Stocks for the Long Run or Liquidity? Tax Data Evidence and Portfolio Choice Implications*

November 18, 2025

Abstract

Do temporary stock price crashes matter for long-term investors? I use over 25 years of U.S. income tax data to characterize the savings behavior and risk exposures of high-income working-age households. Aggregate stock price crashes coincide with persistent declines in wage and private business income for many of these households, who take large drawdowns from their liquid assets – including stocks – in response. I develop a life-cycle model with consumption adjustment frictions to match this observed savings behavior and determine its portfolio choice implications. Investing in stocks is risky when falling income and rigid expenditures may force investors to liquidate their holdings at temporarily-depressed prices, resulting in low optimal portfolio shares. These results challenge the conventional wisdom that the stock market is relatively safe for long-term investors.

*Contact information and acknowledgements have been removed to preserve author anonymity. The U.S. Census Bureau has not reviewed the paper for accuracy or reliability and does not endorse its contents. Any conclusions expressed herein are those of the author and do not represent the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Data Management System (DMS) number: P-7503840, Disclosure Review Board (DRB) approval numbers: CBDRB-FY24-SEHSD003-090, CBDRB-FY25-SEHSD003-121).

1 Introduction

What makes stocks risky for individual investors? Aggregate stock prices are volatile and subject to infrequent but large crashes. However, one of the key facts in empirical asset pricing is that these fluctuations are partly transitory. Following large crashes, stock prices tend to recover to their original pre-crash level within only a few years. More formally, news about expected future returns accounts for a large share of the volatility in realized aggregate stock returns over relatively short holding periods. These returns exhibit negative serial correlations at longer time horizons, making stock returns less risky over long holding periods than their volatile short-term returns might suggest.[1] These facts underlie the conventional wisdom — espoused in popular media such as the book *Stocks for the Long Run* (Siegel, 1994) — that stocks are a relatively safe investment for individuals with long investment horizons, so long as they can avoid selling through short-lived price crashes. In this paper I use over 25 years of individual income tax returns to empirically evaluate whether investors actually do so in practice, then critically evaluate such portfolio choice prescriptions in light of this evidence.

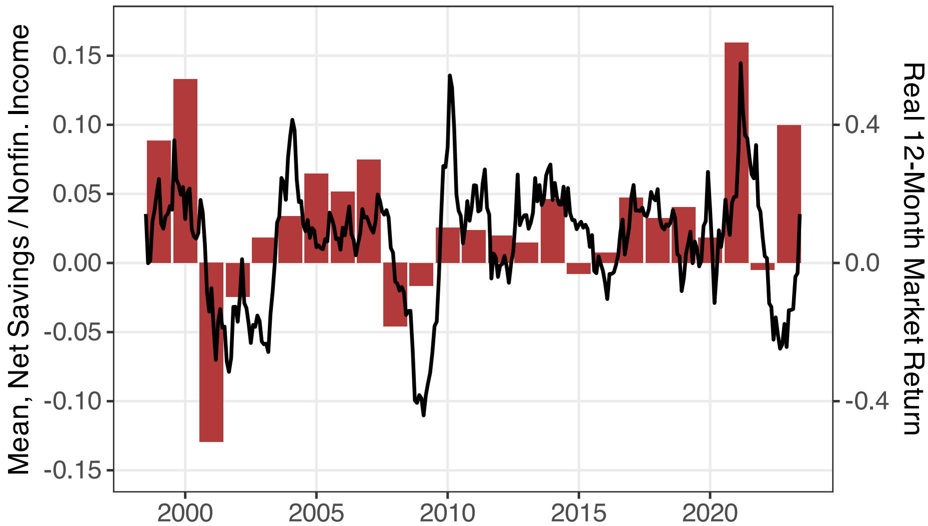
I study the investment decisions and risk exposures of high-income working-age households, in the top fifth of the wage and private business income distribution within their age group. Stocks serve as an important long-run investment vehicle for these investors as they save for retirement, with many accumulating significant holdings in both tax-advantaged retirement accounts (e.g. 401(k)s and IRAs) and taxable brokerage accounts. These taxable stockholdings also represent a large share of many investors' available liquid assets that they can draw down on in a time of financial need, without facing high tax penalties or transactions costs. The income tax data report the dividend and interest income that these investors receive each year from these liquid stock and fixed income investments. I use capitalization methods to estimate these investors' active net savings flows in each asset class from changes in their dividend and interest income across consecutive tax years. These methods estimate the dollar value of investors' stock sales from the observed decline in their dividend income, relative to the dividends paid by the aggregate stock market over the same period. Figure 1 depicts the average estimated flow into stocks among these investors in each year, measured as a fraction of each household's wage and private business income over the prior three years.

These households' average flows into and out of the stock market are strongly procyclical. Panel (a) shows these average net savings flows (red bars) alongside realized returns on the aggregate stock market over the same period (black line). The large stock price crashes occurring around the time of the early 2000s recession and the Global Financial Crisis coincide with large estimated outflows from stocks in investors' taxable accounts, in contrast with steady inflows during expansion periods. These crashes also coincide with sharp declines in average wage and private business income for these investors, shown in Panel (b). Even the

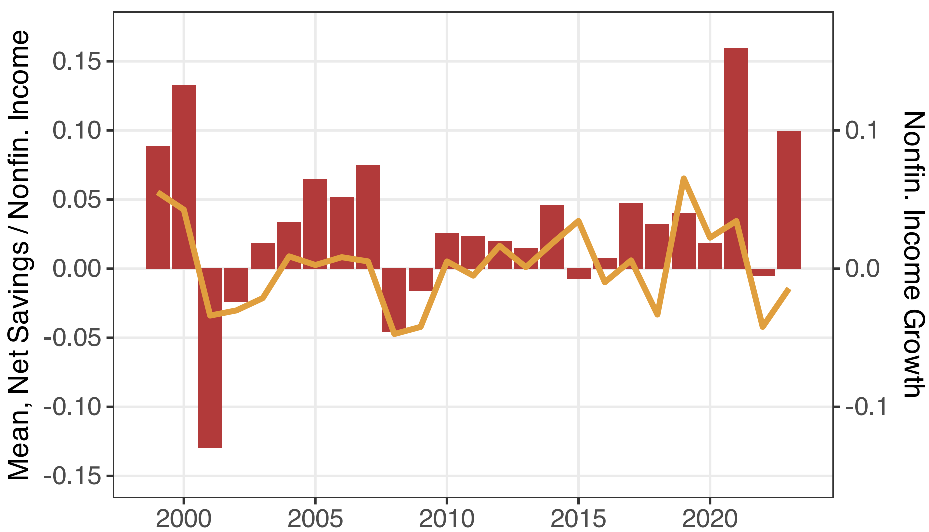
[1] See, for example, Campbell and Shiller (1988) and Cochrane (2011). Lustig and Verdelhan (2012) and Kroencke (2022) provide additional evidence about return predictability over the business cycle. Campbell, Giglio, and Polk (2013) and Martin (2017) provide additional evidence on return predictability over the sample period studied in this paper.

Figure 1: Average Estimated Net Savings Flows into Stock

(a) with Realized Stock Returns (line, right axis)



(b) with Nonfinancial Income Growth (line, right axis)



Notes: This figure depicts mean net savings flows in stocks (bars), aggregate stock returns (line, top panel), and changes in nonfinancial income (wage and private business income; line, bottom panel) across households in the main tax data sample in each year. Net savings flows in stocks are imputed from dividend income across consecutive tax years and scaled by each household's average nonfinancial income over the prior three tax years before computing means. Changes in nonfinancial income across tax years are scaled by the same prior three-year average. Each bar is centered at the transition date between the two tax years used to impute savings, e.g. the bar centered at $t = 2010$ represents flows imputed from income in tax years 2010 and 2011 that primarily reflect transactions taking place around January 1, 2010. Points along the line for income growth (bottom panel) are aligned similarly. For stock returns (top panel), points at each date represent the cumulative real return over the prior 12 months. See Section 3 and Figure 6 for details about the sample construction and savings flow imputation.

more moderate stock price crash in 2022 saw steep decreases in both net flows into stocks and real income relative to the prior year. These aggregate time series present suggestive evidence that declines in investors' nonfinancial income may partly drive the large outflows from stocks observed during these periods. In my empirical analysis, I use a wealth of additional household-level information to more sharply characterize their savings behavior and responses to such shocks.

I first present an overview of balance sheets and expenditures for the high-income working-age households I study, using publicly-available survey data. The typical household in this group has substantial assets, but the majority in illiquid forms such as housing or tax-advantaged retirement accounts (which carry a 10% tax penalty for most withdrawals made before age 59½). Their liquid asset holdings – including stocks, mutual funds, bank account balances, bonds, and other assets held outside retirement accounts – add up to large dollar amounts but typically only a fraction of annual income. In a hypothetical financial emergency most report they would draw down on these liquid assets rather than borrow or cut their spending, much of which consists of difficult-to-adjust expenditures such as housing costs and childcare. But given observed levels of liquid asset holdings, most would need to exhaust a large share of these assets in order to fully fund their typical expenditures over several months. I then turn to the panel income tax data to characterize the *actual* savings and drawdown decisions of millions of households, beyond what can be done with survey data.

Using this measure of active savings flows imputed from individual households' dividend and interest income, I present new stylized facts about the savings behavior of high-income working-age households over the business cycle. Within their taxable accounts, these investors' average flows into stocks are volatile and strongly procyclical, especially for those initially at the top of the nonfinancial income distribution (i.e. wage and private business). The skewness of these flows into stocks across households is also strongly procyclical: the median flow is close to zero in most years, while the share of households with large inflows versus outflows varies significantly across expansions and recessions. In contrast, flows into fixed income are less procyclical and less skewed. These stylized facts for working-age investors flows into stocks mirror many of the stylized facts for individual labor income growth documented in the seminal work of Guvenen, Ozkan, and Song (2014). The tax data allow me to further characterize the joint distribution of individual savings flows and income growth: many of the households that took large drawdowns from their liquid assets in a given year also experienced large declines in their nonfinancial income over the same period.

Motivated by this evidence, I formally estimate the savings response of individual households to exogenous and persistent shocks to their nonfinancial income, similar in nature to the persistent earnings losses observed in recessions. I exploit the fine geographic information available in the tax data and construct these shocks from ZIP-code level aggregate income growth, capturing changes in local economic conditions and the earnings prospects of individuals employed in similar firms and occupations. These local shocks generate large and persistent declines in households' nonfinancial income, and I find that they take large drawdowns from their liquid assets in response. In the first two years following the shock I estimate a "marginal propensity to save" between 0.50 to 0.85 out of each lost dollar of post-tax income, reflecting both larger drawdowns on their initial holdings and lower savings inflows. These liquid asset drawdowns include both stock and fixed income holdings, and are larger among households with higher initial assets.[2] This large savings response to persistent income shocks stands in stark contrast to benchmark consumption-savings models that predict households will primarily adjust *spending* in response to such shocks. For example, in the simplest perfect foresight model where the consumer's time discount rate equals the risk-free interest rate, a permanent shift in the level of income leads to *no* change in savings or wealth, and a one-for-one permanent shift in consumption instead.[3]

Given the observed demand for liquidity by working-age investors at times when risky asset valuations are low, how should they allocate their wealth between asset classes? Is it optimal to invest a large share of liquid wealth in stocks, as many of them do in the data? The second contribution of this paper is to answer these questions in a quantitative study of optimal portfolio choice, disciplined by the empirical evidence from the tax data. My model combines several elements that have typically been studied in isolation in the existing portfolio choice literature: time-varying expected stock returns as in Merton (1973) and Campbell and Viceira (1999); risky labor income that is correlated with stock returns as in Viceira (2001); and frictions in adjusting consumption as in Grossman and Laroque (1990) and Flavin and Nakagawa (2008). Recessions in the model are characterized by sharp declines in stock prices that are partly attributable to news of high future returns and are thus partly transitory, consistent with empirical asset pricing evidence. They are also characterized by high idiosyncratic income risk, with some households suffering large and persistent declines in their nonfinancial income. In the absence of adjustment frictions, households facing an unexpected and persistent decline in their income would cut their consumption sharply. When adjustment frictions prevent them from doing so they must instead draw down on their liquid assets. The larger the share of these liquid assets that are initially invested in stocks, the larger the proportion of liquid wealth that is exhausted as a result. This realistic liquidity demand of working-age investors in the model, at times when risky asset valuations are low and expected future returns are high, serves as a strong deterrent from investing most of their liquid wealth in stocks. This contrasts with typical portfolio choice models without income risk or consumption adjustment frictions, in which transitory stock price fluctuations are relatively inconsequential for long-term investors.

The intuition for these optimal portfolio choice results can be conveyed in a simple illus-

[2]The fact that households draw down on their fixed income holdings (as well as stocks) in response to these income shocks may seem to be in tension with the weak cyclicality of average fixed income savings flows stated in the previous paragraph. These savings responses are estimated from panel regressions with granular time-interacted fixed effects. The average savings flow across all households in a given year reflect both their responses to income shocks, and additional factors common to all households that drive overall savings and rebalancing between asset classes but are absorbed by the time-interacted fixed effects in the panel regressions. Section 4.2 contains additional discussion comparing these results.

[3]See Carroll (2009), who also shows that the marginal propensity to consume out of permanent income shocks is smaller but still close to 1 in the non-perfect foresight case with permanent and transitory income shocks (ranging from 0.75 to 0.92 across parameterizations).

trative example, disciplined by the statistics presented earlier. Consider a married couple that initially earns \$600,000 in combined annual salary, holds \$400,000 in liquid financial assets that are fully invested in stocks, and pays \$200,000 annually in costs such as mortgage interest, utilities, and others that are difficult to reduce or defer. In an unfortunate turn of events the aggregate stock market temporarily declines in value by 50%, similar to the peak-to-trough decline in the value of the S&P500 index during the Great Recession. At the same time both spouses lose their job and income for the year. If they draw down on their stockholdings to fully cover their expenditures, then their liquid assets — now worth only \$200,000 after the stock price crash — will be fully depleted. If they instead invested all of their liquid wealth in safe assets, they would deplete only half of it and could even use the remainder to purchase stocks at low valuations.

Taken together, these results suggest that stocks are riskier for many working-age investors than the conventional wisdom suggests. Stocks serve as both a long-term investment and one of the main liquid assets on the balance sheet of many of these households. The likelihood of needing to tap into this liquidity rises exactly when stock prices are low, since wage and private business income tends to fall at the same time. I provide empirical evidence that many households do indeed appear to liquidate their stockholdings at exactly these times. In a quantitative model that reproduces this large savings response to persistent income shocks through consumption adjustment frictions, households' optimal stock portfolio shares are low even when most fluctuations in stock prices are purely transitory.

Related literature: Empirically, this paper provides new evidence on household savings and investment behavior. I do so by developing new capitalization methods to impute households' net savings flows from their observed dividend and interest income, which build on earlier methods for estimating wealth levels from these income flows (Smith, Zidar, and Zwick, 2023) and for estimating "synthetic" savings flows from group-level average wealth estimates (Saez and Zucman, 2016; Martínez-Toledano, 2023). The savings imputation method I present is similar in spirit to these earlier synthetic savings flow estimates, but improves in several important respects. First, I carefully account for the timing of asset transactions within each year, which is particularly important given the large movements in stock prices occurring within many calendar years. I describe how transactions taking place at different points during the year are reflected in the savings flows imputed from annual income, and adjust the inputs to the capitalization formula appropriately.[4] Second, I estimate these savings flows at the household level using panel data, whereas these synthetic savings flow estimates are constructed from repeated cross-section estimates of group-level averages under the assumption that the composition of the group is stable over time. My household-level estimates allow me to characterize the full cross-sectional distribution of savings flows across households, and the joint dynamics of individual savings flows and nonfinancial income.

The new empirical evidence presented in this paper about the dynamics of household savings flows builds on existing work using a wide variety of different data sources. Gabaix and

[4] See Appendix A.3.

Koijen (2023) use data from the Federal Reserve Board's Flow of Funds accounts to document that U.S. households' aggregate flows into the stock market are highly procyclical. Hoopes, Langetieg, Nagel, Reck, Slemrod, and Stuart (2022) and Armstrong, Hoopes, and Maydew (2025) use alternative tax data from capital gains tax reporting (available through partnership with the IRS) to study investors' gross sales over time and around major stock market crashes. Gabaix, Koijen, Mainardi, Oh, and Yogo (2025a) and Gabaix, Koijen, Mainardi, Oh, and Yogo (2025b) use detailed holdings data available from 2016 to document that high-net-worth households' flows are also highly procyclical.[5] Giglio, Maggiori, Stroebel, and Utkus (2021a) and Giglio, Maggiori, Stroebel, and Utkus (2021b) use holdings data linked to household survey responses to study the role of investor beliefs in driving these flows. Brunnermeier and Nagel (2008) study investors' portfolio rebalancing in response to income shocks using survey data, while Meeuwis (2022) does so with detailed holdings data. Other work outside of the U.S. makes use of population-wide wealth tax registry data, such as Calvet, Campbell, and Sodini (2009). In particular, using asset holdings and income data from Norway, Mogstad, Schmidt, Schwartz, Tiurina, Vestad, and von Turkovich (2025) show that wealthy households make large sales from their stockholdings in response to losses in their wage and private business income. They show that these flows significantly impact aggregate stock prices and risk premia in a general equilibrium model, which may account for part of the decline in stock prices observed during periods of falling nonfinancial income.

I build on this existing work by providing new evidence on individual U.S. households' savings flows over a period spanning more than 25 years, covering multiple boom-bust cycles in both the stock market and the real economy. The tax data used in this paper cover the full U.S. population over a longer sample period than most datasets on individual U.S. investor holdings. Additionally, for a given individual the tax data provide a comprehensive summary of their taxable dividend and interest income received from *all* sources, not just accounts held at a single financial institution. Beyond documenting the properties of households' savings flows, the evidence I present on households' savings responses to income shocks provides additional insight into the *drivers* of these flows. The large adjustment in households' liquid asset holdings that we document in response to persistent income shocks is consistent with complimentary empirical evidence that finds relatively small consumption responses to similar persistent income shocks among households with relatively high liquid asset holdings.[6] I provide additional evidence on the composition of these liquid asset drawdowns, and show that stockholdings account for a significant share.

Theoretically, this paper ties together three main strands of the portfolio choice literature. The first is the literature studying optimal portfolio choice for long-term investors in models where expected excess stock returns vary over time, such as Merton (1973), Campbell and Viceira (1999), Barberis (2000), and Wachter (2002). These models mostly ignore the role of

[5]Balloch and Richers (2023) use these data to study the portfolio composition and returns of high-net-worth households, while Mainardi (2025) use them to study investors' responses to capital gains taxation.

[6]For example, see Baker (2018), Andersen, Jensen, Johannesen, Kreiner, Leth-Petersen, and Sheridan (2023), Patterson (2023), and Fagereng, Onshuus, and Torstensen (2024).

risky nonfinancial income and typically solve the portfolio choice problem of an investor whose consumption is financed exclusively by financial wealth. Conservative long-term investors in these models (with relative risk aversion greater than 1) have an intertemporal hedging motive that leads them to invest a larger share of their wealth in assets whose realized returns covary negatively with news about their expected future portfolio returns, all else equal. As a result, in empirically realistic calibrations where news about expected future returns account for a large share of variation in realized stock returns over short horizons, these models broadly prescribe that conservative long-term investors should hold a larger share of their wealth in stocks on average (relative to a myopic investor who cares only about one-period returns, or to a world in which the observed variance of stock returns is accounted for entirely by permanent revisions in expected future cash flows). These models formalize the conventional wisdom that stocks are "safe in the long run."

I build on these models by incorporating risky nonfinancial income and consumption adjustment frictions. When both of these additional elements are present, transitory stock price crashes become much more painful for long-term investors if large declines in nonfinancial income are likely to occur at the same time. Facing a large and persistent decline in nonfinancial income, an investor would wish to cut her consumption. When short-run adjustment frictions make it difficult to do so, her liquid assets must adjust instead. The larger her initial stock portfolio share, the larger her proportion of initial liquid wealth that is depleted as a result. The intertemporal benefit of high expected future returns is substantially weakened when investors have little financial wealth left to reinvest. The model-implied optimal stock share is much lower for working-age households with typical levels of consumption, income, and liquid wealth than for the canonical long-term investor whose consumption is financed exclusively by financial wealth. Consumption-wealth ratios are low enough for the latter group that the traditional intertemporal hedging motive dominates in their portfolio choice.

The second strand is the literature studying optimal portfolio choice in models with risky human capital. These models incorporate rich characterizations of individual labor income risk and its contemporaneous correlation with asset returns.[7] In more recent work, Catherine (2022) and Shen (2024) show that richer models with countercyclical idiosyncratic earnings risk can quantitatively explain average stock market participation rates and portfolio shares over the life cycle without relying on unreasonable estimates of risk aversion and participation costs. However, in contrast with the models of time-varying expected stock returns discussed earlier and in contradiction with a large body of empirical asset pricing evidence, these models typically assume that realized stock returns are serially uncorrelated (and thus driven entirely by permanent cash flow news rather than news about future expected returns). Investors avoid stocks in these models because of the possibility of a large permanent decline in stock prices that coincides with a large permanent decline in their labor income. I show that investors' aversion to stocks is weaker when these large declines in stock prices are instead partly transitory and

[7] In important early quantitative work, Viceira (2001), Cocco, Gomes, and Maenhout (2005), and Benzoni, Collin-Dufresne, and Goldstein (2007) study models with normally-distributed returns and income shocks, where idiosyncratic earnings risk does not vary over time.

driven by higher expected future returns. Incorporating consumption adjustment frictions overturns this result and makes stocks risky from the perspective of working-age investors. This aversion to stocks is no longer driven primarily by the risk of large permanent crashes in both stock prices and labor income, but rather by an empirically-realistic *demand for liquid wealth* in states of the world with low income, rigid expenditures, and temporarily-depressed stock prices.[8]

The third strand is the literature studying optimal portfolio choice in models with consumption adjustment frictions. Grossman and Laroque (1990), Flavin and Nakagawa (2008), Chetty and Szeidl (2007), and Chetty, Sandor, and Szeidl (2017) all study portfolio choice models in which a subset of consumption goods such as housing cannot be freely or costlessly adjusted each period, while Gomes and Michaelides (2003) study the portfolio choice problem under internal habit formation. As in the portfolio choice models with risky human capital discussed above, these existing models either assume that stock returns are serially uncorrelated or consider static portfolio choice over a single time interval. My contribution here is to incorporate these adjustment frictions into a dynamic setting with time-varying expected stock returns and risky nonfinancial income. These adjustment frictions make *transitory* stock price declines much more painful for working-age investors with high and empirically-realistic ratios of consumption to liquid wealth, since they must potentially exhaust a large share of this wealth to fund their expenditure commitments following a sharp decline in their wage or private business income. For retired investors with low consumption-to-liquid wealth ratios these adjustment frictions are less important, and the model-implied optimal portfolios are similar to the frictionless case. Importantly, I discipline these adjustment frictions using the size of households' observed asset drawdowns in response to large income shocks, which I estimate using the administrative income tax data. My focus on how investors accumulate and allocate their *liquid* wealth builds on existing work that studies the role of illiquid assets such as housing and retirement accounts in household portfolio choice.[9]

The remaining sections of the paper are organized as follows. Section 2 provides an initial overview of households' wealth, income, and expenditures based on survey data. Section 3 describes the administrative income tax panel data and the methodology for estimating savings flows from observed dividend and interest income. Section 4 presents the main empirical results using the tax data. Section 5 describes the portfolio choice model. Finally, Section 6 concludes.

[8]Michaelides and Zhang (2017) study a model with time-varying expected stock returns, but where stock returns and labor income are uncorrelated. Benzoni, Collin-Dufresne, and Goldstein (2007) incorporate return predictability in a model where idiosyncratic earnings risk does not vary over time.

[9]For quantitative portfolio choice models with housing, see Cocco (2005) and Vestman (2019) (among others). For quantitative models with retirement accounts, see Gomes, Michaelides, and Polkovnichenko (2009) and Dahlquist, Setty, and Vestman (2018) (among others). Campanale, Fugazza, and Gomes (2015) study a model where all stock transactions require paying a fixed transaction cost. Duarte, Fonseca, Parker, and Goodman (2022) incorporate both of these features, as well as other features studied in prior work, into a single rich portfolio choice model.

2 An Overview of Household Balance Sheets and Expenditures

Before introducing the administrative income tax data, I first characterize the level and composition of household wealth, income, and expenditures using publicly-available survey data. I present several stylized facts, with a focus on high-income working-age households. Illiquid assets such as real estate and tax-advantaged retirement accounts comprise the majority of assets held by these households. Many of them invest a large share of the liquid assets that they hold in stocks. While total net worth exceeds earnings for many households, the level of liquid assets does not. A large share of these assets would thus be exhausted if they were to be used to cover typical expenditures following a significant income shock, and a large share of high-income households' expenditures are concentrated in categories such as housing, healthcare and education that may be difficult to adjust or defer on short notice. Additional survey questions indicate that most of these households would indeed draw down on their financial assets in response to a hypothetical shock, rather than cutting discretionary spending, borrowing, or postponing payments for existing debt or bills. In the following sections of the paper, I use the panel income tax data to empirically characterize how households adjust their asset holdings in response to *actual* observed income shocks.

The main dataset used in this section is the Survey of Consumer Finances (SCF), a publicly-available dataset based on a triennial survey of U.S. households conducted by the Federal Reserve Board. Each survey provides an annual snapshot of wealth, income, and demographics for a large cross-section of households, with a sample design that deliberately oversamples households at the top of the wealth distribution.[10] These snapshots are available every three years from 1989 to 2022. The SCF provides detailed information on the value and composition of households' assets, liabilities, and income. However, the repeated cross-section structure makes it impossible to follow the same individual across multiple consecutive years in order to characterize the *dynamics* of their income and wealth over time. This can be done with panel administrative income tax data, which also have the benefits of nearly full population coverage and the absence of missing or incorrect survey responses.[11] Nevertheless, the SCF data provide useful information on wealth which cannot be obtained directly from U.S. income tax data.

I define a SCF household's *nonfinancial income* as the sum of wage/salary income and private business income.[12] The inclusion of private business income is motivated by Smith,

[10] See Kennickell (2008). The SCF provides sample weights, which I use to construct a representative weighted sample.

[11] The SCF sample includes only several thousand unique households per year. The number of households included in the SCF started at 3,143 in the 1989 wave, peaked at 6,482 in the 2010 wave, and has since fallen to 4,595 in the 2022 wave.

[12] Wage/salary income is defined as Summary Extract variable `wageinc`, derived from survey variable `X5702`. Private business income is defined as Summary Extract variable `bussefarminc`, derived from survey variables `X5704` for sole proprietorship and farm income reported on IRS Schedule C, and `X5714` for S corporation, partnership, trust, rental, and royalty income reported on IRS Schedule E.

Yagan, Zidar, and Zwick (2019) who show that many working-age individuals at the top of the income distribution derive the majority of their income from pass-through private businesses, and argue that this income primarily reflects returns to the business owner's human capital rather than financial capital. This contrasts sharply with income received from dividend, interest and realized capital gains, which primarily flow from holdings of liquid financial assets such as publicly-traded stocks and bank deposits. Because I effectively treat households' private business holdings as illiquid assets, I exclude the estimated value of these holdings when computing total household assets and net worth.[13]

When ranking households by income or wealth, I divide these measures by 2 in cases where the respondent is married, but multiply the sampling weight for the household by 2. This corresponds to individual-level rankings based on equal splits of tax units/households, which is consistent with the treatment of the income tax data and with prior work such as Smith, Zidar, and Zwick (2023). The age of the household is defined as the age of the reference person. In order to provide a detailed disaggregation across age, wealth, and income groups, I pool individuals across SCF survey waves but adjust the sample weights so that each wave is weighted equally when computing each statistic.

