Center for Financial Security
UNIVERSITY OF WISCONSIN-MADISON

Retirement & Disability Research Center

Motohiro Yogo
Princeton University and NBER

Andrew Whitten
U.S. Department of the Treasury

Natalie Cox
Princeton University

Financial Inclusion Across the United States

This research was conducted while Whitten was an employee of the U.S. Department of the Treasury. 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. This research was derived in part from research activities performed pursuant to a grant from the U.S. Social Security Administration (SSA) funded as part of the Retirement and Disability Research Consortium. 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 Treasury, SSA, or agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of the contents of this paper. Reference herein to any specific commercial product, process or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply endorsement, recommendation or favoring by the U.S. Government or any agency thereof. For comments and discussions, we thank Edith Brashares, Alice Henriques Volz, Neviana Petkova, John Sabelhaus, and Alexander Zentefis. 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, UCLA, University of Michigan, University of Wisconsin, 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, the 2022 Chicago Household Finance Conference, and the 2023 AEA Annual Meeting.

Center for Financial Security

University of Wisconsin-Madison

1300 Linden Drive
Madison, WI 53706

608-890-0229
cfs@mailplus.wisc.edu
cfs.wisc.edu

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 retirement and bank accounts, respectively. For the same group, 38 percent of households had access to an employer retirement plan. Geographic variation in financial participation primarily relates to income rather than racial composition. By instrumental variables, we estimate the causal effect of access to an employer retirement plan. Universal access with automatic enrollment could increase retirement account participation by 17 percentage points in the lowest income quintile over ten years.

JEL classification: D14, G51

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

1. Introduction

An inclusive society should strive for financial participation of all households, regardless of income or race. Survey evidence shows 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 have 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 implies 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 racial composition, average income, 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 average income. Bank account participation is also negatively correlated with the Hispanic and Black population shares, but these correlations disappear conditional on average income. Bank branch density has low explanatory power for bank account participation at the ZCTA level. However, this finding does not rule out the importance of other supply factors such as banking fees (Dlugosz, Melzer, and Morgan 2021) or spatial discrimination at a more granular geographic level (Sakong and Zentefis 2022). Overall, average income is the strongest predictor of ZCTA-level financial participation. In fact, the explanatory power of average income is so high that geography plays little role in financial participation conditional on average 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 that lacked access to an employer retirement plan in 2010. The identifying assumption is that workers did not choose their 2010 employer based on expectation of future retirement benefits at the beginning of the employment relationship. Some workers subsequently gain access if their employer starts a 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 had the worker remained with the same employer in our sample period after 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 15 percentage points in the lowest income quintile. Automatic enrollment further increases retirement account participation by 34 percentage points for these households. On the intensive margin, each additional year of access to an employer retirement plan without automatic enrollment increases retirement account participation by 4 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 estimate 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 with automatic enrollment had already been in effect nationally since 2010. In this counterfactual, retirement account participation increases by 17 percentage points in the lowest income quintile and 11 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 starting a retirement plan. Workers can deduct retirement contributions from taxable income and earn tax-deferred returns. Tax incentives affect even low-income workers, who may face a low marginal income tax rate and a zero capital gains tax rate, 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. Our findings suggest that tax incentives for employers to offer retirement plans may be more effective than those for workers to save in retirement accounts.

Based on the administrative tax data, we can 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. By not participating in retirement accounts, households are forgoing valuable tax savings. In particular, the Saver's Credit is like a matching contribution of 100 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, bank account participation increases wealth accumulation and durable good purchases (Célerier and Matray 2019; Stein and Yannelis 2020). 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 some equity exposure is optimal.

Our study contributes to the literature on wealth inequality. Wealth inequality is greater 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 1 percent. In comparison, the wealth share was 2 percent for the bottom half of households and 33 percent for the top 1 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 1 percent. Because retirement and bank accounts are the most important means of accumulating financial wealth, studying low participation at the bottom of the income distribution is important for a better understanding of wealth inequality.

The remainder of the paper is organized as follows. Section 2 describes how we construct the sample of households and measures of income and financial participation from the administrative tax data. Section 3 summarizes financial participation by income. In Section 4, we study the relation between geographic characteristics and financial participation. In Section 5, 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 6 concludes.

2. Data Construction

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

2.1. 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 2006 to 2019. 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).

2.1.1. Sample

For each year from 2015 to 2019, we sample all individuals aged 50 to 59, who have either a tax return or an information return with a ZIP Code within the U.S. states or Washington, DC. We focus on ages 50 to 59 as the most relevant part of the life cycle for retirement saving. As we show in Appendix A, the sample includes over 43 million individuals each year, 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 by 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). We define a spouse as a current joint filer on a Form 1040 or, for those who do not currently file a Form 1040, a previous joint filer in the previous ten years who currently has the same household identifier (i.e., the same address) or appears on the same Form 1095 as the sampled individual.[2] The latter two

[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, Mortenson, and Splinter (2021) constructed the household identifiers based on a textual analysis of the addresses on tax returns and information returns.

criteria ensure that we do not artificially 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 non-spousal household members (e.g., parents or children living at the same address) share financial accounts.

2.1.2. Income

We construct pre-tax household income following Larrimore, Mortenson, and Splinter (2021). For tax filers who have a Form 1040, we start with total income (e.g., line 7b on the 2019 form), 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, Mortenson, and Splinter (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 2019 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 national distribution of usual household income within each year. 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).

2.1.3. 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 previous 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 from 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 previous 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 an individual's employer in a given year. We search through all

[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.

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 previous ten years. By merging the administrative tax data with the Department of Labor's Form 5500 (see Appendix B), we measure whether the employer retirement plan has automatic enrollment.

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 (access to) an account if either spouse has (access to) an 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 is 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 (Mortgage Interest Statement). Furthermore, we could measure home ownership based on Form 1098 and itemized deductions for property taxes. However, these measures of financial participation are more difficult to interpret because a household may prefer to rent. Similarly, we cannot tell whether a household does not have a mortgage because it does not need one or has been denied. In related work, Lurie and Pearce (2019) use the administrative tax data to study health insurance coverage.

