that a standard deviation increase in income relates to a 10 percentage point increase in retirement account participation. Income has significant explanatory power with the R^2 increasing from 0.46 in column 1 to 0.86 in column 2. Moreover, the coefficients for the Hispanic, black, and other nonwhite population shares significantly decrease in magnitude. The coefficient for the Hispanic population share decreases in magnitude from -0.37 to -0.18, and the coefficient for the black population share decreases in magnitude from -0.29 to -0.07. Thus, race is a much less important determinant of retirement account participation conditional on income.

In column 3 of Table 6, we control for commuting zone fixed effects. Relative to column 2, the coefficients are virtually unchanged, and the R^2 increases slightly from 0.86 to 0.90. Thus, the correlation between retirement account participation and income is about local geographic variation within commuting zones.

Bank Accounts

Analogous to Table 6, Table 7 reports panel regressions of ZCTA-level bank account participation on geographic characteristics. In column 1 of Table 7, race is a significant determinant of bank account participation. A percentage point increase in the Hispanic and black population shares relates respectively to a 6 and 9 basis point decrease in bank account participation. A percentage point increase in the other nonwhite population share relates to a 14 basis point decrease in bank account participation. In contrast, a percentage point increase in the Asian population share relates to a 3 basis point increase in bank account participation.

In column 2 of Table 7, we control for bank branch density at the ZCTA level (see Appendix B), which is standardized. Bank branch density has no effect on bank account participation. Moreover, the coefficients for race are virtually unchanged, implying that the correlation between bank account participation and race is not a simple matter of the location of bank branches. However, we emphasize that our findings do not rule out the importance of supply factors. Using more granular data at the census block level, Sakong and Zentefis (2022) find that bank branch access does not relate to median income but relates to the black population share. Thus, our data at the ZCTA level could hide a more local level of spatial discrimination. Dlugosz et al. (2021) suggest that banking fees are an important supply factor that could explain bank account participation. They find that a cap on overdraft fees constrains the supply of overdraft credit and deposit accounts through higher minimum deposits.

In column 3 of Table 7, we control for average log income, which is standardized. A statistically significant coefficient of 0.04 means a standard deviation increase in income relates to a 4 percentage point increase in bank account participation. Income has significant explanatory power with the R^2 increasing from 0.19 in column 1 to 0.82 in column 2. Moreover, the coefficients for race are all close to zero. Thus, race is not a significant determinant of bank account participation conditional on income.

In column 4 of Table 7, we control for commuting zone fixed effects. Relative to column 3, the coefficients are virtually unchanged, and the R^2 increases slightly from 0.82 to 0.87. Thus, the correlation between bank account participation and income is about local geographic variation within commuting zones.

C. Interpreting Income

At face value, income is the primary determinant of financial participation because households without enough income are unable to save. In addition, fees on accounts with low balances may disincentivize low-income households from opening and keeping a bank account. Low-income households may not have sufficient tax incentives to open a retirement account, despite a Saver's Credit of up to 50 percent of retirement contributions. Other than these direct effects of income, financial participation could be correlated with income for other reasons.

Bumcrot et al. (2013) find that financial literacy is correlated with poverty rates across states. Thus, financial participation could be correlated with income partly through financial literacy. Tables 7 and 6 show significant correlation between financial participation and income within commuting zones, which are smaller geographic areas than states. Unfortunately, the sample size in the 2009 National Financial Capability Study, on which Bumcrot et al. (2013) is based, is too small to tabulate at the ZCTA level. Future research could investigate whether financial literacy partly explains the correlation between financial participation and income by collecting a large sample of financial literacy measures at the ZCTA level.

Mullainathan and Shafir (2009) hypothesize that low-income households are less likely to have bank accounts because of institutions that shape behavior. For example, low-income households may not have incentives to open a bank account if their employers do not use electronic deposits. If their income is too low to file taxes, they have no need for an electronic funds transfer to pay taxes or receive refunds. Under this hypothesis, a nudge such as an electronic deposit of wages as the default option could boost bank account participation.

Peer effects could reinforce the correlation between financial participation and income at the ZCTA level (Duflo and Saez 2002, 2003). In ZCTAs with high average income, lower-income households may be more likely to socialize with higher-income households and learn about retirement and bank accounts. Future research could investigate whether peer effects partly explain the correlation between financial participation and income by measuring social connections within ZCTAs (Bailey et al. 2018).

D. Comparison with the SCF

For retirement account participation, the coefficients for race in the aggregate tax data (i.e., column 1 of Table 6) are similar in magnitude to those in the SCF (i.e., column 3 of Table 5). Controlling for income, retirement account participation is weakly correlated with race in the aggregate tax data (i.e., column 2 of Table 6), but it remains strongly correlated with race in the SCF (i.e., column 4 of Table 5). For bank account participation, the coefficients for race in the aggregate tax data (i.e., column 1 of Table 7) are similar in magnitude to those in the SCF (i.e., column 1 of Table 5). Controlling for income, bank account participation is no longer correlated with race in the aggregate tax data (i.e., column 3 of Table 7), but it remains correlated with race in the SCF (i.e., column 2 of Table 5).

We discuss two hypotheses for why income significantly reduces the correlation between financial participation and race in the aggregate tax data but not in the SCF. The first hypothesis is group effects that are not separately identified from individual effects in the aggregate tax data. The second hypothesis is mismeasurement of income in the SCF. These hypotheses are not mutually exclusive, so both may matter in reality.