Fact 1: Typical working-age households hold most of their assets in illiquid forms (even among top earners). Panel (a) of Figure 2 decomposes the average asset holdings across the income distribution among households age 40-45. I disaggregate assets (excluding private business holdings) into five broad categories, sorted from least to most liquid: real estate; other nonfinancial assets; annuities, trusts, and other miscellaneous financial assets; retirement accounts such as IRAs or 401(k) accounts; and liquid financial assets.[14] This definition of relatively liquid financial assets includes those that can be sold or drawn down upon relatively quickly without large transaction costs or penalties, and includes bank account balances, stocks and mutual funds held in taxable brokerage accounts, and other assets listed in Panel (b). In contrast, selling nonfinancial assets such as real estate often involves delays and high transaction costs; retirement accounts *can* be liquidated if needed, but doing so triggers large tax penalites for working-age households.

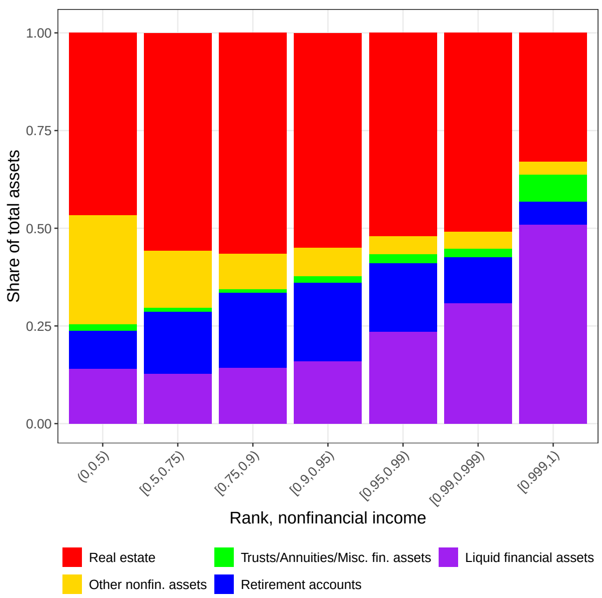
The average working-age household holds the majority of its assets in illiquid forms, even among top earners. Consider the top 1% earners among individuals age 40-45. Nonfinancial assets such as real estate represent more than half of their asset holdings, while liquid financial assets represent roughly one third. Outside of the top 1% earners the share of liquid assets is even lower; other nonfinancial assets such as vehicles account for a larger share of household assets, while a greater share of financial assets are held in retirement accounts. Although I focus on households age 40-45 here, the average liquid share of total assets varies much less across age groups than across the income distribution in a given age group. Next, I unpack the composition of these liquid financial assets.

[13]The value of households' private business holdings is defined as Summary Extract variable `bus`, and includes businesses where household members have either an active or nonactive interest.

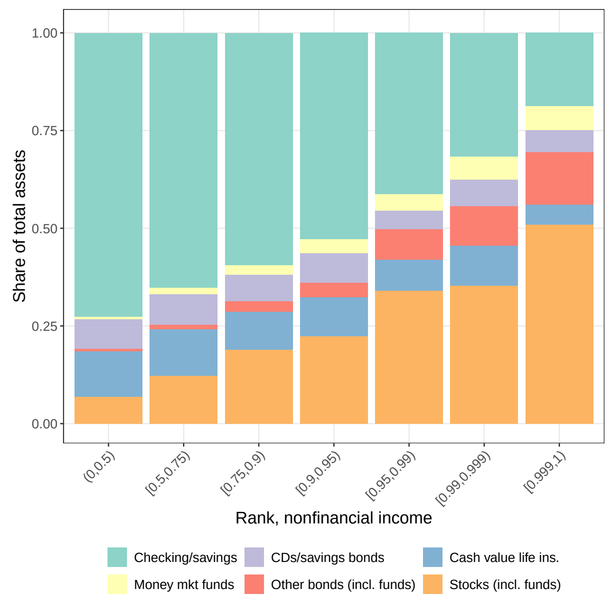
[14]Liquid financial assets are defined as total financial assets (SCF Bulletin variable `fin`) minus quasi-liquid retirement accounts (`retqliq`); trusts and annuities (`othma`); and other/miscellaneous financial assets (`othfin`).

Figure 2: Households' Assets and Liquid Stock Holdings

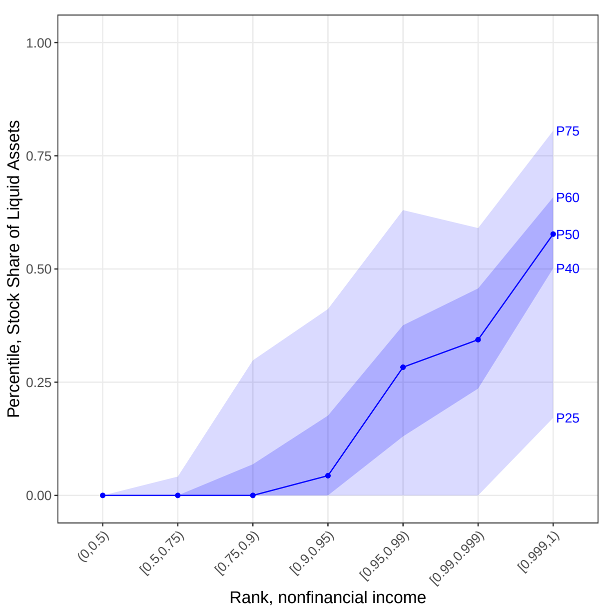
(a) Liquid and Illiquid Assets at Age 40-45



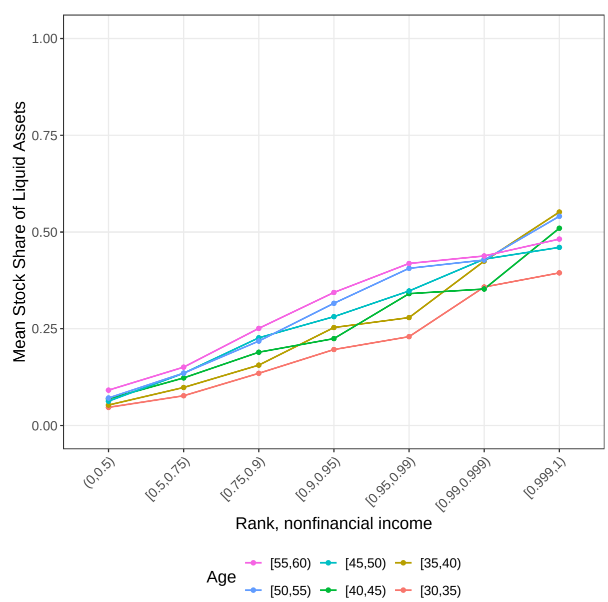
(b) Composition of Liquid Assets at Age 40-45



(c) Stock Share of Liquid Assets, Distribution at Age 40-45



(d) Average Stock Share of Liquid Assets, by Age and Income



Notes: This figure summarizes the composition of assets held by working-age households in the Survey of Consumer Finances. In all panels, the horizontal axis represents the rank of households in the distribution of nonfinancial income (wage/salary and private business income) within their five-year age group. Panel (a) plots average shares for different components of total assets (excluding private business holdings) across the income distribution among households age 40-45. Panel (b) plots average shares for subcomponents of liquid financial assets — defined as total financial assets excluding retirement accounts, trusts, annuities, and miscellaneous financial assets — among households age 40-45. Panel (c) plots percentiles (25, 40, 50/median, 60, 75) of the share of stocks in liquid assets among households age 40-45. Panel (d) plots the mean share of stocks in liquid assets for households age 30-60, split by five-year age groups and income rank within each age group. All panels pool SCF waves 1989 through 2022, and each wave is weighted equally when computing each individual statistic.

Fact 2: A large share of their liquid financial assets are invested in stocks. Panel (b) of Figure 2 decomposes these households' *liquid* asset holdings in a similar manner. I split these into transaction accounts (e.g. bank checking and savings accounts, money market funds, etc.), certificates of deposit and savings bonds, other bond holdings (e.g. Treasury, corporate, and municipal bonds), cash value life insurance, and taxable stock holdings. I include both direct and indirect holdings (e.g. through mutual funds or ETFs) for stocks and bonds.

The top 1% earners at age 40-45 hold 35% of their liquid financial assets in stocks on average. The average share held in stocks is similar to the share held in liquid transaction accounts (38%). Bonds and certificates of deposit account for 17% of liquid wealth, while cash value life insurance accounts for 10%. There is also substantial heterogeneity in portfolios among households at the same point in the income distribution. Panel (c) plots percentiles of the stock share of liquid assets for the same groups of households. Roughly one quarter of the top 1% earners hold no liquid stocks (outside of retirement accounts), while one third hold over half of their liquid wealth in stocks. Panel (d) shows that the mean share of liquid assets invested in stocks rises with both age and nonfinancial income.

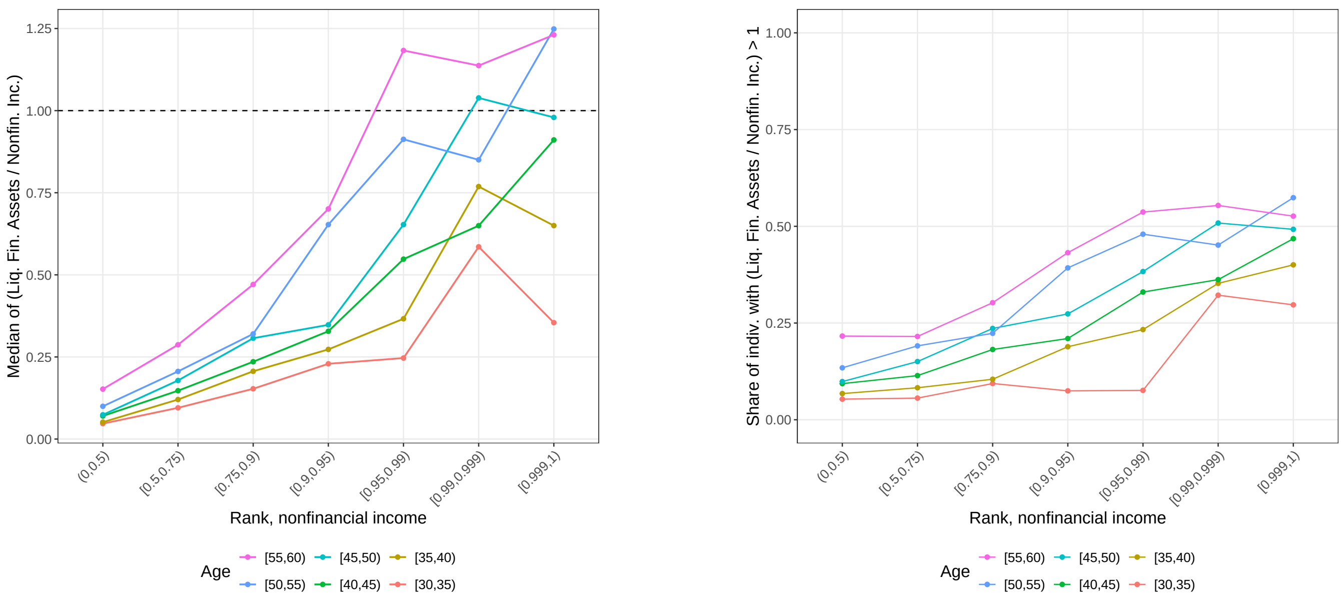
The stock share of liquid assets that I calculate in the SCF differs from measures used in other papers studying U.S. households' finances. I focus on financial assets that working-age households can quickly and easily liquidate or drawn down on without incurring large tax penalties or transaction costs; this excludes tax-advantaged retirement accounts, despite their importance in investors' long-term retirement savings plans. Ameriks and Zeldes (2004) use SCF data to calculate the stock share of total financial assets including retirement accounts (which affects both the denominator and numerator in the calculation), as well as separate panel data on retirement accounts from TIAA-CREF. Parker, Schoar, Cole, and Simester (2024) use data from a large financial services company to calculate the stock share of total assets held in each individual's accounts at the company, which includes retirement accounts but excludes assets held in accounts at other companies (e.g. bank accounts). The measure most similar to the one used in this paper is the risky share of liquid assets computed by Gabaix, Koijen, Mainardi, Oh, and Yogo (2025a) using data from the wealth management platform Addepar.

Fact 3: Most working-age households hold less than one year's earnings in liquid assets. After characterizing the composition of households' liquid wealth, I now characterize its level relative to their (pre-tax) earnings. Panel (a) of Figure 3 plots the median liquid wealth-to-nonfinancial income ratios across individual households by age and their rank in the (nonfinancial) income distribution, while Panel (b) plots the share of households for whom this ratio is greater than 1. Most working-age households have liquid asset holdings that amount to less than their annual pre-tax earnings. This holds across the income distribution: liquid wealth-to-income ratios increase with income, yet are still below 1 for the majority of top earners within all but the oldest age groups.

The observed levels of liquid wealth for working-age households imply that a large unexpected expenditure shock — similar in magnitude to their annual earnings — could force many

Figure 3: Liquid Assets-to-Income Ratios for Individual Households

(a) Median Ratios, by Age and Income **(b)** Share of Individuals with Ratio > 1



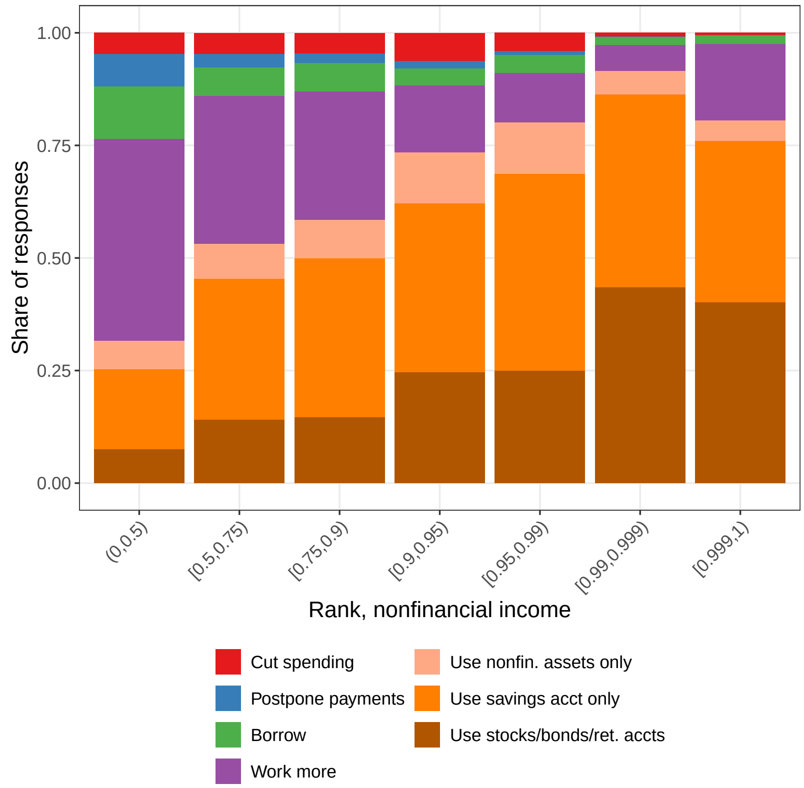
Notes: This figure plots summary statistics for the ratio of liquid financial assets-to-nonfinancial income (wage/salary and private business income, before taxes) among individual working-age households in the Survey of Consumer Finances. In all panels, the horizontal axis represents the rank of households in the distribution of nonfinancial income (wage/salary and private business income) within their five-year age group. Panel (a) plots the median value of this ratio for groups of households split by five-year age groups and income rank within each age group. Panel (b) plots the share of households for whom this ratio exceeds 1 (i.e. their liquid assets exceed their income in the previous year) within each group. All panels pool SCF waves 1989 through 2022, and each wave is weighted equally when computing each individual statistic.

of them to exhaust a large share of their liquid assets. Next I show that survey evidence from the SCF suggests this is in fact the main margin of adjustment that many households would rely on in response to such a shock.

Fact 4: Few households would cut spending, borrow, or postpone payments in response to unexpected emergency expenses. In 2019 and 2022, the SCF included a question asking households how they would respond to an unexpected financial emergency:

> **Question (2022 survey):** If tomorrow you experienced a financial emergency that left you unable to pay all of your bills, how would you deal with it? Would you borrow money, would you spend out of savings or investments, would you postpone paying bills, work more or get an extra job, or would you do something else?
>
> 1. Borrow money
> 2. Spend out of savings/investments
> 3. Postpone payments
> 4. Cut back [on spending]
> 5. Work more/get extra job

Figure 4: How Would Households Adjust in a Hypothetical Financial Emergency?



Notes: This figure plots the share of responses to the SCF's hypothetical financial emergency question for working-age households at different points in the income distribution. The horizontal axis represents the rank of households in the distribution of nonfinancial income (wage/salary and private business income) within their five-year age group. The components of each bar represent the share of households in each group that responded with one of the five options (see text for details). The response "Spend out of savings/investments" is partitioned into three different subcomponents based on answers to follow-up questions: (a) would liquidate financial assets including stocks, bonds, certificates of deposit, and/or retirement accounts (possibly in addition to other assets); (b) would only draw down on savings accounts (the most common response); and (c) would sell or draw down only on nonfinancial assets (including borrowing through a home equity line of credit). Observations are pooled from the 2019 and 2022 SCF waves, and each wave is weighted equally when computing each individual statistic.

Respondents answer with only one of the five options.[15] Detailed follow-up questions ask which specific types of financial assets they would draw down on, sources they would borrow from, discretionary spending categories they would cut back on, or existing debt and bill payments they would postpone.

Figure 4 plots the share of each response among households at different points in the income distribution. Because this question was asked in a consistent manner for only the two most recent survey waves, I pool individuals age 30-60 when computing these shares. Few households report that they would cut spending or postpone existing debt and bill payments. A slightly higher share report that they would borrow. Most young lower-income households

[15]Responses are recorded in the SCF variable `X7775`. This question is only asked to respondents who did not spend more than their income in the previous year (variable `X7508`). A similar question was asked in 2016 but without the "Work more/get extra job" option, so I focus on the 2019 and 2022 surveys that feature the same options.

report that they would increase their labor supply. Among higher-income households, the vast majority report that they would draw down on their savings and investments. The detailed follow-up question asks which specific assets they would draw down on. While the modal respondent would draw down only on their savings accounts, a large share of high-income respondents report that they would liquidate stocks, bonds, certificates of deposit, or assets held in pension and retirement accounts.

These survey questions ask how households would adjust to a hypothetical financial emergency, rather than how households adjusted in response to an *actual* situation they faced in the past. Nevertheless, the responses suggest that the vast majority of high-income households view their financial asset holdings as the main margin of adjustment they would rely on in response to such a shock. In the following sections of the paper I move beyond the SCF's repeated cross-section data and turn to the panel income tax dataset, which I use to provide a rich empirical characterization of how households adjust their financial asset holdings in response to observed fluctuations in their nonfinancial income.

Fact 5: Housing, healthcare, education, and other categories of spending that are difficult to quickly reduce or defer account for a large share of high-income households' expenditures. Few high-income households surveyed in the SCF report that they would cut back on their spending in a hypothetical financial emergency. One reason for this may be that relatively inflexible spending categories such as housing, healthcare, and education account for a large share of their expenditures. To document this fact, I turn to more detailed microdata from the Consumer Expenditure Survey (CEX).

The CEX is a quarterly survey in which households report their spending across detailed individual categories of goods and services, as well as their income and demographic information. I use these data to compute average expenditure shares across major spending categories for households at different points in the income distribution.Unlike the SCF, the CEX does not oversample households at the top of the income distribution, and income is top-coded at levels which affect between 3 to 6% of households in the sample in a given year. Therefore, I consolidate households in the top 5% of the income distribution when computing average expenditure shares by category.

Table 1 reports average expenditure shares by category for households at different points in the income distribution. In a similar analysis that pooled households across the income distribution, Chetty and Szeidl (2007) showed that a large share of expenditures for the typical household are concentrated in categories such as housing that may be difficult or costly to adjust in the short term. The average shares reported here show that this is true even for households in the upper tail of the income distribution. Housing and utilities expenses (including rent, mortgage interest payments, and property taxes) account for the largest share of expenditures for both low- and high-income households, representing more than one third of expenditures even for top-5% earners. Compared to lower-income households, high-income households allocate a smaller share of their spending to food at home, and a slightly larger share to food away from home and discretionary spending such as entertainment. They al-

Table 1: Shares of Households' Consumption Expenditures (Consumer Expenditure Survey)

Expenditure category	(0,0.5)	[0.5,0.75)	[0.75,0.9)	[0.9,0.95)	[0.95,1)
Shelter and utilities	0.375	0.339	0.335	0.339	0.339
Healthcare and life insurance	0.061	0.076	0.074	0.064	0.067
Education and personal services	0.016	0.031	0.043	0.051	0.076
Food at home	0.163	0.130	0.109	0.105	0.094
Food away from home and other discretionary	0.121	0.134	0.142	0.149	0.142
Other household and misc. expend.	0.020	0.020	0.020	0.023	0.020
Apparel, personal care, and household supplies	0.034	0.035	0.040	0.043	0.043
Transportation (gas, car insurance, transit, etc.)	0.105	0.106	0.100	0.092	0.091
Durables purchases (vehicles, appliances, etc.)	0.080	0.099	0.103	0.094	0.094
Cash contributions (gifts, donations, etc.)	0.025	0.029	0.033	0.040	0.034

Notes: This table reports average expenditure shares by category computed from CEX data, among households age 40-45 grouped by nonfinancial income. Observations are pooled across CEX sample waves, and the top 5% is used as the highest income group due to income top-coding.

locate a much larger share of their budget to education (e.g. tuition for children's school or college) and personal services (e.g. child care).

High-income households do allocate a slightly larger share of their spending towards discretionary luxury goods and services such as restaurant dining and entertainment that could be easily reduced if necessary. However, relatively inflexible categories account for a large share of their spending. Housing/utilities, healthcare/life insurance, and education/personal services together account for nearly half of spending for the typical top-5% earner.

Taken together, these facts suggest that the typical high-income working-age household has substantial asset holdings, but primarily in illiquid forms that would be difficult or costly to tap for liquidity if needed. Although many hold large dollar amounts of liquid assets, for most households these balances represent only a fraction of their income or expenditures in a typical year. If they were to draw down on these liquid assets to cover typical expenditures following a large negative shock to their income, they would likely deplete a large share of those balances in doing so. Next, I turn to the administrative income tax data which will allow me to characterize how households respond to shocks of this nature, beyond what can be learned from the repeated cross-section survey data discussed in this section.

3 Overview of Administrative Income Tax Data and Savings Estimates

This section provides an overview of the data and capitalization methods used in this paper. Section 3.1 describes the administrative income tax panel and other linked datasets. Section 3.2 describes the capitalization methods used to estimate active savings flows from observed dividend and interest income.

3.1 Overview of Administrative Income Tax Data and Other Linked Datasets

The main dataset used in my analysis is the U.S. Census Bureau's panel of annual income tax records for the full universe of individual filers from 1998 to 2023. These data have previously been used in papers such as Chetty, Hendren, Jones, and Porter (2020). For each individual, I observe the dollar value of many (but not all) of the pre-tax income line items reported on their individual income tax return Form 1040. The main income variables used:

- Wage and salary income from paid employment, as reported on Form W-2.
- Cash flows received from financial assets held outside of tax-advantaged accounts such as IRAs and 401(k)s. I separately observe:
  - Taxable ordinary dividend income received from publicly-traded stocks, mutual funds, and privately-held C-corporations.[16]
  - Taxable interest income received from bank accounts, directly-held bonds and loans, and other taxable interest-generating fixed income investments.
  - Nontaxable interest income received from municipal bonds held directly or through mutual funds.
- Adjusted gross income (AGI), a comprehensive income measure which includes realized capital gains.
- An alternative comprehensive income measure "total money income" (TMI) used in Census Bureau surveys. Unlike AGI, this measure excludes realized capital gains.[17]

For married individuals who file a joint tax return, I observe the sum total of income in each category received by both spouses.

Although these comprehensive income measures include most important income components, some of the individual components are not available in the dataset as separate line items. Income received from sole proprietorships (reported in Schedule C) and passthrough private businesses (S-corporations and partnerships, reported in Schedule E) is included in both AGI and TMI, but is not reported separately. Realized capital gains are included in AGI, but are not reported separately. However, informative proxies can be constructed for

[16]The single dividend income line item I observe includes both qualified and nonqualified dividend income. The qualified dividend income category was introduced in tax year 2003 and includes dividend income from C-corporation holdings that meet certain holding period requirements. The nonqualified dividend income category includes dividend income from both C-corporation holdings that do not meet these holding period requirements, and taxable bonds held indirectly through mutual funds (which pass the interest received from these holdings to investors in the form of dividends). Additional details are provided in IRS Publication 550 (`https://www.irs.gov/forms-pubs/about-publication-550`).

[17]Total money income includes "taxable wage and salary income, interest (taxable and tax-exempt), dividends, gross Social Security income, unemployment compensation, alimony received, business income or losses (including for partnerships and S-corps), farm income or losses, and net rent, royalty, and estate and trust income. Prior to tax year 2018, TMI also included total pensions and annuities" (Bee, Mitchell, Mittag, Rothbaum, Sanders, Schmidt, and Unrath, 2023).

these unobserved components using the available set of comprehensive income measures and individual line items. Finally, I do not observe individual income taxes paid.

I define "nonfinancial income" as TMI minus taxable dividends, taxable and nontaxable interest, and gross Social Security income. For working-age households, this measure primary reflects wage, salary, and private business income.

In my main analysis, I track each household for at most six consecutive tax years at a time. To construct the main sample used in the analysis, I start by selecting a 10% subset of all Protected Identification Keys, the Census Bureau's identifier for individuals (Wagner and Layne, 2014). In each initial reference year ranging from $t = 2000$ to $t = 2022$, I restrict the main sample to individuals satisfying the following criteria:

- **Age:** I keep individuals whose age at the end of year t is between 30 and 59 (inclusive), based on their birth year reported in the 2000 and 2010 Decennial Censuses. For married individuals filing a joint return, I use the average age of the individual and their spouse; with slight abuse of language, I refer to this as the individual's age.[18]

- **Available tax returns with consistent filing status:** I keep individuals with available tax return data in each year from $t - 2$ through $t + 3$ (or through 2023, the last year of available tax return data).[19] I also require that the individual is either married filing a joint return in all of these consecutive years, or not married in all of these years. I excluding married individuals filing separate tax returns.

- **Rank in initial nonfinancial income distribution:** I compute average nonfinancial income for each individual over tax years $t - 2$ through t. For married individuals, their "equal-split" individual income is defined as half of the combined income of both spouses reported on their joint tax return. I then keep individuals with average initial nonfinancial income in the top 20% of individuals with the same age in year t (e.g. 30-year-olds in the top 20% among other 30-year-olds).

The resulting sample provides broad coverage of working-age individuals with significant income from sources other than financial assets. These individuals also account for the majority of taxable stock ownership among non-retired households. The full sample contains between 1.30 and 1.45 million individual observations in each reference year. Rounded observation counts for the full sample of individuals in each reference year are presented in Table B.1.

[18]If different birth years are reported for the same individual in the 2000 and 2010 Decennial Censuses, an average of the ages implied by each birth year (rounded up to the nearest integer) is used, as long as they differ by less than five years; otherwise, age is set as missing. For married individuals filing a joint return, if age is missing for either spouse then the other non-missing value is used.

[19]I relax this restriction in the earliest initial reference year $t = 2000$ and keep individuals with missing tax return data in year $t - 2 = 1998$. I then compute average initial nonfinancial income based on the available values in years $t - 1 = 1999$ and $t = 2000$. Alexander, Bleckley, Fisher, Genadek, Leonard, and Magganas (2024) document the presence of missing tax returns in the Census Bureau's income tax data for earlier years, most likely due to late filers who were excluded in the data initially provided by the IRS. Because many high-income individuals obtain extensions to file late tax returns, I relax these data availability restrictions in 1998 only in order to avoid excluding them from the main sample.