2.2. Survey of Consumer Finances

To benchmark our findings to survey evidence, we use the 2016 and 2019 Survey of Consumer Finances (Board of Governors of the Federal Reserve System 2016–2019). 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 by including food stamps and other government support that are not observed in the administrative tax data. We define bank account participation as ownership of a checking account, savings account, or money market fund. We define retirement account participation as either ownership of or payments from an employer retirement account or an IRA.

3. 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 relation between household characteristics and financial participation in the SCF, which help us interpret the geographic analysis in Section 4.

3.1. 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.

TABLE 1
RETIREMENT 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	0.22	0.56	0.83	0.94	0.97	0.70
2016	0.21	0.55	0.82	0.94	0.97	0.70
2017	0.21	0.55	0.82	0.94	0.97	0.70
2018	0.21	0.55	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.

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 2 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, 92 percent of households with a member aged 50 to 59 had bank accounts in the administrative tax data, which is close to 93 percent in 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.

Table 2
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	0.70	0.94	0.99	1.00	1.00	0.92
2017	0.68	0.94	0.98	1.00	1.00	0.92
2018	0.67	0.93	0.98	0.99	1.00	0.92
2019	0.70	0.93	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.

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, 70 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 implies 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 2, the SCF uses a broader definition of income that includes 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.

3.2. Additional Facts About Retirement Accounts

Table 3 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, 12 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 employers. Higher-income households are more likely to have both an employer retirement plan and an IRA. In the highest income quintile, 23 percent of households have only an employer retirement plan, 5 percent have only an IRA, and 69 percent have both.[4]

Table 3
Breakdown of Retirement Account Participation

	Percentile of usual income					
Households with	0–20	20–40	40–60	60–80	80–100	All
Employer plan only	0.12	0.35	0.47	0.41	0.23	0.32
IRA only	0.05	0.06	0.06	0.05	0.05	0.05
Both	0.04	0.13	0.28	0.47	0.69	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 4 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 (at any point in the previous ten years). Access to an employer retirement plan increases to 79 percent of households in the second income quintile and 93 percent of households in the third income quintile. Although the time series are short, access to an employer retirement plan apparently declines in the lowest income quintile from 42 percent of households in 2015 to 38 percent of households in 2019.

[4] Table 3 could understate the importance of employer retirement plans if some households have IRAs that were funded entirely by rollovers from employer retirement plans.

Table 4
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	0.96	0.82
2016	0.41	0.80	0.93	0.97	0.96	0.81
2017	0.39	0.80	0.93	0.97	0.96	0.81
2018	0.38	0.80	0.93	0.97	0.96	0.81
2019	0.38	0.79	0.93	0.96	0.96	0.80

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.

3.3. Household Characteristics Related to Financial Participation

To help us interpret the geographical analysis in Section 4, we summarize the relation between household characteristics and financial participation in the SCF (Hogarth, Anguelov, and Lee 2005). Columns 1 to 3 of Table 5 report regressions of retirement account participation on household characteristics for households with a member aged 50 to 59 in 2019. The regression is a linear probability model, where the dependent variable is an indicator variable that is one for participants. In column 1, race is a significant predictor of retirement account participation. Households with Hispanic respondents are 37 percentage points less likely to have a retirement account than households with white respondents. Households with Black respondents are 17 percentage points less likely to have a retirement account. Households with other nonwhite respondents are 4 percentage points less likely to have a retirement account.

In column 2 of Table 5, log income is also a significant predictor of retirement account participation. A standard deviation increase in log income predicts a 24 percentage point increase in retirement account participation.

In column 3 of Table 5, we test whether race or income is a stronger predictor of retirement account participation. Relative to column 1, the coefficients for race significantly decrease in magnitude. The coefficient for Hispanic decreases in magnitude from -0.37 to -0.23. Similarly, the coefficient for Black decreases in magnitude from -0.17 to -0.03. Thus, race is a much weaker predictor of retirement account participation conditional on income.

Columns 4 to 6 of Table 5 report analogous regressions for bank account participation. In column 4, race is a significant predictor of bank account participation. Households with Hispanic respondents are 7 percentage points less likely to have a bank account than households with white respondents. Households with Black respondents are 11 percentage points less likely to have a bank account. Households with other nonwhite respondents are 4 percentage

TABLE 5
HOUSEHOLD CHARACTERISTICS RELATED TO FINANCIAL PARTICIPATION

	Retirement			Bank		
Regressor	(1)	(2)	(3)	(4)	(5)	(6)
Race:						
Hispanic	-0.37		-0.23	-0.07		-0.03
	(0.02)		(0.02)	(0.01)		(0.01)
Black	-0.17		-0.03	-0.11		-0.07
	(0.02)		(0.02)	(0.01)		(0.01)
Other nonwhite	-0.04		-0.04	-0.04		-0.04
	(0.02)		(0.02)	(0.01)		(0.01)
Log income		0.24	0.23		0.07	0.06
		(0.00)	(0.00)		(0.00)	(0.00)
Constant	0.75	0.69	0.72	0.97	0.95	0.96
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
R^2	0.06	0.27	0.29	0.03	0.10	0.11
Observations	7,795	7,795	7,795	7,795	7,795	7,795

This table reports regressions of retirement or bank account participation on household characteristics. The coefficient for log income is standardized. Heteroskedasticity-robust standard errors are reported in parentheses. The sample includes households with a member aged 50 to 59 in the 2019 SCF.

points less likely to have a bank account.

In column 5 of Table 5, log income is also a significant predictor of bank account participation. A standard deviation increase in log income predicts a 7 percentage point increase in bank account participation.

In column 6 of Table 5, we test whether race or income is a stronger predictor of bank account participation. Relative to column 4, the coefficients for race decrease in magnitude but remain statistically significant. The coefficient for Hispanic decreases in magnitude from -0.07 to -0.03. Similarly, the coefficient for Black decreases in magnitude from -0.11 to -0.07. 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 a strong predictor 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 (Brown, Cookson, and Heimer 2019). A second hypothesis is spatial discrimination based on the location of bank branches (Sakong and Zentefis 2022). The cost of accessing banking services may be high for nonwhite households if there are no bank branches nearby. These hypotheses motivate the geographic analysis in the next section.

