CERGE - EI
Center for Economic Research and Graduate Education
–
Economics Institute

Behavior and Complexity in Household Finance

Marta Cota

Dissertation

Prague 2024

Dissertation Committee:
Marek Kapička (CERGE-EI, Chair)
Ctirad Slavík (CERGE-EI)
Filip Matějka (CERGE-EI)

Referees:
Kathrin Schlafmann (Copenhagen Business School)
Fabrizio Mazzonna (Università della Svizzera Italiana)

Contents

Acknowledgements

I want to express my gratitude to my advisors, Marek Kapička and Ctirad Slavík, and my hosts during my research visits, Kathrin Schlafmann, Mariacristina De Nardi, and Jaroslav Borovička, for their support and guidance throughout my Ph.D. studies. I thank the faculty at CERGE-EI for their comments, advice, and answers to questions I had.

My sincere thanks to Marta Morazzoni for her constant and unconditional support. I also appreciate Lidia Cruces, Nicolò Russo, and Katherina Thomas for their time, advice, and patience during my journey—it wouldn't have been the same without them.

I thank my mom, dad, sister, and friends in Croatia and at CERGE-EI for their understanding, long phone calls, and last-minute meet-ups over the years. Finally, a special thank you to my partner and co-author, Ante, for everything. This is for you.

This output was supported by the NPO 'Systemic Risk Institute' number LX22NPO5101, funded by European Union – Next Generation EU (Ministry of Education, Youth and Sports, NPO: EXCELES).

Prague, Czech Republic Marta Cota

July, 2024

Abstract

This thesis explores mechanisms underlying differences in financial decision-making, including retirement saving and mortgage choice. Chapter 1 investigates the long-term effect of extrapolative expectations on retirement savings in subsidized employer-matched 401(k) accounts in the U.S. The chapter introduces a deviation from rational expectations in reproducing worker-level retirement contribution rates over their tenure. Chapter 2 generates a novel U.S. data set and examines the correlation between financial literacy, mortgage rate attainment, and refinancing. Empirical estimates motivate the introduction of financial literacy and the level of search effort as dimensions of heterogeneity that generate differences in mortgage repayments. I find that losses from low financial skill levels and search effort amount to almost 10 percent of the total loan, implying significant effects on the budgets of financially unskilled and inexperienced U.S. households. Chapter 3 leverages these empirical findings and embeds a micro-founded mortgage search framework in a standard heterogeneous agent model of consumption and saving. Financially skilled agents face lower cognitive search costs and thus explore more options, and ultimately achieve lower mortgage rates. Conditional on assets and productivity, consumption choices differ based on expected mortgage rate changes and mortgage uptake. The model quantifies the effects of financial education and mortgage accessibility and suggests that the effectiveness of financial education increases when mortgages are highly accessible. Chapters 2 and 3 underscore the importance of cost-effective access to financial planning information amidst the increased mortgage accessibility in the U.S.

Introduction

These three chapters explore underlying mechanisms in individual financial decision-making regarding retirement savings and mortgage choices. The first chapter focuses on differences in individual income expectations through the lens of extrapolative behavior. Extrapolative expectations give rise to optimism and pessimism effects on saving in illiquid retirement accounts. Analytical findings of a stylized model prove that pessimists prefer to save in liquid accounts that they can tap into at any time. The full life cycle model differentiates between rational and subjective worker savings accumulation. The full model incorporates retirement contribution incentives, including employer rate matches and tax deferrals in 401(k) accounts in the U.S. Even though these incentives sustain stable rates over the rational worker's life, extrapolative workers delay their contributions initially and gradually increase them over their working life, matching the contribution rates data. A policy test shows that mandating automatic enrollment into 401(k) accounts increases saving rates among extrapolative workers at the beginning of their tenure. Long-term effects decrease because workers offset their initial contribution rates by decreasing them later on in work life.

The second chapter, co-authored with Ante Šterc, is an empirical study that aims to explain differences in mortgage rate attainment among otherwise similar borrowers in the U.S. The paper provides novel insights into the interaction between individual financial literacy and loan shopping behavior and its effect on the mortgage interest rate in the U.S. market. We merge two publicly available U.S. data sets and employ statistical methods that account for the uncertainty in the merging procedure. The merged data set contains mortgage and borrower characteristics, followed by survey responses on loan shopping behavior and objective individual financial literacy measures. First, we find that financial literacy changes with age and exhibits a hump-shaped life cycle profile. Second, we find that financial literacy and search effort interaction explains a part of the mortgage variation among otherwise similar borrowers. Specifically, financially skilled borrowers considering multiple lenders attain 13.4 b.p. lower mortgage rates at origination. This mortgage rate spread translates to over \$9,329 higher payments for a \$100,000 loan over the mortgage term. We also show that the interaction coefficient increases over the 2014-2020 period, simultaneously with a steady increase in non-bank lenders in the U.S. mortgage market. Third, our findings suggest that three years after the mortgage originated, financially unskilled borrowers are 35-45% more likely to become delinquent. The paper motivates the mortgage search model in the third chapter of this thesis.

The third chapter, co-authored with Ante Šterc, embeds a micro-founded mortgage search framework in an otherwise standard heterogeneous agents model of consumption and saving. To understand how financial education, more accessible mortgages, and mortgage rate changes affect households with low financial literacy, we formulate and calibrate a heterogeneous search frictions model with endogenous financial skills accumulation. In the model, search costs are cognitive and are contingent on the individual level of financial skills. In this regard, financial skills and search interact and deliver a new dimension of consumption heterogeneity, owing to differences in mortgage repayments. Calibrated to the data set defined in the second chapter, model estimates show that search intensity and financial skill variations contribute to 55% and 10% of mortgage rate variations, respectively. We find that *i*) more accessible mortgages lead to a higher risk of delinquency among low-skilled households, *ii*) financial education mitigates the adverse effects of increased accessibility, and *iii*) low mortgage rates favor high-skilled homeowners and, by reinforcing refinancing activity, deepen consumption differences across different financial skill levels.

1 Extrapolative Expectations and Retirement Savings

1.1 Introduction

Income expectation biases arising from pessimism or optimism may affect the extent of saving, resulting in lower savings rates than rates predicted by the standard rational model of wealth accumulation. This paper investigates the effects of income expectation bias on retirement savings rates over workers' careers. I build a structural life-cycle model with income forecast biases that generates patterns of retirement savings plan contributions during a worker's career. The model mechanism works through the interplay between income level and volatility misperception. While all workers overstate income volatility, future income level is either overstated or understated, resulting in delays in retirement contributions.

The structural life-cycle model relies on three findings in the expectation data of the Michigan Survey of Consumers. First, income forecast bias changes sign across the income distribution, moving from pessimistic low-income to optimistic high-income households. I extend the Michigan Survey of Consumers' data analysis in Schlafmann and Rozsypal (2023) and show that households extrapolate, basing their income predictions on previous income realizations. Next, I show that their income forecast errors decrease as households age. Lastly, I show that households overstate the probability of losing a job, regardless of age and education level. While all households overstate persistent income volatility, their income growth expectations differ, separating pessimists from optimists. Over time, however, the income level biases decrease, leaving out additional precautions due to overstated income volatility.

Using the retirement contribution data in the Survey of Consumer Finances, I show that low-income workers' liquid-to-retirement savings ratio decreases over the life cycle, albeit slower than high-income workers. With age, workers reallocate their savings to illiquid savings accounts, including retirement savings. In line with these findings, Parker et al. (2022) find steadily increasing contribution rates over the life-cycle across all cohorts.

The life-cycle model solution shows that extrapolative workers follow a pattern of gradual increase in contributions over the work life, owing to the interplay between income level and volatility misperception. Whereas rational agents respond to contribution incentives, including employer matches and tax deferrals, expectation biases understate the benefits of unrealized gains in retirement accounts. Expectations-driven savings paths imply greater retirement savings inequality. In addition, accommodating persistent contribution rates across cohorts (Parker et al., 2022), I show that retirement plan reforms do not affect contribution rates with extrapolative households.

I motivate the extrapolation mechanism analytically in a stylized three-period model. The mechanism works through the fear of being borrowing-constrained by persistently low income in the near future. Pessimists contribute less to their illiquid retirement accounts and allocate their savings to liquid accounts they can tap into. On the other hand, optimists postpone their retirement contributions due to optimistic income expectations.

I reinstate the mechanism and develop a structural life-cycle model with extrapolative expectations and two types of assets, built on Schlafmann and Rozsypal (2023). Workers face persistent and transitory income shocks and choose when to start and how much to contribute to their private retirement plans, and how much to keep in liquid savings accounts. Retirement plans are illiquid and imitate private retirement accounts (such as the 401(k)) in the U.S., and therefore include employer's match and tax deferrals. As this paper focuses on the intensive margin, all workers in the model are eligible to participate in the retirement plan.

Using the Method of Simulated Moments, I calibrate the income-forecast misperception to match survey responses from the Michigan Survey of Consumers. The trade-off between saving for the near future and saving for retirement differs from the rational benchmark case. Pessimistic workers are not willing to forego their liquid buffers and hence save less in retirement savings, regardless of the forms of their incentives.[1] On the other hand, optimistic workers postpone their contributions relative to rational counterparts. However, over the work life, both types of workers catch up with retirement savings by increasing contribution rates towards retirement, reflecting firm-level data findings (Parker et al., 2022; Choukhmane, 2021).

Whereas the rational benchmark predicts decreasing liquid saving rates across wealth percentiles, my model solution shows that liquid savings from income remain flat after the 20th wealth percentile, consistent with empirical findings (Fagereng et al., 2019; De Nardi and Fella, 2017). Extrapolation-driven savings paths generate greater retirement savings differences, facilitating the importance of unrealized capital gains in illiquid retirement accounts. Wealthier workers contribute more to retirement plans throughout the work lifespan and when capital gains are finally realized, once workers retire, they consume at significantly higher levels. Rational benchmark produces better responses to saving incentives, restraining retirement savings inequality across the wealth distribution.

After aligning contribution patterns with the extrapolative expectations solution, I test the implications of the 401(k) automatic enrollment policy, recommended by the U.S. government in the employer guide[2]. All employees are enrolled in a plan with automatically, and employers match their employees' contributions up to a certain level (typically, up to 3%). Model simulations show that workers contribute less when approaching retirement and offset their higher contribution rates at the start of their working lives. As a result, adjustments to retirement account enrollment do not result in substantial welfare gains.

Voluntary participation savings paths with extrapolative expectations serve as a baseline for worker behavior and show that automatic participation policy correlates with findings in event study estimates (Choukhmane, 2021; Goda et al., 2020). Initial contribution rates tend to be relatively high but decrease as workers offset the distortion on impact. Policy tests call for novel retirement system adjustments (such as auto-escalation policies) that account for catching-up behavior. Including auto-escalating employer matches may cause additional delays in saving for retirement through 401(k) accounts. I leave it for future research.

[1] I included tax deferrals and benefits functions of different kinds.

[2] Published Guidance from the IRS can be found at https://www.irs.gov/retirement-plans/published-guidance.

1.2 Related literature

Studies by Grevenbrock et al. (2021) and De Nardi et al. (2009) use life expectancy biases to motivate retirement saving decisions. In both of these studies, the consumption paths of assets during retirement are closer to the data than in standard rational expectation models. However, saving for retirement through retirement contribution accounts adheres to the life-cycle income path and other savings decisions from the start of the work life. Duarte et al. (2021) build a rational expectations model with retirement saving portfolio allocation and find that almost all young workers add to their retirement accounts and invest their savings primarily in equity[3]. Contrary to default options, rational workers opt for equity funds and face significant losses over the life cycle.

In contrast to the rational expectations portfolio choice solution, empirical studies that use firm-level data find dominantly low contribution rates and default fund choices (Blanchett et al., 2021; Parker et al., 2022). Contributions are low a few years after the tenure begins, (Choi et al., 2003; Choukhmane, 2021; Devlin-Foltz et al., 2015). As a result, modeling retirement plan contributions includes money (opt-out) costs as a tool for saving for retirement to match the retirement savings patterns in the data (Choukhmane, 2021; Dahlquist et al., 2018; Love, 2006). Moment targeting produces opt-out that can be immensely high (DellaVigna, 2018).

This paper adds to retirement contribution studies using a behavioral assumption that reconciles the dynamics of contribution patterns over the work-life. The data analysis builds on that in Schlafmann and Rozsypal (2023) and sets the ground for the structural life-cycle model. Schlafmann and Rozsypal (2023) focus on future income expectations and find that people tend to overestimate the persistence of their future income. I find evidence suggestive of extrapolation, adding to evidence from panel data in the U.K. (Cocco et al., 2022) and the Netherlands (Massenot and Pettinicchi, 2019). Moreover, data estimates on subjective unemployment probabilities overstate the actual job loss probabilities across the whole sample population, adding to increased precaution documented in the data Blundell et al. (2008). Correspondingly, an increase in idiosyncratic risk shifts beliefs towards pessimism, capturing the mechanism outlined in the theoretical model of expectation bias in Bhandari et al. (2016).

The model in this paper connects expectation biases to workers' retirement plan contributions in the U.S. Recent data findings in Ghilarducci et al. (2018) show that workers extrapolate from their recent past and adjust their savings to ensure living standards. Moreover, Goda et al. (2020) show that contribution behavior varies significantly with financial literacy, while behavioral biases such as present bias and exponential growth bias remain insignificant. Similarly, I highlight the effect of understanding one's income on yearly contribution rates throughout their career.

Aligning early retirement contribution behavior often necessitates behavioral assumptions implying passive behavior, i.e., adding at default rates [4] (Bernheim et al., 2015; Ameriks

[3]Once the worker chooses how much to contribute, she can opt for a type of fund to invest in: equity, bonds or a mixed target-date (TDF) fund. TDF is usually set as a default option in retirement plans.

[4]The default rate is either set to 0% or 3%, depending on the enrollment regime.

et al., 2007; Benartzi and Thaler, 2007). In a dynamic setting, studies incorporating time-preference biases cannot reconcile the retirement contributions found in the data and often turn to contribution adjustment costs (Choukhmane, 2021; Dahlquist et al., 2018). However, cost estimates are large and amount to thousands of dollars each year, discussed Choukhmane (2021) and DellaVigna (2018). I contribute by incorporating expectation patterns that distort individual decisions over their work life. The mechanism in the structural model implies staggered contributions early in the work-life across the income distribution. Moreover, the bias assumption tracks with staggered contributions throughout the work life.

Separating extrapolation from other household characteristics adds to the growing behavioral finance literature. Experimental studies Krijnen et al. (2022) and Goda et al. (2020) argue that correcting workers' expectations regarding savings growth may increase contribution rates. As time inconsistency and other psychological biases require experimental data, this paper contributes to the literature by explaining passive behavior in retirement saving, based on public survey data estimates.

Ultimately, connecting income forecast errors to saving choices steps out of the finance literature that connects portfolio choices and future returns extrapolation (Bordalo et al., 2018, 2019). Adhering to investors' behavior, a new line of research uses extrapolation to account for heterogeneity in the marginal propensity to consume, making income expectation biases relevant for fiscal stimulus (Auclert et al., 2020; Choi and Mertens, 2019). In this respect, agents adding to 401k accounts often do not know a lot about the specifics of the fund stocks-to-bond ratio. Therefore, my model in this paper highlights the difference between rational and subjective behavior, focusing on the liquid-illiquid saving ratio based on differences in future income expectations.

Sticking to liquid savings tools due to pessimism based on past low-income realizations translates to a negligible reaction to retirement plan adjustments for the bottom part of the income distribution. This explains persistent contribution rates across cohorts Parker et al. (2022) and ambiguous effects of retirement plan reforms (Choukhmane, 2021; Beshears et al., 2022; Bernheim et al., 2015). Similarly, the policy exercise with extrapolative workers finds insignificant effects of automatic enrollment on average retirement savings.

1.3 Data

1.3.1 Bias in the income growth forecast

The structural life-cycle model hinges on income expectation patterns in the Michigan Survey of Consumers data. My data analysis builds on and adds to previous findings in Schlafmann and Rozsypal (2023) and Das and Van Soest (1999). I incorporate multiple survey questions to establish useful facts relating to future income misperception and retirement confidence across the survey sample. Income mean and volatility forecast errors across household characteristics exhibit patterns that align with the subjective income process assumption.

First, I reiterate the findings in Schlafmann and Rozsypal (2023), who show that linear regression estimates highlight the importance of a worker's position in the income distribution. Individual income level defines a worker as pessimistic or optimistic. I add two lines of estimates to the MSC analysis in Schlafmann and Rozsypal (2023). First, I employ the logistic regression model on subjective probabilities of income increase and find evidence for extrapolative expectations, corroborating other income expectations analyses (Massenot and Pettinicchi, 2019; dHaultfoeuille et al., 2021). Second, using subjective unemployment probabilities as a proxy for perceived income volatility, I argue that workers overstate job separation rates across all education levels and age groups.

1.3.2 Constructing expectation errors at the household level

Since households are re-interviewed (once) in the MSC survey, I follow the approach in Schlafmann and Rozsypal (2023) and estimate the bias in income growth expectations using the questions

1. *"During the next 12 months, do you expect your (family) income to be higher or lower than during the past year? "*
2. *"By about what percent do you expect your (family) income to (increase/decrease) during the next 12 months?"*

and then compare the answers to realized income responses in the next survey wave. The advantage of the MSC survey is that it asks respondents to specify their income growth forecast percentage from 1986-2012. In the second interview, households are asked to report their last-period income, which may be subject to individual measurement error[5]. Since I cannot attain the objective last-period income value, I use household characteristics to compare reported income with official income statistics to check for robustness.

After denominating income values to 2010 dollars, I evaluate expectation errors using short panels by tracking the household during their re-interview

$$\phi_{i,t} = \frac{\Delta\widehat{Y}_{i,t+1}}{Y_{i,t}} - \frac{\Delta Y_{i,t+1}}{Y_{i,t}} = \widehat{g}_{i,t+1|t} - g_{i,t+1},$$

where $g_{i,t+1}$ is the self-reported real income growth for the previous period (previous year) and $\widehat{g}_{i,t+1|t}$ is the expected income growth. The sign of the error implies pessimism or optimism. If the error is positive, the worker expected a higher income than was realized. The negative error thus implies a pessimistic future income outlook.

The analysis does not include household incomes lower than the estimated unemployment benefits aggregated yearly. Moreover, the sample includes households with no change in socioeconomic characteristics, such as family structure and education. Ideally, households would be responding to survey questions for two consecutive years. However, restricting the sample to respondents re-interviewed during the subsequent year does not change the regression estimates. There are 47,000 re-interviewed households, from which 30,000 respondents

[5] Schlafmann and Rozsypal (2023) give a detailed explanation regarding the survey data and a comparison to other surveys.

gave their first response in June. Specific questions regarding job uncertainty and retirement confidence came later in the MSC survey; hence, sample sizes vary from 20,000 to 37,000, depending on the question.

1.3.3 Income forecast error distribution estimates

Figure 1 shows the difference between forecast error distributions for each income quantile. The mean of the error distribution shifts from left to right, implying a shift from negative to positive forecast errors or from pessimism to optimism. That is, on the low end of the income distribution, workers expect their income to be lower than it actually is. In contrast, high-income workers are more optimistic. In addition, a distribution tail comparison shows a larger mass of pessimists in the first and second income quantiles. Therefore, error distributions shift gradually.

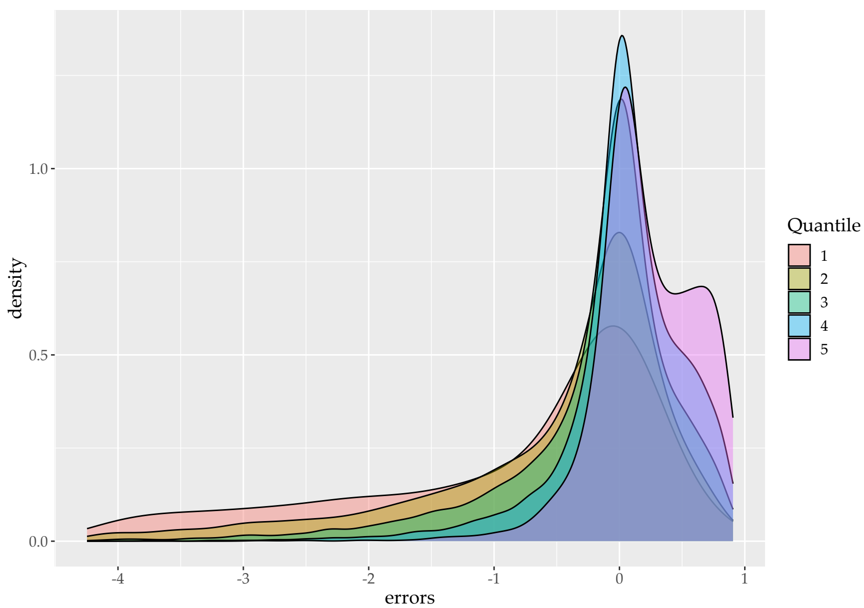


Figure 1: Income growth errors change sign when they move from lower to higher quintiles, MSC data.

1.3.4 Linear regression results

While non-parametric estimates show differences in the error distribution across income quantiles, the linear model estimates control for other household characteristics. The income quantile remains a significant predictor for forecast error while controlling for household characteristics. Estimates include month and year effects and limit the education variable to only three possibilities, clearly distinguishing between high school graduates, college graduates, and those with less education. Income quantile coefficients are significant and large relative to other household characteristics (Table 1).

The majority of regressors are indicator variables, whereas *age* and age^2 are re-scaled, following Gelman (2008). This way, the regression model does not lose interpretability, and the standard errors are not downward biased [6]. As a result, age becomes a significant predictor of the income growth forecast error [7]. Therefore, the structural model incorporates a decrease in the bias later in work life, as a result of deterministic income shape with respect to work experience.

Table 1: Linear Regression Results, MSC data. Source: author's estimates.