3.2 Imputing Savings Flows from Observed Cash Flows

The panel income tax return dataset provides information about the cash flows generated by financial assets held outside of tax-advantaged accounts, such as dividends from taxable stock and mutual fund holdings, taxable interest from savings accounts and directly-held bonds, and nontaxable interest from municipal bonds. However, I do not observe the market value of the financial assets generating these cash flows, nor the actual transactions made by individuals in these assets. Instead, for a given individual I use the *changes* in the cash flows observed across consecutive tax years to impute the value of their net sales or purchases in each asset class. My capitalization approach builds on earlier methods used by Saez and Zucman (2016), Smith, Zidar, and Zwick (2023) and Martínez-Toledano (2023) to estimate wealth levels and active savings flows from these cash flows. In this section I present a brief overview of this approach within a simplified version of the full household accounting framework presented in Appendix A, which features multiple assets and transactions taking place at multiple points within each year.

Consider a household j that holds shares in a broad stock market index fund within their taxable brokerage account. Time t is denominated in years, with integer values denoting the start of the corresponding calendar year (e.g. $t = 2009$ represents the start of calendar year on January 1, 2009). A single share of the fund trades at price $\mathbf{P}_t^D$ at time t. Over the course of the year from time t to $t+1$, the fund pays total dividends $\mathbf{D}_{t+1}^D$.[20] The household holds $Q_{j,t}^D$ shares of the asset at time t, and receives dividend income $D_{j,t+1} \equiv Q_{j,t}^D \overline{\mathbf{D}}_{t+1}^D$ from time t to $t+1$. Similarly, its holdings $Q_{j,t-1}^D$ at time $t-1$ determine the dividend income $D_{j,t} \equiv Q_{j,t-1}^D \mathbf{D}_t^D$ that it receives from time $t-1$ to time t. For the purposes of exposition, asset transactions are assumed to take place only at the start and end of each calendar year.[21] In this case the value of the household's net purchases/sales at time t, defined as $S_{j,t,t+1}^D \equiv \mathbf{P}_t^D \cdot (Q_{j,t}^D - Q_{j,t-1}^D)$, can be backed out from their dividend income as follows:

$$S_{j,t,t+1}^D = \underbrace{\left(D_{j,t+1} - \frac{\mathbf{D}_{t+1}^D}{\mathbf{D}_t^D} D_{j,t} \right)}_{\substack{\text{“excess” change in household } j \\ \text{dividend income}}} \times \underbrace{\frac{\mathbf{P}_t^D}{\mathbf{D}_{t+1}^D}}_{\substack{\text{capitalized by} \\ \text{forward price/div. ratio}}} \tag{1}$$

The first term in parentheses can be interpreted as the "excess" change in the household's dividend income, defined the difference between their observed dividend income over the period from time t to $t+1$, relative to a buy-and-hold projection based on their dividend income in the prior year. When households trade only a single dividend-paying asset, any excess change in their dividend income is solely attributable to changes in their holdings of this asset. Capitalizing this value by the forward price-dividend ratio then converts this excess change in

[20]Note that the time subscript used for taxable income denotes the *end date* of the corresponding tax or calendar year. For example, $\mathbf{D}_{2010}^D$ denotes dividends received from time $t = 2009$ (representing January 1, 2009) up to time $t + 1 = 2010$ (representing January 1, 2010), which are reported in the household's income tax return for tax year 2009.

[21]This assumption is relaxed in Appendix A.3.

dividend income to the dollar value of the corresponding stock purchase or sale.

This example is deliberately simple but captures the key intuition behind the capitalization estimates of households active savings flows. Appendix A.1 generalizes to the case with multiple dividend-paying assets, while Appendix A.3 generalizes to the case where transactions take place at multiple dates within each year. The final capitalization formula used to estimate households' active savings flows $\widehat{S}^D_{j,t,t+1}$ in dividend-generating assets, from dividend income $D_{j,t}$ and $D_{j,t+1}$ over two consecutive tax years, is given by:

$$\widehat{S}^D_{j,t,t+1} = \frac{\overline{\mathbf{P}}^D_t}{\overline{\mathbf{D}}^D_{t+1}} \times \left(D_{j,t+1} - \frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t} D_{j,t} \right) \tag{2}$$

Here $\overline{\mathbf{P}}^D_t$ and $\overline{\mathbf{D}}^D_{t+1}$ represent the average share price and dividends-per-share paid by a representative portfolio of dividend-generating assets, which consists of roughly 90% stocks and 10% dividend-generating fixed income investments, such as bond and money market mutual funds (based on household portfolio shares of taxable dividend-generating assets observed in the SCF). Appendix A.3 describe how this average share price is computed over each time window. This imputed savings flow can be interpreted as a weighted sum of households' asset transactions taking place in the two-year window over which dividend income is measured, with the most weight placed on transactions occurring close to the transition date between the two calendar years.

I use a nearly-identical formula to impute households transactions in nontaxable interest-generating assets from their observed nontaxable interest income. The reference portfolio used to compute the share prices and cash flow growth rates in the capitalization formula is the universe of Vanguard municipal bond funds. To capitalize taxable interest income, I use a method described in Appendix A.4 that takes into account the heterogeneity in returns on interest-generating assets across the wealth distribution documented by Smith, Zidar, and Zwick (2023).

4 Empirical Evidence on Household Savings Behavior

After describing the structure of the panel income tax dataset and the capitalization methods for imputing households' active savings flows from their observed dividend and interest income, this section presents the main empirical results. In Section 4.1 I describe the properties of these estimated savings flows over time and across households. In Section 4.2 I estimate the response of these savings flows to exogenous income shocks at the household level. The large estimated household savings response to these persistent income shocks contrasts with the small response of savings in benchmark consumption-savings models. These estimates will serve as a key moment to discipline the magnitude of the consumption adjustment frictions I incorporate into the portfolio choice model described in Section 5.

Throughout this section, I frequently split the households within the main sample along two dimensions. The first dimension is their percentile in the distribution of average nonfinancial income over the prior three years, $\overline{Y}_{j,t-2:t}$, within their one-year age group. The main sample is constructed by taking households that rank within the top 20% of their age group along this measure. This main sample is further split into subsamples for percentile ranks 80 to 90, 90 to 95, 95 to 99, and 99 to 100 (top 1% earners). For brevity, I refer to these as *income groups* throughout this section.

The second dimension is the ratio of their total capitalized liquid financial wealth in year t, $\overline{W}^{tot}_{j,t}$, relative to their average nonfinancial income over the prior three years. This wealth estimate is obtained by capitalizing households' dividend and interest income in year t, using the capitalization factors described in Appendix A. For brevity, I refer to these as *wealth-to-income ratios* throughout this section.[22] Within each income group, I split households into further subsamples based on the value of this wealth-to-income ratio.

The composition of households' liquid wealth varies across both income and wealth groups. Figure 5 plots average stock shares of capitalized liquid wealth across households, which I define for a given individual based on the stock subcomponent $\overline{W}^{D}_{j,t}$ of this wealth (from their capitalized dividend income). The average stock share of this liquid wealth increases strongly across wealth-to-income groups. Within each of these wealth groups, average stock shares increase with income. Stocks account for less than a quarter of capitalized liquid wealth among households in the sample with less than half of a typical year's pre-tax income in this form of savings, but more than half of this wealth for households at the top of the income and wealth distribution.

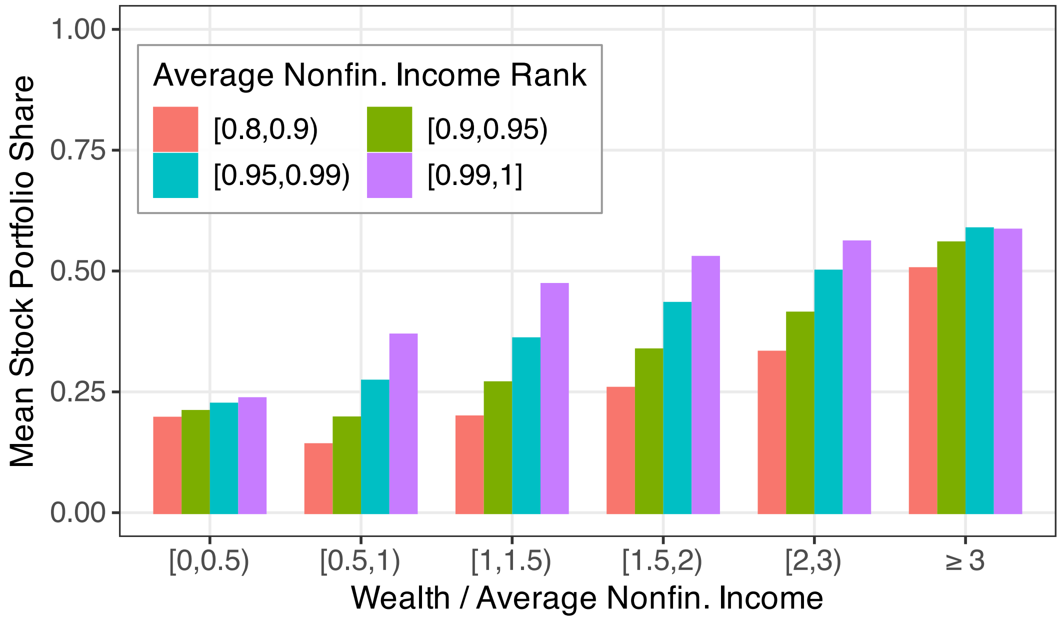
4.1 Time Series and Cross-Sectional Patterns

I now describe the properties of the savings flows estimated from the income tax data, starting with patterns over time and the business cycle. Figure 6 depicts average savings flows by asset class in each year of the sample period. The bars in the top panel depict net flows into stocks (estimated from households' dividend income), while the bottom panel depicts net flows into fixed income (estimated from households' interest income). Each household's savings flow $S^{A}_{j,t,t+1}$ is scaled by their average nonfinancial income over the prior three years (denoted by $\overline{Y}_{j,t-2:t}$) before taking the average of these scaled flows across households. The line in each panel shows the average growth in nonfinancial income $Y_{j,t+1}$ relative to the same prior average $\overline{Y}_{j,t-2:t}$.

Within their taxable investment accounts, households' average flows into the stock market are strongly procyclical. The largest average outflows from stocks occur around the two recessions in the early and late 2000s, when both aggregate stock prices and average nonfinancial income fell sharply. A similar pattern emerges even around the recent stock price crash in 2022, which coincided with more moderate outflows from stocks and declines in real average

[22] This wealth measure $\overline{W}^{tot}_{j,t}$ should not be interpreted as a measure of households' *total* wealth or net worth. It excludes forms of wealth that are relatively illiquid for working-age households but constitute a large share of their asset holdings, such as housing and tax-advantaged retirement accounts (as shown in Section 2). It also excludes forms of liquid wealth that generate no cash flows that appear in households' income tax return, such as checking accounts or physical cash holdings.

Figure 5: Estimated Stock Portfolio Shares Increase with Income and Wealth



Notes: This figure depicts average estimated stock portfolio shares of households' capitalized liquid financial wealth. Average shares are reported separately for households sorted by wealth-to-income ratios (along the horizontal axis) and by income group (adjacent colored bars).

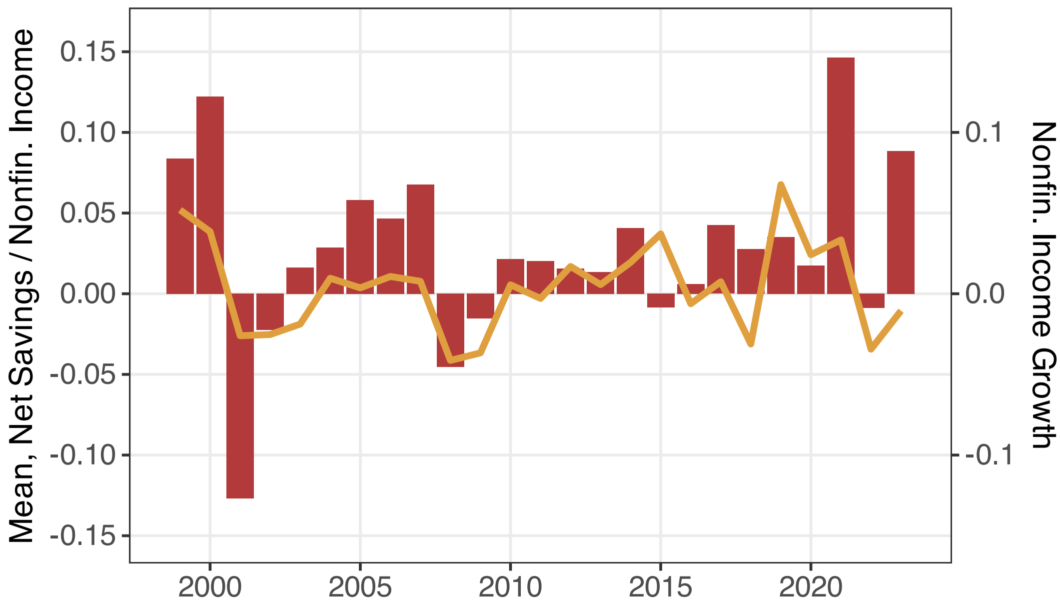
earnings for high-income households. Long economic expansions over this time period, on the other hand, coincide with both large average inflows into the stock market and rising nonfinancial income.

Households' average net flows into fixed income, depicted in the bottom panel of Figure 6, do not exhibit the same strong procyclical patterns as their flows into stocks. Average net flows into fixed income were mostly positive from the late 1990s through the late 2000s, with the largest inflows over the full sample period observed around the time of the late 2000s recession. From the 2010s onwards these average net flows were smaller in magnitude and varied between positive and negative values across years. Households flowed out of fixed income during the low interest rate period following the start of the COVID-19 pandemic, then flowed back in during the following years of rising interest rates and falling aggregate stock prices in 2022.

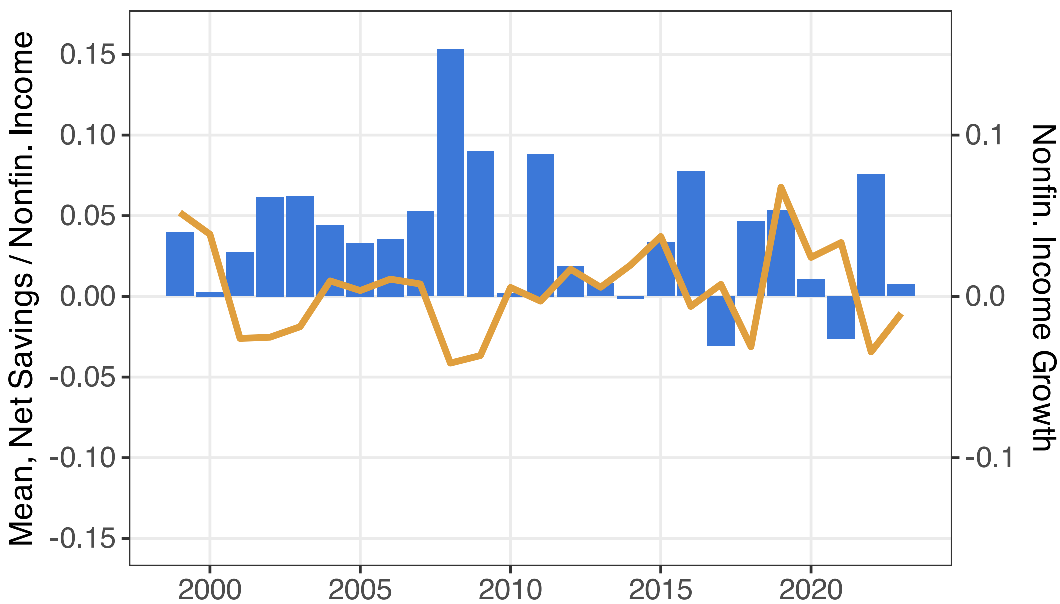
Moving beyond averages, Figure 7 depicts percentiles of the same scaled savings flows across households, by year and asset class. These estimated savings flows are highly dispersed, with both substantial inflows and outflows observed across different households in every year. The 10th and 90th percentiles flows are large in magnitude compared to the average flows shown in Figure 6. However, the most striking feature of Figure 7 is how the tails of the distribution of individual flows into stocks (depicted in the top panel) varies over time. During economic downturns such as the recessions of the early and late 2000s, the 10th percentile of the estimated flows falls sharply as many households shift out of stocks in their taxable accounts and liquidate their existing holdings. During expansions the 90th percentile rises, as many households accumulate significant additional stockholdings in their taxable accounts. These

Figure 6: Average Savings Flows and Income Growth

(a) Stocks



(b) Fixed Income



Notes: This figure depicts mean savings flows by asset class (bars) and changes in nonfinancial income (line) across households in the full sample in each year. For each tax year, each imputed net savings flow $S^A_{j,t,t+1}$ — measured from the change in dividend or interest income relative to the prior tax year — is scaled by average nonfinancial income $\overline{Y}_{j,t-2:t}$ for the same household j over the prior three tax years before computing the mean scaled flow across households. The mean change $(Y_{j,t+1} - Y_{j,t})/\overline{Y}_{j,t-2:t}$ in nonfinancial income relative to the prior tax year, scaled by the same prior three-year average, is computed similarly. Each bar is centered at the transition date between the two tax years used to impute savings, e.g. the bar centered at $t = 2010$ represents flows imputed from income in tax years 2010 and 2011 that primarily reflect transactions taking place around January 1, 2010. Points along the line are aligned similarly. Average income from 1998 to 2000 is used for scaling at the beginning of the sample period when three prior years are not available. Panel (a) reports flows for dividend-generating assets. Panel (b) reports flows for interest-generating assets (both taxable and nontaxable).

patterns hold even around the more moderate run-up and subsequent decline in aggregate stock prices over the 2021-2022 period. The median net flow into stocks is close to zero over the full sample period, implying that much of the cyclical variation in the average flows from Figure 6 is driven by large outflows among some households during downturns and large inflows during expansions, rather than broad but moderate variation in flows across all households.

The distribution of estimated individual flows into stocks varies over the business cycle in a way that mirrors the distribution of individual labor income. In seminal work using administrative wage earnings data from the Social Security Administration, Guvenen, Ozkan, and Song (2014) showed that the skewness of the distribution of individual labor income growth varies significantly over the business cycle. Recessions coincide with a sharp increase in the share of individuals experiencing large earnings declines — especially for individuals at the top of the pre-recession earnings distribution — while expansions coincide with an increase in the share of individuals experiencing large earnings increases. At the same time, many individuals in a typical year have relatively little change in their earnings relative to the prior year, and median earnings growth is considerably more stable than mean earnings growth over the business cycle as a result. These features of the earnings growth distribution are all similar to the features of the distribution of individuals' estimated net flows into stocks depicted in the top panel of Figure 7. In contrast, the distribution of flows into fixed income across individuals (depicted in the bottom panel) does not vary as strikingly across recessions and expansions.

4.2 Household Savings Responses to Income Shocks

The descriptive analysis presented in the previous section provides suggestive evidence that fluctuations in households' nonfinancial income — from jobs or closely-held private businesses — may play a role in generating the large dispersion in estimated savings flows across households, and the cyclical variation in average flows into the stock market across years. In this section I formally analyze the savings response of individual households to persistent unanticipated shocks affecting their nonfinancial income. These estimated responses provide useful information for discriminating between alternative models of households' consumption-savings behavior. I focus on households' response to persistent income shocks given the empirical evidence about the persistence of individual earnings losses in recessions (Guvenen, Ozkan, and Song, 2014).

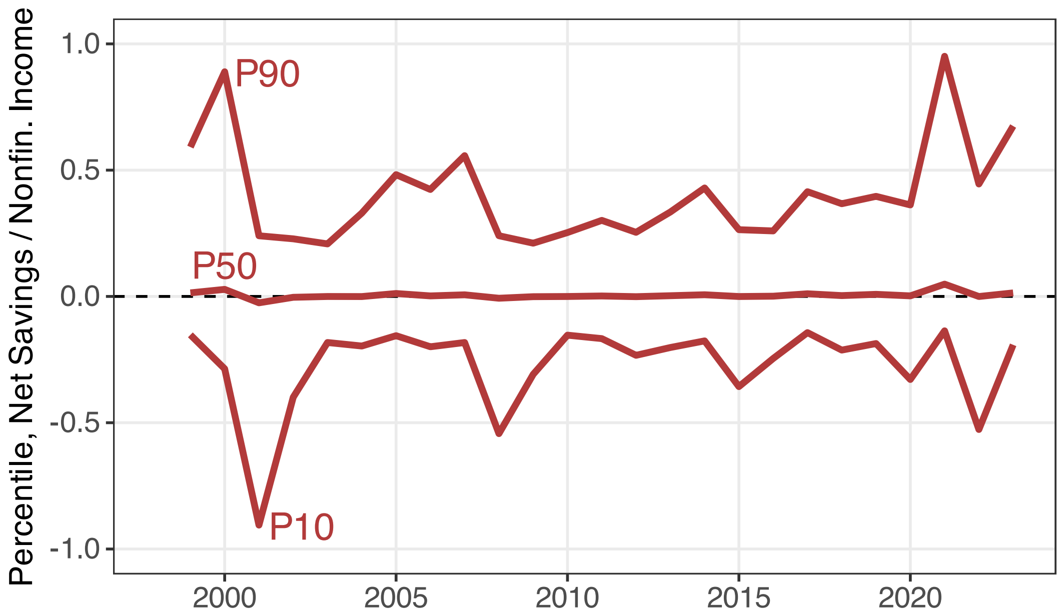
I estimate these household savings responses using panel regressions of the following form:

$$\frac{S^A_{j,t,t+h}}{\overline{Y}_{j,t-2,t}} = \beta^A_h \cdot \frac{\overline{Y}_{j,t+1,t+3} - \overline{Y}_{j,t-2,t}}{\overline{Y}_{j,t-2,t}} + \gamma_h' Z_{j,t+h} + \varepsilon_{j,t+h} \tag{3}$$

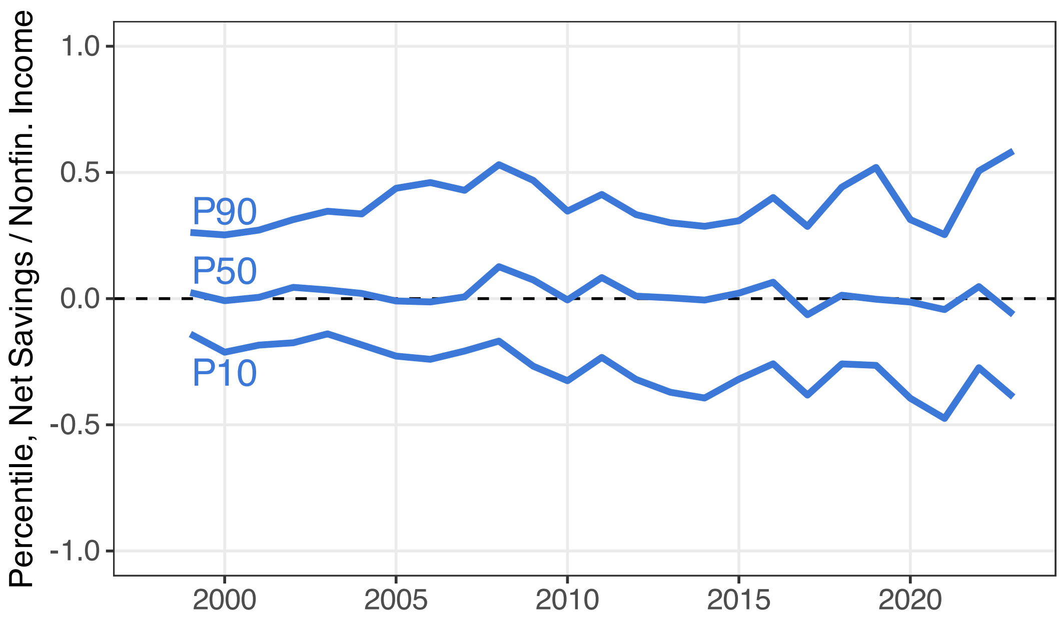
On the left-hand side, $S^A_{j,t,t+h}$ denotes the cumulative imputed net savings flow in asset class A for household j, measured from income spanning years t to $t+h$. This cumulative flow is defined by summing the savings flows measured over each two-year window: $S^A_{j,t,t+h} \equiv \sum_{k=1}^{h} S^A_{j,t+k-1,t+k}$. On the right-hand side, $\overline{Y}_{j,t+1,t+3}$ denotes average nonfinancial income for household j measured over years $t+1$ to $t+3$. I compute the change in this average income

Figure 7: Distribution of Individual Savings Flows

(a) Stocks



(b) Fixed Income



Notes: This figure depicts the 10th, 50th (median), and 90th percentiles of savings flows (scaled by prior income) across households in each year. For each tax year, each imputed net savings flow $S^A_{j,t,t+1}$ — measured from the change in dividend or interest income relative to the prior tax year — is scaled by average nonfinancial income $\overline{Y}_{j,t-2:t}$ for the same household j over the prior three tax years before computing percentiles of scaled flows across households. Points along the line are aligned with the transition date between the two tax years used to impute savings, e.g. the point at $t = 2010$ represents flows imputed from income in tax years 2010 and 2011 that primarily reflect transactions taking place around January 1, 2010. Average income from 1998 to 2000 is used for scaling at the beginning of the sample period when three prior years are not available. Panel (a) reports flows for dividend-generating assets. Panel (b) reports flows for interest-generating assets (both taxable and nontaxable).

relative to the prior three-year average $\overline{Y}_{j,t-2,t}$. I then scale both this income change on the right-hand side, and the savings flow on the left-hand side, by prior average income $\overline{Y}_{j,t-2,t}$. The common scaling makes the coefficient β_h interpretable in "dollar-on-dollar" terms, as a cumulative marginal propensity to save. A household faced with a persistent $1 decrease in their nonfinancial income in each year going forward would be expected to reduce their total net savings over the h years following the initial shock by $\$\beta_h$, either through active drawdowns on their existing liquid assets or smaller savings inflows. The vector $Z_{j,t+h}$ includes additional controls and time-interacted fixed effects, based on household characteristics observed through time t.

There are several reasons why ordinary least squares estimates of Equation (3), using realized changes in households' nonfinancial income as the main regressor, are likely to be uninformative about households' savings responses to the types of persistent unanticipated income shocks that are realized over the business cycle. The main concern is that many large changes in household income may be either anticipated, or correlated with other changes in the household's consumption preferences and liquidity needs. For example, consider the case of a married couple where one spouse voluntarily exits the labor force for several years to care for their childen. This event would lead to a large change in the household's total nonfinancial income. However, it may not necessarily lead to a large unexpected change in the household's liquidity needs. Since this income loss may be anticipated years in advance, the couple can accumulate savings in safe assets to offset the expected gap between their income and expenditures over this period. Second, the household's expenditure needs may fall as certain market consumption expenditures — such as private child care expenditures and food away from home — can be substituted with home production. Evidence from various survey and administrative datasets suggests that many large individual income changes are anticipated or driven by life events that are likely to shift households' consumption preferences and liquidity needs.[23]

To overcome these challenges, I estimate Equation (3) via two-stage least squares. I construct a shock which is more likely to be unanticipated and exogenous to each household's consumption preferences and liquidity needs, and use this to instrument for nonfinancial income growth. Specifically, I exploit the fine geographic information available on each individual's income tax return and construct a shock $\widetilde{g}_{z,t+1}$ from realized nonfinancial income growth in each ZIP-code z. Specifically:

1. I first take the set $\mathcal{J}_{z,t}$ of all working-age individuals (age 30 to 59) from ZIP-code z in year t, who satisfy the criteria for available tax returns with consistent filing status over years $t-2$ to $t+3$ used in constructing the main sample (as described in Section 3.1).

[23]For example, Caplin, Gregory, Lee, Leth-Petersen, and Sæverud (2023) and Wang (2022) use data on subjective expectations of individual earnings growth to argue that anticipated changes (rather than unanticipated idiosyncratic shocks) account for a large share of the cross-sectional dispersion in annual earnings growth observed in administrative datasets. Larrimore, Mortenson, and Splinter (2016) show that various life events which may affect household expenditure needs (such as the birth of a child) are associated with large absolute income changes, while Aguiar and Hurst (2005) show that households adjust to the large but predictable earnings decline at retirement by reducing their market consumption expenditures and substituting towards home production.

