NBER WORKING PAPER SERIES

AI FINANCIAL ADVICE: SUPPLY, DEMAND, AND LIFE CYCLE IMPLICATIONS

Taha Choukhmane
Tim de Silva
Weidong Lin
Matthew Akuzawa

Working Paper 35574
http://www.nber.org/papers/w35574

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2026

First draft: November 2025. We thank John Beshears, James Choi, Anastassia Fedyk, Alberto Rossi, Jon Reuter, Darrell Duffie, Fiona Greig, Jonathan Parker, Chris Tonetti, Tony Cookson, Aizhan Anarkulova, Steffen Andersen, Antonio Coppola, Eric So, and conference and seminar participants at MIT Sloan, Vanguard, Stanford IFDM, Erasmus, CEPR Household Finance, SFS Cavalcade, FIRS, WFA, and the NBER Household Finance and Economics of Aging meetings, for helpful comments and feedback. An earlier version of this paper was circulated with the title "How Good is Generative AI Personal Financial Advice?." The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Taha Choukhmane, Tim de Silva, Weidong Lin, and Matthew Akuzawa. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

AI Financial Advice: Supply, Demand, and Life Cycle Implications
Taha Choukhmane, Tim de Silva, Weidong Lin, and Matthew Akuzawa
NBER Working Paper No. 35574
August 2026
JEL No. D15, G11, G5, G51, G53

ABSTRACT

We ask a representative sample to write prompts seeking spending and investing advice from LLMs, then simulate the lifetime effects of following the advice under realistic asset and labor market conditions. Applying this method to GPT-5.2, we find following the advice would move respondents toward life cycle theory: broader participation in diversified equity funds, age-declining equity shares, and larger savings buffers. Recommendations vary systematically by gender, prior AI experience, and financial literacy. For gender, two-thirds of recommended equity-share differences arise from men and women writing different prompts (demand), while one-third arise from gender labels attached to otherwise identical prompts (supply).

Taha Choukhmane
Massachusetts Institute of Technology
MIT Sloan School of Management
and NBER
tahac@mit.edu

Tim de Silva
Stanford University
Graduate School of Business
and Stanford Institute
for Economic Policy Research
(SIEPR)
tdesilva@stanford.edu

Weidong Lin
Massachusetts Institute of Technology
wel@mit.edu

Matthew Akuzawa
Massachusetts Institute of Technology
maku49@mit.edu

Millions of people now turn to large language models (LLMs) for personal financial advice. Within a few years of these tools becoming widely available, around half of adults in the U.S. and the U.K. report using them for financial guidance (Lloyds Banking Group 2025; J.D. Power 2025), a share that might already exceed the proportion who consult a human financial advisor (Gallup 2025). This rapid adoption raises the prospect that affordable, widely accessible, high-quality financial advice may be closer than ever. However, whether that promise is realized depends on two distinct forces: the content of the advice these models give, and how people change their behavior in response to it. A natural first step, and the focus of this paper, is to characterize the advice itself: what LLMs recommend, and how those recommendations vary across individuals.

We develop and implement a method to describe quantitatively the personal financial advice that LLMs give. Studying this advice systematically is challenging because it depends on what individuals choose to ask about, on features of the models that supply the advice, and on how uncertain financial outcomes accumulate over the life cycle. We address these three challenges by (i) surveying a demographically balanced sample to elicit realistic prompts seeking financial advice from an LLM, (ii) simulating the quantitative implications of the advice received in response to these prompts over the life cycle under realistic asset and labor market conditions, and (iii) evaluating why the advice differs across heterogeneous groups by randomizing features of prompts, which allows us to separate supply-side differences in how models respond to otherwise identical prompts from demand-side differences in what individuals ask.

We apply our method to the spending, saving, and investing advice provided by `GPT-5.2`. We focus on these domains both because they involve some of the most consequential financial decisions households make and because our survey shows they are the topics people most commonly seek AI financial guidance on. We have three main results. First, following LLM advice would move most survey respondents closer to the broad recommendations of standard life cycle theory relative to their observed behavior, including greater participation in diversified equity funds, equity shares that decline with age after 45, and sizable savings buffers. Second, LLM advice departs from life cycle theory in systematic ways: it implies unusually high patience, relies on simple heuristics, smooths consumption imperfectly after income shocks, and allows portfolios to drift passively with realized returns. Third, LLM advice varies systematically with individual characteristics, such as gender, financial literacy, and prior AI experience. These differences accumulate over the life cycle into wealth differences at age 60 of approximately 4–6% between groups. As in the market for human financial advice, such differences can reflect both demand and supply (Reuter and Schoar

2024). For example, we find that men receive higher recommended equity shares than women, and by randomizing gender labels, we show that two-thirds of this difference in investment advice comes from differences in what men and women write in their prompts, while one-third comes from the model supplying different advice to otherwise identical prompts when labeled as coming from men rather than women.

Taken together, our results highlight the potential of generative AI to improve financial decision-making, but suggest that its impact is likely to vary across individuals. The finding that LLM advice broadly aligns with life cycle theory should not be taken for granted. Unlike other forms of automated financial advice, such as robo-advisors, these models are not optimized to improve household financial decisions (D'Acunto et al. 2019; Rossi and Utkus 2020). One might therefore worry that, in pursuit of engagement, LLMs would instead reinforce existing behaviors, validate biases, or simply tell people what they want to hear (Fedyk et al. 2026). Our results suggest a more nuanced conclusion: current general-purpose LLMs often provide advice that aligns with standard financial guidance, but this advice is sensitive to demand and, therefore, differs systematically across users.

While we primarily use `GPT-5.2`, we show that our main results are quantitatively similar using `Gemini 3 Flash`—a similar generation model from a different provider—and `GPT-5.6 Terra`—a more recent model from the same provider. Nevertheless, the advice provided by these models may differ from that of future models. Our contribution is therefore not only to characterize the advice of a particular generation of models, but also to provide a framework for studying LLM financial advice. As these models continue to evolve rapidly, three features of our analysis are likely to remain relevant. First, differences in advice arise not only from model capabilities, but also from demand: what users ask, what information they provide, and how they frame their financial problems. These demand-side constraints are likely to remain important even as models continue to improve rapidly. Second, our survey-and-simulation approach can be reapplied to track how both model and user behavior change over time. Third, progress in applying generative AI to household finance will depend in part on the availability of benchmarks and diagnostic tests against which models can be evaluated and improved. While many such benchmarks exist for other tasks, we provide a set of diagnostics for personal financial advice grounded in normative life cycle theory (e.g., consumption smoothing, diversification, portfolio rebalancing).

Methodology. Our method for studying LLM financial advice proceeds in three steps. First, we survey a demographically balanced sample of 1,000 U.S. adults recruited through Prolific. We ask respondents to write three prompts to an LLM. The first prompt is a description of their financial situation, the second is for advice on spending versus saving,

and the last is for investing advice. This is the main innovation of our approach: using prompts written by regular users introduces meaningful heterogeneity in both the topics people raise and how they express themselves. To our knowledge, our survey provides the first systematic evidence on how prompts vary across individuals seeking financial advice from LLMs. Alongside the prompts, we collect data on respondents' demographics, financial situation, financial literacy, and prior experience seeking financial guidance from LLMs.

In the second step, we build a quantitative life cycle model in the tradition of Gourinchas and Parker (2002) and Cocco et al. (2005) (and most closely related to Choukhmane and de Silva 2026). The model is calibrated to match historical U.S. data and incorporates stochastic income, stochastic asset returns, job-to-job and unemployment transitions, mortality risk, and a tax and social insurance system. We use the model to simulate households' consumption-saving decisions and their allocation of savings across four asset classes: fixed income, diversified equity funds, individual stock holdings, and other risky assets, including crypto, gold, commodities, and collectibles.

The final step in our method combines the prompts written by our survey respondents with our life cycle model. Rather than simulating choices based on policy functions derived from dynamic optimization, we simulate choices based on the LLM recommendations. Specifically, at each age, we randomly match each agent in our simulation with a prompt written by a survey respondent with a similar age, income, and employment status. We then replace the state variables in that prompt with those of the simulated individual and submit two successive LLM queries: first, we submit the prompt and receive textual financial advice from an LLM; then, we use a second LLM to translate this textual advice into quantitative recommendations for saving, consumption, and asset allocations. We repeat this process at each age over the individual's life cycle and across many simulated individuals. Importantly, each query to the LLM is independent, with no memory of past responses; the only link across periods is the evolution of state variables that depend on prior choices.

Our method can be reapplied to other populations, advice topics, or models. A key design choice is the trade-off between specificity and external validity. Asking people to write prompts about defined topics, in our case spending, saving, and investing, allows us to compare responses across individuals and integrate them into a parsimonious life cycle framework. We view this as a feature of the approach that can be adjusted depending on the research question. Greater specificity would allow studying decisions that we abstract from (e.g., insurance, housing, credit management), while more open-ended prompts would better reflect real-world usage at the cost of being harder to compare across individuals and incorporate into a simulation framework.

Main findings. Our main results come from applying this method to `GPT-5.2`, which we refer to as "the LLM" throughout. We chose this LLM because ChatGPT is the most commonly used model for financial advice in our survey (and others, e.g., Lloyds Banking Group 2025; J.D. Power 2025) and `GPT-5.2` was the latest model at the time we ran our analysis; we find quantitatively similar results using `Gemini 3 Flash` and `GPT-5.6 Terra`.

We document three facts about LLM advice over the life cycle. First, following LLM recommendations would move most individuals closer to the broad principles of life cycle theory relative to their observed behavior. On saving, many individuals across all age groups report having accumulated little financial wealth; by contrast, following the advice would lead virtually all simulated individuals to build savings buffers above $10,000 by age 30. On portfolio choice, it would induce near-universal participation in the stock market, with equity shares that decline with age after 45, and risky holdings concentrated in diversified equity funds rather than individual stocks, crypto, gold, commodities, or collectibles. In addition to these broad patterns, the LLM's advice responds to the topics individuals raise in ways that are generally consistent with standard theory. For example, prompts that mention macroeconomic uncertainty lead the LLM to increase saving and reduce equity holdings, consistent with a stronger precautionary motive.

Second, we find that the LLM's advice departs from the more subtle prescriptions of life cycle theory in systematic ways. The advice implies an unusually high discount factor in a standard life cycle model. Instead of tailoring recommendations closely to individuals' evolving economic circumstances, the advice often relies on simple saving and withdrawal heuristics, including bunching at round numbers and recommending fixed withdrawal rules in retirement. Consistent with this reliance on heuristics, consumption is insufficiently smoothed: the LLM recommends too little decumulation in retirement and sharp consumption declines after job loss, even when simulated individuals have sufficient liquid wealth. Asset allocation advice exhibits a similar lack of active adjustment as portfolios drift passively with realized returns rather than being actively rebalanced. These departures appear to reflect, in part, the prompts that people write. A more structured "academic" prompt, which provides details on each agent's financial situation and explicitly references life cycle planning and acting in the user's best interest, improves consumption smoothing and reduces reliance on heuristics, though it does not increase active portfolio rebalancing.

Third, the LLM's advice varies systematically with individual characteristics, such as gender, financial literacy, and prior AI experience. For each of these characteristics, we split survey respondents into two groups and repeat our simulation, sampling only prompts written by individuals in each group (e.g., simulating life cycle paths using only prompts

written by men versus women). The advice varies meaningfully along each dimension: average wealth at age 60 is around 5% higher when sampling prompts written by men, individuals with high financial literacy scores, or those who report prior use of AI for financial guidance. These differences arise because the LLM consistently recommends lower equity shares for prompts written by women or individuals with lower financial literacy, and lower saving rates for prompts written by individuals without prior AI use.

To understand why LLM advice differs across individuals, we develop an approach to separate the roles of demand and supply. Focusing on gender, we use the set of survey prompts that do not explicitly identify the respondent's gender, and randomly add gender labels, such as "I am a man." This separates demand-side differences (i.e., men and women write different prompts) from supply-side differences (i.e., the LLM gives different advice to the same prompt when the stated gender changes). We find that approximately two-thirds of the gender gap in recommended diversified equity shares is demand-driven: it reflects differences in the content of men's and women's prompts that remain even when both are assigned the same gender label. The remaining one-third is supply-driven: the same prompts receive higher equity recommendations when labeled as coming from a man rather than a woman. In contrast, gender differences in recommended holdings of non-diversified assets, such as individual stocks or crypto, are entirely demand-driven. These findings echo evidence from human financial advisors, where gender differences in advice also reflect both demand and supply (Bhattacharya et al. 2024; Bucher-Koenen et al. 2025). Applying the same approach to race, we find smaller differences and no supply-side effect: explicit race labels leave recommendations essentially unchanged.

Related literature. This paper contributes to two strands of literature. First, we contribute to the literature on financial advice and financial innovation by introducing a framework to quantitatively study LLM-generated financial advice. This literature is motivated by the evidence that individuals do not, on their own, always make optimal financial decisions.[1] While this creates a role for professional advice, substantial evidence shows that existing financial advice sources are limited by both demand- and supply-side constraints (see Reuter and Schoar 2024 for a recent review). For example, professional human advisors are often costly, and their advice can be low quality (Linnainmaa et al. 2021; Andries et al. 2025), subject to conflicts of interest (Mullainathan et al. 2012), and affected by misconduct (Egan et al. 2019). Popular personal finance books are cheaper and more

[1]This includes evidence that individuals do not take full advantage of available 401(k) employer matches (Choi et al. 2011; Choukhmane et al. 2025), do not efficiently allocate assets across accounts (Bergstresser and Poterba 2004), or fail to refinance fixed-rate mortgages when beneficial (Andersen et al. 2020). For reviews, see Beshears et al. (2018), Gomes et al. (2021), and Campbell and Ramadorai (2026).

accessible, but their guidance often contains fallacies and departs from the prescriptions of normative economic models (Choi 2022). Automated advice, such as robo-advisors, can improve financial decision-making (D'Acunto et al. 2019; Rossi and Utkus 2020; Reher and Sokolinski 2024), but, relative to LLMs, its uptake has been much slower. One possible reason is that robo-advisors may be more limited in their ability to provide personalized conversational advice, as well as the soft skills and emotional support that clients value in human advisors (Greig et al. 2025). A natural question is therefore whether LLMs can offer advice that is at once affordable, widely accessible, personalized, and high-quality, a combination that no existing source fully provides.

Our contribution is to provide evidence on whether LLM-generated advice can fulfill this promise, taking into account that the advice depends on the questions people ask, the characteristics of the models, and the accumulated impact over the life cycle. In doing so, we extend a longstanding tradition of using quantitative life cycle models to evaluate financial products and policies, with recent applications to adjustable-rate mortgages (Guren et al. 2021; Campbell et al. 2021), target-date funds (Duarte et al. 2025), saving nudges (Choukhmane 2025; Choukhmane et al. 2025), and income-contingent student loans (de Silva 2025). We apply this approach to what is perhaps the fastest-growing area of innovation in personal finance: generative AI.

Second, we contribute to the growing literature on how LLMs are changing economics and finance (see Mo and Ouyang 2025 for a review), and in particular to the emerging literature that studies LLMs as models of human behavior. This literature shows that LLMs often replicate human behavior and biases elicited in surveys and classic economic experiments (Horton et al. 2023; Ross et al. 2024; Bini et al. 2026).[2] A related strand studies LLM behavior in financial settings. Here too, LLMs both resemble and depart from human behavior: Ouyang et al. (2025) find that LLM risk-taking is stable across a variety of behavioral tasks; Fedyk et al. (2026) show that LLM preferences over asset allocations resemble those of younger, high-income individuals, but are less likely to violate transitivity; and Rumpf et al. (2026) compare recommendations from individuals, professional advisors, and LLMs, and find that LLMs are far more sensitive to risk preferences but also exhibit substantial stochasticity. A related set of recent papers studies LLMs directly as providers of financial advice using researcher-designed personas or scenarios, examining whether LLMs can infer investor preferences and tailor recommendations while also documenting limitations such as stochasticity, biases, and sensitivity to demographic attributes (e.g.,

[2]Also related is the strand of this literature that studies agent-based economic models populated by LLM agents, including Hao and Xie (2025), who study two-period consumption-saving problems, and Verstyuk and Douglas (2025), who study a Krusell and Smith (1998) economy.

Fieberg et al. 2025; Takayanagi et al. 2025; Foltyn and Olsson 2026).

Relative to this literature, the methodology that we develop makes two contributions. First, instead of using LLMs to learn about, simulate, or replace human survey responses, we use surveys of humans to learn about LLMs: prompts written by a survey sample allow us to characterize both the supply *and* demand sides of human–LLM interactions. In this sense, our approach uses humans to learn about AI. Second, and more specific to our financial advice application, we embed LLM advice within a dynamic, stochastic life cycle model. This allows us to characterize the cumulative effects of LLM advice over an entire lifetime and compare the resulting behavior with both observed behavior and a quantitative benchmark grounded in economic theory. This life cycle perspective distinguishes our analysis from studies that evaluate LLM advice primarily in static survey or experimental settings.

1 Methodology for Simulating LLM Advice

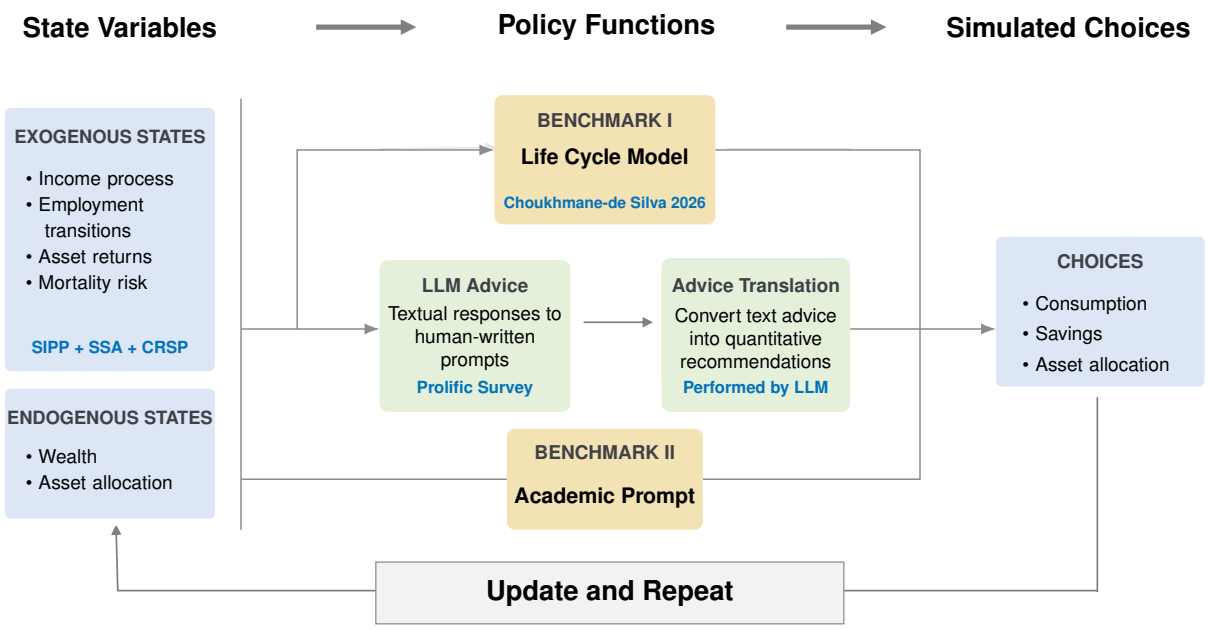
We begin by describing our methodology, which consists of three steps that are summarized in Figure 1. First, we conduct a survey and ask a demographically balanced sample to write prompts seeking saving and investing advice from an LLM (Section 1.1). Second, we calibrate a quantitative life cycle model to U.S. income dynamics, employment transitions, asset returns, and tax rules. This model provides both the economic environment for our simulations and a normative benchmark grounded in economic theory (Section 1.2). Finally, we combine the survey with the model to simulate the life cycle paths of individuals who would follow an LLM's advice each year (Section 1.3). After describing the methodology and benchmarks, which apply in general to any LLM, we describe our model selection.

1.1 Survey

The first step in our methodology is to conduct a survey that elicits prompts in which individuals seek financial guidance from an LLM. We design the survey in Qualtrics and administer it to a demographically balanced U.S. sample recruited through the Prolific online platform. Our initial sample consists of 1,000 complete responses, of which 952 pass both Prolific's authenticity check to screen out AI-generated responses and a minimum specificity requirement. Appendix A.2 provides more details on this sample selection.

The survey asks each respondent to write three free-text prompts to an LLM financial advisor (see Figure E1 for the exact wording). The first prompt asks respondents to describe their financial situation, including whatever details they think would help an LLM provide useful advice. Respondents are told to assume the LLM has no prior information about them. The second prompt asks respondents to seek advice on how much to spend over the

Figure 1. Overview of Methodology for Simulating LLM Advice



Notes: This figure summarizes the methodology for simulating LLM advice. State variables (left) feed into policy functions (center), which produce simulated choices (right). The life cycle model (Benchmark I) provides optimal policy functions derived from dynamic optimization. LLM advice uses survey prompts to generate textual recommendations, which are then translated into quantitative choices. Academic prompts (Benchmark II) replace survey prompts with researcher-designed prompts. Endogenous state variables are updated each period and the process repeats over the life cycle. Section 3.1 also defines a one-shot advice benchmark that is not shown in the schematic.

coming year, while the third asks for advice on how to invest their savings between stocks and safer options. For the second and third prompts, respondents are told that the LLM already has the information from their first response. We focus on spending and investing advice because saving, budgeting, and investing are the domains in which people most commonly seek AI financial guidance, both in our survey and others (J.D. Power 2025).[3]

In addition to the prompts, we survey respondents about the state variables that will be used in our life cycle simulation (including age, income, employment status, and financial wealth), as well as demographic characteristics such as gender, race, financial literacy, and prior experience with AI. Table E1 compares the demographics of our Prolific sample with those of the Current Population Survey (CPS). The sample broadly matches the CPS across age, gender, and income, though it over-represents the unemployed, under-represents the lowest-income earners, and under-represents respondents older than 70.

[3]Among the 48% of respondents to our survey who report having used an AI tool for financial advice or information in the past three months, the most common topics are saving (49%) and investing (49%), followed by budgeting (42%) and financial education (35%) (Figure E2).

Summary statistics for the prompts written by our survey respondents—including word counts, character counts, and the prevalence of numerical and dollar-amount mentions—are reported in Table E2. The financial situation prompt is the longest of the three (mean of 43 words), with 65% of respondents mentioning specific numbers and 34% mentioning dollar amounts. The spending and investment advice prompts are shorter (mean of 27 words each), with lower rates of quantitative detail.

1.2 Life Cycle Model

In the second step, we build a quantitative life cycle model of consumption, saving, and portfolio choices based on Choukhmane and de Silva (2026). Additional model details and the calibration are presented in Appendix B.

Each model period corresponds to one year. Individuals enter working life at age 22, retire deterministically at age 65, and face age-dependent mortality risk based on the Social Security Actuarial Life Tables until a maximum age of 90. They have time-separable CRRA utility over consumption with annual discount factor β and relative risk aversion γ.

Individuals face employment risk and move across four labor-market states: employed, transitioning from job-to-job, unemployed, and retired. While employed, they earn stochastic labor income that combines a deterministic age profile, a persistent AR(1) component, and a transitory shock. Job-to-job transitions are associated, in expectation, with persistent wage gains, whereas unemployment spells lead to persistent wage losses upon reemployment. The parameters of the income process and the employment transition probabilities are estimated using Survey of Income and Program Participation (SIPP) data. The government provides unemployment benefits to the unemployed and Social Security retirement benefits tied to average lifetime earnings to the retired. It also taxes income according to the 2025 U.S. federal schedule and taxes capital returns at a flat rate.

Investors can allocate their savings across four assets: a risk-free bond, a diversified stock market index, an individual stock, and another risky asset representing holdings such as crypto, gold, commodities, and collectibles. The returns of the diversified index are calibrated using historical CRSP total market returns. We refer to the individual stock and the other risky asset as non-diversified assets, since they capture risky positions outside the diversified market index that an LLM can recommend. Both non-diversified assets have the same return process, but independent shock realizations, so they represent distinct risky positions. We calibrate their return process using evidence on individual stock returns from Bessembinder (2018). We construct these returns such that neither can improve the Sharpe ratio of a portfolio that combines the bond and the diversified stock market index.

In each period, investors choose consumption, saving, and portfolio shares across the four assets subject to no-borrowing and no-leverage constraints. We define the following portfolio shares, which we refer to throughout.

$$\text{Equity Share} = \frac{\text{Diversified Stock Holdings} + \text{Individual Stock Holdings}}{\text{Total Wealth}}$$

$$\text{Diversified Equity Share} = \frac{\text{Diversified Stock Holdings}}{\text{Total Wealth}}$$

$$\text{Non-Diversified Asset Share} = \frac{\text{Individual Stock Holdings} + \text{Other Risky Asset Holdings}}{\text{Total Wealth}}$$

We keep track of seven state variables when solving the life cycle model: age, labor productivity, transitory income shock, employment status, tenure, average lifetime income, and savings.

1.3 Simulating Advice over the Life Cycle

The final step in our methodology combines the survey responses with the life cycle model to simulate the life cycle paths of individuals who follow the advice provided by an LLM, as illustrated in Figure 1. Each simulated life path proceeds by iterating, every year from age 22 to 89, through five steps: states, prompt, advice, choices, and shocks. Table 1 illustrates these steps with a concrete example.

Step 1: States. At the beginning of each period, each simulated individual is characterized by the state variables of our life cycle model (Table 1, Panel A). In our example, the individual is 26 years old, employed, earning $68,810, and holding $24,380 in financial assets.

Step 2: Prompt selection and preparation. We assign each simulated individual to one of 12 buckets defined by employment status, age, and income. We then randomly draw a prompt written by a survey respondent in the same bucket, so that the LLM advice assigned to the simulated individual is based on a prompt from someone with similar observable characteristics (Figure A1).[4] Within each prompt bucket, the characteristics of the simulated individual may still differ from those of the sampled survey respondent. We therefore identify any references to state variables in the prompt and replace them with the corresponding values from the simulation. For example, Table 1, Panel B shows the

[4] In choosing which characteristics to match on, we face a trade-off between matching simulated individuals more closely to survey respondents and maintaining sufficient sample size within each bucket.

raw prompt as originally written by a 25-year-old survey respondent earning approximately $75,000 per year. Panel C shows the same prompt after variable insertion: every mention of a state variable—including age, income, wealth levels, and portfolio allocations—is replaced with the simulated individual's current values. Our insertion rules, described in Appendix A, preserve the author's writing style and concerns while ensuring that the quantitative details match the simulated individual's situation. The three survey prompts (financial situation, spending advice, and investment advice) are then concatenated into a single message and sent to an LLM with an instruction to respond in 200 words or fewer.

Step 3: LLM advice. The LLM responds to the prepared prompt (Table 1, Panel D) with textual advice. The qualitative properties of this advice, including which topics the LLM emphasizes and how they compare with what individuals discuss in their prompts, are described in Section 2.1.

Step 4: Translating advice into quantitative choices. Next, the LLM's textual advice is converted into the quantitative recommendations for the choice variables of our model: consumption and savings held in the four asset classes. We do this by passing the LLM's advice back to another LLM, along with deterministic instructions to extract numerical recommendations and output them in a structured plain-text format. This second LLM call maps qualitative advice (e.g., "Keep an emergency fund of 3-6 months of essential expenses in a HYSA/money market" and "Automate investing each payday (e.g., 15–25% of gross)") into dollar amounts for consumption and contributions to each asset class (Table 1, Panel E). The extraction follows a pre-defined ordering; the full translation prompt, including the asset class definitions, extraction ordering, budget constraints, and deterministic rules that ensure all outputs satisfy accounting identities, is in Appendix A.5.2.

Step 5: Shocks and state transitions. After extracting the quantitative recommendations, we draw realizations of the exogenous shocks in our life cycle model: labor market transitions, persistent and transitory income shocks, and asset returns for each of the four asset classes. The realized shocks, combined with the LLM's choices, determine the state variables at the beginning of the next period (Table 1, Panel F). These updated states are then used as the input for the next iteration.

Simulation design. Each query to the LLM is independent, with no memory of past responses; the only link across periods is the evolution of state variables that depend on prior choices. In each run of the model, we simulate the same 1,000 individuals, each starting life at age 22 with no wealth, retiring at age 65, and dying by age 90. A new prompt is drawn and prepared in every period for every individual, and the distribution

Table 1. From States to Choices: Simulation Example for One Period

Panel A: State variables

Age	Employment status	Pre-tax annual income	Post-tax annual income	Bonds	Diversified stocks	Individual stocks	Other risky assets
26	Employed	$68,810	$62,223	$17,938	$6,442	$0	$0

Panel B: Raw survey prompt (as written by respondent)

Financial situation	"I make around 75000 a year before taxes and pay 1700 a month in rent. That is my primary income and I make some money completing paid research surveys. I have around 80000 saved. I am 25."
Spending advice	"I am looking for ways to go out more and enjoy my life without sacrificing my savings. I have a good grip on money and actually struggle to justify spending."
Investment advice	"How should I put away money now that interest rates are going down. Safe bets like my money market and CDs are no longer yield much money."

Panel C: Prepared prompt (after variable insertion)

Financial situation	"I make around **68810** a year before taxes and pay 1700 a month in rent. That is my primary income and I make some money completing paid research surveys. I have around **24380** saved. I am **26**."
Spending advice	"I am looking for ways to go out more and enjoy my life without sacrificing my savings. I have a good grip on money and actually struggle to justify spending."
Investment advice	"How should I put away money now that interest rates are going down. Safe bets like my money market and CDs are no longer yield much money."