4. Geography of Financial Participation

We first construct maps to summarize the geographic variation in financial participation. We then study whether the geographic variation in financial participation relates to racial composition, average income, or access to financial services. Because we do not observe race in the administrative tax data, we cautiously interpret the relation between financial participation and racial composition at the ZCTA level. We conclude this section by reconciling the relation between geographic characteristics and financial participation with the analogous relation between household characteristics and financial participation in the SCF.

4.1. Maps of Financial Participation

Figure 1 is a map of retirement account participation by ZCTA in the lowest income quintile in 2019. We focus on the lowest income quintile because it has the greatest variation in financial participation and because policymakers may be especially interested in these households. 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 represents a populous ZCTA with low retirement account participation. Figure 2 is a similar map of bank account participation by ZCTA in the lowest income quintile in 2019. The colors range from yellow (90–100 percent participation) to red (0–60 percent participation).

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

4.2. Geographic Characteristics Related to Financial Participation

4.2.1. Retirement Accounts

We estimate regressions of ZCTA-level retirement account participation on geographic characteristics in 2019. We weight the observations by the census-derived household count in each cell (see Appendix A). In column 1 of Table 6, racial composition is a significant predictor of retirement account participation. Relative to the omitted category of the white population share, a percentage point increase in the Hispanic population share predicts a 37

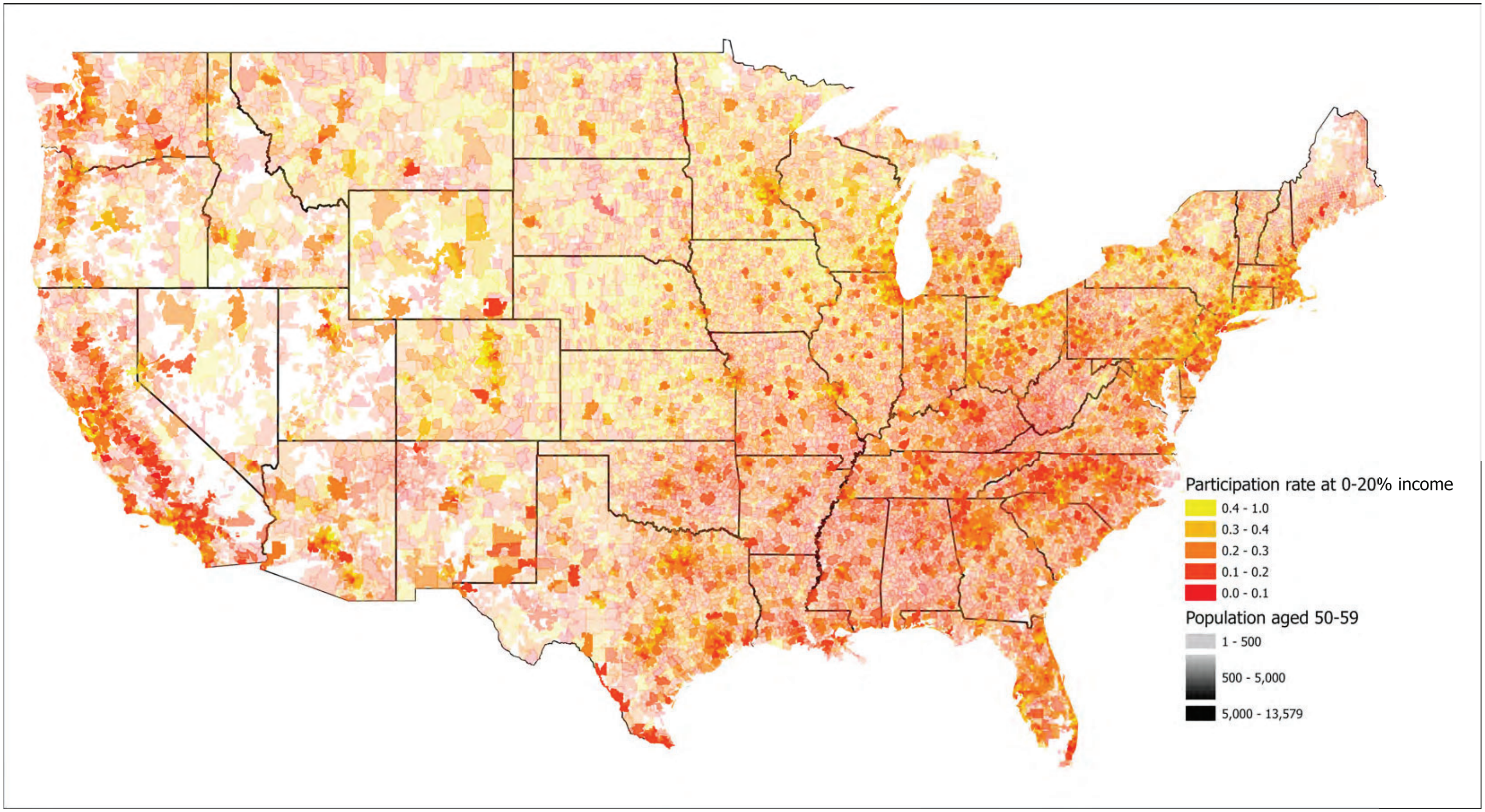

Figure 1. Retirement Account Participation at 0–20 Percentile Income

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.