2. I compute the average of nonfinancial income for this set of individuals in each year $t-2$ through $t+1$, e.g.:

$$\mathbf{Y}_{z,t+1} \equiv \frac{1}{|\mathcal{J}_{z,t}|} \sum_{j \in \mathcal{J}_{z,t}} Y_{j,t+1}$$

   I also compute average income $\overline{\mathbf{Y}}_{z,t-2,t}$ over the prior three years.

3. I compute growth $g_{z,t+1}$ of average income in year $t+1$ relative to the three-year base period:

$$g_{z,t+1} \equiv (\mathbf{Y}_{z,t+1} - \overline{\mathbf{Y}}_{z,t-2,t})/\overline{\mathbf{Y}}_{z,t-2,t}$$

4. I construct the residualized growth rate $g^*_{z,t+1}$ from regressing $g_{z,t+1}$ on two lags of similarly-scaled changes in annual average income, $\Delta\mathbf{Y}_{z,t}/\overline{\mathbf{Y}}_{z,t-2,t}$ and $\Delta\mathbf{Y}_{z,t-1}/\overline{\mathbf{Y}}_{z,t-2,t}$.

5. The final shock is defined as the negative part of the residualized growth rate: $\widetilde{g}_{z,t+1} \equiv \min(g^*_{z,t+1}, 0)$. Only the negative part is kept so that the instrument identifies the savings response to large income declines (or lower-than-expected growth) rather than large income increases.

There are several key advantages to using this ZIP-code-based instrument. The first advantage is statistical power: this instrument can be computed for nearly every individual in the sample, and the dispersion of shocks across ZIP-codes within a given year is high. The second advantage is that the residualization procedure described above helps to purge predictable fluctuations in local income growth, leaving an instrument that is more likely to capture unanticipated shocks based on changes in local economic conditions or among individuals in the same ZIP-code working in similar industries or occupations. The third advantage is that by averaging over many individuals in each ZIP-code, the resulting shock is more likely to be exogenous to the types of individual life events that shift a household's consumption preferences and liquidity needs.

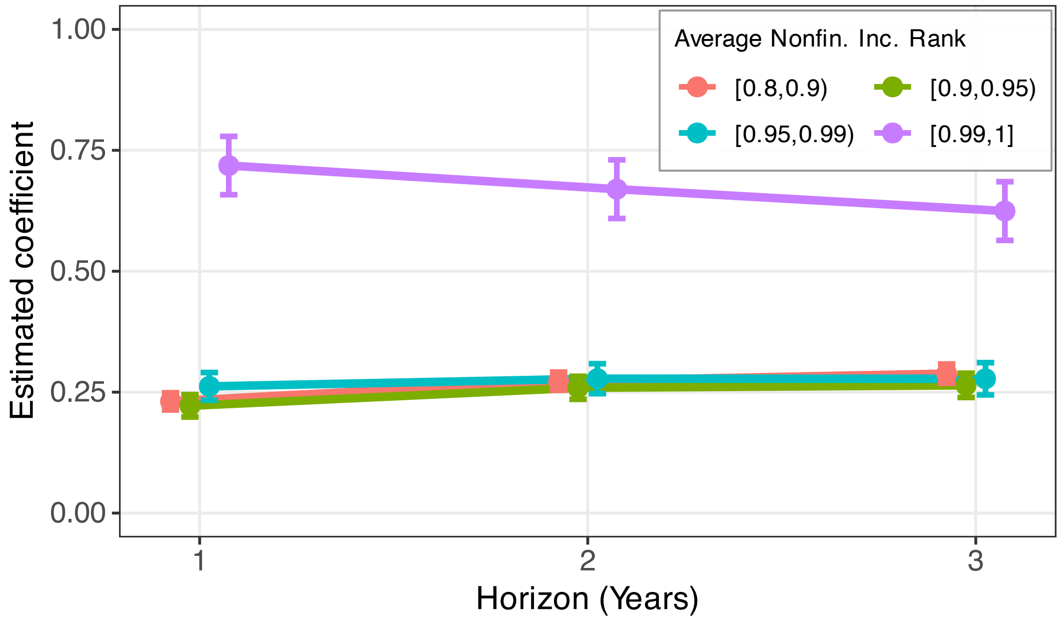
In first-stage regressions, I estimate the income growth exposure of individual households j to the shock $\widetilde{g}_{z(j),t+1}$ in their ZIP-code $z(j)$:

$$\frac{\overline{Y}_{j,t+1,t+h} - \overline{Y}_{j,t-2,t}}{\overline{Y}_{j,t-2,t}} = \pi_h \cdot \widetilde{g}_{z(j),t+1} + \Gamma'_h Z_{j,t+h} + u_{j,t+h} \tag{4}$$

This regression is estimated separately by horizon h — ranging from $h = 1$ to $h = 3$ years — and for individuals in each of the four separate income groups. The set of controls $Z_{j,t+h}$ is the same as used in the second-stage regressions (described below). Figure 8 displays the estimated coefficients $\widehat{\pi}_h$. The estimated passthrough from these local shocks to individual income growth is large, persistent, and highly statistically significant. Top earners are highly exposed to these shocks: the estimated passthrough is twice as large for individuals in the top 1% of the prior earnings distribution relative to the rest of the sample.

Using this instrument $\widetilde{g}_{z(j),t+1}$, I estimate the second-stage regression described in Equation (3). The large sample size and rich set of information available about each household in the tax data allows for a granular set of controls that includes multiple lags of income growth, wealth,

Figure 8: First-Stage Estimates: Individual Exposures to Local Income Shocks



Notes: This figure reports first-stage estimates for each income group of the passthrough from ZIP-code level shocks $\widetilde{g}_{z(j),t+1}$ to individual income growth. The first-stage regression is described in Equation (4). The horizontal axis reports the horizon of h years over which post-period average income $\overline{Y}_{j,t+1,t+h}$ is measured. The vertical axis reports the estimated coefficient $\widehat{\pi}_h$ at each horizon, estimated separately for each of the four income groups. The full set of controls and fixed effects is listed in footnote 24. Bars report ± 2 standard errors, which are clustered by individual j and ZIP-code z.

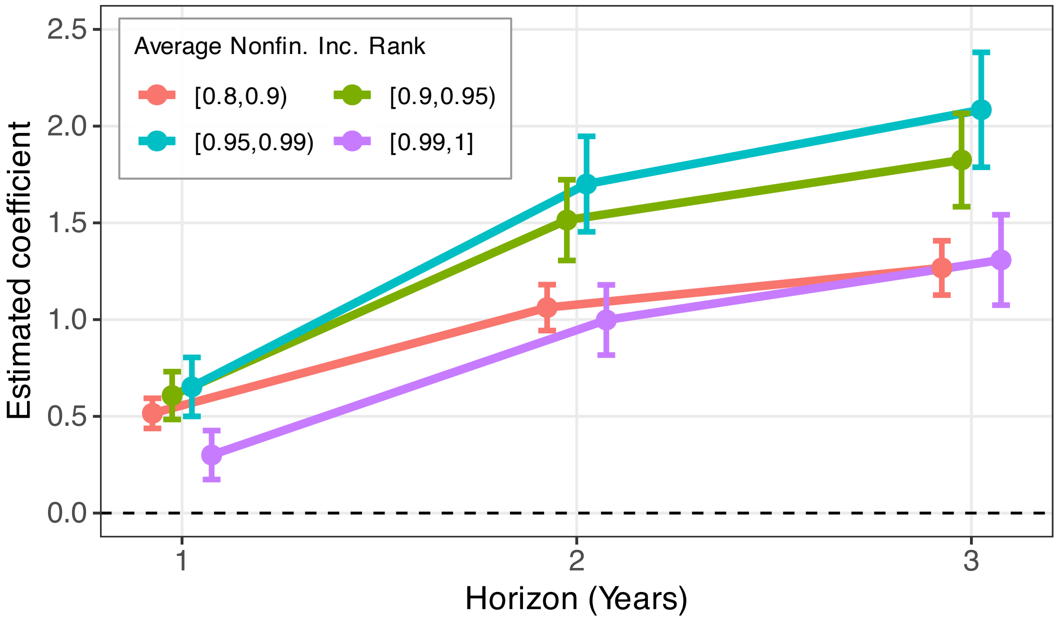
savings flows, and rich time-interacted fixed effects.[24] Intuitively, the second-stage regressions trace out the differences in savings flows over time between ex-ante similar individuals who live in ZIP-codes hit by different ex-post economic shocks.

The second-stage regression in Equation (3) uses growth in pre-tax income as the endogenous regressor. However, households ultimately save in liquid assets and consume out of their *post-tax* income. In order to interpret the estimated coefficient as a marginal propensity to save out of available post-tax income, I divide the estimated coefficients $\widehat{\beta}_h$ by $(1 - \tau_{marg})$, where τ_{marg} is the marginal federal income tax rate for each income group. Because I do not directly observe taxes paid or tax rates in the income tax data, I estimate these marginal

[24] The full set of controls and fixed effects included in all first-stage and second-stage regressions consists of:

- Three-way fixed effects by year, 1-year age group, and wealth-to-income groups
- Four-way fixed effects by year, 5-year age group, nonfinancial income percentile within 1-year age group, and wealth-to-income groups
- Piecewise linear splines in two lags of changes in one-year nonfinancial income, scaled by average initial income. The spline breakpoints are set at -0.5, 0, and +0.5.
- Piecewise linear splines in two lags of imputed net savings for each asset class (dividends, taxable interest and nontaxable interest), scaled by average initial income. The spline breakpoints are set at -0.5, 0, and +0.5.
- Interactions between two lags of capitalized financial wealth by asset class, scaled by average initial income, interacted with two-way fixed effects by 5-year age group and total wealth/income group

Figure 9: Second-Stage Estimates: Savings Responses to Persistent Income Shocks



Notes: This figure reports second-stage estimates for each income group of the savings response to persistent income shocks. The second-stage regression is described in Equation (3). ZIP-code level shocks $\widetilde{g}_{z(j),t+1}$ are used as an instrument for realized nonfinancial income growth in the first-stage regression described in Equation (4), at a horizon of $h = 3$ years. The dependent variable is cumulative imputed net savings in all financial assets (both stocks and fixed income) $S^{tot}_{j,t,t+h}$ through year $t + h$, scaled by average nonfinancial income $\overline{Y}_{j,t-2:t}$ over tax years $t - 2$ to t. The horizontal axis reports the horizon of h years over which the savings response is measured. To convert estimates to a marginal propensity to save out of post-tax income, marginal tax rates τ_{marg} are estimated for each income group using data from the SCF and NBER's TAXSIM calculator. The vertical axis reports the estimated coefficient $\widehat{\beta}_h$ at each horizon divided by $(1 - \tau_{marg})$, estimated separately for each of the four income groups. The full set of controls and fixed effects is listed in footnote 24. Bars report ±2 standard errors, which are clustered by individual j and ZIP-code z.

tax rates for each income group using data from the Survey of Consumer Finances and the NBER's TAXSIM calculator.

Figure 9 displays the estimated tax-adjusted coefficients from the second-stage regressions, with total imputed net savings $S^{tot}_{j,t,t+h}$ in all liquid financial assets as the dependent variable at each horizon. These estimates imply a large marginal propensity to save (or draw down on existing savings) out of persistent income shocks. In response to a shock that lowers a household's annual post-tax wage or business income by \$1000 in each future year, households cut their active savings and draw down on their existing liquid assets by between \$500 and \$850 in each year. Most of the cumulative savings response over the full three-year period after the initial shock is accounted for by the response in the first two years, indicating that households may eventually rely on other margins of adjustment (such as reducing expenditures) in the longer-run.

The large household savings responses to persistent income shocks that I estimate are at odds with the predictions of canonical consumption-savings models that predict *spending* should serve as the main margin of adjustment in response to such shocks. In the simplest configuration of the deterministic permanent income hypothesis model, in which the time discount factor equals the inverse of the gross risk-free interest rate, a one-time permanent shock to the level of income would be expected to generate *no* savings response, with households adjusting spending one-for-one in every year that follows. In the following section I show that consumption adjustment frictions provide one way to rationalize these large estimated savings responses, and trace out the implications of these frictions for households' optimal portfolio choices.

5 Portfolio Choice Model

In this section I present the life-cycle portfolio choice model. After describing the structure of the model, I compare the model-implied household savings response to persistent income shocks against the estimates obtained from the income tax data. Large consumption adjustment frictions are needed to match this large estimated savings response. Finally, I study the model's implications for optimal portfolio shares and wealth accumulation over the life cycle given these adjustment costs.

5.1 Model Description

Life cycle and household preferences: I model the decisions of households from labor market entry at age $a_0 = 25$ up to a maximum age of $a_{max} = 99$. I use subscript j to denote an individual household, $t_0(j)$ to denote the period when the household enters the labor market, and $t_{max}(j) \equiv t_0(j) + a_{max} - a_0$ to denote the period when the household reaches the maximum age. The household works and receives nonfinancial income in each period from the time it enters the labor market through the time when it reaches the retirement age $a_{ret} = 60$ in period $t_{ret}(j) \equiv t_0(j) + a_{ret} - a_0$.

The household enjoys flow consumption $C_{j,t}$ in each period t. It has Epstein-Zin-Weil preferences (Epstein and Zin, 1989; Weil, 1990) over these consumption streams, with utility over current and future consumption represented by a value function $V_{j,t}$ that satisfies the recursion:

$$V_{j.t} = \left\{ (1-\beta) C_{j,t}^{1-\psi^{-1}} + \beta E_t \left[V_{j,t+1}^{1-\gamma} \right]^{\frac{1-\psi^{-1}}{1-\gamma}} \right\}^{\frac{1}{1-\psi^{-1}}} \tag{5}$$

Here $\beta \in (0,1)$ denotes the household's time discount factor, $\gamma > 0$ denotes the coefficient of relative risk aversion, and $\psi > 0$ denotes the elasticity of intertemporal substitution. The terminal condition for periods after the household reaches the maximum age a_{max} is $E_{t_{max}(j)} \left[V_{j,t_{max}(j)+1}^{1-\gamma} \right] = 0$, as commonly used in life-cycle models with recursive preferences (Cocco, Gomes, and Maenhout, 2005; Choukhmane and de Silva, 2025; Calvet, Campbell, Gomes, and Sodini, 2025).

Asset Returns: The household can invest its liquid wealth in two financial assets: a risk-free one-period bond with constant time-invariant log return r_f, and a risky stock (representing an aggregate stock market fund) with realized log return r_{t+1} over the holding period from t to $t+1$. Expected stock returns vary over time, as in classic models of intertemporal portfolio choice (Merton, 1973; Campbell and Viceira, 1999; Barberis, 2000) and consistent with key facts established in the empirical asset pricing literature (Campbell and Shiller, 1988; Cochrane, 2011).

I closely follow Campbell (1991) and Campbell (2018) in specifying the dynamics of realized and expected stock returns. Investment opportunities are summarized by the state variable x_t that represents the one-period-ahead equity premium. x_t evolves according to a first-order autoregressive process:

$$x_{t+1} = \overline{x} + \phi_x (x_t - \overline{x}) + \xi_{x,t+1} \tag{6}$$

Here $\overline{x}$ denotes the unconditional mean of x_t, $\phi_x \in [0, 1)$ denotes its first-order autocorrelation, and $\xi_{x,t+1}$ denotes the innovation. The level of x_t determines the one-period-ahead expected log stock return:

$$E_t[r_{t+1}] = r_f + x_t - \frac{\sigma_r^2}{2} \quad \Leftrightarrow \quad x_t = E_t[r_{t+1}] + \frac{\sigma_r^2}{2} - r_f \tag{7}$$

where $\sigma_r^2 \equiv Var_t\left(r_{t+1}\right)$ denotes the conditional variance of the realized future return. The shocks determining the unexpected component of r_{t+1} in the definitions below all have constant variance over time, so that the conditional variance $Var_t\left(r_{t+1}\right)$ is constant over time and σ_r^2 can be written without a time subscript.

The unexpected component of the realized log stock return r_{t+1} is determined by separate cash flow news and discount rate news components:

$$r_{t+1} - E_t\left[r_{t+1}\right] = N_{CF,t+1} - N_{DR,t+1} \tag{8}$$

The discount rate news component $N_{DR,t+1}$ reflects innovations to expected future stock returns, based on the Campbell and Shiller (1988) decomposition for the realized log return:

$$N_{DR,t+1} \equiv \left(E_{t+1} - E_t\right) \left[\sum_{h=1}^{+\infty} \rho_{CS}^h \; r_{t+1+h}\right] \tag{9}$$

The parameter ρ_{CS} from the Campbell-Shiller approximation is set to $\rho_{CS} = 0.96$, consistent with typical calibrations for annual asset returns.[25] Given the dynamics of the one-period-ahead equity risk premium x_t specified in Equation (6), $N_{DR,t+1}$ is determined entirely by its innovation $\xi_{x,t+1}$:

$$N_{DR,t+1} = \frac{\rho_{CS}}{1 - \rho_{CS} \cdot \phi_x} \xi_{x,t+1} \tag{10}$$

A positive innovation $\xi_{x,t+1} > 0$ to the equity premium x_{t+1} raises expectations of *future* stock returns, but lowers the *contemporaneous* stock return r_{t+1}.

In addition to discount rate news, the realized log stock return r_{t+1} also reflects a separate

[25] See Chapter 5.3 of Campbell (2018). The value $\rho_{CS} = 0.96$ for the log-linearization parameter in the Campbell-Shiller decomposition implies a steady-state value for the dividend-price ratio of roughly 4%.

cash flow news component $N_{CF,t+1}$. In the Campbell-Shiller decomposition this component reflects revisions in expectations about stock dividends (cash flows) paid in periods $t+1$ and onwards, and is uncorrelated with all past information available through period t by construction. Rather than directly modeling the time series process for stock dividends, the distribution of $N_{CF,t+1}$ is specified directly in order to obtain an appropriate distribution for the realized stock return r_{t+1}.

While the two components $N_{DR,t+1}$ and $N_{CF,t+1}$ of the realized return are unpredictable from all information available to the household at time t, these components may be correlated both with each other, and with contemporaneous shocks to the household's nonfinancial income. Stock market crashes are incorporated into the model by jointly drawing the realized return components from a mixture over two multivariate normal distributions with different means:

$$(N_{CF,t}, N_{DR,t})' \sim \begin{cases} N\left((\mu_{CF-}, \mu_{DR-})', \widetilde{\Sigma}_{CF,DR}\right) & \text{with probability } p_{r-} \\ N\left((\mu_{CF+}, \mu_{DR+})', \widetilde{\Sigma}_{CF,DR}\right) & \text{with probability } 1-p_{r-} \end{cases} \tag{11}$$

The mixture probability p_{r-} determines the probability of a large stock market crash, while μ_{CF-} and μ_{DR-} correspond to the expected values of the cash flow and discount rate news components (respectively) of the market return conditional on a crash occurring. If no crash occurs then the expected values of these two components are given by μ_{CF+} and μ_{DR+}, with $\mu_{CF+} > \mu_{CF-}$ and $\mu_{DR-} < \mu_{DR+}$. $\widetilde{\Sigma}_{CF,DR}$ represents the covariance matrix of cash flow and discount rate news conditional on the mixture draw (i.e. the covariance between $N_{CF,t}$ and $N_{DR,t}$ across all crash periods, or across all non-crash periods).[26]

Nonfinancial income: The household works and receives pre-tax nonfinancial income $Y_{j,t}$ in each period from the age a_0 that it initially enters the labor market through age a_{ret} when it retires. Consumption in retirement, from age $a_{ret}+1$ to a_{max}, is financed by the household's liquid financial wealth $W_{j,t}$ and flows from illiquid assets (described below). This income represents the wages or profits that accrue to the household from a job or closely-held private business. This income is modeled as an exogenous process disciplined by administrative data, rather than explicitly incorporating the household's labor supply choice and a more detailed description of the labor market or entrepreneurship.

Income $Y_{j,t}$ in period t is determined by a persistent component $\overline{Y}_{j,t}$ and transitory shock $\varepsilon_{j,t}$:

$$Y_{j,t} = \overline{Y}_{j,t} \exp(\varepsilon_{j,t}) \tag{12}$$

The transitory shock $\varepsilon_{j,t}$ is drawn independently from all other shocks to income and asset returns, from a normal distribution with mean zero and variance σ_ε^2.[27]

[26]For simplicitly, the two mixture components are restricted to have the same covariance matrix. The model could be easily modified to allow these covariances to differ across the two mixture components, although it may be more difficult to discipline this richer parameterization using the data.

[27]It would be straightforward to allow for a state-dependent distribution of the transitory income shock, similar to the state-dependent distribution of the persistent income shock $\eta_{j,t}$.

The log of the persistent component $\overline{Y}_{j,t}$ follows a random walk with age-dependent drift $g_{a_j(t)}$, where $a_j(t)$ denotes the household's age in period t:

$$\log \overline{Y}_{j,t} = \log \overline{Y}_{j,t-1} + g_{a_j(t)} + \eta_{j,t} \tag{13}$$

The persistent income shock $\eta_{j,t}$ satisfies $E_{t-1}[\eta_{j,t}] = 0$ and is drawn from a normal mixture distribution that varies with the contemporaneous realizations of both the cash flow news $N_{CF,t}$ and discount rate news $N_{DR,t}$ components of the realized stock return r_t:

$$\eta_{j,t} \sim \begin{cases} N\left(\mu_{\eta-}(N_{CF,t}, N_{DR,t}), \widetilde{\sigma}_\eta^2\right) & \text{with probability } p_{\eta-} \\ N\left(\mu_{\eta+}(N_{CF,t}, N_{DR,t}), \widetilde{\sigma}_\eta^2\right) & \text{with probability } 1 - p_{\eta-} \end{cases} \tag{14}$$

where $p_{\eta-}$ can be interpreted as the probability of infrequent but adverse employment events, such as job separations or persistent shocks to firm profitability that pass through to worker pay. The mixture component means are given by:

$$\begin{aligned} \mu_{\eta-}(N_{CF,t}, N_{DR,t}) &= \overline{\mu}_{\eta-} + \lambda_{\eta,CF-} N_{CF,t} + \lambda_{\eta,DR-} N_{DR,t} \\ \mu_{\eta+}(N_{CF,t}, N_{DR,t}) &= \overline{\mu}_{\eta+} + \lambda_{\eta,CF+} N_{CF,t} + \lambda_{\eta,DR+} N_{DR,t} \end{aligned} \tag{15}$$

where the intercept parameters $(\overline{\mu}_{\eta-}, \overline{\mu}_{\eta+})$ determine the average values of the mixture component means in a typical period, while the slope parameters $(\lambda_{\eta,CF-}, \lambda_{\eta,DR-}, \lambda_{\eta,CF+}, \lambda_{\eta,CF+})$ govern the dependence of the persistent income shock distribution on the contemporaneous stock return components.[28] This mixture specification captures two crucial features of individual labor income growth documented by Guvenen, Ozkan, and Song (2014), Guvenen, Schulhofer-Wohl, Song, and Yogo (2017), and Guvenen, Karahan, Ozkan, and Song (2021) using administrative earnings data. The first is the average skewness and high kurtosis of individual income growth: in a typical year a large share of individuals experience relatively little change in their earnings relative to the prior year, while a small subset experience large positive or negative changes.[29] The second is how the mean and skewness of the income growth distribution varies over time, in particular following large declines in aggregate stock returns: the distribution of individual income growth becomes more negatively-skewed as the share of individuals experiencing large earnings losses rises, while the share experiencing large earnings

[28]Because the cash flow and discount rate news components of the realized stock return satisfy $E_t[N_{DR,t+1}] = 0$ and $E_t[N_{CF,t+1}] = 0$, the mixture component means satisfy $E_t[\mu_{\eta-}(N_{CF,t+1}, N_{DR,t+1})] = \overline{\mu}_{\eta-}$ and $E_t[\mu_{\eta+}(N_{CF,t+1}, N_{DR,t+1})] = \overline{\mu}_{\eta+}$. The distribution of the persistent income shock $\eta_{j,t+1}$ depends only on the realization of the *contemporaneous* stock return r_{t+1}. The model could be extended to allow the distribution of $\eta_{j,t+1}$ to vary with other aggregate state variables such as the equity risk premium x_t. Schmidt (2025) presents a general equilibrium asset pricing model where this relationship arises endogenously as the result of exogenous time-varying idiosyncratic earnings risk. Additionally, heteroskedasticity in the conditional distributions of $N_{CF,t+1}$ and $N_{DR,t+1}$ (which are drawn from a time-invariant distribution in the model presented here) would generate time-varying income risk through the dependence of the mixture means in Equation (15).

[29]Although the same variance $\widetilde{\sigma}_\eta^2$ is used for both components of the mixture distribution in Equation (14), the randomly-drawn mixture mean generates excess skewness and kurtosis relative to a normal distribution with the same variance equal to $Var_t(\eta_{j,t}) > \widetilde{\sigma}_\eta^2$.

gains falls. The mixture distribution used here, and its dependence on realized stock returns, is similar to the specification used by Catherine (2022).

Consumption adjustment costs: In contrast to traditional consumption-savings models where flow consumption $C_{j,t}$ is chosen freely each period, households in the model face frictions in reducing consumption below its level $C_{j,t-1}$ from the previous period. These frictions take the form of an asymmetric quadratic adjustment cost paid by the household in period t, as a function of its consumption in the current and previous periods:

$$\Phi_C(C_{j,t}, C_{j,t-1}) \equiv \frac{\varphi_C}{2} \left[\min\left(\frac{C_{j,t}}{C_{j,t-1}} - 1, 0 \right) \right]^2 C_{j,t-1} \tag{16}$$

The quadratic functional form is similar to the functional form used for price and wage adjustment costs in macroeconomics (Rotemberg, 1982) and for capital adjustment costs in corporate finance (Strebulaev and Whited, 2012). A useful interpretation of Equation (16) is that the household must pay a fraction of the prior consumption level $C_{j,t-1}$ to reduce $C_{j,t}$ below this level, with the fraction $\Phi_C(C_{j,t}, C_{j,t-1})/C_{j,t-1}$ determined solely by the consumption growth rate $C_{j,t}/C_{j,t-1}$. Increasing $C_{j,t}$ above its prior level does not entail any immediate adjustment cost, although optimizing households will internalize *future* adjustment costs that they anticipate paying when choosing their current consumption level.

The parameter $\varphi_C \geq 0$ determines the overall magnitude of the adjustment costs. A value of $\varphi_C = 5$, for example, implies that reducing consumption by 10% relative to its previous level $C_{j,t-1}$ requires paying an additional cost equal to 2.5% of $C_{j,t-1}$. Larger reductions in consumption entail greater adjustment costs: for the same parameter values, a 20% reduction in consumption requires paying an additional cost equal to 10% of $C_{j,t-1}$.

Illiquid assets: To keep the model as parsimonious as possible while focusing on how working-age households allocate their liquid wealth, I employ a similar approach to Gourinchas and Parker (2002) in modeling relatively illiquid forms of wealth such as housing and tax-advantaged retirement accounts (e.g. 401(k)s and IRAs), and other retirement income streams such as pensions and Social Security.

During working life — between ages a_0 and a_{ret} — households invest a fixed fraction $s \in [0, 1)$ of their pre-tax income $Y_{j,t}$ in illiquid assets in each period t. This fraction captures forms of investment that increase households' net worth but are difficult or costly to liquidate before retirement age, such as contributions to tax-advantaged retirement accounts and housing purchases or mortgage principal payments that increase home equity. It also captures other payments made by working age households, such as other pension contributions and Social Security payroll taxes, that contribute to future retirement income streams.