To illustrate group effects, we posit an econometric model for bank account participation. Let $y_{i,n}$ be an indicator variable that is one if household i in ZCTA n has a bank account. Let $x_{i,n}$ be log income of household i in ZCTA n. Let $z_{i,n}$ be an indicator variable that is one if household i in ZCTA n is black. Let $Z_n = \sum_{i=1}^{I_n} z_{i,n}/I_n$ be the black population share in ZCTA n, where I_n is the household count in ZCTA n. Suppose that bank account participation is determined by

$$y_{i,n} = \alpha + \beta x_{i,n} + \gamma z_{i,n} + \Gamma Z_n + \omega_{i,n}, \tag{1}$$

where $\omega_{i,n}$ is an error term. The coefficient β represents the individual effect of income on participation. The coefficients γ and Γ represent respectively the individual and group effects of race on participation. Based on Table 5 for the SCF, we assume that $\beta > 0$ and $\gamma < 0$. We cannot estimate Γ because the SCF does not have geographic identifiers.

Aggregating equation (1) over ZCTA n,

$$Y_n = \alpha + \beta X_n + (\gamma + \Gamma) Z_n + \Omega_n, \tag{2}$$

where $Y_n = \sum_{i=1}^{I_n} y_{i,n}/I_n$, $X_n = \sum_{i=1}^{I_n} x_{i,n}/I_n$, and $\Omega_n = \sum_{i=1}^{I_n} \omega_{i,n}/I_n$. Bank account participation at the ZCTA level depends on the black population share with a coefficient of $\gamma + \Gamma$ that represents the sum of individual and group effects. Based on Table 7 for the aggregate tax data, we know that $\beta > 0$ and $\gamma + \Gamma = 0$. Thus, $\Gamma = -\gamma > 0$ could explain the difference between the aggregate tax data (i.e., column 3 of Table 7) and the SCF (i.e., column 2 of Table 5). However, this hypothesis requires an unlikely knife-edge scenario that the group effect exactly offsets the individual effect. Moreover, $\Gamma > 0$ means that white households are *more* likely to participate in black neighborhoods. Opposite-signed individual and group effects are possible, but we would have expected peer effects to be conforming rather than polarizing.

An alternative hypothesis is that financial participation depends on income but not race (i.e., $\gamma = \Gamma = 0$), and income is mismeasured in the SCF. If race correlates with true income, the coefficients for Hispanic and black in Table 5 for the SCF would be downward biased (i.e., more negative). This bias is significantly reduced in Tables 7 and 6 for the aggregate tax data because income is better measured. Under this hypothesis, the unconditional correlation between financial participation and race is not directly about race but about economic factors that are associated with income.

IV. Access to an Employer Retirement Plan

As reported in Table 2, most households have retirement accounts through employer retirement plans rather than only through IRAs. Opening an IRA requires more effort and financial literacy than enrolling in an employer retirement plan. Therefore, access to an employer retirement plan could be a primary determinant of retirement account participation. Moreover, universal access to an employer retirement plan could be an effective policy intervention that boosts retirement account participation for low- and middle-income households, who have relatively low access to employer retirement plans according to Table 3.

A. Identifying Assumptions

Retirement account participation and access to an employer retirement plan could be jointly endogenous. Workers who care about retirement savings may choose to work for an employer with a retirement plan, leading to a positive selection bias. Conversely, some workers may already have retirement security through an IRA, their spouse's retirement savings, or Social Security. Workers with retirement security may choose to work for an employer without a retirement plan (e.g., a small employer or self-employment), leading to a negative selection bias.

We use the panel dimension to estimate the causal effect of access to an employer retirement plan on retirement account participation. We start with a sample of individuals aged 50 to 59 in 2019, who did not have access to an employer retirement plan in 2010. The identifying assumption is that workers do not sort into employers in 2010 based on expectations of whether their employer would eventually offer a retirement plan. This assumption is plausible insofar as the employment relations started long before 2010 for most workers. Whether their employer would eventually offer a retirement plan would have been difficult to forecast or not relevant to their decision to join the firm.

Some workers subsequently gain access if their employer introduces a new retirement plan.

For example, Panel A of Figure 3 illustrates the case of a worker who stays with employer A from 2010 to 2019 and gains access to an employer retirement plan in 2016. Alternatively, some workers switch to an employer that offers a retirement plan. For example, Panel B illustrates the case of a worker who switches from employer A to employer B and gains access to an employer retirement plan in 2013. The intent-to-treat instrument is the counterfactual access to an employer retirement plan if the worker had remained with the same employer since 2010. This instrument addresses the endogeneity problem that arises from workers switching employers to gain access to a retirement plan. In Figure 3, the actual access to an employer retirement plan is in black, and the counterfactual access is in red.

We estimate the instrumental variables regression separately by income quintile to allow for heterogeneous treatment effects. The sample size is larger for lower income quintiles because low-income households are less likely to have had access to an employer retirement plan in 2010. The dependent variable is retirement account participation. The first endogenous regressor is an indicator variable for access to an employer retirement plan by either spouse since 2010. The second endogenous regressor is the number of years that either spouse had access to an employer retirement plan since 2010. This specification captures a potential nonlinearity from the first year of access having a larger impact than each subsequent year of access to an employer retirement plan.

We construct two instruments corresponding to the two endogenous regressors, based on the counterfactual access to an employer retirement plan. For example, in Panel B of Figure 3, the endogenous regressors would be 1 for the indicator variable and 7 for the number of years. The corresponding instruments would be 1 for the indicator variable and 4 for the number of years.