Panel D: LLM advice

"Given your income/savings, you can safely 'permission' yourself to spend more without derailing goals by separating **spending from saving**.

1) **Set a floor, then automate:** Keep an emergency fund of 3-6 months of essential expenses in a HYSA/money market (don't chase yield; it's for safety). Automate investing each payday (e.g.,15-25% of gross, adjust to comfort). What's left is guilt-free.
2) **Create a "fun" system:** Open a separate checking/savings called "FUN". Auto-transfer a fixed amount (start $300-$600/month). If it's there, you must use it or roll it to a bigger experience.
3) **Where to put money as rates fall:** For long-term (5+ years): low-cost broad index funds/ETFs (e.g., total US + total international) in 401(k)/Roth IRA first. For medium-term goals (1-5 years): HYSA + Treasuries/T-bill ladder (often tax-advantaged vs CDs), or a short-term bond fund if you can tolerate modest fluctuation.
4) **Psychology:** Treat spending from the FUN account as a goal, not a failure-your system already protected savings.

"

Panel E: Extracted choices

Consumption	Total saving	Bonds contribution	Diversified stocks contribution	Individual stocks contribution	Other risky assets contribution
$44,378	$17,845	+$5,400	+$12,445	$0	$0

Panel F: Updated states (beginning of next period)

Age	Employment status	Pre-tax annual income	Post-tax annual income	Bonds	Diversified stocks	Individual stocks	Other risky assets
27	Employed (no transition)	$73,016 (income shocks)	$65,504	$23,805	$19,505	$0	$0

Notes: This table illustrates the five steps used to simulate LLM advice for a single individual in a single period. Panel A shows the state variables at the start of the period. Panel B shows the raw survey prompt as originally written by a respondent. Respondent text is reproduced verbatim; spelling and grammar errors are preserved. Panel C shows the same prompt after variable insertion, with boldface values indicating state variables substituted into the respondent's text. Panel D shows the full LLM advice. Panel E shows the quantitative choices extracted from the translation LLM call that are required to satisfy the budget constraint: consumption + total net contributions = post-tax annual income. Panel F shows the state variables at the start of the next period after labor market, income, and asset return shocks are realized.

from which prompts are drawn evolves as an individual's state variables change.

1.4 Benchmarks to Compare with LLM Advice

We compare our simulated LLM advice to three benchmarks. The first is respondents' status-quo self-reported behavior. The second is optimal behavior in the life cycle model, which provides a normative benchmark. The third is LLM advice generated from a structured researcher-designed prompt, which tests whether providing more complete and standardized information moves advice closer to the model benchmark. The latter two benchmarks are illustrated in Figure 1.

Empirical benchmark: observed behavior. The observed-behavior benchmark uses respondents' self-reported savings and portfolio choices from our survey. The behavior of our survey respondents is broadly consistent with evidence on saving and investing over the life cycle prior to the widespread adoption of AI financial advice tools, such as the Survey of Consumer Finances (Gomes et al. 2021).

Advice benchmark I: life cycle model. The second benchmark is constructed by simulating the life cycle model described above and solving for optimal behavior under the same realizations of exogenous shocks used in the LLM simulation. Because the aggregate stock index is calibrated to be mean-variance efficient, we assume that individuals in this benchmark hold neither non-diversified asset class. Portfolio choice, therefore, reduces to an allocation decision between the risk-free bond and the diversified stock index.

Advice benchmark II: academic prompts. The third benchmark replaces the respondent-written survey prompts with a more structured prompt that provides the LLM with explicit information about the individual's current state variables and makes explicit assumptions about the economic environment. The prompt includes individual-specific information on all the state variables, along with baseline assumptions about expected asset returns, taxes, and Social Security rules. It also instructs the LLM to act as a professional financial advisor trained in portfolio theory and life cycle planning. The full prompt and assumptions are presented in Appendix A.5.1 and Table A3.[5]

1.5 LLM Selection

We choose an OpenAI model as our primary LLM because survey evidence—both from our survey (Figure E3) and others (Lloyds Banking Group 2025; J.D. Power 2025)—indicates

[5]For simplicity, the academic prompt assumes the individual is single with no dependents. In contrast, household composition does vary across respondents and we incorporate changes in household composition by age in the life cycle model through an equivalence scale.

that ChatGPT is the most commonly used LLM for financial advice.[6] Among OpenAI's model family, we use `GPT-5.2` (OpenAI 2025), which was the latest available model at the time we ran our analysis. We set its reasoning effort to "Low", which approximates the default ChatGPT experience: most consumer queries are served by fast, low-compute models. For the task in Step 4 of Section 1.3 that translates the LLM's textual advice into quantitative recommendations, we use `GPT-5 Mini`, which is a smaller and faster model in the same family. We choose this smaller model because this task involves deterministic extraction rules rather than open-ended reasoning. As a robustness check, we repeat our main analysis using `Gemini 3 Flash` (Google DeepMind 2025), a model of similar vintage from the second-most-commonly-used model provider in our survey, and `GPT-5.6 Terra` (OpenAI 2026), a more recent model from the same provider. We report these results in Appendix E and summarize them in Section 3.4.

2 Describing Prompts and Baseline LLM Advice

This section first describes the qualitative content of the prompts written by our survey respondents and the LLM's corresponding textual responses. We then describe the LLM's baseline advice over the life cycle, and examine its robustness to alternative methodological choices.

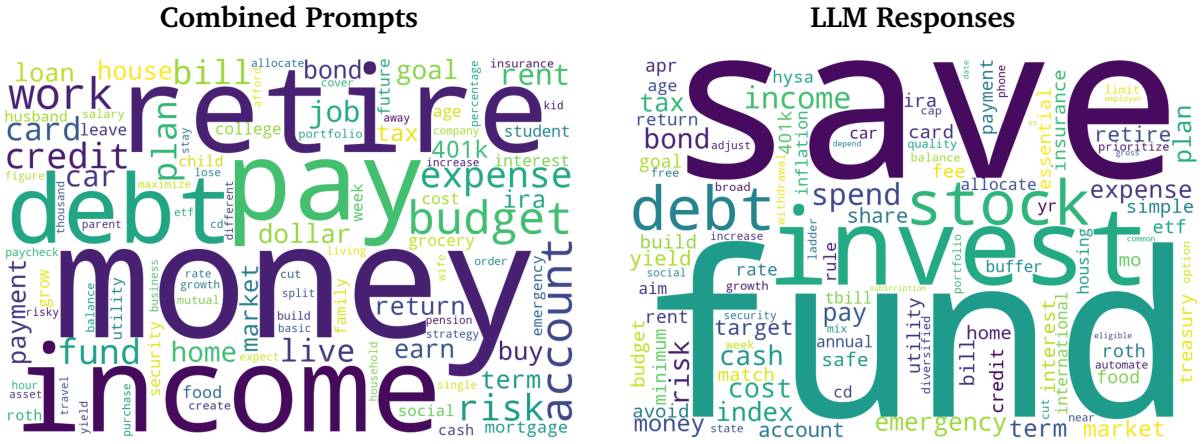
2.1 What Respondents Write and What the LLM Recommends

Measuring topics in prompts and advice. We perform a simple textual analysis by matching keywords in each prompt and LLM response to a dictionary of 27 pre-defined topic categories. For each prompt (or LLM response) and category, we record whether any keyword from that category appears. When comparing prompts with LLM responses, we report the share of respondents who mention each category in their prompts and the corresponding mean mention rate in the advice generated from their prompts, weighting each respondent equally. Full methodological details, including the complete dictionary and topic categories (Table C1), and prompt-level summary statistics (Table E2), are in Appendix C and Appendix E.

What respondents write about. Figure 2 shows word clouds for the survey prompts and LLM's responses. The most prominent words in the prompts are "money," "debt," "income," "pay," and "retire." The topics respondents discuss vary across the three prompts they write (Figure E4). When respondents describe their financial situation, the most common

[6] Using a closed-source model has clear disadvantages from the perspective of reproducibility. However, we chose to prioritize using a model that households are more likely to use in practice.

Figure 2. Word Clouds: Respondent Prompts and LLM Advice



Notes: This figure shows word clouds for respondent-written prompts, pooling across the three prompt questions (left) and for the corresponding LLM advice (right). Word size is proportional to word frequency. Common stopwords and words that mechanically reflect the survey questions are excluded. Figure E5 shows analogous word clouds separately for each prompt question. Details are provided in Appendix C.

categories are Income (35%) and Retirement (30%). When asking for spending advice, Budgeting is the most frequent (45%). When asking for investment advice, Investment Assets (53%) and Risk Preference (40%) are the most common categories.

How LLM advice responds to these prompts. The LLM advice closely reflects the topics survey respondents write about: the five most common topics in the prompts (i.e., Budgeting, Investment Assets, Income, Risk Preference, and Retirement) all appear at substantially higher rates in the LLM's advice (Figure E6). At the same time, the LLM also introduces topics that are less commonly mentioned by respondents. For example, Liquidity appears in 84% of advice responses despite only 6% of respondents mentioning it explicitly, reflecting the model's high propensity to recommend building an emergency fund. Consistent with this pattern, the most prominent words in the LLM advice word cloud are "fund," "invest," "save," "stock," and "debt," along with terms such as "essential," "insurance," and "emergency" (Figure 2). Despite these differences, topic rankings remain broadly aligned between prompts and responses: the rank correlation across all 27 categories is 0.72.

LLM advice provides specific investment recommendations. One important way in which the LLM advice goes beyond respondents' prompts is by naming specific asset classes, financial products, and providers (Figure 3). For example, high-yield savings accounts appear in 59% of advice responses, compared with 3.5% of prompts; bonds appear in 55%, compared with 10%; and Treasury bills, notes, or bonds appear in 40%, compared with 1%. At the product level, fewer than 3% of respondents name any specific ticker, yet LLM advice sometimes recommends branded products: Vanguard investment products appear in 7% of

responses; iShares appears in 3.4%; and cryptocurrency tokens such as Bitcoin or Ethereum appear in 1.9%. The fact that this advice often contains specific recommendations about asset classes, account types, commercial products, and providers raises broader questions about how AI financial advice might affect product competition and the regulation of new sources of financial advice.

2.2 Baseline Quantitative Recommendations

Figure 4 plots the median consumption and equity share by age for individuals following the LLM's advice, along with 25th–75th percentile bands.[7] Table E3 presents additional summary statistics.

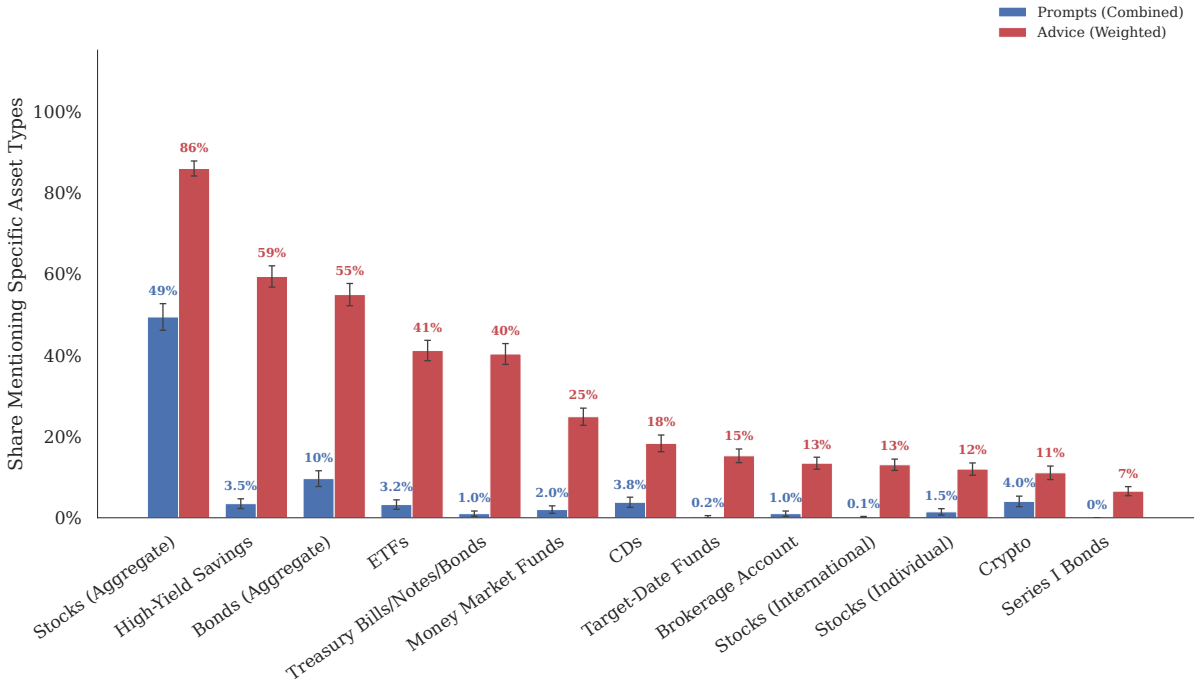
Consumption and saving advice. The left panel of Figure 4 shows that the LLM recommends that individuals smooth their income over working life: consumption is flatter than income, though less flat than our life cycle model prescribes, as we show in Section 3.2. Individuals following the LLM's advice achieve this by saving more during working life and consuming less than their income, while withdrawing and consuming (slightly) more than their income during retirement. This is reflected in their wealth accumulation, with individuals building up a large stock of over $1 million in assets by the time they retire at age 65 (Table E3).

Investment advice. The right panel of Figure 4 shows that the recommended equity shares are relatively high, around 65% on average, and decline with age after approximately age 45. Qualitatively, these portfolio choices are consistent with textbook theories of portfolio choice with human capital (e.g., Merton 1969; Gomes 2020). While less steep than a typical target-date fund glide path, the LLM's recommended equity allocation in working life is similar to the average equity shares that retirement savers choose when making active portfolio choices (Choukhmane and de Silva 2026). The LLM also recommends small allocations to non-diversified assets (e.g., individual stocks and other risky assets such as crypto, gold, commodities, and collectibles). As shown in Figure E8, participation in these asset classes increases with age, reaching approximately 35% for individual stocks and 50% for other risky assets by retirement. However, conditional portfolio shares remain small, averaging 2–3% of financial wealth, so that the vast majority of the LLM's recommended risky asset allocation is in diversified equities.

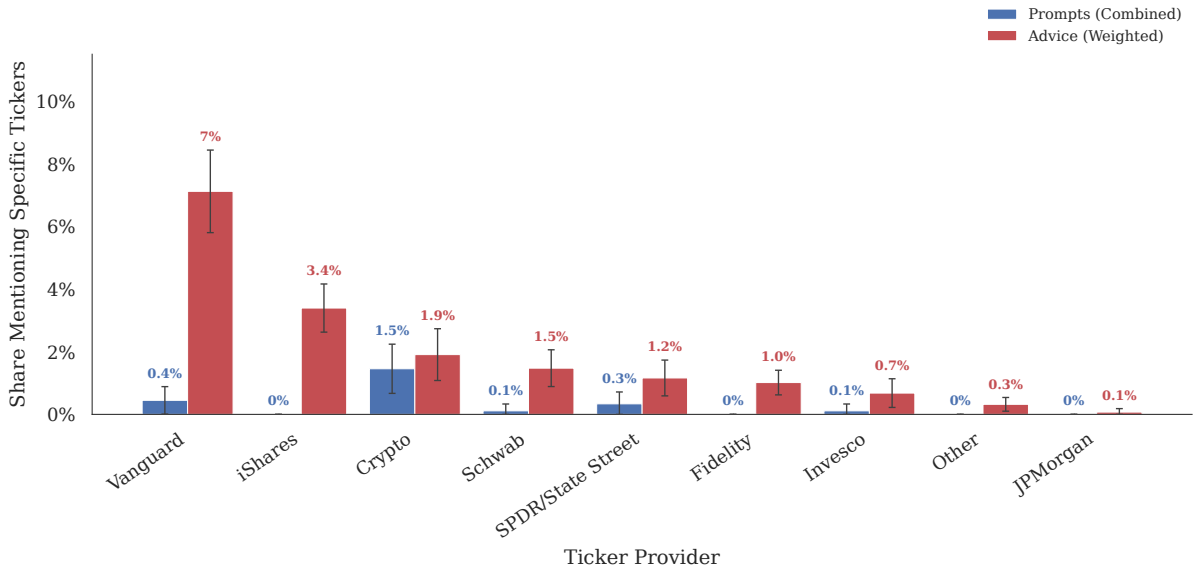
[7] We find quantitatively similar results when using `Gemini 3 Flash` and `GPT-5.6 Terra`, which are shown in Figure E7.

Figure 3. Asset Classes and Products: Prompts vs. Advice

Panel A: Investment Asset Classes

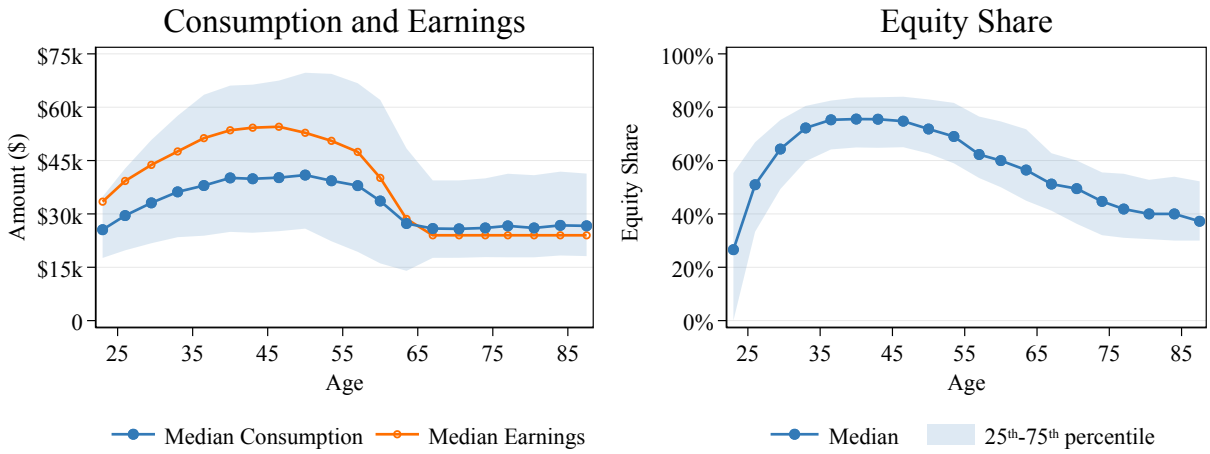


Panel B: Specific Products by Provider



Notes: This figure shows grouped bar charts comparing the mention rates of asset classes and products in respondents' prompts and in baseline LLM advice, as described in Section 1.3. Panel A shows specific investment asset classes with at least 5% mention rate in either prompts or advice, ordered by descending advice mention rate. Panel B shows specific financial products, including tickers and fund names, aggregated by issuing provider and ordered by descending advice mention rate. Blue bars show the share of respondents mentioning the category in any of the three prompts. Red bars show the share of advice responses mentioning the category, with each respondent's prompt weighted equally. Error bars denote 95% confidence intervals.

Figure 4. Life Cycle Profiles of LLM-Recommended Consumption and Equity Shares



Notes: This figure plots the life cycle profiles for individuals following the LLM's recommended choices using the survey prompts, as described in Section 1.3. Ages are grouped into 20 equally sized bins. The left panel shows consumption and post-tax earnings, and the right panel shows the equity share. Dots denote median values within each bin, and shaded areas represent the interquartile range (25^{th}–75^{th} percentile). All values are in 2025 dollars.

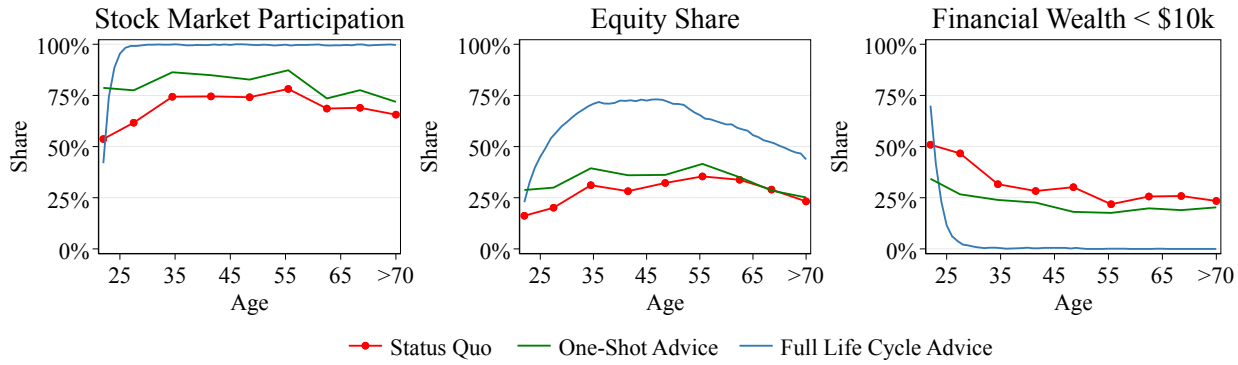
3 LLM Advice and Life Cycle Theory

This section evaluates the financial behavior implied by following LLM advice against the prescriptions of life cycle theory. We first show that following the advice would move respondents toward broad theoretical prescriptions: higher participation in diversified equities, equity shares that decline with age later in life, and larger savings buffers. We then document systematic departures from the normative theory's more subtle prescriptions: the advice implies unusually high patience, relies on simple saving and withdrawal heuristics, smooths consumption imperfectly, and allows portfolios to drift passively with realized returns. Finally, we show that a more structured prompt reduces some (but not all) of these departures.

3.1 Fact 1: Following LLM Advice Would Move Individuals Toward Life Cycle Theory

We begin by asking whether following LLM advice would move respondents closer to the broad prescriptions of life cycle theory relative to observed behavior. This comparison is intentionally qualitative: precise quantitative recommendations from a life cycle model depend on the model specification, preference parameters, the calibration of asset and income risk, and other features of the environment. Therefore, we focus on general principles that are robust across calibrations of standard life cycle portfolio choice models:

Figure 5. Observed Behavior vs. LLM-Recommended Behavior



Notes: This figure compares respondents' observed financial behavior with the behavior implied by following our simulated LLM advice. The three panels show average stock market participation rate (left), average equity share (middle), and the share of individuals with financial wealth below $10,000 (right). Ages are grouped into eight 6-year bins and one last bin that captures every observation age 70 and above. Red lines show respondents' observed behavior. Green lines show the one-shot advice benchmark, which applies the LLM advice generated from each respondent's own prompt to that respondent's current financial situation. Blue lines show the full life cycle simulation, in which simulated individuals follow LLM advice in each period, and prompts are updated to reflect the evolving state variables in the simulation, as described in Section 1.3.

broad participation in diversified equities, equity shares that decline later in life, and the accumulation of precautionary savings buffers (e.g., Gomes 2020).

We compare respondents' observed behavior with two versions of LLM-recommended behavior. The first is a one-shot benchmark, which uses the advice generated from each respondent's original survey prompt and current self-reported financial situation, without replacing state variables or simulating subsequent choices.[8] The second is the full life cycle simulation from Section 1.3, in which individuals follow LLM advice each period using prompts drawn and updated to reflect evolving state variables in the life cycle simulations. Figure 5 plots respondents' observed behavior in red, the one-shot advice benchmark in green, and the full life cycle simulation in blue.[9]

Advice would raise participation in diversified equities. As shown in the left panel of Figure 5, one-shot advice already points respondents toward broader stock market participation. Consistent with the empirical evidence on limited stock market participation, one-third of respondents in our survey report holding no equities, and their average equity share is around 30%. Relative to observed behavior, the green line shows that one-shot advice would raise participation by approximately 5–10 percentage points and raise average equity shares by approximately 5 percentage points. The full life cycle simulation amplifies

[8]We convert the qualitative advice into numerical choices using each respondent's self-reported age, income, employment status, and asset holdings, together with the same translation procedure described in Step 4 of Section 1.3.

[9]We find quantitatively similar results when using `Gemini 3 Flash` and `GPT-5.6 Terra`, as shown in Figure E9.

this movement: following the advice each period leads to near-universal participation and raises equity shares by as much as 40 percentage points relative to observed behavior.

Advice would generate higher equity shares that decline later in life. As shown in the middle panel of Figure 5, recommended portfolio allocations also move closer to the age profile implied by standard life cycle theory. Respondents' observed equity shares are roughly flat by age and, if anything, increase through age 55. By contrast, in the full life cycle simulation, equity shares are higher, rising early in working life as individuals build emergency buffers, then declining after approximately age 45. As we discuss in the next section, however, this decline is quantitatively more modest than what a standard life cycle model predicts.

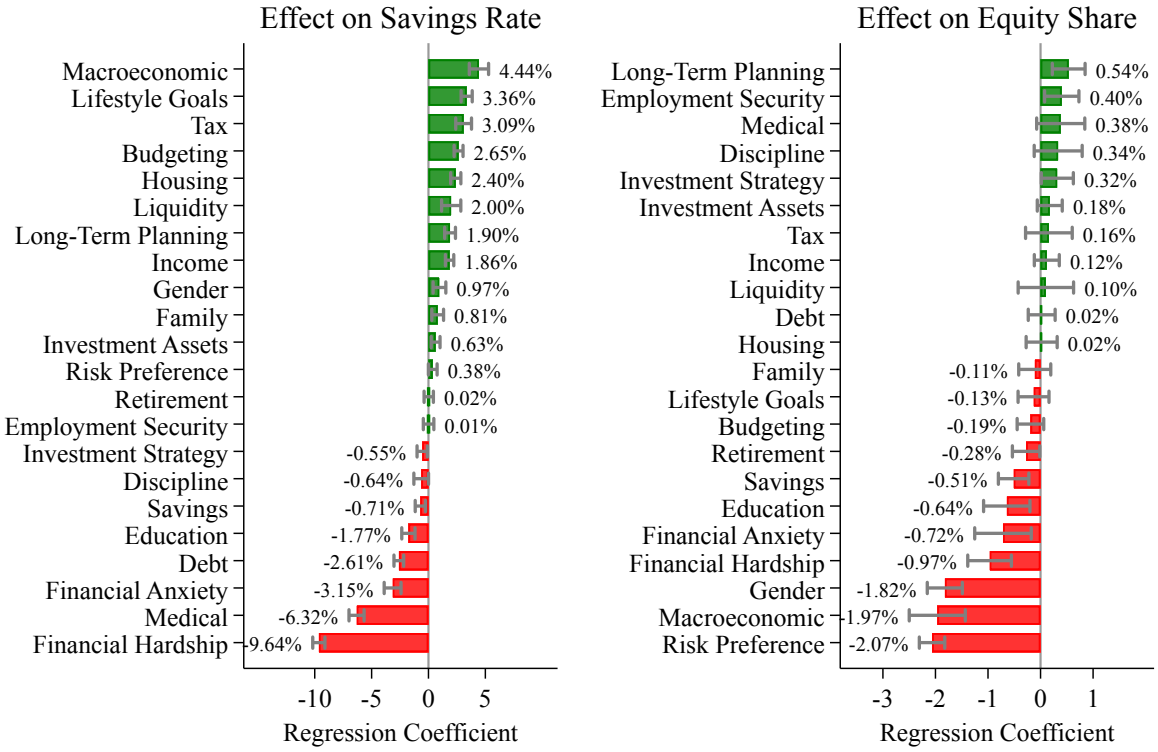
Advice would build larger savings buffers. The right panel of Figure 5 shows that following LLM advice would lead respondents to build financial buffers more quickly. More than 20% of respondents in every age group report holding less than $10,000 in financial wealth. One-shot recommendations reduce this share in all age groups, but cannot fully offset low accumulated wealth within a single year. In contrast, when we simulate individuals following LLM advice every period, virtually all build a savings buffer above $10,000 by age 30.

Advice responds sensibly to the topics respondents raise. Beyond these average patterns, LLM recommendations also vary systematically with the topics respondents mention in their prompts, often in ways that align with life cycle theory. To document this, we exploit the fact that, within each employment-by-income-by-age prompt bucket, the specific prompt assigned to a simulated individual is randomly drawn. This randomness allows us to compare recommendations generated by different prompts for otherwise similar simulated individuals. Specifically, we regress two outcomes, the net saving rate and the change in equity share, on indicators for each of the dictionary categories defined in Section 2.1, controlling for income, wealth, age fixed effects, and prompt-bucket fixed effects.[10] Figure 6 reports coefficients for categories mentioned in at least 5% of prompts.

The results suggest that the LLM responds to prompt content in ways that are broadly consistent with life cycle theory. When prompts mention macroeconomic conditions, typically uncertainty about inflation or recessions, the LLM raises the recommended saving rate by around 4pp and lowers the recommended equity share by 2pp, consistent with a greater precautionary motive. Mentions of financial hardship lower the recommended saving rate by around 10pp and the equity share by 1pp, consistent with immediate liquidity

[10] We use the change in equity share rather than its level because of the substantial inertia documented in Figure 9.

Figure 6. Association between Prompt Topics and LLM Recommendations



Notes: This figure plots the estimated association between the dictionary categories mentioned in each prompt and the LLM's recommended net savings rate (left) and change in equity share (right), using the baseline results described in Section 1.3. Each bar is the coefficient on an indicator for whether the prompt mentions the category, from a regression (one per outcome) that includes all category indicators and controls for income, wealth, age fixed effects, and prompt-bucket fixed effects. Dictionary categories and corresponding keywords are reported in Table C1. Only categories that appear in at least 5% of prompts are included in the regression and in the figure. Error bars denote 95% confidence intervals.

needs. Mentions of risk preferences modestly raise the saving rate by 0.4pp but lower the recommended equity share by 2.1pp. This last pattern echoes Rumpf et al. (2026), who find that LLM portfolio recommendations are more sensitive to stated risk preferences than human advisors.