	Dependent variable:
	Income Growth Forecast Errors
q_2	0.206***
	(0.008)
q_3	0.286***
	(0.008)
q_4	0.327***
	(0.009)
q_5	0.393***
	(0.014)
male	−0.012*
	(0.006)
Education: no HS	0.039***
	(0.006)
college	−0.046***
	(0.003)
age	−0.156***
	(0.026)
age^2	0.152***
	(0.034)
HH size: 1 *adult*	0.096***
	(0.005)
> 2 *adults*	−0.035***
	(0.008)
Constant	−0.303***
	(0.012)
Observations	47,341
Note: heteroskedasticity robust SE, standardized age.	*p<0.1; **p<0.05; ***p<0.01

[6] Reported standard errors are heteroskedasticity-robust.

[7] Furthermore, estimates of kernel density for separate age bins show that the forecast error decreases in mean and variance. Estimates are given in the appendix.

1.3.5 Probability estimations - extrapolation and overstating income volatility

While forecast error signs are dominantly explained by household income, the fact that households extrapolate from the near past is not obvious. Current empirical studies find evidence of extrapolation in household expectations surveys (Massenot and Pettinicchi, 2019; Ghilarducci et al., 2018). In line with empirical findings, my estimates show that households extrapolate based on their recent income realization (i.e, the income growth error), regardless of their income level.

During the survey, households are asked to assign probabilities to the rise in personal income during the next year and evaluate the likelihood of their five-year job retention:

- *What do you think is the percent chance that your income in the next twelve months will be higher than your income in the past twelve months?*

Using the ordered logistic model, I show that the most recent forecast error explains the worker's outlook on future income growth.

The regression coefficients presented in Table 2 indicate that recent forecast errors affect individuals' expectations regarding future income growth. Specifically, households experiencing larger-than-expected income increases ($\text{error}_{t-1} > 0$) are less inclined to anticipate future income growth (i.e., assign greater likelihood to income increase).

Regardless of the error, high-income workers are more likely to expect income increases in the upcoming year. Therefore, workers do not base their expectations on transitory shocks fully but make inferences from their income history. Overall, this suggests that individuals extrapolate from their recent income history regardless of their income level, but their expectations are also shaped by their earnings paths over their life cycles.

Next, I use the workers' job loss predictions elicited in the MSC responses. Comparison of subjective job loss probabilities to empirical estimates in labor studies show that workers overstate the probability of losing a job, regardless of age or education levels.

I leverage the survey question

- *During the next 5 years, what do you think the chances are that you (or your husband/wife) will lose a job you wanted to keep?*

and retrieve subjective job loss predictions over the subsequent five years, $\widehat{p}_i$. Following Love (2006), I assume a constant year-ahead perceived unemployment probability. A worker's subjective probability of losing a job next year is

$$\widehat{\mathbb{P}} = 1 - (1 - \widehat{p}_i)^{\frac{1}{5}}.$$

I take sample averages across age groups and education level, and compare my estimates with job separation data in Love (2006) and Farber et al. (2005), and I outline them in Table 3. Moreover, I perform a t-test and show that workers significantly overstate their unemployment probability, regardless of their age and education background.

Table 2: Income increase likelihood, ordered logistic regression results, MSC data.

	P(income increase in t | $t-1$)
Region: North Central	-0.080^{**}
	(0.039)
Northeast	-0.032
	(0.043)
South	-0.075^{**}
	(0.038)
age	-0.617^{***}
	(0.033)
errors_{t-1}	-0.298^{***}
	(0.016)
Income quintile: q_2	0.300^{***}
	(0.054)
q_3	0.623^{***}
	(0.056)
q_4	0.708^{***}
	(0.059)
q_5	0.815^{***}
	(0.062)
Male	-0.139^{***}
	(0.028)
Education: No High School	-0.181^{**}
	(0.073)
College	0.214^{***}
	(0.030)
Observations	18,997

Note: Controlled for time and family characteristics. *p<0.1; **p<0.05; ***p<0.01

Table 3: Left: subjective job loss probabilities from the MSC data, right: empirical estimates (Farber et al., 2005; Love, 2006).

$\widehat{\mathbb{P}}$	$ed < 12$	$12 \leq ed \leq 15$	$ed > 15$	$ed < 12$	$12 \leq ed \leq 15$	$ed > 15$
age $25-34$	0.087^{*}	0.074^{***}	0.063^{***}	0.068	0.052	0.035
age $35-44$	0.123^{***}	0.078^{***}	0.065^{***}	0.058	0.043	0.030
age $45-54$	0.115^{***}	0.090^{***}	0.068^{***}	0.053	0.039	0.028
age $55-66$	0.067	0.071^{***}	0.056^{***}	0.057	0.039	0.027

$N = 15676$, NBER recession years not included. t-test based p-values. *p<0.1, **p<0.05, ***p<0.01.

In the model, income process misperception includes overstating income volatility throughout the income distribution[8].

[8] Of course, the income quantile does predict the probability stated in the survey. However, subjective probability is still significantly larger than the true one.

Next, I build on Schlafmann and Rozsypal (2023) and relate the pattern in income growth forecast errors to the life cycle model with extrapolative expectations.

Before going through the retirement contribution data findings, I outline expectation assumptions that generate the error patterns in the MSC data. I relate the pattern in income growth forecast errors in the spirit of Schlafmann and Rozsypal (2023), by assuming the misperception in the auto-regression coefficient λ.

1.3.6 Relating forecast errors to the model

The extrapolative expectations model hinges on expectation patterns in the data. The incorrectly perceived income process separates low-income pessimists and high-income optimists. Throughout their career, agents can change their outlook based on their income history. That is, the subjective life-cycle income process incorporates three key findings in the MSC data:

1. income level expectations transition from pessimistic on the left part to optimistic on the right part of the income distribution
2. persistent income volatility is overstated across all workers
3. workers extrapolate from their recent income realizations.

The structural model assumes that the true income process satisfies

$$Y_{it} = A_i G(t) \mathit{\Gamma}_{it} P_{it}, \quad \log A_i \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2), \quad \mathit{\Gamma}_{it} \sim \log \mathcal{N}\bigg(-\frac{\sigma_{\mathit{\Gamma}}^2}{2}, \sigma_{\mathit{\Gamma}}^2\bigg),$$

where $\log P_{it}$ follows an AR(1) process

$$\log P_{it} = \lambda \log P_{it} + \xi_{it}, \quad \xi_{it} \sim \mathcal{N}(\mu_\xi, \sigma_\xi^2).$$

Given the true income process, assume that agents mispercieve the persistence of their income, regardless of their age or individual effects. This implies

$$\hat{Y}_{it} = A_i G(t) \mathit{\Gamma}_{it} P_{it}^{\hat{\lambda}} \implies \hat{\mathbb{E}}_t Y_{i,t+1} = A_i G(t) P_{i,t}^{\hat{\lambda}}, \tag{1}$$

whereas rational agents know the true income process, so

$$\mathbb{E}_t Y_{i,t+1} = A_i G(t) P_{i,t}^{\lambda}. \tag{2}$$

$\hat{\lambda}$ implies the perceived log-income process

$$y_{it} = \alpha_i + g(t) + p_{it} + \gamma_{it} \text{ and } p_{it} = \hat{\lambda} p_{i,t-1} + \xi_{it}, \xi_{it} \sim \mathcal{N}\Big(\mu_\xi, \sigma_\xi^2\Big), \quad \gamma_{it} \sim \mathcal{N}\Big(\mu_\gamma, \sigma_\gamma^2\Big).$$

The differences in expected and realized income are

$$\begin{aligned} \mathbb{E}_t^*[y_{i,t+T}] - \mathbb{E}[y_{i,t+T}] &= \mathbb{E}^*[p_{i,t+T}] - \mathbb{E}[p_{i,t+T}] \\ &= (\hat{\lambda}^T - \lambda^T)(p_{i,t} - \mu_\xi), \forall T \end{aligned}$$

and change depending on

$$p_{i,t} \lesseqgtr \mu_{\xi}.$$

For a large enough realization of persistent income, the subjective future income is higher than the rational one, i.e., the agent is an optimist. In contrast, if persistent income is sufficiently low, the agent becomes pessimistic and expects lower future income. Given that the persistent component is a sum of all previous income realizations, the worker may change their outlook over their career.

Following the MSC data evidence on unemployment probability pessimism among all workers, the structural model includes the persistent component volatility. The data shows that all workers overstate their persistent income volatility, regardless of age or other characteristics. In model terms, using the AR(1) persistent income process

$$p_{it} = \hat{\lambda} p_{i,t-1} + \xi_{i,t}, \forall t = 1, \ldots, T_{ret} - 1,$$

where $\hat{\lambda} > \lambda$ implies the income growth forecast errors in expression (1), conditional volatility satisfies, $\forall t = 1, \ldots, T_{\text{ret}}$

$$\mathbb{V}_t[p_{t+T}] = \sigma_\xi^2 \frac{1-\lambda^{2T}}{1-\lambda} < \hat{\mathbb{V}}_t[p_{t+T}] = \sigma_\xi^2 \frac{1-\hat{\lambda}^{2T}}{1-\hat{\lambda}},$$

regardless of the previous realization of p_t.

For $T = 1$ and $\hat{\lambda} = \lambda + \varepsilon < 1$

$$\begin{aligned} \hat{\mathbb{E}}_t[p_{i,t+1}] - \mathbb{E}_t[p_{i,t+1}] &= \hat{\lambda} p_{i,t} - \lambda p_{i,t} \\ &= \varepsilon \left(\sum_{s=0}^{t-1} \hat{\lambda}^{t-s} \xi_{i,s} + \hat{\lambda}^t p_{i,0} \right), \end{aligned} \tag{3}$$

Thus, persistent income realizations determine the sign of the difference, separating pessimists from optimists, and corresponding to the infinite horizon outlook in Schlafmann and Rozsypal (2023). Differences in signs align with empirical separation of optimists and pessimists based on the income distribution position and the effect of recent income realization on income outlook. Over the life cycle, the deterministic component outweighs the persistent one when considering the income level. As a result, the model forecast error accounts for heterogeneity over time without imposing ad hoc constraints on the income process.

The two stage calibration of the lifecycle model entails calibrating $\hat{\lambda}$ using the MSC income growth error data. Calibration procedure is explained in detail in the next section. The estimation is robust to income growth error outlier specification and always yields $\hat{\lambda} > \lambda$.

1.3.7 Calibrating the income growth bias

Since the MSC does not include panel data, I use the stand-in values for the income process parameters in the Panel Study of Income Dynamics (PSID) data[9]. The deterministic

[9] Schlafmann and Rozsypal (2023) show that objective income growth rates from the PSID align with patterns in the MSC.

income growth parameters follow estimates in Cocco et al. (2005), satisfying the typical hump shape of workers' income over their life cycle. The stochastic part of the income process contains both transitory and persistent components, so I use the estimates in Storesletten et al. (2004). Specifically, the true persistence parameter is set to 0.972. In this way, true $\lambda = 0.972$ becomes the lower bound for mispercieved $\widehat{\lambda}$ [10]

The parametrized income process

$$y_{i,t} = \alpha_i + \text{const.} + g_1 t + g_2 t^2 + g_3 t^3 + p_{i,t} + \gamma_{i,t}, \quad \text{and} \quad p_{i,t} = \lambda p_{it} + \xi_{i,t} \tag{4}$$

is defined with Each grid element $\widehat{\lambda}$ defines the objective function for income persistence bias

Table 4: Income process parameter values.

σ^2_α	const	g_1	g_2	g_3	σ^2_ξ	σ^2_γ	λ
0.27	-2.1700	0.1682	-0.0323	0.0020	0.0737	0.0106	0.972

calibration. The Method of Simulated Moments minimizes the difference between simulated income forecast errors implied by the current $\widehat{\lambda}$ and 4, and the empirical income growth forecast errors in the MSC data.

For each income quantile, the perceived persistence parameter minimizes the mean forecast error. Each grid member represents the sample's income error forecast. After taking out age effects, the residuals are used as the dependent variable in a linear regression that makes predictions for income growth forecast errors at a given income quantile. This way, simulated residuals correspond to the income forecast error in the data and define a loss function.

The calibration includes 50 000 households with separate income processes. The loss function is

$$L(\hat{\lambda}) = \sqrt{\sum_{i=1}^{5} w_i \left(\text{err}(\hat{\lambda})_{q_i} - \text{err}(\lambda)_{q_i}\right)^2}, \tag{5}$$

where w_i is the weight of a given quantile, and is inversely proportional to the life-cycle variance of the each income quantile.

The optimal $\widehat{\lambda} = 0.99$ minimizes the loss function at the value $L = 0.0007$. Calibration results do not depend on the outlier criteria for empirical forecast error data. Moreover, results do not depend on the choice of the grid for $\widehat{\lambda}$ and yield $\widehat{\lambda} > \lambda$. Most importantly, quantile-based forecast error means change sign from low-income (pessimistic) to high-income (optimistic) quantiles (Table 5).

1.3.8 Retirement contributions data

The model relates income expectations to retirement contribution patterns in the data. Using public datasets, I outline two relevant facts of a contribution plan take-up in

[10]Different estimates in the literature reproduce contribution patterns fairly well. That is, model solution does not depend on the change in the parameter values.

Table 5: Mean income growth forecast error by income quantile for true and the mispercieved persistent income process.

mean error quantile	q_1	q_2	q_3	q_4	q_5
$\lambda = 0.972$	-0.1230	-0.0039	0.0506	0.0831	0.1314
$\hat{\lambda} = 0.99$	-0.0881	-0.0106	0.0255	0.0407	0.0326

the U.S. First, the number of low-income workers eligible for a 401(k) plan is persistently increasing (Figure 2, Bureau of Labor Statistics data). Second, at the same time, low income workers tend to keep their savings in liquid accounts even when approaching retirement (Figure 3, the Survey of Consumer Finances data). Only recently, the U.S. Government instructed employers to allow part-time workers to open up a 401(k) account. Among full-time workers, the number of eligible workers in the BLS has increased over the last couple of years. Nevertheless, only 30% of low-wage workers have access to this type of account (Figure 2).

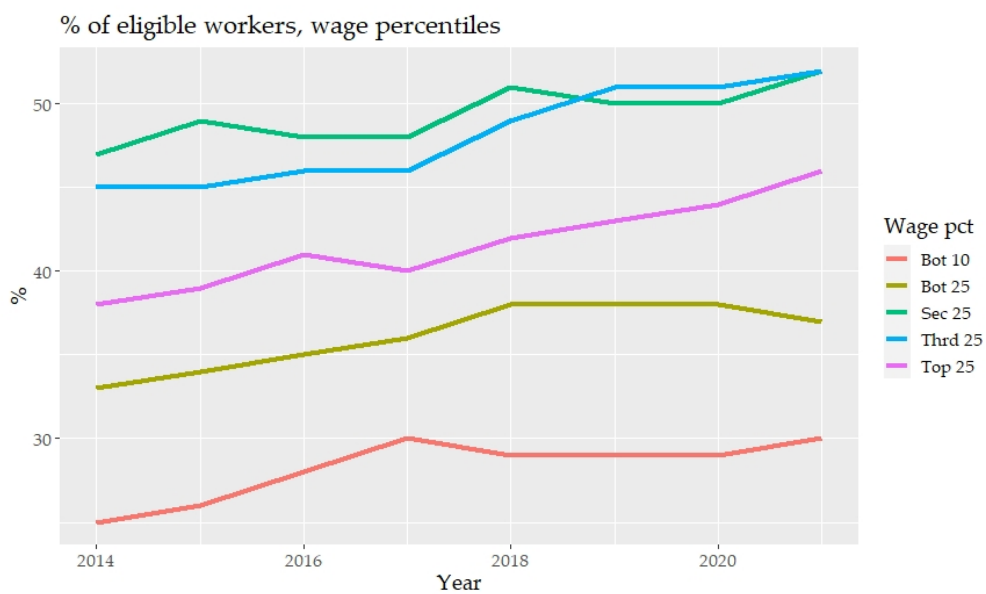


Figure 2: Workers eligible for 401(k), wage quartiles. *Bureau of Labor Statistics data.*

This paper focuses on the intensive margin and considers workers who are eligible to participate. Specifically, the life cycle model aims to capture the differences between rational and extrapolative expectations savings paths, in both liquid and illiquid retirement accounts. Expectation-driven savings paths generate the liquid savings share at the cross section. The paper assesses the performance of each expectations model by comparing liquid savings ratio to it's empirical counterpart. Therefore, the relevant data measure is the share of liquid savings in overall savings accounts, across age groups. Following Bhutta et al. (2022), liquid savings include transaction, checking and savings accounts together with directly held stocks, bonds and other financial assets. Controlling for worker demographic and financial characteristics[11], figure 3 plots the predicted share of liquid savings in all savings accounts against wage percentiles.

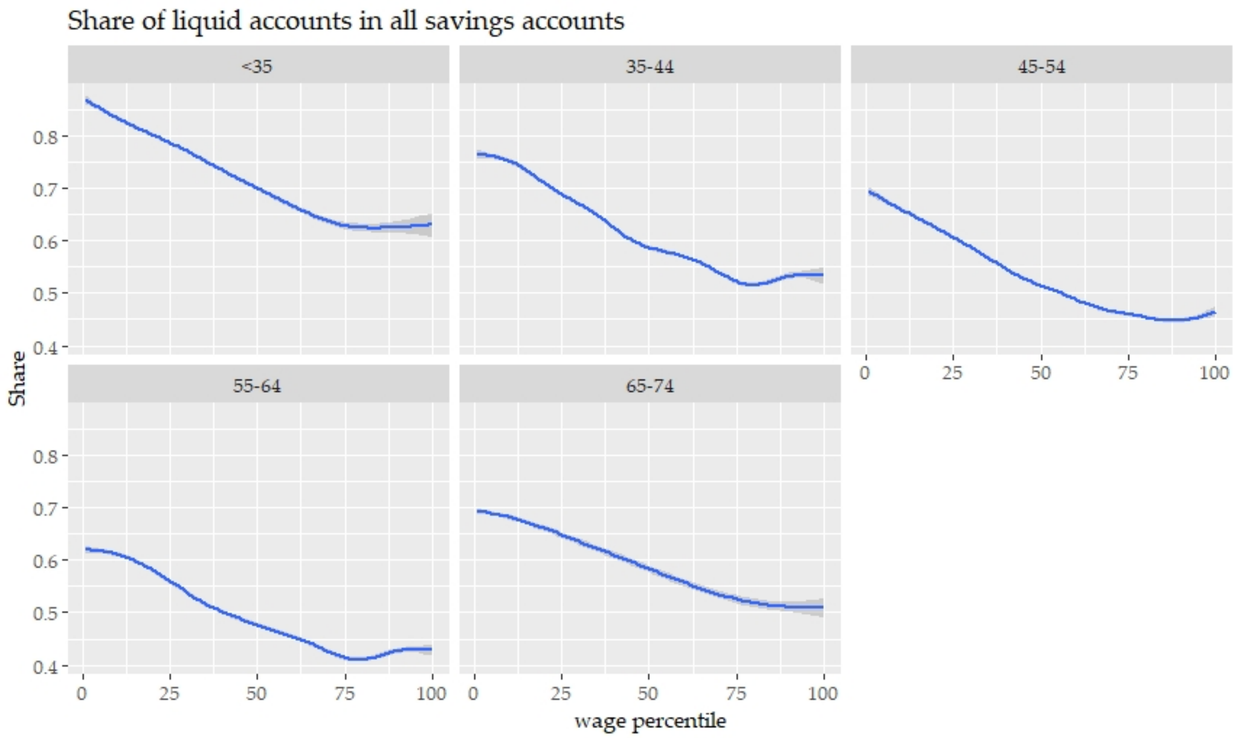


Figure 3: Share of liquid savings in overall savings by age group and wage percentile. *Survey of Consumer Finances data*, own calculations.

Figure 3 shows that the share in liquid savings remains substantial throughout the work life of low-to-middle-income workers. The flattening at the right end of the wage distribution with younger workers provides suggestive evidence for optimism-driven retirement saving delay. In contrast, low- and middle- income workers draw resources from liquid savings, at the expense of attaining higher returns in illiquid accounts. Overall, the share remains high towards retirement, amounting to 40% just before retirement age.[12] That is, the data shows that workers tend to stick to their liquid savings and decrease their liquid savings share only slightly before retirement.

In explaining the high share in liquid savings, this paper's narrative relies on income growth expectations. In reality, low retirement contributions may not only be driven by income growth misperceptions across the income distribution. Taking care of housing is an example of a retirement saving delay. The MSC data analysis in the appendix supports abstracting from housing in the structural model. First, the retirement confidence measure does not vary significantly with home ownership. Moreover, it does not vary with home value. Renters and homeowners perceive their retirement equally.

[11] Experienced bankruptcy, foreclosure, debt payments, real estate equity amount etc.

[12] I repeat the analysis with workers that own defined contribution accounts - the fitted lines across wealth percentiles retain similar shape.

The average worker pays off their mortgage for 30 years after buying the house. The model interpretation views mortgage and other mandatory payments as parts of the spending plan that in turn requires liquidity. The pessimistic worker understates future income and saves more to ensure liquidity. Consequently, the paper makes the case that homeowners and renters ensure liquidity in the same manner.

1.4 Three-period model

The stylized version of the model analytically proves that pessimism affects the decomposition of agents' savings (liquid-to-illiquid savings ratio). Pessimism induces workers to reallocate their savings from illiquid to liquid accounts.

Each agent is endowed with y_1 in period one and decides how much to consume and allocate to their savings accounts, liquid (s_1) and illiquid s_1^R. 2^{nd} period income follows a Bernoulli distribution

$$y_2 \sim \begin{pmatrix} y_L & y_H \\ p & 1-p \end{pmatrix}, \quad y_L < y_1 < y_H.$$

Agents can allocate part of their 2^{nd} period resources to liquid savings b_2 and consume the rest. In the third period, agents consume what they saved from both liquid and illiquid assets. When optimizing, agents form subjective expectations about the second period income. Pessimists assign greater probability to the bad outcome y_L, so

$$\hat{\mathbb{E}}[y_2] = \tilde{p} y_L + (1-\tilde{p}) y_H, \quad \tilde{p} > p.$$

The maximization problem is

$$\max_{s_1, s_1^R \geq 0} u(c_1) + \mathbb{E}[u(c_2) + u(c_3)] \text{ such that } \begin{aligned} c_1 + s_1 + s_1^R &= y_1, \\ c_2 + s_2 &= y_2, \\ c_3 &= R s_1^R + s_2, \end{aligned}$$

where $u(c) = \dfrac{c^{1-\gamma}}{1-\gamma}$.