B. *Causal Effects of Access to an Employer Retirement Plan*

Table 8 reports the instrumental variables regression of retirement account participation on access to an employer retirement plan. In the lowest income quintile, the test statistic of 24,845 is greater than the critical value of 7.03 for rejecting the null of weak instruments at the 5 percent significance level (Stock and Yogo 2005, table 5.2). Similarly, we reject the null of weak instruments for the other income quintiles.

The constant in the model is the baseline participation rate for households with no access to an employer retirement plan. The baseline participation rate increases from 16 percent in the lowest income quintile to 69 percent in the highest income quintile. Higher-income households are more likely to have retirement accounts through IRAs or existing employer retirement plans before 2010.

On the extensive margin, access to an employer retirement plan increases retirement account participation by 17 percentage points in the lowest income quintile. On the intensive margin, each additional year of access increases retirement account participation by 3 percentage points. The treatment effects are larger for higher income quintiles, implying higher takeup rates of employer retirement plans. In the third income quintile, access to an employer retirement plan increases retirement account participation by 24 percentage points.

C. State-Sponsored Retirement Savings Programs

The causal effects in Table 8 suggest that universal access to an employer retirement plan is a policy intervention that could boost retirement account participation. Table 9 lists ten states that have mandates requiring most employers to enroll all workers in a state-sponsored retirement savings program if they do not already offer a retirement plan.[5] The mandates apply to all employers with a minimum number of workers and minimum years in business. The state-sponsored retirement savings programs are legally IRAs and subject to the IRA contribution limits. The default contribution rate is 3 or 5 percent (depending on the state), but workers can adjust the contribution rate or entirely opt out. Chalmers et al. (2021) find a 34 percent participation rate for OregonSaves, which implies that a majority of workers actually opt out. The mandates do not have any costs to the employers, except for the administrative costs of compliance.

Because these mandates are relatively recent, we cannot yet evaluate their long-run impact on retirement account participation. Moreover, these mandates have a limited scope to a subset of employers in the ten participating states. Based on the instrumental variables regression model in Tables 8, we estimate the potential impact of universal access to an employer retirement plan. We estimate the counterfactual retirement account participation in 2019 if universal access to an employer retirement plan had already been in place in 2010. For each worker in each year, we define counterfactual access to an employer retirement plan as one if working (i.e., receiving a Form W-2) and zero otherwise. For each household, we then compute the indicator variable for access to an employer retirement plan by either spouse since 2010 and the number of years that either spouse had access to an employer retirement plan since 2010. Finally, we use the instrumental variables regression model to predict the counterfactual probability of participation for each household.

Table 10 reports the predicted change in retirement account participation by income quintile. Over ten years, retirement account participation would increase by 10 percentage points in the lowest income quintile and 8 percentage points in the second income quintile. In

[5]In September 2021, the House Ways and Means Committee released legislative text that would require retirement plans for all employers with at least five workers and in business for at least two years. It is too early to tell whether universal access to an employer retirement plan could become federal law.

the third to the highest income quintiles, the predicted change declines respectively to 4, 1, and 1 percentage points. Universal access has the largest impact on low- and middle-income households, who have relatively low access to an employer retirement plan. However, universal access does not achieve universal participation because not all households are working, and some households would not take up an employer retirement plan. Universal access has a much smaller impact on high-income households, many of whom already have access to an employer retirement plan.

V. Conclusion

We study retirement and bank account participation for the universe of U.S. households with a member aged 50 to 59 in the administrative tax data. In the lowest income quintile in 2019, 21 percent of households had retirement accounts, and 70 percent of households had bank accounts. The heterogeneity in financial participation conditional on income implies that low participation is not a simple matter of not having enough income to save. An important policy question is how to increase financial participation for low-income households.

In hope of improving financial participation for low-income households, we have constructed interactive maps of financial participation in the lowest income quintile. The interactive maps for retirement accounts and bank accounts are available from the authors' webpage. Users can search for specific locations or zoom in and out to visualize heterogeneity in financial participation across the United States. We hope that this tool is useful for researchers, policymakers, banks, and financial advisors to identify geographic areas with the greatest opportunity for improvement.

Access to an employer retirement plan is a primary determinant of retirement account participation. On the extensive margin, access to an employer retirement plan increases retirement account participation by 17 percentage points in the lowest income quintile. On the intensive margin, each additional year of access increases retirement account participation by 3 percentage points. Universal access to an employer retirement plan could be an effective policy intervention that boosts retirement account participation for low- and middle-income households. Over ten years, universal access to an employer retirement plan could increase retirement account participation by 10 percentage points in the lowest income quintile and 8 percentage points in the second income quintile.

References

Bailey, Michael, Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong, "Social Connectedness: Measurement, Determinants, and Effects," *Journal of Economic Perspectives*, 2018, *32* (3), 259–80.

Bricker, Jesse, Sarena Goodman, Kevin B. Moore, and Alice Henriques Volz, "Wealth and Income Concentration in the SCF: 1989–2019," 2020. FEDS Notes.

Bumcrot, Christopher, Judy Lin, and Annamaria Lusardi, "The Geography of Financial Literacy," *Numeracy*, 2013, *6* (2), 1–16.

Célerier, Claire and Adrien Matray, "Bank-Branch Supply, Financial Inclusion, and Wealth Accumulation," *Review of Financial Studies*, 2019, *32* (12), 4767–4809.