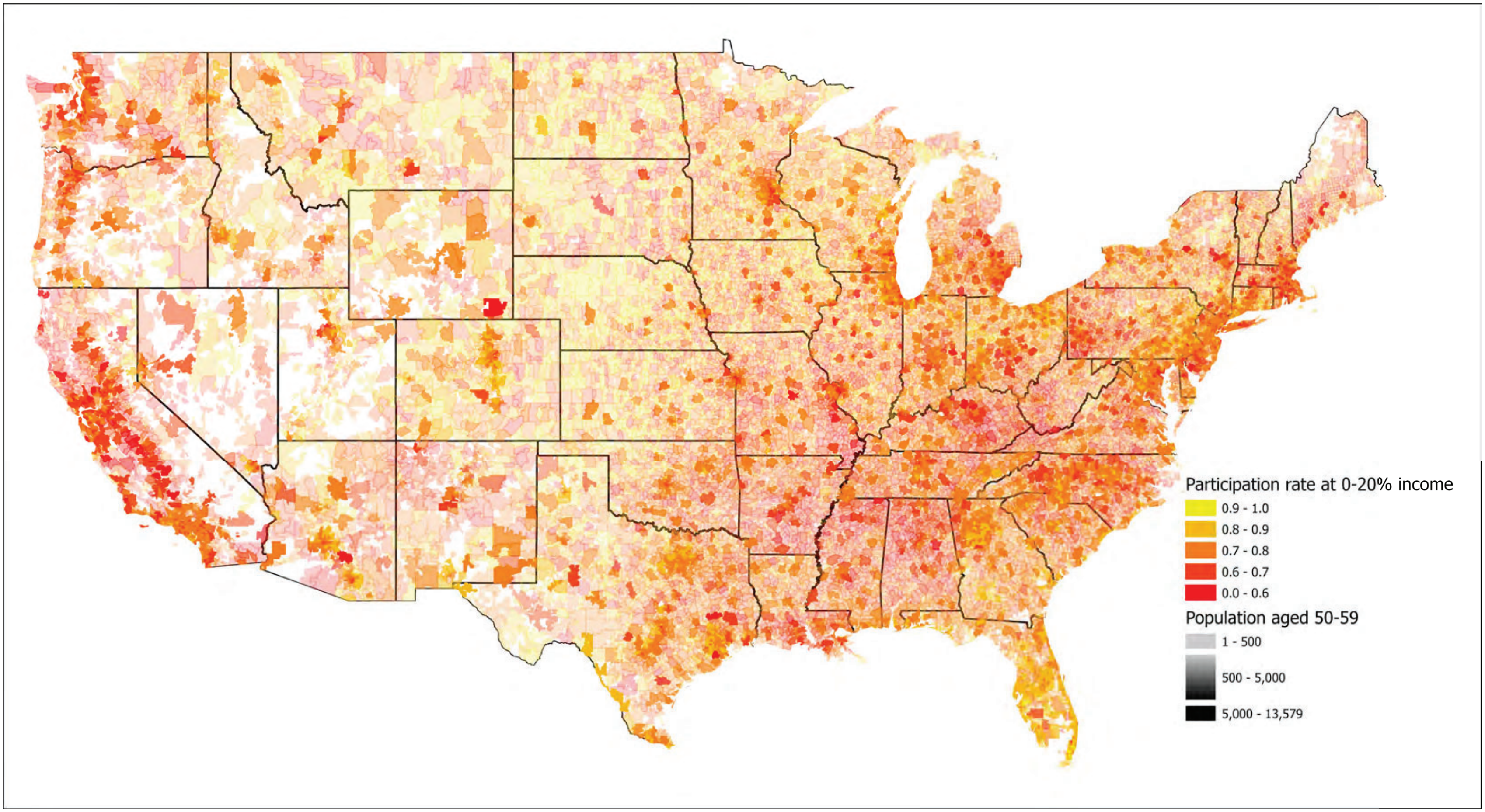

Figure 2. Bank Account Participation at 0–20 Percentile Income
The sample includes all households in the lowest income quintile with a member aged 50 to 59 in the 2019 administrative tax data.

basis point decrease in retirement account participation. A percentage point increase in the Black population share predicts a 30 basis point decrease in retirement account participation. A percentage point increase in the other nonwhite population share predicts a 30 basis point decrease in retirement account participation. In contrast, a percentage point increase in the Asian population share predicts a 7 basis point increase in retirement account participation.

Table 6
Geographic Characteristics Related to Retirement Account Participation

Regressor	(1)	(2)	(3)	(4)
Race:				
Hispanic	-0.37		-0.18	-0.23
	(0.01)		(0.01)	(0.01)
Black	-0.30		-0.06	-0.06
	(0.01)		(0.00)	(0.01)
Asian	0.07		-0.05	-0.12
	(0.02)		(0.01)	(0.01)
Other nonwhite	-0.30		0.00	-0.06
	(0.01)		(0.01)	(0.01)
Average log income		0.12	0.10	0.10
		(0.00)	(0.00)	(0.00)
Constant	0.81	0.70	0.74	0.76
	(0.00)	(0.00)	(0.00)	(0.00)
Commuting zone fixed effects				Y
R^2	0.45	0.80	0.86	0.91
Observations	31,724	31,724	31,724	31,724

This table reports regressions of ZCTA-level retirement account participation on geographic characteristics. The coefficient for average log income is standardized. Heteroskedasticity-robust standard errors are reported in parentheses. The sample includes all households with a member aged 50 to 59 in the 2019 administrative tax data. The observations are weighted by the census-derived household count in each cell.

In column 2 of Table 6, average log income is also a significant predictor of retirement account participation. A standard deviation increase in average log income predicts a 12 percentage point increase in retirement account participation. In fact, average log income alone explains 80 percent of the geographic variation in retirement account participation.

In column 3 of Table 6, we test whether racial composition or average income is a stronger predictor of retirement account participation. Relative to column 1, the coefficients for racial composition significantly decrease in magnitude. The coefficient for the Hispanic population share decreases in magnitude from -0.37 to -0.18. The coefficient for the Black population share decreases in magnitude from -0.30 to -0.06. The coefficient for the other nonwhite population share decreases in magnitude from -0.30 to 0.00. Thus, racial composition is a much weaker predictor of retirement account participation conditional on average income.