Upon realization of the 2^{nd}-period income, all uncertainty is resolved. If the agent does not run down her liquid assets, she is able to divide resources evenly across period 2 and 3, choosing $s_2 = \dfrac{y_H + s_1 - R s_1^r}{2}$ so that $c_2 = c_3$. However, if she is constrained, $s_2 = 0$ and she consumes $c_2^L = y_L + s_1$. During retirement, she consumes retirement savings $R s_1^R$, where $R > 1$. Given all the assumptions, two lemmas hold:

Lemma 1. *If* $R < (1 + R^{\frac{\gamma-1}{\gamma}})$ *and* $y_2 = y_H \implies$ *borrowing constraint does not bind (i.e.,* $s_2 > 0$*, in the high income state).*

Lemma 2. *If the agent chooses to allocate to both liquid and illiquid savings and* $R \neq 1$*, the borrowing constraint binds in the low income state* y_L*.*

Agents know that they will be constrained in the low income state. Assume that agents aren't "too hungry" in the third period in the low income state ($u'(c_3^L) \leq 2u'(c_3^H) \implies c_3^L \geq \frac{c_3^H}{2^\gamma}$) for fixed 1^{st} period allocations s_1, s_1^R. Then, using the implicit function theorem it can be shown that the following result holds.

Proposition 1. *Suppose that retirement savings exhibit greater returns than liquid assets, $R > 1$, but are not too large, satisfying $R < (1 + R^{\frac{\gamma-1}{\gamma}})$. Define $s_1(p)$ an $s_1^R(p)$ as optimal liquid and illiquid savings. Given that the uncertainty in second period income is large enough, $s_1(p) > 0$, the following inequality holds:*

$$\frac{\partial s_1}{\partial p} > 0 > \frac{\partial s_1^R}{\partial p}.$$

That is, an increase in pessimism (assigning $\tilde{p} > p$) implies an increase in liquid asset holdings, and a decrease retirement savings.

At the expense of getting higher returns in retirement savings account, pessimist reallocate their savings to low-return, deposit-like accounts thus have less resources in the retirement. In the structural model, this mechanism sustains allocation to liquid accounts at the start of the work-life. As workers age, the mechanism dies out, and agents gradually increase their contribution to retirement account.

The same result does not apply to high-income workers, as their borrowing constraint is not binding, so the same set of results do not yield the same conclusion. On an intuitive level, high-income workers are not constrained, but are optimistic and may decide to delay their saving for retirement for the second period, when they expect that their income will be relatively higher. The additional return they would get if they started adding today instead of tomorrow may not incentivize them to forego spending more today. In the full life cycle model version, I show that, at both ends of the income distribution, workers delay their saving for retirement. This suggests that the intuition that holds for optimistic high-income workers may hold.

1.5 Full life cycle model

Pessimism reinforces precautionary motives analytically in a three-period simple version of the model. The full life-cycle model exhibits more trade-offs due to multiple sources of income shocks and a longer time horizon. The model solution therefore relies on computational methods. The computational solution uses a transformation of the two-dimensional endogenous grid method, allowing faster computation, as shown in (Druedahl, 2020). This section outlines model assumptions and potential trade-offs workers face throughout their career.

1.5.1 Defined contribution account

The retirement account represents the private contribution account [13], including 401(k) and 403(b). All workers are eligible to open the DC account. A contribution rate choice d_t entails transferring $d_t y_t$ to the DC account. The benefit function incentivizes small deposits [14]

$$h(d_t y_t) = \chi \log(1 + d_t y_t), \quad \forall t \leq T_{ret}.$$

Retirement savings can be used only once the worker retires. Assets in the DC account exhibit a return R_b, which is assumed to be higher than the standard deposit account return. Let b_t denote assets in the DC account after contribution $d_t y_t$. The law of motion for b_t is

$$b_t = R_b n_{t-1} + d_t y_t + h(y_t d_t), \quad \forall t = 1, \ldots, T_{ret} - 1.$$

b_t is the amount of savings after the contribution is made; thus it as a *post-decision* variable, whereas retirement account at the beginning of the period is denoted with n_t (*pre-decision* variable). Different timing notation connects the numerical solution method to Druedahl (2020).

Setting up an account does not yield any costs and may be postponed to a later point in the work life. The minimum contribution rate is set out to be 0%, thus equals the minimum rate for 401(k) in the U.S. The maximum contribution rate is fixed throughout the work life following the U.S. regulation, and corresponds to a specified dollar amount each year:

$$d_t y_t \leq m.$$

Participation in the DC account is voluntary, allowing all employees to catch up with their contributions as they approach retirement. Workers cannot, however, opt out and take the resources once they have created an account. [15]. As a result, when optimizing, the worker chooses between consuming out of assets when retired and being able to tap into the liquid account in case of an income shock. The pessimistic outlook of low-income workers incentivizes delays and low contributions in retirement plans.

1.5.2 Liquid savings account

A standard saving instrument is a liquid account with the return the $R_a < R_b$ on accumulated liquid assets a_t. Lower return of liquid assets encompasses the fact that retirement accounts are tax-deferred. Liquid savings fund current spending and can be accessed at any time. The volatility misperception has the same effect on optimists, up to a point where the income level bias effect outweighs the volatility bias effect. Following Druedahl and Jørgensen (2020), the model solution separates *pre-decision* liquid assets m_t (cash on hand) from the *post-decision* variable a_t.

[13] Abbreviated as DC account.

[14] Model estimates in section 1.6 take the smooth approximation of the step function used in a standard 401(k) employer-employee matching schedule.

[15] In the U.S., the worker can take money from the 401(k), albeit with a penalty of 10% of all illiquid savings. Correspondingly, SCF data shows an insignificant amount of withdrawals.

1.5.3 Retiree's problem

Retirement age is deterministic T_{ret}. A defined contribution account is an additional liquid resource throughout retirement. That is, accumulated savings in the DC account are paid out as annuity payments y_{an}. The amount left after the annuity payment does not exhibit a return. Because not all workers invest in retirement accounts, a retirement benefit is provided, $\bar{b}$. The retiree's problem boils down to a standard consumption saving problem, conditional on having assets in the DC account:

$$\max_{\{c_t, a_t \geq 0\}} u(c_t) \ \text{ s.t. } \ c_t \leq m_t - a_t$$

$$m_{t+1} = R_a a_t + \mathbf{1}_{\{DC\}} y_{\text{an}} + (1 - \mathbf{1}_{\{DC\}})\bar{b}.$$

1.5.4 Worker's problem

While employed, workers receive labor income y_t at the beginning of the employment year and choose the allocation of liquid savings a_t, retirement contribution d_t and consumption c_t, bringing utility

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}.$$

Workers face a persistent and a transitory shock each period. Subjective workers do not fully understand their income process and extrapolate from previous income realizations, whereas rational workers perceive their income correctly. Problem equations hold for both rational and subjective expectations, commonly denoted with $\mathbf{E}_t$.

State variables are current labor income, cash on hand at the beginning of period and accumulated retirement savings (p_t, ξ_t, m_t, n_t). Cash-on-hand consists of accumulated liquid savings and current labor income $y_t = \alpha_i + g_{i,t} + \lambda p_{i,t} + \gamma_{i,t}$ together:

$$m_{i,t} = R_a a_{i,t-1} + y_{i,t}.$$

Throughout the rest of the paper, subscript i is omitted.

The indicator function tracks DC account participation:

$$\forall t = 1, \dots, T_{ret} - 1, z_t = \begin{cases} 1; & \text{has DC acc} \\ 0; & \text{no DC acc} \end{cases}.$$

Opening the retirement account is a one-time decision, i.e.:

$$\mathcal{Z}(z_{t-1}) = \begin{cases} 1 & , z_{t-1} = 1 \\ \{0, 1\} & , z_{t-1} = 0. \end{cases}$$

Each worker maximizes the value function that is the maximum of two conditional value functions. If she did not start contributing, for $z_{t-1} = 0$

$$V(0, p_t, \zeta_t, \xi_t, m_t, n_t) = \max_{z \in \{0,1\}} \begin{cases} v_t(1, p_t, \zeta_t, \xi_t, m_t, n_t), \\ v_t(0, p_t, \zeta_t, \xi_t, m_t, 0), \end{cases}$$

or the worker already contributes, so $z_t = 1$ and

$$v_t(1, p_t, \zeta_t, \xi_t, m_t, n_t) = \max_{0<d_t\leq 1, c_t\geq 0} u(c_t, d_t) + \beta \mathbb{E}_t\Big[V_t(1, p_{t+1}, \zeta_{t+1}, \xi_{t+1}, m_{t+1}, n_{t+1})\Big]$$

such that

$$\begin{aligned} c_t + y_t d_t &\leq m_t - a_t \\ b_t &= R_b b_{t-1} + y_t d_t + h(y_t d_t) \\ m_{t+1} &= R_a a_t + y_{t+1}. \end{aligned}$$

If the worker postpones DC participation, she chooses her consumption and liquid assets to transfer to the next period, earning return R_a and reconsiders adding to retirement savings in the next period.

$$v_t(0, p_t, \zeta_t, \xi, t, m_t, 0) = \max_{c_t, a_t\geq 0} u(c_t) + \beta \mathbb{E}_t[V_t(0, p_{t+1}, \zeta_{t+1}, m_{t+1}, 0)]$$

such that

$$\begin{aligned} c_t &\leq m_t - a_t \\ m_{t+1} &= R_a a_t + y_{t+1} \end{aligned}$$

Interior solution to the DC participant problem satisfies

$$\begin{aligned} \frac{c_t^{1-\gamma}}{1-\gamma} &= \beta \mathbb{E}_t[v_{m,t+1}] \\ d_t y_t &= \frac{\chi}{R_a \mathbb{E}_t[v_{m,t+1}] - R_b \mathbb{E}_t[v_{n,t+1}]}. \end{aligned} \tag{6}$$

Expression 6 represents the trade-off workers face each period. Contribution rate increases with benefits χ, and decreases whenever the current marginal value of liquidity exceeds the marginal value of saving in retirement. Pessimistic expectations overstate low income realizations and thus increase the marginal value of liquidity, which drives the difference between rational and subjective savings paths. However, as the worker approaches retirement, the value of adding to the retirement plan increases.

1.6 Estimation

The model estimation follows two steps. The first step uses sample estimates of the income process in the PSID data (Cocco et al., 2005), and calibrates the income growth bias using the MSC forecast errors at every income quantile.

The benefit function matches the amount added to the retirement savings account, approximating the matching schedule in the employer-employee level data (Choukhmane, 2021; Beshears et al., 2020). Yearly contribution thresholds are calibrated to match the income process in the MSC data.

Table 6: Maximum contribution limits - calibrated thresholds.

Dollar terms	Amount	Percentile
2015 $	18 000	0.08
	24 000	0.135

1.6.1 Contribution match schedule

Contribution limits correspond to the cap given by U.S. law in $2015 terms. In 2015 the maximum contribution amount was $18000 for workers under 50 and $24000 after. Each threshold corresponds to the sample percentile in the MSC income data. Both percentiles from the simulated income sample of 10000 agents using the parametrized income process serve as the constraint in the model (Table 6).

Finally, the contribution function $h(y_t d_t)$ corresponds to the matching schedule among U.S. employers. Most employers match their employee contribution rate up to 6% of wage. That is, as long as the worker contributes less than 3% of her wage, her employer will match with the same amount. If the contribution rate is higher than 3%, her employer matches with 3% of the employee's pre-tax wage. In the baseline case, 3% is **the default rate** and is a subject of the policy change in this paper. Since the benefit function is a smooth approximation for the employer matching schedule, the parameters are pinned down via curve fitting (Table 7).

$$h(d_t, y_t) = \chi \log(a d_t y_t + b)$$

Table 7: Benefit function parameters approximation.

χ	a	b
0.34	5.63	1

1.6.2 Other model parameters

The retirement savings return corresponds to the average return of a standard lifecycle fund, which is known to be the default and most popular choice among 401(k) contributors (Mitchell et al., 2006; Parker et al., 2022). The return on liquid assets incorporates a tax differential, since gains on 401(k) savings are tax-deferred. The rest of the model parameters correspond to standard values found in the literature.

Once simulated, consumption and savings paths define the calibration objective for the risk aversion parameter, as preferences are independent of the income expectations formation (Table 8).

Table 8: Fixed parameters in the model.

Fixed parameter	Source	Value
β	Cocco et al. (2005)	0.98
T_{ret}	Love (2006)	70
T	Love (2006)	90
R_a	*exogenous parameter*	1.02
R_b	*target-date fund performance average*	1.04
γ	*calibrated to match illiquid-to-liquid ratio in the SCF*	3.7

1.7 Solution method

The worker's function is non-convex due to opting into the retirement account. I solve the problem by utilizing the model's upper-envelope property, where the consumption choice is solved independently of the contribution choice. The upper envelope defines values comparable across workers' DC participation choices. The solution algorithm uses the endogenous grid over assets;[16] thus, it is computationally faster. For a fixed contribution rate, the worker consumes out of her net labor income and current liquid savings.

The next section shows the shape and differences in policy functions across the income distribution. The key finding - the consumption and savings plans differences between subjective and rational workers are outlined with life-cycle simulation comparisons after.

1.7.1 Policy functions

Consumption and savings are functions of current liquid and illiquid account balances. In figures 4 and 5, the left axis represents current liquid savings, while the right axis represents current retirement savings. For a fixed level of income persistence, each plane represents consumption and contribution rates. Different income quantiles are represented as different figures and are labeled as low- and high-income, respectively. The difference between rational and subjective worker policies lies in the shape and level of the policy plane, which is shown in the appendix.

As shown in figure 4, young low-income workers (left) consume less than their high-income workers (right).

Figure 5 depicts differences in contribution rates between low- and high-income young workers. Low-income workers are only incentivized to contribute more to retirement accounts if they are low on retirement savings, whereas high-income workers contribute at higher rates even for substantial retirement saving levels.

[16] The model solution builds on Druedahl (2020) and allows for both the persistent and transitory components in the income process.

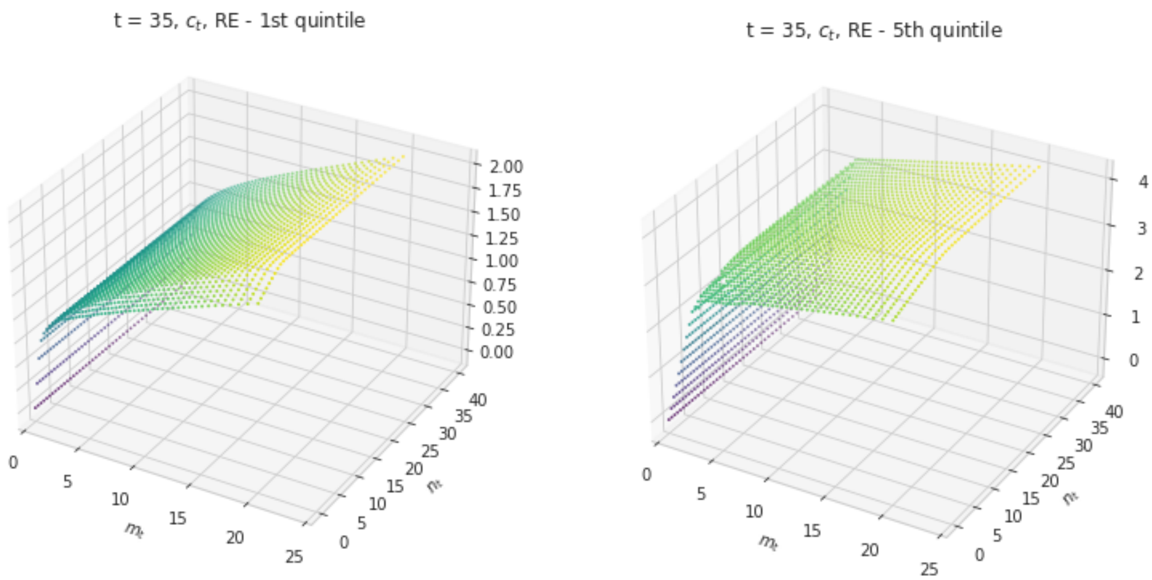


Figure 4: Rational expectations consumption policies for first (left) and fifth (right) income quintile.

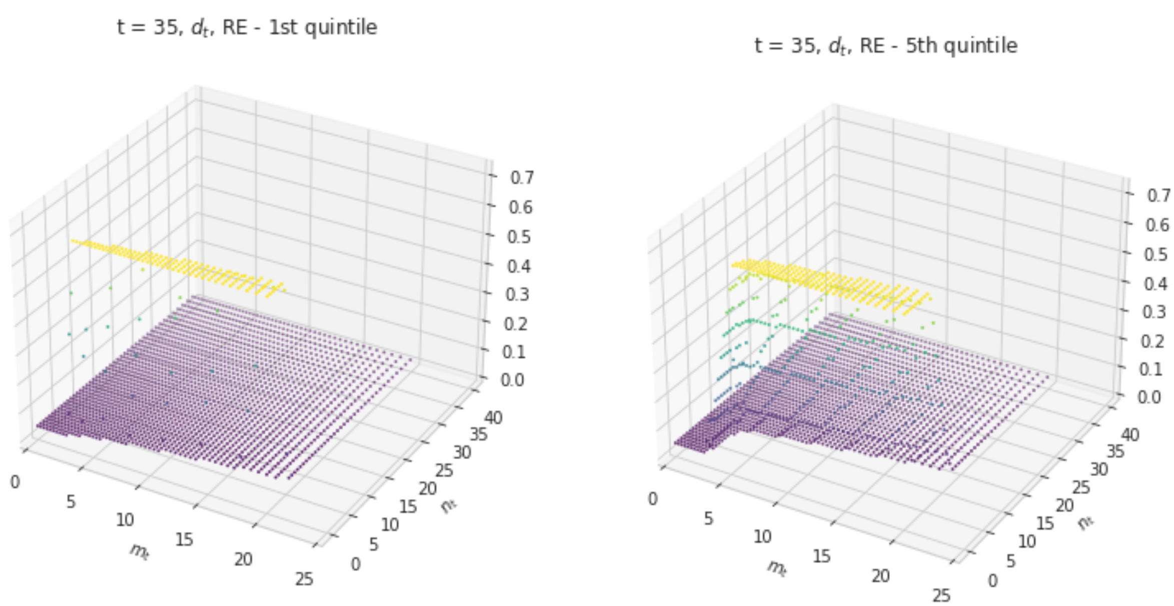


Figure 5: Rational expectations, contribution rate policies, for first (left) and fifth (right) income quintile.

1.7.2 Rational and subjective worker - comparison

In the remaining part of the paper, all simulations compare the subjective expectations solution ($\mathbb{E}^*$) to the rational expectations ($\mathbb{E}$). Consumption and savings policy differences between rational and subjective workers accumulate over the life-cycle, and overall have a different effects on retirement savings. On average, subjective workers save less in retirement accounts, and rely on liquid savings instead (Figure (6), left). In addition, subjective workers consume less than their rational counterparts only to consume more once their income level bias overcomes the uncertainty misperception (Figure (6), right).

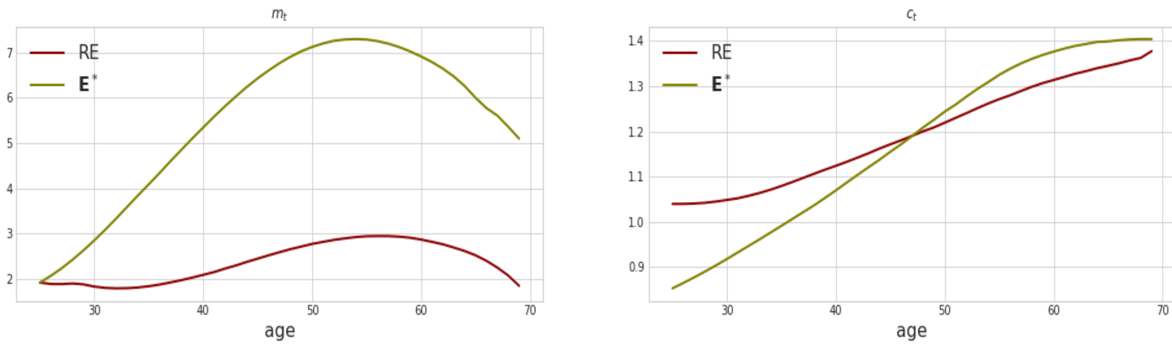


Figure 6: Liquid savings (left) and consumption (right) paths comparison for bottom 25%-income workers.

Even though the calibration targets liquid-to-illiquid assets ratios, the subjective expectations solution aligns with the data on contribution rates over the work-life (Choukhmane, 2021; Parker et al., 2022). Contribution rates increase steadily only to decrease just before retirement; Figure 7 contains dotted yearly contribution rates, averaged on a smaller sample of "middle-class investors" from financial institution data in Parker et al. (2022). In general, subjective expectations substantiate a slow increase in contribution rates over the working period, whereas rational workers decrease their contributions over the work life. Further analysis shows that, corresponding to Parker et al. (2022), the steady increase in contributions show up regardless of the initial income.

At the bottom of the income distribution, rational workers keep their contribution rates lower and do not change over the work life (Figure 8, red line). In contrast, pessimists delay their contributions at the start of the work life, only to increase their contributions after (Figure 8, green line). That is, subjective workers slowly increase their contribution towards the end of work life. Due to the data unavailability on contribution rates for low-income workers[17], subjective expectations solution establishes important facts for low income workers - even though there is a delay in contributions, this delay is offset by increased contributions later on in the work life. A few years before retirement, workers decrease their contributions due to lower incentives once the retirement year is closer, corresponding the bottom tercile estimates in Parker et al. (2022).

[17] Parker et al. (2022) outline their estimates for middle class workers using the data from a financial institution. Their analysis for the bottom tercile supports the findings in this paper.