Chalmers, John, Olivia S. Mitchell, Jonathan Reuter, and Mingli Zhong, "Auto-Enrollment Retirement Plans for the People: Choices and Outcomes in OregonSaves," 2021. NBER Working Paper 28469.

Chetty, Raj, John N. Friedman, Søren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen, "Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark," *Quarterly Journal of Economics*, 2014, *129* (3), 1141–1219.

Dlugosz, Jennifer, Brian Melzer, and Donald P. Morgan, "Who Pays the Price? Overdraft Fee Ceilings and the Unbanked," 2021. Working paper, Board of Governors of the Federal Reserve System.

Duflo, Esther and Emmanuel Saez, "Participation and Investment Decisions in a Retirement Plan: The Influence of Colleagues' Choices," *Journal of Public Economics*, 2002, *85* (1), 121–148.

_ and _ , "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *Quarterly Journal of Economics*, 2003, *118* (3), 815–842.

Hayashi, Fumiko and Sabrina Minhas, "Who Are the Unbanked? Characteristics Beyond Income," *Economic Review: Federal Reserve Bank of Kansas City*, 2018, *2*, 55–70.

Hogarth, Jeanne M., Christoslav E. Anguelov, and Jinhook Lee, "Who Has a Bank Account? Exploring Changes Over Time, 1989–2001," *Journal of Family and Economic Issues*, 2005, *26* (1), 7–30.

Larrimore, Jeff, Jacob Mortenson, and David Splinter, "Household Incomes in Tax Data Using Addresses to Move from Tax-Unit to Household Income Distributions," *Journal of Human Resources*, 2021, *56* (2), 600–631.

Lurie, Ithai Z. and James Pearce, "Health Insurance Coverage from Administrative Tax Data," 2019. Office of Tax Analysis Working Paper 117.

Madrian, Brigitte C. and Dennis F. Shea, "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 2001, *116* (4), 1149–1187.

Mullainathan, Sendhil and Eldar Shafir, "Savings Policy and Decisionmaking in Low-Income Households," in Rebecca M. Blank and Michael S. Barr, eds., *Insufficient Funds: Savings, Assets, Credit, and Banking Among Low-Income Households*, The National Poverty Center Series on Poverty and Public Policy, New York: Russell Sage Foundation, 2009, chapter 5, pp. 121–146.

Ramnath, Shanthi, "Taxpayers' Responses to Tax-based Incentives for Retirement Savings: Evidence from the Saver's Credit Notch," *Journal of Public Economics*, 2013, *101*, 77–93.

Sakong, Jung and Alexander K. Zentefis, "Bank Access Across America," 2022. Working paper, Federal Reserve Bank of Chicago.

Stock, James H. and Motohiro Yogo, "Testing for Weak Instruments in Linear IV Regression," in Donald W. K. Andrews and James H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, Cambridge: Cambridge University Press, 2005, chapter 5, pp. 80–108.

Table 1
Retirement Account Participation by Income

Year	0–20	20–40	40–60	60–80	80–100	All
	Percentile of usual income					
Panel A. Administrative tax data						
2015	0.22	0.55	0.82	0.94	0.97	0.70
2016	0.22	0.55	0.82	0.94	0.97	0.70
2017	0.21	0.54	0.81	0.94	0.97	0.70
2018	0.21	0.54	0.81	0.93	0.97	0.69
2019	0.21	0.54	0.81	0.93	0.97	0.69
Panel B. SCF						
2016	0.33	0.62	0.83	0.90	0.95	0.72
2019	0.24	0.63	0.76	0.92	0.93	0.69

This table reports retirement account participation for households with a member aged 50 to 59. Retirement accounts include employer retirement plans (i.e., defined benefit and defined contribution plans) and IRAs.

Table 2
Breakdown of Retirement Account Participation

Households with	0–20	20–40	40–60	60–80	80–100	All
	Percentile of usual income					
Employer plan only	0.11	0.35	0.47	0.42	0.24	0.32
IRA only	0.05	0.06	0.06	0.04	0.05	0.05
Both	0.04	0.13	0.28	0.47	0.68	0.32
Total	0.21	0.54	0.81	0.93	0.97	0.69

This table reports a breakdown of retirement account participation for households with a member aged 50 to 59 in the 2019 administrative tax data.

Table 3
Access to an Employer Retirement Plan

	Percentile of usual income					
Year	0–20	20–40	40–60	60–80	80–100	All
2015	0.42	0.81	0.93	0.97	0.97	0.82
2016	0.41	0.80	0.93	0.97	0.97	0.82
2017	0.40	0.80	0.93	0.97	0.97	0.81
2018	0.39	0.80	0.93	0.97	0.96	0.81
2019	0.38	0.80	0.93	0.97	0.97	0.81

This table reports access to an employer retirement plan (i.e., defined benefit and defined contribution plans) for households with a member aged 50 to 59 in the administrative tax data.

Table 4
Bank Account Participation by Income

	Percentile of usual income					
Year	0–20	20–40	40–60	60–80	80–100	All
Panel A. Administrative tax data						
2015	0.71	0.94	0.99	1.00	1.00	0.93
2016	0.71	0.94	0.98	1.00	1.00	0.93
2017	0.69	0.94	0.98	1.00	1.00	0.92
2018	0.68	0.94	0.98	0.99	1.00	0.92
2019	0.70	0.94	0.98	0.99	1.00	0.92
Panel B. SCF						
2016	0.76	0.93	0.98	0.99	1.00	0.93
2019	0.80	0.95	1.00	1.00	1.00	0.95

This table reports bank account participation for households with a member aged 50 to 59.