Summary. Taken together, these results show that following LLM advice would move respondents toward broad prescriptions of life cycle theory with higher participation in diversified equities, high equity shares that decline later in life, and larger savings buffers. This is not only an average pattern: the LLM also adjusts its recommendations for saving and risk-taking in response to economically relevant information in respondents' prompts. These findings should be interpreted as the effect of following advice, not the effect of receiving advice. Whether LLM recommendations change actual financial decisions, and how those effects compare to other forms of financial advice, are important questions for

future work.

3.2 Fact 2: LLM Advice Departs from Life Cycle Theory in Systematic Ways

We next compare LLM advice to the more quantitative predictions from our calibrated life cycle model. While Fact 1 shows that following the advice would move respondents toward broad principles of life cycle theory, this comparison highlights some systematic departures from the model's more specific prescriptions. Specifically, the advice implies unusually high patience, relies on simple saving and withdrawal heuristics, smooths consumption imperfectly in response to income shocks, and allows portfolios to drift passively rather than be actively rebalanced.

Advice implies unusually high patience. We summarize the distance between LLM advice and the life cycle model by estimating the preference parameters that best fit the LLM's recommendations. Specifically, we use the simulated method of moments (SMM) to estimate the intertemporal discount factor (β) and coefficient of relative risk aversion (γ) that minimize the distance between model-generated and LLM-generated life cycle profiles.[11] We target two sets of moments: the wealth-to-income ratio during working years and the equity share over the full life cycle. Additional details are provided in Appendix D.

The first row of Table 2 reports estimates targeting moments from the baseline advice simulation with survey prompts. The estimated risk aversion, $\hat{\gamma} = 5.3$, is moderate and within the range of standard estimates. By contrast, the estimated discount factor, $\hat{\beta} = 1.034$, is above one. This high value reflects the LLM's recommendation to consume less than income throughout working life, which generates substantial wealth accumulation by retirement. One possible explanation is unmodeled bequest motives, although bequests are rarely mentioned in respondents' prompts. Another possible interpretation is that the LLM advice reflects a form of paternalism: it may recommend a higher saving target than the standard model would prescribe because individuals may systematically undershoot the targets they receive. In that sense, conservative advice can partly offset imperfect implementation. However, if recommendations are followed exactly, a standard life cycle model can rationalize this advice only with an unrealistically high discount factor.

Panel A of Figure E10 shows the fit of the estimated model on the targeted moments. The model matches the LLM's wealth accumulation during working life and its equity shares

[11] This approach is similar to Cook et al. (2026), who estimate structural preference parameters from LLM behavior in dictator games and job search scenarios. Abdel Haq et al. (2026) pursue an alternative approach, eliciting LLM preferences directly, and show they align closely with those of human subjects.

Table 2. Simulated Method of Moments Estimation Results

Prompt Type	LLM	$\hat{\beta}$	$\hat{\gamma}$
Survey	`GPT-5.2`	1.034	5.3
Academic	`GPT-5.2`	0.990	4.7
Survey	`Gemini 3 Flash`	1.054	5.1
Survey	`GPT-5.6 Terra`	1.032	5.3

Notes: This table reports simulated method of moments (SMM) estimates of the annual time discount factor, β, and coefficient of relative risk aversion, γ, for each prompt-model combination shown. Survey prompts are respondent-written prompts from the survey; academic prompts are researcher-designed prompts described in Section 1.4. The estimates minimize the weighted sum of squared errors between a set of moments from simulations with LLM advice and another from simulations using the model's policy functions. For a full description of the targeted moments and estimation procedure, see Appendix D.

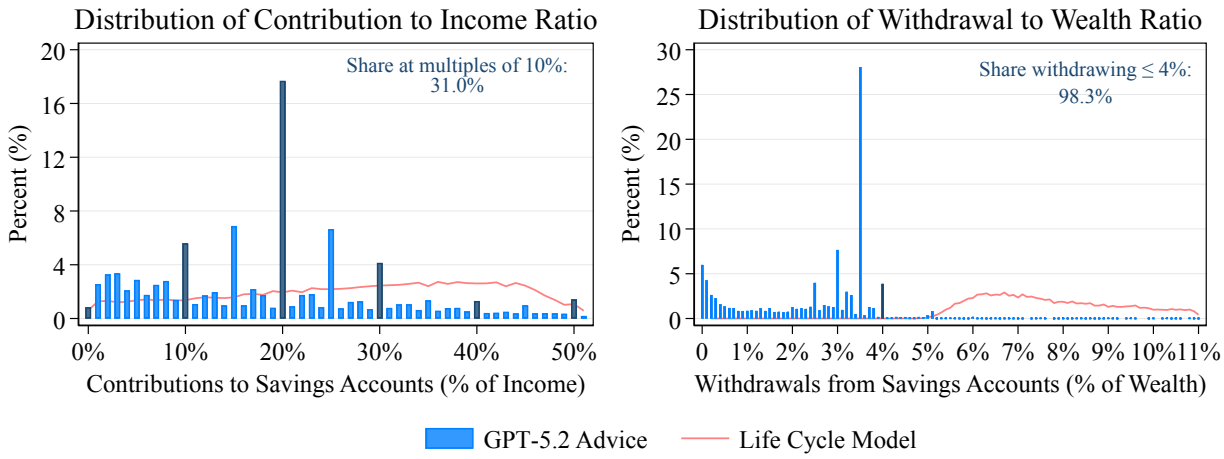
after age 45, but cannot match the relatively low and upward-sloping equity shares that the LLM recommends between ages 25 and 45.[12]

Advice relies on simple saving and withdrawal heuristics. We next examine the LLM's individual-level recommendations that underlie the aggregate life cycle profiles. Figure 7 plots the distributions of LLM-recommended saving rates, saving amounts, and withdrawal rates in the baseline advice simulation (blue histograms), along with the corresponding distributions implied by the life cycle model estimated in Table 2 (red lines). The LLM recommendations exhibit bunching at round numbers: 31% of saving rates are multiples of 10%, and 34% of saving amounts are multiples of $5,000. These spikes are absent by construction in the life cycle model, where choices vary continuously with state variables. Withdrawal advice displays an even stronger heuristic pattern. More than 98% of LLM-recommended withdrawals are at or below 4% of assets, echoing the 4% retirement withdrawal rule popularized by Cooley et al. (1998) and commonly used by financial advisors. By contrast, almost no individuals in the life cycle model choose withdrawal rates below 4%, consistent with the model prescribing faster decumulation than the LLM.

Consumption is insufficiently smoothed. As shown in the left panel of Figure 8, the LLM's recommended consumption (blue line) closely tracks income (orange line) over the life cycle. The right panel shows that this limited consumption smoothing also appears around simulated transitions into unemployment. After job loss, simulated income falls by roughly 50%, and the LLM's recommended consumption falls by a similar percentage, consistent with empirical evidence on consumption responses to unemployment in Gruber (1997) and Ganong and Noel (2019). In the life cycle model, by contrast, consumption barely falls after job loss because individuals draw down buffer stocks to absorb the shock

[12]This mismatch reflects a standard implication of life cycle portfolio models: absent participation costs (Gomes and Michaelides 2005) or correlated labor income risk (Benzoni et al. 2007; Catherine 2022), optimal equity shares are high early in life.

Figure 7. Saving and Withdrawal Heuristics in LLM Advice: Survey Prompts



Notes: This figure plots the distributions of saving rates and withdrawal-to-wealth ratios implied by the baseline LLM advice described in Section 1.3. The left panel restricts the sample to ages below 65 and plots saving contributions as a share of income. The right panel restricts the sample to ages 65 and above and plots withdrawals as a share of wealth. Blue histograms show the choices implied by the LLM advice. Dark blue bars highlight mass at common heuristic values, including saving rates at multiples of 10% and withdrawal-to-wealth ratios at 4%. Red lines show the corresponding distributions generated by the estimated life cycle model using the preference parameters reported in Table 2.

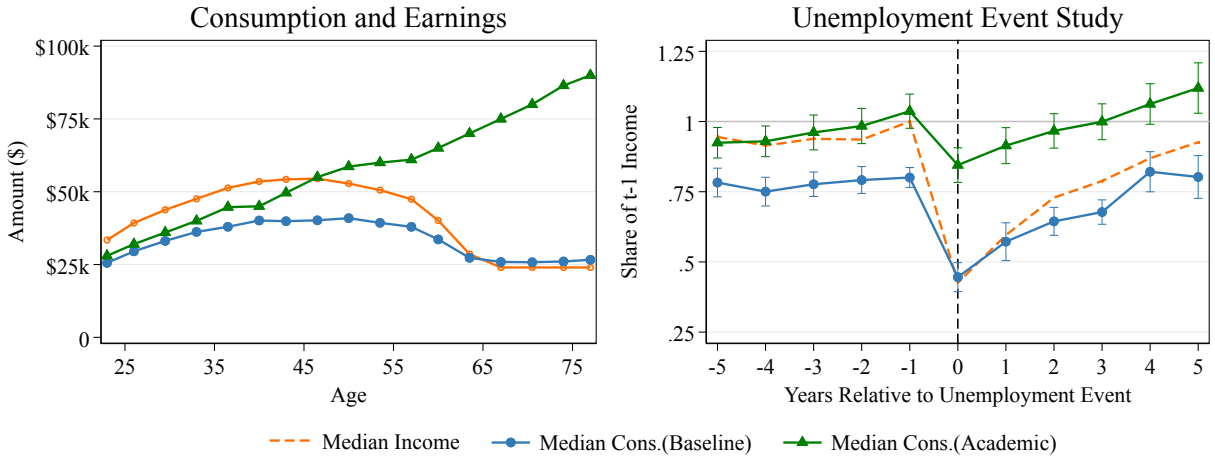
(Figure E11). The fact that LLM advice smooths consumption insufficiently both over predictable life cycle changes in income and around employment and income shocks is particularly notable, given that simulated individuals hold sizable liquid balances.[13]

Portfolio recommendations exhibit passive drift. Following Calvet et al. (2009), we measure passive portfolio drift as the change in an individual's equity share from period t to $t+1$ that would occur if the individual made no active portfolio adjustment and simply let returns change the portfolio shares. Figure 9 plots actual changes in equity shares against this passive-drift measure. In the life cycle model with CRRA preferences, these two measures are essentially uncorrelated: target allocations are not sensitive to wealth, and individuals actively choose their portfolio each period. In contrast, LLM-recommended portfolio changes move almost one-for-one with passive drift, exhibiting substantial inertia consistent with evidence from household portfolios (Calvet et al. 2009; Choukhmane and de Silva 2026).[14] This inertia persists even when we skip the translation step and provide the LLM with all relevant state variables, including the existing portfolio allocation (as

[13] Results are similar in our baseline specification, where asset holdings are mentioned in the prompt only if the original survey respondent included such information, and in a robustness exercise that appends all relevant state variables to the prompt; see discussion in Section 3.4 and Figure E12.

[14] Quantitatively, Calvet et al. (2009) estimate a regression coefficient of 0.5 between actual and passive portfolio changes using observational data, suggesting that LLM recommendations exhibit roughly twice as much inertia.

Figure 8. Consumption Smoothing: Survey vs. Academic Prompts



Notes: This figure compares consumption profiles from the baseline using the survey prompts with profiles from the academic-prompt benchmark described in Section 1.4. The left panel plots median consumption and median income over the life cycle. The right panel plots the median ratio of consumption relative to income one year prior to unemployment within five years before and after the unemployment event, restricting to individuals between ages 30 and 55. The blue and green lines show consumption implied by the baseline LLM advice and the academic benchmark, respectively; the orange line shows the corresponding median income. Error bars denote 95% confidence intervals. All values are in 2025 dollars.

we discuss later in Section 3.4). This suggests that the inertia in the LLM's advice is not an artifact of respondents omitting information about their current portfolios. Instead, the LLM tends to advise individuals on how to allocate new savings rather than explicitly recommending that they rebalance existing holdings.

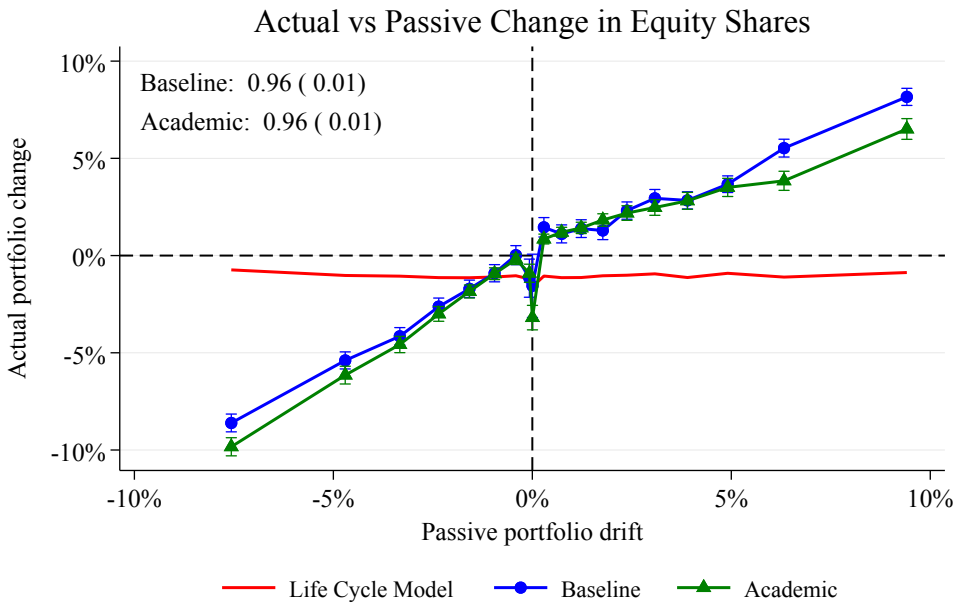
3.3 Structured Prompts Reduce Some Departures from Life Cycle Theory

We next examine whether replacing the prompts written by survey respondents with a more structured researcher-designed prompt, which we call the *academic prompt*, moves LLM advice closer to life cycle theory. As described in Section 1.4, the academic prompt provides explicit values for all current state variables, clarifies assumptions about the economic environment, and frames the task as providing professional financial advice.[15]

Structured prompts reduce heuristics. The LLM's advice produced in response to the academic prompt exhibits less reliance on simple saving and withdrawal heuristics. Figure 10 shows that the distribution of recommended saving rates exhibits less bunching at round

[15]The academic prompt also skips the translation step by instructing the LLM to provide quantitative recommendations directly. As we discuss in Section 3.4, skipping the translation step and providing all state variables while using survey prompts generates advice similar to our baseline. Therefore, the differences in this section can primarily be attributed to the content and framing of the academic prompt, rather than to differences in the translation pipeline or information about state variables. The corresponding life cycle profiles and comparison with observed behavior are shown in Figure E13 and Figure E14.

Figure 9. Passive Portfolio Drift in LLM Advice



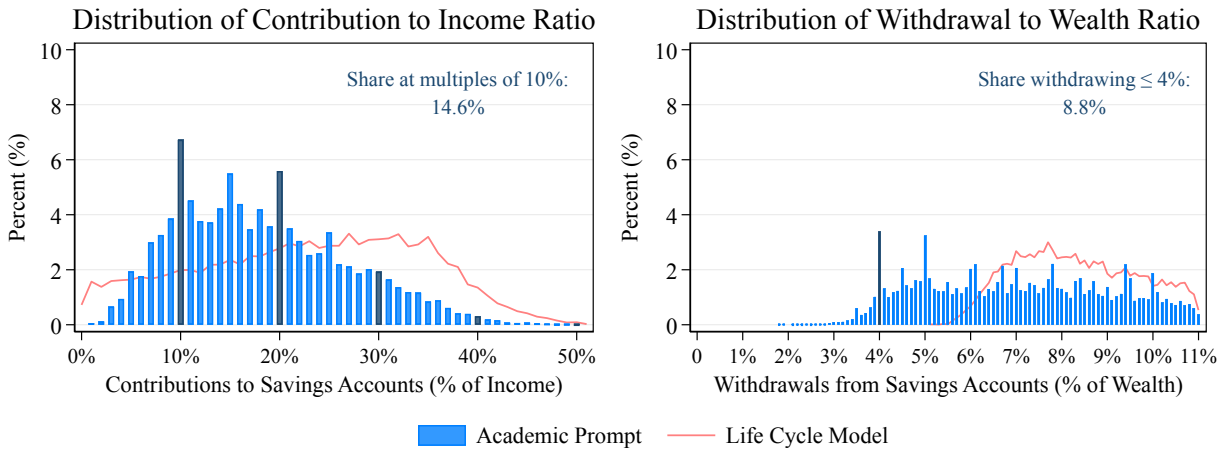
Notes: This figure compares actual portfolio changes with passive portfolio drift for the baseline LLM advice, the academic-prompt benchmark, and the estimated life cycle model. Blue denotes baseline advice using survey prompts, green denotes advice using the academic prompt, and red denotes the life cycle model. Actual portfolio change is the adjustment in the equity share before returns are realized from one period to the next. Passive portfolio drift is the mechanical change in the equity share induced by asset returns over the period, assuming no rebalancing. The sample is restricted to observations with equity shares strictly between 0% and 100%, and observations are grouped into 20 equal-sized bins of passive portfolio drift. Error bars denote 95% confidence intervals. Regression slopes estimated on the raw data and standard errors for the survey and academic prompts are reported in the top left.

numbers than under respondent-written survey prompts, and withdrawal recommendations are less concentrated at values at or below 4% of assets. These changes translate into a more reasonable estimated discount factor: the academic prompt estimate is $\hat{\beta} = 0.99$, relative to $\hat{\beta} = 1.034$ under survey prompts (second row of Table 2). Portfolio choices, by contrast, remain broadly similar and imply a similar estimate of risk aversion, $\hat{\gamma} = 4.7$.

Structured prompts improve consumption smoothing. The advice produced in response to the academic prompt exhibits greater consumption smoothing over the life cycle and in response to income shocks. The left panel of Figure 8 shows that consumption no longer tracks income over the life cycle (green line), and individuals decumulate more substantially in retirement. The right panel shows a parallel improvement around transitions into unemployment. With the academic prompt (green), the drop in consumption is smaller and reverses more quickly, consistent with greater use of accumulated savings buffers to absorb income shocks.

Structured prompts do not reduce portfolio inertia. In contrast, structured prompts do not meaningfully reduce portfolio inertia. As shown in Figure 9 (green line), the relationship

Figure 10. Saving and Withdrawal Heuristics in LLM Advice: Academic Prompts



Notes: This figure reproduces Figure 7 using the LLM advice generated in response to the academic prompt. See the notes to Figure 7 for additional details.

between actual changes in equity shares and passive portfolio drift remains close to one under both survey and academic prompts. This suggests that the lack of active rebalancing is not primarily due to missing state variables or the informal wording of respondent prompts. Instead, it appears to reflect a more prevalent pattern in the LLM's financial advice.

3.4 Robustness of the Patterns in Baseline LLM Advice

Next, we show that the broad patterns of advice documented above are robust to alternative methodological choices, including using a different model, eliciting quantitative recommendations directly, or repeating identical queries multiple times.

Alternative models. To assess whether our findings are specific to `GPT-5.2`, we rerun the advice-generation step using `Gemini 3 Flash` and `GPT-5.6 Terra` while retaining `GPT-5 Mini` for the translation task. The advice provided by these alternative models yields very similar patterns. Relative to respondents' observed behavior, both models recommend higher stock market participation, higher equity shares, and larger savings buffers (Figure E9). Quantitatively, the advice provided by both models implies similar structural time and risk preference parameters (Table 2) and a similar model fit (Figure E10, Panels C and D). Like `GPT-5.2`, `Gemini 3 Flash` and `GPT-5.6 Terra` rely on simple saving and withdrawal heuristics (Figure E15), and exhibit limited consumption smoothing around job loss and substantial passive portfolio drift (Figure E16). Collectively, these results suggest that our findings are not specific to a single model.

Eliciting quantitative advice directly. Our baseline simulation first elicits qualitative advice using survey-based prompts and then uses a second LLM call to translate that advice into quantitative choices, with the full set of state variables entering only at the translation stage. As a robustness check of this two-step pipeline, we consider an alternative pipeline that elicits quantitative recommendations directly by appending the state variables to the survey prompt and instructing the LLM to return recommendations, without a separate translation step. Figure E12 shows that the direct quantitative pipeline reproduces the patterns we document in this section for our baseline advice. Consumption and stock-share life cycle profiles are very similar to the baseline two-step pipeline (Panel A), passive portfolio drift is essentially unchanged, and the consumption response to job loss is nearly identical. Reliance on heuristics persists, though it is somewhat attenuated (Panel B): the share of saving rates at multiples of 10% falls from 31.0% to 18.7%, and the share of withdrawals at or below 4% of assets falls from 98.3% to 92.4%. Thus, our main results are not driven by the separate translation step or by withholding state variables during advice generation.

Repeated queries. LLM advice is inherently stochastic: the same prompt may produce different textual advice across queries, and the same textual advice may be translated into different quantitative recommendations.[16] To quantify the importance of randomness in different steps in our methodology, we repeat the advice-generation and translation steps for the one-shot advice (described in Section 3.1).[17] For each individual, we conduct two exercises. First, we pass the same prompt through the full pipeline five times. Second, we hold the textual advice fixed and translate it into quantitative recommendations five times. In each exercise, we calculate the standard deviation of the five resulting quantitative recommendations. Figure E17 reports the distribution of these standard deviations across individuals. When we repeat the full pipeline, the median standard deviation is $3,220 for recommended annual consumption and 6.5 percentage points for the recommended equity share. When we repeat only the translation of a fixed piece of advice, the corresponding medians fall to $335 and 1.6 percentage points. At the median, translation-only variation is therefore approximately one-tenth as large as full-pipeline variation for consumption and one-quarter as large for equity shares. Thus, although the translation step introduces some randomness, most of the variation in quantitative recommendations arises when the model generates new textual advice in response to the same prompt.

[16] `GPT-5.2` is a reasoning model and does not allow users to set the temperature to zero. Rumpf et al. (2026) document similar stochasticity in the financial advice provided by LLMs.

[17] We use one-shot advice due to its significant cost advantage.

4 Heterogeneity in LLM Advice: Supply and Demand

Our analysis thus far has focused on characterizing the average properties of LLM advice. In this section, we examine whether this advice varies systematically across groups of individuals, and how much of this variation arises from demand or supply. Demand-side differences reflect variation in the prompts respondents write: what topics they raise, what information they provide, and how they frame their financial decisions. Supply-side differences reflect how the LLM responds to the same or similar prompts when observable characteristics change. We first show that simulated individuals accumulate more wealth when advice is generated from prompts written by respondents with higher financial literacy or prior AI experience, or by men. We then use randomized variation in prompt labels to determine the roles of demand and supply in determining gender and racial differences in advice.

4.1 Fact 3: LLM Advice Varies Systematically Across Groups of Individuals

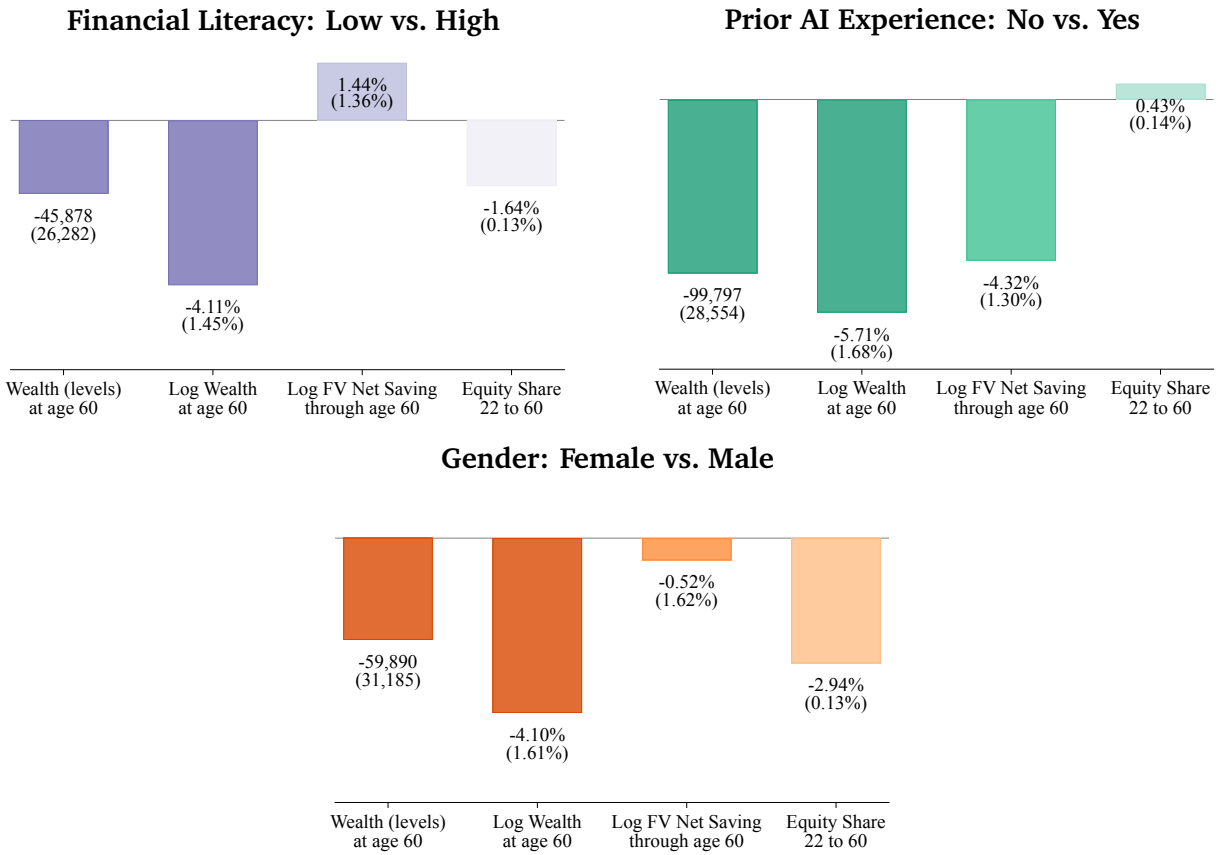
To quantify heterogeneity in LLM advice, we repeat the life cycle simulation separately for groups of survey respondents based on their financial literacy, prior AI experience, and gender. For each characteristic, we split respondents into two groups and simulate life cycle paths using only prompts written by respondents in one group at a time. Importantly, all simulations use the same realizations of exogenous shocks (i.e., income, employment, mortality, and asset returns), so any differences in simulated outcomes are driven entirely by differences in the prompts written by each group and the LLM's responses.

Figure 11 summarizes the resulting differences in life cycle outcomes. In each panel, the first two bars report the implied difference in average wealth at age 60 across the two groups, in levels and in logs. The next two bars relate these wealth differences to differences in cumulative savings, measured as the future value of net saving compounded at the risk-free rate, and differences in portfolio allocation, measured by the average equity share between ages 22 and 60.[18]

Low-financial-literacy prompts accumulate less wealth. The top-left panel of Figure 11 splits respondents into high- and low-literacy groups based on their performance on the Big Five financial literacy test (Lusardi and Mitchell 2023). Prompts written by respondents

[18]We compound net saving at the risk-free rate rather than at ex-post portfolio returns, so that this measure is not directly affected by portfolio choices. Additionally, we stop at age 60 to ensure that these differences entirely reflect working-life advice.

Figure 11. Heterogeneity in LLM Financial Advice



Notes: This figure compares life cycle outcomes implied by LLM advice generated from prompts written by different groups of survey respondents. All exogenous shocks are held fixed across prompt groups. The top-left panel compares simulations with prompts written by respondents with high versus low financial literacy, measured using the Big Five financial literacy questions. The top-right panel compares simulations with prompts written by respondents who have versus have not previously used AI for financial advice. The bottom panel compares simulations with prompts written by male versus female respondents. In each panel, the four bars report the difference in simulated wealth at age 60 (in levels), log wealth at age 60, the log future value of net saving through age 60 cumulated at the risk-free rate, and the average equity asset share between ages 22 and 60. All values are in 2025 dollars. Standard errors are reported in parentheses.

who answer at least one of these questions incorrectly generate advice that results in wealth at age 60 that is $46,000, or 4.1%, lower on average. This difference is driven primarily by portfolio choice: prompts from the less financially literate group lead to a 1.64pp lower average allocation to equity between ages 22 and 60. Differences in cumulative savings go in the opposite direction and would, if anything, partially offset the wealth difference, since the less financially literate group is advised to save more through age 60. The word clouds in Figure E18 suggest that this difference in investment advice reflects both what the two groups ask about in their prompts and the textual advice they receive; Figure E19 provides illustrative examples.

Prompts without AI experience accumulate less wealth. In our survey, 48% of respondents report having used an AI tool for financial advice or information in the past three

months (Figure E20), a share consistent with other recent industry surveys (Lloyds Banking Group 2025; J.D. Power 2025). The top-right panel of Figure 11 shows that prompts written by respondents who have not previously used AI for financial advice generate recommendations that result in wealth at age 60 that is $100,000, or 5.7%, lower on average. Unlike the financial literacy result, this difference is driven almost entirely by saving behavior: the future value of cumulative net saving is 4.32% lower in the non-AI-user simulation, while the difference in average equity asset shares is economically small (0.43pp).[19] Figure E21 provides illustrative prompts from each group, and word clouds in Figure E22 show the corresponding differences in the LLM's textual advice.