In retirement — from ages $a_{ret} + 1$ to a_{max} — households receive flow income each period from these illiquid assets and other retirement income streams, in addition to returns on their liquid wealth $W_{j,t}$. This flow income captures cash payments the household receives such as distributions from tax-advantaged retirement accounts, income received from other pensions, and Social Security payments. It can also be thought to capture rental income from real estate properties owned by the household, or the owner-equivalent rent from the household's primary residence.[30] The flow income received each period in retirement is equal to $S\overline{Y}_{j,t_{ret}(j)}$, where $\overline{Y}_{j,t_{ret}(j)}$ is the level of persistent earnings in the last period that the household works and $S \geq 0$ is a constant. The terminal persistent earnings $\overline{Y}_{j,t_{ret}(j)}$ serve as a proxy for the household's cumulative earnings throughout working age, which determines the amount contributed to retirement accounts and benefits received from other pensions and Social Security. Indexing this flow income to $\overline{Y}_{j,t_{ret}(j)}$ avoids the need to track cumulative earnings as an additional state variable, but has the drawback of overstating uncertainty about retirement income for households that are very close to retirement.

Liquid wealth dynamics and budget constraint: The household enters period t with liquid wealth $W_{j,t}$, after the realization of the stock return r_{t+1}. During working age it receives pre-tax nonfinancial income $Y_{j,t}$. A fraction s of this income is invested in illiquid assets and a fraction $\tau \in (0,1)$ is paid in taxes, leaving the household with $(1-s-\tau)Y_{j,t}$ to allocate towards consumption and liquid savings. In retirement the household receives flow income $S\overline{Y}_{j,t_{ret}(j)}$ from its illiquid asset holdings. I use $M_{j,t}$ to denote the total resources available to the household, or "cash-on-hand", which is defined as the sum of its liquid financial wealth and any additional funds available to consume and invest in liquid assets:

$$M_{j,t} = W_{j,t} + \begin{cases} (1-s-\tau)Y_{j,t} & \text{while working:} \quad t \leq t_{ret}(j) \\ S\overline{Y}_{j,t_{ret}(j)} & \text{while retired:} \quad t > t_{ret}(j) \end{cases} \tag{17}$$

Given $M_{j,t}$, the household chooses its consumption level $C_{j,t}$. It pays adjustment costs $\Phi_C(C_{j,t}, C_{j,t-1})$ based on its prior consumption level $C_{j,t-1}$. Its total expenditure is thus $C_{j,t} + \Phi_C(C_{j,t}, C_{j,t-1})$, and the remainder $\widetilde{W}_{j,t} \equiv M_{j,t} - C_{j,t} - \Phi_C(C_{j,t}, C_{j,t-1})$ is the household's liquid wealth balance at the end of the period.

The household allocates its liquid wealth $\widetilde{W}_{j,t}$ between stocks and bonds by choosing its stock portfolio share $\theta_{j,t}$. The realized log stock return r_{t+1} then determines the household's gross portfolio return from period t to $t+1$, and its liquid wealth balance at the start of period $t+1$:

$$W_{j,t+1} = \underbrace{[\theta_{j,t}\exp(r_{t+1}) + (1-\theta_{j,t})\exp(r_f)]}_{\text{gross portfolio return}} \widetilde{W}_{j,t} \tag{18}$$

[30]Since the model includes a single homogenous consumption good with adjustment costs, the household's flow consumption $C_{j,t}$ should be interpreted as including consumption of housing services. For working-age households in the model that are assumed to either rent or owe a mortgage on their primary residence, expenditures on housing services should be interpreted as including rent and mortgage interest payments but not mortgage principal payments, which I consider to be included in illiquid investment $sY_{j,t}$. This interpretation is consistent with the Consumer Expenditure Survey's classification of mortgage interest and principal payments as consumption expenditures versus investment. Retired households in the model are assumed to own their primary residence outright with no mortgage. While this primary residence requires no rent or interest expenditures and generates no rental income stream, I interpret the flow income from illiquid assets as including the owner-equivalent rent that offsets the household's consumption of housing services from their primary residence.

Combining this with Equations (17) and (18) yields the law of motion for cash-on-hand:

$$M_{j,t+1} =$$

$$[\theta_{j,t}\exp(r_{t+1}) + (1-\theta_{j,t})\exp(r_f)] \cdot [M_{j,t} - C_{j,t} - \Phi_C(C_{j,t}, C_{j,t-1})] + \begin{cases} (1-s-\tau)Y_{j,t} & \text{if } t \leq t_{ret}(j) \\ S\overline{Y}_{j,t_{ret}(j)} & \text{if } t > t_{ret}(j) \end{cases} \tag{19}$$

Solving the model: Solving the household's problem requires keeping track of the household's age a and the value of four state variables at the start of each period t:

1. The equity risk premium x_t, which summarizes investment opportunities.
2. Cash-on-hand M_t, which summarizes the household's available resource to consume and invest.
3. Lagged consumption $C_{j,t-1}$, which determines the adjustment costs $\Phi_C(C_{j,t}, C_{j,t-1})$ paid at each level of current consumption $C_{j,t}$.
4. The persistent earnings component $\overline{Y}_{j,t}$ (and in retirement, the terminal value $\overline{Y}_{j,t_{ret}(j)}$ from the household's final year of work).

Given the random walk dynamics of $\log \overline{Y}_{j,t}$ in Equation (13), the household's problem can be solved without tracking the persistent earnings component as an additional state variable by solving for the scaled value function $v_{j,t} \equiv V_{j,t}/\overline{Y}_{j,t}$ (as in e.g. Gourinchas and Parker, 2002). The original individual-level state variables are rescaled to track the ratio of cash-on-hand to persistent earnings $m_{j,t} \equiv M_{j,t}/\overline{Y}_{j,t}$ (or to terminal persistent earnings $\overline{Y}_{j,t_{ret}(j)}$ during retirement), and the ratio of lagged consumption-to-current cash-on-hand $\breve{c}m_{j,t} \equiv C_{j,t-1}/M_{j,t}$, in addition to the exogenous state variable x_t.

Uncertainty is driven by four shocks that determine the equity risk premium x_t, the realized stock return r_t, individual income $Y_{j,t}$, and its persistent component $\overline{Y}_{j,t}$:

1. The innovation $\xi_{x,t}$ to the equity risk premium x_t, which determines the discount rate news component $N_{DR,t}$ of the realized stock return.
2. The cash flow news component $N_{CF,t}$ of the realized stock return.
3. The persistent income shock $\eta_{j,t}$. The distribution of this shock depends on the two components $N_{DR,t}$ and $N_{CF,t}$ of the contemporaneous stock return r_t.
4. The transitory income shock $\varepsilon_{j,t}$.

Gaussian quadrature is used to integrate over each of these shocks. When simulating the model, these shocks are drawn from continuous normal distributions.

Parameter values: Each period in the model represents one year, which allows for a straightforward mapping between model parameters and observed annual income data. Throughout the rest of this section, model solutions are presented under the following parameter values. For preferences, the coefficient of relative risk aversion is set to $\gamma = 5$ and the elasticity of intertemporal substitution to $\psi = 0.5$. The time discount factor is set to $\beta = 0.85$. Because the model does not explicitly include mortality risk, this value should be interpreted as implicitly incorporating the additional time discounting based on average mortality rates over the full life cycle.[31]

Households face a marginal tax rate of $\tau = 0.35$ while working. They invest a fraction $s = 0.15$ of each period's pre-tax earnings in assets that are assumed to be illiquid during working life, such as housing and tax-advantaged retirement accounts. In retirement, these assets generate flow income in each year equal to a fraction $S = 0.60$ of persistent earnings $\overline{Y}_{j,t_{ret}(j)}$ at retirement.

The annual real log return on the safe risk-free bond is set to $r_f = 0.02$, i.e. an annual return of 2%. The average equity premium is set to $\overline{x} = 0.05$, i.e. 5% per year. The equity premium's annual persistence x_t is set to $\phi_x = 0.85$, which implies that a positive 1% shock to its level *lowers* the contemporaneous stock return r_t by approximately 5.22% through the discount rate news component $N_{DR,t}$. The standard deviation of the risk premium shocks $\xi_{x,t}$ is set to 2.3%, which implies a standard deviation of discount rate news equal to $\sigma_{DR} = 12.0\%$. The standard deviation of cash flow news is set to $\sigma_{CF} = 7.5\%$. The correlation between cash flow and discount rate news is set to $\rho_{CF,DR} = -0.75$. Together, these values imply an annual stock return volatility of approximately $\sigma_r = 18.3\%$. These values are in line with the values implied by vector autoregression-based forecasts and decompositions of stock returns surveyed in Chapters 5 and 9 of Campbell (2018).

The parameters governing the mixture components for the cash flow and discount rate news components of the realized stock return are set so that large stock market crashes occur with an annual probability of 0.15, and conditional on a crash occurring the realized log stock return falls by $r_{t+1} - E_t[r_{t+1}] = -0.25$ relative to its expected value from the prior period (i.e. the realized return is $\exp(-0.25) - 1 = -22.1\%$ lower than expected). The values of the mixture component means $\mu_{DR-} = 0.158$ for discount rate news and $\mu_{CF-} = -0.092$ for cash flow news conditional on a crash occurring are obtained from a linear projection of each component on the realized stock return, based on the unconditional moments stated in the previous paragraph. The mixture component means μ_{DR+} and μ_{CF+} for the non-crash state are then set so that $E_t[N_{DR,t+1}] = 0$ and $E_t[N_{CF,t+1}] = 0$.

For individual income, the age-dependent growth rates of persistent earnings g_a in Equation (13) is taken from the life cycle labor earnings profiles estimated by Guvenen, Ozkan, and Song (2014). The standard deviation of transitory earnings shocks $\varepsilon_{j,t}$ is set to $\sigma_\varepsilon = 0.20$, or 20% per year. The distribution of the persistent earnings shock $\eta_{j,t}$ depends on the large set of

[31] It would be straightforward to incorporate age-specific mortality rates into the model. This value of β falls within the range used in prior papers that estimate this parameter by matching observed average levels of total household wealth over the life cycle, including both liquid and illiquid assets. For example, Fagereng, Gottlieb, and Guiso (2017) estimate a value of $\beta = 0.77$ to match average wealth levels for Norwegian households in a model with a bequest motive, while Catherine (2022) estimates a value of $\beta = 0.91$ to match wealth levels for U.S. households in a model with similar cyclical income risk and no stock market participation frictions. These values exclude discounting due to mortality risk (which is explicitly incorporated into the model). These models assume stock returns are serially uncorrelated, which lowers households' optimal stock portfolio shares and thus their wealth accumulation over the life cycle.

parameters presented in Equation (14), which determine both the shape of the unconditional earnings growth distribution and its dependence on the components of the contemporaneous realized stock return r_t. I specify these parameters so that these features are similar to the persistent earnings shock distribution used in Catherine (2022). I set the annual probability of drawing from the low-mean persistent earnings growth mixture component to $p_{\eta-} = 0.15$, and the average persistent earnings growth conditional on drawing from this component to $\overline{\mu}_{\eta-} = -0.10$. I choose the remaining parameters appearing in Equation (14) so that:

- The overall standard deviation of the persistent earnings shock (including the dispersion attributable to the uncertain mixture mean) is set to $\sigma_\eta = 0.25$, or 25% per year.
- A 1% increase in the cash flow news component $N_{CF,t}$ of the realized return r_t increases average persistent earnings across individuals (equal to $p_{\eta-}\mu_{\eta-}(N_{CF,t}, N_{DR,t}) + (1 - p_{\eta-})\mu_{\eta+}(N_{CF,t}, N_{DR,t})$) by 0.2%, and increases average earnings growth for the low-mean mixture component $\mu_{\eta-}(N_{CF,t}, N_{DR,t})$ by 0.9%. I impose the same effect for a 1% decrease in the discount rate news component $N_{DR,t}$, so that both components of the realized stock return have the same effect on the earnings growth distribution.

These targets are chosen to match the implied values from the calibration used by Catherine (2022). The calibrated earnings process captures the key empirical fact that the left tail of the earnings growth distribution shifts more strongly with realized stock returns than the mean.

With all of these parameter values fixed, I solve the model under different values of the consumption adjustment cost parameter φ_C, to determine how this feature of the model shapes households' optimal savings and portfolio choices.

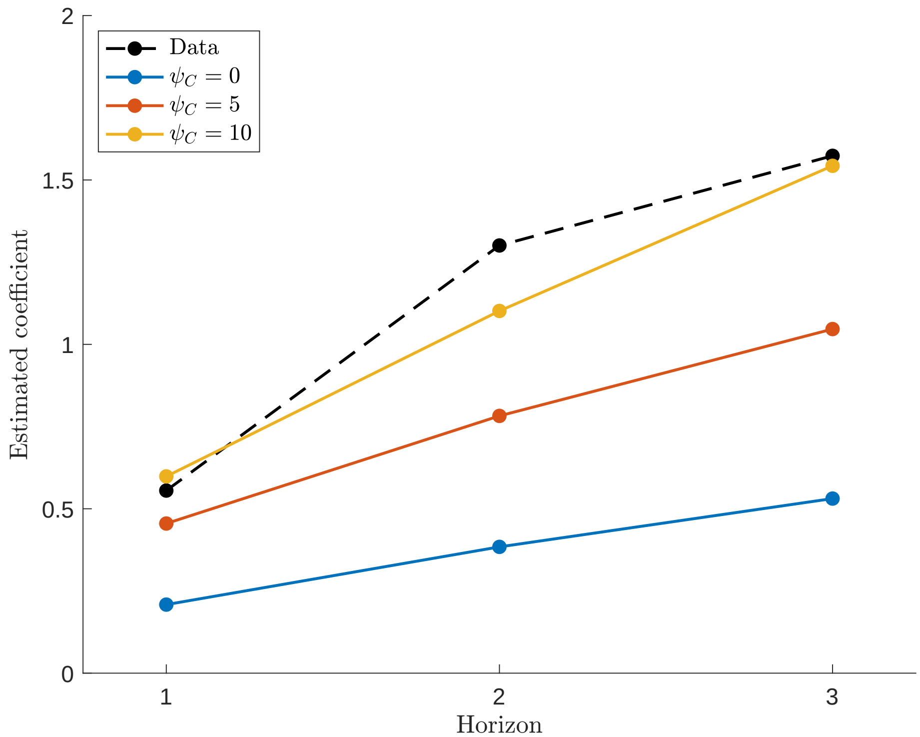
5.2 Adjustment frictions, savings responses, and optimal portfolio choices

I now compare the model-implied household savings response to persistent income shocks against the estimates presented in Section 4.2. In the data, I estimated the savings response regression in Equation (3) using the ZIP-code-based instrument for individual income growth. This instrument induces large and persistent changes in individual income. In the model, I simulate many cohorts of working-age individuals and estimate a similar regression, using the change in the persistent component of individual earnings $\overline{Y}_{j,t+1} - \overline{Y}_{j,t}$ as the analogous regressor:

$$\frac{S_{j,t,t+h}}{\overline{Y}_{j,t}} = \beta_h \cdot \frac{\overline{Y}_{j,t+1} - \overline{Y}_{j,t}}{\overline{Y}_{j,t}} + \delta_{age(j),t} + \varepsilon_{j,t+h} \tag{20}$$

Here $S_{j,t,t+1} \equiv \widetilde{W}_{j,t+1} - W_{j,t+1}$ is the model counterpart of the measured active savings flow $S^{tot}_{j,t,t+1}$, and $S_{j,t,t+h}$ denotes the cumulated savings flow through period $t+h$. $\delta_{age(j),t}$ denote age-by-time fixed effects. I simulate the model under several alternative values of the adjustment cost parameter φ_C to determine which value produces individual savings responses consistent with the estimates from the tax data.

Figure 10: Large Adjustment Costs Needed to Match the Estimated Savings Response



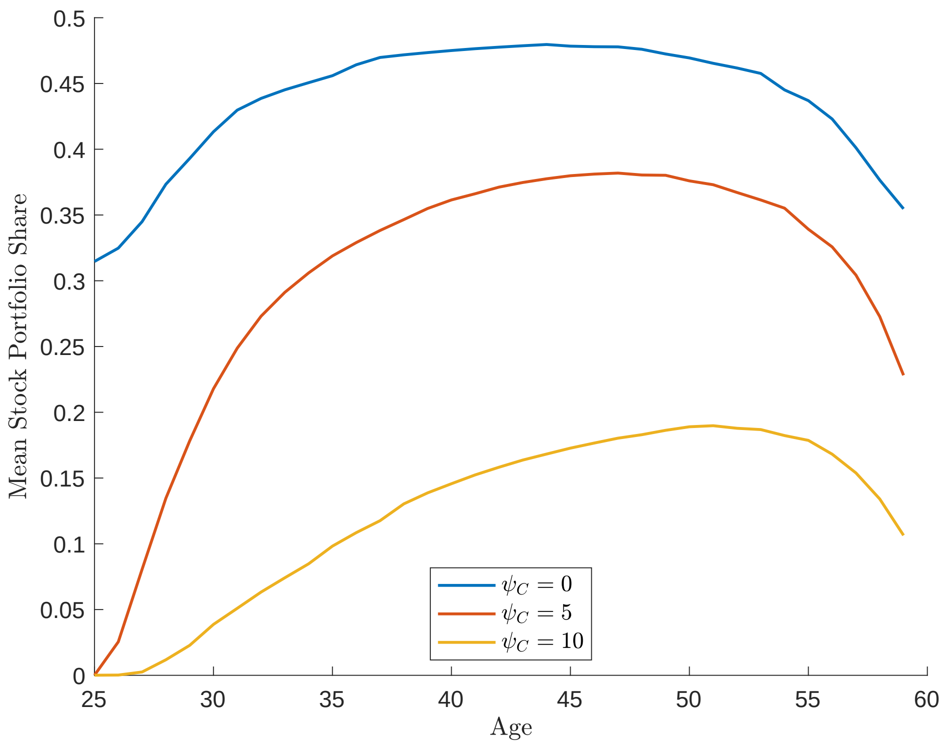
Notes: This figure compares the household savings response to persistent earnings shocks estimated from the income tax data (dashed black line) against the model-implied responses (solid colored lines). The estimated savings responses (black) are an average of the second-stage estimates of the savings regression in Equation (3) reported in Figure 9, weighted by each income group's share of observations in the tax data sample. The model-implied responses (colored) are estimates from the analogous regression in Equation (20) using model-simulated data. Each colored line corresponds to model-implied estimates for a different value of the adjustment cost parameter φ_C. The horizontal axis reports the horizon h (in years) at which the savings response is measured. Estimated savings responses from the tax data are adjusted to a post-tax basis as described in Section 4.2, and model-implied responses are adjusted similarly by multiplying the estimated coefficients β_h from Equation (20) by $1/(1-\tau)$.

Figure 10 compares the model-implied savings response to persistent income shocks against my empirical estimates. The model with no adjustment costs ($\varphi_C = 0$) generates a small savings response to these shocks which is substantially smaller than the empirical estimates. Intuitively, the model with frictionless consumption adjustment generates consumption and savings responses to persistent income shocks which are qualitatively similar to the canonical permanent income hypothesis. When faced with a persistent decline in their nonfinancial income, households primarily reduce their *spending* rather than their savings.

I find that large consumption adjustment costs are required in order to generate a model-implied savings response to persistent income shocks that is quantitatively consistent with estimates from the income tax data. An adjustment cost value of $\varphi_C = 5$ generates a contemporaneous savings response similar to the empirical estimate, but fails to match the additional estimated savings response in the years following the initial shock. With high adjustment costs of $\varphi_C = 10$ the model-implied savings responses are much closer to the empirical estimates, both for the initial year following the shock and at a longer horizon of $h = 3$ years after the initial shock. These adjustment frictions imply significant costs for households to rapidly cut their consumption: the parameter value $\varphi_C = 10$ implies that a 10% reduction in consumption requires paying an additional cost equal to 5% of previous year's consumption.

What are the implications of these large consumption adjustment frictions for households' optimal portfolio choices? Figure 11 presents the model-implied average optimal stock portfolio shares over households' working-age period. Without consumption adjustment frictions, the optimal stock portfolio share is roughly 50% across most ages. This value is considerably higher than would be implied if stock returns were serially uncorrelated with the same annual volatility. Risk-averse long-term investors in the model (with relative risk aversion of $\gamma = 5$) perceive stocks as significantly less risky over long investment horizons, given the discount rate news account for a substantial share of realized stock return volatility.

Figure 11: Adjustment Costs Lower Optimal Stock Portfolio Shares

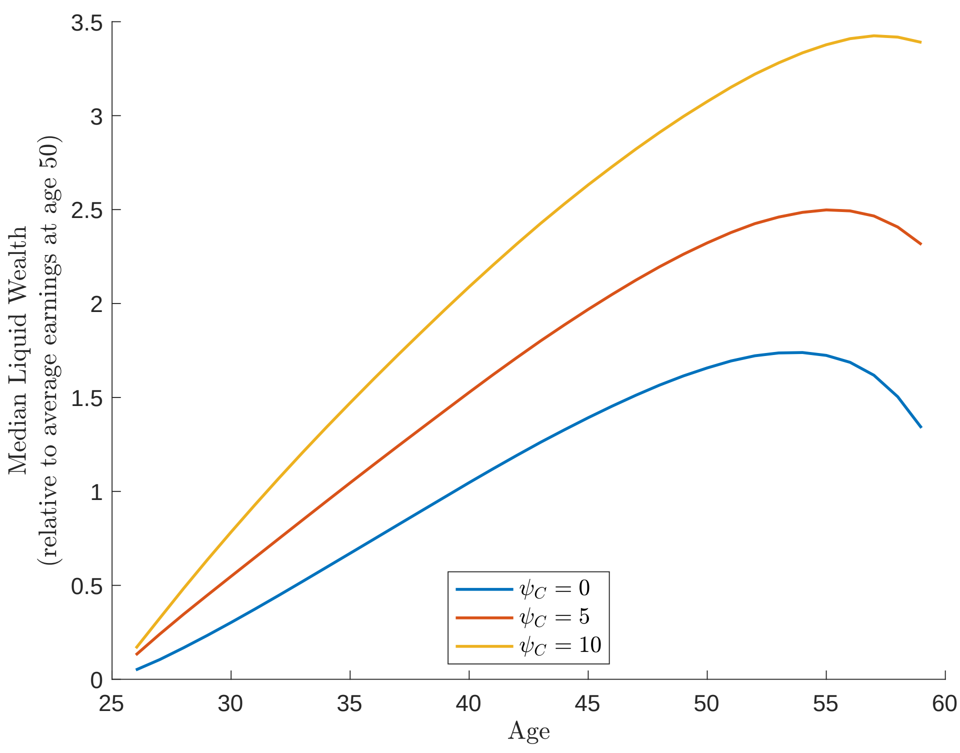


Notes: This figure reports average model-implied optimal stock portfolio shares $\theta_{j,t}$ across working ages under different values of the consumption adjustment cost parameter φ_C. For each value of φ_C, observations are simulated for consecutive cohorts of individuals under the consumption/savings and portfolio policies, then compute the average stock portfolio share $\theta_{j,t}$ (vertical axis) across households at each age (horizontal axis). Each line represents average stock portfolio shares corresponding to a different value of the adjustment cost parameter φ_C.

In contrast, optimal stock portfolio shares are significantly lower when households face adjustment costs that produce realistic savings responses to persistent income shocks. This difference is particularly pronounced at young ages, when households have little liquid wealth and must finance their consumption primarily through their nonfinancial income. Under the high value $\varphi_C = 10$ for the adjustment cost parameter that was required to closely match the savings response estimated in the tax data, the model-implied optimal stock portfolio shares are less than a third of the optimal value from the frictionless model (with $\varphi_C = 0$) at all points in households' working life. Even for the more moderate value $\varphi_C = 5$ for the adjustment cost parameter, optimal stock portfolio shares are significantly lower than in the frictionless model, particularly early in the life cycle.

High consumption adjustment costs also significantly increase households' optimal wealth accumulation. Figure 12 depicts the median simulated path of household wealth throughout working age, under the model-implied optimal consumption and savings policies. Households accumulate substantially more liquid financial wealth at all points in the life cycle when faced with high consumption adjustment costs. The inability to easily cut consumption creates significant liquidity needs for households after they experience a negative income shock. This in turn generates a strong precautionary savings motive that induces households to accumulate more liquid wealth in anticipation of these future liquidity needs.

Figure 12: Adjustment Costs Increase Wealth Accumulation



Notes: This figure reports model-implied median liquid wealth $W_{j,t}$ (relative to average nonfinancial income $Y_{j,t}$ at age 50) across working ages under different values of the consumption adjustment cost parameter φ_C. For each value of φ_C, observations are simulated for consecutive cohorts of individuals under the consumption/savings and portfolio policies, then compute median liquid wealth $W_{j,t}$ (vertical axis) across households at each age (horizontal axis). Wealth is normalized as a fraction of average pre-tax nonfinancial income $Y_{j,t}$ of households at age 50, pooled across all years in the simulation. Each line represents median wealth values corresponding to a different value of the adjustment cost parameter φ_C.

For a given set of parameter values, optimal stock portfolio shares in my model are increasing in liquid wealth. Under their optimal savings and portfolio policies, households choose low optimal stock portfolio in the presence of high consumption adjustment costs (as depicted in Figure 11) *despite* the fact that these costs lead them to accumulate substantially more liquid wealth. Consumption adjustment costs thus play a substantial role in determining households' optimal portfolio choices, even when the long-run risk of investing in stocks is relatively low (as it is in the model). More broadly, these adjustment costs offer a potential alternative explanation for several observed patterns in household savings and portfolio choices over the life cycle. Many structural life-cycle models reported in the literature use relatively high values of the risk aversion parameter γ or fixed costs associated with participation in the stock market to explain low observed stock portfolio shares, especially among young households with low levels of financial wealth. My results suggest an alternative explanation for these observed portfolio choices, through a combination of a lower "pure" risk aversion parameter γ and high adjustment costs ψ_C that increase households' *effective* risk aversion, especially at low levels of liquid wealth. Similarly, precautionary savings motives are considered to be a significant determinant of households' savings choices and wealth accumulation over the life cycle. My results suggest that these precautionary savings motives may be driven by a *need* to avoid sharp cuts in spending due to the high adjustment costs associated with such a change, rather than a *desire* to avoid sharp cuts in spending arising from a high risk aversion coefficient γ.

6 Conclusion

In this paper, I argue that stocks are a riskier investment for many working-age households than their relatively safe long-run returns would seem to suggest. Empirically, I use over 25 years of individual income tax data to characterize how high-income, working-age investors invest and utilize their liquid wealth. Using new capitalization methods, I estimate households' active savings flows in taxable accounts by asset class, and find that these aggregate flows into the stock market are highly procyclical. Large stock price crashes coincide with both large outflows from the stock market, and large declines in wage and private business income for many working-age investors. Individual households take large drawdowns from their liquid wealth — including both stocks and fixed income — after experiencing persistent losses in their nonfinancial income. This behavior is at odds with the predictions of canonical consumption-savings models that predict *spending* will serve as the main margin of adjustment in response to such shocks.

Theoretically, I incorporate consumption adjustment frictions into a rich portfolio choice model to explain this observed savings behavior and trace out its portfolio choice implications. High consumption adjustment cost are required to match the large estimated savings response that I document in the data. These adjustment costs imply optimal stock portfolio shares that are substantially lower than in the case with no adjustment frictions, despite the fact that they also lead households to accumulate substantially more liquid wealth. The model implies that stocks are a relatively risky investment for many working-age investors even when a large share of variation in stock prices is purely transitory. More broadly, my empirical evidence and quantitative modeling point of the importance of investors' income risk, expenditure commitments, and resulting liquidity needs as first-order considerations for their long-term investment decisions.

References

Aguiar, M., and E. Hurst (2005): "Consumption versus Expenditure," *Journal of Political Economy*, 113(5), 919–948.

Alexander, J. T., D. Bleckley, J. Fisher, K. Genadek, S. H. Leonard, and A. Magganas (2024): "Newly Available Individual-Level U.S. Tax Data from 1969–1994," U.S. Census Bureau, ADEP Working Paper 2024-02.

Ameriks, J., and S. P. Zeldes (2004): "How Do Household Portfolio Shares Vary With Age?," Working Paper.

