# Financial Inclusion Across the United States*

Motohiro Yogo† Andrew Whitten‡ Natalie Cox§

September 30, 2021

PRELIMINARY DRAFT – PLEASE DO NOT CITE

**Abstract**

We document new facts about participation in bank and retirement accounts, based on the universe of U.S. households with a member aged 50 to 59 in administrative tax data. Financial participation is much higher than that reported in survey data, especially for low-income households. However, financial participation declines among low-income households from 2008 to 2018. Geographic variation in financial participation relates to income inequality rather than racial segregation or access to financial services. Based on instrumental variables, we estimate the long-run impact of access to employer retirement plans on retirement account participation, which is especially large for low- and middle-income households.

*JEL*: D14, G51

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

*For comments and discussions, we thank Edith Brashares and Neviana Petkova. We also thank seminar participants at the U.S. Department of the Treasury, University of Michigan, and the 2021 Office of Tax Analysis Research Conference. The views expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Department of the Treasury or the National Bureau of Economic Research.

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

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

§Princeton University (nbachas@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 participation in bank and retirement accounts is 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 document new facts about participation in bank and retirement accounts, based on the universe of U.S. households with a member aged 50 to 59 from 2008 to 2018 in administrative tax data (containing virtually all tax returns and information returns).

First, we find that financial participation is much higher than that reported in the Survey of Consumer Finances (SCF), especially for low-income households. In 2016, 82 percent of households in the lowest income quintile had bank accounts in the tax data, compared with 72 percent in the SCF. In the same year, 42 percent of households in the lowest income quintile had retirement accounts in the tax data, compared with 27 percent in the SCF. Throughout the paper, we define retirement accounts comprehensively to include employer retirement plans (both defined benefit and defined contribution plans) and individual retirement arrangements (IRAs). Financial participation is higher in the tax data than in the SCF for the other income quintiles as well, but the gap narrows for high-income households.

Second, we find that financial participation declines for low-income households from 2008 to 2018. In the lowest income quintile, bank account participation declines from 85 percent in 2008 to 79 percent in 2018. In the same income quintile, retirement account participation also declines from 49 percent in 2008 to 41 percent in 2018. The declining retirement account participation mirrors access to employer retirement plans, which declines in the lowest income quintile from 60 percent in 2008 to 50 percent in 2018. Since we fix the age window at 50 to 59, these trends reflect cohort effects in which younger cohorts are less likely to participate than older cohorts at a given age. The fact that financial participation and access to employer retirement plans are declining for low-income households adds to the mounting evidence on growing inequality.

Third, we exploit the large sample to study the geographic determinants of financial participation. In particular, we examine whether the large variation in financial participation across ZIP Code Tabulation Areas (ZCTAs) relates to racial segregation, income inequality, or access to financial services. Bank account participation is negatively correlated with 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, which appears to rule out spatial discrimination based on the geographic distribution of

bank branches. Retirement account participation is also negatively correlated with Hispanic and black population shares. However, the correlation significantly weakens for Hispanics and disappears for blacks conditional on income. The overall evidence suggests that the geographic variation in financial participation is most closely related to income inequality rather than racial segregation or access to financial services.

Fourth, we exploit the panel dimension to estimate the long-run impact of access to employer retirement plans on retirement account participation. Retirement account participation and access to employer retirement plans are jointly endogenous if workers choose employers based on their desire for retirement benefits. Our instrument for access to employer retirement plans is the two-digit NAICS sector code for the first observed employer that issued a Form W-2 to a worker, conditional on that employer not offering a retirement plan. The identifying assumption is that a worker does not choose her sector of employment based on the relative prevalence of employer retirement plans across sectors. Although the worker does not have access to an employer retirement plan through the first observed employer, she is more likely to eventually have access if she stays in a sector where employer retirement plans are more prevalent. We find that access to employer retirement plans (through any employer in the last ten years) increases retirement account participation by 57 percentage points in the lowest income quintile. The average treatment effect diminishes monotonically to 56, 53, 41, and 30 percentage points respectively in the second to the highest income quintiles. The marginal impact of employer retirement plans diminishes for high-income households because they already participate through IRAs or existing employer retirement plans.

The tax data allow us to measure the extensive margin of whether a household has a bank or a retirement account but not the intensive margin of the account balance. The extensive margin is important from the perspective of life-cycle saving for low-income households. In the presence of behavioral biases, participation in a bank account with electronic deposit of wages or a retirement account with automatic contributions could get households in the habit of saving (Mullainathan and Shafir 2009). Savings are important for buffering consumption from income shocks when working or health shocks in retirement. Bank account participation could also lead to more wealth accumulation and purchases of durable goods (Célerier and Matray 2019). By not participating in retirement accounts, households are potentially forgoing valuable tax savings.

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 2016 SCF, the wealth share was 2 percent for the bottom half of households and 34 percent for the top one percent. In comparison, the income share was 13 percent for the bottom half of households

and 24 percent for the top one percent. Therefore, the bottom half of households owns only 3 percent of wealth and earns 17 percent of income in the subpopulation that excludes the top one percent. Because bank and retirement accounts are the most important means of accumulating financial wealth, low participation at the bottom of the income distribution is important for better understanding wealth inequality.

Policymakers try to 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 limited causal effect on retirement contributions. We find that access to employer retirement plans has a large impact on retirement account participation. Thus, tax incentives for employers to offer retirement plans may be more effective than tax incentives for workers to save in retirement plans. This implication of our findings is consistent with the evidence that small nudges (e.g., easy access to a retirement account through an employer) can have large effects on participation (Madrian and Shea 2001; Chetty et al. 2014).

The remainder of the paper is organized as follows. Section I describes how we construct our sample and measures of income and financial participation from the tax data. Section II presents new facts about financial participation. In Section III, we study the geographic determinants of financial participation. In Section IV, we estimate the long-run impact of access to employer retirement plans on retirement account participation. Section V concludes.

## I. Data Construction

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

### *A. Tax Data*

We use administrative tax data of the Internal Revenue Service, which contain tax returns (Form 1040) and information returns for tax years 1999 to 2018. 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 1098 (mortgage interest statement).[1]

*1. Sample*

We sample all individuals aged 50 to 59 from 2008 to 2018, who have a record in the tax data (i.e., either a tax return or an information return) and have a ZIP Code within the U.S. states. 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, our sample includes nearly 40 million individuals, which covers 96 percent of the census population in 2010. Relative to the census, we miss individuals who do not enter the tax data, including undocumented workers and some foreign residents.

For each sampled individual, we also obtain their tax data for the previous nine years. So for a sampled individual in 2008, we obtain their tax data for 1999 to 2008. This nine-year lookback explains why our empirical analysis starts in 2008. We also sample the spouses of sampled individuals (regardless of age), defined as a current joint filer on Form 1040 or a previous joint filer within the last nine years who currently has the same household identifier. The latter criterion ensures that we do not break up households that stop filing taxes. Larrimore et al. (2021) constructed the household identifiers based on a textual analysis of the addresses on tax returns and information returns.

*2. 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 6 on the 2018 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.

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

For nonfilers who do not have a Form 1040, we again follow Larrimore et al. (2021), defining pre-tax individual income as 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 Form 1099-MISC income (assuming 70 percent for offsetting expenses). For nonfilers who form households through the household identifer, we define pre-tax household income as the sum of the pre-tax individual incomes.

To adjust for inflation, we measure income in 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.

### *3. Financial Participation*

We measure bank account participation based on electronic funds transfer for payment of taxes or receipt of refunds on Form 1040. We also measure bank account participation based on taxable (box 1) or tax-exempt (box 8) interest on Form 1099-INT.[2]

We define retirement account participation comprehensively to include employer retirement plans and IRAs. We measure participation in an employer retirement plan if box 13 for retirement plan 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 reports contributions, rollovers, conversions, and recharacterizations during the tax year. We also measure retirement account participation based on a retirement distribution on Form 1099-R. For part of our analysis, we use the distribution codes on Form 1099-R to distinguish an employer retirement plan versus an IRA.

For individuals who do not participate in an employer retirement plan, we measure access to an employer retirement plan if box 13 for retirement plan 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.