In column 4 of Table 6, we control for commuting zone fixed effects. Relative to column 3, the coefficients on racial composition increase in magnitude, and the coefficient on average income remains unchanged. The R^2 increases slightly from 86 to 91 percent. This result implies that the correlation of retirement account participation with racial composition and average income is about local geographic variation within commuting zones.

4.2.2. Bank Accounts

Analogous to Table 6, Table 7 reports regressions of ZCTA-level bank account participation on geographic characteristics in 2019. In column 1 of Table 7, racial composition is a significant predictor of bank account participation. Relative to the omitted category of the white population share, a percentage point increase in the Hispanic population share predicts a 5 basis point decrease in bank account participation. A percentage point increase in the Black population share predicts a 9 basis point decrease in bank account participation. A percentage point increase in the other nonwhite population share predicts a 15 basis point decrease in bank account participation. In contrast, a percentage point increase in the Asian population share predicts a 5 basis point increase in bank account participation.

In column 2 of Table 7, average log income is also a significant predictor of bank account participation. A standard deviation increase in average log income predicts a 4 percentage point increase in bank account participation. In fact, average log income alone explains 82 percent of the geographic variation in bank account participation.

In column 3 of Table 7, the regressors are an indicator variable for the presence of a bank branch and the bank branch density at the ZCTA level (see Appendix B). On the extensive margin, bank account participation is 1 percent higher in ZCTAs with at least one bank branch. On the intensive margin, bank branch density has no explanatory power for bank account participation. Overall, bank branch density explains only 1 percent of the geographic variation in bank account participation.

Our finding is a cross-sectional correlation and does not rule out a causal effect of branch access on bank account participation. In fact, Célerier and Matray (2019) find that deregulation between 1994 and 2005 led to an expansion of bank branches and increased bank account participation among low-income households. Our finding also does not rule out the importance of other supply factors. Dlugosz, Melzer, and Morgan (2021) find that banking fees are an important supply factor that could explain bank account participation. In particular, a cap on overdraft fees could constrain the supply of overdraft credit and deposit accounts through a higher minimum deposit requirement. Using more granular data at the census block-group level, Sakong and Zentefis (2022) find worse bank branch access in areas with a higher Black population share. However, they also find worse bank branch access in

Table 7
Geographic Characteristics Related to Bank Account Participation

Regressor	(1)	(2)	(3)	(4)	(5)
Race:					
Hispanic	-0.05			0.03	0.02
	(0.00)			(0.00)	(0.00)
Black	-0.09			0.01	0.01
	(0.00)			(0.00)	(0.00)
Asian	0.05			0.00	0.02
	(0.00)			(0.00)	(0.00)
Other nonwhite	-0.15			-0.02	-0.02
	(0.01)			(0.00)	(0.00)
Average log income		0.04		0.05	0.05
		(0.00)		(0.00)	(0.00)
Bank branch presence			0.01	0.00	0.00
			(0.00)	(0.00)	(0.00)
Bank branch density			0.00	0.00	0.00
			(0.00)	(0.00)	(0.00)
Constant	0.95	0.93	0.92	0.93	0.93
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Commuting zone fixed effects					Y
R^2	0.18	0.82	0.01	0.84	0.89
Observations	31,724	31,724	31,724	31,724	31,724

This table reports regressions of ZCTA-level bank account participation on geographic characteristics. The coefficients for bank branch density and average log income are standardized. Heteroskedasticity-robust standard errors are reported in parentheses. The sample includes all households with a member aged 50 to 59 in the 2019 administrative tax data. The observations are weighted by the census-derived household count in each cell.

higher income areas, which is offset by a higher demand for banking services. Thus, their finding is consistent with an overall positive correlation between bank account participation and average income due to demand factors.

In column 4 of Table 7, we test whether racial composition or average income is a stronger predictor of bank account participation. Relative to column 1, the coefficients for racial composition significantly decrease in magnitude. The coefficient for the Hispanic population share switches sign from -0.05 to 0.03. The coefficient for the Black population share decreases in magnitude from -0.09 to 0.01. The coefficient for the other nonwhite population share decreases in magnitude from -0.15 to -0.02. Thus, racial composition is a much weaker predictor of bank account participation conditional on average income.

In column 5 of Table 7, we control for commuting zone fixed effects. Relative to column 4, the coefficients are virtually unchanged. The R^2 increases slightly from 84 to 89 percent. This result implies that the correlation between bank account participation and average income is about local geographic variation within commuting zones.

4.3. Interpreting Income

At face value, income is the strongest predictor of financial participation because households without enough income are unable to save. Moreover, 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. In addition to these direct effects of income, financial participation could be correlated with income for more subtle reasons.

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.

Bumcrot, Lin, and Lusardi (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 6 and 7 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, used by Bumcrot, Lin, and Lusardi (2013), 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.

Peer effects could reinforce the geographic correlation between financial participation and income (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).

4.4. 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 broadly similar in magnitude to those in the SCF (i.e., column 1 of Table 5). Controlling for income, the correlation between retirement account participation and income overall weakens more in the aggregate tax data (i.e., column 3 of Table 6) than in the SCF (i.e., column 3 of Table 5).

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

We discuss two hypotheses for why income could have different effects on the correlation between financial participation and race in the aggregate tax data versus 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 bank account participation. The coefficients γ and Γ represent the individual and group effects, respectively, of race on bank account 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, we have

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

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 and the SCF. 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 result in conformity rather than polarization.

An alternative hypothesis is that financial participation depends on income but not race (i.e., $\gamma = \Gamma = 0$), and reported income in the SCF contains measurement error. If race correlates with true income, the coefficients for Hispanic and Black in the SCF would be downward biased (i.e., more negative). This bias is reduced in 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 relate to income.

5. Access to an Employer Retirement Plan

As reported in Table 3, the vast majority of households have retirement accounts through employer retirement plans rather than through only 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, which have lower access to employer retirement plans according to Table 4.