Prompts from women accumulate less wealth. The bottom panel of Figure 11 compares advice given by the LLM to men and women. Prompts written by women generate recommendations that result in wealth at age 60 that is $60,000, or 4.1%, lower on average than prompts written by men. This difference is mainly driven by portfolio choice: the average equity share recommended to women is 2.94pp lower than the share recommended to men, while the future value of cumulative net saving is nearly identical across the two groups. Figure E23 shows that this difference in equity share recommendations appears across all age and income groups, and that men are also advised to rebalance their portfolios more actively. This gender difference in investment advice is consistent with Foltyn and Olsson (2026), who find an average gender gap in equity allocations of 1.8pp across 33 LLMs in researcher-designed scenarios. It is also consistent with patterns observed in observational data (Agnew et al. 2003) and in professional financial advice (Bucher-Koenen et al. 2025).

4.2 Demographic Differences Reflect Both Demand and Supply

The heterogeneity results above raise a natural question: do differences in LLM advice reflect differences in what respondents ask (demand), or differences in how the LLM responds to otherwise similar prompts (supply)? We answer this question by first focusing on gender specifically because it has been a focus of study in the literature on human financial advice (Bucher-Koenen et al. 2025), and because our prompt design allows us to experimentally separate demand- from supply-side drivers of gender differences in a way that is difficult in human-advisor settings (Reuter and Schoar 2024). Two features of this exercise are worth stating upfront. First, it separates the two channels rather than decomposing the gap in Figure 11: identifying the supply channel requires inserting a gender label that respondents did not write, so the experiment measures the LLM's response

[19] Although AI use and financial literacy are correlated, the distinct mechanisms behind the wealth gaps (i.e., portfolio allocation versus cumulative saving) suggest that these two comparisons capture different sources of heterogeneity in the advice.

to an explicit gender statement rather than to the subtler ways gender may be conveyed in naturally written prompts (or be reflected in an LLM memory of past interactions with the user). Second, the estimates capture the effect of gender on the advice generated in a single period, which are not directly comparable to the cumulative life cycle differences reported above.

Men and women write prompts and receive advice with different language. Figure 12 shows word clouds of terms that are overrepresented in prompts (Panel A) written by men versus women and in the corresponding advice (Panel B). Men are much more likely to mention stocks and discuss investment strategies, while women are more likely to explicitly mention their gender, discuss debt, and mention safe, liquid investment options such as certificates of deposit (CDs), deposits, and credit unions. The LLM's responses also differ systematically, with more aggressive investment language in advice to men. One of the most common words overrepresented among men is "rebalance," consistent with the gender differences in recommended portfolio rebalancing shown in Figure E23.

Randomized labels to separate demand from supply. To separate demand from supply, we use the 83% of prompts that do not explicitly reveal the respondent's gender, whether directly or through gendered self-descriptions (e.g., prompts written by a woman referencing being a "mother" or a "wife"), and randomly insert either "I am a man" or "I am a woman" at the beginning of the prompt each time it is used in the simulation. Because the inserted label is randomized, its effect identifies the supply channel: how the LLM responds to gender holding the content of the prompt fixed. The effect of the prompt author's gender, holding the inserted label fixed, identifies the demand channel: how prompts written by men and women differ in ways that affect advice.

Using these labeled prompts, we simulate life cycles as in the baseline and estimate the following regression during working life

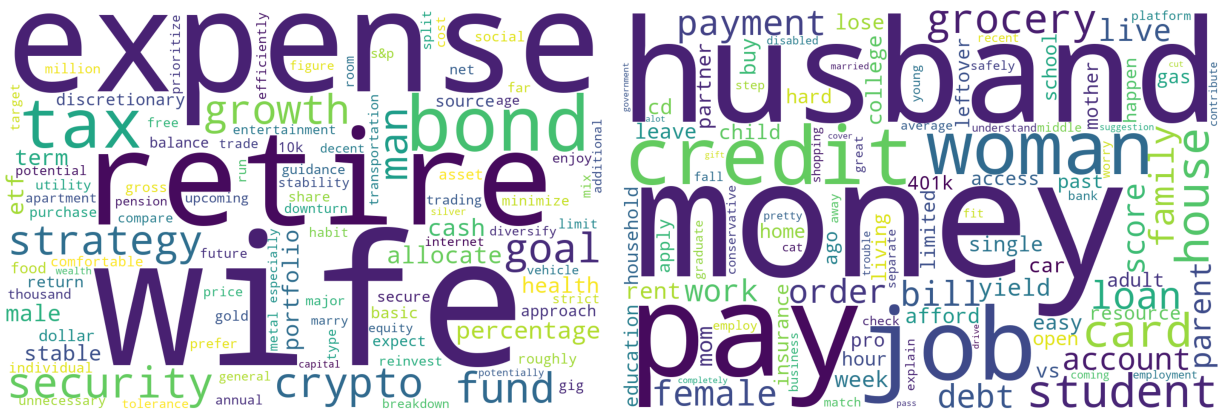
$$Y_{i,t} = \beta_D \,\text{FemaleAuthor}_{i,t} + \beta_S \,\text{FemaleLabel}_{i,t} + \delta_{i,t}^{\text{bucket}} + \varepsilon_{i,t} \tag{1}$$

where $Y_{i,t}$ is the outcome of interest for individual i and age t, $\text{FemaleAuthor}_{i,t}$ indicates that the prompt used to generate advice in period t was written by a woman, and $\text{FemaleLabel}_{i,t}$ indicates that the prompt used to generate advice was randomly assigned a female label. The fixed effects $\delta_{i,t}^{\text{bucket}}$ correspond to the age × income × employment cells within which prompts are randomly drawn, ensuring that comparisons are made among prompts assigned to similar simulated individuals. The coefficient β_D captures demand (the effect of prompts written by women rather than men, holding labeled gender fixed), while β_S captures supply

Figure 12. Gender Differences in Prompts and Advice

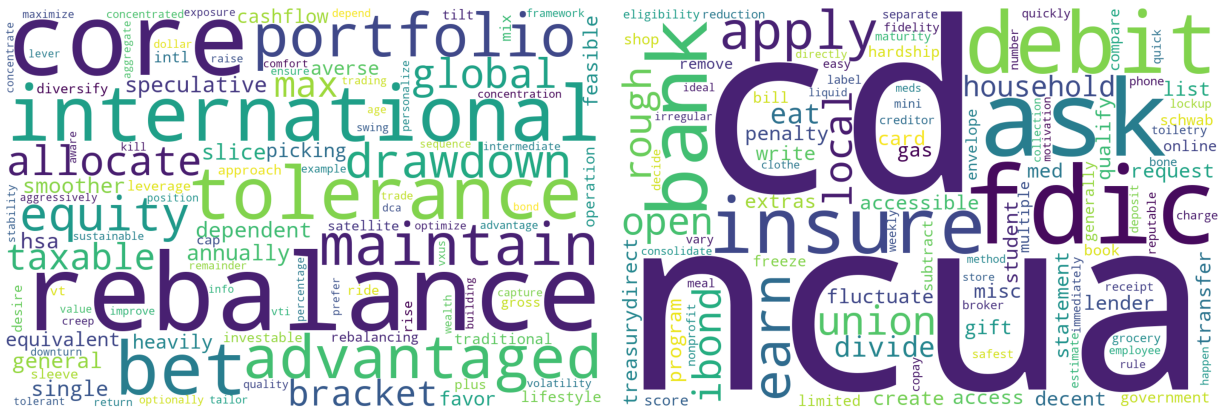
Panel A: Words Overrepresented in Prompts by Gender

Men **Women**



Panel B: Words Overrepresented in Advice by Gender

Men **Women**



Notes: This figure compares gender differences in respondent-written prompts and the corresponding LLM advice. Panel A pools the three prompt questions and shows words overrepresented in prompts written by men (left) and women (right). Panel B shows words overrepresented in the corresponding LLM advice for prompts written by men (left) and women (right). Word size is proportional to the difference in word frequency between groups. Common stopwords and words that mechanically reflect the survey questions are excluded.

(the effect of a gender label, holding prompt content fixed). Both coefficients measure period-by-period effects on recommended choices from a single prompt draw; they do not cumulate advice differences over the life cycle as in Figure 11.

Gender gap in diversified holdings is mostly demand-driven. The first three columns in the top row of Table 3 show the results from estimating (1) with the change in diversified equity share as the dependent variable.[20] Moving from a prompt written by a man and assigned a male label to a prompt written by a woman and assigned a female label reduces

[20]We use the change in diversified equity share rather than its level because of the substantial inertia documented in Figure 9. Results are quantitatively similar when using the level of diversified equity shares.

Table 3. Separating Demand and Supply Drivers of Demographic Differences in Advice

	Diversified Equity Share			Nondiversified Asset Share			Net Saving Rate		
	Author (Demand)	Label (Supply)	Total	Author (Demand)	Label (Supply)	Total	Author (Demand)	Label (Supply)	Total
Female (vs. Male)	-0.96 (0.15)	-0.54 (0.15)	-1.50 (0.21)	-0.19 (0.03)	0.01 (0.03)	-0.19 (0.04)	-0.88 (1.67)	-1.76 (1.64)	-2.63 (2.34)
Black (vs. White)	0.66 (0.21)	0.03 (0.18)	0.69 (0.28)	-0.11 (0.05)	-0.02 (0.04)	-0.13 (0.06)	3.09 (3.39)	0.14 (2.89)	3.24 (4.45)
Asian (vs. White)	-0.65 (0.26)	-0.16 (0.18)	-0.81 (0.31)	-0.16 (0.05)	-0.02 (0.04)	-0.18 (0.07)	1.21 (4.13)	-2.44 (2.89)	-1.23 (5.04)

Notes: This table separates demographic differences in LLM recommendations into differences associated with prompt content and effects of demographic labels, using the specification in (1). Each row reports one comparison: Female relative to Male, Black relative to White, or Asian relative to White. For each comparison, the sample excludes prompts that explicitly mention the relevant demographic characteristic, and labels are randomly assigned as described in Section 4.2. The three sets of columns report changes in diversified equity share, non-diversified asset share, and the net saving rate, defined as post-tax earnings minus consumption divided by post-tax earnings. All regressions are restricted to working-life observations. "Author" reports β_D, the coefficient on the author's actual demographic characteristic, holding the assigned label fixed; it captures differences associated with prompt content (the demand channel). "Label" reports β_S, the coefficient on the demographic label inserted into the prompt, holding content fixed; it captures the causal effect of the label on the LLM's response (the supply channel). "Total" reports $\beta_D + \beta_S$, the combined effect of a prompt being both authored by and labeled as belonging to the demographic group. Standard errors are reported in parentheses.

the recommended diversified equity share by 1.50pp in that period. Two-thirds of this combined effect is demand-driven: drawing a prompt written by a woman reduces the recommended diversified equity share by $\hat{\beta}_D = -0.96$pp relative to drawing a prompt written by a man, holding the randomized gender label fixed. The remaining one-third is supply-driven: holding prompt content fixed, a female label reduces the recommended diversified equity share by $\hat{\beta}_S = -0.54$pp. Therefore, both what respondents write and how the LLM responds to stated gender contribute to the gender gap in portfolio advice, with demand accounting for more of the difference.

Gender gap in non-diversified holdings is entirely demand-driven. We next estimate (1) with the change in non-diversified assets as the dependent variable, which includes individual stocks, crypto, gold, commodities, and collectibles, and accounts for approximately 1% of baseline allocations. In contrast to the gender gap in diversified holdings, the middle three columns in the top row of Table 3 show that the gender difference in non-diversified holdings is entirely demand-driven ($\hat{\beta}_D = -0.19$pp, $\hat{\beta}_S \approx 0$pp). Finally, in the last three columns, we estimate the same specification for the net saving rate. This gap is not statistically distinguishable from zero and is not the result of offsetting forces: both the demand coefficient ($\hat{\beta}_D = -0.88$pp) and the supply coefficient ($\hat{\beta}_S = -1.76$pp) are negative, though neither is statistically distinguishable from zero.

Racial differences are smaller and are demand-driven. Finally, we apply the same analysis to two comparisons based on race: Black versus White respondents, and Asian versus White respondents.[21] As with gender, we drop any prompts that explicitly mention the respondent's race, which respondents disclose far less often than gender, and we randomly insert "I am Black", "I am Asian", or "I am White" at the beginning of each prompt. Using these labeled prompts, we simulate life cycles as in the baseline and estimate two versions of (1): one comparing prompts written by Black and White respondents, and another comparing prompts written by Asian and White respondents.

The second and third rows of Table 3 show that the LLM's portfolio recommendations also differ by race: moving from a White-written, White-labeled prompt to a Black-written, Black-labeled prompt raises the recommended diversified equity share by 0.69pp, while the corresponding Asian–White difference is −0.81pp. Both differences are smaller than the gender difference of −1.50pp and are driven by demand. For the Black–White comparison, the demand effect is 0.66pp while the supply effect is approximately zero. For the Asian–White comparison, the supply effect is also small and statistically insignificant, though it is estimated less precisely than the gender benchmark. Differences in recommended non-diversified asset shares are similar in magnitude to the gender differences and are entirely demand-driven, whereas differences in saving rates are not statistically distinguishable from zero.

Interpretation. Distinguishing demand- versus supply-driven differences in LLM advice matters because the two channels have distinct interpretations. The demand-driven differences across gender and race reflect real differences in what these demographic groups choose to write about, which could be amplified or dampened by model design. The supply-driven difference for gender, in which the LLM gives different advice to identical prompts labeled as coming from a man versus a woman, may reflect implicit inference: the model may use gender as a signal for characteristics correlated with gender, such as risk preferences or labor-market risk (Croson and Gneezy 2009). But even when such correlations exist, gender is a coarse signal, and this type of inference would ideally be made explicit to users (Filippin and Crosetto 2016). Alternatively, supply-driven differences may reflect algorithmic bias, with the model reproducing gender stereotypes present in its training data.

The near-absence of a supply-side effect for race, in contrast to gender, could reflect how these models are trained: providers oftentimes evaluate and align their models to

[21] We do not have data on Hispanic ethnicity, so White, Black, and Asian are the only racial or ethnic groups in our survey with a sufficient number of prompts to perform the separate simulations by group.

avoid differential treatment on the basis of race, and these safeguards may be enforced more stringently for race than for gender. However, this interpretation comes with a caveat. An explicit statement, such as "I am Black", is precisely the kind of salient signal such safeguards are designed to detect, so the null supply effect does not rule out differential responses to subtler markers of race, such as names, dialect, or contextual cues, that are more common in actual usage.

5 Conclusion

Large language models have, within a few years, become a major source of personal financial advice that around half of respondents report using in both our survey and industry surveys (Lloyds Banking Group 2025; J.D. Power 2025). This paper develops a method to characterize the advice that these models provide and documents three facts. First, following LLM advice would move most respondents closer to the broad prescriptions of life cycle theory: they would hold larger savings buffers, be more likely to own diversified equity funds, and reduce their equity shares as they get older. Second, the advice departs from the theory's more demanding prescriptions: it implies unusually high patience, relies on simple saving and withdrawal heuristics, smooths consumption imperfectly after income shocks, and allows portfolios to drift passively with realized returns. Third, the advice varies systematically across groups, including by gender, financial literacy, and prior AI experience. These differences reflect both demand (what respondents write) and supply (how the model responds to otherwise identical prompts).

These findings suggest that generative AI could offer an affordable, widely accessible source of financial guidance and help overcome the costs, biases, and conflicts of interest associated with traditional human financial advisors (Reuter and Schoar 2024). The fact that LLM advice broadly aligns with life cycle theory is promising, especially because these general-purpose models are not explicitly optimized to improve household financial decisions. At the same time, our results show that the content of AI financial advice depends not only on model capabilities but also on what users ask. The departures from life cycle theory that we document provide concrete diagnostics to guide model development, while the demand-side constraints we identify are likely to limit the quality of the advice users receive even as models improve. More broadly, our methodology, which combines survey responses with simulations, provides a systematic way to track how AI financial advice changes as models evolve, as users learn to write different prompts, and as firms and regulators decide whether this advice should mirror, amplify, or mitigate heterogeneity in what individuals ask. Our results characterize the effect of following LLM advice,

not of receiving it: whether people act on these recommendations, and how any effects compare with those of human advisors, robo-advisors, and other sources of guidance, remain important open questions for future work.[22]

[22] Moss et al. (2026) provide early evidence: using data from an AI-powered investment adviser, they find that investors act on its recommendations, but that greater user intervention is associated with worse risk-adjusted performance.

References

Abdel Haq, Omar, Amitabh Chandra, Tomáš Jagelka, Erzo F. P. Luttmer, and Joshua Schwartzstein (2026), "Revealing life preferences through LLMs." Working Paper 35185, National Bureau of Economic Research.

Agnew, Julie, Pierluigi Balduzzi, and Annika Sundén (2003), "Portfolio choice and trading in a large 401(k) plan." *American Economic Review*, 93, 193–215.

Andersen, Steffen, John Y. Campbell, Kasper Meisner Nielsen, and Tarun Ramadorai (2020), "Sources of inaction in household finance: Evidence from the Danish mortgage market." *American Economic Review*, 110, 3184–3230.

Andries, Marianne, Maxime Bonelli, and David Sraer (2025), "Financial advisors and investors' bias." *The Review of Financial Studies*. Advance article, published October 27, 2025.

Benzoni, Luca, Pierre Collin-Dufresne, and Robert S. Goldstein (2007), "Portfolio choice over the life-cycle when the stock and labor markets are cointegrated." *The Journal of Finance*, 62, 2123–2167.

Bergstresser, Daniel and James Poterba (2004), "Asset allocation and asset location: Household evidence from the survey of consumer finances." *Journal of Public Economics*, 88, 1893–1915.

Beshears, John, James J. Choi, David Laibson, and Brigitte C. Madrian (2018), "Behavioral Household Finance." *Handbook of Behavioral Economics*, 1, 177–276.

Bessembinder, Hendrik (2018), "Do stocks outperform Treasury bills?" *Journal of Financial Economics*, 129, 440–457.

Bhattacharya, Utpal, Amit Kumar, Sujata Visaria, and Jing Zhao (2024), "Do women receive worse financial advice?" *The Journal of Finance*, 79, 3261–3307.

Bini, Pietro, Lin William Cong, Xing Huang, and Lawrence J. Jin (2026), "Behavioral economics of AI: LLM biases and corrections." Working Paper 34745, National Bureau of Economic Research.

Braxton, J. Carter, Kyle Herkenhoff, Jonathan Rothbaum, and Lawrence D. W. Schmidt (2025), "Changing income risk across the US skill distribution: Evidence from a generalized Kalman filter." *American Economic Review*, 115, 4438–4475.

Bucher-Koenen, Tabea, Andreas Hackethal, Johannes Koenen, and Christine Laudenbach (2025), "Gender differences in financial advice." *American Economic Review*, 115, 4218–4252.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini (2009), "Fight or flight? Portfolio rebalancing by individual investors." *The Quarterly Journal of Economics*, 124, 301–348.

Campbell, John Y., Nuno Clara, and João F. Cocco (2021), "Structuring Mortgages for Macroeconomic Stability." *The Journal of Finance*, 76, 2525–2576.

Campbell, John Y. and Tarun Ramadorai (2026), "Household finance in retrospect and prospect." Working Paper 34621, National Bureau of Economic Research. Journal of Finance: Insights and Perspectives, forthcoming.

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

Choi, James J. (2022), "Popular personal financial advice versus the professors." *Journal of Economic Perspectives*, 36, 167–192.

Choi, James J., David Laibson, and Brigitte C. Madrian (2011), "$100 bills on the sidewalk: Suboptimal investment in 401(k) plans." *Review of Economics and Statistics*, 93, 748–763.

Choukhmane, Taha (2025), "Default options and retirement saving dynamics." *American Economic Review*, 115, 3749–3787.

Choukhmane, Taha and Tim de Silva (2026), "What drives investors' portfolio choices? separating risk preferences from frictions." *The Journal of Finance*, 81, 5–48.

Choukhmane, Taha, Lucas Goodman, and Cormac O'Dea (2025), "Efficiency in household decision-making: Evidence from the retirement savings of U.S. couples." *American Economic Review*, 115, 1485–1519.

Cocco, João F., Francisco J. Gomes, and Pascal J. Maenhout (2005), "Consumption and portfolio choice over the life cycle." *The Review of Financial Studies*, 18, 491–533.

Cook, Thomas R., Sophia Kazinnik, Zach Modig, and Nathan M. Palmer (2026), "What do LLMs want?" Finance and Economics Discussion Series 2026-006, Board of Governors of the Federal Reserve System.

Cooley, Philip L., Carl M. Hubbard, and Daniel T. Walz (1998), "Retirement savings: Choosing a withdrawal rate that is sustainable." *AAII Journal*, 20, 16–21. Available via AAII.

Croson, Rachel and Uri Gneezy (2009), "Gender differences in preferences." *Journal of Economic Literature*, 47, 448–474.

D'Acunto, Francesco, Nagpurnanand Prabhala, and Alberto G. Rossi (2019), "The Promises and Pitfalls of Robo-Advising." *The Review of Financial Studies*, 32, 1983–2020.

de Silva, Tim (2025), "Insurance versus moral hazard in income-contingent student loan repayment." *The Quarterly Journal of Economics*, 140, 2851–2905.

Duarte, Victor, Julia Fonseca, Aaron S. Goodman, and Jonathan A. Parker (2025), "Simple allocation rules and optimal portfolio choice over the lifecycle." Working Paper 29559, National Bureau of Economic Research. Originally issued December 2021.

Egan, Mark, Gregor Matvos, and Amit Seru (2019), "The market for financial adviser misconduct." *Journal of Political Economy*, 127, 233–295.

Fedyk, Anastassia, Ali Kakhbod, Peiyao Li, and Ulrike Malmendier (2026), "AI and perception biases in investments: An experimental study." *Journal of Financial Economics*. Accepted.

Fieberg, Christian, Lars Hornuf, Maximilian Meiler, and David J. Streich (2025), "Using large language models for financial advice." CESifo Working Paper 11666, CESifo.

Filippin, Antonio and Paolo Crosetto (2016), "A reconsideration of gender differences in risk attitudes." *Management Science*, 62, 3138–3160.

Foltyn, Richard and Jonna Olsson (2026), "The worth of a "wo": Gender bias in financial advice from LLMs." Discussion Paper 21323, Centre for Economic Policy Research.

Gallup (2025), "Americans still turn to people for financial advice." Gallup News. Gallup Panel survey of U.S. adults.

Ganong, Peter and Pascal Noel (2019), "Consumer spending during unemployment: Positive and normative implications." *American Economic Review*, 109, 2383–2424.

Gomes, Francisco and Alexander Michaelides (2005), "Optimal life-cycle asset allocation: Understanding the empirical evidence." *The Journal of Finance*, 60, 869–904.

Gomes, Francisco J. (2020), "Portfolio Choice over the Life Cycle: A Survey." *Annual Review of Financial Economics*, 12, 277–304.

Gomes, Francisco J., Michael Haliassos, and Tarun Ramadorai (2021), "Household Finance." *Journal of Economic Literature*, 59, 919–1000.

Google DeepMind (2025), "Gemini 3 Flash." https://blog.google/products-and-platforms/products/gemini/gemini-3-flash/.

Gourinchas, Pierre-Olivier and Jonathan A. Parker (2002), "Consumption over the life cycle." *Econometrica*, 70, 47–89.

Greig, Fiona, Tarun Ramadorai, Alberto G. Rossi, Stephen P. Utkus, and Ansgar Walther (2025), "Human financial advice in the age of automation." Working paper, SSRN. Revise and resubmit at the Journal of Finance; SSRN version last revised July 7, 2025.

Gruber, Jonathan (1997), "The Consumption Smoothing Benefits of Unemployment Insurance." *American Economic Review*, 87, 192–205.

Guren, Adam M., Arvind Krishnamurthy, and Timothy J. McQuade (2021), "Mortgage Design in an Equilibrium Model of the Housing Market." *The Journal of Finance*, 76, 113–168.

Hao, Yuzhi and Danyang Xie (2025), "A multi-LLM-agent-based framework for economic and public policy analysis." *arXiv preprint arXiv:2502.16879*.

Horton, John J., Apostolos Filippas, and Benjamin S. Manning (2023), "Large language models as simulated economic agents: What can we learn from Homo Silicus?" Working Paper 31122, National Bureau of Economic Research. Revised February 2026.

J.D. Power (2025), "As more U.S. consumers struggle with rising prices, many turn to artificial intelligence for financial advice." Banking and Payments Intelligence Report - Press Release. Published 28 August 2025; based on survey of approximately 4,000 U.S. consumers.

Krusell, Per and Anthony A. Smith, Jr. (1998), "Income and Wealth Heterogeneity in the Macroeconomy." *Journal of Political Economy*, 106, 867–896.

Linnainmaa, Juhani T., Brian T. Melzer, and Alessandro Previtero (2021), "The misguided beliefs of financial advisors." *The Journal of Finance*, 76, 587–621.

Lloyds Banking Group (2025), "Over 28 million adults now using AI tools to help manage their money." Press Release. Published 03 November 2025.

Lusardi, Annamaria, Pierre Carl Michaud, and Olivia S. Mitchell (2017), "Optimal financial knowledge and wealth inequality." *Journal of Political Economy*, 125, 431–477.

Lusardi, Annamaria and Olivia S. Mitchell (2023), "The importance of financial literacy: Opening a new field." *Journal of Economic Perspectives*, 37, 137–154.

Merton, Robert C. (1969), "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case." *Review of Economics and Statistics*, 51, 247–257.

Mo, Hongwei and Shumiao Ouyang (2025), "(generative) AI in financial economics." *Journal of Chinese Economic and Business Studies*, 23, 509–587.

Moss, Austin, Jackie Wegner, and Sarah L. C. Zechman (2026), "AI meets DIY: The impact of human intervention on AI-assisted investing." *SSRN Electronic Journal*.

Mullainathan, Sendhil, Markus Noeth, and Antoinette Schoar (2012), "The market for financial advice: An audit study." NBER Working Paper 17929, National Bureau of Economic Research.

OpenAI (2025), "Introducing GPT-5.2." https://openai.com/index/introducing-gpt-5-2/. OpenAI Blog.

OpenAI (2026), "GPT-5.6: Frontier intelligence that scales with your ambition." https://openai.com/index/gpt-5-6/. OpenAI Blog.

Ouyang, Shumiao, Hayong Yun, and Xingjian Zheng (2025), "AI as decision-maker: Ethics and risk preferences of LLMs." Working paper, SSRN; last revised December 8, 2025.

Reher, Michael and Stanislav Sokolinski (2024), "Robo advisors and access to wealth management." *Journal of Financial Economics*, 155, 103829.

Reuter, Jonathan and Antoinette Schoar (2024), "Demand-side and supply-side constraints in the market for financial advice." *Annual Review of Financial Economics*, 16, 391–411.

Ross, Jillian, Yoon Kim, and Andrew W. Lo (2024), "LLM economicus? mapping the behavioral biases of LLMs via utility theory." *arXiv preprint arXiv:2408.02784*.

Rossi, Alberto G. and Stephen P. Utkus (2020), "Who Benefits from Robo-advising? Evidence from Machine Learning." *SSRN Electronic Journal*. Last revised February 1, 2021.

Rumpf, Matthias, Michael Haliassos, Tetyana Kosyakova, and Thomas Otter (2026), "From humans to algorithms: How financial advice differs across professionals, peers, and LLMs." Working paper, SSRN.

Takayanagi, Takehiro, Kiyoshi Izumi, Javier Sanz-Cruzado, Richard McCreadie, and Iadh Ounis (2025), "Are generative AI agents effective personalized financial advisors?" In *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 286–295.

Verstyuk, Sergiy and Michael R. Douglas (2025), "Consumption and savings with large language model agents." Working paper, Harvard University; SSRN version posted January 8, 2026 and last revised January 16, 2026.

Internet Appendix to "AI Financial Advice: Supply, Demand, and Life Cycle Implications"

For Online Publication Only

Contents

A Prompt Construction and Design **A.2**

A.1 Survey Prompt Variable Insertion A.2

A.2 Prompt Selection A.4

A.3 Prompt Bucketing A.4

A.4 LLM Advice Pipeline A.5

A.5 LLM Prompts A.8

A.5.1 Academic Prompt A.8

A.5.2 Translation Prompt A.11

B Life Cycle Model Details **A.13**

B.1 Demographics and Preferences A.13

B.2 Labor Market A.13

B.3 Savings Account A.14

B.4 Government A.15

B.5 Summary of Investors' Problem A.15

B.6 Calibration A.16

C Textual Analysis Details **A.18**

D Additional Details on Simulated Method of Moments Estimation **A.23**

E Additional Tables and Figures **A.24**

Appendix A. Prompt Construction and Design

This appendix describes how the respondent-written prompts introduced in Section 1.1 are processed and integrated into the life cycle simulation framework. The survey questions are shown in Figure E1 and sample demographics are in Table E1.

A.1 Survey Prompt Variable Insertion

Whenever an individual prompt mentions a state variable from the life cycle model (income, age, wealth, tenure, or asset allocation), we replace that mention with a placeholder that is updated to the simulated individual's state. Table A1 lists representative replacement patterns for each state variable. The rest of this section describes the rules governing these replacements.