Table 5
Determinants of Financial Participation in the SCF

	Retirement		Bank	
Regressor	(1)	(2)	(3)	(4)
Race:				
Hispanic	-0.34	-0.21	-0.11	-0.06
	(0.02)	(0.01)	(0.01)	(0.01)
Black	-0.21	-0.08	-0.12	-0.08
	(0.01)	(0.01)	(0.01)	(0.01)
Other nonwhite	-0.06	-0.05	-0.05	-0.05
	(0.01)	(0.01)	(0.01)	(0.01)
Log income		0.21		0.07
		(0.00)		(0.00)
Constant	0.79	0.76	0.96	0.95
	(0.01)	(0.01)	(0.00)	(0.00)
R^2	0.06	0.26	0.04	0.12
Observations	16,496	16,496	16,496	16,496

This table reports regressions of retirement or bank account participation on household characteristics. The coefficient for log income is standardized. All specifications include year fixed effects, which are not reported for brevity. Heteroskedasticity-robust standard errors are reported in parentheses. The sample includes households with a member aged 50 to 59 in the 2016 and 2019 SCF.

Table 6
Geographic Determinants of Retirement Account Participation

Regressor	(1)	(2)	(3)
Race:			
Hispanic	-0.37	-0.18	-0.22
	(0.01)	(0.01)	(0.01)
Black	-0.29	-0.07	-0.06
	(0.01)	(0.01)	(0.01)
Asian	0.04	-0.03	-0.11
	(0.03)	(0.01)	(0.01)
Other nonwhite	-0.29	-0.01	-0.06
	(0.01)	(0.01)	(0.01)
Average log income		0.10	0.10
		(0.00)	(0.00)
Constant	0.82	0.75	0.76
	(0.00)	(0.00)	(0.00)
Commuting zone fixed effects			Y
R^2	0.46	0.86	0.90
Observations	159,077	159,077	159,077

This table reports panel regressions of ZCTA-level retirement account participation on geographic characteristics. The coefficient for average log income is standardized. All specifications include year fixed effects, which are not reported for brevity. Standard errors, reported in parentheses, are robust to heteroskedasticity and clustering by ZCTA. The sample includes all households with a member aged 50 to 59 in the 2015 to 2019 administrative tax data. The observations are weighted by the census-derived household count in each cell.

Table 7
Geographic Determinants of Bank Account Participation

Regressor	(1)	(2)	(3)	(4)
Race:				
Hispanic	-0.06	-0.06	0.03	0.02
	(0.00)	(0.00)	(0.00)	(0.00)
Black	-0.09	-0.09	0.01	0.01
	(0.00)	(0.00)	(0.00)	(0.00)
Asian	0.03	0.03	0.00	0.01
	(0.00)	(0.00)	(0.00)	(0.00)
Other nonwhite	-0.14	-0.15	-0.02	-0.02
	(0.01)	(0.01)	(0.00)	(0.00)
Bank branch density		0.00		
		(0.00)		
Average log income			0.04	0.04
			(0.00)	(0.00)
Constant	0.96	0.96	0.93	0.93
	(0.00)	(0.00)	(0.00)	(0.00)
Commuting zone fixed effects				Y
R^2	0.19	0.20	0.82	0.87
Observations	159,077	159,077	159,077	159,077

This table reports panel regressions of ZCTA-level bank account participation on geographic characteristics. The coefficients for bank branch density and average log income are standardized. All specifications include year fixed effects, which are not reported for brevity. Standard errors, reported in parentheses, are robust to heteroskedasticity and clustering by ZCTA. The sample includes all households with a member aged 50 to 59 in the 2015 to 2019 administrative tax data. The observations are weighted by the census-derived household count in each cell.

Table 8

Instrumental Variables Regression for Retirement Account Participation

Regressor	Percentile of usual income 0–20	20–40	40–60	60–80	80–100
Access to employer plan	0.17	0.18	0.24	0.25	0.24
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Number of years with access	0.03	0.03	0.03	0.02	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log income	-0.03	0.04	0.04	0.03	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.16	0.25	0.37	0.54	0.69
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Weak instrument test	24,845	22,670	23,421	18,495	33,718
Observations	5,157,974	2,950,170	1,480,696	713,610	687,616

The first endogenous regressor is an indicator variable for access to an employer retirement plan by either spouse in the last ten years. The second endogenous regressor is the number of years that either spouse had access to an employer retirement plan over the last ten years. The intent-to-treat instrument is the counterfactual access to an employer retirement plan had the worker remained with the same employer since 2010. The coefficient for log income is standardized. Heteroskedasticity-robust standard errors are reported in parentheses. The critical value for a test of weak instruments at the 5 percent significance level is 7.03 (Stock and Yogo 2005, table 5.2). The sample includes all households with a member aged 50 to 59 in the 2019 administrative tax data, who did not have access to an employer retirement plan in 2010.