5.1. 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 of our data 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 lacked access to an employer retirement plan in 2010. The identifying assumption is that workers did not choose their 2010 employer based on expectation of future retirement benefits at the beginning of the employment relationship. For example, the worker would not have been able to predict that her employer would eventually start a retirement plan after 2010 at the time that she joined the firm (long before 2010 for most workers in our sample). Thus, we assume that the introduction of the employer retirement plan is an unexpected treatment.

Some of these workers subsequently gain access if their employer starts a retirement plan after 2010. 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 had the worker remained with the same employer in our sample period after 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.

The employer's decision to start a retirement plan is not random and could be related to its outcomes. For example, employer A in Figure 3 could have started a retirement plan in 2016 as a consequence of fast growth between 2010 and 2016. Our identifying assumption is that this growth would not have been predictable at the time that the worker joined the firm. The threat to identification is that the employer's propensity to start a retirement plan is correlated with worker characteristics. For example, more productive workers could cause the firm to grow faster, and these workers have a stronger saving motive and are more likely to enroll in an employer retirement plan. We control for income to address this particular

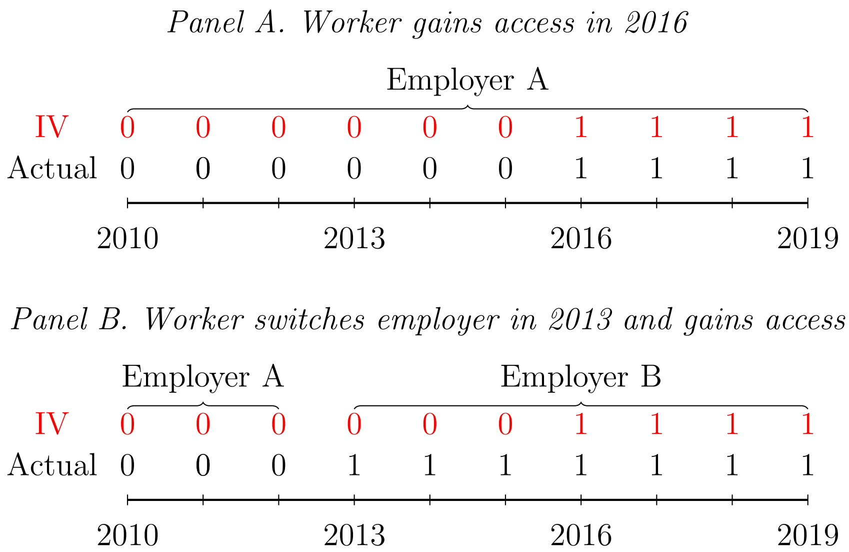

Figure 3. Illustration of the Instrument

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.

channel. However, any residual variation in worker productivity that is orthogonal to income remains a threat to identification.

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 more of these households lacked access to an employer retirement plan in 2010. The dependent variable is retirement account participation in 2019. The first endogenous regressor is an indicator variable for access to an employer retirement plan by either spouse from 2011 to 2019. The second endogenous regressor is additional years (beyond a year) that either spouse had access to an employer retirement plan without automatic enrollment. 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. The third endogenous regressor is an indicator variable for automatic enrollment in an employer retirement plan by either spouse from 2011 to 2019.

We construct three instruments corresponding to the three endogenous regressors, based on the counterfactual access to an employer retirement plan. For example, in Panel B of Figure 3, the endogenous regressors are 1 for the indicator variable for access to an employer retirement plan, 6 for the additional years of access, and either 0 or 1 for the indicator variable for automatic enrollment (depending on employer B's plan). The corresponding instruments are 1 for the indicator variable for access to an employer retirement plan, 3 for the additional years of access, and either 0 or 1 for the indicator variable for automatic enrollment (depending on employer A's plan).

5.2. 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. Before we discuss the estimates, we check for weak identification. In all income quintiles, the test static for rejecting the null of weak instruments is far greater than the critical value at the 5 percent significance level (Stock and Yogo 2005).[5] Furthermore, we report the first-stage regressions in Appendix C.

Table 8
Instrumental Variables Regression for Retirement Account Participation

	Percentile of usual income				
Regressor	0–20	20–40	40–60	60–80	80–100
Access to employer plan	0.15	0.18	0.24	0.28	0.31
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Additional years of access	0.04	0.04	0.04	0.02	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Automatic enrollment	0.34	0.32	0.28	0.15	0.05
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Log income	0.01	0.05	0.04	0.04	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.07	0.16	0.28	0.46	0.61
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Weak instrument test	10,036	11,335	8,857	5,688	9,523
Observations	5,837,709	3,538,967	1,709,152	792,476	719,025

The three endogenous regressors are an indicator variable for access to an employer retirement plan by either spouse from 2011 to 2019, additional years (beyond a year) that either spouse had access to an employer retirement plan without automatic enrollment, and an indicator variable for automatic enrollment in an employer retirement plan by either spouse. 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.

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 7 percent in the lowest income quintile to 61 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.

[5] Although Stock and Yogo (2005, Table 5.2) do not report the critical values for the case of three endogenous regressors, the critical value is less than 7.03 for the reported case of two endogenous regressors.

On the extensive margin, access to an employer retirement plan increases retirement account participation by 15 percentage points in the lowest income quintile. Automatic enrollment further increases retirement account participation by 34 percentage points for these households. On the intensive margin, each additional year (beyond a year) of access without automatic enrollment increases retirement account participation by 4 percentage points. Thus, automatic enrollment has a cumulative effect (i.e., $15+34$) that is close to that (i.e., $15+4\times 8$) for nine years of active enrollment. These estimates confirm the effectiveness of automatic enrollment for retirement account participation, especially for low- and middle-income households (Madrian and Shea 2001; Derby, Mackie, and Mortenson 2022).

On the one hand, the treatment effects for access to an employer retirement plan increase in income, implying higher takeup rates for high-income households. The coefficient on the indicator variable for access to an employer retirement plan increases from 0.15 for the lowest income quintile to 0.31 for the highest income quintile. On the other hand, the treatment effects for automatic enrollment decrease in income, implying that most high-income households enroll regardless of automatic enrollment. The coefficient on the indicator variable for automatic enrollment decreases from 0.34 for the lowest income quintile to 0.05 for the highest income quintile. Indeed, automatic enrollment has a larger treatment effect for households in the lowest income quintile (i.e., $0.15 + 0.34$) than for those in the highest income quintile (i.e., $0.31 + 0.05$).