Household rules. Our model simulates outcomes at the individual level. When a prompt instead describes income or wealth at the household level, we scale up the simulated amount before inserting it: by default we double the income and wealth that are inserted unless the prompt mentions more than two working adults, in which case those numbers are scaled by the number of working adults. When a prompt provides information at the individual level, we use it directly. If the prompt mentions the income of another household member, we leave that income unchanged and insert household income as the sum of that income and the simulated individual income. If the prompt specifies each member's balance in an asset class, we assume holdings in that class are proportional to the state balances, so household holdings equal the simulated individual holdings divided by the respondent's share of total balance.

Income values. We replace each income amount with its simulated counterpart, using gross or after-tax income depending on the tax treatment specified in the prompt. If the prompt does not specify the tax treatment, we insert the after-tax amount.[1] If the prompt reports an hourly wage and a number of hours worked per week, denoted by N, we compute hourly income as the simulated income divided by $52 \times N$, and insert that amount. If hours or income are reported as a range, we use the midpoint.[2] If hours are not reported, we set $N = 40$ if the job is full-time, or $N = 20$ if the job is part-time. When income comes from multiple sources or is reported at multiple frequencies (e.g., monthly and hourly), we annualize each amount, preserve the implied ratios across sources, and rescale to the original frequency.

Wealth values. Many prompts describe financial wealth as dollar amounts held in several accounts. If the prompt mentions a single account with an ambiguous label such as "savings," we interpret that amount as total wealth. If a single account clearly corresponds to one asset class, we interpret the amount as wealth in that class. For prompts with multiple accounts, let $\{D_n\}_{n \in N_D}$, $\{I_n\}_{n \in N_I}$,

[1] Phrases such as "I bring home $X" or "my monthly salary is $X" can be read as either gross or net income, by both the respondents and the LLM.

[2] When replacing a range with a single value would be ungrammatical, we replace both bounds with the same value.

$\{S_n\}_{n\in N_S}$, and $\{O_n\}_{n\in N_O}$ denote dollar amounts in diversified stock, individual stock, non-stock, and other risky assets. We classify holdings as diversified when the prompt mentions exchange-traded funds, index funds, or tickers of assets in these categories, and as individual stocks when the prompt refers to specific stocks or a small bundle of stocks chosen by the writer. When the prompt refers to stock market investments without distinguishing between the two, we treat them all as diversified and assume no holdings of individual stocks. Instead of mentioning asset classes directly, individuals often mention holdings in accounts that can contain multiple asset classes, such as retirement or brokerage accounts. We refer to these accounts as "mixed accounts", denoting their holdings by $\{M_n\}_{n\in N_M}$. With these mixed accounts, total wealth is defined as $W \equiv \sum_{n\in N_D} D_n + \sum_{n\in N_I} I_n + \sum_{n\in N_S} S_n + \sum_{n\in N_O} O_n + \sum_{n\in N_M} M_n$. We also denote θ_x as the share of wealth in asset class x.

We now describe how we replace the dollar holdings in the different asset classes, distinguishing between two cases with and without mixed accounts. In the notation below, variables with a tilde (e.g., $\tilde{W}$) refer to the values in our simulation, while variables without a tilde (e.g., W) refer to values mentioned in prompts.

- **Scenario 1: Well-defined asset classes.** If the prompt mentions only asset classes that fall within our four asset classes, then the wealth in each asset is distributed to all of the prompt-mentioned asset classes in ratios that match the original within-class ratios:

$$\tilde{X}_n = \tilde{\theta}_x \frac{X_n}{\sum_{n\in N_X} X_n} \tilde{W}, \qquad X \in \{D, I, S, O\}. \tag{2}$$

  Example: Suppose that the prompt mentions that the original prompt writer has \$1,200 in a checking account and \$1,800 in a savings account. If the simulated agent that draws this prompt has \$10,000 in the non-stock asset class, then the prompt will state that there is \$4,000 in a checking account and \$6,000 in a savings account.

- **Scenario 2: Mixed accounts.** If the prompt mentions an asset that may contain multiple asset classes (e.g., a 401k or mutual fund), we first allocate all single-asset class holdings according to (2), and then allocate whatever balance remains across the mixed accounts to match the original ratios:

$$\tilde{X}_n = \min\left(\theta_x, \tilde{\theta}_x\right) \frac{X_n}{\sum_{n\in N_X} X_n} \tilde{W}, \qquad X \in \{D, I, S, O\}, \tag{3}$$

$$\tilde{M}_n = \left(\tilde{W} - \sum_{n\in N_D} \tilde{D}_n - \sum_{n\in N_I} \tilde{I}_n - \sum_{n\in N_S} \tilde{S}_n - \sum_{n\in N_O} \tilde{O}_n\right) \frac{M_n}{\sum_{n\in N_M} M_n}. \tag{4}$$

  Example 1: Suppose that the prompt mentions that the original prompt writer has \$1,200 in a checking account and \$1,800 in a 401(k). If the simulated agent has \$5,000 in the non-stock asset and \$5,000 in all other asset classes combined, then the prompt will state the following

values:

$$\text{Checking Account} = \min\left(\frac{\$5,000}{\$10,000}, \frac{\$1,200}{\$1,200 + \$1,800}\right) \cdot \$10,000 = \$4,000,$$
$$\text{401(k) Account} = \$10,000 - \text{Checking Account} = \$6,000.$$

Example 2: Suppose that the prompt mentions that the original prompt writer has $1,800 in a checking account and $1,200 in a 401(k). If the simulated agent has $5,000 in the non-stock asset and $5,000 in all other asset classes combined, then the prompt will state the following values:

$$\text{Checking Account} = \min\left(\frac{\$5,000}{\$10,000}, \frac{\$1,800}{\$1,800 + \$1,200}\right) \cdot \$10,000 = \$5,000,$$
$$\text{401(k) Account} = \$10,000 - \text{Checking Account} = \$5,000.$$

In this example, the minimum in (3) binds: the simulated individual cannot have a 60% allocation in checking accounts because that would exceed the $5,000 held in the non-stock asset class.

One-time lump sum payments. When a prompt refers to a one-time lump-sum payment (e.g., an inheritance, settlement, or house sale), we replace the stated dollar value only if the payment has already occurred; expected future payments are left unchanged. We replace the value by first computing its share of the respondent's self-reported financial wealth, then applying that same share to simulated wealth.[3] For example, if a respondent reports having received a $10,000 inheritance in their prompt, and separately reports total financial wealth of $100,000, we replace the stated $10,000 value with 10% of total simulated wealth.

A.2 Prompt Selection

As described in Section 1.1, out of 1,000 complete Qualtrics responses, we removed 33 that failed Prolific's platform-level authenticity checks and 15 that failed a minimum specificity requirement, leaving 952 responses. To pass the minimum specificity check, each of the respondent's three prompts had to address its assigned topic: the first the respondent's situation, the second general spending or saving, and the third investing.

A.3 Prompt Bucketing

We divide prompts into 12 buckets defined by employment status, age, and income. We first group prompts by the respondent's employment status (employed, unemployed, or retired). Within each employment group, we then split by age: into halves for the unemployed, into terciles for the

[3] Since self-reported financial wealth is categorical, we use the midpoint of each range, and $50,000 when the respondent selected "I don't know/prefer not to answer."

employed, and into a single bucket for the retired. For the employed, we further split each age tercile into income terciles, yielding 2 (unemployed) + 3×3 (employed) + 1 (retired) = 12 buckets. Figure A1 illustrates the bucketing procedure, and Table A2 reports the distribution of prompt sets across buckets.[4] Because the survey elicits income in discrete brackets, ties are common and can make tercile divisions unbalanced. We assign tied respondents to the lower income tier, except when so many respondents select the highest bracket that this rule would leave the top tercile too small, in which case we assign ties to the higher tier instead.

A.4 LLM Advice Pipeline

We convert LLM responses into consumption and portfolio decisions in two steps. In the first step, we (i) assign each simulated agent-age pair to a prompt bucket based on employment status, age, and income, (ii) draw a prompt set at random from that bucket, (iii) construct the prepared message, and (iv) send it to the LLM, saving its response.[5] In the second step, we pass the response back to a second LLM along with deterministic instructions for converting the qualitative advice into quantitative choices (in JSON format). These quantitative outputs are then used to update the endogenous state variables in our simulation.

[4] In our heterogeneity analysis (e.g., using only prompt sets written by women), we repeat the sorting and bucketing procedure separately for each subgroup.

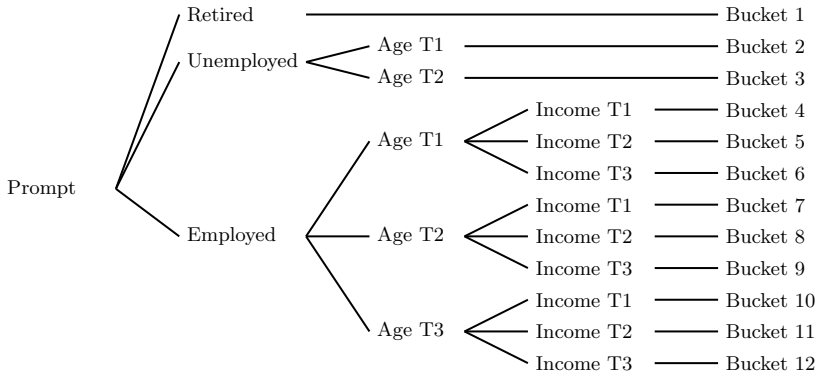
[5] We instruct the LLM to provide responses in at most 200 words.

Table A1. Mapping between State Variables and Respondent-Written Prompts

Original text	Replacement text
	Age
"X years old"	"{age} years old"
"in my Xs"	"in my {math.floor(age/10)*10}s"
"I retire in X years"	"I retire in {65 - age} years"
	Income
"$X a year before taxes"	"${pretaxincome:,.0f} a year before taxes"
"$X a month before taxes"	"${pretaxincome/12:,.0f} a month before taxes"
"$X biweekly before taxes"	"${pretaxincome/26:,.0f} biweekly before taxes"
"$X a week before taxes"	"${pretaxincome/52:,.0f} a week before taxes"
"$X a day before taxes"	"${pretaxincome/(52*5):,.0f} a day before taxes"
"$X an hour before taxes"	"${pretaxincome/(52*40):,.2f} an hour before taxes"
"$X a year after taxes"	"${income:,.0f} a year after taxes"
"$X a month after taxes"	"${income/12:,.0f} a month after taxes"
"$X biweekly after taxes"	"${income/26:,.0f} biweekly after taxes"
"$X a week after taxes"	"${income/52:,.0f} a week after taxes"
"$X a day after taxes"	"${income/(52*5):,.0f} a day after taxes"
"$X an hour after taxes"	"${income/(52*40):,.2f} an hour after taxes"
	Stock Allocation
"X% of wealth in savings"	"{100 - current_stock_allocation}% of wealth in savings"
"X% of wealth in stock"	"{current_stock_allocation}% of wealth in stock"
	Miscellaneous
"X years into a job"	"{tenure} years into a job"
"past annual income of $X"	"past annual income of ${avgpastincome:,.0f}"

Notes: This table summarizes text replacement used to map respondent-written prompts to simulated state variables. The left column shows examples of original text in the survey prompts; the right column shows the corresponding replacement text used in the LLM prompt, where placeholders are filled with the simulated individual's age, income, portfolio allocation, or other state variables.

Figure A1. Prompt Bucketing Tiers



Notes: This figure outlines the tiered system used to assign survey prompts to buckets. Prompts are bucketed by respondent employment status, age tercile, and, for employed respondents, income tercile, yielding 12 buckets in total. The age and income cutoffs are heuristic thresholds based on the survey sample. In each period, the simulated individual's employment status, age, and income determine the relevant bucket, and a prompt is randomly drawn from that bucket.

Table A2. Prompt Bucket Distribution by Heterogeneity Cut

Prompt Distribution by Bucket and Cut											
Employment	Age	Income	Bucket	Baseline	Financial Literacy		AI Usage		Gender		No Explicit
Status	Bucket	Bucket	Number		Below-perfect	Perfect	No	Yes	Male	Female	Gender
Retired			1	132	68	64	97	35	57	75	97
Unemployed	1		2	71	55	16	42	32	30	42	63
	2		3	68	54	14	35	30	27	40	57
Employed	1	1	4	117	90	57	60	61	58	63	105
	1	2	5	60	31	19	29	29	38	25	51
	1	3	6	56	24	14	20	33	32	25	46
	2	1	7	84	59	51	53	69	38	44	69
	2	2	8	94	37	14	25	23	45	45	72
	2	3	9	49	48	25	36	25	25	23	39
	3	1	10	101	65	31	53	40	40	59	88
	3	2	11	77	25	29	16	50	48	16	67
	3	3	12	43	42	20	31	28	28	29	35
			Total	952	598	354	497	455	466	486	789

Notes: This table reports the number of respondent-level prompt sets assigned to each prompt bucket. The baseline column uses the full survey sample and the bucket structure shown in Figure A1. The remaining columns report the same distribution separately by financial literacy, prior AI experience, and gender. The financial-literacy columns compare below-perfect (0–4 of 5 Big Five questions correct) with perfect (5 correct). Within each group, age and income cutoffs are computed separately within the relevant prompt-set sample. The final column reports the subset of prompt sets that do not explicitly mention gender. Go back.

A.5 LLM Prompts

This subsection presents the full text of the two researcher-written prompts used in our analysis. The academic prompt (Section 1.4) is used to elicit financial advice from `GPT-5.2`, while the translation prompt is used to convert the LLM's textual advice into quantitative choices via `GPT-5 Mini`, as described in Step 4 of Section 1.3.

A.5.1 Academic Prompt

The following is the full text of the academic prompt used to elicit financial advice from `GPT-5.2`, as described in Section 1.4. Individual-specific state variables (age, employment status, income, account balances) are inserted into the User section of the prompt each period.

PROMPT EXAMPLE

SYSTEM

You are a U.S. financial advisor acting in the best interest of your clients. You are academically trained in household finance, modern portfolio theory, and life cycle planning. Your objective is to produce high-quality personalized financial advice tailored to my circumstances under these baseline assumptions:

- Normal life expectancy, living expenditures, retirement age, employment risk, and income risk
- Current U.S. tax law and Social Security rules stay unchanged
- Risk-free savings earn 2.0% real return annually
- Real stock returns match the 60-year U.S. total stock market historical average
- I am single with no dependents and have no bequest motive. I do not care about wealth after death

Given my **age, employment status, annual unemployment benefits, average annual post-tax income (since age 22), taxable account balances, and total net wealth**, give the best financial advice for allocating my resources across annual spending and the following four taxable accounts:

- D=diversified stock (index funds/ETFs/mutual funds; "S&P 500/total market")
- I=individual stocks (single-company stocks or a small bundle (<5))
- N=nonstock (cash/savings/HYSA/money market/CDs/bonds/treasuries/buffer/emergency fund)
- O=other (anything not in D, I, N; crypto/gold/commodities/collectibles)

Based on my information and these assumptions, think step by step about your best advice for addressing the following questions (but do not output your reasoning):

1. How much should I consume this year in dollar amounts taking into account all my living expenses over the year?
2. Should I contribute to, withdraw from, or keep unchanged each of my 4 taxable accounts?
3. Should I transfer money between each of my four accounts?

INTERNAL COMPUTATION STRUCTURE (do not output):

INPUTS: sN,sD,sI,sO,total_wealth,income_post_tax,avgpastincome,tenure.

VARIABLES (annual):
Contribution (cN,cD,cI,cO); withdrawals (wN,wD,wI,wO); transfers tXY for X,Y in (N,D,I,O), X≠Y.
C = cN+cD+cI+cO
W = wN+wD+wI+wO

FUNDING
Contributions (cN,cD,cI,cO) use income_post_tax
Withdrawals (wN, wD, wI, wO) fund spending; reduce holdings
Transfers (tXY) move existing holdings between buckets

SOURCE LIMITS
wN+tND+tNI+tNO<=sN
wD+tDN+tDI+tDO<=sD
wI+tIN+tID+tIO<=sI
wO+tON+tOD+tOI<=sO

BUDGET IDENTITY

consume_amt = income_post_tax + W - C
Constraints: consume_amt>=0; C<=income_post_tax+W

OUTPUT INSTRUCTIONS

Provide this advice ONLY as JSON with double quotes with the format described below without other explanations. All values must be numeric (no dollar signs, no commas, no text). If no action is taken in a category, return 0.

Before producing the JSON, verify internally that:

- The budget identity and all source constraints are satisfied
- All contribution, withdrawal, and transfer values must be >= 0.

If any constraint is violated, correct the values before output.

Prompt Example

System (continued)

OUTPUT JSON KEYS (exactly these, in this order): {
"consume_amt": <dollar amount to spend this year; must satisfy the budget identity and be >=0>,
"nonstock_contrib": <dollar amount contributed to NONSTOCK using income>,
"nonstock_withdraw": <dollar amount withdrawn from NONSTOCK to spend>,
"divstock_contrib": <dollar amount contributed to DIVERSIFIED STOCK using income>,
"divstock_withdraw": <dollar amount withdrawn from DIVERSIFIED STOCK to spend>,
"indstock_contrib": <dollar amount contributed to INDIVIDUAL STOCK using income>,
"indstock_withdraw": <dollar amount withdrawn from INDIVIDUAL STOCK to spend>,
"other_contrib": <dollar amount contributed to OTHER using income>,
"other_withdraw": <dollar amount withdrawn from OTHER to spend>,
"transfer_nonstock_to_divstock": <dollar amount transferred from NONSTOCK to DIVERSIFIED STOCK>,
"transfer_divstock_to_nonstock": <dollar amount transferred from DIVERSIFIED STOCK to NONSTOCK>,
"transfer_nonstock_to_indstock": <dollar amount transferred from NONSTOCK to INDIVIDUAL STOCK>,
"transfer_indstock_to_nonstock": <dollar amount transferred from INDIVIDUAL STOCK to NONSTOCK>,
"transfer_divstock_to_indstock": <dollar amount transferred from DIVERSIFIED STOCK to INDIVIDUAL STOCK>,
"transfer_indstock_to_divstock": <dollar amount transferred from INDIVIDUAL STOCK to DIVERSIFIED STOCK>,
"transfer_nonstock_to_other": <dollar amount transferred from NONSTOCK to OTHER>,
"transfer_other_to_nonstock": <dollar amount transferred from OTHER to NONSTOCK>,
"transfer_divstock_to_other": <dollar amount transferred from DIVERSIFIED STOCK to OTHER>,
"transfer_other_to_divstock": <dollar amount transferred from OTHER to DIVERSIFIED STOCK>,
"transfer_indstock_to_other": <dollar amount transferred from INDIVIDUAL STOCK to OTHER>,
"transfer_other_to_indstock": <dollar amount transferred from OTHER to INDIVIDUAL STOCK>
}

User

HERE IS MY INFORMATION:

- age: 39 years old
- I have been working at the same firm for 2 years
- My annual take-home pay income (after taxes) is $43,188 this year
- My average annual income since age 22 is $16,329. Going forward, my average income will determine my social security benefits in retirement.
- Amount held in taxable diversified stock assets (mutual funds, stock ETFs, stock index funds, diversified brokerage holdings): $13,536
- Amount held in taxable individual stocks (singular stocks or a small bundle (<5)): $0
- Amount held in taxable non-stock assets (cash, checking, savings, HYSA, money market, CDs, bonds, Treasuries, emergency fund): $15,208
- Amount held in other assets (e.g., crypto, gold, commodities, collectibles): $0

Respond with JSON only. No prose.

Table A3. Academic Prompt: Assumptions and Individual-Specific Statements

Panel A: Baseline Assumptions	
Academic Prompt Assumption	**Life Cycle Model Implementation**
"Normal life expectancy, living expenditures, retirement age, employment risk, and income risk"	Mortality risk calibrated to SSA life tables; exogenous retirement at age 65; persistent income risk and employment transitions calibrated to SIPP data
"Current U.S. tax law and Social Security rules stay unchanged"	2025 Federal income tax schedule (single filer) and Social Security benefit formula
"Risk-free savings earn 2.0% real return annually"	2% annual risk-free return
"Real stock returns match the 60-year U.S. total stock market historical average"	Log-normal returns calibrated to the 1925–2006 CRSP value-weighted index: 6.4% equity premium, 20% standard deviation. Thus, the model's historical window is longer than the 60-year window stated in the prompt
"I am single with no dependents and have no bequest motive. I do not care about wealth after death"	One individual per household and zero utility from bequests
Panel B: Individual-Specific Statements	
Category	**LLM Prompt Statement**
Case 1: Employed	
Employment	"I have started a new job this year" / "I have been working at the same firm for {tenure} years"
Current Income	"My annual take-home pay income (after taxes) is $X this year"
Past Income	"My average annual income since age 22 is $Y. Going forward, my average income will determine my social security benefits in retirement"
Case 2: Unemployed	
Employment	"I am currently unemployed"
Current Income	"My annual unemployment benefit income (after taxes) is $X this year"
Case 3: Retired (Age ≥ 65)	
Employment	"I am retired"
Current Income	"My annual Social Security benefit income (after taxes) is $X this year"

Notes: This table summarizes the construction of the academic prompt benchmark. Panel A maps the baseline assumptions stated in the prompt to their implementation in the life cycle model. Panel B lists the individual-specific statements included in the prompt each period. Go back.

A.5.2 Translation Prompt

The following is the full text of the translation prompt used to convert textual LLM advice into quantitative choices via `GPT-5 Mini`, as described in Step 4 of Section 1.3.

PROMPT EXAMPLE

USER

You are a deterministic extractor. Convert `ADVICE_TEXT` into ONE-YEAR USD flows between annual spending and 4 saving buckets:

BUCKETS:

- `consume_amt` = annual spending (rent, groceries, travel, medical bills, subscriptions, debt payments, expenses)
- D = diversified stock (index funds/ETFs/mutual funds; "S&P 500/total market")
- I = individual stocks (single-company stocks or a small bundle, < 5)
- N = nonstock (cash/savings/HYSA/money market/CDs/bonds/Treasuries/buffer/emergency fund)
- O = other (anything not in D, I, N; crypto/metals/commodities/collectibles)

INPUTS (truth): `sN`, `sD`, `sI`, `sO`, `total_wealth`, `income_post_tax`. Use INPUTS as truth even if `ADVICE_TEXT` mentions different income/wealth.

VARIABLES (annual):
Contributions (`cN`, `cD`, `cI`, `cO`); withdrawals (`wN`, `wD`, `wI`, `wO`); transfers `tXY` for $X, Y \in \{N, D, I, O\}$, $X \neq Y$.
C=`cN+cD+cI+cO`
W=`wN+wD+wI+wO`

FUNDING

- Contributions (`cN`, `cD`, `cI`, `cO`) use `income_post_tax`.
- Withdrawals (`wN`, `wD`, `wI`, `wO`) fund spending; reduce holdings.
- Transfers (`tXY`) move existing holdings ONLY if text explicitly says move/transfer/rebalance/sell/buy OR invest "savings/-portfolio/windfall/remaining capital."
- "Open Roth IRA/brokerage" is not a bucket; bucket comes from what is bought/held.

ACTIONABLE-NUMBER FILTER:
Use a number only if tied to action verbs: spend/budget/save/invest/contribute/deposit/withdraw/transfer/move/rebalance/buy/sell/pay. Ignore yields/returns and time horizons.
Itemized budgets (housing $X, food $Y, ...): sum all spending line items → `consume_amt`. "Left over"/"remainder"/"surplus" → contributions. Debt payments (credit card, student loan, car loan) are spending, not contributions.

ANNUALIZE
Ranges → midpoint:

- $A/year or "annual" → A
- $m/month or "monthly" → $12m$
- $w/week or "weekly" → $52w$
- $q/quarter or "quarterly" → $4q$
- $d/day or "daily" → $365d$
- $p/paycheck: biweekly ×26, semi-monthly ×24, weekly ×52, else assume ×26

LIMITED DURATION RULE (one-year conversion):
If text says "for N months/weeks/paychecks":

- If N < annual count (months < 12, weeks < 52), total = rate × N.
- If N ≥ annual count, total = annualized full-year rate.

MULTI-YEAR GOAL RULE:
Convert H to months (years*12). year1_total = round_half_up(T * min(12, H_months) / H_months).

PERCENTAGE BASE RULES
% rules: if mentions portfolio/wealth/assets OR withdrawal rate, => `round(pct * total_wealth)`; otherwise, use `round(pct* income_post_tax)`.

- "Save/invest X%" → X% of `income_post_tax` (unless portfolio/wealth is referenced).
- "Withdraw X%" or "4% rule" → X% of `total_wealth`.
- "X% in stocks" or "X/Y split" → X% of `total_wealth` (allocation target, implies transfers).

STOCK-SPLIT DEFAULT
When advice says "stocks/equities" without D versus I and does not change the split:
D_share = sD/(sD+sI) if sD+sI>0 else 1; I_share=1-D_share.
Any stock_total → D=round(D_share*stock_total), I=stock_total-D.
No D↔I transfers unless explicitly instructed.

COMMON PATTERN RULES

- "Emergency fund of X months": targetN = X * (income_post_tax/12).If `sN` < `targetN`, prioritize N contributions until target met.
- If "emergency fund before investing" with $ target and recurring amount: A=recurring annual amount, `cN` = min(A, max(0,`targetN`-`sN`)).
- "50/30/20 rule" or similar: needs+wants = spending (`consume_amt`); savings portion = total contributions to N/D/I/O.
- "4% rule"/"X% withdrawal rate": annual withdrawal = rate × `total_wealth`.
- RMDs (Required Minimum Distributions) → withdrawal from total wealth, not a transfer.
- Target date fund without an explicit allocation should be: 90% in D + 10% in N if age < 40 // $(170 - 2 \times \text{age})$% in D and the rest in N if age ≥ 40 and age ≤ 70 // 30% in D + 70% in N if age > 70

EXTRACTION + CONFLICTS (choose one directive per field):
Priority: annual $ > lump-sum $ > monthly $ > weekly $ > paycheck $ > %.
Tie-break: prefer "Recommendation/Summary/Action section," else last mention. Map "stocks" to (D+I) and "safer/cash/bonds" to N unless stated.

SOURCE LIMITS `wN+tND+tNI+tNO` ≤ `sN`;
`wD+tDN+tDI+tDO` ≤ `sD`;
`wI+tIN+tID+tIO` ≤ `sI`;
`wO+tON+tOD+tOI` ≤ `sO`.

BUDGET IDENTITY `consume_amt` = `income_post_tax`+W-C.
Constraints: `consume_amt` ≥ 0; C ≤ `income_post_tax`+W.

SPENDING RULE (`repair_ind`):
- If explicit total spending (amount, % of income, or summed itemized budget): set `consume_amt` to that; set `repair_ind=0`.
- Else if any actionable saving directive exists: derive `consume_amt` via the budget identity; set `repair_ind=0`.
- Else if no mention of spending or saving amount, set spending `consume_amt=income_post_tax`; all flows = 0; `repair_ind=1`.

REPAIR (deterministic):
A) If any source limit is violated: reduce that asset's outgoing transfers first (output-key order), then its withdrawal.
B) If $C >$ `income_post_tax` $+ W$: scale contributions proportionally to fit (round N, then D, then I; O absorbs remainder).
Recompute `consume_amt`.

OUTPUT JSON KEYS (exactly these, in this order): {{
"consume_amt": <dollar amount to spend this year; must satisfy the budget identity and be >=0>,
"nonstock_contrib": <dollar amount contributed to NONSTOCK using income>,
"nonstock_withdraw": <dollar amount withdrawn from NONSTOCK to spend>,
"divstock_contrib": <dollar amount contributed to DIVERSIFIED STOCK using income>,
"divstock_withdraw": <dollar amount withdrawn from DIVERSIFIED STOCK to spend>,
"indstock_contrib": <dollar amount contributed to INDIVIDUAL STOCK using income>,
"indstock_withdraw": <dollar amount withdrawn from INDIVIDUAL STOCK to spend>,
"other_contrib": <dollar amount contributed to OTHER using income>,
"other_withdraw": <dollar amount withdrawn from OTHER to spend>,
"transfer_nonstock_to_divstock": <dollar amount transferred from NONSTOCK to DIVERSIFIED STOCK>,
"transfer_divstock_to_nonstock": <dollar amount transferred from DIVERSIFIED STOCK to NONSTOCK>,
"transfer_nonstock_to_indstock": <dollar amount transferred from NONSTOCK to INDIVIDUAL STOCK>,
"transfer_indstock_to_nonstock": <dollar amount transferred from INDIVIDUAL STOCK to NONSTOCK>,
"transfer_divstock_to_indstock": <dollar amount transferred from DIVERSIFIED STOCK to INDIVIDUAL STOCK>,
"transfer_indstock_to_divstock": <dollar amount transferred from INDIVIDUAL STOCK to DIVERSIFIED STOCK>,
"transfer_nonstock_to_other": <dollar amount transferred from NONSTOCK to OTHER>,
"transfer_other_to_nonstock": <dollar amount transferred from OTHER to NONSTOCK>,
"transfer_divstock_to_other": <dollar amount transferred from DIVERSIFIED STOCK to OTHER>,
"transfer_other_to_divstock": <dollar amount transferred from OTHER to DIVERSIFIED STOCK>,
"transfer_indstock_to_other": <dollar amount transferred from INDIVIDUAL STOCK to OTHER>,
"transfer_other_to_indstock": <dollar amount transferred from OTHER to INDIVIDUAL STOCK>}}

Appendix B. Life Cycle Model Details

This appendix provides additional details on and the calibration of the life cycle model described in Section 1.2.