Thus, we have constructed three indicator variables for bank account participation, re-

[2]Form 1099-INT has incomplete coverage because it is only required 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.

tirement account participation, and access to employer retirement plans. However, these variables could have gaps in the panel dimension if an individual does not file taxes or receive a particular information return in a given year. Therefore, we use a nine-year lookback to improve these variables. For example, we measure bank account participation in 2008 if the criteria for having a bank account are satisfied in any year between 1999 and 2008. Thus, the precise definition of bank account participation in 2008 is having an account in the last ten years, even if that account is closed as of 2008. 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 that has ever had a bank account is different from one that has never had a bank account.

The lookback procedure implies that the indicator variable for access to employer retirement plans is one if any employer in the last ten years offered a retirement plan. This measurement choice is intentional so that we estimate the long-run impact to employer retirement plans on retirement account participation in Section IV. We cannot measure whether a worker has automatic enrollment in an employer retirement plan, which has a large effect on short-run participation (Madrian and Shea 2001). However, automatic enrollment appears to matter less for long-run participation since the ten-year participation rate under elective enrollment is close to the participation rate under automatic enrollment (Madrian and Shea 2001, Figure 1).

We define indicator variables at the household level by taking the maximum over the spouses' indicator variables for bank account participation, retirement account participation, and access to employer retirement plans. That is, a household has a retirement account if either spouse has a retirement account.

During the initial research design, we considered two other measures of financial participation. Participation in stocks and equity mutual funds would have been interesting. However, the tax returns and information returns do not contain any information about stocks and equity mutual funds in retirement accounts. We could measure mortgage participation based on Form 1098. However, mortgage participation may not be a relevant welfare measure if households have preferences for renting or have paid off their mortgage. In a related paper, Lurie and Pearce (2019) use tax data to study health insurance coverage.

*4. Aggregation*

We use a crosswalk file to map the ZIP Codes to ZIP Code Tabulation Areas (ZCTAs). 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 analysis, we use commuting zone fixed effects. The Census Bureau defines 709 commuting zones by aggregating counties according to local labor markets.

We construct five income groups based on the distribution of usual household income over the entire United States. 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).

As we describe in Appendix A, we construct three aggregate data files from the household-level data. The first data file is aggregated by year and income quintile over the entire United States. The second data file is aggregated by year and ZCTA. The third data file is aggregated by year, ZCTA, and income quintile. The aggregation serves two purposes. First, the aggregate data by ZCTA allows us to study the geographic determinants of financial participation. Second, we cannot disclose household-level data, but we could share aggregate data publicly to facilitate replication and future research. However, we must mask small cells to avoid revealing information about specific households, which we describe in Appendix A.

### *B. 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, and other nonwhite. Other nonwhite includes American Indian, Alaska Native, Native Hawaiian, other Pacific Islander, and multiple race. To control for housing costs, we use the median home value at the ZCTA level from the American Community Survey Selected Housing Characteristics Five-Year Estimates.

### *C. 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 to reduce the impact of outliers.

### *D. Survey of Consumer Finances*

To benchmark our findings to survey evidence, we use the 2010, 2013, and 2016 SCF. We restrict our sample to households with a member aged 50 to 59 in these years. 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 tax data by including realized capital gains, food stamps, and other government support. We construct participation in

bank and retirement accounts, following the same definitions that we use for the tax data. The SCF groups race into white, Hispanic, black, and other nonwhite (including Asian).

## II. Facts about Financial Participation

We present new facts about financial participation based on the tax data. We first show that financial participation in the tax data is much higher than that reported in the SCF, especially for low-income households. We then show that financial participation declines for low-income households from 2008 to 2018. Finally, we replicate known results about the determinants of financial participation in the SCF, which help us interpret the geographic analysis in Section III.

### *A. Comparing Financial Participation with the SCF*

Table 1 reports bank account participation for households with a member aged 50 to 59 in the SCF and the tax data. In 2016, 82 percent of households in the lowest income quintile had bank accounts in the tax data, compared with 72 percent in the SCF. Bank account participation is higher in the tax data than the SCF for the second and the third income quintiles as well. At the highest two income quintiles, participation reaches 100 percent in both data sets. Thus, the SCF appears to underreport bank account participation relative to the tax data for low-income households.

Table 2 reports retirement account participation for households with a member aged 50 to 59 in the SCF and the tax data. In 2016, 42 percent of households in the lowest income quintile had retirement accounts in the tax data, compared with 27 percent in the SCF. Similarly, 75 percent of households in the second income quintile had retirement accounts in the tax data, compared with 55 percent in the SCF. Retirement account participation is higher in the tax data than the SCF for the other income quintiles as well, but the gap narrows for high-income households. Our finding is consistent with Dushi et al. (2011) and Bee and Mitchell (2017), who find that household surveys as part of the census underreport retirement account participation relative to the tax data.

The underreporting of financial participation in the SCF could be due to sampling bias, recall bias, or strategic misreporting. The SCF has sampling bias if noncompliance among potential survey participants is correlated with income or financial participation. Households may not recall that they have a retirement account, especially low-income households that have small accounts or do not contribute regularly. Households may be worried about their privacy and strategically misreport financial participation. None of these potential problems arises in the tax data.

To better understand retirement account participation, Table 3 breaks it down to employer retirement plans versus IRAs for households with a member aged 50 to 59 in 2016. Of the 42 percent of households in the lowest income quintile that have a retirement account, 25 percent have only an employer retirement plan, 6 percent have only an IRA, and 11 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. Of the 99 percent of households in the highest income quintile that have a retirement account, 19 percent have only an employer retirement plan, 3 percent have only an IRA, and 76 percent have both.

### *B. Historical Trends in Financial Participation*

Panel A of Figure 1 reports bank account participation by income quintile from 2008 to 2018. In the lowest income quintile, bank account participation declines from 85 percent in 2008 to 79 percent in 2018.[3] Since we fix the age window, this trend reflects a cohort effect in which younger cohorts are less likely to have a bank account than older cohorts at a given age. For the higher income quintiles, bank account participation is essentially constant over time.

Panel B of Figure 1 reports retirement account participation by income quintile from 2008 to 2018. In the lowest income quintile, retirement account participation declines from 49 percent in 2008 to 41 percent in 2018. In the second income quintile, retirement account participation also declines from 83 percent in 2008 to 74 percent in 2018. That is, younger cohorts are less likely to have a retirement account than older cohorts at a given age.

This trend in retirement account participation mirrors declining access to employer retirement plans in Panel C of Figure 1. In the lowest income quintile, access to employer retirement plans declines from 60 percent in 2008 to 50 percent in 2018. In this income quinitle, retirement account participation in Panel B has declined nearly one-to-one with access to employer retirement plans in Panel C. In the second income quintile, access to employer retirement plans declines from 90 percent in 2008 to 85 percent in 2018. In this income quintile, retirement account participation in Panel B has declined slightly faster than access to employer retirement plans in Panel  due to declining take-up rate for employer retirement plans.

[3]In Appendix A, we discuss a potential measurement issue that the declining bank account participation may be related to a declining tax filing rate.

### *C. Determinants of Financial Participation in the SCF*

To help us interpret the geographical analysis in Section III, we replicate known results about the determinants of financial participation in the SCF (Hogarth et al. 2005). Table 4 reports regressions of participation in bank or retirement accounts on household characteristics for households with a member aged 50 to 59 in the 2010, 2013, and 2016 SCF. The regression is a linear probability model because 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 4, race is a significant determinant of bank account participation. Hispanics and blacks are 13 percentage points less likely than whites to have a bank account. Other nonwhites are 3 percentage points less likely than whites to have a bank account. Controlling for log income in column 2, the coefficients for race decrease in magnitude but remain statistically significant. Hispanics and blacks are respectively 9 and 8 percentage points less likely than whites to have a bank account conditional on income. Hayashi and Minhas (2018) also find that Hispanics and blacks are less likely to have a bank account conditional on income in the 2015 FDIC Survey of Unbanked and Underbanked Households.

Columns 3 and 4 of Table 4 report analogous regressions for retirement account participation. In column 3, race is a significant determinant of retirement account participation. Hispanics and blacks are respectively 33 and 22 percentage points less likely than whites to have a retirement account. Other nonwhites are 16 percentage points less likely than whites to have a retirement account. Controlling for log income in column 4, the coefficients for race decrease in magnitude but remain statistically significant. Hispanics and blacks are respectively 22 and 9 percentage points less likely than whites to have a retirement account conditional on income. Other nonwhites are 11 percentage points less likely than whites to have a retirement account conditional on income.