5.3. 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. The mandates apply to all employers with a minimum number of workers and minimum years in business.[6] 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. Because the mandates do not require employer contributions, employers incur only the administrative costs of compliance.

Because these mandates are relatively recent, we cannot yet estimate their long-run

[6] In December 2022, the Senate released legislative text that would require retirement plans for all employers with at least ten workers and in business for at least three years. It is too early to tell whether universal access to an employer retirement plan could become federal law.

Table 9
State-Sponsored Retirement Savings Programs

State	Program	Effective year	Default contribution rate (%)	Employers with at least: Workers	Employers with at least: Years in business
California	CalSavers	2020	5	5	
Colorado	SecureSavings	2023	5	5	2
Connecticut	MyCTSavings	2022	3	5	
Illinois	Secure Choice	2018	5	5	2
Maine	MaineSaves	2023	5	5	2
Maryland	MarylandSaves	2022	5	1	2
New Jersey	Secure Choice	2022	3	25	2
New York	Secure Choice	2023	3	10	2
Oregon	OregonSaves	2017	5	5	
Virginia	VirginiaSaves	2023	5	25	2

The features of the state-sponsored retirement savings programs are reported as of December 2022.

impact on retirement account participation. Moreover, these mandates have a limited scope, applying to a subset of employers in the ten participating states. However, we could estimate the counterfactual retirement account participation in 2019 if universal access to an employer retirement plan with automatic enrollment had already been in effect nationally since 2010. For each worker in each year, we define counterfactual access to an employer retirement plan as 1 if she worked (i.e., received a Form W-2) and 0 otherwise. For each household, we then compute the indicator variable for access to an employer retirement plan by either spouse from 2011 to 2019. Finally, we use the instrumental variables regression model in Table 8 to predict the counterfactual probability of participation under automatic enrollment for each household.

Table 10 reports the predicted change in retirement account participation by income quintile. Had universal access to an employer retirement plan with automatic enrollment already been in effect nationally since 2010, retirement account participation in 2019 would have increased by 17 percentage points in the lowest income quintile and 11 percentage points in the second income quintile. In the third to the highest income quintiles, the predicted change declines to 5, 1, and 1 percentage points, respectively. Universal access has the largest impact on low- and middle-income households, which currently have lower access to an employer retirement plan. However, universal access only raises the participation rate to 37 percent in the lowest income quintile. Universal access does not achieve universal participation because not all households are working, and the takeup rate is far less than one.

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	0.21	0.54	0.81	0.93	0.97
Counterfactual	0.37	0.65	0.85	0.95	0.98
Predicted change	0.17	0.11	0.05	0.01	0.01
Observations	7,214,428	7,214,428	7,214,428	7,214,428	7,214,428

In the counterfactual, workers have access to an employer retirement plan with automatic enrollment during all working years. The sample includes all households with a member aged 50 to 59 in the 2019 administrative tax data.

The causal effects in Table 8 are based on employers that voluntarily start retirement plans. Therefore, these causal effects may not have external validity for a policy that mandates universal access through all employers. First, employers that voluntarily start retirement plans may be different from other employers, perhaps viewing retirement plans more positively and nudging workers into participation. Second, universal access to an employer retirement plan could have macroeconomic effects, such as a higher takeup rate from both implicit and explicit government endorsement and greater public awareness. For these reasons, it would be interesting to revisit this question ten years later when we have enough data to evaluate the state-sponsored retirement savings programs.

6. 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.

Access to an employer retirement plan is a primary determinant of retirement account participation. By instrumental variables, we estimate that access to an employer retirement plan increases retirement account participation by 15 percentage points in the lowest income quintile. Automatic enrollment further increases retirement account participation by 34 percentage points for these households. Therefore, universal access to an employer retirement plan could be an effective policy intervention that boosts retirement account participation for low- and middle-income households. Universal access to an employer retirement plan with automatic enrollment could increase retirement account participation by 17 percentage points in the lowest income quintile and 11 percentage points in the second income quintile over ten years.

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 first author's 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.

Our findings have important implications for the Social Security program. First, 79 percent of households in the lowest income quintile without retirement accounts primarily rely on Social Security benefits during retirement. The wellbeing of these households is especially sensitive to Social Security programs that provide a safety net for low-income workers, such as the special minimum benefit and Supplemental Security Income. Policymakers need to be especially careful when considering changes that could cut benefits to these programs. Second, the geographic disparities in retirement account participation imply different degrees of reliance on Social Security benefits across the United States. Policymakers could use our interactive maps to target advertisement and outreach to communities that are in most need of Social Security benefits, making sure that all households that are eligible apply for the benefits. Third, universal access to an employer retirement plan could boost retirement account participation and reduce reliance on Social Security benefits. However, households that are not able to save enough or do not have sufficient work history will continue to rely on Supplemental Security Income as a safety net. Policymakers need to continue evaluating the relevance of these programs in the changing retirement landscape.

References

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

Board of Governors of the Federal Reserve System. 2016–2019. "Survey of Consumer Finances." Https://www.federalreserve.gov/econres/scfindex.htm.

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

Brown, James R., J. Anthony Cookson, and Rawley Z. Heimer. 2019. "Growing Up Without Finance." *Journal of Financial Economics* 134 (3):591–616.

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

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

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

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

Derby, Elena, Kathleen Mackie, and Jacob Mortenson. 2022. "Worker and Spousal Responses to Automatic Enrollment." Working paper, Joint Committee on Taxation.

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

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

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

Federal Deposit Insurance Corporation. 2005–2019. “Annual Survey of Branch Office Deposits.” Https://banks.data.fdic.gov/docs/.