B.1 Demographics and Preferences

Each period corresponds to one year, and working life starts at $t = 0$ and lasts for T_w periods. Retirement starts at $t = T_w$, and agents can live at most T periods. Before their certain death in period $t = T$, investors face age-dependent mortality risk with survival probability in period $t + 1$ conditional on survival in period t denoted by m_t. We denote an investor's age as $a_t = t + a_0$, where a_0 is the age investors enter working life. Investors have time-separable CRRA utility over consumption streams. We denote investors' annual discount factor as β and relative risk aversion as γ.

B.2 Labor Market

At any point in time, investors can be in one of four employment states, denoted emp_t: E = employed by the same employer as in the previous period, JJ = employed by a different employer than in the previous period, U = unemployed in the current period, and Ret = retired. The fact that investors face uncertainty about their future employment status, in addition to earnings risk, is an important feature of our model because it introduces deviations in income shocks from normality, which Braxton et al. (2025) highlight as important empirically.

Employment: $emp_t = E$**.** While working, investors earn an exogenous income w_t. The log income process consists of a deterministic component that is cubic in age, a stochastic component that follows an AR(1) process with normally distributed innovations, and a transitory shock that is normally distributed:

$$\begin{aligned} \ln w_t =& \delta_0 + \delta_1 a_t + \delta_2 a_t^2 + \delta_3 a_t^3 + \eta_t + \varepsilon_t, \quad \eta_t = \rho \eta_{t-1} + \xi_t^E, \\ & \xi_0^E \sim N(0, \sigma^2_{\xi_0}), \quad \xi_t^E \sim N(0, \sigma^2_{\xi}), \quad \varepsilon_t \sim N(0, \sigma^2_{\varepsilon}) \qquad \forall t > 0. \end{aligned} \tag{5}$$

Investors' tenure status evolves according to $ten_t = ten_{t-1} + 1$ if they remain employed by the same employer. We assume that the initial distribution of ξ_0^E is different from that of ξ_t^E for $t > 0$ to account for heterogeneity in initial-period incomes.

Job transition: $emp_t = JJ$**.** While in the employed state (E), an investor may make a job-to-job transition with probability $\pi^{JJ}(t, ten_t)$ that depends on both their age and tenure at the current job. After a job-to-job transition, income evolves according to:

$$\begin{aligned} \ln w_t =& \delta_0 + \delta_1 a_t + \delta_2 a_t^2 + \delta_3 a_t^3 + \eta_t + \varepsilon_t, \quad \eta_t = \rho \eta_{t-1} + \xi_t^{JJ}, \\ & \xi_t^{JJ} \sim N(\mu^{JJ}, \sigma^2_{\xi}), \qquad \varepsilon_t \sim N(0, \sigma^2_{\varepsilon}). \end{aligned} \tag{6}$$

This earnings process captures a wage premium associated with switching jobs. Investors' tenure is reset to $ten_t = 0$ following a job-to-job transition.

Unemployment: $emp_t = U$**.** While in the employed state (E), an investor may become unemployed with probability $\pi^{EU}(t, ten_t)$ that depends on both their age and tenure at their current job. When investors are unemployed, they receive unemployment benefits equal to $ui_t = ui(\eta_t)$, where $ui(\eta_t)$ is described below. If investors become employed at $t+1$ after being unemployed in period t, income at $t+1$ evolves according to

$$\begin{aligned} \ln w_{t+1} &= \delta_0 + \delta_1 a_{t+1} + \delta_2 a_{t+1}^2 + \delta_3 a_{t+1}^3 + \eta_{t+1} + \varepsilon_{t+1}, \\ \eta_{t+1} &= \rho\eta_t + \xi_{t+1}^U, \quad \xi_{t+1}^U \sim N(-\mu^{EU}, \sigma_\xi^2), \quad \varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2). \end{aligned} \tag{7}$$

This earnings process captures the persistent wage reduction associated with experiencing unemployment.

Retirement: $emp_t = Ret$**.** In period $t = T_w$, all investors retire deterministically. During retirement in periods $t \in [T_w, T-1]$, investors earn public pension benefits denoted by ss_t, which are described below.

B.3 Savings Account

Investors start with zero assets at $t = 0$ and cannot borrow. They can accumulate assets inside a savings account that can be invested in one of two financial assets for the life cycle model and one of four financial assets for the prompts to the LLMs. First, there is a risk-free bond that has a constant gross return of $R_t^B = R_f$ per year. Second, there is a risky asset that corresponds to a diversified stock market index and pays a stochastic i.i.d. gross return of R_t^D per year, where

$$\ln R_t^D = \ln R_f + \mu_D + u_t^D, \quad u_t^D \sim \mathcal{N}(0, \sigma_D^2). \tag{8}$$

Additionally, there are two non-diversified assets, each with its own stochastic gross return, R_t^I and R_t^O, corresponding to the individual stock bundle and other risky assets (e.g., crypto, gold, commodities, and collectibles), respectively. Their joint return process is defined relative to the diversified market return:

$$\ln R_t^x = \alpha_x + b_x\big(\ln R_t^D - \ln R_f\big) + \sigma_{\varepsilon,x} u_{x,t}, \quad u_{x,t} \sim \mathcal{N}(0,1), \quad \forall x \in \{I, O\}, \tag{9}$$

where $u_{I,t}$ and $u_{O,t}$ are drawn independently of each other and of the market shock u_t^D in (8). The balance of the savings account, denoted by L_t, then evolves according to:

$$L_{t+1} = \big(L_t + s_t^l\big)\big[1 + (R_{t+1}^\theta - 1)(1 - \tau_c)\big], \quad L_0 = 0, \tag{10}$$

where s_t^l is the net saving that the investor places in this account, τ_c is the rate of capital taxation, and R_t^θ is the investors' portfolio return that depends on their portfolio share in the risky assets.

$$R_{t+1}^\theta = R_f + \theta_t^D \left(R_{t+1}^D - R_f\right) + \theta_t^I \left(R_{t+1}^I - R_f\right) + \theta_t^O \left(R_{t+1}^O - R_f\right), \quad \theta_t^D + \theta_t^I + \theta_t^O \in [0,1]. \tag{11}$$

When solving the life cycle model, we impose $\theta_t^I = \theta_t^O = 0$ because, by construction, neither non-diversified asset can improve the Sharpe ratio of a portfolio that combines the bond and diversified stock index (see Appendix B.6). While this is an approximation because the CAPM would not arise in equilibrium in our model, we make this assumption for parsimony when solving the model. In the LLM simulation, however, all four assets are available, and the LLM is free to allocate wealth to the non-diversified assets.

B.4 Government

Unemployment benefits. Investors receive an unemployment benefit of $ui(\eta_t)$ when their employment ends. This benefit depends on the labor productivity, η_t, from the last period in which the agent was employed.

Retirement benefits. After retirement, investors receive Social Security benefits, denoted by $ss_t = ss(ae_{T_w-1})$. ae_{T_w-1} is the investor's average lifetime earnings at the time of retirement. Let ae_t denote the average of wages up to and including period t. Initialize $ae_0 = w_0$. Then

$$ae_{t+1} = \begin{cases} \frac{w_{t+1}+(t+1)\cdot ae_t}{t+2}, & \text{if } t < T_w - 1, \\ ae_{T_w-1} & \text{else.} \end{cases} \tag{12}$$

Taxation. Investors face a nonlinear income tax schedule $tax_i(\cdot)$, which depends on their taxable income. For employed individuals or those in job transitions, their taxable income is their wages. For unemployed or retired individuals, their taxable income is the benefits they receive from the government.

B.5 Summary of Investors' Problem

Investors face a dynamic optimization problem with seven state variables: a_t = age, η_t = labor productivity, ε_t = transitory income shock, emp_t = employment status, ten_t = tenure, ae_t = average lifetime income, and L_t = savings. Denote the vector of these state variables as $\mathbf{x}_t$. Investors have three controls: c_t = consumption, θ_t^D = portfolio share, and s_t^l = savings. In choosing these controls, we restrict investors from borrowing and engaging in any margin trading (i.e., short-selling or taking leveraged positions):

$$L_t \geq 0, \quad \theta_t^D \in [0,1]. \tag{13}$$

When simulating LLM advice, the only difference is that there are five controls, including the two non-diversified asset shares θ_t^I and θ_t^O.

B.6 Calibration

We calibrate our model parameters following Choukhmane and de Silva (2026) where possible. We provide a brief overview here and refer the reader to that paper for a more detailed discussion.

Demographics. We set the length of one period in the model to one year and set $a_0 = 22$, $T_w = 43$, and $T = 68$, such that workers enter working life at age 22, retire at 65, and live their final year of life at 89. For each age, we calibrate mortality risk to match the 2015 U.S. Social Security Actuarial Life Tables. We use the equivalence scale estimated in Lusardi et al. (2017) to capture changes in household composition over the life cycle.

Labor income process. We use data from the Survey of Income and Program Participation (SIPP) to estimate parameters of the labor income process and transition probabilities at the annual frequency. This income process has several components. First, we estimate an earnings process for workers staying in the same job, corresponding to (5), which contains a deterministic and stochastic component.[6] Second, we use data on employment transitions from SIPP to estimate the median salary increase following a job-to-job transition, μ^{JJ}, and the median salary decrease when workers transition back to employment after an unemployment spell, $-\mu^{EU}$. Third, we use SIPP microdata to estimate the three transition probabilities between the three working-age labor market states. Finally, we set the initial unemployment rate equal to 22%, which is the share of age-22 individuals in SIPP who are unemployed.

Tax and benefit system. Investors' tax liability, $tax_i(\cdot)$, is calculated according to the 2025 U.S. federal income tax schedule for single filers who claim only the standard deduction. We calculate Social Security benefits according to the 2025 formula with a Supplemental Security Income program floor. Unemployment benefits are computed with a replacement rate of 40%. We set the capital return tax rate, τ_c, to 21%.

Bond and diversified stock returns. We set the net risk-free rate to be constant at 2%, a standard value in life cycle models (Gomes 2020). We set the equity premium to be 6.4%, which is equal to the average CPI-adjusted return on the CRSP Value-Weighted Index between 1925 and 2006 minus our 2% risk-free rate. We set the volatility of log stock returns to 20%, which matches that of the CRSP Value-Weighted Index. We assume that asset returns are uncorrelated with shocks to labor income and employment transition probabilities.

Non-diversified asset returns. Both non-diversified assets are calibrated to have the same underlying return process, but independent idiosyncratic shock realizations, so they represent distinct

[6] Relative to Choukhmane and de Silva (2026), we also estimate the variance of the transitory shock as part of this estimation.

risky positions. We calibrate this process such that neither non-diversified asset can improve the Sharpe ratio of a portfolio that combines a bond and a diversified stock market index, implying that a CAPM investor would optimally hold zero portfolio shares in both assets. Formally, we start by taking the average arithmetic return and standard deviation, μ_x and σ_x, for individual stocks from Bessembinder (2018), which gives $\mu_x = 0.1474$ and $\sigma_x = 0.819$. Assuming the CAPM holds and denoting μ_m as the arithmetic mean diversified market return, we can compute the non-diversified asset beta using:

$$\beta_x^* = \frac{(1+\mu_x) - R_f}{(1+\mu_m) - R_f}. \tag{14}$$

Next, note that log return parameters for the non-diversified asset are given by:

$$\sigma_{x,\log} = \sqrt{\log\left(1 + \left(\frac{\sigma_x}{1+\mu_x}\right)^2\right)}, \tag{15}$$

$$\mu_{x,\log} = \log(1+\mu_x) - \tfrac{1}{2}\,\sigma_{x,\log}^2. \tag{16}$$

We assume $\beta_x^* = b_x$, and then compute the idiosyncratic log volatility using

$$\sigma_{\varepsilon,x} = \sqrt{\sigma_{x,\log}^2 - b_x^2 \sigma_D^2} \tag{17}$$

and the log intercept as

$$\alpha_x = \mu_{x,\log} - b_x \mu_D, \tag{18}$$

where μ_D is the mean log excess return of the diversified asset in (8).[7] This then implies that, given a simulated market gross return R_t^D and an independent draw $u_{x,t} \sim \mathcal{N}(0,1)$, we can construct each non-diversified asset's return as

$$R_t^x = \exp\Big(\alpha_x + b_x\big(\log R_t^D - \log R_f\big) + \sigma_{\varepsilon,x}\, u_{x,t}\Big). \tag{19}$$

[7] We assume $b_x = \beta_x^*$ for simplicity. In reality, the arithmetic beta implied by a log-loading b_x is $\frac{\text{cov}(R_t^x, R_t^D)}{\text{var}(R_t^D)} = \frac{1+\mu_x}{1+\mu_m}\frac{\exp(b_x\sigma_D^2)-1}{\exp(\sigma_D^2)-1}$, which implies $b_x = \frac{1}{\sigma_D^2}\log\Big(1 + \beta_x^*\frac{1+\mu_m}{1+\mu_x}\big(\exp(\sigma_D^2)-1\big)\Big)$. However, given our calibration, these two are quite similar.

Appendix C. Textual Analysis Details

This appendix provides methodological details for the dictionary-based textual analysis described in Section 2.1.

Text preprocessing. Each free-text response is processed through a three-step pipeline. First, finance-specific terms are normalized using pattern-based rewriting (e.g., "t-bills" becomes "tbill"). Second, the text is tokenized and lemmatized using spaCy's `en_core_web_sm` model, which splits text into individual words and reduces them to root forms (e.g., "worried" becomes "worry"). We use the small spaCy model because testing across all three model sizes (`sm`, `md`, `lg`) showed similar results. Third, we add a set of manual root reductions specific to financial words that the spaCy model misses (e.g., "retirement" to "retire"). Importantly, finance-relevant strings—including account types (e.g., 401k), fund tickers (e.g., VTI), and acronyms (e.g., ETF)—are preserved as-is without lemmatization, but plural forms are normalized (e.g., "401ks" to "401k," "ETFs" to "ETF"). The output is one sequence of root-form words per respondent per prompt.

Dictionary construction. Based on reading the prompts, we manually define 27 topic dictionaries covering the following categories: Debt, Savings, Retirement, Housing, Investment Strategy, Income, Employment Security, Budgeting, Education, Family, Gender, Financial Hardship, Financial Anxiety, Risk Preference, Medical, Long-Term Planning, Tax, Lifestyle Goals, Macroeconomic, Investment Assets, Providers, Products, Liquidity, Discipline, Diversification, Insurance, and Inheritance. Each category contains a list of keywords, which are listed in Table C1. Importantly, categories are not mutually exclusive: keywords can appear in multiple dictionaries (e.g., "mortgage" appears in both Debt and Housing). Additionally, keywords are processed through the same pipeline as respondent text, ensuring consistent matching between the prompts and categories. Multi-word keywords (e.g., "credit card," "high yield savings account") are matched by checking for consecutive words in the respondent's processed text.

Stopwords. When making word clouds, a list of stopwords—a union of scikit-learn's default English stopwords, custom generic terms, and survey-instrument words (e.g., "financial," "advice," "situation")—is removed. Prompt-specific tautological terms are additionally removed from the prompt word clouds: "spend" is excluded from the spending prompt and "invest" from the investment prompt. The combined prompt word cloud (Figure 2) excludes the union of all prompt-specific stopwords, while the individual prompt word clouds (Figure E5) exclude only the stopwords specific to each prompt.

Table C1. Dictionary Categories and Keywords

Category	Keywords
Debt	bankruptcy, car loan, consolidate, credit, credit card, credit card payment, credit score, debt, debt payment, loan, loan payment, loan principal, minimum payment, mortgage, mortgage payment, owe, pay debt, payoff, personal loan, refinance, student debt, student loan
Savings	checking account, emergency fund, emergency savings, high yield savings, high yield savings account, nest egg, rainy day, save money, savings account, set aside
Retirement	401k, annuity, ira, pension, required minimum distribution, retire, retirement, retirement account, retirement savings, rmd, roth, roth ira, social security, ssdi, ssi
Housing	apartment, buy house, condo, down payment for house, down payment on house, home down payment, home equity, home loan, house, house down payment, house fund, housing, mortgage, mortgage down payment, own home, property, rent, rent payment
Investment Strategy	allocate, capital appreciation, compound interest, compounding, day trade, day trading, dividend, growth, portfolio, profit, return
Income	annual income, bonus, disability, disposable income, earning, fixed income, household income, income, limited income, low income, monthly income, passive income, paycheck, salary, stipend, take home, wage
Employment Security	change job, employment, freelance, gig, job, job stability, jobless, layoff, main provider, provider, self-employed, sole provider, stable employment, stable income, stable job, unemployed, unemployment
Budgeting	afford, bill, budget, cost, cut back, discretionary, entertainment, essential, expense, food, groceries, living expenses, monthly budget, necessities, spend less, spending plan, subscription, utilities
Education	college, degree, education, school, student, tuition

Category	Keywords
Family	boyfriend, brother, childcare, children, dad, daughter, dependent, divorced, family, father, fiance, fiancee, girlfriend, grandchildren, grandfather, grandmother, household, husband, kids, married, mom, mother, parent, partner, sister, son, spouse, widow, widower, wife
Gender	boyfriend, dad, father, female, gentleman, girlfriend, grandfather, grandmother, husband, lady, male, man, mom, mother, pregnant, single dad, single father, single mom, single mother, stay at home dad, stay at home mom, widow, widower, wife, woman
Financial Hardship	am broke, bankruptcy, barely, be broke, concerned about bills, concerned about money, difficult, ends meet, financial strain, go broke, hardship, paycheck to paycheck, poverty, shortfall, struggle, tight, worried about bills, worried about money
Financial Anxiety	afraid, anxiety, anxious, concern, fear, hesitant, nervous, overwhelm, panic, panicking, stress, stressed, stressful, uncertain, unsure, worry
Risk Preference	aggressive, conservative, gamble, guaranteed, high risk, low risk, moderate risk, risk adverse, risk averse, risk free, risk profile, risk tolerance, risk-averse, risky, safer, secure, speculative, volatile
Medical	copay, dental, doctor, health, health insurance, healthcare, medical, medication, medicine, premium, prescription, therapy
Long-Term Planning	financial goal, financial security, future, goal, horizon, long term, retire early, retirement goal, short term, timeline
Tax	capital gains, property tax, tax, tax bracket, tax-advantaged, tax-free, taxable
Lifestyle Goals	car, car insurance, car lease, car payment, commute, new car, transportation, travel, trip, vacation, vehicle
Macroeconomic	cost of living, crash, downturn, economy, fed funds rate, federal reserve, high prices, inflation, interest rate, market crash, price increase, purchasing power, recession, rising prices

Category	Keywords
Investment Assets	ada, altcoin, annuity, avax, balanced fund, bitcoin, blue chip, blue chip stocks, bnb, bond, bond etf, bond fund, bond index, bond mutual fund, brokerage, brokerage account, btc, cardano, cash equivalent, cd, cds, chainlink, commodities, commodity, corporate bond, crypto, dividend etf, dividend fund, dividend stock, dogecoin, etf, eth, ethereum, gold, government bond, growth fund, growth stock, high yield account, high yield savings, high yield savings account, high-yield savings, high-yield savings account, hysa, ibond, income fund, index etf, index fund, individual stock, international stock, junk bond, large cap, litecoin, ltc, market fund, matic, mid cap, money market, money market account, money market fund, muni bond, municipal bond, mutual fund, penny stock, precious metal, preferred stock, real estate, reit, reit etf, rental property, ripple, s&p, savings bond, sector fund, silver, single stock, small cap, sol, solana, stablecoin, stock, stock etf, stock fund, stock market, stock mutual fund, t-bill, target date, target date fund, target-date, target-date fund, tbill, three fund portfolio, three-fund portfolio, total bond market, total international, total market, total stock market, treasury, treasury bill, treasury bills, treasury bond, treasury note, usdc, usdt, value fund, value stock, xrp
Providers	acorns, ally, ally bank, ameriprise, bank of america, betterment, bilt, bofa, capital one, cash app, cashapp, charles schwab, chase account, chase bank, chase card, chase checking, chase credit, chase freedom, chase sapphire, chase savings, coinbase, discover bank, discover financial, discover online savings, e*trade, edward jones, etrade, fidelity, goldman sachs, jpmorgan chase, m1 finance, marcus, merrill, merrill lynch, morgan stanley, navy federal, paypal, robin hood, robinhood, schwab, sofi, synchrony, td ameritrade, td bank, usaa, vanguard, wealthfront, webull, wells fargo, zelle
Products	agg, avax, avuv, bil, binance coin, bitcoin, bnb, bnd, btc, cardano, chainlink, dgro, dogecoin, eth, ethereum, fskax, ftihx, fxaix, fxnax, fzilx, fzrox, gld, iau, ief, itot, ivv, ixus, jepi, jepq, litecoin, matic, polygon, qqq, qqqm, ripple, schb, schd, schh, schp, schz, sgov, slv, sol, solana, soxx, spaxx, splg, spy, swisx, swppx, swtsx, swvxx, tlt, usdc, usdt, vbtlx, vfiax, vgit, vgsh, vgt, vig, vmfxx, vnq, voo, vt, vti, vtiax, vtip, vtsax, vtwax, vug, vwo, vxus, vym, xlk, xrp

Category	Keywords
Liquidity	accessible, available funds, buffer, cash buffer, cash on hand, cash reserve, cushion, easy access to funds, emergency cash, emergency fund, emergency savings, hold cash, liquid, liquid cash, liquidity, rainy day, withdraw, withdrawal
Discipline	cut back, discipline, frivolous, guilty, habit, impulsive, indulge, overspend, restrictive, self discipline, shopping, splurge, unnecessary, wasteful
Diversification	asset allocation, asset mix, balance between different investment options, balance between investment options, balance between investments, balanced portfolio, between stocks and bonds, diversification, diversified, diversify, mix, portfolio allocation, rebalance, split between, stocks and bonds
Insurance	auto insurance, car insurance, coverage, deductible, homeowner insurance, life insurance, renter insurance
Inheritance	beneficiary, estate planning, inherit, inheritance, trust fund

Notes: This table summarizes the dictionary categories and keywords used in the textual analysis in Section 2.1. Go back.

Appendix D. Additional Details on Simulated Method of Moments Estimation

To estimate the preference parameters β and γ for each version of the LLM output, we use simulated method of moments on a discrete grid. Our estimation uses a set of 324 moments. We use the 25th, 50th, and 75th percentiles of the wealth-to-income ratio from ages 23 to 64, and the 25th, 50th, and 75th percentiles of equity share from ages 23 to 88. Let $\mathcal{W}_{i,t}$ and $\theta_{i,t}$ be the wealth-to-income ratio and equity share, respectively, for agent i at time t. The former is defined as $\mathcal{W}_{i,t} \equiv \frac{L_{i,t}}{w_{i,t} - tax_i(w_{i,t})}$.

For the wealth-to-income ratio moments, we omit the first year, when wealth is zero by construction. We also restrict these moments to working ages, which isolates saving accumulation before retirement: LLM withdrawal advice in retirement bunches at fixed rules, such as 4% of wealth, that the two-parameter life cycle model is not designed to reproduce. For the equity share moments, we omit the first and last periods of life.

We estimate the preferences by solving the life cycle model on a 201-by-121 grid of β and γ values ranging from 0.8 to 1.2 and 2 to 14, respectively. In the β dimension, we use a step size of 0.002 while in the γ dimension, the step size is 0.1. After solving for the life cycle moments for every point on this grid, we then calculate an error term between the targeted moments and every point on the grid as follows:

$$Z(\beta,\gamma) = X_{\text{LC}}(\beta,\gamma) - X_{\text{LLM}}, \qquad \text{SMM}_{\text{error}}(\beta,\gamma) = Z(\beta,\gamma)' \cdot W \cdot Z(\beta,\gamma) \tag{20}$$

In the above equation, X_{LLM} is the vector of LLM moments and $X_{\text{LC}}(\beta,\gamma)$ is the corresponding vector of life cycle moments when solving the life cycle model using β and γ as the inputted preference parameters. W is the weighting matrix. For our estimation, we use the identity matrix as the weighting matrix as it has better small sample properties than the optimal weighting matrix.

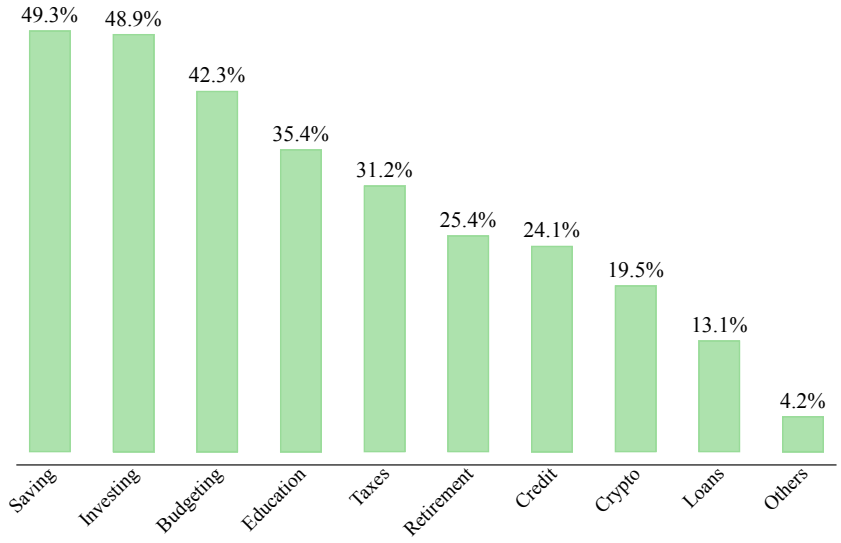
Appendix E. Additional Tables and Figures

Figure E1. Survey Questions for Eliciting LLM Prompts



Notes: This figure displays the three survey questions used to collect respondent-written LLM prompts. Go back.

Figure E2. Personal Finance Topics Discussed with AI Tools



Notes: This figure shows the personal finance topics that respondents report discussing with AI tools, among those who have previously used AI for financial advice or information. Respondents could select multiple topics. Go back.

Table E1. Demographic Comparison: CPS vs. Prolific Survey

	CPS Data	Prolific Survey
Male	49%	49%
20–29 years old	17%	17%
30–39 years old	18%	19%
40–49 years old	17%	16%
50–59 years old	16%	21%
60–69 years old	16%	18%
70–79 years old	11%	8%
80+ years old	5%	0%
$0-25,000	35%	27%
$25,001-50,000	25%	27%
$50,001-75,000	16%	20%
$75,001-100,000	10%	12%
$100,001-150,000	8%	9%
$150,000+	6%	5%
White	76%	63%
Black/African-American	13%	11%
Asian/Asian-American	7%	6%
Mixed	2%	11%
Other	2%	8%
Employed	62%	72%
Unemployed	3%	15%
Retired	36%	14%

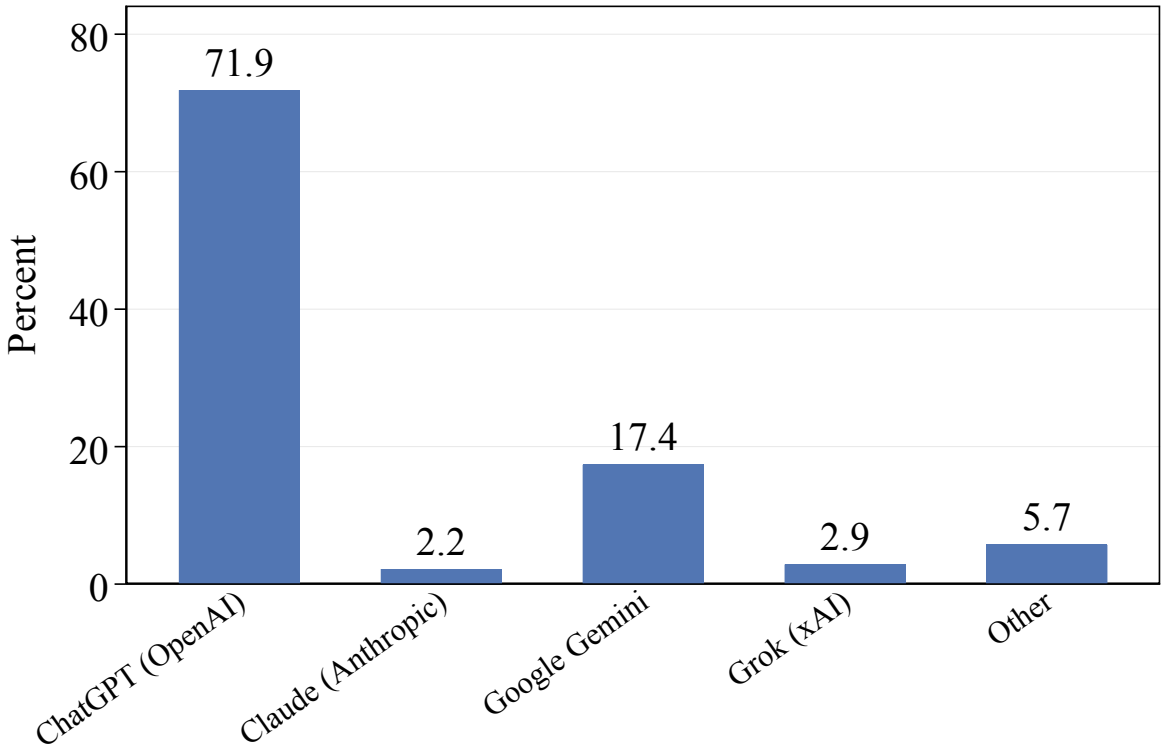
Notes: This table compares the demographic composition of the Prolific survey sample with the March 2025 Current Population Survey (CPS). The race categories follow the response options collected in the Prolific survey. The income demographics from the Prolific survey are calculated on the subset of responses that report their income. This subset makes up 96% of the total sample. Go back.