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. Even if there is no longer explicit discrimination, mistrust of financial institutions could persist as a cultural norm after centuries of discrimination. A second hypothesis is spatial discrimination based on the geographic distribution of bank branches. The cost of accessing banking services may be high for low-income households in nonwhite neighborhoods if there are no bank branches nearby. These hypotheses motivate the geographic analysis in the next section.

## III. Geography of Financial Participation

We first summarize the geographic variation in financial participation across ZCTAs by income quintile. We then examine whether the large variation in financial participation across ZCTAs relates to racial segregation, income inequality, or access to financial services.

### *A. Geographic Variation in Financial Participation*

Figure 2 reports the distribution of bank account participation across ZCTAs by income quintile in 2018, based on the aggregate data by ZCTA and income quintile described in Section I. The distribution essentially has a single peak at one for the fourth and the highest income quintiles. That is, higher-income households have bank accounts regardless of where they live. The distribution not only has a lower mean but also larger variance for the lower income quintiles. In particular, whether households in the lowest income quintile have bank accounts depends to a large extent on where they live.

Figure 3 reports the distribution of retirement account participation across ZCTAs by income quintile in 2018. As with bank account participation, the distribution essentially has a single peak at one for the highest income quintile. In addition, the distribution not only has a lower mean but also larger variance for the lower income quintiles.

Figures 2 and 3 imply that income is an important determinant of financial participation. ZCTAs with higher-income households should have higher financial participation. However, these figures also suggest that other determinants of financial participation could explain the geographic variation within income quintiles. In particular, we explore racial segregation and access to financial services as potential determinants of financial participation.

### *B. Geographic Determinants of Financial Participation*

Table 5 reports regressions of bank account participation on ZCTA-level characteristics, estimated on the aggregate data by year and ZCTA described in Section I. We estimate the panel regression with year fixed effects on a sample that covers 2008 to 2018. We weight the observations by the census-derived household count in each cell (see Appendix A for details).

In column 1 of Table 5, race is a significant determinant of bank account participation. A percentage point increase in the Hispanic and black population shares relates respectively to a 4 and 6 basis point decrease in bank account participation. A percentage point increase in the other nonwhite population share relates to a 9 basis point decrease in bank account participation. In contrast, a percentage point increase in the Asian population share relates to a 4 basis point increase in bank account participation.

In column 2 of Table 5, we add commuting zone fixed effects and find that the $R^2$ increases from 0.27 to 0.48. Moreover, the coefficients for race increase in magnitude for Hispanics and blacks. A percentage point increase in the Hispanic and black population shares relates to a 7 basis point decrease in bank account participation. The coefficient for Asians is zero when we control for commuting zone fixed effects. The remainder of the specifications in Table 5 control for commuting zone fixed effects to emphasize that the correlation between bank account participation and race is about local geographic variation.

In column 3 of Table 5, we add bank branch density at the ZCTA level, which has a precisely estimated coefficient of zero. The coefficients on race are virtually unchanged, implying that the correlation between bank account participation and race is not explained by a low access to banking services in nonwhite neighborhoods. This finding appears to rule out spatial discrimination based on the geographic distribution of bank branches.

In column 4 of Table 5, we add average log income at the ZCTA level, which has a statistically significant coefficient of 0.02. That is, a standard deviation increase in income relates to a 2 percentage point increase in bank account participation. Relative to column 2, the coefficient for Hispanics becomes a precisely estimated zero, and the coefficient for blacks becomes a precisely estimated $-0.01$. Thus, race is not a determinant of bank account participation conditional on income.

Savings ultimately depend on income relative to the cost of living, which could vary across ZCTAs within a communing zone because of housing costs. In column 5 of Table 5, we control for the log home value at the ZCTA level, which has a precisely estimated coefficient of zero. Thus, housing costs do not explain the local geographic variation in bank account participation within commuting zones.

Analogous to Table 5, Table 6 reports regressions of retirement account participation on ZCTA-level characteristics, estimated on the aggregate data by year and ZCTA. In column 1, race is a significant determinant of retirement account participation. A percentage point increase in the Hispanic and black population shares relates respectively to a 31 and 19 basis point decrease in retirement account participation. A percentage point increase in the Asian and other nonwhite population shares relates respectively to a 7 and 14 basis point decrease in retirement account participation.

In column 2 of Table 6, we add commuting zone fixed effects and find that the $R^2$ increases from 0.49 to 0.66. Moreover, the coefficients for race increase in magnitude throughout. A percentage point increase in the Hispanic and black population shares relates respectively to a 38 and 20 basis point decrease in retirement account participation. A percentage point increase in the Asian and other nonwhite population shares relates respectively to a 18 and 16 basis point decrease in retirement account participation. The remainder of the specifications

in Table 6 control for commuting zone fixed effects to emphasize that the correlation between retirement account participation and race is about local geographic variation.

In column 3 of Table 6, we add average log income at the ZCTA level, which has a statistically significant coefficient of 0.08. That is, a standard deviation increase in income relates to a 8 percentage point increase in retirement account participation. Relative to column 2, the coefficient for blacks becomes a precisely estimated 0.01, and the coefficient for other nonwhites becomes a precisely estimated 0.02. In addition, the coefficient for Hispanics decreases in magnitude from $-0.38$ to $-0.14$, and the coefficient for Asians decreases in magnitude from $-0.18$ to $-0.15$. Thus, race is a much less important determinant of retirement account participation conditional on income.

In column 4 of Table 6, we control for the log home value at the ZCTA level, which has a statistically significant coefficient of $-0.03$. That is, a standard deviation increase in home values relates to a 3 percentage point decrease in retirement account participation. The sign of the coefficient implies that households that live in ZCTAs with higher housing costs are less likely to participate in retirement accounts. However, the coefficients for race and average log income do not change much relative to column 3.

### *C. Interpreting Income*

At face value, income is the primary determinant 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 keeping a bank account. Low-income households may not have sufficient tax incentives to open a retirement account. 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 the U.S. states. Thus, financial participation could be correlated with income partly through financial literacy. Tables 5 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 may have no need for an electronic funds transfer to pay taxes or receive refunds. Under this hypothesis, small nudges such as an electronic deposit of wages as a default option could boost financial 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 bank and retirement accounts. Future research could investigate whether peer effects partly explain the correlation between financial participation and income by collecting a large sample of social linkages within ZCTAs.

### *D. Comparison with the SCF*

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

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 that income is mismeasured 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 $Z_n = \sum_{i=1}^{I} z_{i,n}/I$ be the black population share 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 $\beta$ represents the individual effect of income on participation. The coefficients $\gamma$ and $\Gamma$ represent respectively the individual and group effects of race on participation. Based on Table 4 for the SCF, we assume that $\beta > 0$ and $\gamma < 0$. We cannot estimate $\Gamma$ 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 $Y_n = \sum_{i=1}^{I} y_{i,n}/I$, $X_n = \sum_{i=1}^{I} x_{i,n}/I$, and $\Omega_n = \sum_{i=1}^{I} \omega_{i,n}/I$. 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 5 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 4 of Table 5) and the SCF (i.e., column 2 of Table 4). 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.

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 Hispanics and blacks in Table 4 for the SCF would be downward biased (i.e., more negative). This bias is significantly reduced in Tables 5 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. Impact of Employer Retirement Plans

As reported in Table 3, most households have retirement accounts through employer retirement plans rather than IRAs. Opening an IRA requires a higher level of effort and financial knowledge than enrolling in an employer retirement plan. Thus, access to employer retirement plans could be a primary determinant of retirement account participation, especially for low-income households. In this section, we estimate the impact of access to employer retirement plans on retirement account participation.

### *A. Identifying Assumptions*

Retirement account participation and access to employer retirement plans could be jointly endogenous. Workers who care about retirement saving 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 address the endogeneity problem through instrumental variables probit, where retirement account participation is the outcome variable and access to employer retirement plans is the endogenous regressor. As we discussed in Section I, the indicator variable for access to employer retirement plans is one if any employer in the last ten years offered a retirement plan. Thus, we estimate the long-run impact to access to employer retirement plans over the last ten years on current retirement account participation.