Table 9

State-Sponsored Retirement Savings Programs

State	Program	Default contribution rate (%)	Employers with at least: Workers	Employers with at least: Years in business
California	CalSavers	5	5	
Colorado	Secure Savings	5	5	2
Connecticut	Secure Choice	3	5	
Illinois	Secure Choice	5	25	2
Maine	MaineSaves	5	25	
Maryland	MarylandSaves	5	1	2
New Jersey	Secure Choice	3	25	2
New York	Secure Choice	3	10	2
Oregon	OregonSaves	5	1	
Virginia	VirginiaSaves		25	2

Table 10
Predicted Change in Retirement Account Participation

	Percentile of usual income				
Participation rate	0–20	20–40	40–60	60–80	80–100
Actual	0.31	0.62	0.84	0.94	0.97
Counterfactual	0.41	0.70	0.88	0.96	0.98
Predicted change	0.10	0.08	0.04	0.01	0.01
Observations	6,716,121	6,716,144	6,716,097	6,716,121	6,716,120

The counterfactual is the predicted retirement account participation rate if workers were to have access to an employer retirement plan during all working years. The sample includes all households with a member aged 50 to 59 in the 2019 administrative tax data.

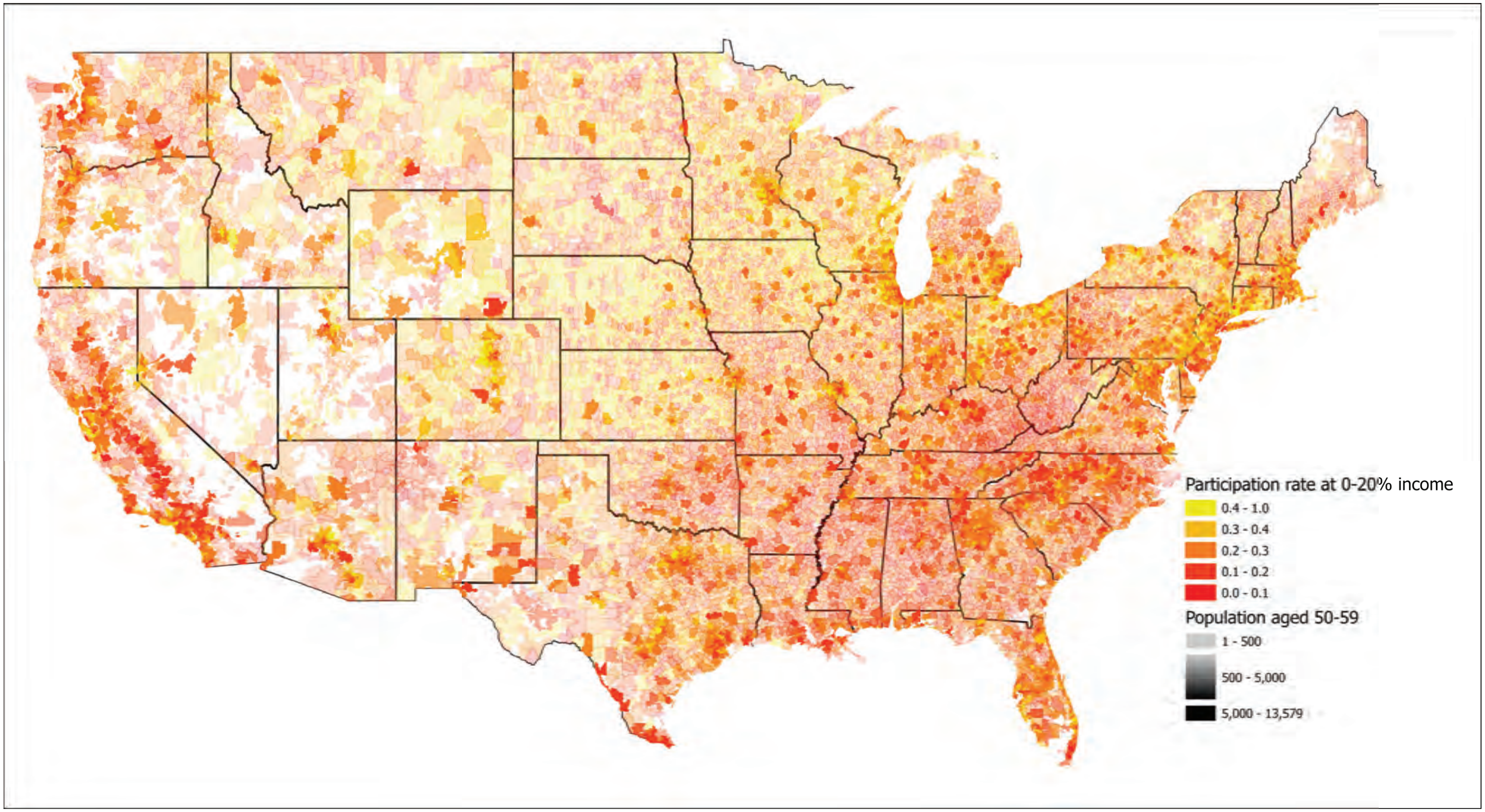

Figure 1. Retirement Account Participation at 0–20 Percentile Income

Retirement accounts include employer retirement plans (i.e., defined benefit and defined contribution plans) and IRAs. The sample includes all households in the lowest income quintile with a member aged 50 to 59 in the 2019 administrative tax data.