Andersen, A. L., A. S. Jensen, N. Johannesen, C. T. Kreiner, S. Leth-Petersen, and A. Sheridan (2023): "How Do Households Respond to Job Loss? Lessons from Multiple High-Frequency Datasets," *American Economic Journal: Applied Economics*, 15(4), 1–29.

Armstrong, D. M., J. L. Hoopes, and E. L. Maydew (2025): "Stock Trading by U.S. Individual Investors: Evidence from Administrative Tax Data," Working Paper.

Baker, S. R. (2018): "Debt and the Response to Household Income Shocks: Validation and Application of Linked Financial Account Data," *Journal of Political Economy*, 126(4), 1504–1557.

Balloch, C. M., and J. Richers (2023): "Asset Allocation and Returns in the Portfolios of the Wealthy," Working Paper.

Barberis, N. (2000): "Investing for the Long Run when Returns Are Predictable," *The Journal of Finance*, 55(1), 225–264.

Bee, A., J. Mitchell, N. Mittag, J. Rothbaum, C. Sanders, L. Schmidt, and M. Unrath (2023): "National Experimental Wellbeing Statistics," SEHSD Working Paper 2023-02; CES Working Paper 23-04.

Benartzi, S. (2001): "Excessive Extrapolation and the Allocation of 401(k) Accounts to Company Stock," *The Journal of Finance*, 56(5), 1747–1764.

Benzoni, L., P. Collin-Dufresne, and R. S. Goldstein (2007): "Portfolio Choice over the Life-Cycle when the Stock and Labor Markets Are Cointegrated," *The Journal of Finance*, 62(5), 2123–2167.

Brunnermeier, M. K., and S. Nagel (2008): "Do Wealth Fluctuations Generate Time-Varying Risk Aversion? Micro-evidence on Individuals," *American Economic Review*, 98(3), 713–36.

Calvet, L. E., J. Y. Campbell, F. J. Gomes, and P. Sodini (2025): "The Cross-Section of Household Preferences," *The Journal of Finance*, Forthcoming.

Calvet, L. E., J. Y. Campbell, and P. Sodini (2009): "Fight or Flight? Portfolio Rebalancing by Individual Investors," *The Quarterly Journal of Economics*, 124(1), 301–348.

Campanale, C., C. Fugazza, and F. Gomes (2015): "Life-cycle portfolio choice with liquid and illiquid financial assets," *Journal of Monetary Economics*, 71, 67–83.

Campbell, J. Y. (1991): "A Variance Decomposition for Stock Returns," *The Economic Journal*, 101(405), 157–179.

——— (2018): *Financial Decisions and Markets: A Course in Asset Pricing*. Princeton University Press.

Campbell, J. Y., S. Giglio, and C. Polk (2013): "Hard Times," *The Review of Asset Pricing Studies*, 3(1), 95–132.

Campbell, J. Y., and R. J. Shiller (1988): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *The Review of Financial Studies*, 1(3), 195–228.

Campbell, J. Y., and L. M. Viceira (1999): "Consumption and Portfolio Decisions when Expected Returns are Time Varying," *The Quarterly Journal of Economics*, 114(2), 433–495.

Caplin, A., V. Gregory, E. Lee, S. Leth-Petersen, and J. Sæverud (2023): "Subjective Earnings Risk," Working Paper 31019, National Bureau of Economic Research.

Carroll, C. D. (2009): "Precautionary Saving and the Marginal Propensity to Consume out of Permanent Income," *Journal of Monetary Economics*, 56(6), 780–790.

Catherine, S. (2022): "Countercyclical Labor Income Risk and Portfolio Choices over the Life Cycle," *The Review of Financial Studies*, 35(9), 4016–4054.

Catherine, S., P. Sodini, and Y. Zhang (2024): "Countercyclical Income Risk and Portfolio Choices: Evidence from Sweden," *The Journal of Finance*, 79(3), 1755–1788.

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

Chetty, R., L. Sandor, and A. Szeidl (2017): "The Effect of Housing on Portfolio Choice," *The Journal of Finance*, 72(3), 1171–1212.

Chetty, R., and A. Szeidl (2007): "Consumption Commitments and Risk Preferences," *The Quarterly Journal of Economics*, 122(2), 831–877.

Choukhmane, T., and T. de Silva (2025): "What Drives Investors' Portfolio Choices? Separating Risk Preferences from Frictions," *The Journal of Finance*, Forthcoming.

Cocco, J. F. (2005): "Portfolio Choice in the Presence of Housing," *The Review of Financial Studies*, 18(2), 535–567.

Cocco, J. F., F. J. Gomes, and P. J. Maenhout (2005): "Consumption and Portfolio Choice over the Life Cycle," *The Review of Financial Studies*, 18(2), 491–533.

Cochrane, J. H. (2011): "Presidential Address: Discount Rates," *The Journal of Finance*, 66(4), 1047–1108.

Dahlquist, M., O. Setty, and R. Vestman (2018): "On the Asset Allocation of a Default Pension Fund," *The Journal of Finance*, 73(4), 1893–1936.

Døskeland, T. M., and H. K. Hvide (2011): "Do Individual Investors Have Asymmetric Information Based on Work Experience?," *The Journal of Finance*, 66(3), 1011–1041.

Duarte, V., J. Fonseca, J. A. Parker, and A. Goodman (2022): "Simple Allocation Rules and Optimal Portfolio Choice Over the Lifecycle," Working Paper.

Eisfeldt, A. L., A. Falato, D. Lee, and M. Z. Xiaolan (2024): "Equity Pay Beyond the C-Suite," Working Paper.

Epstein, L. G., and S. E. Zin (1989): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57(4), 937–969.

Fagereng, A., C. Gottlieb, and L. Guiso (2017): "Asset Market Participation and Portfolio Choice over the Life-Cycle," *The Journal of Finance*, 72(2), 705–750.

Fagereng, A., M. B. Holm, B. Moll, and G. Natvik (2021): "Saving Behavior Across the Wealth Distribution: The Importance of Capital Gains," Working Paper.

Fagereng, A., H. Onshuus, and K. N. Torstensen (2024): "The consumption expenditure response to unemployment: Evidence from Norwegian households," *Journal of Monetary Economics*, 146, 103578.

Flavin, M., and S. Nakagawa (2008): "A Model of Housing in the Presence of Adjustment Costs: A Structural Interpretation of Habit Persistence," *American Economic Review*, 98(1), 474–95.

Gabaix, X., and R. Koijen (2023): "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," Working paper.

Gabaix, X., R. Koijen, F. Mainardi, S. Oh, and M. Yogo (2025a): "Asset Demand of U.S. Households," Working Paper.

——— (2025b): "Limited Risk Transfer Between Investors: A New Benchmark for Macro-Finance Models," Working Paper.

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus (2021a): "Five Facts about Beliefs and Portfolios," *American Economic Review*, 111(5), 1481–1522.

——— (2021b): "The joint dynamics of investor beliefs and trading during the COVID-19 crash," *Proceedings of the National Academy of Sciences*, 118(4), e2010316118.

Gomes, F., and A. Michaelides (2003): "Portfolio choice with internal habit formation: a life-cycle model with uninsurable labor income risk," *Review of Economic Dynamics*, 6(4), 729–766, Finance and the Macroeconomy.

Gomes, F., A. Michaelides, and V. Polkovnichenko (2009): "Optimal savings with taxable and tax-deferred accounts," *Review of Economic Dynamics*, 12(4), 718–735.

Gourinchas, P.-O., and J. A. Parker (2002): "Consumption Over the Life Cycle," *Econometrica*, 70(1), 47–89.

Grinblatt, M., and M. Keloharju (2001): "How Distance, Language, and Culture Influence Stockholdings and Trades," *The Journal of Finance*, 56(3), 1053–1073.

Grossman, S. J., and G. Laroque (1990): "Asset Pricing and Optimal Portfolio Choice in the Presence of Illiquid Durable Consumption Goods," *Econometrica*, 58(1), 25–51.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021): "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?," *Econometrica*, 89(5), 2303–2339.

Guvenen, F., S. Ozkan, and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122(3), 621–660.

Guvenen, F., S. Schulhofer-Wohl, J. Song, and M. Yogo (2017): "Worker Betas: Five Facts about Systematic Earnings Risk," *American Economic Review*, 107(5), 398–403.

Hoopes, J. L., P. Langetieg, S. Nagel, D. Reck, J. Slemrod, and B. A. Stuart (2022): "Who Sells During a Crash? Evidence from Tax Return Data on Daily Sales of Stock," *The Economic Journal*, 132(641), 299–325.

Kennickell, A. B. (2008): "The Role of Over-Sampling of the Wealthy in the Survey of Consumer Finances," in *IFC Bulletin No. 28: IFC's Contribution to the 56th ISI Session, Lisbon, 2007*. Irving Fisher Committee on Central Bank Statistics.

Kroencke, T. A. (2022): "Recessions and the stock market," *Journal of Monetary Economics*, 131, 61–77.

Larrimore, J., J. Mortenson, and D. Splinter (2016): "Income and Earnings Mobility in U.S. Tax Data," in *Economic Mobility: Research & Ideas on Strengthening Families, Communities & the Economy*, ed. by F. R. B. of St. Louis, and the Board of Governors of the Federal Reserve System, pp. 481–516. Federal Reserve Bank of St. Louis and Board of Governors of the Federal Reserve System.

Lustig, H., and A. Verdelhan (2012): "Business cycle variation in the risk-return trade-off," *Journal of Monetary Economics*, 59, S35–S49, Supplement issue:October 15-16 2010 Research Conference on 'Directions for Macroeconomics: What did we Learn from the Economic Crises' Sponsored by the Swiss National Bank (http://www.snb.ch).

Mainardi, F. (2025): "Who Harvests? Tax Alpha and Heterogeneous Responses to Capital Gains Taxation," Working paper.

Mariano, R. S., and Y. Murasawa (2003): "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18(4), 427–443.

Martin, I. (2017): "What is the Expected Return on the Market?," *The Quarterly Journal of Economics*, 132(1), 367–433.

Martínez-Toledano, C. (2023): "House Price Cycles, Wealth Inequality and Portfolio Reshuffling," Working Paper.

Massa, M., and A. Simonov (2006): "Hedging, Familiarity and Portfolio Choice," *The Review of Financial Studies*, 19(2), 633–685.

Meeuwis, M. (2022): "Wealth Fluctuations and Risk Preferences: Evidence from U.S. Investor Portfolios," Working Paper.

Merton, R. C. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41(5), 867–887.

Michaelides, A., and Y. Zhang (2017): "Stock Market Mean Reversion and Portfolio Choice over the Life Cycle," *Journal of Financial and Quantitative Analysis*, 52(3), 1183–1209.

Mogstad, M., L. Schmidt, M. Schwartz, M. Tiurina, O. Vestad, and N. von Turkovich (2025): "Risk Exposures and Diversification at the Top of the Wealth Distribution: Evidence from Norway," .

Parker, J. A., A. Schoar, A. Cole, and D. Simester (2024): "Household Portfolios and Retirement Saving over the Life Cycle," *Journal of Finance*, forthcoming.

Patterson, C. (2023): "The Matching Multiplier and the Amplification of Recessions," *American Economic Review*, 113(4), 982–1012.

Piketty, T., E. Saez, and G. Zucman (2018): "Distributional National Accounts: Methods and Estimates for the United States," *The Quarterly Journal of Economics*, 133(2), 553–609.

Rotemberg, J. J. (1982): "Sticky Prices in the United States," *Journal of Political Economy*, 90(6), 1187–1211.

Saez, E., and G. Zucman (2016): " Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data ," *The Quarterly Journal of Economics*, 131(2), 519–578.

Schmidt, L. D. (2025): "Climbing and falling off the ladder: Asset pricing implications of labor market event risk," *Journal of Financial Economics*, 172, 104131.

Shen, J. (2024): "Countercyclical Risks, Consumption, and Portfolio Choice: Theory and Evidence," *Management Science*, 70(5), 2862–2881.

Siegel, J. (1994): *Stocks for the Long Run: The Definitive Guide to Financial Market Returns and Long-term Investment Strategies*. McGraw-Hill.

Smith, M., D. Yagan, O. Zidar, and E. Zwick (2019): "Capitalists in the Twenty-First Century," *The Quarterly Journal of Economics*, 134(4), 1675–1745.

Smith, M., O. Zidar, and E. Zwick (2023): "Top Wealth in America: New Estimates Under Heterogeneous Returns," *The Quarterly Journal of Economics*, 138(1), 515–573.

Strebulaev, I. A., and T. M. Whited (2012): "Dynamic Models and Structural Estimation in Corporate Finance," *Foundations and Trends in Finance*, 6(1–2), 1–163.

Vestman, R. (2019): "Limited Stock Market Participation Among Renters and Homeowners," *The Review of Financial Studies*, 32(4), 1494–1535.

Viceira, L. M. (2001): "Optimal Portfolio Choice for Long-Horizon Investors with Nontradable Labor Income," *The Journal of Finance*, 56(2), 433–470.

Wachter, J. A. (2002): "Portfolio and Consumption Decisions under Mean-Reverting Returns: An Exact Solution for Complete Markets," *The Journal of Financial and Quantitative Analysis*, 37(1), 63–91.

Wagner, D., and M. Layne (2014): "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' (CARRA) Record Linkage Software," CARRA Working Paper 2014-01, U.S. Census Bureau.

Wang, T. (2022): "Perceived versus calibrated income risks in heterogeneous-agent consumption models," Discussion paper.

Weil, P. (1990): "Nonexpected Utility in Macroeconomics," *The Quarterly Journal of Economics*, 105(1), 29–42.

A Details on income tax data methodology

This section describes the details of the capitalization method I apply to the income tax data, and the underlying assumptions necessary to estimate the response of net savings flows to nonfinancial income shocks. The administrative income tax data used in this paper report the total taxable dividend income (paid by e.g. stocks or fixed income mutual funds), taxable interest income (paid by e.g. savings accounts or directly-held bonds), and nontaxable interest income (paid by directly-held municipal bonds or municipal bond funds) received by each tax unit (hereafter referred to as a *household*) over a given calendar year. The reported income is aggregated over all assets paying each form of income; for example, total dividend income is aggregated across all stocks and mutual funds. There is no directly-reported information about the income received from each individual asset, proceeds from sales or purchases of these assets, or any characteristics of these assets.

I first introduce an accounting framework to determine conditions under which households' unobserved net savings flows can be inferred from their observed dividend income. I build on the capitalization methods developed by Saez and Zucman (2016) and Smith, Zidar, and Zwick (2023) for estimating wealth levels and synthetic savings rates from capital income flows observed in tax data.

A.1 Accounting framework and imputed net savings

Consider a household j. The household trades a set of individual assets a which each generate either taxable dividends, taxable interest income, or nontaxable interest income. Let $\mathcal{A}_D$, $\mathcal{A}_I$, and $\mathcal{A}_N$ denote each of these three asset classes, and $Q^a_{j,t}$ denote the quantity of shares in asset a held by household j at time t (denominated in years).[32] For the purposes of exposition, in this section all asset transactions are assumed to take place at the start of each calendar year, so that holdings are constant throughout the year ($Q^a_{j,\tau} = Q^a_{j,t}$ for times $\tau \in [t, t+1)$). A formal treatment of time aggregation is presented in Appendix A.3, where asset transactions which take place at the start of the calendar year are shown to have the largest effect on the measured change in dividend income relative to the prior year and the savings flow imputed from this change. For the rest of this section, the time t takes only integer values (e.g. $t = 2009$ at the start of calendar year 2009).

First consider the case of a single dividend-paying asset a. Each share of asset a trades at price $\mathbf{P}^a_t$ at time t. Over the course of the year from time t to $t+1$, it pays total dividends $\mathbf{D}^a_{t+1}$. The dividend income received by household j from asset a over the same time period is given by $D^a_{j,t+1} = Q^a_{j,t}\mathbf{D}^a_{t+1}$. Note that this timing convention differs from the dating of annual income tax returns: $D^a_{j,2009}$ represents dividend income received in tax year 2008, which is known only at the end of the calendar year (i.e. at $t = 2009$, the start of calendar year 2009). I define $S^a_{j,t} \equiv \mathbf{P}^a_t(Q^a_{j,t} - Q^a_{j,t-1})$ as household j's *net savings in asset* a, the dollar value of its net purchases or sales at the start of the calendar year t (which take place at price $\mathbf{P}^a_t$). The

[32]Short positions are ruled out by restricting $Q^a_{j,t} \geq 0$.

total dividend income received by household j is obtained by aggregating its income across all dividend-paying assets: $D_{j,t} \equiv \sum_{a\in\mathcal{A}_D} D^a_{j,t}$. The *net savings in dividend-paying assets* is defined similarly: $S^D_{j,t} \equiv \sum_{a\in\mathcal{A}_D} S^a_{j,t}$.

The change in dividend income received by household j from asset a between consecutive calendar years can be attributed to either a change in its holdings $Q^a_{j,t}$, or to a change in dividends $\mathbf{D}^a_{t+1}$ paid by the asset:

$$D^a_{j,t+1} - D^a_{j,t} = Q^a_{j,t}\mathbf{D}^a_{t+1} - Q^a_{j,t-1}\mathbf{D}^a_t = \underbrace{(Q^a_{j,t} - Q^a_{j,t-1})\mathbf{D}^a_{t+1}}_{\text{due to change in holdings}} + \underbrace{Q^a_{j,t-1}(\mathbf{D}^a_{t+1} - \mathbf{D}^a_t)}_{\substack{\text{due to change in} \\ \text{asset dividends}}} \tag{A.1}$$

If the dividend income in each year was observed, then the net savings in asset a could be calculated as

$$S^a_{j,t} = \frac{\mathbf{P}^a_t}{\mathbf{D}^a_{t+1}}\left(D^a_{j,t+1} - \frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t}D^a_{j,t}\right) \tag{A.2}$$

I refer to the term in parentheses as the *excess change in dividend income* received from asset a. It represents the component of household j's dividend income from asset a in year $t+1$ that can be attributed only to changes in its asset holdings: $D^a_{j,t+1} - \frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t}D^a_{j,t} = (Q^a_{j,t} - Q^a_{j,t-1})\mathbf{D}^a_{t+1}$. Scaling this excess change by the appropriate capitalization factor — the current price-to-future dividend ratio $\mathbf{P}^a_t/\mathbf{D}^a_{t+1}$ — yields the value of the net savings flow $S^a_{j,t} = \mathbf{P}^a_t(Q^a_{j,t} - Q^a_{j,t-1})$.

In practice, I observe total dividend income aggregated across all assets rather than asset-specific dividend income for each household. I derive a similar decomposition for dividend income changes that accounts for the composition of these asset holdings. To aid with this derivation, I introduce the following useful notation:

- Let $W^a_{j,t} \equiv \mathbf{P}^a_t Q^a_{j,t-1}$ denotes the value of household j's holdings of asset a at the start of year t *before* any rebalancing (but reflecting the current price $\mathbf{P}^a_t$). Similarly, let $\widetilde{W}^a_{j,t} \equiv \mathbf{P}^a_t Q^a_{j,t} = W^a_{j,t} + S^a_{j,t}$ denote the value of its holdings *after* rebalancing.
- Let $W^D_{j,t} \equiv \sum_{a\in\mathcal{A}_D} W^a_{j,t}$ denotes the total value of household j's dividend-paying assets *before* any rebalancing or net saving (but reflecting current prices). Similarly, let $\widetilde{W}^D_{j,t} \equiv \sum_{a\in\mathcal{A}_D} \widetilde{W}^a_{j,t} = W^D_{j,t} + S^D_{j,t}$ denote the value of its holdings *after* any rebalancing and saving.
- Let $\omega^a_{j,t} \equiv W^a_{j,t}/W^D_{j,t}$ denote household j's portfolio share for asset a among its dividend-paying assets at time t, *before* rebalancing (with $\omega^a_{j,t} = 0$ if $W^D_{j,t} = 0$). Let $\widetilde{\omega}^a_{j,t} \equiv \widetilde{W}^a_{j,t}/\widetilde{W}^D_{j,t}$ denote its portfolio share *after* rebalancing.
- Let $\alpha^a_{j,t} \equiv D^a_{j,t}/D_{j,t}$ denote household j's share of its total dividend income received from asset a (with $\alpha^a_{j,t} = 0$ if $D_{j,t} = 0$). Note that $\alpha^a_{j,t} = D^a_{j,t}/D_{j,t}$

- These definitions imply the following useful expression for net savings in asset a:

$$\begin{aligned} S^a_{j,t} &= \widetilde{W}^a_{j,t} - W^a_{j,t} \\ &= \widetilde{\omega}^a_{j,t}\widetilde{W}^D_{j,t} - \omega^a_{j,t}W^D_{j,t} \\ &= \widetilde{\omega}^a_{j,t}\widetilde{W}^D_{j,t} - \omega^a_{j,t}\left(\widetilde{W}^D_{j,t} - S^D_{j,t}\right) \\ &= \omega^a_{j,t}S^D_{j,t} + \left(\widetilde{\omega}^a_{j,t} - \omega^a_{j,t}\right)\widetilde{W}^D_{j,t} \\ &= \omega^a_{j,t}S^D_{j,t} + \left(\widetilde{\omega}^a_{j,t} - \omega^a_{j,t}\right)\left(W^D_{j,t} + S^D_{j,t}\right) \end{aligned} \tag{A.3}$$

Using these definitions, the change in household j's total dividend income between calendar years can be decomposed as follows:

$$\begin{aligned} &D_{j,t+1} - D_{j,t} \\ &= \sum_{a\in\mathcal{A}_D}\left(D^a_{j,t+1} - D^a_{j,t}\right) \\ &= \sum_{a\in\mathcal{A}_D}(Q^a_{j,t} - Q^a_{j,t-1})\mathbf{D}^a_{t+1} + \sum_{a\in\mathcal{A}_D}Q^a_{j,t-1}(\mathbf{D}^a_{t+1} - \mathbf{D}^a_t) \\ &= \sum_{a\in\mathcal{A}_D}S^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t} + \sum_{a\in\mathcal{A}_D}W^a_{j,t}\frac{\mathbf{D}^a_{t+1} - \mathbf{D}^a_t}{\mathbf{P}^a_t} \\ &= \underbrace{S^D_{j,t}\sum_{a\in\mathcal{A}_D}\omega^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}}_{\text{due to overall savings}} + \underbrace{\left(W^D_{j,t} + S^D_{j,t}\right)\sum_{a\in\mathcal{A}_D}\left(\widetilde{\omega}^a_{j,t} - \omega^a_{j,t}\right)\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}}_{\text{due to rebalancing}} + \underbrace{W^D_{j,t}\sum_{a\in\mathcal{A}_D}\omega^a_{j,t}\frac{\mathbf{D}^a_{t+1} - \mathbf{D}^a_t}{\mathbf{P}^a_t}}_{\text{due to change in asset dividends}} \end{aligned} \tag{A.4}$$

This decomposition illustrates the three possible ways that household j's total dividend income could increase from year t to year $t+1$:

- The household saves more in dividend-paying assets ($S^D_{j,t} > 0$), but *without* rebalancing towards assets with higher forward dividend yields $\mathbf{D}^a_{t+1}/\mathbf{P}^a_t$. For example, suppose the household simply purchases each individual asset a in proportion to its initial share of the value of its dividend-paying assets, so that $\widetilde{\omega}^a_{j,t} = \omega^a_{j,t}$. If the dividends paid out for each individual asset are unchanged ($\mathbf{D}^a_{t+1} = \mathbf{D}^a_t$), the household would still receive higher dividend income $D_{j,t+1}$ relative to the prior year t.
- The household actively rebalances its holdings of dividend-paying assets towards those with higher forward dividend yields. For example, suppose the household sells its holdings of one low-yield asset a in order to purchase another high-yield asset a' with $\mathbf{D}^{a'}_{t+1}/\mathbf{P}^{a'}_t > \mathbf{D}^a_{t+1}/\mathbf{P}^a_t$, and makes no other change in its holdings (so that $-S^a_{j,t} = S^{a'}_{j,t}$ and $S^D_{j,t} = 0$). The household will receive higher dividend income $D_{j,t+1}$ relative to the prior year t, despite not investing any *additional* wealth in dividend-paying assets.
- The assets initially held by the household pay higher dividends in year $t+1$, while holdings are unchanged ($S^a_{j,t} = 0$ for every asset a).

This decomposition is used below to clarify the assumptions underlying the capitalization method.

Motivated by the expression for net savings $S^a_{j,t}$ in a single asset a from Equation (A.2), I now describe my approach for estimating *total* net savings across all dividend-paying assets, $S^D_{j,t}$. The *reference portfolio* of dividend-paying assets with weights $\varpi^a_{j,t}$ is taken as given.Given these weights, the forward dividend yield and realized dividend growth of the reference portfolio are computed as

$$\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{P}}^D_t} = \sum_{a\in\mathcal{A}_D} \varpi^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}, \qquad \frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t} = \sum_{a\in\mathcal{A}_D} \varpi^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} \tag{A.5}$$

Given this reference portfolio, I define household j's *imputed net savings for dividend-paying assets* as

$$\widehat{S}^D_{j,t} = \frac{\overline{\mathbf{P}}^D_t}{\overline{\mathbf{D}}^D_{t+1}}\left[D_{j,t+1} - \frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t}D_{j,t}\right] \tag{A.6}$$

I refer to this as *imputed* savings because it may differ from the true unobserved savings $S^D_{j,t}$. A convenient expression for this discrepancy can be derived in three steps, by first writing the change in total dividend income as:

$$D_{j,t+1} - D_{j,t} = S^D_{j,t}\sum_{a\in\mathcal{A}_D}\omega^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t} + \left(W^D_{j,t} + S^D_{j,t}\right)\sum_{a\in\mathcal{A}_D}\left(\widetilde{\omega}^a_{j,t} - \omega^a_{j,t}\right)\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t} + \sum_{a\in\mathcal{A}_D} D^a_{j,t}\left(\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} - 1\right) \tag{A.7}$$

The excess change in total dividend income is then written as:

$$\begin{aligned}
&D_{j,t+1} - \frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t}D_{j,t} \\
&= D_{j,t+1} - D_{j,t} - \underbrace{D_{j,t}}_{=\sum_{a\in\mathcal{A}_D} D^a_{j,t}}\left(\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t} - 1\right) \\
&= S^D_{j,t}\sum_{a\in\mathcal{A}_D}\omega^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t} + \left(W^D_{j,t} + S^D_{j,t}\right)\sum_{a\in\mathcal{A}_D}\left(\widetilde{\omega}^a_{j,t} - \omega^a_{j,t}\right)\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t} + \sum_{a\in\mathcal{A}_D}\underbrace{D^a_{j,t}}_{=\alpha^a_{j,t}D_{j,t}}\left(\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} - \frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t}\right)
\end{aligned} \tag{A.8}$$

Finally, imputed net savings can be written as

$$
\begin{aligned}
\widehat{S}^D_{j,t} &= \frac{\overline{\mathbf{P}}^D_t}{\overline{\mathbf{D}}^D_{t+1}} \left[D_{j,t+1} - \frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t} D_{j,t} \right] \\
&= S^D_{j,t} \left(\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{P}}^D_t} \right)^{-1} \sum_{a \in \mathcal{A}_D} \omega^a_{j,t} \frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t} \\
&\quad + \left(\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{P}}^D_t} \right)^{-1} \left(W^D_{j,t} + S^D_{j,t} \right) \sum_{a \in \mathcal{A}_D} \left(\widetilde{\omega}^a_{j,t} - \omega^a_{j,t} \right) \frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t} \\
&\quad + \left(\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{P}}^D_t} \right)^{-1} D_{j,t} \sum_{a \in \mathcal{A}_D} \alpha^a_{j,t} \left(\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} - \frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t} \right)
\end{aligned}
\tag{A.9}
$$

This decomposition highlights three sufficient conditions under which imputed net savings $\widehat{S}^D_{j,t}$ for household j exactly matches its true unobserved net savings $S^D_{j,t}$:

1. The dividend yield of the reference portfolio $\overline{\mathbf{D}}^D_{t+1}/\overline{\mathbf{P}}^D_t$ is equal to the average dividend yield of household j's initial portfolio of dividend-paying assets $\sum_{a \in \mathcal{A}_D} \omega^a_{j,t} \mathbf{D}^a_{t+1}/\mathbf{P}^a_t$. Note that these weights $\omega^a_{j,t}$ are defined *before* the household rebalances at time t, based on holdings at time $t-1$ (but asset prices at time t).