To construct the instruments, we start with a sample of individuals aged 50 to 59 in 2018 and pull their first available Form W-2 in 1999.[4] We focus on the sub-sample of individuals who did not have access to an employer retirement plan in 1999. Our instruments are the indicator variables corresponding to the two-digit NAICS sector of their employer in 1999. The identifying assumption is that a worker does not choose her sector of employment based on the relative prevalence of employer retirement plans across sectors. However, the worker may choose her employer based on whether it offers a retirement plan, which is the reason for conditioning on the sub-sample of individuals who did not have access through the first observed employer.

Two conditions must be satisfied for the instruments to be relevant. First, the prevalence of employer retirement plans must be sufficiently different across sectors. Second, the first observed sector of employment must be sufficiently predictive of the actual sector in which the worker eventually earns retirement benefits. Although the worker does not have access to an employer retirement plan in 1999, she is more likely to eventually have access if she stays in a sector where employer retirement plans are more prevalent. Thus, the strength of the treatment depends on the first observed sector of employment, which we assume to be exogenous.

We note three implementation choices in constructing the instruments. First, Form W-2 in 1999 for a small share of workers (about 5 percent) is issued by an employer with a missing NAICS code. We assign these workers to a separate "unknown" sector. Second, we assign individuals without a Form W-2 in 1999 to a random sector based on the empirical distribution of workers across sectors conditional on income quintile. The randomness ensures that identification does not come from individuals who choose to stay out of the labor market. Third, if the sectors are different in a two-member household, we randomly draw one of the sectors and assign it to the household.

[4]Ideally, we would go back further, but the digitized data are not fully available before 1999. For individuals aged 50 to 59 in 2018, the first Form W-2 that we observe is at age 31 to 40 in 1999.

### *B. Average Treatment Effects of Employer Retirement Plans*

We estimate the instrumental variables probit model separately by income quintile to allow for heterogenous treatment effects. As we discussed above, we focus on the sub-sample of households that did not have access to an employer retirement plan in 1999. Therefore, the sample size is smaller for higher income quintiles because high-income households are more likely to have had access to an employer retirement plan in 1999. For completeness, Table B1 in Appendix B reports the results for the complement sub-sample of households that had access to an employer retirement plan in 1999.

Table 7 reports the estimated coefficients for the instrumental variables probit model. The take-up rate of employer retirement plans determines the coefficient on access to employer retirement plans. Participation through an IRA or an existing employer retirement plan before the last ten years determines the constant in the model. The coefficients on access to employer retirement plans are positive and significant for all income quintiles.

To translate these coefficients to economically interpretable magnitudes, Table 8 reports the average treatment effects by income quintile. The first row reports the average predicted probability conditional on access to employer retirement plans. The second row reports the average predicted probability conditional on no access. The average treatment effect is the difference between these rows. In the lowest income quintile, access to employer retirement plans increases retirement account participation by 57 percentage points. The average treatment effect diminishes monotonically to 56, 53, 41, and 30 percentage points respectively in the second to the highest income quintiles. The marginal impact of employer retirement plans diminishes for high-income households because they already participate through IRAs or existing employer plans before the last ten years.

A number of states including California, Colorado, Connecticut, Illinois, Maryland, New Jersey, Oregon, and Virginia mandate that employers enroll workers in a state retirement plan (structured as an IRA with payroll deduction) if they do not already offer a retirement plan. These states essentially make access to employer retirement plans universal for all workers. Chalmers et al. (2021) find a participation rate of 34 percent for OregonSaves (i.e., Oregon's retirement plan). Although this number is in the ballpark of the average treatment effects in Table 8, they are not directly comparable for two reasons. First, some workers may already have retirement accounts before enrollment in OregonSaves, so 34 percent is not the marginal impact of employer retirement plans. Second, OregonSaves is a relatively recent program that started in 2018, so its long-run impact over ten years is likely to be higher.

Although it is too early to tell, universal access to employer retirement plans could become federal law. In September 2021, the House Ways and Means Committee released legislative text that would mandate employer retirement plans for all businesses that are

at least two-years old and have more than five workers. The average treatment effects in Table 8 inform us about the potential long-run impact of such legislation for retirement account participation.

## V. Conclusion

Based on a sample of all U.S. households with a member aged 50 to 59 in administrative tax data, we find that participation in bank and retirement accounts is much higher than that reported in the SCF, especially for low-income households. However, a worrying trend is that financial participation and access to employer retirement plans are declining for low-income households, adding to the mounting evidence on growing inequality. The geographic variation in financial participation is more closely related to income inequality rather than racial segregation or access to financial services. Finally, we find that access to employer retirement plans is a primary determinant of retirement account participation, especially for low-income households. Thus, policies that improve access to employer retirement plans could create more opportunities for low-income households to save for retirement.

In hope of improving financial participation for low-income households, we have constructed interactive maps of financial participation for the lowest income quintile. Figure 4 is a snapshot of the interactive map for bank accounts. The colors range from yellow (90–100 percent participation) to red (0–60 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 5 is a snapshot of the interactive map for retirement accounts. 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.

## References


**Bee, Adam and Joshua Mitchell**, "Do Older Americans Have More Income Than We Think?," *Proceedings: Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, 2017, *110*, 1–85.

**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.

**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.

**Dushi, Irena, Howard M. Iams, and Jules Lichtenstein**, "Assessment of Retirement Plan Coverage by Firm Size, Using W-2 Tax Records," *Social Security Bulletin*, 2011, *71* (2), 53–65.

**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.

**Ramnath, Shanthi P. and Patricia K. Tong**, "The Persistent Reduction in Poverty from Filing a Tax Return," *American Economic Journal: Economic Policy*, 2017, *9* (4), 367–394.

Table 1
Comparing Bank Account Participation

| | Percentile of usual income | | | | | |
|---|---|---|---|---|---|---|
| Year | 0–20 | 20–40 | 40–60 | 60–80 | 80–100 | All |
| *Panel A. SCF* | | | | | | |
| 2010 | 0.74 | 0.90 | 0.96 | 0.99 | 1.00 | 0.94 |
| 2013 | 0.73 | 0.83 | 0.92 | 0.99 | 1.00 | 0.92 |
| 2016 | 0.72 | 0.91 | 0.96 | 1.00 | 1.00 | 0.93 |
| *Panel B. Tax data* | | | | | | |
| 2010 | 0.86 | 0.97 | 0.99 | 1.00 | 1.00 | 0.96 |
| 2013 | 0.84 | 0.97 | 0.99 | 1.00 | 1.00 | 0.96 |
| 2016 | 0.82 | 0.96 | 0.99 | 1.00 | 1.00 | 0.95 |

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

Table 2
Comparing Retirement Account Participation

| | Percentile of usual income | | | | | |
|---|---|---|---|---|---|---|
| Year | 0–20 | 20–40 | 40–60 | 60–80 | 80–100 | All |
| *Panel A. SCF* | | | | | | |
| 2010 | 0.27 | 0.58 | 0.70 | 0.86 | 0.92 | 0.72 |
| 2013 | 0.25 | 0.52 | 0.69 | 0.88 | 0.92 | 0.70 |
| 2016 | 0.27 | 0.55 | 0.74 | 0.90 | 0.93 | 0.71 |
| *Panel B. Tax data* | | | | | | |
| 2010 | 0.47 | 0.81 | 0.94 | 0.98 | 0.99 | 0.84 |
| 2013 | 0.44 | 0.77 | 0.93 | 0.98 | 0.99 | 0.82 |
| 2016 | 0.42 | 0.75 | 0.91 | 0.97 | 0.99 | 0.81 |

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 3
Breakdown of Retirement Account Participation

| | Percentile of usual income | | | | | |
|---|---|---|---|---|---|---|
| Participation in | 0–20 | 20–40 | 40–60 | 60–80 | 80–100 | All |
| Employer plan only | 0.25 | 0.45 | 0.46 | 0.35 | 0.19 | 0.34 |
| IRA only | 0.06 | 0.05 | 0.04 | 0.03 | 0.03 | 0.04 |
| Both | 0.11 | 0.25 | 0.42 | 0.59 | 0.76 | 0.43 |
| Total | 0.42 | 0.75 | 0.91 | 0.97 | 0.99 | 0.81 |

This table reports a breakdown of retirement account participation for households with a member aged 50 to 59 in 2018.