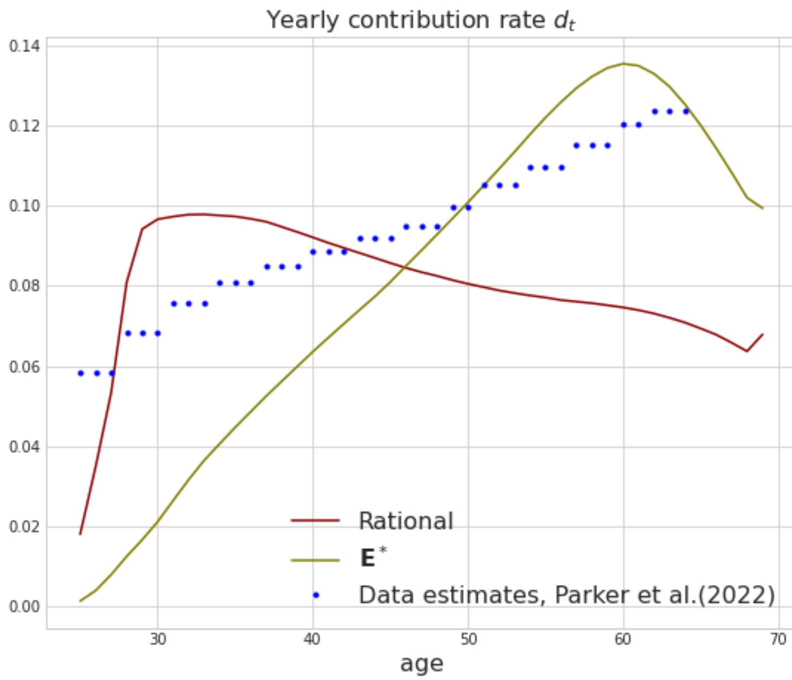


Figure 7: Contribution rates at the mean of the income distribution, rational (red) and subjective (green) life cycle paths.

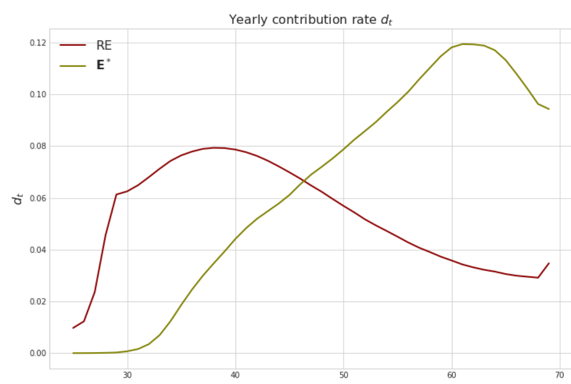


Figure 8: Rational (red) and subjective (green) lifecycle contribution paths for bottom 25%.

1.7.2.1 Liquid-to-illiquid savings ratio across the work life

In addition to contribution rates, subjective expectation model performs well in fitting liquid-to-illiquid savings ratios to SCF data estimates throughout workers' careers. In contrast, rational expectations model understates liquid savings across the life cycle. Throughout this section I compare savings ratios to SCF data estimates from the first part of the paper (Figure 3).

Figure 9 depicts savings ratios across wage percentiles for the two models, for workers age 45-54. Subjective expectations capture the shape and slightly overstate savings ratios in

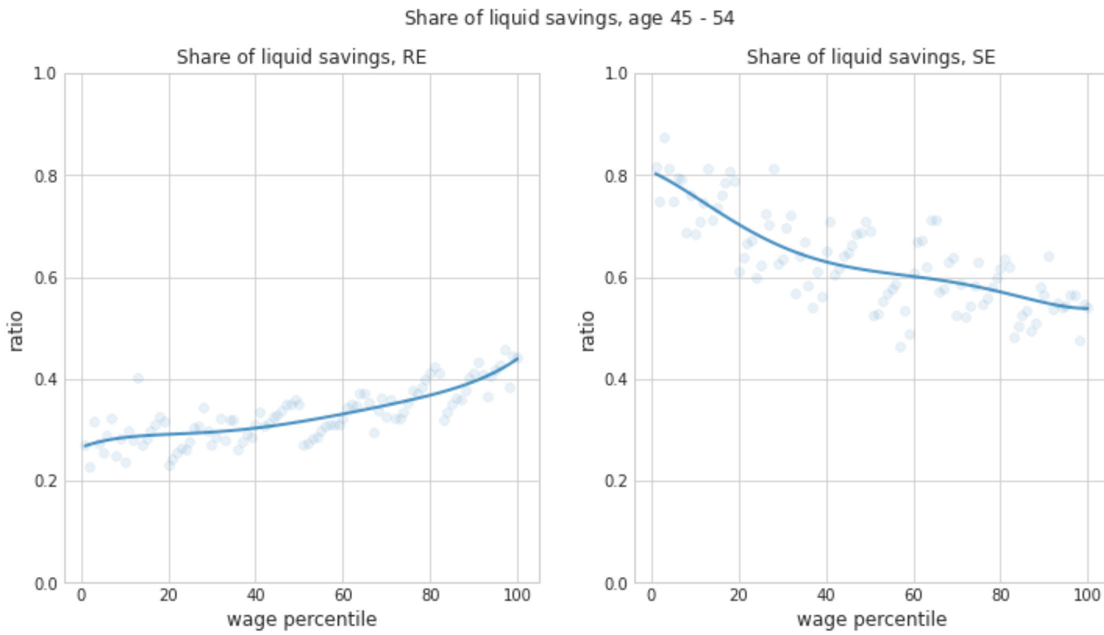


Figure 9: Savings ratios for workers age 45-54, model simulations. Rational expectations; left, and subjective expectations; right.

comparison to SCF data estimates for the same age group (Figure 3, top panel, right graph). Moving one cohort up (Figure 10), subjective expectations capture the ratios even better, both with shape and size (Figure 3, bottom panel, left graph).

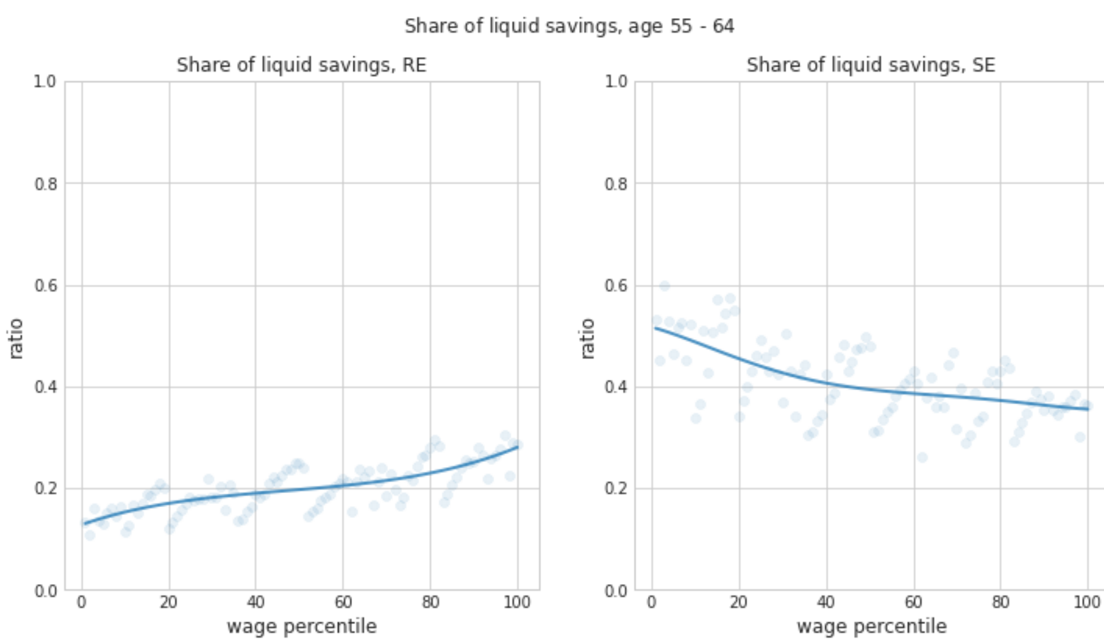


Figure 10: Savings ratios for workers age 55-64, model simulations. Rational expectations; left, and subjective expectations; right.

1.7.2.2 Subjective expectations and saving rates implications

The effects of extrapolation are also attributable to saving rates findings in the data. Specifically, Fagereng et al. (2019) find that the net saving rate (i.e., liquid savings from income) remains flat from the 20^{th} wealth percentile onward. Subjective expectations simulations support this empirical fact, whereas rational expectations imply increasing the net saving rate across the wealth distribution (Figure 11, left). Extrapolative expectations generate wealthy workers with incentives to save due to the misperceived volatility of income.

Including gross saving rates (i.e., savings that include retirement accounts), extrapolation implies a larger difference between the two saving rates across the wealth distribution (Figure 11, right), which is consistent with empirical findings on capital gains differences across the wealth distribution. In contrast, rational expectations solution exhibits stark differences even for the bottom 20% of the wealth distribution, which is not supported in the data (De Nardi and Fella, 2017; Fagereng et al., 2019)[18]. Even when all workers are eligible to contribute, the disparity in gross saving rates across the wealth distribution highlights the effect of unrealized capital gains in retirement accounts. These gains materialize once workers reach retirement and affect retirement consumption inequality.

Figure 11: Net and gross saving rates, rational expectations (RE, left) and subjective expectations ($\mathbf{E}^*$, right) simulations.

All agents in the model are eligible to save in employer-matched retirement accounts. Even with the default rate and employer matching, subjective expectations create a lack of incentives to save in retirement accounts. However, the increased eligibility and growing interest in retirement savings incentives provide a foundation for specific policy evaluation that may reflect workers' responses.

[18]Since future spending plans (including mortgage, rent, etc.) draw out of liquid savings, liquid savings rates high.

1.8 Policy experiment - automatic enrollment

As a way of ensuring retirement security, automatically enrolling workers into their retirement plans has been encouraged by U.S. legislation. The majority of empirical studies estimate differences between two types of enrollment: active enrollment, in which workers actively choose to begin contributing to their 401(k) (benchmark model), and automatic enrollment, which enrolls workers automatically. Employers can thus add 401(k) accounts in their employees' names through automatic enrollment.

The long-term effects of automatic enrollment with default rates cannot be evaluated simply due to the recent policy introduction, and the resulting findings discuss the short-term effect (5 to 7 years after the enrollment). Choukhmane (2021) and this paper are among the first ones to test the potential effects of automatic enrollment throughout work life. The default rate set remains at the standard and is 3%, leaving workers to adjust their contributions without any costs.

Only in the first year of employment do employers make the contribution in workers' names. Given that the subjective expectations model recreates the patterns in contributions found in microdata (catching up, increasing contributions, starting with low contributions), I test for policy effects with workers who extrapolate. Policy tests imply that automatic enrollment has an insignificant effect on retirement savings right before retirement.

Figure 12 shows rational and subjective workers' consumption and savings paths across the life cycle. Subjective workers maintain their liquid buffers and thus consume less. Aside from a decrease in consumption due to the exogenous default rate in the first year of tenure, differences between the two expectation solutions remain the same.

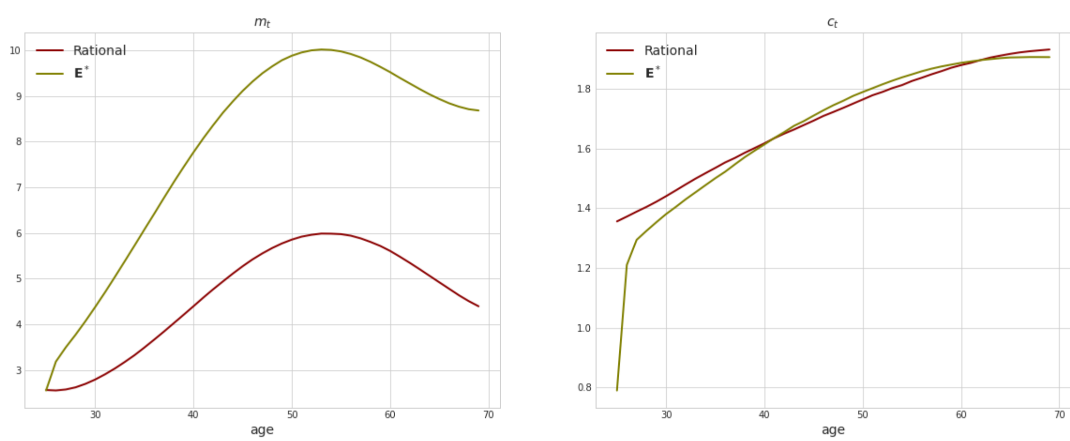


Figure 12: Rational (RE) and subjective (E) liquid savings and consumption under auto-enrollment.

Even though consumption and liquid savings initially adjust to the automatic enrollment, retirement contributions in later work life are offset by the initial increase. Knowing that they need to add substantially more than preferred in the first year of tenure, workers offset first-year contributions by delaying their contributions and, ultimately, catching up with a slightly lower contribution rate towards the end of the work-life (13, right). On the other hand, rational workers do not change their savings paths because they are already responding to incentives in the voluntary setting.

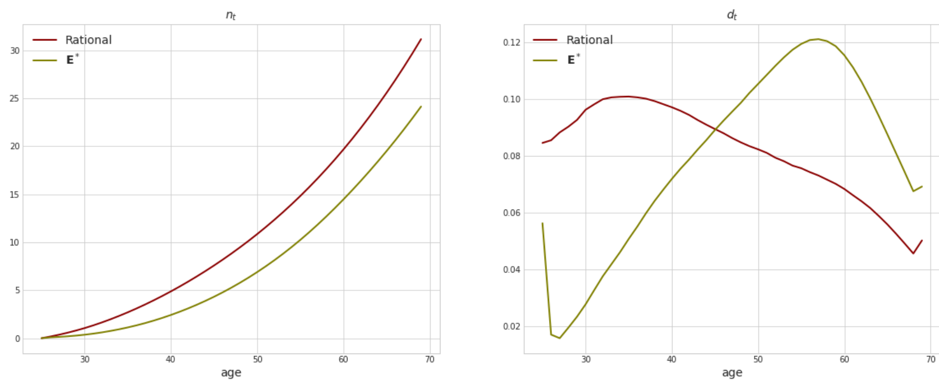


Figure 13: Rational (RE) and subjective (E) retirement savings under auto-enrollment,bottom 25%.

1.8.1 Subjective workers under auto-enrollment

Comparing subjective workers' savings paths across voluntary and automatic enrollment renders the effect of auto-enrollment negligible. While voluntary contributions remain flat (non-existent) at the beginning of tenure, automatic contributions decrease right after the initial contribution made in the worker's name (Figure 14, bottom right plot). Therefore, contribution rates increase steadily. Moreover, liquid buffer amounts remain high (Figure 14, top right plot), in line with empirical findings on the insignificant effect of auto-enrollment on other financial decisions (Beshears et al., 2022). Therefore, gains obtained from a first-year accumulating throughout the work life

Specifically, even though contribution rates are higher for all levels of liquid and illiquid savings (Figure 15, points in green), they are later offset, and the contribution under voluntary policy prevails (points in red).

While I assume no switching costs whatsoever, the average contribution rate for automatic enrollment never goes back to zero during the first five years of tenure (Figure 14, bottom right graph). The policy introduces the trade-off represented by equation 6 because now the value of adding to the retirement account depends on gains accumulating from the mandatory first-year contribution. Over time, however, workers adjust. In sum, automatic enrollment has short-term positive effects, whereas long-term contributions decrease relative to voluntary contributors savings rates.

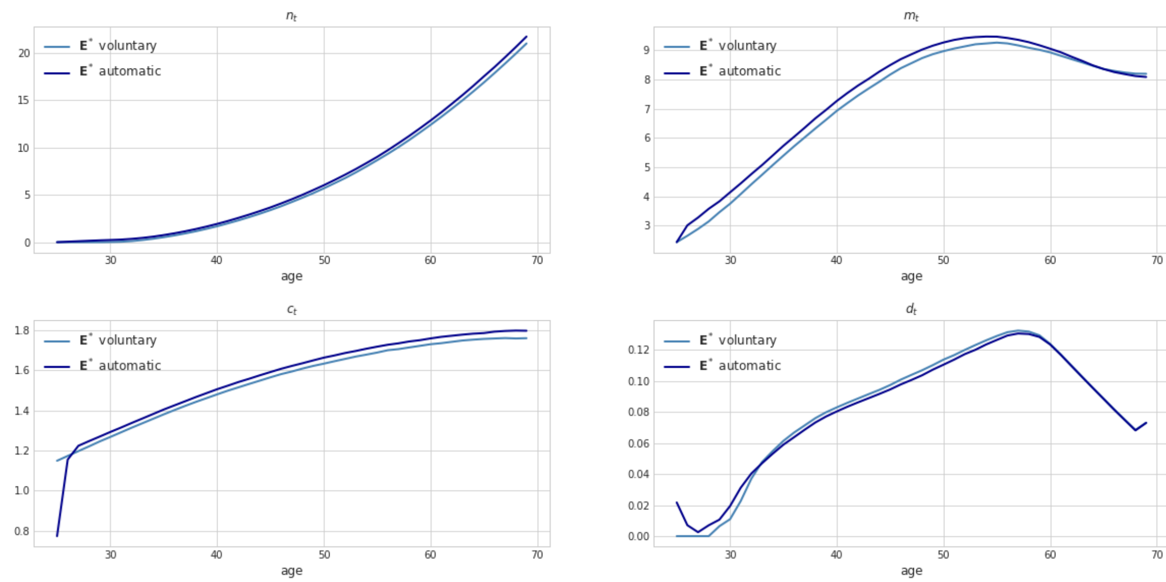


Figure 14: Subjective (E) solution under voluntary and auto-enrollment setting.

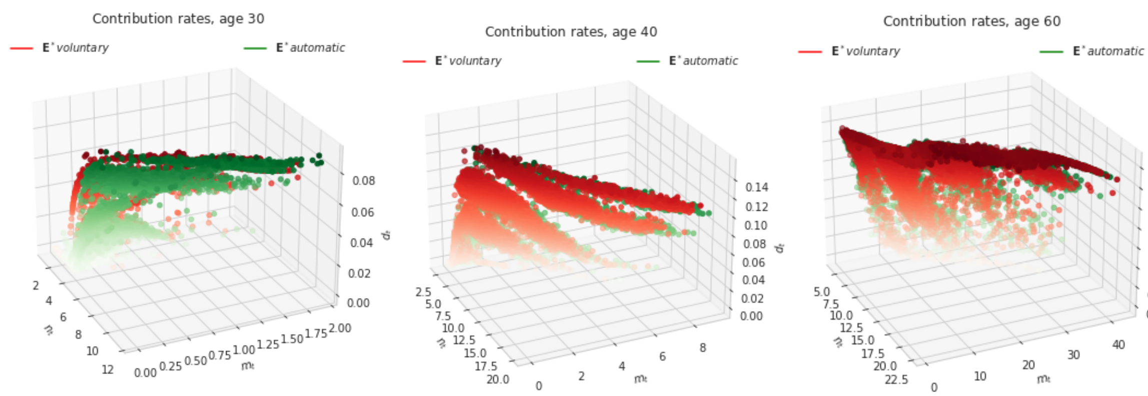


Figure 15: Simulated contribution rates under automatic enrollment, subjective expectations (E).

Table 9 shows the effect of automatic enrollment on retirement savings in the last year of tenure, for each quantile of the income distribution. Even though the average effect is negligible, the income quantile-breakdown shows a larger increase in retirement savings for bottom 50% of wage-earners (Table 9). Therefore, the size of the welfare effects varies with social planner preferences.

	q_1	q_2	q_3	q_4
retirement savings increase (%)	3,75	3,9	2,2	1,8

Table 9: Retirement savings increase under automatic enrollment.

Finally, figure 16 shows consumption differences in retirement for workers of different earnings paths. Based on the median income within last 5 years of tenure, I plot consumption paths under voluntary and automatic enrollment. Across all income quantiles, consumption differences are small. Consumption shifts upwards throughout retirement, owning to a slight increase in annuity payments each year.

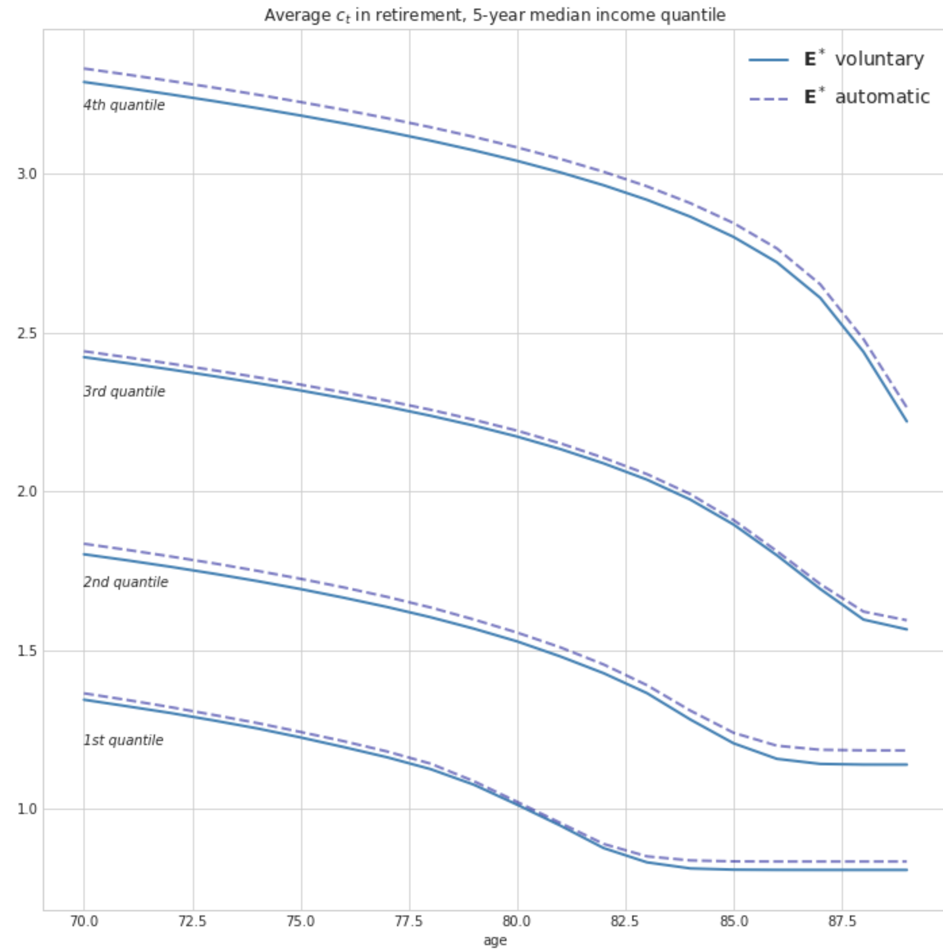


Figure 16: Subjective (E) consumption in retirement under the voluntary and auto-enrollment settings.