2. The household does not rebalance into assets with a different forward dividend yield than the reference portfolio: $\sum_{a \in \mathcal{A}_D} \widetilde{\omega}^a_{j,t} \mathbf{D}^a_{t+1}/\mathbf{P}^a_t = \sum_{a \in \mathcal{A}_D} \omega^a_{j,t} \mathbf{D}^a_{t+1}/\mathbf{P}^a_t$. Note that this does not rule out rebalancing *across* dividend-paying assets vs. other assets; it only restricts rebalancing *within* the set of dividend-paying assets.

3. The household's initial portfolio of dividend-paying assets has the same weighted average dividend growth rate as the reference portfolio, where the weights are determined by each asset's share of total dividend income in year t: $\sum_{a \in \mathcal{A}_D} D^a_{j,t} \frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} = D_{j,t} \frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t}$.

These assumptions are unlikely to all hold for a single household j, and thus $\widehat{S}^D_{j,t}$ is likely to be an imprecise estimate of $S^D_{j,t}$ for a single household. However, I will ultimately be computing averages of $\widehat{S}^D_{j,t}$ over many households (or using it as the dependent variable in regressions with data from many households). In Appendix A.2 I present sufficient conditions under which the difference in these average imputed savings across groups is a valid estimate of the true unobserved difference in average savings across groups.

The previous discussion focused solely on the case of dividend-paying assets, and imputing net savings $S^D_{j,t}$ in these assets from observed dividend income $D_{j,t}$. For assets $\mathcal{A}_N$ that pay nontaxable interest $\mathbf{N}^a_{j,t}$, I define a reference portfolio and estimate household j's net savings $S^N_{j,t,t+1}$ from its observed nontaxable interest income $N_{j,t} = \sum_{a \in \mathcal{A}_N} Q^a_{j,t-1} \mathbf{N}^a_t$ as

$$
\widehat{S}^N_{j,t,t+1} = \frac{\overline{\mathbf{P}}^N_t}{\overline{\mathbf{N}}^N_{t+1}} \left[N_{j,t+1} - \frac{\overline{\mathbf{N}}^N_{t+1}}{\overline{\mathbf{N}}^N_t} N_{j,t} \right] \tag{A.10}
$$

For assets $\mathcal{A}_I$ that pay taxable interest income $\mathbf{I}^a_t$, I instead directly estimate each household's average taxable interest-generating wealth from time $t-1$ to t — denoted by $\widehat{\overline{W}}^I_{j,t}$ — using the method described in Appendix A.4, then define the imputed savings flow in taxable interest-generating assets as the difference in imputed average wealth:

$$\widehat{S}^I_{j,t,t+1} = \widehat{\overline{W}}^I_{j,t+1} - \widehat{\overline{W}}^I_{j,t} \tag{A.11}$$

This definition assumes that households earn no capital gains or losses on their taxable interest-generating assets, so that changes in imputed wealth are entirely attributable to active savings rather than passive asset price revaluation effects. This assumption is reasonable for most households, since bank savings accounts and nonmarketable U.S. Treasury savings bonds account for a large share of all taxable interest-generating assets held by households. However, the rate of return used to capitalize this taxable interest income varies both over time and across households (consistent with the empirical evidence from Smith, Zidar, and Zwick, 2023). Appendix A.4 contains further details.

A.2 Imputed savings for groups of households

In the data, I ultimately use the millions of tax unit observations in the data to compute averages of imputed savings $\widehat{S}^D_{j,t}$ across many households in a given year. In this section, I establish conditions under which these averages provide an unbiased estimate of the average unobserved net savings $S^D_{j,t}$ for the same group of households.

I first introduce additional notation. For a group of households $j \in \mathcal{J}$, let $\mathbb{E}_{\mathcal{J}}[X_j] \equiv \frac{1}{|\mathcal{J}|}\sum_{j\in\mathcal{J}} X_j$ denote the cross-sectional average of some variable X_j defined at the household level. Using this notation, the cross-sectional average of the imputed savings measure $\widehat{S}^D_{j,t}$ can be written as:

$$\begin{aligned} &\mathbb{E}_{\mathcal{J}}\left[\widehat{S}^D_{j,t}\right] \\ &= \left(\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{P}}^D_t}\right)^{-1} \mathbb{E}_{\mathcal{J}}\left[S^D_{j,t}\sum_{a\in\mathcal{A}_D}\omega^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}\right] \\ &\quad + \left(\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{P}}^D_t}\right)^{-1} \mathbb{E}_{\mathcal{J}}\left[\widetilde{W}^D_{j,t}\sum_{a\in\mathcal{A}_D}\left(\widetilde{\omega}^a_{j,t}-\omega^a_{j,t}\right)\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}\right] \\ &\quad + \left(\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{P}}^D_t}\right)^{-1} \mathbb{E}_{\mathcal{J}}\left[D_{j,t}\sum_{a\in\mathcal{A}_D}\alpha^a_{j,t}\left(\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t}-\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t}\right)\right] \end{aligned} \tag{A.12}$$

From this expression, a sufficient set of assumptions can be derived that will guarantee $\mathbb{E}_{\mathcal{J}}\left[\widehat{S}^D_{j,t}\right] = \mathbb{E}_{\mathcal{J}}\left[S^D_{j,t}\right]$, so that this cross-sectional average is an unbiased estimate of the unobserved average savings $S^D_{j,t}$.

Assumptions to estimate $\mathbb{E}_{\mathcal{J}}\left[S^D_{j,t}\right]$:

1. Across households j, the unweighted average of the forward dividend yield $\sum_{a\in\mathcal{A}_D}\omega^a_{j,t}\mathbf{D}^a_{t+1}/\mathbf{P}^a_t$

is equal to the forward dividend yield $\overline{\mathbf{D}}^D_{t+1}/\overline{\mathbf{P}}^D_t$ of the reference portfolio:

$$\frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{P}}^D_t} = \mathbb{E}_{\mathcal{J}}\left[\sum_{a\in\mathcal{A}_D} \omega^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}\right]$$

2. Across households j, the forward dividend yield is uncorrelated with net savings $S^D_{j,t}$:

$$\mathbb{E}_{\mathcal{J}}\left[S^D_{j,t}\sum_{a\in\mathcal{A}_D} \omega^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}\right] = \mathbb{E}_{\mathcal{J}}\left[S^D_{j,t}\right]\cdot\mathbb{E}_{\mathcal{J}}\left[\sum_{a\in\mathcal{A}_D} \omega^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}\right]$$

3. Across households j, the wealth-weighted average of the forward dividend yield is the same before and after rebalancing at time t (where wealth weights are based on post-rebalancing wealth $\widetilde{W}^D_{j,t}$):

$$\mathbb{E}_{\mathcal{J}}\left[\frac{\widetilde{W}^D_{j,t}}{\mathbb{E}_{\mathcal{J}}\left[\widetilde{W}^D_{j,t}\right]}\sum_{a\in\mathcal{A}_D} \widetilde{\omega}^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}\right] = \mathbb{E}_{\mathcal{J}}\left[\frac{\widetilde{W}^D_{j,t}}{\mathbb{E}_{\mathcal{J}}\left[\widetilde{W}^D_{j,t}\right]}\sum_{a\in\mathcal{A}_D} \omega^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{P}^a_t}\right]$$

4. Define the counterfactual dividend income growth for household j as

$$\sum_{a\in\mathcal{A}_D}\alpha^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} = \sum_{a\in\mathcal{A}_D}\frac{D^a_{j,t}}{D_{j,t}}\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} = \sum_{a\in\mathcal{A}_D}\frac{Q^a_{j,t-1}\mathbf{D}^a_t}{D_{j,t}}\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} = \frac{1}{D_{j,t}}\sum_{a\in\mathcal{A}_D}Q^a_{j,t-1}\mathbf{D}^a_{t+1}$$

Across households j, the total dividend income-weighted average ($D_{j,t}$) of this counterfactual dividend income growth is equal to the dividend growth of the reference portfolio $\overline{\mathbf{D}}^D_{t+1}/\overline{\mathbf{D}}^D_t$:

$$\mathbb{E}_{\mathcal{J}}\left[\frac{D_{j,t}}{\mathbb{E}_{\mathcal{J}}\left[D_{j,t}\right]}\sum_{a\in\mathcal{A}_D}\alpha^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t}\right] = \frac{\overline{\mathbf{D}}^D_{t+1}}{\overline{\mathbf{D}}^D_t}$$

These assumptions allow individual households' asset holdings and portfolio weights within the set of dividend-paying assets — $\{Q^a_{j,t-1}\}_{a\in\mathcal{A}_D}$ and $\{\omega^a_{j,t} = \mathbf{P}^a_t Q^a_{j,t-1}\}_{a\in\mathcal{A}_D}$ — to deviate from the reference portfolio weights $\{\overline{\omega}^a_{j,t}\}_{a\in\mathcal{A}_D}$. These deviations do not generate bias in estimates of average net savings, as long as the average forward dividend yield and ex post dividend growth rates of those portfolios are the same as the values for the reference portfolio used to compute $\widehat{S}^D_{j,t}$ (Assumptions 1 and 4). These assumptions also allow for households to rebalance or purchase assets with a different composition in the reference portfolio, provided that high-wealth households do not rebalance into other dividend-paying assets with much higher or lower yields (Assumption 3). Assumption 2 rules out the possibility that most saving within group $\mathcal{J}$ is done by households with unusually high or low dividend yields (which would imply that these savings flows are either under-reflected or over-reflected in the group's future dividend income).

In the empirical work, I frequently compare average imputed savings between two different groups of households $\mathcal{J}$ and $\mathcal{J}'$, either explicitly or implicitly in regressions with $\widehat{S}^D_{j,t}$ as the dependent variable. In many cases these comparisons involve groups of households that are similar up through a particular point in time t — in terms of characteristics such as past nonfinancial income $Y_{j,t}$ or dividend income $D_{j,t}$ — but experience different outcomes at time $t+1$. For example, group $\mathcal{J}$ might consist of households where the employer of the household's primary earner experiences a large decline in total employment or revenue from time t to $t+1$, while group $\mathcal{J}'$ consists of ex ante similar households with no large decline in their employer's total employment or revenue over the same period. For these comparisons, the last assumption above can be relaxed to allow some common misspecification in the dividend growth rate of the reference portfolio $\overline{\mathbf{D}}^D_{t+1}/\overline{\mathbf{D}}^D_t$ across both households, which will be "differenced out" of the average savings flows (or absorbed by a time fixed effect in the regression).

Assumptions to estimate $\mathbb{E}_{\mathcal{J}}\left[S^D_{j,t}\right] - \mathbb{E}_{\mathcal{J}'}\left[S^D_{j,t}\right]$ for two groups with the same initial dividend income $\mathbb{E}_{\mathcal{J}}\left[D_{j,t}\right] = \mathbb{E}_{\mathcal{J}'}\left[D_{j,t}\right]$:

1. The previous Assumptions 1, 2 and 3 hold within each group $\mathcal{J}$ and $\mathcal{J}'$.
2. The total dividend income-weighted counterfactual dividend income growth rate is the same for both groups:

$$\mathbb{E}_{\mathcal{J}}\left[\frac{D_{j,t}}{\mathbb{E}_{\mathcal{J}}\left[D_{j,t}\right]}\sum_{a\in\mathcal{A}_D}\alpha^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t}\right] = \mathbb{E}_{\mathcal{J}'}\left[\frac{D_{j,t}}{\mathbb{E}_{\mathcal{J}}\left[D_{j,t}\right]}\sum_{a\in\mathcal{A}_D}\alpha^a_{j,t}\frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t}\right]$$

Discussion of assumptions: I now discuss the assumptions underlying the group-level estimates of net savings, together with relevant empirical evidence from datasets other than US administrative income tax records. It will be useful to continue with the example of household groups $\mathcal{J}$ and $\mathcal{J}'$ introduced above, which share similar ex ante characteristics (including average taxable dividend income) through time t but then differ in the ex post outcomes of their employer at time $t+1$ (with employers of group-$\mathcal{J}$ households reducing their total employment sharply, and employers of group-$\mathcal{J}'$ households not reducing their employment).

Assumption 1 requires that the average dividend yield of households' holdings matches the dividend yield of the reference portfolio, which is necessary to correctly capitalize the excess change in dividend income and impute the dollar value of households' average net savings. If the reference portfolio's dividend yield was misspecified but all other assumptions held, then the average estimated savings $\widehat{S}^D_{j,t}$ for a given group would be distorted by a multiplicative factor; the magnitude of true unobserved savings $S^D_{j,t}$ would be overestimated if the reference portfolio's dividend yield was too low, and underestimated if it was too high.

When comparing average imputed savings across groups, Assumption 1 may be violated if there is significant heterogeneity across groups in the composition of their taxable dividend-paying assets. Differences in portfolio allocation *across* asset classes — to taxable dividend-paying assets vs. interest-paying assets — do not pose a problem for the capitalization approach. However, households employed at cyclically-sensitive firms may earn their dividend income primarily from safe low-yield bond funds or growth stocks, while households employed at cyclically-insensitive firms may earn their dividend income from risky higher-yield assets. I argue that such heterogeneity is unlikely to affect my estimates of relative savings flows across groups, for two reasons.

First, the comparisons in the empirical analysis primarily focus on groups of households experiencing different ex post shocks (e.g. to their employer), while *ex ante* risk exposures determine their portfolio allocation. Formally, suppose that a measure of ex ante risk exposure $risk_{f,t-1}$ is observed for firm f. Each household j takes into account the risk of its employer $f(j)$ when making its initial portfolio allocation decision at time $t-1$, and the dividend yield of its portfolio is given by the equation

$$\sum_{a \in \mathcal{A}_D} \omega_{j,t}^a \frac{\mathbf{D}_{t+1}^a}{\mathbf{P}_t^a} = \alpha_{\mathbf{D}/\mathbf{P}|risk} + \beta_{\mathbf{D}/\mathbf{P}|risk} \cdot risk_{f(j),t} \tag{A.13}$$

The ex post growth $\Delta emp_{f,t+1}$ in the total employment of firm f is related to its ex ante risk exposure by the following regression:

$$\Delta emp_{f,t+1} = \alpha_{\Delta emp|risk} + \beta_{\Delta emp|risk} \cdot risk_{f(j),t-1} + \varepsilon_{f,t+1} \tag{A.14}$$

A regression of household j's dividend yields on the ex post employment growth of their employer $f(j)$ would yield the coefficient

$$\begin{aligned}
\beta_{\mathbf{D}/\mathbf{P}|\Delta emp} &= \frac{Cov_j\left(\sum_{a \in \mathcal{A}_D} \omega_{j,t}^a \mathbf{D}_{t+1}^a/\mathbf{P}_t^a,\ \Delta emp_{f(j),t+1}\right)}{Var_j(\Delta emp_{f(j),t+1})} \\
&= \frac{Cov_j\left(\beta_{\mathbf{D}/\mathbf{P}|risk} \cdot risk_{f(j),t-1},\ \beta_{\Delta emp|risk} \cdot risk_{f(j),t-1} + \varepsilon_{f,t+1}\right)}{Var_j(\Delta emp_{f(j),t+1})} \\
&= \beta_{\mathbf{D}/\mathbf{P}|risk} \cdot \beta_{\Delta emp|risk} \cdot \frac{Var_j\left(risk_{f(j),t-1}\right)}{Var_j(\Delta emp_{f(j),t+1})} \\
&= \beta_{\mathbf{D}/\mathbf{P}|risk} \cdot \frac{Cov_j\left(\Delta emp_{f(j),t+1},\ risk_{f(j),t-1}\right)}{Var_j(\Delta emp_{f(j),t+1})} \\
&= \beta_{\mathbf{D}/\mathbf{P}|risk} \cdot \frac{[Cov_j\left(\Delta emp_{f(j),t+1},\ risk_{f(j),t-1}\right)]^2}{Var_j(\Delta emp_{f(j),t+1}) \cdot Var_j\left(risk_{f(j),t-1}\right)} \cdot \frac{Var_j\left(risk_{f(j),t-1}\right)}{Cov_j\left(\Delta emp_{f(j),t+1},\ risk_{f(j),t-1}\right)} \\
&= \beta_{\mathbf{D}/\mathbf{P}|risk} \cdot R^2_{\Delta emp|risk}/\beta_{\Delta emp|risk}
\end{aligned} \tag{A.15}$$

where $R^2_{\Delta emp|risk}$ denotes the R-squared value from the regression (A.14), equal to the squared correlation between the two variables. As long as this R^2 is relatively low — that is, there is enough idiosyncratic variation $\varepsilon_{f,t+1}$ in ex post employment growth which is not predicted by the ex ante risk measure $risk_{f(j),t-1}$ that households use in making their portfolio decisions — then average dividend yields should not differ significantly for groups of households sorted on their employer's ex post employment growth $\Delta emp_{f(j),t+1}$.

Second, my empirical work focuses on households that receive taxable dividend income, the vast majority of whom are stockholders.[33] Existing empirical studies estimate a relatively small correlation between measures of cyclical labor income risk exposure and stock shares of *total financial wealth* among households that participate in the stock market, and a larger correlation among all households attributable to differences in stock market participation.[34] These small estimates serve as a likely upper bound on differences in the share of *dividend-paying assets* invested in stocks vs. safe dividend-paying assets (e.g. money market and other fixed income mutual funds), since safe interest-paying assets such as savings accounts account for a large share of the non-stock component of total financial wealth. This evidence suggests that there are unlikely to be significant differences in the composition of dividend-paying assets across the groups of households that are compared in the empirical analysis.

Assumption 2 imposes that savings flows in dividend-paying assets are uncorrelated with portfolio dividend yields across households *within* each group. If households within the group with higher net savings flows also have higher dividend yields, then the average active savings flow would be overestimated from the change in dividend income (and underestimated if the highest savers have lower dividend yields). Most existing evidence on differential savings behavior and investment returns focuses on comparing households with very different ex ante characteristics, such as different age, wealth, or income groups.[35] Within a group of households that are initially similar along these dimensions, annual savings flows in dividend-paying assets are likely driven by idiosyncratic rebalancing decisions or liquidity needs, and thus seem plausibly uncorrelated with portfolio dividend yields.

Assumption 3 imposes that households do not actively rebalance their portfolio of dividend-paying assets in a way that substantially changes the wealth-weighted average dividend yield. This assumption may be violated if households actively rebalance in response to an income shock. For example, if households facing a large decline in income rebalance their dividend-paying assets towards safer low-yield assets (such as money market mutual funds), then their decline in dividend income would be partly misattributed to an active liquidation of their dividend-paying assets (rather than rebalancing). Existing estimates suggest that such rebalancing in response to income shocks is likely to be small. Using panel data on both retirement and non-retirement (taxable) accounts from a large U.S. financial institution, Meeuwis (2022) finds that a persistent 10% shock to an investor's wage income is associated with a change in the stock share of only 0.3 percentage points.[36] These data are informative about rebalancing

[33]

[34]Based on Swedish administrative data, Figure 1 of Catherine, Sodini, and Zhang (2024) shows a difference in conditional stock shares of financial wealth (among households that participate in the stock market) of roughly 5 percentage points across workers in education-by-industry groups with the highest versus lowest cyclical skewness in labor income growth rates, relative to an average stock share of 45% among participants. The difference in unconditional stock shares (among all households) between the same groups is roughly twice as large, due to a nearly 20 percentage point difference in participation rates.

[35]See e.g. Saez and Zucman (2016), Fagereng, Holm, Moll, and Natvik (2021), and Smith, Zidar, and Zwick (2023).

[36]See the estimates reported in Table 4 of Meeuwis (2022). Table 1 reports an average stock share of 0.77 for individuals' wealth invested in the account, so the estimated coefficient of 0.04 from regressing log stock shares on persistent income changes implies a change in the stock share of $[\exp(0.10 \cdot 0.04) - 1] \cdot 0.77 \approx 0.003$ or 0.3 percentage points in response to a 10% income shock.

within the set of dividend-paying assets because the assets in these accounts all pay dividends to investors (either qualified or non-qualified). While these estimates imply large expected differences in portfolio shares across the full income distribution, they also imply that such rebalancing within the set of dividend-paying assets is unlikely to distort my imputed savings estimates.

Assumption 4 — that average counterfactual dividend growth rates are the same for the initial portfolios held by each group — may be violated if a portion of households' taxable stockholdings are held in shares of their employer (for example, through equity-based compensation packages) or other local firms.[37] For group-$\mathcal{J}$ households, if the same employers that reduced employment also reduced their dividend payouts to shareholders, and employer stock represents a large share of households' taxable stockholdings, then the passive decline in dividend income received by these households from their employer stock would be misattributed to active liquidation of their stockholdings. If $\overline{\alpha}_t^{empl}$ is the share of each household's total dividend income received from its employer stockholdings in year t and $\mathbf{D}_{t+1}^{empl(j)}/\mathbf{D}_t^{empl(j)}$ is the dividend growth of household j's employer, then this bias is given by

$$\begin{aligned}&\mathbb{E}_{\mathcal{J}}\left[\widehat{S}_{j,t}^{D}\right]-\mathbb{E}_{\mathcal{J}'}\left[\widehat{S}_{j,t}^{D}\right]\\&=\mathbb{E}_{\mathcal{J}}\left[S_{j,t}^{D}\right]-\mathbb{E}_{\mathcal{J}'}\left[S_{j,t}^{D}\right]+\overline{\alpha}_t^{empl}\left(\frac{\overline{\mathbf{D}}_{t+1}^{D}}{\overline{\mathbf{P}}_t^{D}}\right)^{-1}\mathbb{E}_{\mathcal{J}}\left[D_{j,t}\right]\left\{\mathbb{E}_{\mathcal{J}}\left[\frac{\mathbf{D}_{t+1}^{empl(j)}}{\mathbf{D}_t^{empl(j)}}\right]-\mathbb{E}_{\mathcal{J}'}\left[\frac{\mathbf{D}_{t+1}^{empl(j)}}{\mathbf{D}_t^{empl(j)}}\right]\right\}\end{aligned} \tag{A.16}$$

Importantly, the degree of bias is determined by the *dividend income share* $\overline{\alpha}_t^{empl}$ of households' own-employer stockholdings, rather than the *portfolio value share* ω_t^a. Eisfeldt, Falato, Lee, and Xiaolan (2024) show that equity-based compensation is much more prevalent among publicly traded firms that do not pay dividends.

A.3 Accounting for time aggregation

This section relaxes the assumption from Appendix A.1 that all asset transactions take place at the start of each calendar year. I show that the excess change in a household's dividend income between consecutive calendar years t and $t+1$ (as defined previously) is equal to a weighted sum of all of the household's asset transactions $dQ_{j,\tau}^a$ over both calendar years. These weights are largest for transactions that take place near the end of calendar year t or the start of calendar year $t+1$, since they have a large effect on dividend income received in year $t+1$ but not in year t. This suggests a time-averaging adjustment to the capitalization factor used to impute net savings from excess changes in dividend income, which is then used when computing these capitalization factors in the data. The imputed savings measure $\widehat{S}_{j,t}^D$ estimates a similar weighted average of savings flows $dS_{j,\tau}^a$ over the full two-year period, with

[37] Benartzi (2001), Grinblatt and Keloharju (2001), Massa and Simonov (2006), and Døskeland and Hvide (2011) provide evidence of households' bias in their stock portfolio allocation towards their current and past employers, other firms in the same industry, and local firms.

the most weight on savings flows near the turn of the calendar year.

Throughout this section, t continues to denote integer values corresponding to the start of a calendar year. Non-integer values corresponding to other times within the calendar year are denoted by τ (e.g. $t = 2009$ for the start of the 2009 calendar year and the end of the 2008 calendar year, and $\tau = 2008.5$ for the midpoint of the 2008 calendar year). Each asset a pays dividends as a continuous flow with rate δ_τ^a per share at time τ, so that the total dividends paid out to one share of asset a from time $t-1$ to t is given by

$$\mathbf{D}_t^a = \int_{t-1}^{t} \delta_\tau^a d\tau \tag{A.17}$$

$Q_{j,\tau}^a$ denotes the shares of asset a held by household j at time τ within the calendar year, and $dQ_{j,\tau}^a$ to denote the change at time τ in the quantity of shares held. For convenience, trading is assumed to take place at fixed times $\tau = t, t+1/N, \ldots, t+(N-1)/N$ within each calendar year. I write the Riemann-Stieltjes integral

$$\int_{\tau_1}^{\tau_2} F_\tau dQ_{j,\tau}^a \equiv \sum_{i=i_1}^{i=i_2} F_{t+i/N}(Q_{j,t+i/N}^a - Q_{j,t+(i-1)/N}^a) \tag{A.18}$$

where $t + i_1/N < \tau_1 < t + (i_1+1)/N$ and $t + i_2/N < \tau_2 < t + (i_2+1)/N$. I define $dS_{j,\tau}^a = \mathbf{P}_\tau^a dQ_{j,\tau}^a$ as net savings in asset a from transactions at time τ.