Table 4
Determinants of Financial Participation in the SCF

| | Bank | | Retirement | |
|---|---|---|---|---|
| Regressor | (1) | (2) | (3) | (4) |
| Race: | | | | |
| Hispanic | -0.13 | -0.09 | -0.33 | -0.22 |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Black | -0.13 | -0.08 | -0.22 | -0.09 |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Other nonwhite | -0.03 | -0.01 | -0.16 | -0.11 |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Log income | | 0.08 | | 0.20 |
| | | (0.00) | | (0.00) |
| Constant | 0.97 | 0.96 | 0.79 | 0.76 |
| | (0.00) | (0.00) | (0.01) | (0.01) |
| $R^2$ | 0.04 | 0.12 | 0.06 | 0.24 |
| Observations | 21,631 | 21,631 | 21,631 | 21,631 |

This table reports regressions of participation in bank or retirement accounts 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 2010, 2013, and 2016 SCF.

Table 5
Geographic Determinants of Bank Account Participation

| Regressor | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Race: | | | | | |
| Hispanic | -0.04 | -0.07 | -0.07 | 0.00 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Black | -0.06 | -0.07 | -0.07 | -0.01 | -0.01 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Asian | 0.04 | 0.00 | 0.00 | 0.01 | 0.02 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Other nonwhite | -0.09 | -0.07 | -0.08 | -0.02 | -0.02 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Bank branch density | | | 0.00 | | |
| | | | (0.00) | | |
| Average log income | | | | 0.02 | 0.02 |
| | | | | (0.00) | (0.00) |
| Log home value | | | | | 0.00 |
| | | | | | (0.00) |
| Constant | 0.97 | 0.98 | 0.98 | 0.96 | 0.96 |
| | (0.00) | (0.00) | (0.00) | (0.00) | (0.00) |
| Commuting zone fixed effects | | Y | Y | Y | Y |
| $R^2$ | 0.27 | 0.48 | 0.48 | 0.73 | 0.73 |
| Observations | 350,156 | 350,156 | 350,156 | 350,156 | 341,100 |

This table reports panel regressions of bank account participation on ZCTA-level characteristics, estimated on aggregate data by year and ZCTA. The coefficients for bank branch density, average log income, and log home value 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 2008 to 2018 tax data. The observations are weighted by the census-derived household count in each cell.

Table 6
Geographic Determinants of Retirement Account Participation

| Regressor | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Race: | | | | |
| Hispanic | -0.31 | -0.38 | -0.14 | -0.14 |
| | (0.01) | (0.01) | (0.00) | (0.00) |
| Black | -0.19 | -0.20 | 0.01 | 0.00 |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Asian | -0.07 | -0.18 | -0.15 | -0.12 |
| | (0.02) | (0.02) | (0.01) | (0.01) |
| Other nonwhite | -0.14 | -0.16 | 0.02 | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Average log income | | | 0.07 | 0.08 |
| | | | (0.00) | (0.00) |
| Log home value | | | | -0.03 |
| | | | | (0.00) |
| Constant | 0.92 | 0.94 | 0.87 | 0.87 |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Commuting zone fixed effects | | Y | Y | Y |
| $R^2$ | 0.49 | 0.66 | 0.88 | 0.90 |
| Observations | 350,156 | 350,156 | 350,156 | 341,100 |

This table reports panel regressions of retirement account participation on ZCTA-level characteristics, estimated on aggregate data by year and ZCTA. The coefficients for average log income and log home value 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 2008 to 2018 tax data. The observations are weighted by the census-derived household count in each cell.

Table 7
Instrumental Variables Probit for Retirement Account Participation

| | Percentile of usual income | | | | |
|---|---|---|---|---|---|
| Regressor | 0–20 | 20–40 | 40–60 | 60–80 | 80–100 |
| Access to employer plans | 1.08 | 1.89 | 2.00 | 1.49 | 1.23 |
| | (0.07) | (0.04) | (0.05) | (0.06) | (0.04) |
| Log income | -0.04 | 0.14 | 0.14 | 0.13 | -0.03 |
| | (0.01) | (0.00) | (0.00) | (0.00) | (0.00) |
| Constant | -1.07 | -1.09 | -0.70 | 0.27 | 0.74 |
| | (0.04) | (0.03) | (0.04) | (0.05) | (0.03) |
| Observations | 4,699,009 | 3,399,980 | 2,275,572 | 1,467,732 | 1,268,825 |

The instruments for access to employer retirement plans are indicator variables corresponding to the two-digit NAICS sector code of the employer that issued a Form W-2 in 1999. 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 2018 tax data, who did not have access to an employer retirement plan in 1999.

Table 8
Average Treatment Effects of Employer Retirement Plans

| Participation rate | Percentile of usual income | | | | |
|---|---|---|---|---|---|
| conditional on | 0–20 | 20–40 | 40–60 | 60–80 | 80–100 |
| Access to employer plans | 0.65 | 0.76 | 0.89 | 0.96 | 0.98 |
| No access | 0.08 | 0.19 | 0.36 | 0.55 | 0.68 |
| Average treatment effect | 0.57 | 0.56 | 0.53 | 0.41 | 0.30 |
| Observations | 4,699,009 | 3,399,980 | 2,275,572 | 1,467,732 | 1,268,825 |

The average treatment effect is the difference in the average predicted probability conditional on access to employer retirement plans versus no access. The sample includes all households with a member aged 50 to 59 in the 2018 tax data, who did not have access to an employer retirement plan in 1999.

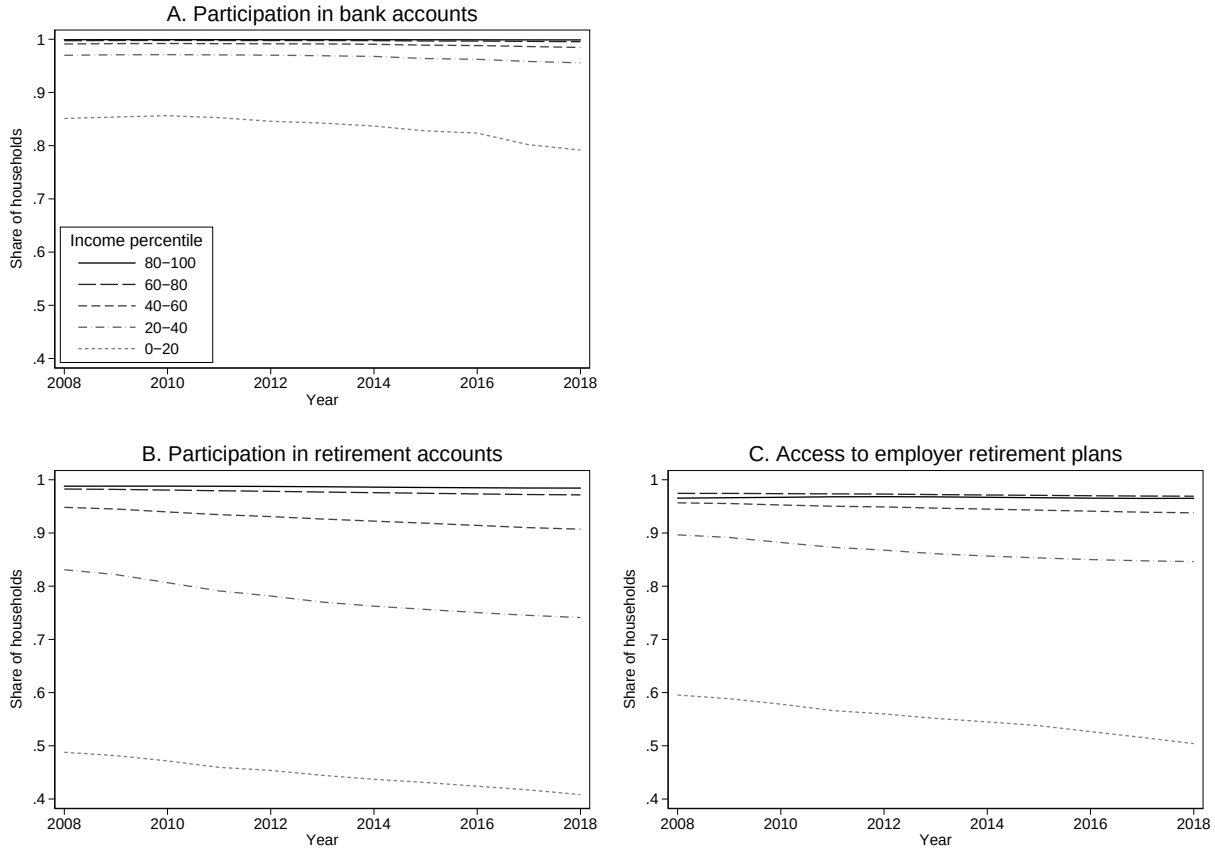



Figure 1. Financial Participation by Income Quintile

Retirement accounts include employer retirement plans (i.e., defined benefit and defined contribution plans) and IRAs. The sample includes all households with a member aged 50 to 59 in the 2008 to 2018 tax data.