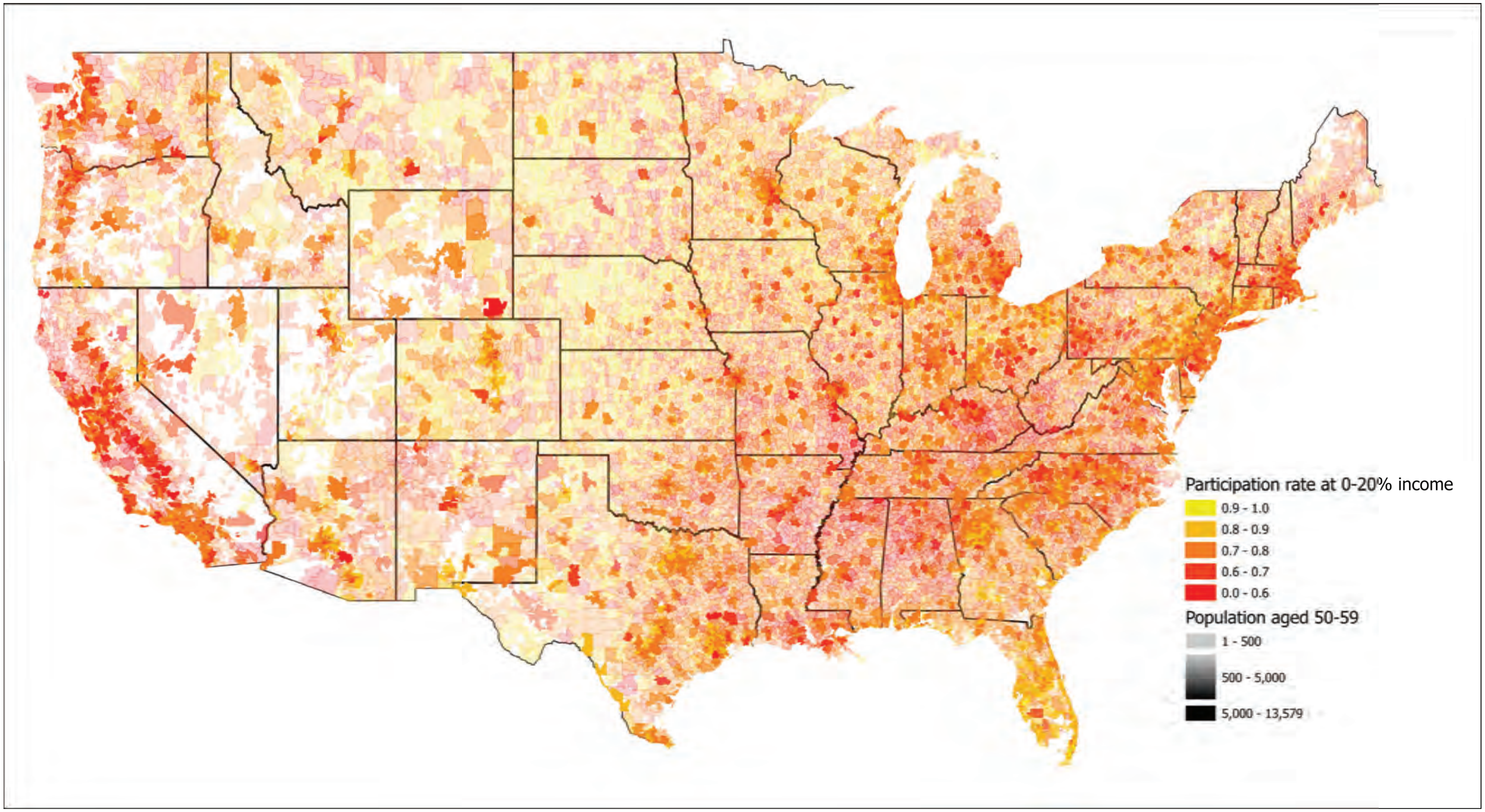

Figure 2. Bank Account Participation at 0–20 Percentile Income
The sample includes all households in the lowest income quintile with a member aged 50 to 59 in the 2019 administrative tax data.

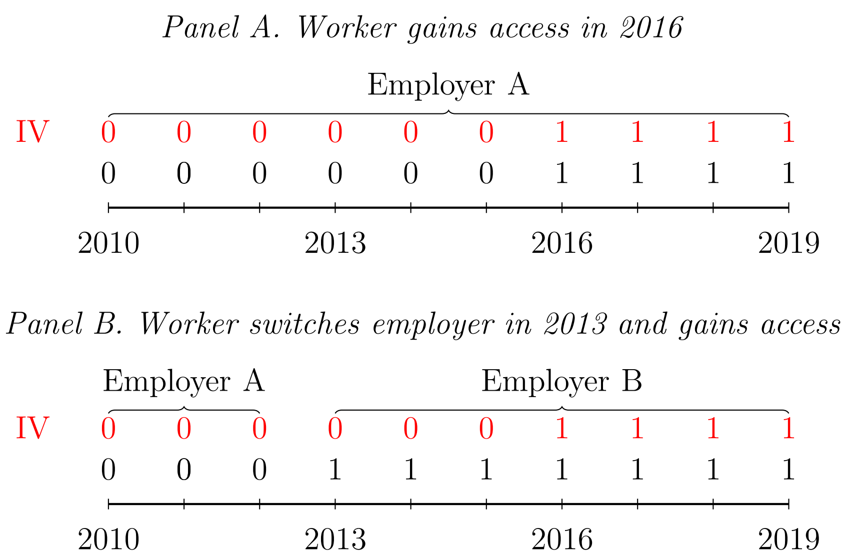

Figure 3. Illustration of the Instrument

In Panel A, the worker stays with employer A from 2010 to 2019 and gains access to an employer retirement plan in 2016. In Panel B, the worker switches from employer A to employer B and gains access to an employer retirement plan in 2013. The intent-to-treat instrument in red is the counterfactual access to an employer retirement plan had the worker remained with employer A from 2010 to 2019.