Ultimately, agents who participate in retirement plans catch up with workers who start adding from the beginning and consume similarly throughout retirement. Including workers in retirement plans right from the beginning yields insignificant effects due to extrapolation and additional precautionary motives. The share of workers who participate increases with age and conforms to the data. In contrast, rational workers start adding from the beginning, utilizing their benefits, and thus leaving automatic enrollment testing redundant.

1.9 Conclusion

This paper introduces a deviation from rational expectations in a life-cycle model with liquid and illiquid savings accounts to explain retirement contribution patterns over the work life. The structural life-cycle model builds on individual income forecast errors found in the Michigan Survey of Consumers data. In the model, agents extrapolate from their past income realization and base their consumption and (illiquid) savings decisions on biased income projections.

The model's expectations incorporate household income forecast biases estimated from the Michigan Survey of Consumers. I expand the data analysis in Schlafmann and Rozsypal (2023) and show that households tend to extrapolate from past income growth to form expectations about their future income. Second, subjective unemployment probabilities imply volatility overstating across all workers. Third, the income forecast bias decreases over the work life.

The three-period stylized model analytically proves that pessimism induces reallocation to liquid savings at the expense of saving for retirement. Consequently, retirees consume out of liquid savings accounts. My findings suggest that pessimists require more significant incentives to save in illiquid accounts. The full life-cycle model connects extrapolation to savings allocation over time, with the presence of transitory and persistent income shocks. Saving for retirement is possible through a private retirement account closely following savings incentives in the employer-employee data.

The extrapolative expectations solution matches the empirical contribution rates pattern in the data. Workers delay participating in retirement accounts only to increase their contribution throughout their careers. Contrary to workers who extrapolate, rational workers contribute at higher rates from the start of the work life and keep their contribution rates flat later on, which is inconsistent with empirical findings (Choukhmane, 2021; Parker et al., 2022).

Even though the benefits of saving in a 401(k) plan include tax deferrals, employer matching the contribution rate, and higher expected returns, retirement studies find that workers do not add to their accounts. Therefore, automatic enrollment remains to be the policy encouraged in U.S. legislation. The recentness of the auto-enrollment policy does not allow testing for long-term effects. Since the extrapolative expectations solution captures contribution patterns across cohorts, I test for automatic enrollment policy effects throughout a worker's career. The effect of auto-enrollment on total retirement savings decreases from 3,75% at the bottom to 1,8% at the top of the income distribution. As the bottom quantile continues to add to their liquid accounts, retirement consumption does not increase significantly. That is, throughout their career, workers save in the same way, in line with short-term effects in the U.S. data (Beshears et al., 2022).

This paper is the first to incorporate extrapolation in the life-cycle model to explain retirement contribution patterns. Policy tests call for novel retirement system adjustments (such as auto-escalation policies) that account for catching-up behavior. However, without simulating auto-escalating employers' contribution matches, this paper cannot make a quantitative statement on welfare gains in retirement. Including auto-escalating matches may induce additional delays in saving for retirement through 401(k) accounts. I leave it for future research.

2 Mortgage Shopping Behavior in the U.S. - Stochastic Record Linkage

- co-authored with Ante Šterc

2.1 Introduction

Following increasing availability of data, the literature on financial behavior has moved towards empirical estimates of cognitive and monetary costs of individual investing and saving. In an effort to measure cognitive costs and differences in understanding, Lusardi et al. (2010) proposed to measure the financial knowledge of individuals using a set of three survey questions (*"The Big Three"*). These questions define the **objective financial literacy score** and are related to differences in saving and consumption behavior.

Whereas most of the literature focuses on the correlation between financial literacy score and debt or asset levels, our paper aims to uncover the mechanism underlying the positive correlation between financial knowledge and individual debt management. We focus on the mortgage rate attainment in the U.S. and, using a stochastic imputation procedure, show that individual loan shopping behavior and financial skill level interact and explain a part of the residual mortgage rate variation after accounting for observables.

The U.S. mortgage market has undergone significant structural changes and advancements in digital mortgage advertising and undertaking. With a steady increase in non-bank online lenders, the mortgage market experienced increasing competition, which has been elicited through the increase in mortgage eligibility of borrowers with modest credit scores (Zhou, 2022; Bhattacharya et al., 2021). With increasing loan options, individual loan shopping behavior and financial knowledge became significantly more important for mortgage attainment. We focus on the demand side while controlling for the other contract specifics.

Limited data availability does not allow connections between individual financial knowledge to loan shopping behavior. To circumvent public data limitations, we employ the Stochastic Record Linkage (Enamorado et al., 2019) and impute individual financial literacy scores for borrowers in the National Survey of Mortgage Originations (NSMO). The stochastic linking method allows us to control for the uncertainty in the financial skill level we obtain from the external data set. In this way, for every borrower in the NSMO, we estimate a distribution of the financial skill level that depends on her respective match to a record in the Survey of Consumer Finances (SCF).

The objective measure of financial skills offers unique insights into individual mortgage attainment. Our findings surpass subjective perceptions of financial knowledge and risk aversion. Our first line of findings uses the SCF sample and suggests that financial literacy exhibits a hump-shaped profile over the life cycle. Moreover, we show that financially skilled borrowers are 20-30% more likely to refinance their mortgage, irrespective of their income, education, and mortgage size.

Next, we turn to our new merged data set and measure the borrower's effort using a survey question on the number of mortgage lenders that borrowers considered in their mortgage shopping process. Our estimates show that, among similar mortgage applicants, financially savvy ones are 5% more likely to consider one additional lender. Moreover, we show that search effort effectiveness increases with skill level and predicts a 13.4 b.p. lower mortgage rate for financially savvy borrowers who exert more effort in the mortgage acquisition process.

The sample period from 2014 to 2021 provides a window to observe variations in financial skills and search effects within each origination year. We find that the interaction effect increases over this timeframe. This period aligns with a simultaneous increase in the presence of non-bank lenders in the U.S. mortgage market. Our findings indicate that, when controlling for year effects, the influence of search efforts among financially savvy borrowers increases over time. Consequently, we argue that financial skills and search activity are increasingly pivotal in explaining mortgage rate disparities among U.S. households.

In our estimates, we go beyond the mortgage origination and observe borrowers' loan performance scores over time. We find that financially unskilled borrowers are 35-45% more likely to be late on payments three years after the mortgage originated. Given that our estimates control for the mortgage amount, credit score, and payment-to-income ratio of every mortgage, we interpret this result as a consequence of poor budgeting and low savings buffers against individual payment shocks.

Our findings represent a set of stylized facts for the mortgage attainment process in the U.S. In our subsequent work, we introduce a set of assumptions that correspond to our findings on the importance of individual search behavior and financial knowledge for mortgage rate attainment in the U.S.

2.2 Related Literature

This paper contributes to empirical studies on mortgage undertaking, refinancing, and financial literacy effects on individual mortgage performance. Our paper leverages the current way U.S. households face the mortgage process.

The empirical literature argues that financial literacy explains financial behavior in the credit market. Bhutta et al. (2020) use mortgage origination platform data and show that, even within the specific loan officer, there is a considerable amount of dispersion in interest rates among otherwise comparable borrowers[19]. Moreover, Gerardi et al. (2023) find significant race differences in mortgage prices, pertaining to more than income and education differences.

The losses from the mortgage contract go beyond the choice at origination and may come from refinancing mistakes. Our estimates from the SCF data corroborate findings in Agarwal et al. (2016) and show that financially unskilled households do not refinance as often. In the Danish environment, Andersen et al. (2020) attribute the mistakes to refinancing to individual inattention. Keys et al. (2016) find that more than 20% of U.S. borrowers did not refinance at the optimal time, when interest rates were low, and relate individual suboptimal

[19]Specifically, Bhutta et al. (2020) compare borrowers with similar credit scores and characteristics searching for the same loan amount.

behavior to procrastination and financial sophistication. We estimate individual refinancing probability differences across financial literacy scores while controlling for other observables.

Owing to the series of seminal papers (Lusardi et al., 2010; Lusardi and Mitchell, 2014; Lusardi et al., 2020), the correlation between individual financial literacy and portfolio choice and saving behavior has been well documented. Bhutta et al. (2022) focus on liquid savings and show that financially unskilled households more often face liquidity constraints due to their low liquid buffers. Through the lens of our estimates, lower buffers may be coming from poor practice in mortgage choice.

Following empirical findings, Jappelli and Padula (2017) and Lusardi et al. (2017) introduce financial literacy in the portfolio allocation model while assuming that individual returns depend on the level of financial literacy. Our estimates suggest a mechanism that relates mortgage rate attainment and individual financial literacy through search effort. In this way, we introduce a search mechanism that we model in our subsequent paper.

In the European contexts, where the number of potential lenders is significantly lower, Damen and Buyst (2017) show that borrowers can save more than €7,078 over the mortgage term by shopping and comparing different mortgage products. Additionally, U.K. estimates show that young and inexperienced borrowers make costly mortgage choices (Coen et al., 2023).

Our estimates underscore the effectiveness of the mortgage search depending on individual financial literacy scores, as low-income borrowers may be searching out of fear and make costly choices (Agarwal et al., 2020). In this regard, the sign of the interaction between search and financial skills changes as individual incentives change.

2.3 Data analysis and stylized facts

Leveraging on the robustness of stochastic imputations, we outline the set of estimates that highlight the importance of financial skills and search behavior in mortgage attainment. Whereas most of our inference is correlational, the dataset allows exploring mortgage performance a couple of years after the mortgage originated. First, we introduce the SCF data and present three stylized facts related to financial literacy and a broad definition of mortgage refinancing. Next, we introduce the second data source (NSMO) and later proceed to present the findings of the novel U.S. dataset (NSMO+) generated using the stochastic merging method.

2.3.1 The Survey of Consumer Finances

The SCF, a triennial survey of randomly chosen U.S. households, captures data on investment, housing, and debt. These responses construct a comprehensive balance sheet for typical U.S. households, vital for empirical household finance studies. Our analysis focuses on a SCF subset with a ”financial literacy score,” from the 2016 and 2019 waves, comprising 60,125 responses. By incorporating data on credit search behavior and mortgage refinancing, akin to the NSMO data, we explore credit shopping patterns among 41,788 first-lien mortgage holders and renters, aligning with NSMO standards. The evidence on variation in individual financial literacy provides three key insights that form foundational assumptions for our mortgage search model.

Financial literacy score is based on a set of three questions (*The Big Three*) that are shown to be efficient in comprehensively evaluating individual financial skills (Lusardi et al., 2010; Lusardi and Mitchell, 2014; Bhutta et al., 2022). The set of questions tests individual understanding of inflation, risk diversification and compounding:

1. Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow?
   - More**/Exactly/Less than $102
   - Do not know/Refuse to answer
2. Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account?
   - More/Exactly/Less** than today
   - Do not know/Refuse to answer
3. Please tell me whether this statement is true or false. "Buying a single company's stock usually provides a safer return than a stock mutual fund."
   - True
   - False**
   - Do not know
   - Refuse to answer

Unlike perceived financial knowledge, which signifies confidence, these objective scores provide insight into actual financial planning and behavior (Bhutta et al., 2022; Lusardi et al., 2010). To explore this, we employ a stochastic merging procedure, integrating mortgage data with the SCF. This approach allows us to discern collective patterns in *objective financial skills*, search effort, and mortgage rates among comparable borrowers.

First, we highlight essential household characteristics pertaining to financial literacy. Utilizing an ordered logistic model, we predict financial literacy scores based on borrower attributes. Table 10 presents personal attributes associated with financial literacy. Model-generated probabilities indicate that college graduates correctly respond to all financial literacy questions with a probability of 77%, while high-school graduates do so with a probability of 52%. Additionally, Figure 17 offers empirical evidence demonstrating a positive correlation between educational attainment and financial literacy.

Although education explains a considerable portion of the variation in financial literacy, as evident from the significant coefficients in Table 10, income, age, and race also play significant roles. These factors highlight additional dimensions crucial for skills and,

Table 10: Ordered logistic model, personal characteristics correlating with financial literacy. Source: SCF, 2016-2019, authors' calculations.

	Dependent variable:
	Financial literacy score
Worker	0.041^{*}
	(0.025)
Married	0.111^{***}
	(0.024)
Non-white	-0.392^{***}
	(0.019)
Female	-0.474^{***}
	(0.025)
Education: High-school	0.211^{***}
	(0.031)
Some college	0.599^{***}
	(0.031)
College degree	1.123^{***}
	(0.033)
Income percentile: 20^{th} - 40^{th}	0.049^{*}
	(0.028)
40^{th} - 60^{th} 3	0.073^{**}
	(0.031)
60^{th} - 80^{th}	0.179^{***}
	(0.035)
80^{th} - 90^{th}	0.349^{***}
	(0.043)
90^{th} - 100^{th}	0.649^{***}
	(0.048)
Pseudo R^2	0.134
Observations	60,125
Note: Controlling for age and asset amount.	*p<0.1; **p<0.05; ***p<0.01

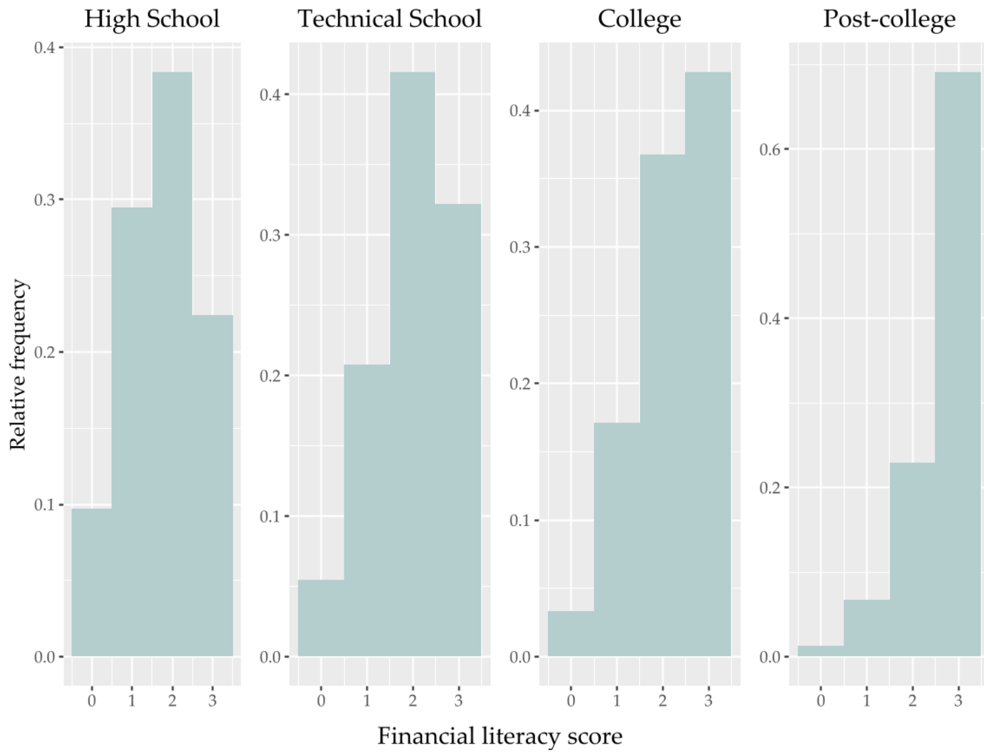


Figure 17: Financial literacy distribution by education level. Source: SCF, 2016-2019, authors' calculations.

consequently, individual saving and borrowing behaviors. We consider financial skills as a dimension that encompasses these conventional explanatory variables, albeit imperfectly, due to the impacts of learning by doing and unexpected expense shocks, as discussed in studies such as Agarwal et al. (2007) and Lusardi and Mitchell (2014).

2.3.2 Stylized facts from the SCF

While the separation of financial literacy from other household characteristics falls beyond the scope of this paper, we present key data patterns shedding light on individual financial skills and their potential impacts on mortgage shopping behavior. These patterns define a set of three empirical facts important for our model assumptions and validity.

First, we document that financial skills vary with age. We apply a polynomial fit to the standardized skill score across age groups. Although Figure 18 can not account cohort effects, the hump-shaped fit corresponds to panel data estimates depicting skill variations over time (see Agarwal et al. (2007) and Lusardi et al. (2010)). Indicative of a decline in consumer finance knowledge with approaching retirement, Figure 18 illustrates skill depreciation, corroborating findings from panel-data studies on financial sophistication.

The second empirical fact underscores the positive correlation between refinancing probability and financial literacy. Our analysis reveals that the likelihood of mortgage refinancing increases with higher financial skills and mortgage payments, holding other charac-

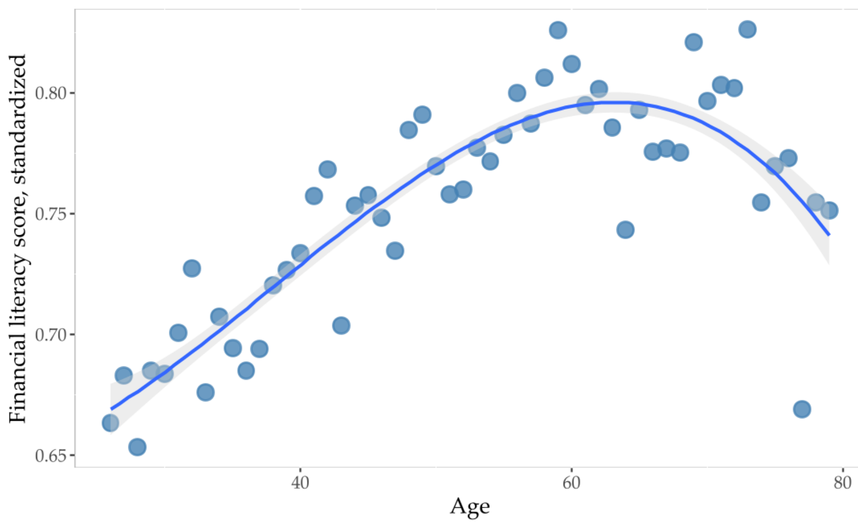


Figure 18: Average financial literacy by age groups, polynomial fit. Source: SCF 2016-2019, authors' calculations.

teristics constant. Variations in these probabilities are illustrated in the heatmap depicting predicted refinancing probabilities in Figure 19.

We evaluated the likelihood of mortgage refinancing among borrowers based on their self-reported search efforts in making borrowing decisions. With borrower attributes and mortgage size held constant, greater financial literacy, income, and effort imply a greater likelihood of mortgage refinancing (as illustrated in Table 33 in the Appendix). In contrast, Table 11 demonstrates that education does not significantly influence refinancing. Thus, financial skills emerge as a distinct dimension significantly impacting refinancing decisions within the SCF dataset.

Overall, coefficients in Table 11 imply that, across all income categories, financially savvy borrowers are 20%-30% more likely to refinance their mortgage.

Our third finding highlights a positive correlation between financial skills and the time households dedicate to credit shopping. Employing an ordered logistic model, we find that financially savvy renters and homeowners invest a significant amount of time in credit shopping, regardless of their housing expenses. The coefficient estimates are detailed in Table 12, and Figure 20 illustrates a heatmap showing model-predicted probabilities of spending a considerable amount of time searching for credit among renters. Households with strong financial skills tend to allocate more time to exploring credit opportunities, with a 15% increase in the likelihood of spending additional time for mortgage owners and a 10% increase for renters. Furthermore, our estimates indicate that renters, on average, dedicate less time to search efforts, and their search intensity shows a more gradual growth with higher levels of financial skills[20].

In the SCF, an average homeowner has over 70% of their total monthly debt obligations dedicated to mortgage repayments. Consequently, the specifics of a mortgage con-

[20] The heatmap of predicted probabilities for homeowners is available in Appendix B.3, Figure 57.

Table 11: Binary regression estimates, likelihood of refinancing. Source: SCF 2016-2019, authors' calculations.

	Dependent variable:
	Ever refinanced their mortgage
Financial literacy score: low	0.093
	(0.122)
medium	0.262^{**}
	(0.116)
high	0.478^{***}
	(0.115)
Search effort, borrowing: medium	0.055
	(0.056)
high	0.125^{**}
	(0.058)
Education: high school	−0.106
	(0.081)
some college	-0.222^{***}
	(0.081)
college degree	−0.089
	(0.080)
Female	0.103^{*}
	(0.057)
non-white	-0.280^{***}
	(0.037)
Mortgage size: $83,000 - $159,000	-0.170^{***}
	(0.047)
$159,001 - $ 297,000	-0.360^{***}
	(0.049)
$ 297,001 - $ 1,450,000	-0.394^{***}
	(0.054)
Constant	-0.869^{***}
	(0.175)
Pseudo R^2	0.077
Observations	18,702
Note: Controlled for age, income, family structure and survey wave effects.	*p<0.1; **p<0.05; ***p<0.01

Table 12: Ordinal logistic regression, time spent shopping for credit. Source: SCF 2016-2019, authors' calculations.