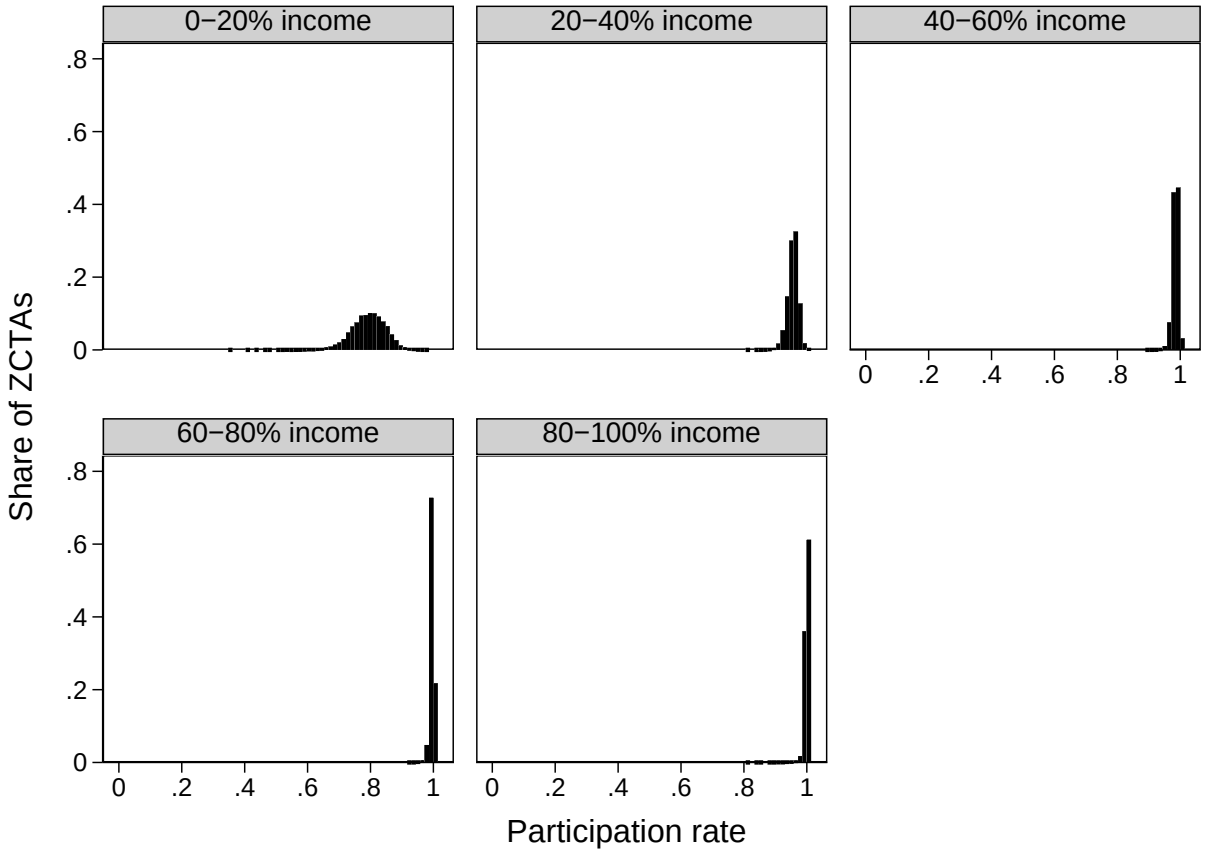



Figure 2. Geographic Variation in Bank Account Participation

This figure reports the distribution of bank account participation across ZCTAs by income quintile in 2018. The sample includes ZCTAs with population of at least 1,000 at age 50 to 59.

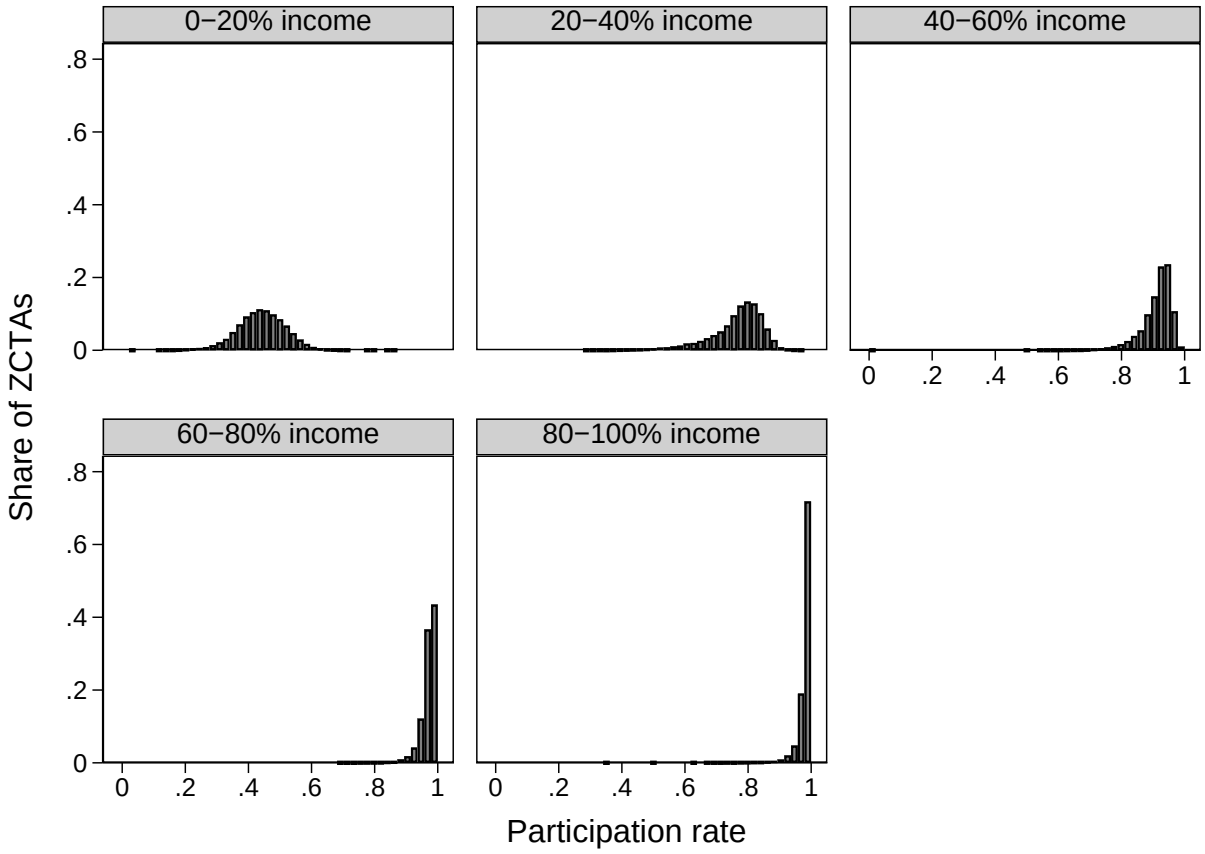



Figure 3. Geographic Variation in Retirement Account Participation

This figure reports the distribution of retirement account participation across ZCTAs by income quintile in 2018. Retirement accounts include employer retirement plans (i.e., defined benefit and defined contribution plans) and IRAs. The sample includes ZCTAs with population of at least 1,000 at age 50 to 59.