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

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

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

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

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

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

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

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

Stein, Luke C. D. and Constantine Yannelis. 2020. “Financial Inclusion, Human Capital, and Wealth Accumulation: Evidence from the Freedman’s Savings Bank.” *Review of Financial Studies* 33 (11):5333–5377.

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

U.S. Census Bureau. 2005–2019. "American Community Survey Demographic and Housing Five-Year Data." Https://www.census.gov/data/developers/data-sets/acs-5year.html.

U.S. Department of Labor. 2009–2019. "Form 5500 Annual Report." Https://www.dol.gov/agencies/ebsa/about-ebsa/our-activities/public-disclosure/foia/form-5500-datasets.

Appendix A. Administrative Tax Data

A1. Population Count

For each year from 2015 to 2019, we sample all individuals aged 50 to 59 on April 1 (i.e., census day), who have either a tax return or an information return with 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. The ratio of the population count in the administrative tax data to that in the census is slightly greater than one from 2015 to 2019.

Table A1
Comparing the Population Count with the Census

Year	Tax data	Census	Ratio
2015	44,221	43,985	1.01
2016	44,202	43,738	1.01
2017	43,920	43,278	1.01
2018	43,463	42,777	1.02
2019	43,331	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 with a ZIP Code within the U.S. states or Washington, DC. The census population count is the resident population aged 50 to 59.

A potential reason for the slightly higher population count in the administrative tax data is the treatment of part-time U.S. residents and migrant workers. These individuals are in our sample if they receive information returns or file tax returns with a U.S. address, but they are not part of the census population if they do not reside in the United States on April 1. The difference in the population counts is sufficiently small that we are not concerned for the purposes of this paper.

A2. 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, we use the ZCTA-level data to study the relation between geographic characteristics and financial participation. Second, we cannot disclose household-level data, but we can share aggregate data 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 2 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 one that has 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 one that has fewer than 20 households or 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 161,051 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×md_incm/(1+hh2).

Appendix B. Other Data

B1. Department of Labor's Form 5500

We use the Forms 5500 and 5500-SF Annual Reports (U.S. Department of Labor 2009–2019). We focus on the sample of defined contribution plans with active participants at either the beginning of the plan year (e.g., box 6a(1) on the 2019 form) or the end of the plan year (e.g., box 6a(2) on the 2019 form). For each employer, we construct an indicator variable for automatic enrollment if the pension feature code (e.g., box 8a of the 2019 form) is 2S on any of its active plans. To catch cases where the pension feature code is incomplete, we do a textual analysis of the plan description in the actual filing to search for "automatic enrollment", "auto enrollment", or "default enrollment".

As described in Section 2, we construct an indicator variable for access to an employer retirement plan if the box for retirement plan is checked on any Form W-2 issued by the employer. For this subset of employers, we construct an indicator variable for automatic enrollment by merging with the Form 5500 data by employer identification number (EIN) and tax year. If the EIN on the Form W-2 fails to merge with a Form 5500, we use the parent firm's EIN (if available) on Form 851 (Affiliations Schedule) to merge with the Form

5500 data. Thus, we attribute the parent firm's retirement plan to their subsidiaries in cases where the subsidiary's EIN does not match with a Form 5500.

B2. 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 Data (U.S. Census Bureau 2005–2019). 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.

B3. Federal Deposit Insurance Corporation

We count the number of bank branches by ZCTA, based on the Annual Survey of Branch Office Deposits (Federal Deposit Insurance Corporation 2005–2019). We construct bank branch density as the logarithm of the number of branches divided by the population within a ZCTA. We winsorize the right tail at 10 branches per 1,000 residents (about 7 percent of observations) to reduce the impact of outliers. We set bank branch density to zero for ZCTAs without a bank branch.

Appendix C. Additional Results

Table C1
First-Stage Regression for Access to an Employer Retirement Plan

	Percentile of usual income				
Regressor	0–20	20–40	40–60	60–80	80–100
Instrument for					
Access to employer plan	0.14	0.22	0.25	0.29	0.38
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Additional years of access	0.05	0.02	0.01	0.01	0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Automatic enrollment	0.23	0.10	0.06	0.05	0.03
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log income	0.09	0.07	0.01	-0.01	-0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.20	0.52	0.63	0.61	0.53
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	5,837,709	3,538,967	1,709,152	792,476	719,025

The endogenous regressor is an indicator variable for access to an employer retirement plan by either spouse from 2011 to 2019. 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 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 C2
First-Stage Regression for Additional Years of Access to an Employer Retirement Plan

	Percentile of usual income				
Regressor	0–20	20–40	40–60	60–80	80–100
Instrument for					
Access to employer plan	-0.01	0.26	0.37	0.47	0.53
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Additional years of access	0.18	0.26	0.28	0.30	0.41
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Automatic enrollment	0.47	0.21	-0.16	-0.46	-0.58
	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)
Log income	0.13	0.34	0.05	-0.08	-0.08
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.22	1.17	1.74	1.68	1.16
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	5,837,709	3,538,967	1,709,152	792,476	719,025

The endogenous regressor is additional years (beyond a year) that either spouse had access to an employer retirement plan without automatic enrollment from 2011 to 2019. 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 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 C3
First-Stage Regression for Automatic Enrollment in an Employer Retirement Plan

	Percentile of usual income				
Regressor	0–20	20–40	40–60	60–80	80–100
Instrument for	0.00	0.01	0.00	0.00	-0.03
Access to employer plan	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	0.01	0.01	0.01	0.01	0.01
Additional years of access	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
	0.19	0.29	0.41	0.47	0.58
Automatic enrollment	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log income	0.01	0.03	0.01	0.00	-0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.02	0.09	0.15	0.16	0.12
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Observations	5,837,709	3,538,967	1,709,152	792,476	719,025

The endogenous regressor is an indicator variable for automatic enrollment in an employer retirement plan by either spouse from 2011 to 2019. 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 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.

Center for Financial Security

School of Human Ecology
University of Wisconsin-Madison

1300 Linden Drive
Madison, WI 53706

608-890-0229
cfs@mailplus.wisc.edu
cfs.wisc.edu