Table E2. Prompt Summary Statistics

Prompt	Subgroup	Level	N	Mean Word Count	Mean Char Count	Share Numbers (%)	Share Dollars (%)
All	All	All	952	96.7	499.2	76.3	45.1
Financial Situation	All	All	952	43.0	218.3	65.2	34.4
Spending Advice	All	All	952	26.7	137.0	31.1	17.0
Investment Advice	All	All	952	27.0	142.0	27.7	11.9
All	Income	Low	490	96.9	497.2	74.3	44.5
All	Income	Mid	296	96.0	498.4	78.7	45.3
All	Income	High	130	93.4	486.5	80.0	50.8
All	Financial Literacy	Low	253	88.6	450.0	70.4	41.1
All	Financial Literacy	Mid	345	100.8	519.6	79.1	45.2
All	Financial Literacy	High	354	98.5	514.6	77.7	47.7
All	AI Use	No	497	90.3	463.3	73.6	41.9
All	AI Use	Yes	455	103.7	538.5	79.1	48.6
All	Gender	Female	486	98.2	501.3	74.5	44.7
All	Gender	Male	466	95.2	497.2	78.1	45.5
All	Employment	Unemployed	139	105.1	544.2	68.4	36.7
All	Employment	Retired	132	87.2	451.1	75.0	46.2
All	Employment	Employed	681	96.9	499.4	78.1	46.6

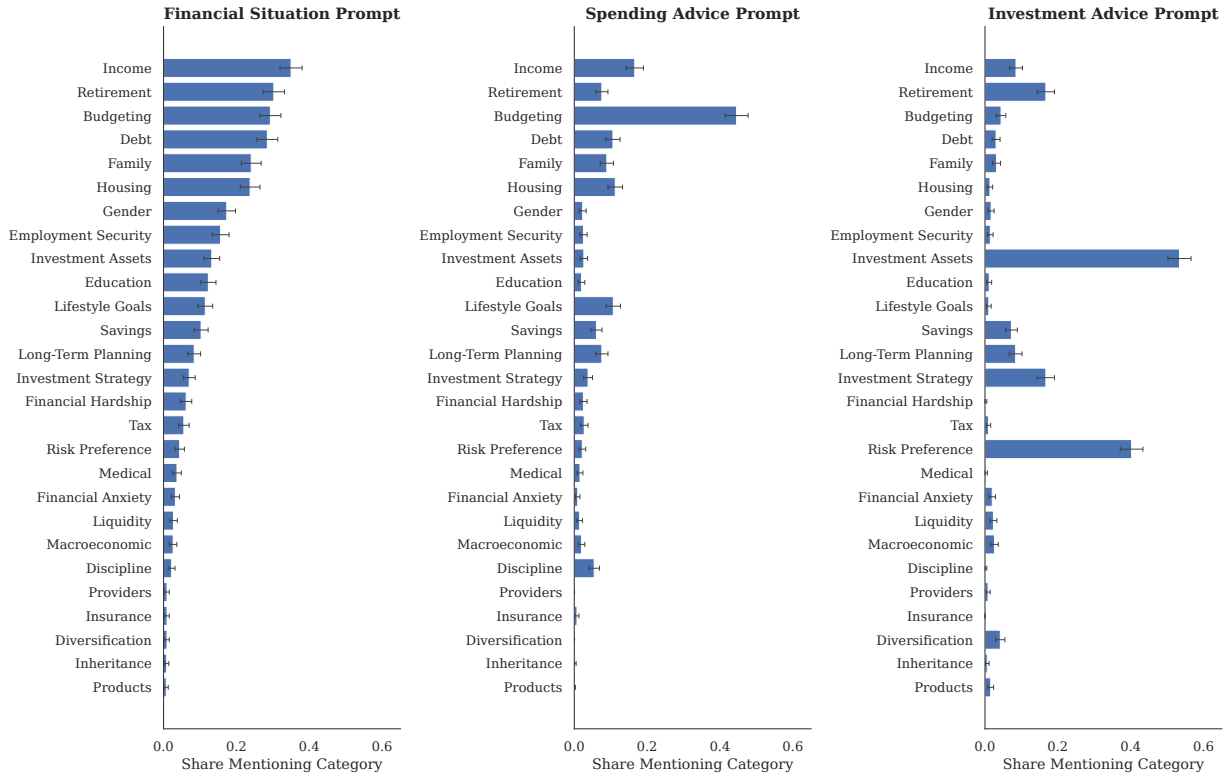
Notes: This table reports descriptive statistics for the respondent-written individual prompts. The first row combines the three prompt questions for each respondent; the next three rows report statistics separately for the financial situation, spending advice, and investment advice prompts. The remaining rows pool the three prompts and report statistics by demographic subgroup. "Share Numbers" is the share of responses containing any digit. "Share Dollars" is the share containing "$" or the word "dollar." Income groups are low (<$50k), middle ($50k–$100k), and high (>$100k); income is not reported for 36 respondents. Financial literacy groups are low (0–3 correct), middle (4 correct), and high (5 correct out of 5 quiz questions). AI Use indicates whether the respondent has previously used AI for financial advice. Go back.

Figure E3. AI Model Usage Among Survey Respondents



Notes: This figure reports the AI model used by respondents who have previously used AI for financial advice. Bars show the share of these respondents reporting each model. Go back.

Figure E4. All Dictionary Categories by Prompt



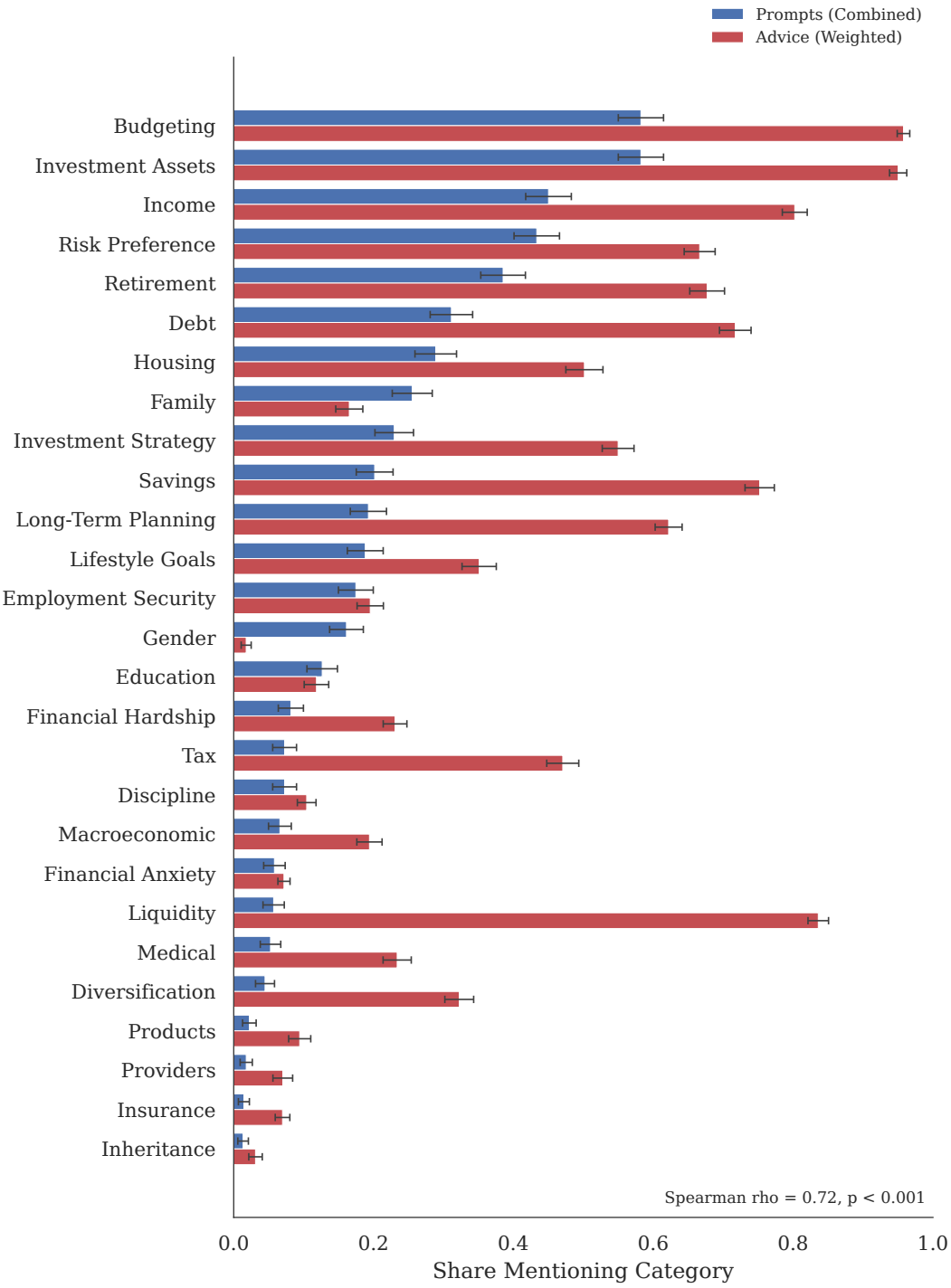
Notes: This figure reports dictionary category mention rates separately for each of the three prompt questions. Bars show the share of respondents whose prompt contains at least one keyword from the category; capped lines denote 95% confidence intervals. Categories are ordered by their mention rate in the financial situation prompt. Go back.

Figure E5. Word Clouds by Individual Prompt



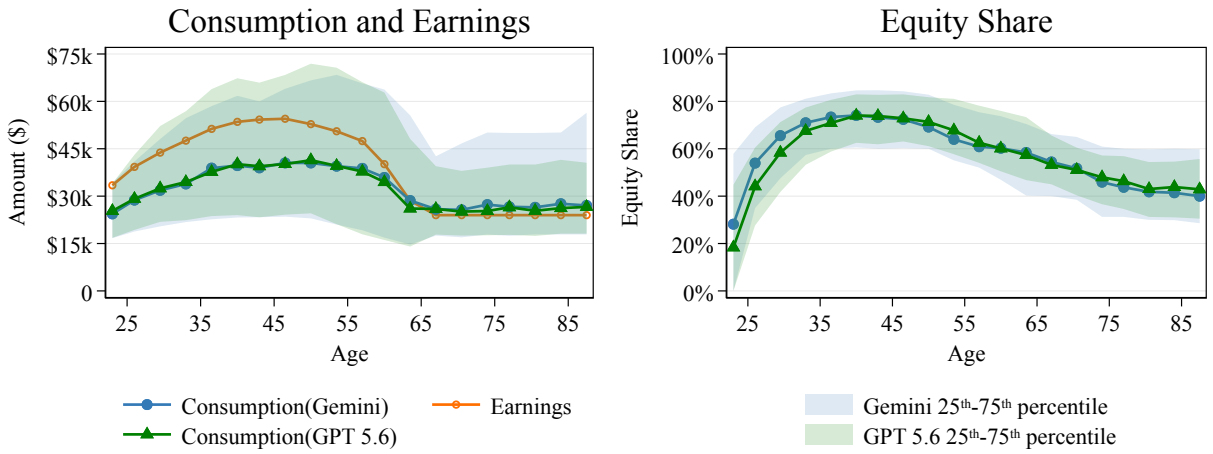
Notes: This figure reproduces the analysis in Figure 2, separating out the three prompts that respondents write. Go back.

Figure E6. All Dictionary Categories: Prompts vs. Advice



Notes: This figure compares dictionary category mention rates in respondents' combined prompts with the corresponding LLM advice. Blue bars show the share of respondents whose combined prompt text contains at least one keyword from the category. Red bars show the share of advice responses mentioning the category, with each respondent's prompt weighted equally. Categories are ordered by their mention rate in the combined prompts. Capped lines denote 95% confidence intervals. The Spearman rank correlation between prompt and advice mention rates is reported in the bottom-right corner. Go back.

Figure E7. Life Cycle Profiles of LLM-Recommended Consumption and Equity Shares: Alternative Models



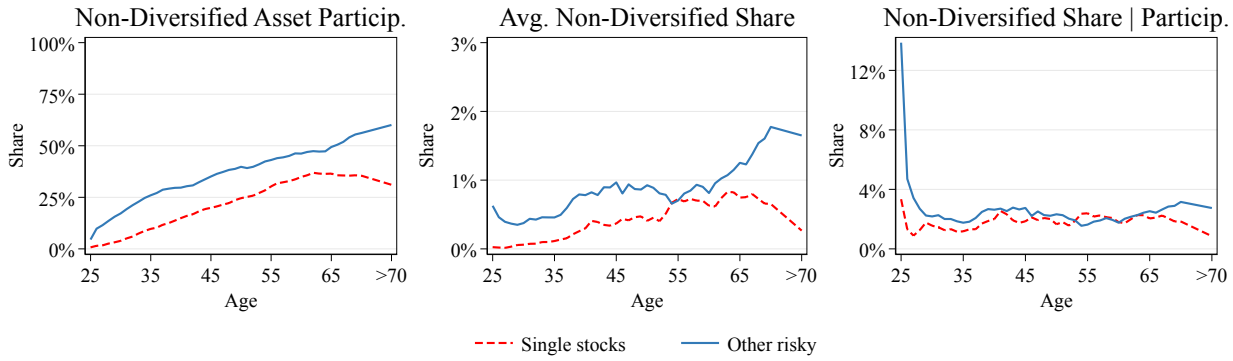
Notes: This figure plots the life cycle profiles for individuals following the LLM's recommended choices using `Gemini 3 Flash` and `GPT-5.6 Terra` with the survey prompts, as described in Section 1.3 and Section 3.4. Ages (22–89) are grouped into 20 equally sized bins. The left panel shows consumption and post-tax earnings, and the right panel shows the equity share. Markers denote median values within each bin, and shaded areas represent the interquartile range (25^{th}–75^{th} percentile). All values are in 2025 dollars. Go back.

Table E3. Summary Statistics from LLM and Life Cycle Model Simulations

Statistic	Employment Status	Age					
		22–29	30–39	40–49	50–59	60–64	Total
Panel A: Simulated Inputs							
Post-tax Earnings ($)	Employed	44,145	60,884	68,857	66,335	60,191	60,782
	Unemployed	16,449	22,017	22,838	21,782	19,249	20,456
Employment rate (%)	-	90	91	91	85	61	86
Avg. Tenure (years)	-	1.9	3.1	3.9	4.0	3.3	3.3
Panel B: LLM Endogenous Choices							
Consumption ($)	Employed	34,637	47,804	53,918	54,914	53,437	48,842
	Unemployed	13,363	19,294	21,539	24,952	27,202	22,816
Liquid Wealth ($)	Employed	46,964	215,949	618,130	1,350,137	2,147,938	701,998
	Unemployed	29,315	163,963	487,529	1,233,073	1,931,464	1,032,009
Stock participation (%)	Employed	87	100	100	100	100	97
	Unemployed	84	99	100	100	100	98
Equity share (%)	Employed	45	69	73	66	59	64
	Unemployed	42	65	72	66	59	62
Net Saving Rate	Employed	21	19	18	14	7	17
(% of earnings)	Unemployed	17	9	-2	-32	-62	-25
Panel C: Life Cycle Model (LCM) Endogenous Choices							
Consumption ($)	Employed	25,450	43,469	64,936	85,830	99,249	59,578
	Unemployed	19,687	33,185	48,613	63,660	77,342	55,546
Liquid Wealth ($)	Employed	70,262	334,886	736,493	1,124,386	1,315,676	644,512
	Unemployed	44,536	241,214	583,203	881,494	1,041,613	684,088
Stock participation (%)	Employed	86	100	100	100	100	97
	Unemployed	78	100	100	100	100	97
Equity share (%)	Employed	86	97	82	67	57	81
	Unemployed	78	99	82	65	59	72
Net Saving Rate	Employed	40	23	-4	-48	-98	-7
(% of earnings)	Unemployed	-25	-62	-131	-251	-390	-220

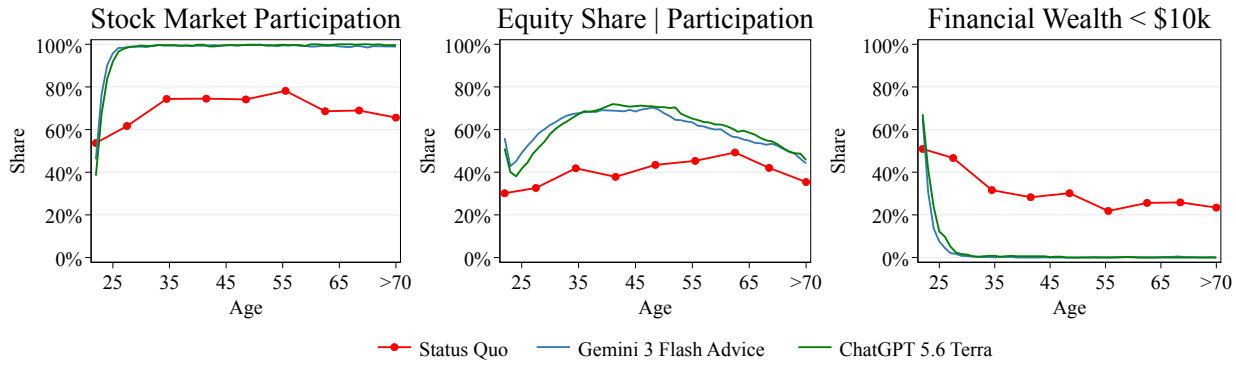
Notes: This table reports mean values for key simulated inputs and outcomes by age group. The Total column reports averages over ages 22–64. Panel A presents exogenous inputs that are identical across both LLM and life cycle model simulations. Post-tax earnings correspond to wage and unemployment insurance earnings; employment rate is the share of individuals employed; and tenure is the average number of years at the current employer. Panels B and C report endogenous outcomes from the LLM and life cycle model, respectively, separately by employment status. Net saving rate is defined as post-tax earnings minus consumption, divided by post-tax earnings. All monetary values are in 2025 dollars. Go back.

Figure E8. Non-Diversified Asset Participation and Portfolio Shares



Notes: This figure plots the life cycle profiles of non-diversified asset holdings—individual stocks and other risky assets (e.g., crypto, gold, commodities, collectibles)—for individuals following the LLM's advice using the survey prompts, as described in Section 1.3. The left panel shows unconditional participation rates by age; the middle panel shows the average portfolio share of financial wealth allocated to each asset class; and the right panel shows the average portfolio share conditional on participation. Ages (22–89) are grouped into 20 equally sized bins. Go back.

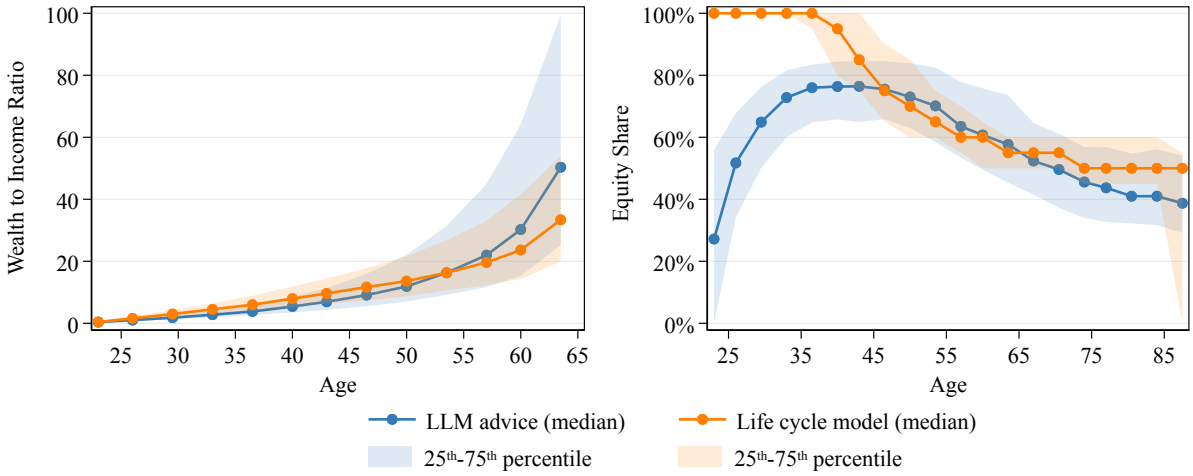
Figure E9. Observed Behavior vs. LLM-Recommended Behavior: Alternative Models



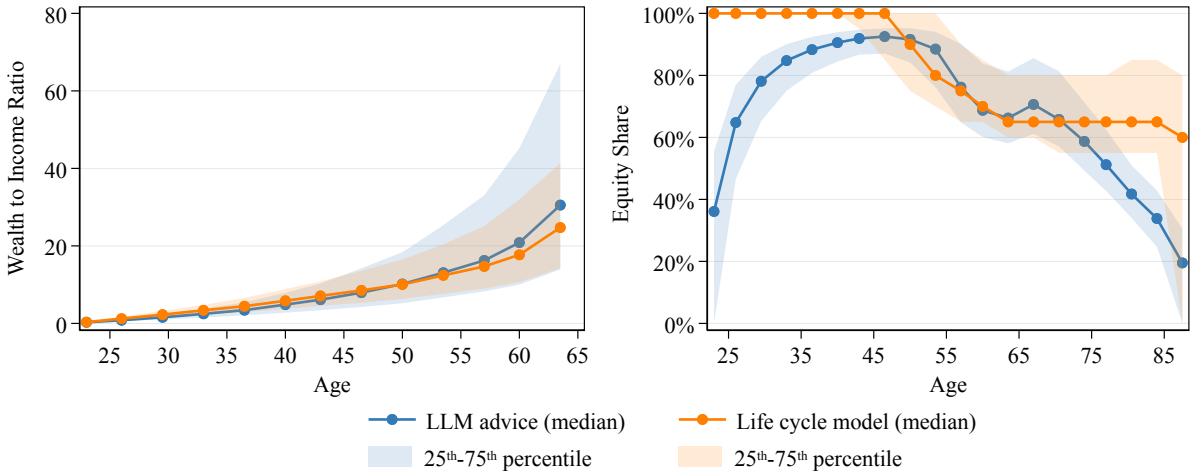
Notes: This figure compares survey respondents' observed financial behavior with the behavior implied by following advice from `Gemini 3 Flash` and `GPT-5.6 Terra` in the full life cycle simulation described in Section 1.3. The three panels show the average stock market participation rate (left), average equity share conditional on participation (middle), and the share of individuals with financial wealth below $10,000 (right). Ages are grouped into eight 6-year bins and one last bin that captures every observation age 70 and above. All values are in 2025 dollars. Go back.

Figure E10. Estimated Life Cycle Model Fit on Targeted Moments

Panel A: Survey Prompts, `GPT-5.2`



Panel B: Academic Prompts, `GPT-5.2`



Panel C: Survey Prompts, `Gemini 3 Flash`

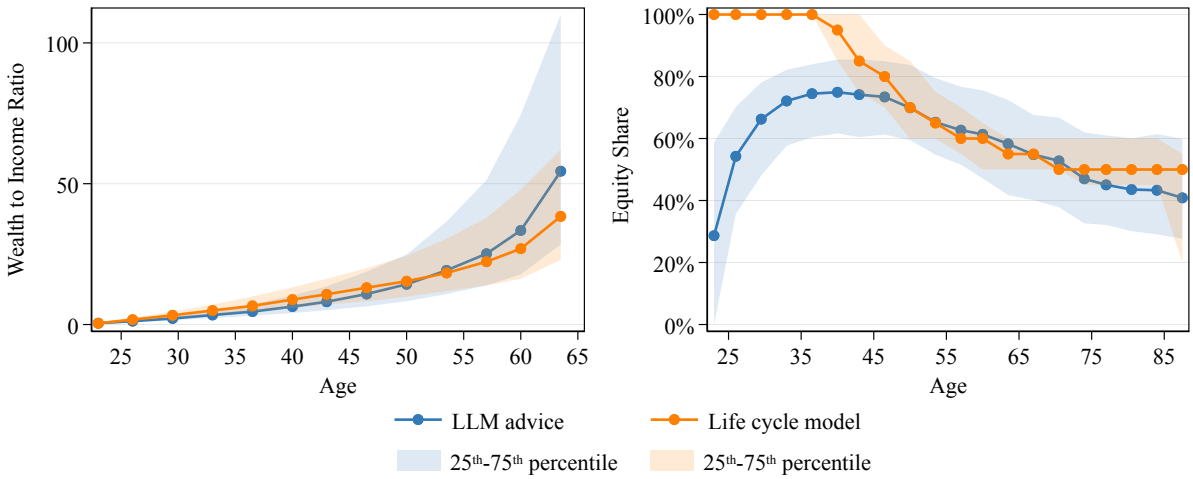
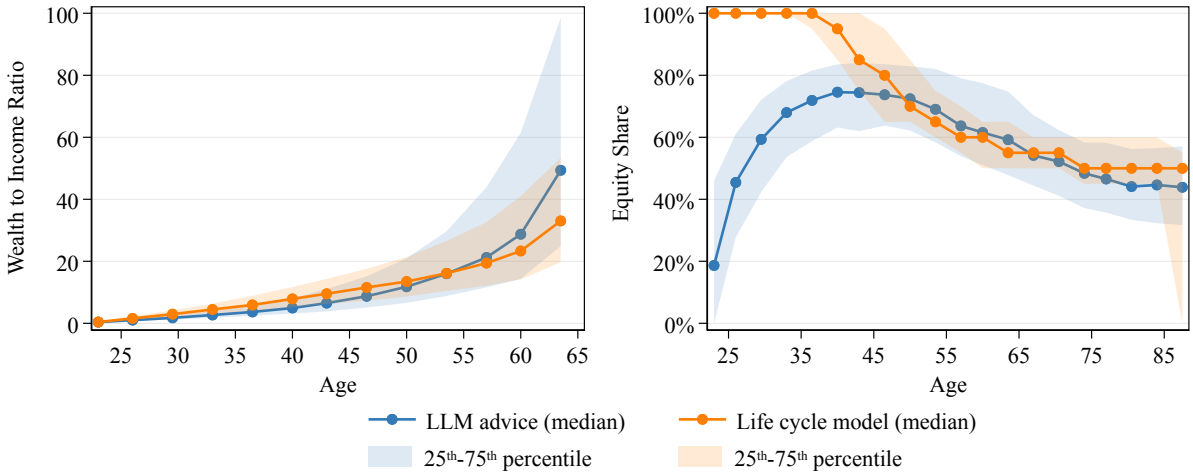


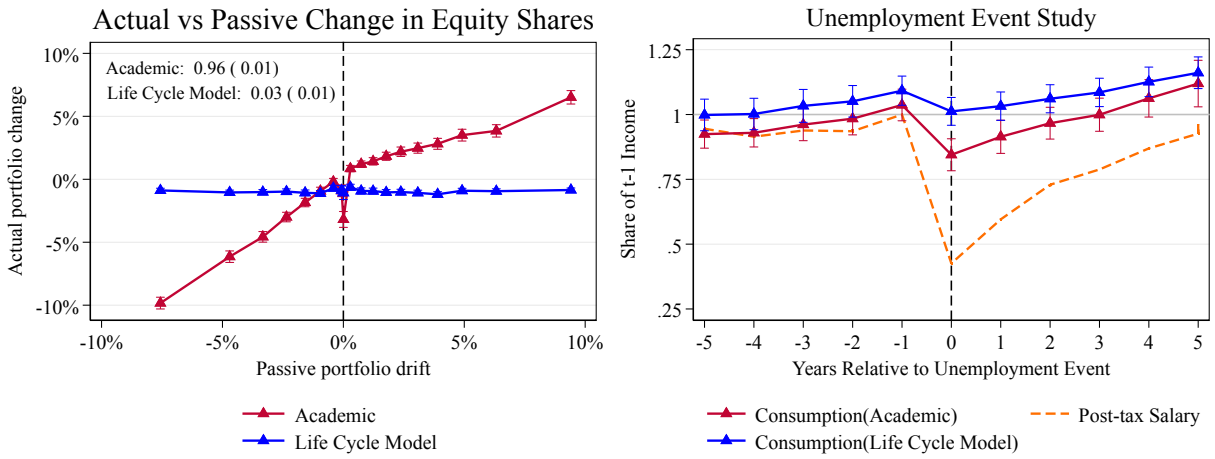
Figure E10. Estimated Life Cycle Model Fit on Targeted Moments (Continued)

Panel D: Survey Prompts, `GPT-5.6 Terra`



Notes: This figure shows the fit of the estimated life cycle model, using the preference parameters reported in Table 2, to the targeted SMM moments. Panels A and B use `GPT-5.2` with the survey and academic prompts, respectively. Panels C and D use the survey prompts with `Gemini 3 Flash` and `GPT-5.6 Terra`, respectively. Go back.