Appendix A. Administrative Tax Data

A. Population Count

We sample all individuals aged 50 to 59 from 2015 to 2019, who have either a tax return or an information return and have a ZIP Code within the U.S. states or Washington, DC. Table A1 compares the population count in the administrative tax data with the resident population aged 50 to 59 in the census. In 2019, the administrative tax data includes 43.193 million individuals, which is 1.02 times the census population. The ratio of the population count in the administrative tax data to that in the census is constant at 1.01 between 2015 and 2018.

Table A1
Comparing the Population Count with the Census

Year	Tax data	Census	Ratio
2015	44,219	43,985	1.01
2016	44,199	43,738	1.01
2017	43,916	43,278	1.01
2018	43,459	42,777	1.02
2019	43,324	42,355	1.02

The population counts are reported in thousands. The population count in the administrative tax data includes all individuals aged 50 to 59, who have either a tax return or an information return and have a ZIP Code within the U.S. states or Washington, DC. The census population count is the resident population aged 50 to 59.

In theory, the population counts should be identical. In practice, the population count in the administrative tax data could be larger than that in the census if foreign residents file taxes or receive information returns at a U.S. mailing address. The discrepancy in the population counts is sufficiently small that we are not concerned for the purposes of this paper.

B. Aggregation

We construct three aggregate data sets from the household-level data. The first data set (Data_incm) is aggregated by year and income quintile. The second data set (Data_ZCTA) is aggregated by year and ZCTA. The third data set (Data_ZCTA_incm) is aggregated by year, ZCTA, and income quintile. The aggregation serves two purposes. First, the ZCTA-level data allow us to study the geographic determinants of financial participation. Second, we cannot disclose household-level data, but we can share aggregate data publicly to facilitate replication and future research.

For each of these data sets, we construct the following variables. We take three steps to avoid revealing information about specific households and to comply with data sharing requirements. First, we round usual household income to the nearest $100 after aggregation. We refer to Section I for the definition of usual household income. Second, we mask observations that would otherwise be derived from cells with fewer than 100 households by aggregating with other cells. Data_incm is not subject to masking because the cells are sufficiently large. Third, we do not report the household count in each cell and instead estimate it based on census data.

- incm_usual: Average usual household income.
- Lincm_usual: Average log usual household income.
- d_bank: Share of households that have a bank account.
- d_retire: Share of households that have a retirement account.
- d_emp_part: Share of households that have an employer retirement plan.
- d_emp_access: Share of households that have access to an employer retirement plan.
- d_emp_only: Share of households that have only an employer retirement plan.
- d_ira_only: Share of households that have only an IRA.
- d_emp_ira: Share of households that have both an employer retirement plan and an IRA.
- hh2: Share of households with two people aged 50 to 59.
- md_incm (for Data_ZCTA_incm only): Share of households in a given income quintile within a ZCTA.
- obs: Census-derived household count in each cell.

For Data_ZCTA, we define a cell as a {year, ZCTA} couplet and a small cell as that with fewer than 100 households. We sort all small cells by the average of the variable that is being constructed. We group adjacent cells so that each group has between 100 and 300 households. By sorting before grouping, we maximize the chance that cells with similar average values are grouped together. We then calculate the weighted average of the variable within each group and assign it to all cells in that group. Finally, we discard the household count in each cell that was used for the weighted average. We repeat the masking procedure for all variables. About 28 percent of the 32,850 cells in 2019 are subject to masking.

For Data_ZCTA_incm, we apply the same masking procedure for all variables with three changes. First, we define a cell as a {year, ZCTA, income quintile} triplet. Second, we define a small cell as that (a) with fewer than 20 households or (b) is nested in a {year, ZCTA} couplet with fewer than 100 households. Thus, we do not mask a cell if and only if it has at least 20 households and is part of a ZCTA with at least 100 households. Third, we group adjacent cells in two steps. We first group cells that are small according to definition (b) so that each group has between 100 and 300 households. We then group the remaining cells that are small according to definition (a) so that each group has between 20 and 60 households. About 31 percent of the 160,914 cells in 2019 are subject to masking.

For analysis that requires household counts, we estimate them based on the population aged 50 to 59 by ZCTA in the American Community Survey Demographic and Housing Five-Year Estimates. Because these data are individual counts, we need to make an adjustment for households that have two people aged 50 to 59 to avoid double counting. For Data_ZCTA, we approximate the household count by ZCTA as Population/(1+hh2). For Data_ZCTA_incm, we approximate the household count by ZCTA and income quintile as Population$\times$md_incm/(1+hh2).

Appendix B. Other Data

A. American Community Survey

We construct the population shares by race at the ZCTA level, based on the American Community Survey Demographic and Housing Five-Year Estimates. We group race into white, Hispanic, black, Asian, or other nonwhite. Other nonwhite includes American Indian, Alaska Native, Native Hawaiian, other Pacific Islander, and multiple race.

B. Federal Deposit Insurance Corporation

We count the number of bank branches by ZCTA, based on the Federal Deposit Insurance Corporation's Annual Survey of Branch Office Deposits. We construct bank branch density as the number of branches divided by the population within a ZCTA. We winsorize the right tail at one branch per 1,000 residents (about 7 percent of observations) to reduce the impact of outliers.