The total dividend income received by household j from asset a over the period $t-1$ to t is given by

$$\begin{aligned} D_{j,t}^a &= \int_{t-1}^{t} Q_{j,\tau}^a \delta_\tau^a d\tau \\ &= Q_{j,t-1}^a \mathbf{D}_t^a + \int_{t-1}^{t} (Q_{j,\tau}^a - Q_{j,t-1}^a)\delta_\tau^a d\tau \\ &= Q_{j,t-1}^a \mathbf{D}_t^a + \int_{t-1}^{t} \left(\int_{t-1}^{\tau} dQ_{j,\tau'}^a \right) \delta_\tau^a d\tau \\ &= \underbrace{Q_{j,t-1}^a \mathbf{D}_t^a}_{\text{due to initial holdings}} + \underbrace{\int_{t-1}^{t} \left(\int_{\tau}^{t} \delta_{\tau'}^a d\tau' \right) dQ_{j,\tau}^a}_{\text{due to change in holdings}} \end{aligned} \tag{A.19}$$

where the order of integration is swapped so that this income can be separately attributed to initial holdings $Q_{j,t-1}^a$ and transactions $dQ_{j,\tau}^a$ that take place throughout the year. Transactions that take place early in the calendar year are reflected the most in annual income $D_{j,t}^a$ because they affect flow dividends received throughout nearly the entire calendar year. On the other hand, transactions that take place late in the calendar year are reflected the least in annual income $D_{j,t}^a$ because they only affect flow dividends received over a small portion of the calendar year. I use a similar decomposition for year $t+1$ income:

$$\begin{aligned}
D^a_{j,t+1} &= \int_t^{t+1} Q^a_{j,\tau}\delta^a_\tau d\tau \\
&= Q^a_{j,t-1}\mathbf{D}^a_{t+1} + \int_t^{t+1} (Q^a_{j,\tau} - Q^a_{j,t-1})\delta^a_\tau d\tau \\
&= Q^a_{j,t-1}\mathbf{D}^a_{t+1} + \int_t^{t+1} \left(\int_{t-1}^{\tau} dQ^a_{j,\tau'}\right)\delta^a_\tau d\tau \\
&= \underbrace{Q^a_{j,t-1}\mathbf{D}^a_{t+1}}_{\text{due to initial holdings at } t-1} + \underbrace{\mathbf{D}^a_{t+1}\int_{t-1}^{t} dQ^a_{j,\tau} + \int_t^{t+1}\left(\int_\tau^{t+1}\delta^a_{\tau'}d\tau'\right)dQ^a_{j,\tau}}_{\text{due to change in holdings}}
\end{aligned} \tag{A.20}$$

With $t-1$ used as the base period in this decomposition, transactions that take place before time t are reflected in flow dividends that accrue over the full year from time t to $t+1$. Transactions that take place after time t are reflected in flow dividends that accrue over only the remainder of the same calendar year. Excess dividend growth can then be written as a weighted sum of transactions $dQ^a_{j,\tau}$ from $t-1$ to $t+1$:

$$\begin{aligned}
&D^a_{j,t+1} - \frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t}D^a_{j,t} \\
&= \left[Q^a_{j,t-1}\mathbf{D}^a_{t+1} + \mathbf{D}^a_{t+1}\int_{t-1}^{t} dQ^a_{j,\tau} + \int_t^{t+1}\left(\int_\tau^{t+1}\delta^a_{\tau'}d\tau'\right)dQ^a_{j,\tau}\right] \\
&\quad - \frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t}\left[Q^a_{j,t-1}\mathbf{D}^a_t + \int_{t-1}^{t}\left(\int_\tau^{t}\delta^a_{\tau'}d\tau'\right)dQ^a_{j,\tau}\right] \\
&= \mathbf{D}^a_{t+1}\left[\int_{t-1}^{t}\left(1 - \frac{\int_\tau^t \delta^a_{\tau'}d\tau'}{\mathbf{D}^a_t}\right)dQ^a_{j,\tau} + \int_t^{t+1}\left(\frac{\int_\tau^{t+1}\delta^a_{\tau'}d\tau'}{\mathbf{D}^a_{t+1}}\right)dQ^a_{j,\tau}\right] \\
&= \mathbf{D}^a_{t+1}\int_{t-1}^{t+1}\chi^a_{\tau|t}dQ^a_{j,\tau}
\end{aligned} \tag{A.21}$$

where

$$\chi^a_{\tau|t} = \begin{cases} 1 - \left(\int_\tau^t \delta^a_{\tau'}d\tau'\right)/\mathbf{D}^a_t & \text{if } t-1 \le \tau \le t \\ 1 - \left(\int_t^\tau \delta^a_{\tau'}d\tau'\right)/\mathbf{D}^a_{t+1} = \left(\int_\tau^{t+1}\delta^a_{\tau'}d\tau'\right)/\mathbf{D}^a_{t+1} & \text{if } t < \tau \le t+1 \end{cases} \tag{A.22}$$

If the flow rate of dividends paid per share δ^a_τ is constant within each calendar year (i.e. $\delta^a_\tau = \mathbf{D}_t$ for $\tau \in [t-1,t)$ and $\delta^a_\tau = \mathbf{D}_{t+1}$ for $\tau \in [t,t+1)$) then the weights take the simple form $\chi^a_{\tau|t} = 1 - |\tau - t|$. Transactions that take place at time t are reflected the most in the excess change in dividend income because they affect $D^a_{j,t+1}$ but not $D^a_{j,t}$. Transactions that take place exactly at time $t-1$ are not reflected in the excess change because they affect dividend income in both years, while transactions that take place exactly at time $t+1$ are not reflected because they do not affect income in either year.[38]

[38] The structure of these weights is similar to the approximations derived by Mariano and Murasawa (2003) to relate the growth of low-frequency time-aggregated data series to higher-frequency growth rates (e.g. the change in annual GDP between two calendar years and growth rates in each quarter).

The excess change in dividend income can alternatively be written in terms of savings flows $dS^a_{j,\tau} = \mathbf{P}^a_\tau dQ^a_{j,\tau}$ rather than transactions:

$$D^a_{j,t+1} - \frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} D^a_{j,t} = \mathbf{D}^a_{t+1} \int_{t-1}^{t+1} \frac{1}{\mathbf{P}^a_{j,\tau}} \chi^a_{\tau|t} dS^a_{j,\tau} \tag{A.23}$$

If all of household j's savings flows $dS^a_{j,\tau}$ were observed, a weighted average of these transactions could be recovered by multiplying this excess change in annual dividend income by an "ideal" household-specific capitalization factor $\mathbf{P}^{a*}_{j,t}/\mathbf{D}^a_{t+1}$:

$$\int_{t-1}^{t+1} \chi^a_{\tau|t} dS^a_{j,\tau} = \frac{\mathbf{P}^{a*}_{j,t}}{\mathbf{D}^a_{t+1}} \left[D^a_{j,t+1} - \frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} D^a_{j,t} \right] \tag{A.24}$$

where the "ideal" average share price $\mathbf{P}^{a*}_{j,t}$ is a weighted harmonic average with weights $\chi^a_{\tau|t} dS^a_{j,\tau}$ determined by savings flows at each date (some of which may be negative):

$$\mathbf{P}^{a*}_{j,t} = \frac{\int_{t-1}^{t+1} \chi^a_{\tau|t} dS^a_{j,\tau}}{\int_{t-1}^{t+1} \frac{1}{\mathbf{P}^a_{j,\tau}} \chi^a_{\tau|t} dS^a_{j,\tau}} \tag{A.25}$$

In practice, I approximate the weights as $\chi^a_{\tau|t} \approx 1 - |t-\tau|$ and replace the (unobserved) savings component of the weight $dS^a_{j,\tau}$ with the time increment $d\tau$ to obtain a simpler share price index:[39]

$$\overline{\mathbf{P}}^a_t = \left(\int_{t-1}^{t+1} \frac{1}{\mathbf{P}^a_{j,\tau}} (1 - |t-\tau|) d\tau \right)^{-1} \tag{A.26}$$

These approximations are reasonable as long as the flow rate of dividends-per-share δ^a_τ does not vary significantly within each calendar year, and the savings flows $dS^a_{j,\tau}$ that are weighted heavily (for periods $\tau \approx t$) are not clustered in periods where $\mathbf{P}^a_{j,\tau}$ differs significantly from $\overline{\mathbf{P}}^a_t$. Note that while the share price is averaged over the full window from time $t-1$ to $t+1$, the annual dividends-per-share $\mathbf{D}^a_t$ and $\mathbf{D}^a_{t+1}$ are aggregated only within each calendar year.

The final estimate $\widehat{S}^a_{j,t}$ of the weighted-average savings flow in asset a — based only on household j's annual dividend income $D^a_{j,t}$ and $D^a_{j,t+1}$ from the asset in each year, the asset's weighted average share price $\overline{\mathbf{P}}^a_t$, and its total dividends $\mathbf{D}^a_t$ and $\mathbf{D}^a_{t+1}$ paid in each year — is given by:

$$\widehat{S}^a_{j,t} \equiv \frac{\overline{\mathbf{P}}^a_t}{\mathbf{D}^a_{t+1}} \left[D^a_{j,t+1} - \frac{\mathbf{D}^a_{t+1}}{\mathbf{D}^a_t} D^a_{j,t} \right] \approx \int_{t-1}^{t+1} (1 - |t-\tau|) dS^a_{j,\tau} \tag{A.27}$$

A.4 Capitalizing taxable interest income

Individuals receive taxable interest income primarily from their holdings of bank deposits; savings bonds, such as Series EE or I savings bonds issued by the U.S. Treasury; other government, corporate, and foreign bonds; and other boutique fixed income investments.[40] When

[39] The weights sum to $\int_{t-1}^{t+1} (1 - |t-\tau|) d\tau = 1$.

[40] These asset classes are listed in decreasing order of their aggregate share of all interest-generating assets.

capitalizing this interest income to estimate the average level of interest-generating wealth in a given tax year and net savings flows across two tax years for a given individual, two challenges arise. The first challenge is that average returns on interest-generating assets vary significantly across the income distribution. Smith, Zidar, and Zwick (2023) — referred to as SZZ throughout the rest of this section — show that individuals at the top of the income or wealth distribution earn higher returns on these fixed income investments, even within asset classes such as bank deposits. The second challenge is that many existing capitalization approaches impose sharp discontinuities in returns at specific points in the income or wealth distribution. These discontinuities may have little effect on the estimated top wealth shares that this existing literature has focused on. However, the resulting time series of capitalized wealth estimates for a given individual will exhibit large swings near these discontinuities, which can add significant noise to estimated net savings flows. Simple interpolation schemes which smooth the estimated rates of return over these discontinuities may imply an average rate of return across all households which differs from estimates based on aggregate data, which would distort the estimated level of wealth and net savings flows by a scale factor even if the relative rates of return across households are correctly estimated.

To confront the first challenge, I use SZZ's two-tier classical minimum-distance estimate of the average return on interest-generating fixed income investments for households ranking in the top 0.1% of the taxable interest income distribution in a given year. This estimate is a weighted sum of the 5-year U.S. Treasury yield and Moody's BAA bond yields, with weights estimated to match both the time series properties of this group's aggregate interest income, and average rate of return and level of this group's fixed income asset holdings from the SCF over the 1989-2016 period.[41] To confront the second challenge, I impute returns on taxable interest-generating wealth for the remaining 99.9% of households using a method that ensures estimated wealth is strictly increasing in interest income, estimated returns are strictly *decreasing* in interest income, and the implied wealth-weighted average return across all households is consistent with estimates from aggregate data. I also adjust for the non-negligible share of interest-generating wealth held by households that report no taxable interest income, since capitalization approaches that allocate zero interest-generating wealth to these households may significantly overstate the wealth of households that report low but positive interest income.

Accounting framework: I describe the capitalization method within an accounting frame-

[41] In the published paper and appendix, SZZ use two- and three-tier classical minimum-distance (CMD) estimates which separate households based on the estimated level of their non-interest-generating wealth. The administrative income tax data used in this paper do not contain the detailed information necessary to estimate important components of non-interest wealth such as private business wealth and housing. However, the publicly-posted replication files provide additional CMD parameter estimates based on groupings of households according to taxable interest income, total adjusted gross income, and adjusted gross income excluding interest income. I use the estimated weights for households grouped by taxable interest income provided in the replication file `cmd_replication/matlab/tex/top01-intinc-ust5-baa-btstrp200/parameters4.txt`. I use the original estimated weights over the original 1989-2016 sample period, but with updated data for the 5-year Treasury yield and Moody's BAA bond yield.

work similar to the ones used in previous sections. Consider a household $j \in \mathcal{J}_t$ in year t, where $\mathcal{J}_t$ denotes the set of all households filing taxes in year t. Let $I_{j,t}$ denote the household's taxable interest income in year t. If $I_{j,t} > 0$, I use $q^I_{j,t} \in [0,1]$ to denote the household's rank in the taxable interest income distribution in year t. Let $W^I_{j,t}$ denote the household's level of taxable interest-generating wealth, which is not observed and will be estimated, and let $\iota_{j,t} \equiv I_{j,t}/W^I_{j,t}$ denote the implied rate of return on these assets.

The rate of return and wealth share estimates used to discipline the capitalization estimates of individuals' interest-generating wealth are wealth-weighted average rates of return or shares of aggregate interest-generating wealth for households falling in different segments $\mathcal{Q} \subseteq [0,1]$ of the taxable interest income distribution. For households j in a given segment $\mathcal{Q}$, with taxable interest income rank $q^I_{j,t} \in \mathcal{Q}$, let $s^{WI}_t(\mathcal{Q})$ denote their share of total aggregate interest-generating wealth $\sum_{j\in\mathcal{J}_t} W^I_{j,t}$, let $s^I_t(\mathcal{Q})$ denote their share of aggregate taxable interest income $\sum_{j\in\mathcal{J}_t} I_{j,t}$, and let $\bar{\iota}_t(\mathcal{Q})$ denote the wealth-weighted average rate of return within the segment $\mathcal{Q}$. This within-group average return is related to the wealth-weighted average return across *all* households, denoted by $\bar{\iota}_t([0,1])$, by:

$$\bar{\iota}_t(\mathcal{Q}) = \frac{\sum_{j:q_{j,t}\in\mathcal{Q}} I_{j,t}}{\sum_{j:q_{j,t}\in\mathcal{Q}} W^I_{j,t}} = \frac{s^I_t(\mathcal{Q}) \cdot \sum_{j\in\mathcal{J}_t} I_{j,t}}{s^{WI}_t(\mathcal{Q}) \cdot \sum_{j\in\mathcal{J}_t} W^I_{j,t}} = \frac{s^I_t(\mathcal{Q})}{s^{WI}_t(\mathcal{Q})}\bar{\iota}_t([0,1]) \tag{A.28}$$

Similarly, let $\overline{W}_t(\mathcal{Q}) \equiv \frac{1}{|\{j:q_{j,t}\in\mathcal{Q}\}|}\sum_{j:q_{j,t}\in\mathcal{Q}} W^I_{j,t}$ denote average taxable interest-generating wealth for individuals in the segment, and $\overline{I}_t(\mathcal{Q}) \equiv \frac{1}{|\{j:q_{j,t}\in\mathcal{Q}\}|}\sum_{j:q_{j,t}\in\mathcal{Q}} I_{j,t}$ denote average taxable interest income. The wealth-weighted average return is equal to their ratio: $\bar{\iota}_t(\mathcal{Q}) = \overline{I}_t(\mathcal{Q})/\overline{W}_t(\mathcal{Q})$.

Observable average rates of return and wealth shares: I make use of the following estimated rates of return and aggregate wealth shares from sources other than the administrative income tax data, derived from datasets where aggregate or individual wealth is observed directly:

- The average wealth-weighted return across all households, denoted by $\widehat{\bar{\iota}}_t([0,1])$. I update the aggregate return series constructed by Piketty, Saez, and Zucman (2018) through tax year 2023, using data on aggregate asset holdings data from the Federal Reserve's U.S. Financial Accounts and aggregate interest income totals from the IRS.[42]
- The average return for households in the top 0.1% of the taxable interest income distribution, denoted by $\widehat{\bar{\iota}}_t([0.999,1])$. I use the SZZ CMD estimate.

[42]The most recently-updated capitalization factor series provided on the authors' webpage (`https://gabriel-zucman.eu/usdina/`) were made available in February 2022. The authors' series extend only through 2021, and because of tax filing delays after the end of the calendar year many of the input variable values for 2020 and 2021 are based on projections. I reconstruct the authors' capitalization factors using updated data for both the U.S. Financial Accounts series and IRS aggregate taxable interest income totals (`https://www.irs.gov/pub/irs-soi/22intaba.xls`). Due to tax filing delays, the final IRS aggregate income statistics for tax year 2023 were not available when these data were retrieved. To obtain an estimate of the final aggregate interest income totals for tax year 2023, I retrieved the aggregate interest income totals based on returns filed through November 2024 (`https://www.irs.gov/pub/irs-soi/24inweek47.xls`), and divided this value by the share of aggregate interest income for tax year 2022 reported on returns filed through the November 2023.

- The share of aggregate taxable interest-generating wealth owned by households that report no taxable interest income. This share is denoted by $\widehat{s}^{WI}_t([0, \underline{q}^I_t))$, where $\underline{q}^I_t$ denotes the share of households with zero taxable interest income $I_{j,t} = 0$ (so that $I_t(\underline{q}^I_t) = \min\{I_{j,t} \mid I_{j,t} > 0\}$).

Given these estimates, together with estimated shares of aggregate taxable interest income computed directly from the individual income tax data, the implied wealth share of all households in the bottom 99.9% of the taxable interest income distribution is given by:

$$\widehat{s}^{WI}_t([0.999, 1]) \equiv \overbrace{\widehat{s}^{I}_t([0.999, 1])}^{\text{from indiv. tax data}} \frac{\overbrace{\widehat{\imath}_t([0, 1])}^{\text{agg. return}}}{\underbrace{\widehat{\imath}_t([0.999, 1])}_{\text{SZZ CMD return}}}, \quad \widehat{s}^{WI}_t([0, 0.999)) \equiv 1 - \widehat{s}^{WI}_t([0.999, 1]) \tag{A.29}$$

The implied wealth share for households in the bottom 99.9%, excluding households that report no taxable interest income, is given by:

$$\widehat{s}^{WI}_t([\underline{q}^I_t, 0.999)) \equiv \widehat{s}^{WI}_t([0, 0.999)) - \underbrace{\widehat{s}^{WI}_t([0, \underline{q}^I_t))}_{\text{from SCF data}} \tag{A.30}$$

The implied rate of return for the same group of households is then given by:[43]

$$\widehat{\imath}_t([\underline{q}^I_t, 0.999)) = \frac{\overbrace{\widehat{s}^{I}_t([0, 0.999))}^{\text{from indiv. tax data}}}{\widehat{s}^{WI}_t([\underline{q}^I_t, 0.999))} \widehat{\imath}_t([0, 1]) \tag{A.31}$$

Additionally, I compute the implied average level of taxable interest-generating wealth for individuals reporting no taxable interest income:

$$\widehat{\overline{W}}_t([0, \underline{q}^I_t)) = \frac{\widehat{s}^{WI}_t([0, \underline{q}^I_t))}{\underline{q}^I_t} \cdot \widehat{\overline{W}}([0, 1]) = \frac{\widehat{s}^{WI}_t([0, \underline{q}^I_t))}{\underline{q}^I_t} \cdot \frac{\overbrace{\widehat{\overline{I}}_t([0, 1])}^{\text{from indiv. tax data}}}{\widehat{\imath}_t([0, 1])} \tag{A.32}$$

Capitalization approach: For households in the top 0.1% of the taxable interest income distribution, I estimate their level of interest-generating wealth by capitalizing their interest income using the SZZ CMD estimate of the group average return $\widehat{\imath}_t([0.999, 1])$. For households with no taxable interest income, I impute their level of interest-generating wealth as the estimated group average level $\widehat{\overline{W}}_t([0, \underline{q}^I_t))$ defined above. For households with positive taxable interest income below the top 0.1% cutoff, I estimate their wealth from their observed interest

[43] For convenience I write $\widehat{s}^I_t([0, 0.999))$ in the numerator of these expressions. Because households at ranks $q^I_{j,t} \leq \underline{q}^I_t$ earn no taxable interest income by definition, it follows that $\widehat{s}^I_t([\underline{q}^I_t, 0.999)) = \widehat{s}^I_t([0, 0.999))$.

income $I_t(q^I_{j,t})$ by scaling their interest income:

$$\widehat{W}^I_t(q^I_{j,t}) \equiv \begin{cases} I_t(q^I_{j,t}) \ / \ \widehat{\iota}_t([0.999,1]) & \text{if } q^I_{j,t} \geq 0.999 \\ \Big(I_t(q^I_{j,t})/I_t(0.999)\Big)^{1-\alpha^I_t} \cdot \widehat{W}^I_t(0.999) & \text{if } q^I_{j,t} \in [\underline{q}^I_t, 0.999) \\ \widehat{\overline{W}}_t([0,\underline{q}^I_t)) & \text{if } q^I_{j,t} < \underline{q}^I_t \end{cases} \tag{A.33}$$

Here $\alpha^I_t \in [0,1]$ is a parameter that determines the degree of heterogeneity in implied rates of return on interest-generating wealth across households. The implied rate of return $\iota_t(q^I_{j,t})$ is given by:

$$\iota_t(q^I_{j,t}) \equiv I_t(q^I_{j,t})/W^I_t(q^I_{j,t}) = \begin{cases} \widehat{\iota}_t([0.999,1]) & \text{if } q^I_{j,t} \geq 0.999 \\ \Big(I_t(q^I_{j,t})/I_t(0.999)\Big)^{\alpha^I_t} \cdot \widehat{\iota}_t([0.999,1]) & \text{if } q^I_{j,t} \in [\underline{q}^I_t, 0.999) \\ 0 & \text{if } q^I_{j,t} < \underline{q}^I_t \end{cases} \tag{A.34}$$

When $\alpha^I_t = 0$, this collapses to an equal-returns specification where the interest income of all households is capitalized using the top 0.1% return, with $\iota_t(q^I_{j,t}) = \widehat{\iota}_t([0.999,1])$. When $\alpha^I_t = 1$, all households in the bottom 99.9% with positive interest income are assigned the same wealth as the top 0.1% cutoff, with $\widehat{W}^I_t(q^I_{j,t}) = \widehat{W}^I_t(0.999)$, and all differences in interest income across households within the bottom 99.9% are attributed to large differences in returns $\iota_t(q^I_{j,t})$.

This capitalization method implies a wealth-weighted average rate of return for households in the bottom 99.9% of the taxable interest income distribution in year t, excluding households with zero income, which is equal to:

$$\overline{\iota}_t([\underline{q}^I_t, 0.999)) = \frac{\sum_{j:\underline{q}^I_t \leq q^I_{j,t} < 0.999} I_{j,t}}{\sum_{j:\underline{q}^I_t \leq q^I_{j,t} < 0.999} \widehat{W}^I_t(q^I_{j,t})} = \frac{\sum_{j:\underline{q}^I_t \leq q^I_{j,t} < 0.999} I_{j,t}}{\sum_{j:\underline{q}^I_t \leq q^I_{j,t} < 0.999} (I_{j,t}/I_t(0.999))^{1-\alpha^I_t} \cdot \widehat{W}^I_t(0.999)} \tag{A.35}$$

This implied return is decreasing in α^I_t. The value of α^I_t is chosen so that the average return implied by the capitalization method is exactly equal to the rate $\widehat{\iota}_t([\underline{q}^I_t, 0.999))$ defined in Equation (A.31).[44] The capitalization method then matches the average returns and shares of aggregate interest-generating wealth for all three groups of households (with zero taxable interest income, the remaining bottom 99.9%, and the top 0.1%, respectively) implied by the three average return and wealth share input estimates taken from other data sources.

[44] In practice, instead of computing the sums over households j that appear in the numerator and denominator of Equation (A.35), I use a trapezoid rule approximation to the integrals $\int_{\underline{q}^I_t}^{0.999} I_t(q^I)dq^I$ and $\int_{\underline{q}^I_t}^{0.999} W_t(q^I)dq^I$ over taxable interest income ranks q^I, using the observed quantiles of taxable interest income computed at 10,000 equally-spaced indices $q^I \in (0,1]$.

B Additional Income Tax Data Summary Statistics and Results

This section presents additional summary statistics and results using the administrative income tax data. All individual-level variables used in the regressions are winsorized in the full sample within each year, by replacing the bottom and top 1% of observations for each variable within each year with the 1st and 99th percentile values (respectively).

The full set of controls and fixed effects included in all first-stage and second-stage regressions consists of:

- Three-way fixed effects by year, 1-year age group, and wealth-to-income groups
- Four-way fixed effects by year, 5-year age group, nonfinancial income percentile within 1-year age group, and wealth-to-income groups
- Piecewise linear splines in two lags of changes in one-year nonfinancial income, scaled by average initial income. The spline breakpoints are set at -0.5, 0, and +0.5.
- Piecewise linear splines in two lags of imputed net savings for each asset class (dividends, taxable interest and nontaxable interest), scaled by average initial income. The spline breakpoints are set at -0.5, 0, and +0.5.
- Interactions between two lags of capitalized financial wealth by asset class, scaled by average initial income, interacted with two-way fixed effects by 5-year age group and total wealth/income group

Table B.1: Individual Observation Counts in Full Sample by Year

Reference Year	Observations
2000	1,448,000
2001	1,382,000
2002	1,452,000
2003	1,428,000
2004	1,446,000
2005	1,448,000
2006	1,448,000
2007	1,441,000
2008	1,437,000
2009	1,437,000
2010	1,441,000
2011	1,432,000
2012	1,424,000
2013	1,408,000
2014	1,396,000
2015	1,386,000
2016	1,368,000
2017	1,363,000
2018	1,346,000
2019	1,330,000
2020	1,319,000
2021	1,392,000
2022	1,466,000

Notes: This table presents individual observation counts for the full income tax data sample in each reference year. The main sample is constructed from an initial random 10% subsample of all individuals who filed a tax return over the sample period. The sample in each reference year t is then further restricted to: (1) working-age individuals, (2) satisfying data availability restrictions over years $t-2$ to at most $t+3$, (3) who rank in the top 20% of the nonfinancial income distribution among individuals with the same age. Full details of the sample construction are described in Section 3.1. Observation counts are rounded in accordance with U.S. Census Bureau disclosure guidelines.

Table B.2: First-Stage Estimates: Individual Exposures to Local Income Shocks

	Income rank [0.8,0.9)			Income rank [0.9,0.95)		
	$h=1$	$h=2$	$h=3$	$h=1$	$h=2$	$h=3$
ZIP-code shock	0.231	0.273	0.288	0.222	0.259	0.264
	(0.009)	(0.009)	(0.010)	(0.012)	(0.012)	(0.013)
Observations (mil.)	14.08m	14.08m	14.08m	7.096m	7.096m	7.096m
R^2	0.124	0.124	0.123	0.154	0.160	0.161
Fixed Effects/Controls	(see caption)					
Clustering	Individual; ZIP-code					

	Income rank [0.95,0.99)			Income rank [0.99,1)		
	$h=1$	$h=2$	$h=3$	$h=1$	$h=2$	$h=3$
ZIP-code shock	0.262	0.278	0.278	0.719	0.670	0.625
	(0.014)	(0.015)	(0.017)	(0.030)	(0.030)	(0.030)
Observations (mil.)	5.68m	5.68m	5.68m	1.409m	1.409m	1.409m
R^2	0.182	0.191	0.192	0.200	0.202	0.196
Fixed Effects/Controls	(see caption)					
Clustering	Individual; ZIP-code					

Notes: This table reports first-stage estimates for each income group of the passthrough from ZIP-code level shocks $\widetilde{g}_{z(j),t+1}$ to individual income growth, as depicted in Figure 8. The first-stage regression is described in Equation (4). Each column reports the estimated coefficient $\widehat{\pi}_h$ for each horizon of h years over which post-period average income $\overline{Y}_{j,t+1,t+h}$ is measured, estimated separately for each of the four income groups. The full set of controls and fixed effects is listed at the start of Appendix B. Standard errors are clustered by individual j and ZIP-code z and reported in parentheses. Observation counts are rounded in accordance with U.S. Census Bureau disclosure guidelines.

Table B.3: Second-Stage Estimates: Savings Responses to Persistent Income Shocks

	Income rank [0.8,0.9)			Income rank [0.9,0.95)		
	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
Income growth (inst.)	0.516	1.062	1.267	0.607	1.514	1.825
	(0.039)	(0.059)	(0.070)	(0.062)	(0.104)	(0.121)
Observations (mil.)	14.08m	14.08m	14.08m	7.096m	7.096m	7.096m
First-stage F	833.3	833.3	833.3	444.2	444.2	444.2
Fixed Effects/Controls	(see caption)					
Clustering	Individual; ZIP-code					

	Income rank [0.95,0.99)			Income rank [0.99,1)		
	$h = 1$	$h = 2$	$h = 3$	$h = 1$	$h = 2$	$h = 3$
Income growth (inst.)	0.652	1.701	2.084	0.300	0.998	1.308
	(0.076)	(0.123)	(0.148)	(0.063)	(0.091)	(0.117)
Observations (mil.)	5.68m	5.68m	5.68m	1.409m	1.409m	1.409m
First-stage F	277.1	277.1	277.1	424	424	424
Fixed Effects/Controls	(see caption)					
Clustering	Individual; ZIP-code					

Notes: This figure reports second-stage estimates for each income group of the savings response to persistent income shocks, as depicted in Figure 9. The second-stage regression is described in Equation (3). ZIP-code level shocks $\widetilde{g}_{z(j),t+1}$ are used as an instrument for realized nonfinancial income growth in the first-stage regression described in Equation (4), at a horizon of $h = 3$ years. The dependent variable is cumulative imputed net savings in all financial assets (both stocks and fixed income) $S^{tot}_{j,t,t+h}$ through year $t + h$, scaled by average nonfinancial income $\overline{Y}_{j,t-2:t}$ over tax years $t - 2$ to t. To convert estimates to a marginal propensity to save out of post-tax income, marginal tax rates τ_{marg} are estimated for each income group using data from the SCF and NBER's TAXSIM calculator, and estimated coefficients $\widehat{\beta}_h$ for pre-tax income are divided by $(1 - \tau_{marg})$. Each column reports the marginal tax rate-adjusted coefficient for each horizon of h years over which the savings response is measured, estimated separately for each of the four income groups. The full set of controls and fixed effects is listed at the start of Appendix B. Standard errors are clustered by individual j and ZIP-code z and reported in parentheses. Observation counts are rounded in accordance with U.S. Census Bureau disclosure guidelines.