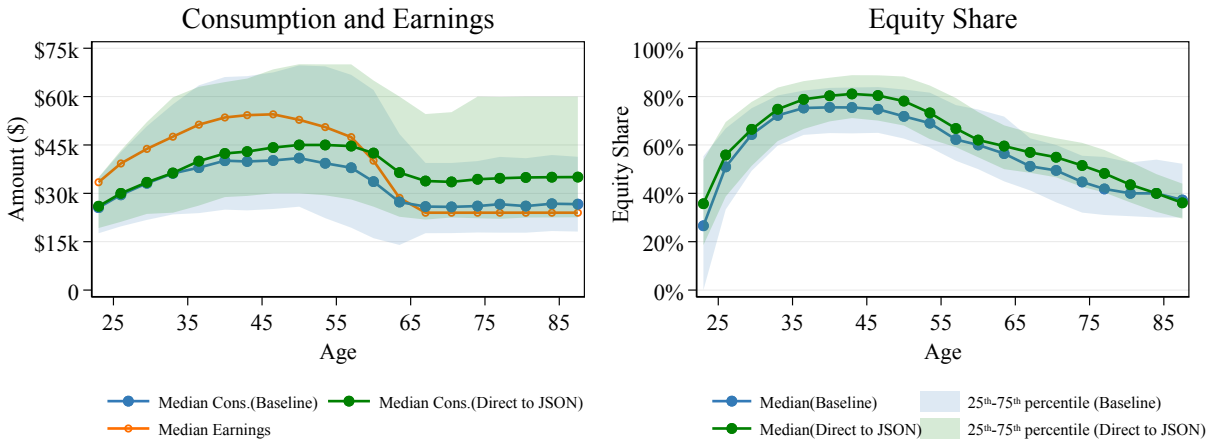
Figure E11. Responsiveness of LLM Advice to Shocks: Academic Prompt



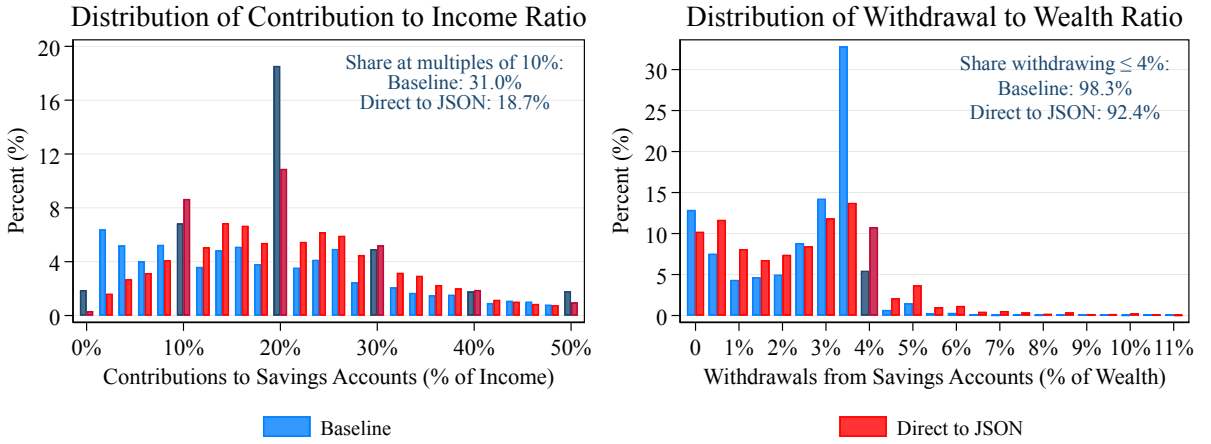
Notes: This figure compares the responsiveness to shocks between the LLM advice and the life cycle model using the academic prompt described in Section 1.4. Red triangles denote the LLM advice, blue triangles denote the life cycle model results, and the orange dashed line indicates the post-tax earnings path implied by the life cycle model described in Section 1.2. The left panel presents a binscatter plot of average actual portfolio change against passive portfolio drift. Passive portfolio drifts are grouped into 20 equally sized bins. The right panel plots the unemployment event study, constructed in the same way as the right panel of Figure 8. Go back.

Figure E12. Baseline Two-Step Pipeline vs. Direct Quantitative Pipeline

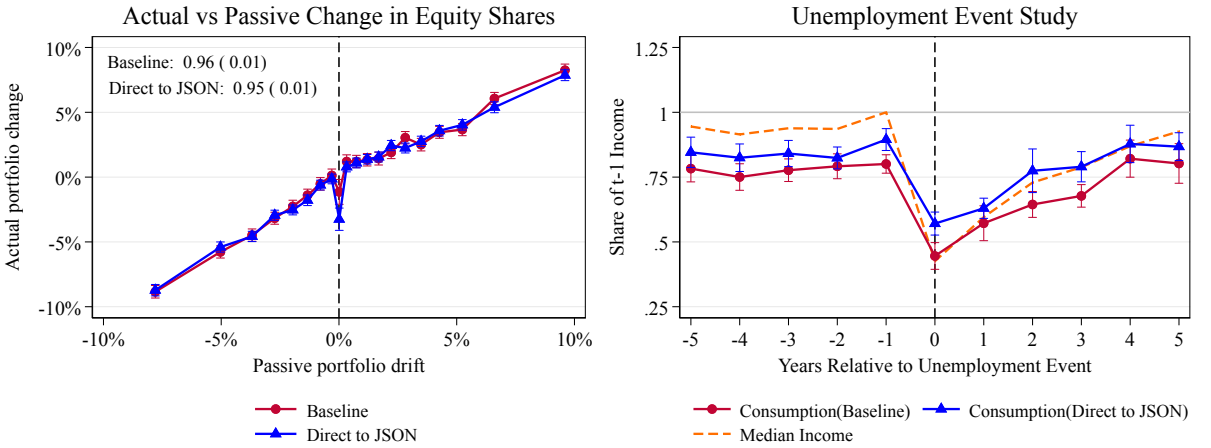
Panel A: Consumption and Equity Share Life Cycle Profiles



Panel B: Saving and Withdrawal Heuristics

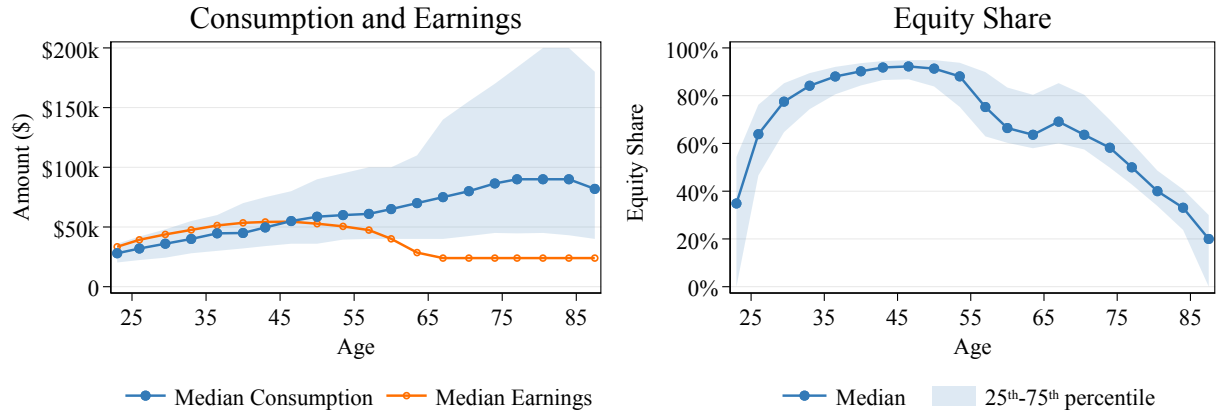


Panel C: Shock Response



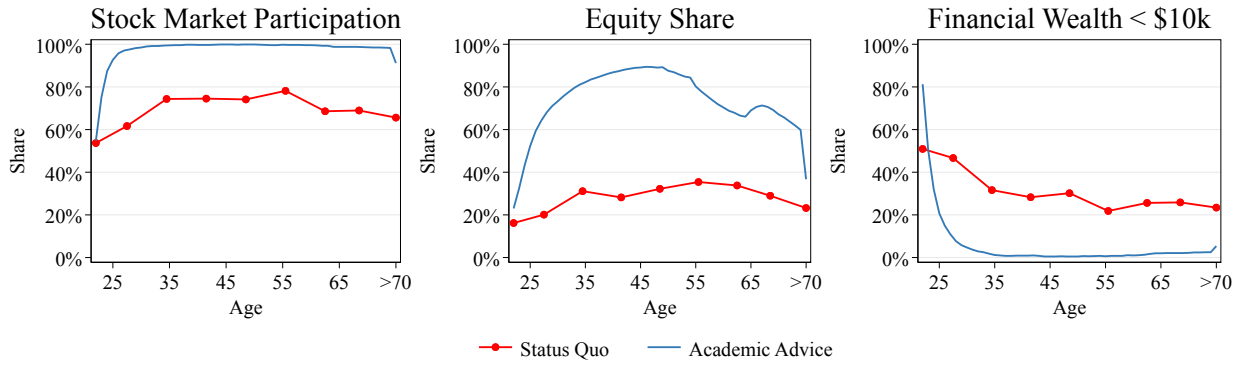
Notes: This figure compares our main results under the baseline two-step pipeline with those from the alternative direct quantitative pipeline described in Section 3.4. Go back.

Figure E13. Life Cycle Profiles of LLM-Recommended Consumption and Equity Shares: Academic Prompt



Notes: This figure plots the life cycle profiles for individuals following the LLM's recommended choices using the academic prompt, as described in Section 1.4. Ages (22–89) are grouped into 20 equally sized bins. The left panel shows consumption and post-tax earnings, and the right panel shows the equity share. Dots denote median values within each bin, and shaded areas represent the interquartile range (25^{th}–75^{th} percentile). All values are in 2025 dollars. Go back.

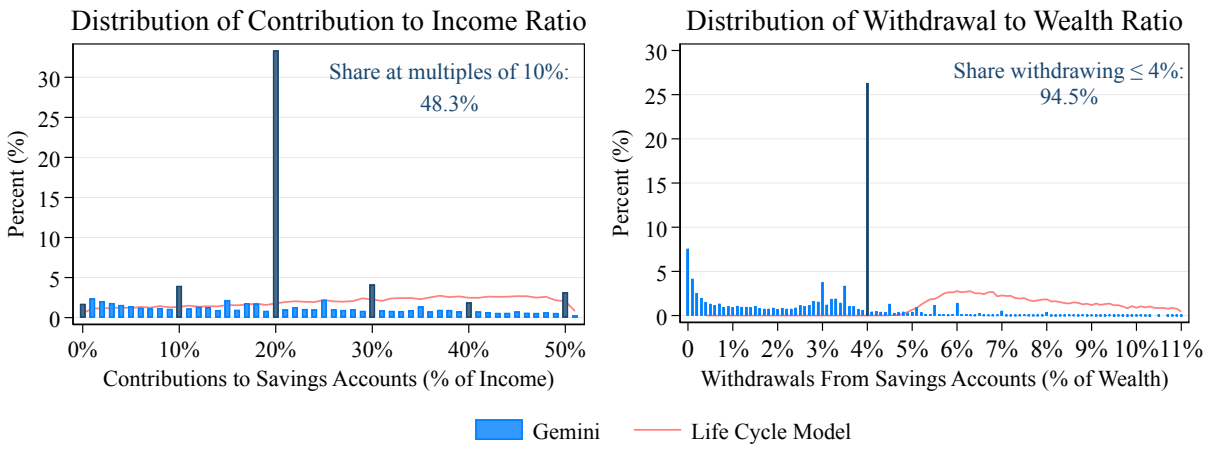
Figure E14. Observed Behavior vs. LLM-Recommended Behavior: Academic Prompt



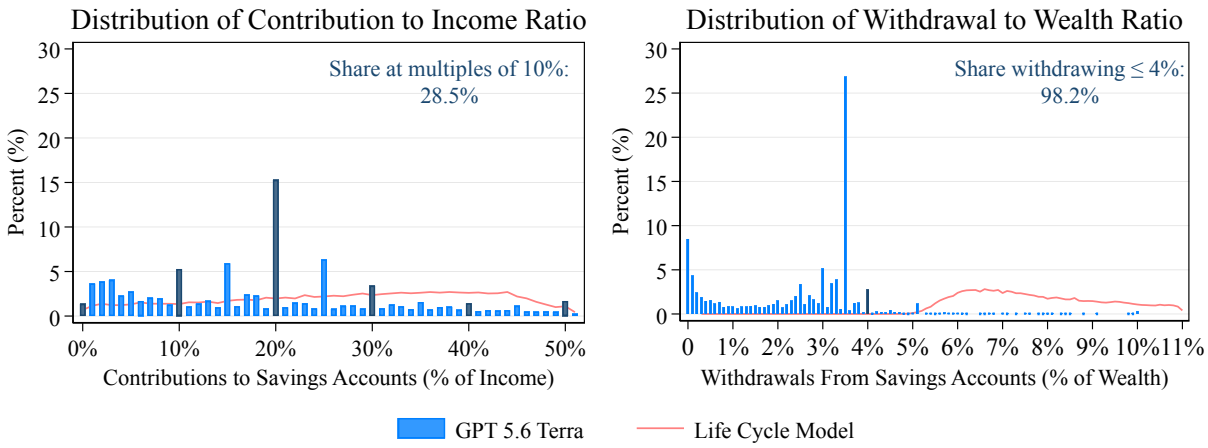
Notes: This figure compares survey respondents' observed financial behavior with the LLM's recommendations using the academic prompt described in Section 1.4. Observed behavior, as reported by individuals in our Prolific survey, is shown in red; recommendations using the academic prompt are shown in blue. Ages are grouped into eight 6-year bins and one last bin that captures every observation age 70 and above. All values are in 2025 dollars. Go back.

Figure E15. Saving and Withdrawal Heuristics in LLM Advice: Alternative Models

Panel A: `Gemini 3 Flash`



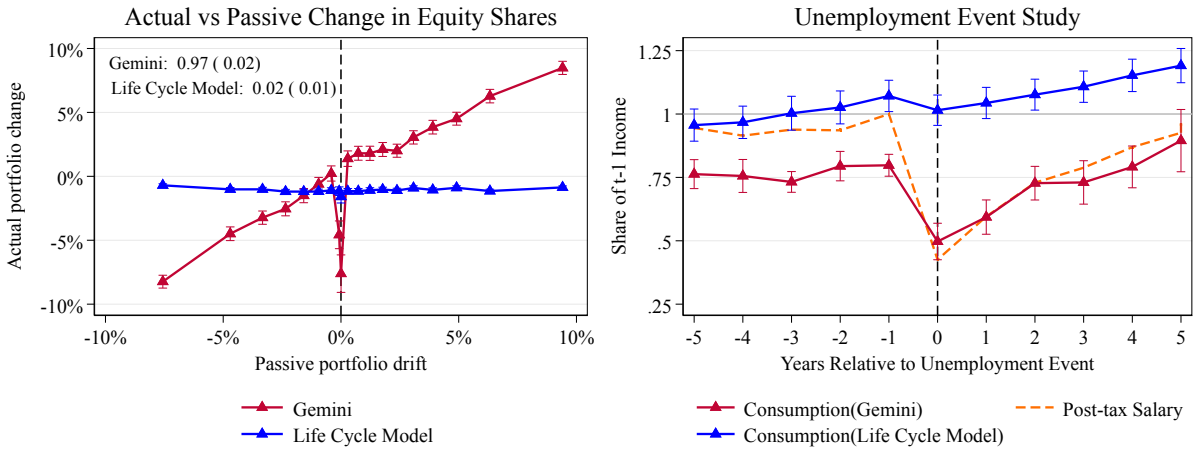
Panel B: `GPT-5.6 Terra`



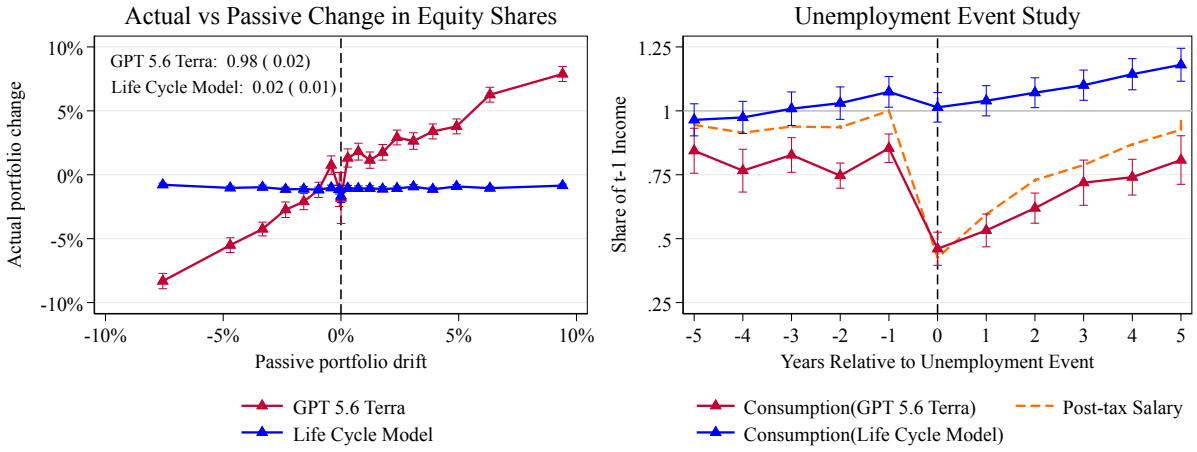
Notes: This figure plots the distributions of saving contributions as a share of income (left) and retirement withdrawals as a share of wealth (right) implied by advice from `Gemini 3 Flash` (Panel A) and `GPT-5.6 Terra` (Panel B). The blue histograms show the choices implied by the LLM advice, dark blue bars highlight mass at common heuristic values, and red lines show the corresponding distributions generated by the estimated life cycle model described in Section 1.2. The LLM advice is generated using the survey prompts as described in Section 1.3 and Section 3.4. Go back.

Figure E16. Responsiveness of LLM Advice to Shocks: Alternative Models

Panel A: `Gemini 3 Flash`

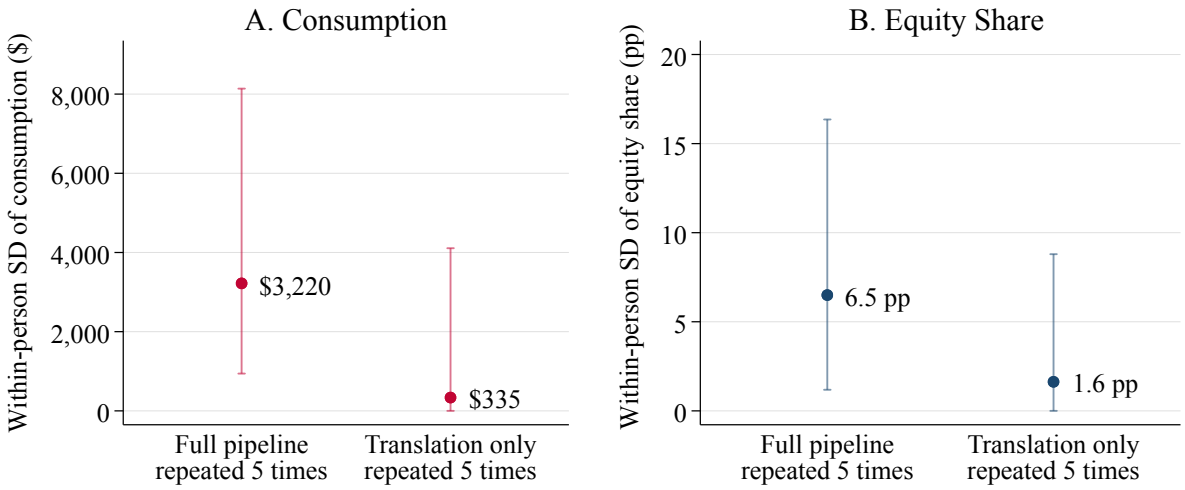


Panel B: `GPT-5.6 Terra`



Notes: This figure compares the responsiveness to shocks of advice from `Gemini 3 Flash` (Panel A) and `GPT-5.6 Terra` (Panel B) with the estimated life cycle model. Within each panel, the left plot shows average actual portfolio changes against passive portfolio drift, grouping observations into 20 equal-sized bins of passive drift. The right plot shows the unemployment event study constructed in the same way as the right panel of Figure 8. Red lines denote the LLM advice, blue lines denote the life cycle model, and the orange dashed line denotes the post-tax earnings path implied by the life cycle model described in Section 1.2. Go back.

Figure E17. Variation in Consumption and Equity Share Across LLM Queries



Notes: This figure measures the variability of the one-shot recommendations described in Section 3.1. We conduct two exercises for each of the 952 respondents. In the "Full pipeline" exercise, we hold the respondent's prompt fixed and repeat the entire procedure five times: each repetition generates new textual advice and then translates that advice into quantitative recommendations. In the "Translation only" exercise, we hold both the prompt and the textual advice fixed and repeat only the translation into quantitative recommendations five times. For each individual and exercise, we calculate the standard deviation across the five recommendations. Dots report the median of these individual-level standard deviations, and whiskers report the 25th and 75th percentiles. Panel A reports variation in recommended annual consumption in dollars. Panel B reports variation in the recommended equity share in percentage points. Go back.

Figure E18. Differences in Prompts and Advice by Financial Literacy

Panel A: Words Overrepresented in Prompts by Financial Literacy

High Financial Literacy | **Low Financial Literacy**



Panel B: Words Overrepresented in Advice by Financial Literacy

High Financial Literacy | **Low Financial Literacy**



Notes: This figure compares differences in respondent-written prompts and corresponding LLM advice by financial literacy. Panel A pools the three prompt questions and shows words overrepresented in prompts written by respondents with high financial literacy (left; 5 of 5 Big Five questions correct) and with low financial literacy (right; 0–4 correct). Panel B shows words overrepresented in the corresponding LLM advice for respondents with high financial literacy (left) and with low financial literacy (right). Word size is proportional to the difference in word frequency between groups. Common stopwords and words that mechanically reflect the survey questions are excluded. Go back.

Figure E19. Example Prompts by Financial Literacy

High Financial Literacy User:

I am a short term futures trader usually executing **3 to 4 trades per daily session**. I do not hold any positions after the market closes. I **trade the futures micro contracts of indexes, crude oil, gold, currencies and bonds**. I also place option trades on the SPX. I am looking to imptove my strategy which is based on Cumalative Delta and Orderflow. My account size is 10,000 USD

I'd like to know how much to spend over the coming year.

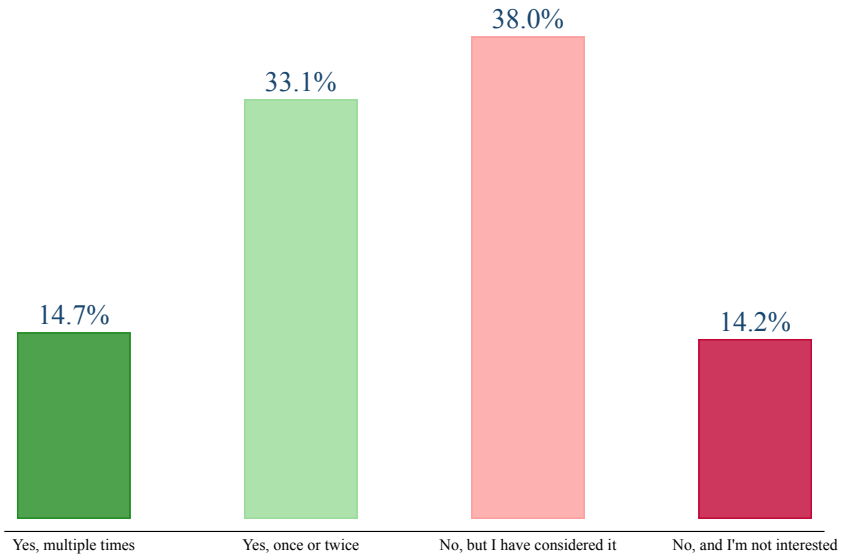
I'd like to know how to invest to improve my strategy and reduce my risk exposure.

Low Financial Literacy User:

Hello….Im in need of some financial advice. I need to invest some money into something that will make me money in the future. […] I have current stocks invested, however, **I know very little about stocks** and how and when to buy/sell, and **what stocks I need to buy and/or avoid**? Some advice would be very helpful! Thank you! :)

Notes: This figure presents examples of respondent-written prompts from individuals with high and low financial literacy, as measured by the Big Five financial literacy questions in the survey. Respondent text is reproduced verbatim; spelling and grammar errors are preserved. Red text highlights phrases related to financial knowledge and investment sophistication. Go back.

Figure E20. Prior AI Use for Financial Advice Among Survey Respondents



Notes: This figure reports responses to the survey question, "In the past 3 months, have you used an AI tool to get financial advice or information?" in the Prolific survey. Bars show the share of respondents selecting each response option. Go back.

Figure E21. Example Prompts by Prior AI Experience

AI User:

Assume you are a CFA with 20+ years of experience. I am needing advice to help create a sustainable investment portfolio using E*trade or similar self-select investment platforms. I have $5K to invest currently. I have another $10k in savings and am already investing in an employer sponsored retirement account contributing the max allowing amount.

Using your CPA knowledge, and our investment strategy determined previously, how much would I bel able to continually spend to add to my investment portfolio?

What additional investment vehicles should I be looking add for my further spending? Should reinvest in high return vehicles or expand and diversify?

AI Non-User:

I would like the AI tool to create a budget for my new pay rate as I am starting a new job. I will be making $20 an hour and working 40-45 hours a week. I pay $1000 in rent & around $500 in utilities/ household need bills.

How much could I spend freely with this salary while maintaining funding for rent, household needs, & $200 savings a month?

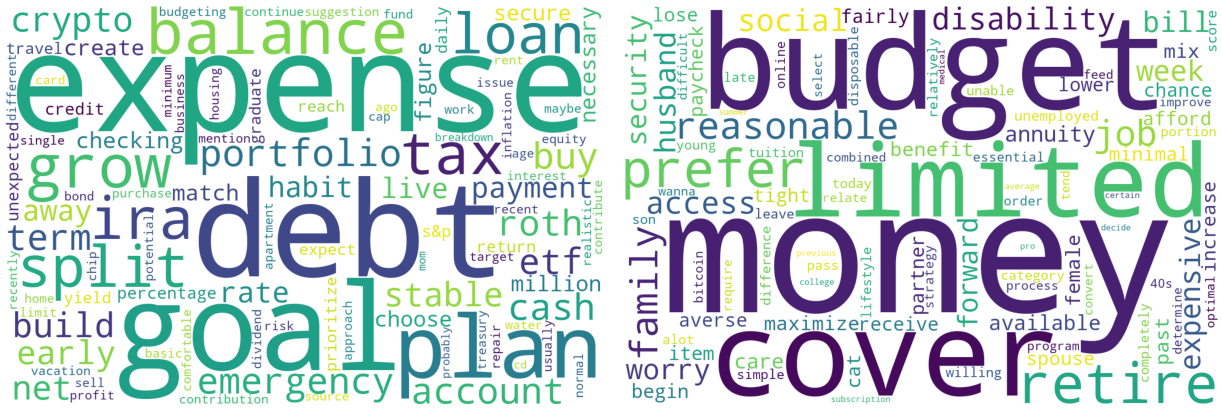
Where should I invest starting with $50 & consistently adding $25 a month after?

Notes: This figure presents examples of respondent-written prompts from individuals with and without prior AI use for financial advice. Respondent text is reproduced verbatim; spelling and grammar errors are preserved. Red text highlights phrases that directly reference AI or expert-advisor framing. Go back.

Figure E22. Differences in Prompts and Advice by Prior AI Experience

Panel A: Words Overrepresented in Prompts by Prior AI Experience

Prior AI Use | **No Prior AI Use**



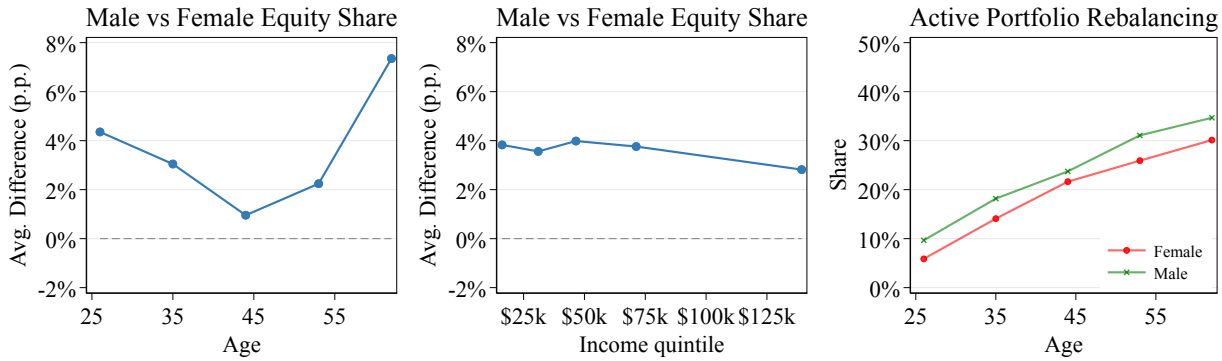
Panel B: Words Overrepresented in Advice by Prior AI Experience

Prior AI Use | **No Prior AI Use**



Notes: This figure compares differences in respondent-written prompts and corresponding LLM advice by prior AI use for financial advice. Panel A pools the three prompt questions and shows words overrepresented in prompts written by respondents with prior AI use (left) and without prior AI use (right). Panel B shows words overrepresented in the corresponding LLM advice for respondents with prior AI use (left) and without prior AI use (right). Word size is proportional to the difference in word frequency between groups. Common stopwords and words that mechanically reflect the survey questions are excluded. Go back.

Figure E23. Gender Gap in Recommended Equity Shares and Portfolio Rebalancing



Notes: This figure examines gender differences in LLM-recommended equity shares and in active rebalancing. The left panel plots the average difference in recommended equity shares between men and women by age. The middle panel plots the same difference by income quintile; the x-axis labels give the dollar cutoffs for the five quintile bins. The right panel plots the share of recommendations involving active portfolio rebalancing, rather than passive drift, separately for women and men by age. Age and income are each grouped into five bins. Go back.