	Low-to-great deal of spent in shopping for credit(1-3)	
	Homeowners	Renters
Low|Medium	-15.343^{***}	0.439^{***}
	(0.236)	(0.086)
Medium|Great	-18.042^{***}	-1.748^{***}
	(0.237)	(0.090)
Mort. payment per month: -$750-$1150	-0.017	
	(0.049)	
$1150-$1700	0.038	
	(0.053)	
$1700-$2700	0.0314	
	(0.060)	
$2700+	0.071^{***}	
	(0.056)	
Rent payment per month: $500-$690		-0.132^{**}
		(0.046)
$690-$920		-0.058
		(0.047)
$920-$1300		0.029
		(0.048)
$1300+		0.0385
		(0.052)
Education: HS	0.421^{***}	0.373^{***}
	(0.074)	(0.048)
some college	0.436^{***}	0.612^{***}
	(0.074)	(0.048)
college degree	0.437^{***}	0.565^{***}
	(0.075)	(0.053)
Wage percentile: 20-40	-0.0368	0.147^{**}
	(0.059)	(0.051)
40-60	-0.016	0.140^{*}
	(0.061)	(0.056)
60-80	-0.051	0.122^{*}
	(0.063)	(0.058)
80-100	-0.097	0.260^{***}
	(0.068)	(0.062)
Financial literacy: level 1	0.256	0.090
	(0.112)	(0.065)
level 2	0.400^{***}	0.161^{***}
	(0.106)	(0.062)
level 3	0.350^{***}	0.360^{***}
	(0.105)	(0.064)
Pseudo R^2	0.015	0.029
Observations	22,178	19,610

Note: Controlled for gender, race, age, debt-to-income, risk attitudes, assets and survey wave effects. *p<0.1; **p<0.05; ***p<0.01

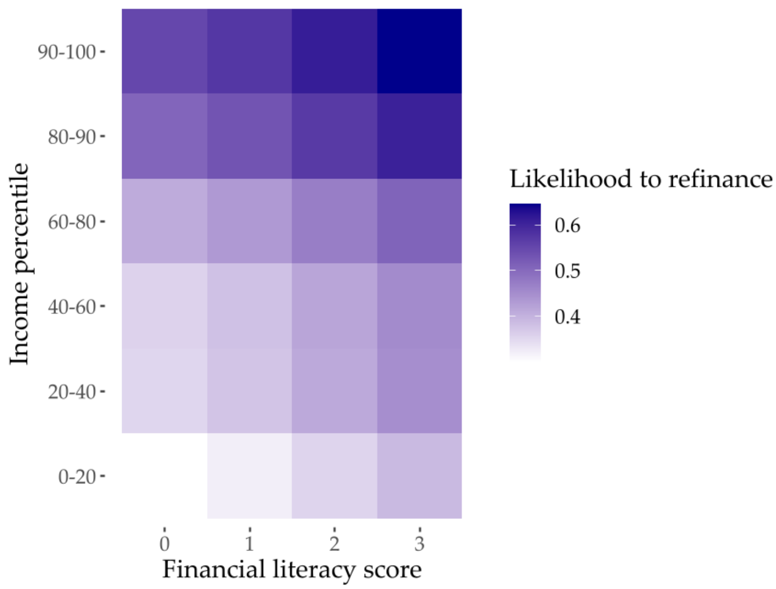


Figure 19: Mortgage refinance likelihood across income percentiles and financial literacy scores. Source: SCF 2016-2019, authors' calculations.

Figure 20: Great deal of time spent shopping for credit, ord. logit predictions, renters only. Source: SCF 2016-2019, authors' calculations.

tract significantly influence expenditure and savings patterns throughout their working years, deeply impacting available liquidity. In this context, we obtain a dataset that is comprehensive, encompassing detailed information on both the mortgage contract and household characteristics. Shifting our attention to mortgage data, we gain insights into individual mortgage shopping behavior. Individual shopping behavior, coupled with a standard set of observable factors, determines the mortgage interest rate, which frequently remains fixed over the mortgage term. Through our model, shopping behavior shapes spending and saving patterns over the 30-year mortgage duration. To substantiate our assumptions regarding mortgage search, we base the majority of our model assumptions on our new U.S. data findings.

2.3.3 The National Survey of Mortgage Originations (NSMO)

Our novel data set leverages the amount of information within the NSMO. For a representative sample of U.S. population, NSMO connects mortgage registry data to the survey on mortgage acquisition experience, spanning mortgage originations from 2013 to 2021. The dataset includes newly originated first-lien residential mortgages, covering both initial acquisitions and refinanced mortgages.

Important for our paper, the survey inquires about loan shopping behavior and the overall consumer experience during the mortgage process. All survey responses are matched with institutional lender data, providing specific details of the mortgage contract, including locked-in mortgage rates, government sponsorship, low-income area indicators, loan-to-value ratios (LTVs), borrower's payment-to-income ratio, credit score, education, and income. We limit the data to home purchases and refinancing, resulting in a survey sample of 43,094 mortgages, each weighted to ensure representativeness in our analysis.

Our focus revolves around borrowers' search behavior prior to the mortgage application. We use the question

- *How many different mortgage lenders/brokers did you seriously consider before choosing where to apply for this mortgage?*

The individual survey responses serve as a proxy variable for the cognitive search effort. Instead of relying on the number of formal mortgage applications, we analyze the number of lenders considered. We argue that the response reveals the variation in the cognitive search effort **prior to the application process**.

While the majority of borrowers tend to submit formal applications to a single lender – resulting in over 35,000 mortgages being obtained from that chosen lender – the number of lenders seriously taken into account varies across the sample. We assert that, due to the expense associated with the application process, borrowers concerned about rejection are more likely to formally apply to multiple lenders, driven by fear of being declined, similar to the mechanism discussed in works such as Agarwal et al. (2020). Consequently, the number of lenders considered reveals shopping behavior that provides deeper insights into cognitive efforts invested in the attainment process. Important for our paper, approximately 70 percent of the survey respondents undergo the mortgage process without the use of a mortgage broker.

Furthermore, the number of lenders considered reflects the contemporary approach to mortgage search. Online applications typically compare various lenders and "recommend" the optimal choice, considering the borrower's credit score, income, and down payment options [21].

[21] For instance, a consumer can visit https://www.bankrate.com/mortgages/mortgage-rates/ and list their current or desired mortgage amount to compare rates across lenders.

In the paper appendix, we demonstrate how search effort relates to differences in individual characteristics, and show that educated borrowers more often consider multiple lenders before submitting a formal application. We also show that locked-in mortgage rates fluctuate in relation to education and search effort, and highlight the importance of search effort in mortgage rate attainment.

Our mechanism goes beyond education and focuses on individual financial knowledge. Education is positively however not perfectly correlated with financial knowledge, and the latter is shown to significantly affect individual financial decisions. Leveraging the matched dataset, we introduce the concept of **effective search** among borrowers with higher financial skills. Figure 21 depicts search variation related to the education level of individuals. The rest of our analysis remains concentrated on financial skills and utilizes the matched data.

After the mortgage origination, the NSMO tracks individual mortgage performance until loan closure. Conditional on averages in other borrower characteristics, our estimates use the merged data set and underline financial skills and search behavior as being significant in predicting meeting payment due dates.

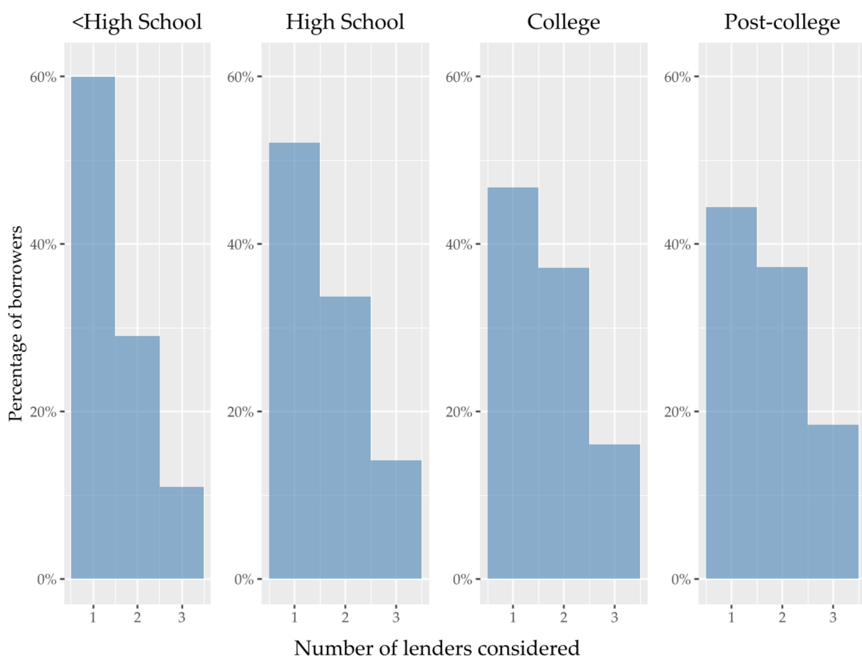


Figure 21: Number of lenders considered by education level. Source: NSMO data set, authors' calculations.

2.4 Stochastic imputation, mortgage data extended (NSMO+)

Information regarding individual mortgages is limited within the SCF. Beyond mortgage payments and past refinancing behavior, data on a mortgage contract is unavailable. To overcome this limitation, we employ stochastic matching to integrate the two datasets. By doing so, we maximize the utility of publicly accessible information about mortgage contract specifics and individual skills, and account for the uncertainty inherent in the matching process.

Instead of imputing financial literacy scores deterministically, the BRL method estimates the distribution of financial skill level for every borrower in the NSMO. Based on the set of mutual observables, we obtain Bayesian weights for every match between NSMO and the SCF and use them later for making statistical inferences. This method has been analytically shown to reduce the biases in coefficient estimates in linear models and preserve asymptotic normality and consistency in non-linear estimation (Enamorado et al., 2019). We outline the BRL assumptions and likelihood formulation in section B.4 of the Appendix.

Our paper is the first to link SCF and NSMO. Record matching allows us to estimate the financial skill distribution, for every NSMO borrower. While Bayesian weights control for the imputation-driven bias, details of the mortgage contract allow us to control our estimates for other borrowers and mortgage specifics. In this way, our estimates reflect potential sources of the mortgage rate dispersion among otherwise similar borrowers who apply for similar contracts. Table 13 outlines population shares in respective data sources[22].

Common observables that define the likelihood of record match are measures related to individual financial skills, commonly used in correlation analyses in empirical studies (Bhutta et al., 2022; Lusardi and Mitchell, 2014). We depict their relative importance using the R^2 decomposition in the paper appendix, depicted in table 36. Common observables include income, education, gender, age, race, occupation, family characteristics, and presence of the retirement plan and asset holdings. Once we have a borrower-specific skill distribution, our estimates separate skilled and unskilled borrowers who search more or less, and keep the lender's side of the contract fixed (term, amount, government sponsorship, origination year, etc.)

2.5 NSMO+ data findings

In this section, we outline joint patterns in mortgage rates, individual search efforts, and financial skills, and discuss individual mortgage performance across skill levels. Initially, we discuss the importance of financial skills and their role in how much search effort is exerted **prior to** the mortgage application. Next, explore the interplay between financial skills, search effort, and mortgage rates, and introduce the concept of **effective** search among skilled borrowers. Lastly, we focus on repayment behavior heterogeneity across different skill levels. We return to our empirical estimates in the model's steady-state analysis and align the model-driven patterns to our merged data findings.

Our findings boil down to a set of correlations that motivate a mortgage search model which is the topic of the third chapter. We control for many characteristics, both on the household and mortgage side, to appropriate the discussion on causality. Without a time dimension in financial literacy data, we cannot interpret our results as causal. We plan to make a step ahead using life-cycle events in the SCF data.

[22]To circumvent a part of the self-selection issue in mortgage uptake, we match borrowers to the NSMO with first-lien mortgage owners in the SCF. This way, we make two data sets more comparable.

Table 13: Population shares in the respective sample. Source: NSMO 2013-2022 and SCF 2016-2019, authors' calculations.

	Data set	
	NSMO	SCF
income brackets	[6%, 9% , 18%, 19%, 30%, 18%]	[13%, 8%, 13% ,11%,20%, 35%]
education brackets	[1%, 10%, 5%, 20%, 35%, 29%]	[6%, 18%, 9%, 15%, 27%, 25%]
gender (Female, Male)	[44%, 55%]	[17%,83%]
age (<35,35-44,45-54,55-64,65-74,>=75)	[18%, 22%, 22%, 21%, 14% ,3%]	[8%, 14%, 20%, 26% , 20%, 12%]
race (Caucasian, African-American, other)	[84%, 6%, 10%]	[82%, 7%, 11%]
occupation (Employed, Self-employed, Retired/Student, Other)	[68%, 10%, 19% ,2%]	[47%, 26%, 25%, 2%]
has children (Yes, No)	[64%, 36%]	[60% , 40%]
owns financial assets (Yes, No)	[57%, 43%]	[58% 42%]
retirement plan participation (Yes, No)	[86%, 14%]	[62%, 38%]
Number of observations	43,094	40,515

2.5.1 Search, financial skills and locked-in mortgage rates

Using imputed financial skills, we find that financially savvy borrowers consider more lenders on average, and show that search effort variation patterns resemble the breakdown by education level (see Figure 54 in section B.2 of the Appendix). Moreover, we find that savvy applicants search more effectively and generally secure lower mortgage rates in comparison to their comparable counterparts.

In our sample, we redefine the number of lenders considered and bin 3, 4, and 5+ together, and represent it with 3+. Our estimates show that while 60% of low-skilled borrowers focus on only one lender, and only 10% on three or more lenders, 58% of financially savvy borrowers consider multiple lenders (Table 14).

Table 14: Number of lenders considered across financial skills, weighted frequencies. Source: merged dataset, authors' calculations.

	Number of lenders considered		
	1	2	3+
Financial Literacy			
Low	58.48%	41.52%	0
High	41.37%	36.42%	22.21%

Next, we estimate a ordinal logistic model that assumes latent thresholds for every observation ij in the merged data set

$$\mathbb{P}(\text{num\_cons}_{ij} = k) = p_{ij,k} = \mathbb{P}(-\kappa_{k-1} < \beta X_i + \beta^f \text{fin\_skills}_j + u_{ij,k} < \kappa_k), \quad k \in \{1, 2, 3+\}.$$

We adjust our estimates with borrower-skill specific distributional weights that account for match uncertainty in the inflated set of 155,500 observations[23].

Table 15 depicts the explanatory power of each borrower characteristic. Important to our narrative, our estimates imply that financially skilled borrowers (top tercile) are 4% more likely to consider more lenders i.e., search more. Moreover, we find that females and borrowers living in non-metropolitan areas are 30 and 5 percent less likely to consider multiple lenders. Additionally, education significantly affects search effort, as we find that college graduates and post-college borrowers are 40% and 50% more likely to search more, respectively.

Search effort correlates negatively with low-to-moderate non-metropolitan areas, known as low-shopping areas, which are often subject to mortgage overpricing (Bartlett et al., 2022). Notably, the effect of financial skills is of the same magnitude as income or credit score, or the geo-location effect[24]. Abstracting from all standard observables leaves a significant residual effect of financial skills. However, the skills effect in our estimates remains conservative due to the nature of our merging process and strong correlations between skills and gender, income, education, etc. outlined in the SCF data analysis.

2.5.2 Residual mortgage rate dispersion and repayment costs heterogeneity

Controlling for the loan amount, term (30 years), borrower's credit score (*"Very good" and "excellent"*) and the origination year (fixed to 2016), we compare the residual mortgage rate dispersion across different levels of financial skills. Even though these borrowers are comparable to mortgage lenders, financially savvy ones tend to lock in at lower rates. Figure 22 shows that the interest rate density for the savviest borrowers (denoted with the blue curve) has a lower mean, and is thicker towards lower interest rates. On the other hand, unskilled borrowers are more likely to end up with higher interest rates, as shown in Figure 22 with the red density graph.

Using the 2020 origination sub-sample, we show that, for a \$200,000 loan, the top tercile of financially skilled borrowers secured mortgages with a **20 percent lower spread in the mortgage rate distribution**, underscoring the larger variation in interest rates obtained by low-skilled borrowers, depicted by Figure 22. This pattern holds consistently over time, with the usual spread difference ranging between 15% and 20%.

Next, we regress the locked-in interest rate on a set of borrower characteristics X_i, mortgage contract specifics M_i and match-based financial skills fin_skills_i:

$$\text{rate}_i = \alpha + \beta X_i + \beta^m M_i + \beta^f \text{fin\_skills}_i + \gamma \text{fin\_skills}_i \times \text{num\_len}_i + \varepsilon_i,$$

and estimate the rate-based losses over the mortgage duration.

[23] We repeat the analysis with the linear probability model that does not require weights inclusion and obtain similar results

[24] In addition, our SCF analysis shows significant variation of credit search effort with financial literacy, with 20% greater likelihood for high-skilled borrowers to spend more time in loan shopping. The two findings together support our search model assumptions.

	Dependent variable: # of lenders considered		
	Coefficient	SE	z score
(Intercept):1|2	-0.4515^{***}	0.0947	-4.7665
(Intercept):2|3	-2.1960^{***}	0.0950	-23.1239
Financial literacy	0.0444^{**}	0.0216	2.0616
Age	-0.1603^{***}	0.0143	-11.1923
Credit score	0.0515^{***}	0.0146	3.5298
Female	-0.2904^{***}	0.0141	-20.5282
Race: non-white	0.2426^{***}	0.0198	12.2247
Income:			
$\$35,000 - \$49,999$	-0.0262	0.0379	-0.6922
$\$50,000 - \$74,999$	-0.0312	0.0356	-0.8767
$\$75,000 - \$99,999$	-0.0172	0.0364	-0.4734
$\$100,000 - \$174,999$	-0.0351	0.0362	-0.9685
$\$175,000+$	-0.0227	0.0401	-0.5659
Metropolitan area:			
Low-to-moderate income	-0.0176	0.0215	-0.8195
Non-metropolitan area	-0.0517^{*}	0.0237	-2.1834
Loan Amount:			
$\$100,000$-$\$199,999$	0.0852^{***}	0.0231	3.6859
$\$200,000$-$\$299,999$	0.1864^{***}	0.0260	7.1664
$\$300,000$-$\$399,999$	0.2337^{***}	0.0305	7.6579
$> \$400,000$	0.3157	0.0324^{***}	9.7351
Education:			
some college	0.2657^{***}	0.0249	10.6772
college	0.4228	0.0247^{***}	17.1297
post-college	0.5302^{***}	0.0264	20.0973
Pseudo R^2			0.012
Observations			155,500
Note: controlled for year effects.			$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 15: Ordered logit with imputed financial literacy and weights.

Table 16 displays coefficients for two sets of estimates, with the first column focusing solely on first originations. In both regressions, we account for mortgage specifics, including loan type, amount, term, sponsorship, number of underwriters, and loan-to-value ratios. Notably, both sets of estimates reveal an interaction between financial literacy and search effort, significantly contributing to the explanation for locked-in mortgage rates.

Initially, our findings align with those of Agarwal et al. (2020), showing that fear of application rejection mechanically amplifies search efforts among first originations, ultimately leading to higher average rates. This is highlighted in Table 16, which reveals a significant and positive coefficient of 0.220 for search effort within the context of first originations. Upon

Table 16: Mortgage rate regression, controlling for loan and borrower characteristics. Source: merged data set, authors' calculations.

	mortgage rate	
	(First origination)	(All mortgages)
#Lenders considered: two	0.034	−0.006
	(0.087)	(0.062)
#Lenders considered: three	0.220*	0.125
	(0.120)	(0.083)
Financial skills	0.017	−0.016
	(0.088)	(0.060)
Considered 2 lenders× fin skills	−0.072	−0.023
	(0.113)	(0.080)
Considered 3 lenders × fin skills	−0.354**	−0.220**
	(0.153)	(0.106)
Age	0.044***	0.062***
	(0.010)	(0.007)
Metro area - LMI tract	0.033**	0.022**
	(0.013)	(0.009)
Non-metro area	−0.018	0.003
	(0.015)	(0.010)
Female	0.032***	0.030***
	(0.009)	(0.006)
African-American	−0.005	0.007
	(0.019)	(0.013)
Asian	−0.021	−0.036***
	(0.020)	(0.013)
Other (including hispanic)	0.069***	0.051***
	(0.025)	(0.017)
Income: $35,000-$50,000	0.007	−0.043**
	(0.024)	(0.017)
$50,000-$75,000	0.036	−0.018
	(0.023)	(0.016)
$75,000-$100,000	0.034	−0.011
	(0.024)	(0.017)
$100,000-$175,000	0.064***	0.004
	(0.024)	(0.017)
$175,000 and more	0.054**	−0.00004
	(0.027)	(0.019)
Education: high-school	−0.054***	−0.033***
	(0.017)	(0.011)
college graduate	−0.105***	−0.071***
	(0.017)	(0.012)
post-college graduate	−0.131***	−0.090***
	(0.019)	(0.012)
Refinancing		−0.074***
		(0.007)
Credit score	−0.263***	−0.247***
	(0.010)	(0.007)
Constant	5.269***	4.955***
	(0.099)	(0.066)
Observations	21,461	43,084
R^2	0.369	0.440
Adjusted R^2	0.368	0.439
Residual Std. Error	23.662 (df = 21412)	22.325 (df = 43034)
F Statistic	260.809*** (df = 48; 21412)	689.013*** (df = 49; 43034)

Note: Controlled for loan type, government-sponsored enterprise, loan amount, number of borrowers, time effects, LTV and term. *p<0.1; **p<0.05; ***p<0.01

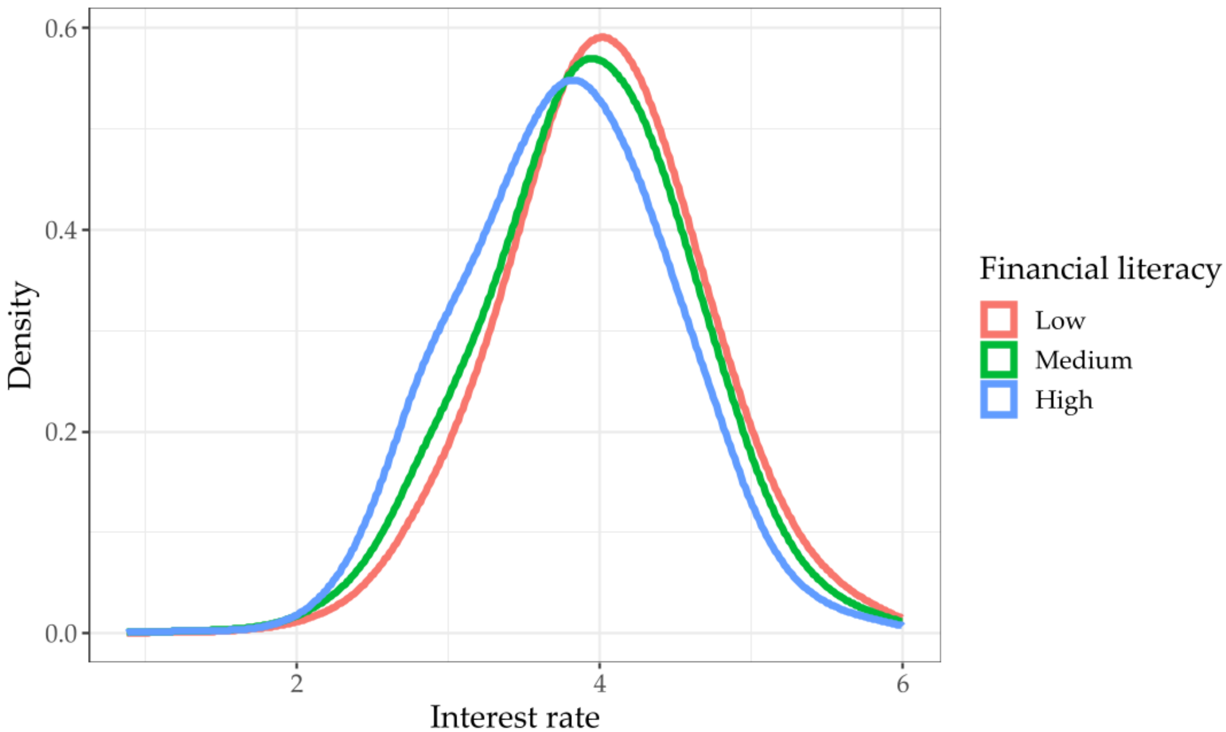


Figure 22: Mortgage rate variation in 2020 among best credit score borrowers. Source: merged data set, authors' calculation.

interaction with skills, the intensity of the search assumes the role of an informed mortgage search. Financially skilled borrowers who explore a wider range of lenders tend to secure lower mortgage rates. This translates to an average rate difference of 13.4 basis points (with a corresponding coefficient of 0.220-0.354=-0.134).