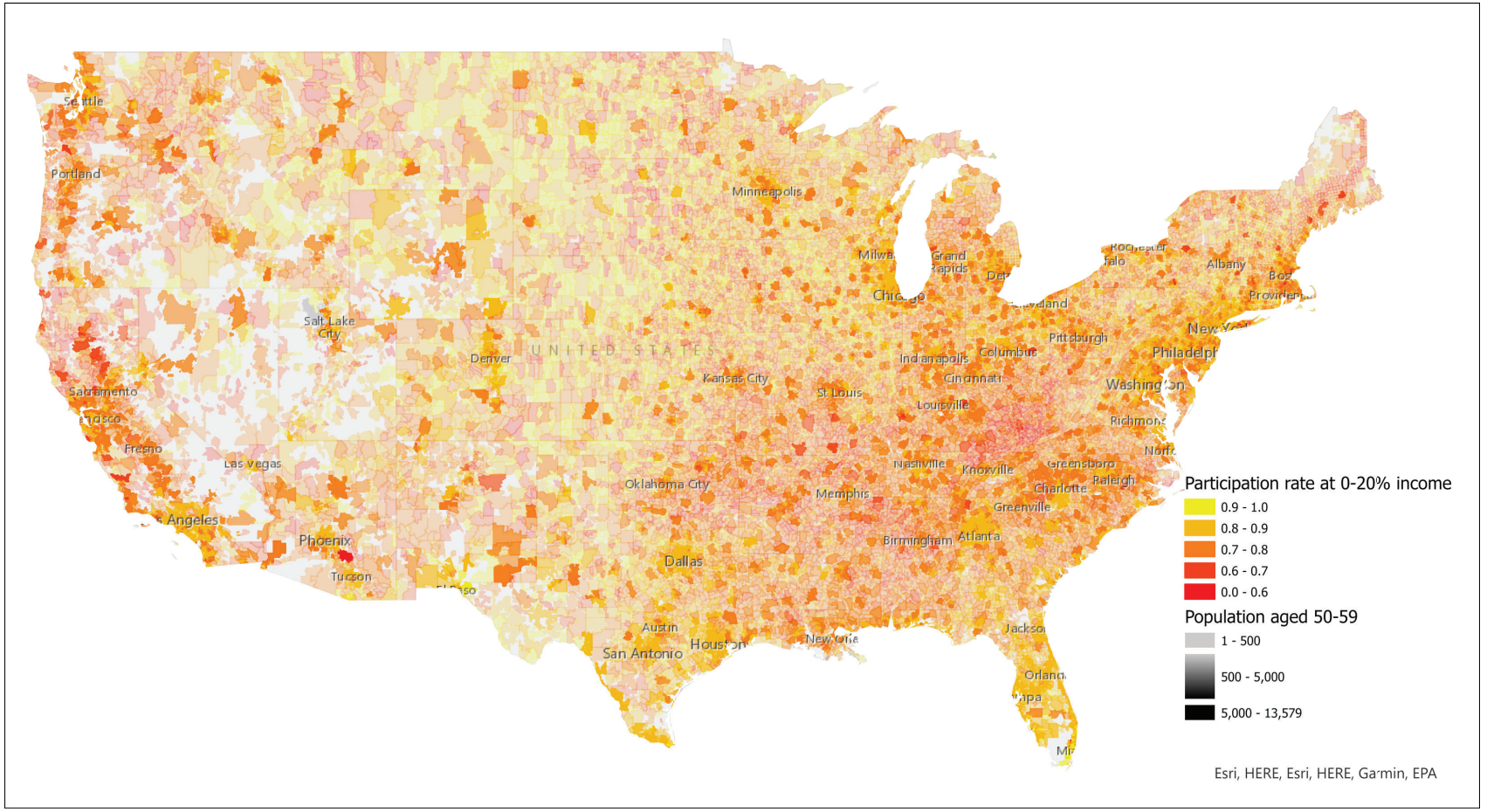



Figure 4. Bank Account Participation at 0–20 Percentile Income

The sample includes all households in the lowest income quintile with a member aged 50 to 59 in the 2018 tax data.

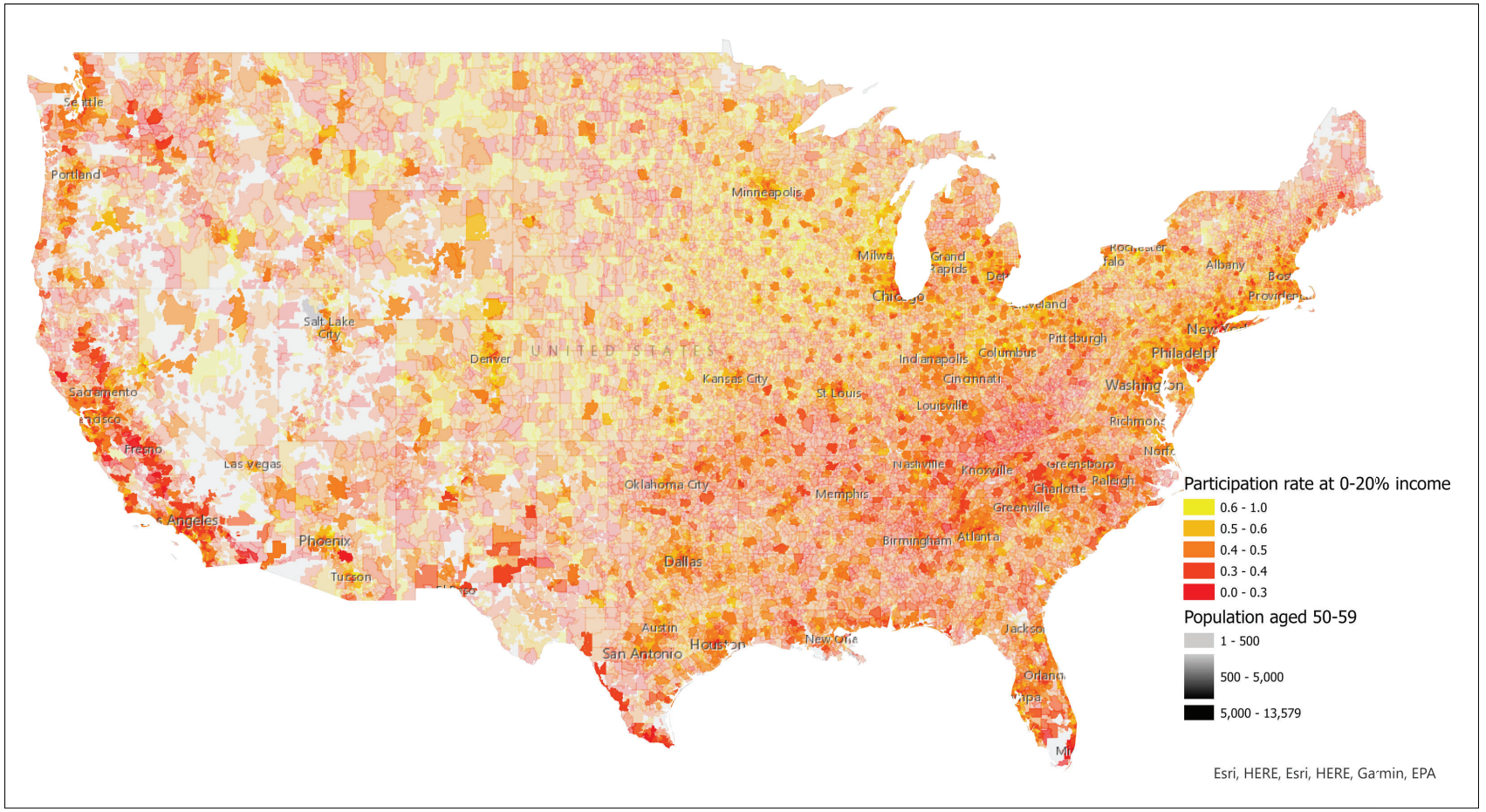



Figure 5. Retirement Account Participation at 0–20 Percentile Income

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 2018 tax data.