Our supplementary findings align with existing research employing loan-level data, underscoring that female and Hispanic borrowers often encounter higher mortgage rates. On the flip side, individuals with higher education enjoy, on average, a reduction of 13.1 basis points in rates during initial originations, though this effect decreases during refinancing. As we consider the intricate interplay among skills, gender, race, and education, our estimates concerning skill disparities present a cautious estimate of the minimum divergence in mortgage repayments, subsequently impacting differences in consumption after accounting for mortgage payments.

Nevertheless, when we analyze the variations in search effort and interest rate regressions, it becomes evident that the extent and effectiveness of search effort differs based on financial skills. This implies the likelihood of lower mortgage payments among financially skilled yet comparable borrowers.

2.5.3 Effective search

We emphasize the role of effective search and compare our predicted distributions of locked-in rates between borrowers who engage in extensive searches and those who consider one lender only. Figure 23 depicts mortgage rate distributions across two scenarios. Low-skilled borrowers that search more effectively do not gain from the search, as the mortgage rate distribution stays the same (left panel in Figure 23). In contrast, high-skilled borrowers who search more end up with lower rates (depicted by the blue curve in the right panel of Figure 23), rendering their search as effective. Our findings on search effectiveness, coupled with a significant and positive search coefficient in the interest rate regression (Table 16), align with the fear of rejection mechanism among low-income borrowers in Agarwal et al. (2020). Less financially savvy borrowers search more because they fear rejection. As a result, this does not significantly change their mortgage rates compared to those who put in less effort.

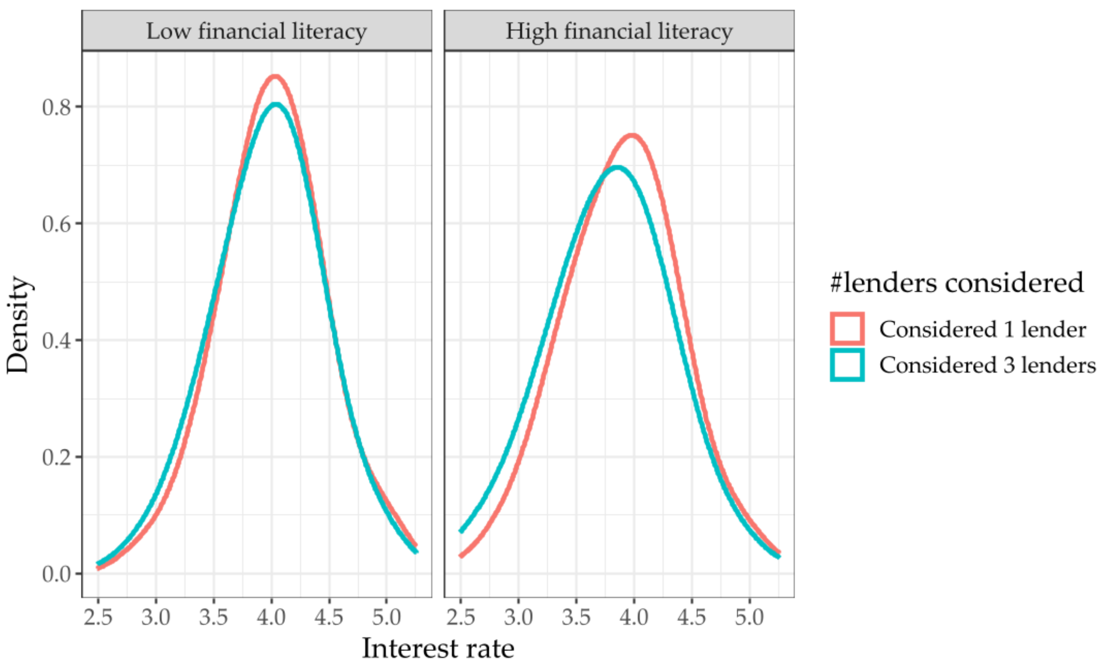


Figure 23: Mortgage rate dispersion; interaction of search effort and financial skills. High skilled borrowers who exert more search effort generally lock in at lower mortgage rates. Source: merged data set, authors' calculation.

The disparities observed in lock-in rates during the origination phase ultimately translate into compounded losses over the entire mortgage term[25]. To illustrate, for a $100,000 loan with a standard duration, an average borrower with high financial skills can secure a rate of approximately 3.8%, compared to 4.05% for those with lower financial skills. This sets the lower boundary for cumulative losses at $6,693 over the mortgage term. Moreover, the additional impact of low search effort introduces more than $2,636 in costs throughout

[25]Over 75% of mortgages in our sample are 30 years fixed-rate mortgages.

the mortgage term. These estimates, though not accounting for other correlations among borrower characteristics, stand as conservative approximations for losses in the mortgage market, amounting to at least $9,329. Notably, this represents a significant proportion of the losses derived from institutional data and subjective insights into the mortgage process (Bhutta et al., 2020). Given that mortgage repayments accounts for over 70% of monthly debt payments, addressing these losses is an imperative for bolstering liquidity for all households, especially those with lower incomes.

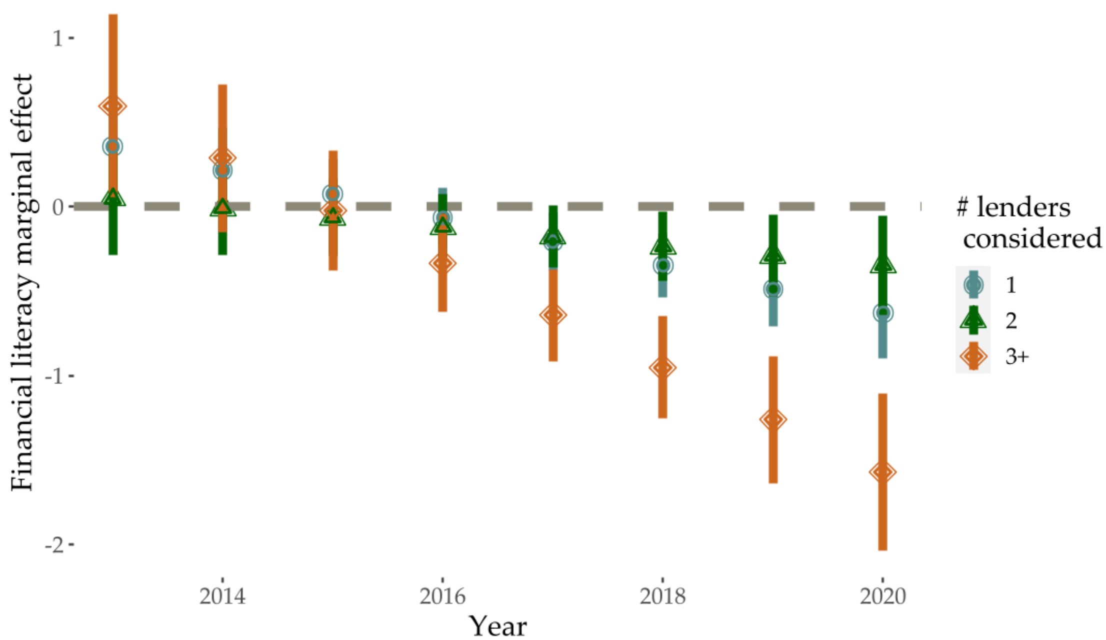


Figure 24: Financial skill coefficient in the mortgage rate regression, differences over the sample period. Source: merged data set, authors' calculations.

Figure 24 represents the year and financial skills interaction coefficient over the sample period. Relative to the first year in the sample, 2013, later mortgage origination years show signs of increasing significance of both financial skills and search effort for mortgage rate attainment. Our sample period is marked by the steady increase in non-bank lenders share in the mortgage market. As these lenders turn to online advertising and borrowing (Bhattacharya et al., 2021), our findings are suggestive of increasing effects of skilled search effort amidst the mortgage options expansion.

2.5.4 Mortgage performance after origination

NSMO+ tracks the individual mortgage performance until the loan closure, with scores denoting missing repayment due dates up to and over 180 days, bankruptcy levels based on U.S. law, and regular payments made on time. Specifically, the data set separates scores for late payments up to 150 days, and the worst scores indicates mortgage payments later than 150 days and defaults.[26].

The sample size constrains our analysis of the default and late payment indicators, so we separate the score values for late payments and defaults from regular payments and define the indicator variable $\mathbf{1}_{\{\text{late payments or defaults}\}}$. We quantify the effect of individual financial skills and search effort at the time of origination using the linear probability model estimation that controls for other observables.

We model the probability as

$$\mathbb{P}(\text{late with payments}) = \alpha + \beta X_i + \beta^f \text{fin\_lit}_i + \beta^s \text{search\_effort}_i + \varepsilon_i,$$

where fin_lit_i is the average skill amount across all matches[27]. We regress the indicator on a set of borrower observables, mortgage characteristics, individual financial skills, and search effort at the time of origination.

We standardize all continuous regressors (age, credit score, payment-to-income ratio) and compare the size of the coefficients. Our estimates are presented in Table 17.

Table 17 conforms to the standard intuition regarding household characteristics prevalent for mortgage performance. While borrowers with greater payment-to-income ratio are more likely to be late, those with higher credit scores are more likely to meet their payment due dates. In line with Gerardi et al. (2023) and Bhutta et al. (2020), we find that non-white borrowers are more likely to be late with payments. Importantly for our paper, financially skilled borrowers who exerted more effort are less likely to have been late on payments two years after mortgage origination.

Figure 25 plots default prediction differences across different skill and search levels. Specifically, our predictions state that financial unskilled face a 1.6 p.p. greater likelihood of being late with mortgage payments. Added to this, borrowers who considered one lender are 0.2 p.p. more likely to be late with payments, possibly because they secured their mortgages at higher rates. Put differently, getting one more question wrong in the financial literacy test corresponds to being 40%-50% more likely to not meet mortgage repayment dates three years after the origination.

The patterns identified through our analysis of the SCF and NSMO+ serve as the foundation for a mortgage search model that accounts for the variation in search costs contingent on individual financial skills. We revisit each of these findings within the framework of our model in our subsequent work.

[26]According to the Home Mortgage Disclosure Act data, delinquency rates are reliable indicators of mortgage default. https://www.consumerfinance.gov/data-research/mortgage-performance-trends/mortgages-30-89-days-delinquent/

[27]We perform a separate, score-based analysis that shows significance and a similar effect size.

Table 17: Late payment probability, linear model. *Source:* merged data set, authors' calculation.

	$\mathbb{P}$*(Late payment)*
Loan Amount: \$100,000-\$199,999	0.0001
	(0.002)
\$200,000-\$299,999	-0.004^{**}
	(0.002)
\$300,000-\$399,999	-0.004^{**}
	(0.002)
> \$400,000	-0.005^{***}
	(0.002)
Financial literacy	-0.017^{**}
	(0.007)
Multiple lenders considered	-0.002^{**}
	(0.001)
Female	0.002^{*}
	(0.001)
Education: high-school	0.003
	(0.002)
college	-0.0001
	(0.002)
post-college	-0.0002
	(0.002)
Race: non-white	0.005^{***}
	(0.001)
Age	0.002^{*}
	(0.001)
Payment-to-income	0.005^{***}
	(0.001)
Credit Score	-0.020^{***}
	(0.001)
Constant	0.023^{***}
	(0.005)
Observations	43,084
Adjusted R^2	0.017
F Statistic	54.783^{***} (df = 14; 43069)
Note: all variables are standardized to preserve interpretability.	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

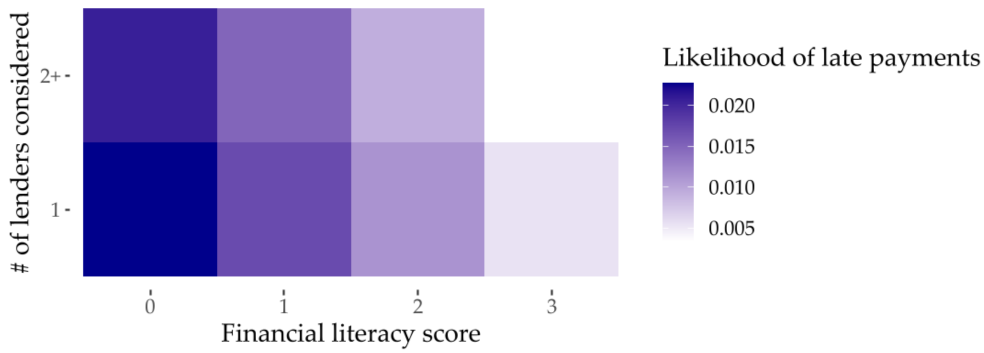


Figure 25: Likelihood of late payments across effort and financial skills. Source: Probability model predictions, merged data set, authors' calculation.

2.6 Conclusion

Our paper contributes to the empirical literature on mortgage undertaking in two ways. First, we employ the stochastic record linkage procedure and merge the National Survey of Mortgage Originations with the Survey of Consumer Finances. This effectively creates a new data set on mortgages that incorporates objective financial literacy scores. Second, we leverage the statistical properties of the merging procedure and investigate the joint correlation between individual financial literacy and search effort in the mortgage undertaking process while accounting for specific record link uncertainty. Third, our findings motivate a novel search mechanism that connects individual financial literacy to mortgage repayment.

Our data estimates show that financially skilled households more often seriously consider multiple lenders, showing signs of an effective search procedure. Moreover, we show that financial literacy and loan search effort interact and explain a part of the mortgage rate variation. Specifically, skilled borrowers who search more end up getting a 13.4 b.p. on average lower interest rate at the time of the loan origination. Using back-of-the-envelope calculations, we estimate the lower bound for potential losses from an unskilled search - for a $100,000 loan, financially unskilled borrowers lose at least $9,329 over the thirty-year mortgage span.

Our paper speaks to behavior after a mortgage is originated. Using our novel data set, we show that financially unskilled households face a 34-45% greater likelihood of becoming delinquent three years after the mortgage originated, irrespective of their payment-to-income ratio. This finding, coupled with our findings on lower refinancing probability among financially unskilled households, motivates the importance of the mortgage search mechanism for consumption differences across similar borrowers.

3 Financial Skills and Search in the Mortgage Market

- co-authored with Ante Šterc

3.1 Introduction

The period of low interest rates spanning from 2010 to 2020 witnessed a notable rise in the mortgage market and accelerated entry of non-bank lenders, leading to increased accessibility and speed in mortgage acquisition (McCafrey, 2021). This trend was marked by a significant shift towards non-banks, with non-bank lenders capturing 70% of the first-lien U.S. mortgage market in 2021 (Degerli and Wang, 2022). The entry of new lenders led to more relaxed requirements for potential borrowers, including lower credit score thresholds for mortgage approval (Cornelli et al., 2022). The combination of greater availability and relaxed requirements enabled younger and less experienced borrowers to enter the U.S. housing market.

Chapter 2 of this thesis presents empirical estimates derived from a unique U.S. dataset and indicates that discrepancies in borrower search effort and financial knowledge contribute to residual variations in mortgage rates. We integrate endogenous investment in financial skills and search effort into the mortgage search model to characterize variations in mortgage choice.

Our novel search framework, embedded within a heterogeneous agents model, allows us to link differences in financial skills to consumption disparity via the mortgage repayment channel. In the model, individual mortgage attainment depends on endogenous financial skills and search intensity, conditional on the borrower's savings level. Our framework generates empirically plausible disparities in non-durable consumption and aligns with refinancing and delinquency probabilities across borrowers' financial skills.

Subsequently, we use our model to conduct counterfactual analyses that underscore the potential impact of financial education on mortgage attainment and repayment ability. We show that financial education mitigates the adverse effects of higher mortgage accessibility on less financially skilled homeowners. Finally, our findings reveal that low mortgage rates disproportionately favor highly skilled homeowners, leading to increased refinancing activity and perpetuating consumption disparities across different financial skill levels.

Our paper uniquely contributes to the existing literature through a novel structural framework incorporating our key findings on heterogeneous **cognitive** search costs. Borrowers invest in financial skills, which reduce the cognitive costs of searching for mortgages and affect subsequent mortgage performance. In the steady state, borrowers with greater financial skills actively search for mortgages, explore a broader range of offers, and lock in at lower mortgage rates. Conversely, financially unskilled borrowers are less inclined to participate in the mortgage search. When they do engage, their limited search efforts result in random, higher mortgage rates, making the mental effort of search comparable to the advantages of renting. In conjunction with individual search and skills, the distribution of mortgage offers endogenously defines the lock-in mortgage rate distribution we use for model calibration.

The model delivers a mortgage repayment schedule across the joint distribution of financial skills and assets among otherwise similar borrowers (intensive margin). At the extensive margin, the model delivers differential housing costs that, together with savings choices, collectively describe consumption differences between renters and homeowners.

Our analytical results tease out the effect of individual financial skill levels on consumption growth. We break down consumption growth into three channels: time preference, expected mortgage rate change, and precaution due to potential future expenses. While expected changes in mortgage repayments disincentivize saving, potential future expense shocks induce saving, with the most substantial effect on the financially savviest borrowers.

We calibrate the model using a set of key data moments from our previous empirical work and perform validity checks using non-durable consumption data from the external Bureau of Labor Statistics data set. Our model reproduces empirical patterns in mortgage rate attainment, with search effort and skills explaining 55% and 10% of the mortgage rate dispersion. Financially skilled homeowners engage in more intensive searches and are 30% more likely to refinance. On average, renters accumulate lower skill levels, reflecting data patterns from the SCF.

Our model experiments deliver key findings on changes in consumption inequality prompted by search cost changes and exogenous mortgage rate shifts. First, we introduce financial education, effectively reducing skill investment costs among low-skilled agents. We set up a policy test to approximate a 90-minute course in financial planning for low-skilled renters. Promoting financial skill accumulation increases the average skill level by 9% on average and reinforces search intensity and mortgage take-up, leading to a 1.6% greater share of homeowners overall. As relatively more skilled renters enter the mortgage market, the average delinquency rate is 2.8% lower than the benchmark. Moreover, because investment costs flatten out across all agents, the consumption inequality is, relatively, 1.4% lower.

Our second experiment accommodates mortgage market advancements and increases mortgage availability, effectively decreasing search costs for all agents in the economy. Corresponding to empirical findings (Degerli and Wang, 2022), accessible mortgages reflect a relative increase in search intensity by 7.8% for renters and 16.9% for homeowners. Mortgage accessibility mainly works in favor of current homeowners. We also show that accessible mortgages expose households to delinquency due to low incentives for skill accumulation (with a relative increase of 1.1%). The relative increase in the delinquency rate is 1.7%.

The relative increase in the average delinquency rates reflects the adverse effect of increased mortgage access. We show that financial education has a stronger effect within the accessible mortgage environment, leading to a relatively higher (0.4 p.p.) average financial skill level. Lower cognitive search costs reincentivize skill accumulation. Increases in mortgage availability render financial education effective in reducing consumption inequality by 1.5% and decreasing the average delinquency rate by 2.7%.

Our third experiment relates exogenous average mortgage rate change to changes in consumption inequality. We compare two scenarios: a low-mean rate scenario, marked by a 20 b.p. decrease in the average mortgage rate, and a high-mean rate scenario, characterized by a 10 b.p. increase in the average rate.

We show that the low-rate scenario benefits existing homeowners, leading to a 64.9% increase in refinancing activity. Therefore, homeowners secure lower mortgage rates and reduce their housing expenses. However, renters invest only 1.4% more in search activity, and often end up with higher rates or staying in rentals. This increases consumption by 1.4%. Lower mortgage rates, in this context, perpetuate the gap in consumption between renters and homeowners.

Conversely, the high-rate scenario exhibits a 36.5% decrease in search intensity among current homeowners. The increase in mortgage rates narrows the consumption disparity between renters and homeowners, leading to a 5.6% reduction in consumption inequality. Both scenarios underscore the crucial role of search intensity and the sensitivity of credit searches to interest rates.