## Appendix A. Data Construction

### *A. Sample*

We sample all individuals aged 50 to 59 from 2008 to 2018, who have either a tax return or an information return and have a ZIP Code within the U.S. states. Table A1 compares the population count in the tax data with that in the census. Our sample includes nearly 40 million individuals, which covers 96 percent of the census population in 2010. Coverage, which is the ratio of the population count in the tax data to that in the census, is stable between 94 and 97 percent across years.

Table A1
Comparing the Population Count with the Census

| Year | Census | Tax data | Coverage |
|---|---|---|---|
| 2008 | 39,605 | 37,999 | 0.96 |
| 2009 | 40,409 | 38,676 | 0.96 |
| 2010 | 41,032 | 39,480 | 0.96 |
| 2011 | 41,519 | 40,274 | 0.97 |
| 2012 | 42,842 | 40,700 | 0.95 |
| 2013 | 43,266 | 41,006 | 0.95 |
| 2014 | 43,464 | 41,171 | 0.95 |
| 2015 | 43,652 | 41,123 | 0.94 |
| 2016 | 43,722 | 40,892 | 0.94 |
| 2017 | 43,201 | 40,527 | 0.94 |
| 2018 | 42,521 | 40,246 | 0.95 |

The population counts are reported in thousands. The census population count is the civilian non-institutionalized population at age 50 to 59 in the Current Population Survey. The population count in the 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. Coverage is the ratio of the population count in the tax data to that in the census.

### *B. Tax Filing*

Panel A of Figure A1 reports the share of households filing a tax return (Form 1040) by income quintile from 2008 to 2018. The tax filing rate is declining for all households, and it is declining faster for lower-income households. In the lowest income quintile, the tax filing rate declines from 59 percent in 2008 to 47 percent in 2018. Panel B reports the share of households that have ever filed a tax return in the last ten years by income quintile from 2008 to 2018. In the lowest income quintile, the ever-filing rate declines from 92 percent in 2008 to 89 percent in 2016. The ever-filing rate drops to 82 percent in 2017 because the filing rate was unusually high in 2007 in response Economic Stimulus Act of 2008 (Ramnath and Tong

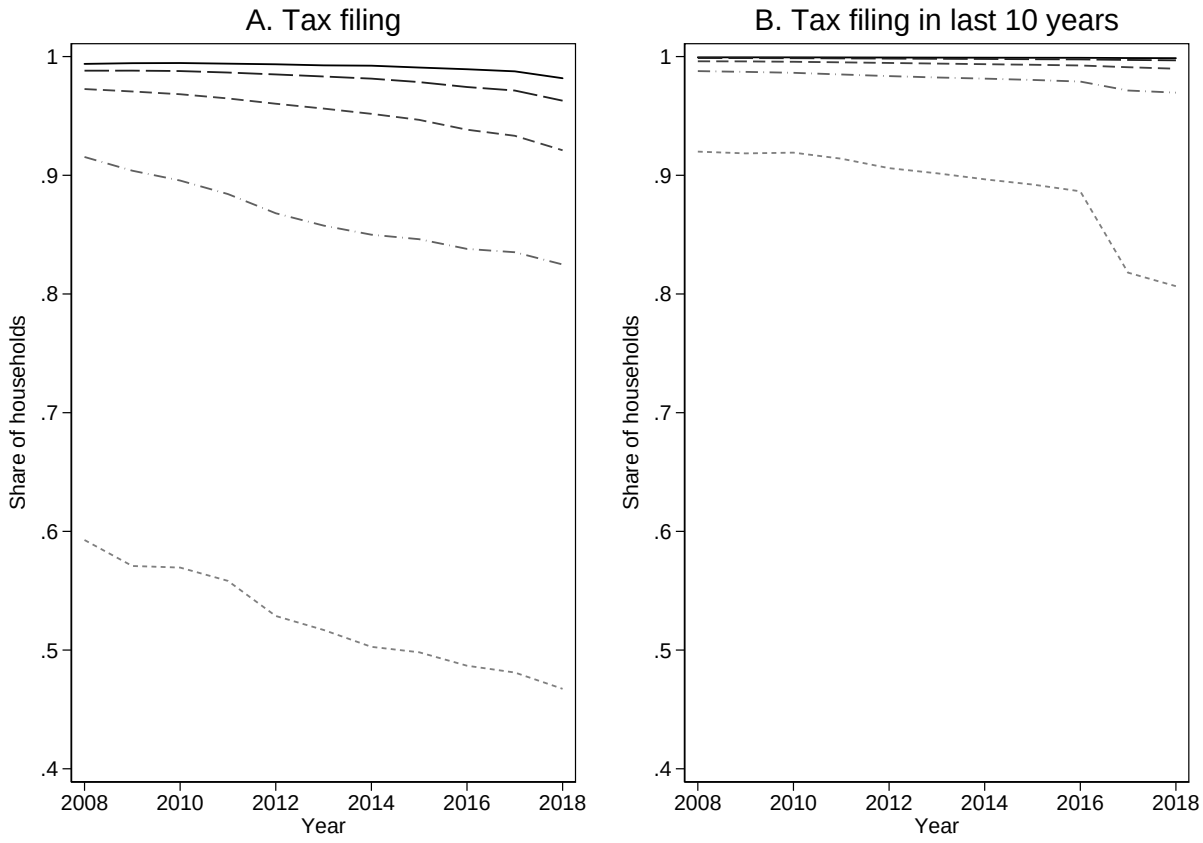



Figure A1. Tax Filing Rate by Income Quintile
The sample includes all households with a member aged 50 to 59 in the 2008 to 2018 tax data.

2017). The high filing rate in 2007 increases the ever-filing rate from 2008 to 2016 through the nine-year lookback.

As we described in Section I, we measure bank account participation primarily through electronic funds transfer as part of a tax filing. Therefore, one concern in interpreting Panel A of Figure 1 is that the apparent decline in bank account participation in the lowest income quintile may reflect the declining ever-filing rate. Therefore, we cannot rule out an alternative hypothesis that bank account participation is constant over time in the lowest income quintile, and we fail to catch households that have a bank account but did not file a tax return over the last ten years. This concern does not apply to the decline in retirement account participation, which we measure entirely from information returns. Although not a focus of this paper, the declining ever-filing rate could be another symptom of worsening financial inclusion if nonfilers are foregoing refunds due. Further research is necessary to determine the causes and consequences of the declining tax filing.

### *C. Aggregation*

We construct three aggregate data files from the household-level data.

- Data_incm: Aggregated by year and income quintile over the entire United States.
- Data_ZCTA: Aggregated by year and ZCTA.
- Data_ZCTA_incm: Aggregated by year, ZCTA, and income quintile.

For each of these data files, 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 before 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 less 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_filer: Share of households that filed a tax return.
- d_filerL: Share of households that filed a tax return in the last ten years.
- 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 less 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 29 percent of the 32,870 cells in 2018 are subject to masking.

For Data_ZCTA_incm, we apply the 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 less than 20 households or (b) is nested in a {year, ZCTA} couplet with less 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. Around 32 percent of the 160,810 cells in 2018 are subject to masking.

For results that require household counts, we estimate it 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. Additional Results

TABLE B1
AVERAGE TREATMENT EFFECTS FOR HOUSEHOLDS WITH ACCESS TO AN EMPLOYER RETIREMENT PLAN IN 1999

| Participation rate conditional on | Percentile of usual income | | | | |
|---|---|---|---|---|---|
| | 0–20 | 20–40 | 40–60 | 60–80 | 80–100 |
| Access to employer plans | 0.81 | 0.90 | 0.97 | 0.99 | 1.00 |
| No access | 0.16 | 0.32 | 0.51 | 0.69 | 0.81 |
| Average treatment effect | 0.65 | 0.58 | 0.45 | 0.30 | 0.18 |
| Observations | 2,040,243 | 3,339,272 | 4,463,682 | 5,271,518 | 5,470,426 |

The average treatment effect is the difference in the average predicted probability conditional on access versus no access to employer retirement plans. The sample includes all households with a member aged 50 to 59 in the 2018 tax data, who had access to an employer retirement plan in 1999.