Although changes in the U.S. mortgage market have tightened the gap between mortgage rates among similar borrowers, they need to be more effective with low-skilled borrowers. Our model experiments offer compelling evidence showing that promoting individual investments in financial skills could be crucial in addressing these persistent disparities. With accessible mortgages and a better understanding of the mortgage process, attaining lower mortgage repayments reduces the exposure of financially unskilled households to liquidity constraints.

Lastly, within the pool of financially savvier households, the diminishing utility cost of searching for new mortgage options reinforces refinancing activity. This observation hints at the amplified potency of the refinancing channel of monetary policy. A richer set of sources of heterogeneity and careful outlining of the mortgage supply can yield insights into mortgage market responses to financial education and monetary policy.

3.2 Related Literature

This paper contributes theoretical studies on mortgage undertaking and financial literacy effects on consumption and budgeting, leveraging the current way U.S. households face the mortgage process.

Following the structural changes in mortgage lending, the main focus has been put on consumer choice and search. The closest two papers to ours introduce hidden information or heterogeneity in rate beliefs while keeping i.i.d costs of search. Whereas Agarwal et al. (2020) introduces a model with search and screening and reproduces "the searching for approval" mechanism, we leverage FinTech algo pricing and assume perfect screening. Alexandrov and Koulayev (2018) incorporate a static framework with borrowers who hold beliefs about the interest rate dispersion, while we assume perfectly informed borrowers. In this respect, we complement Alexandrov and Koulayev (2018) in two ways. First, we add structure to search cost variation as opposed to taking an i.i.d. cost assumption. Second, we endogenize search costs as they depend on individual accumulation of financial skills. We add to the line of search models go beyond the mortgage take-up, and include the choice to refinance.

The data availability during the low interest rate for the last ten years shifted focus on refinancing. Andersen et al. (2020) argue that search frictions induce failure to refinance, attributing search frictions to behavioral factors such as inattention. Keys et al. (2016)

find that more than 20% of U.S. borrowers did not refinance at the optimal time, when interest rates were low, and relate individual sub-optimality to procrastination and financial sophistication. Gerardi et al. (2023) and Agarwal et al. (2017) discuss race and age disparities in mortgage refinancing, and argue that sophistication may be the underlying source. Our data analysis complements Andersen et al. (2020) and Keys et al. (2016), and is supportive of the view in Gerardi et al. (2023), showing that financial skills increase search effectiveness and the likelihood of refinancing, further supporting our model's assumptions.

While standard measures like loan-to-value constraints and income uncertainty disincentivize home ownership (Paz-Pardo, 2024), recent studies argue that behavioral assumptions affect mortgage take-up and subsequent performance. While Schlafmann (2020) underscores the importance of self-control in mortgage undertaking, Bailey et al. (2018) focus on leverage choice pertaining to individual house price beliefs. Moreover, Exler et al. (2021) highlights the difference in income risk perception for default and consumer scoring. In this regard, our paper introduces individual financial sophistication and search intensity as additional drivers of heterogeneity in mortgage undertaking.

Finally, our novel approach to modeling mortgage search leverages digital advancements in the era of increasing the market share of non-bank lenders. Empirical studies show that these lenders most often operate online and frequently make use of FinTech algorithms for mortgage pricing. The no-contact evaluation reduces the mortgage rate dispersion (Fuster et al., 2019; Zhou, 2022), albeit not fully. The U.S. law of fair pricing allows lenders to utilize other borrowers' observables to evaluate risks associated with the specific mortgage origination. In this regard, lenders are free to use any data that may inform about delinquency risk. Bartlett et al. (2022) show that county-based characteristics, including search effort and sophistication, add to the final mortgage price.

Adding to debt behavior literature, our model introduces endogenous financial skills accumulation and captures skill depreciation (Agarwal et al., 2007; Lusardi et al., 2017). Lusardi et al. (2017) models endogenous financial literacy accumulation and delivers plausible heterogeneity in wealth returns. In their paper, Mazzonna and Peracchi (2023) show that cognitive decline significantly affects wealthier households who misperceive their cognitive abilities, supporting our assumption of skill depreciation over time. Jappelli and Padula (2017) relate consumption growth differences to financial sophistication through investment choice subject to financial literacy effects. We contribute to this line of literature by modeling the mortgage choice subject to search frictions that depend on the individual level of financial literacy.

To that end, financial education policy that targets households who cannot keep up with skills may have heterogeneous effects across older cohorts. In the model with lenders who score their consumers, financial education significantly increases welfare (Exler et al., 2021). In our context, financial education alleviates search costs and implicitly affects household's liquidity through lower mortgage repayments.

3.3 Empirically motivated mortgage search model

3.3.1 Model setup

A continuum of risk-averse agents solves an infinite horizon problem in continuous time. Agents are heterogeneous with respect to financial skills $f_0 \sim \Gamma(f_0)$, labor productivity $z \in \{z_L, z_H\}$, and assets $a \sim \Gamma_a$. Upon income realization, agents pay their housing costs, consume c and save a. While renting, the agent continues to pay the rent cost κ. At any point, agents may take up or refinance their mortgage and adjust their housing costs to support their preferred level of consumption.

In our model, housing preferences correspond to willingness to invest in skills and put in search effort when acquiring a mortgage. In this regard, our model accounts for the cognitive complexity surrounding mortgage undertaking and introduces housing preferences through willingness to learn and search. The trade-off preceding the decision to own a home includes the possibility of facing a large expense shock once becoming a homeowner.

The change in housing status requires exerting search effort s that increases the number of mortgage offers the agent receives. In this way, the search effort corresponds to our data measure that uses the number of lenders considered as a search proxy. As the survey question focused on the consideration rather than formal application, out search costs are modeled as utility costs. The agent faces mortgage offers every period, corresponding to the current lender's web advertising practice in the U.S. On top of arrivals, the agent chooses search intensity that effectively increases the number of sample draws, rendering the mortgage arrival rate as endogenous.

The search cost depends on individual financial skills and thus changes over time. Conditional on searching, agents can take up a mortgage proportional to their income wz. We set the mortgage size to amount to 4 times the borrower's current income, capturing median-to-upper quartile mortgage amounts. The endogeneity of individual search intensity gives rise to the endogenous lock-in rate distribution G[28]. The mean and the variance of the lock-in distribution serve as calibration targets for the model solution.

Individual search intensity s, together with consumption and saving choices, comprises the set of individual policies that maximize expected future utility. Conditional on their optimal choice, borrowers "sort" into mortgage rates based on the number of offers drawn. After taking up a mortgage, homeowners face an expense shock that depends on their financial skills and assets. The expense shock represents any event that triggers losing a house, such as health, divorce, or other shock that prevents the owner from repaying their mortgage. These shocks are rare but serve as a reason to precaution among current homeowners. After the shock, the agent returns to renting and can undergo a relatively more costly mortgage take-up.

With the goal of decreasing their monthly repayments, homeowners may choose to refinance at any point in time. Refinancing carries an upfront cost c_{ref}, equivalent to 5% of the mortgage amount[29]. In addition, refinancing requires search effort, which corresponds to

[28] We derive the expression for the lock-in rate distribution in the appendix (Expression 48).

[29] According to Freddie Mac, refinancing costs range from 3-6% of the mortgage size. (Source:

meeting more lenders. Our model assumes that the homeowner's primary goal is to attain the lower mortgage rate, corresponding to our survey analysis (87% of the NSMO respondents state a lower interest rate as the primary benefit from refinancing. In addition, 68% render lower monthly payments as their priority).

3.3.2 Financial skills accumulation

The financial skill investment assumption closely follows the standard assumption in human capital accumulation literature (Browning et al., 1999; Kapička and Neira, 2019). Agents invest in financial skills that depreciate with exogenous rate δ. Each period, agents decide to invest $i \geq 0$ in financial skills f, facing a utility cost $c^f(i,z)$. The choice i represents the share of current financial skills invested into the next period skill level. Utility costs depend on the agent's productivity and increase with the share i:

$$c^f(i,z) = i_0 \frac{i^{1+\frac{1}{\gamma_i}}}{1+\frac{1}{\gamma_i}} \frac{1}{1+z},$$

where γ_i is the elasticity of investment cost with respect to investment i, and i_0 is the scaling parameter. Attaining financial skills implies lower search costs, which, through the amount of sampling from mortgage offers, generates a better position in the mortgage market. Corresponding to the life cycle pattern (the fit in Figure (18)), financial skills depreciate at rate δ. Overall, choosing i yields utility cost $c^f(i,z)$, adding to individual financial skills level according to

$$\dot{f} = \frac{\mu}{\eta}(if)^\eta - \delta f.$$

Similar to human capital, the curvature parameter η characterizes the returns to additional investment in financial skills. When choosing the optimal i, the agent includes the gains characterized by η and utility loss generated by the elasticity parameter γ_i.

3.3.3 Refinancing - decision and options

Homeowners face expense shocks and ensure liquidity through savings accumulation and mortgage refinancing. On a period basis, the agent chooses to refinance a mortgage or become a homeowner to ensure lower housing payments. Refinancing a mortgage or selecting into homeownership requires exerting search effort that effectively increases the amount of mortgages drawn from the exogenous distribution Φ. Search costs enter the utility, and are explicitly modeled as

$$c^m(s,f) = c_0 \frac{s^{1+\frac{1}{\gamma_s}}}{1+\frac{1}{\gamma_s}} \frac{1}{(1+f)^{\gamma_f}}, \quad \frac{1}{\gamma_s}, \gamma_f > 0,$$

where m stands for the mortgage. The coefficient γ_s represents the search cost elasticity with respect to search effort s, c_0 is the scaling parameter, and γ_f characterizes the effect of individual financial skills on the mortgage search process.

https://myhome.freddiemac.com/refinancing).

3.3.3.1 Expense shock

Expense shocks proxy for a homeowner's poor financial management and losing a house. The probability of facing a financial shock $p(f, a)$ decreases with the level of financial skills and assets. $p(f, a)$ serves as an additional incentive to accumulate financial skills or save. When the shock hits, homeowners lose their house and switch to renting with cost κ. Later, we externally estimate the parameters of the logistic probability model that captures the dependence on individual financial skills and assets.

3.3.4 The agent's problem

Denoting the housing state with $\theta_t \in \{\text{ho}, \text{ren}\}$, the most general formulation for the agent's problem is

$$\max_{\{c_t, s_t, i_t\}} \mathbb{E}_0 \int_0^\infty e^{-\rho t} [u(c_t, i_t, s_t) - c^f(i_t, z_t) - c^m(s_t, f_t)]_{\theta_t} dt, \text{ s.t.}$$

$$\begin{aligned}
&\dot{a}_t = Ra_t + wz_t - \mathbf{1}_{\{\theta_t = \text{ho}\}} Mr_t - \mathbf{1}_{\{\theta_t = \text{ren}\}} \kappa - c_t, \\
&\dot{f}_t = \frac{\mu}{\eta}(i_t f_t)^\eta - \delta f_t, \\
&h \to r \quad \text{with intensity} \quad p(f, a), \\
&z_t \text{ is a Poisson process with intensities } \omega_1 \text{ and } \omega_2, \\
&a_t \geq 0.
\end{aligned}$$

Recursive formulation of the problem with respective first order conditions reveal the salient trade-offs for individual consumption and search choice.

3.3.4.1 Value functions

The recursive problem form consists of Hamilton-Jacobi-Bellman (HJB) equations, housing type-flow equations and boundary constraints, separately for renters and homeowners. The flow of homeownership at different mortgage rates combines the distributions of homeowners and renters across their financial skills and assets.

Renters pay fixed rent cost κ, save in liquid accounts a_t and accumulate financial skills f_t. They engage in costly searches to get mortgage options and may decide to move to a house. Prior to the first origination, renters face additional search frictions ϕ. Dropping the time subscript, the HJB equation for renters is

$$\rho V^R(f,a,z) = \max_{\{c,s,i\}} \Big\{ u(c) - c^f(i,z) - c^m(s,f) + \frac{\partial V^R}{\partial f}(f,a,z)\dot{f} + \frac{\partial V^R}{\partial a}(f,a,z)\dot{a} \tag{7}$$
$$+ \lambda \phi s(f,a,z) \int_{\underline{r}}^{\overline{r}} \max\{V^H(f,a,z,r') - V^R(f,a,z), 0\} d\Phi(r')$$
$$+ \sum_{z'} \omega(z,z')(V^R(f,a,z') - V^R(f,a,z)) \Big\}$$

subject to

$$\dot{a} = Ra + wz - \kappa - c,$$
$$\dot{f} = \frac{\mu}{\eta}(if)^\eta - \delta f,$$
$$a \geq 0.$$

The homeowner's problem is defined with

$$\rho V^H(f,a,z,r) = \max_{\{c,s,i\}} \Big\{ u(c) - c^f(i,z) - c^m(s,f) + \frac{\partial V^H}{\partial f}(f,a,z,r)\dot{f} + \frac{\partial V^H}{\partial a}(f,a,z,r)\dot{a} \tag{8}$$
$$+ \lambda s(f,a,z,r) \int_{\underline{r}}^{\overline{r}} \max\{V^H(f,a - c_{\text{ref}}, z, r') - V^H(f,a,z,r), 0\} d\Phi(r')$$
$$+ \sum_{z'} \omega(z,z')(V^H(f,a,z',r) - V^H(f,a,z,r))$$
$$+ p(f,a)(V^R(f,a,z) - V^H(f,a,z,r)) \Big\}$$

subject to

$$\dot{a} = y(a,s) + wz - Mr - c,$$
$$\dot{f} = \frac{\mu}{\eta}(if)^\eta - \delta f,$$
$$a_t \geq 0.$$

Every row in equation 8 represents possible transitions into different productivity or homeownership states.

The state constraint $a \geq 0$ gives rise to the *boundary constraint* in the continuous time setup. That is, the FOC $u'(c(a)) = V'^{R,H}(a)$ holds everywhere (Achdou et al., 2022), so we include the boundary condition for assets $u'(c) \leq \dfrac{\partial V^H(f,0,z,r)}{\partial a}$. Optimal search effort, financial skill investment, and consumption satisfy the set of first order and boundary conditions. Moreover, the policy functions are consistent with Kolmogorov Forward Equations, i.e., they respect flows in and out of the mortgage market.

3.3.5 Kolmogorov Forward Equations

Flow changes work through exogenous separations (financial shocks and shocks to productivity) or are endogenously driven by search intensity and mortgage offer arrival rates.

Therefore, the distribution of homeowners with financial skills f, assets a, productivity $z_i, i \in \{L, H\}$, who repay mortgage at rate r satisfies the Kolmogorov Forward Equation:

$$\begin{aligned}
0 = &- \frac{\partial g^H(f,a,z_i,r)}{\partial f}\dot{f} - \frac{\partial g^H(f,a,z_i,r)}{\partial a}\dot{a} - (p(f,a) + \lambda s^H(f,a,z_i,r)\Phi(r))g^H(f,a,z_i,r) + \\
&+ \lambda \int_{\underline{r}}^{\overline{r}} s^H(f,a+c_{\text{ref}},z_i,r')g^H(f,a+c_{\text{ref}},z_i,r')d\Phi(r') + \lambda\phi s^R(f,a,z_i)g^R(f,a,z_i) + \\
&+ \omega_i(g^H(f,a,z_{-i},r) - g^H(f,a,z_i,r)).
\end{aligned} \tag{9}$$

The distribution of renters with financial skills f, assets a, productivity $z_i, i \in \{L, H\}$, satisfies the Kolmogorov Forward Equation:

$$\begin{aligned}
0 = &- \frac{\partial g^R(f,a,z_i)}{\partial f}\dot{f} - \frac{\partial g^R(f,a,z_i)}{\partial a}\dot{a} + p(f,a)\int_{\underline{r}}^{\overline{r}} g^H(f,a,z_i,r')d\Phi(r') + \\
&- \lambda\phi s^R(f,a,z_i)g^R(f,a,z_i) + \omega_i(g^R(f,a,z_{-i}) - g^R(f,a,z_i)).
\end{aligned} \tag{10}$$

3.3.6 Partial equilibrium properties

The exogenous interest rate distribution $\mathit{\Phi}$, rental rate κ, and the interest rate on liquid deposits R define the partial equilibrium of the model. Joint distribution of assets, skills, and housing costs arises endogenously, through individual search intensity and locked-in mortgage rates. In this section, we refer to partial equilibrium as an equilibrium.

Assuming heterogeneous lenders (i.e., heterogeneous mortgage offers $\mathit{\Phi}(r)$), the equilibrium consists of values $V^R(f,a,z)$, $V^H(f,a,z,r)$ defined with equations (7) and (8), respectively, and optimal policies for search intensity, financial skill investment and consumption $s^H(f,a,z,r)$, $i^H(f,a,z,r)$, $c^H(f,a,z,r)$ for homeowners and $s^R(f,a,z,r)$, $i^R(f,a,z,r)$, and $c^R(f,a,z,r)$ for renters. Policy functions define distributions of homeowners $g^H(f,a,z,r)$ and renters $g^R(f,a,z)$ across financial skill level, assets, productivity and mortgage rates. These satisfy Kolmogorov Forward Equations (9) and (10). The object of our interest is the model fit across the (f,r) subspace, as we aim to capture the patterns in the data from our empirical analysis.

In the following sections, we present equilibrium properties for both the benchmark version and the simplified version of the model. The derivations and propositions presented in this context do not consider income uncertainty and instead focus on outlining model properties related to consumption and savings effects.

3.3.7 Mortgage reservation value

We define mortgage reservation rate $\tilde{r}(\cdot)$ as the rate that leaves renters indifferent between taking up a mortgage and remaining renters: $V^R(f,a) = V^H(f,a,\tilde{r}(f,a))$. In addition to the rental rate κ, search costs that depend on skills represent an additional value of being a renter. Therefore, the reservation value is pinned down by the rent-to-mortgage rate ratio, conditional on the level of skills and assets. Because the value function strictly decreases with the interest rate, the mortgage reservation rate represents the highest interest rate at which the renter is willing to borrow, and thus to transition into homeownership.

Proposition 3.0.1. *The reservation mortgage rate ($\tilde{r}(f,a)$) is heterogeneous across assets and financial skills and is implicitly given with an equation*

$$-c^m(s(f,a,\tilde{r}(f,a))) + c^m(s(f,a,\kappa)) + u'(c(f,a,\tilde{r}(f,a)))[\kappa - \tilde{r}(f,a)M] \\ + \lambda[s(f,a,\tilde{r}(f,a)) - \phi s(f,a,\kappa)] \int \max\Big\{V^H(f,a,r') - V^H(f,a,\tilde{r}(f,a)), 0\Big\} d\Phi(r') = 0$$

Abstracting from additional frictions upon first-time mortgage origination (i.e., setting $\phi = 1$ and $c_{\text{ref}} = 0$) simplifies the reservation value equation. Particularly, there is no additional value in remaining a renter, other than paying rent costs κ. Therefore, across the asset-skill distribution, the reservation mortgage payment $\tilde{r}(f,a)$ corresponds to the rent price κ.

Corollary 3.0.1. *Abstracting from mortgage adjustment frictions ($\phi = 1, c_{ref} = 0$), the reservation interest rate $\tilde{r}(f,a)$ does not depend on assets or financial skills; it is constant across borrowers and corresponds to renting costs κ: $\tilde{r}(f,a)M = \kappa$.*

In this simplified setting, the reservation rate strategy reduces the complexity of the value function expression, and can be used to infer mortgage performance effects on consumption growth.

Corollary 3.0.2. *Excluding external search frictions, variations in consumption growth can be attributed to three factors: patience, expected future mortgage rates, and precautionary measures in response to expense shocks.*

$$\frac{\dot{c}}{c} = \frac{1}{\sigma}\left[R - \rho - \lambda s\left(\int_{\underline{r}}^{r}\left(1 - \frac{u'(c(f,a,r'))}{u'(c(f,a,r))}\right) d\Phi(r')\right) + p\left(\frac{u'(c(f,a,\kappa))}{u'(c(f,a,r))} - 1\right)\right] \tag{11}$$

Expression (39) disentangles three channels that influence consumption growth. The initial segment represents the conventional impact of impatience, while the second term reflects the effect of expected mortgage rate attainment. This effect is especially significant for high-mortgage rate payers, as they primarily depend on their search efforts without emphasizing savings. Considering states of skills and assets that dictate the level of search effort, borrowers possess knowledge of the offer rate distribution, and thus rely on future search outcomes. However, in the absence of any effort exerted by a borrower, expected future mortgage rates do not influence consumption growth.

The third segment in equation (39) corresponds to the precautionary effect triggered by the possibility of an expense shock. Precautionary motives diminish as the mortgage rate decreases. When mortgage conditions are favorable, the loss of a house carries significant negative consequences. In this regard, the model captures the increasing propensity to save along the percentiles of the asset distribution, as documented in Mian et al. (2020).

3.4 Quantification

In our approach to a quantitative solution, we utilize the finite difference method for continuous time models, following the methodology described in Achdou et al. (2022). While certain model parameters can be directly obtained from the merged dataset described in Chapter 2, we categorize them into *exogenously set* parameters and *calibrated* parameters. The calibration targets involve essential data moments that capture distinctions in homeownership and mortgage rates for first-time borrowers and upon refinancing. When we target data averages and medians, we evaluate the model's ability to match the rate-skill-assets distribution. Specifically, we establish the validity of the model using consumption and housing expenditure inequality measures.

We describe the two steps in model calibration and outline the simulated patterns relevant to our data findings.

3.4.1 Parametrization

The model is parameterized at the annual level. The first set of parameters is exogenous and combines our data estimates with literature standards.

3.4.1.1 External parameters

Utility is CRRA and the coefficient of risk aversion is set to the standard in the literature $\sigma = 2$. The time preference rate is set to $\rho = 0.05$, and the risk-free rate is $R = 0.04$ (Achdou et al., 2022). Individual productivity follows a Poisson process with transition rates estimated in Guerrieri and Lorenzoni (2017), $z \in \{0.8, 1.3\}$ with intensities $\omega_1 = \omega_2 = \frac{1}{3}$. Wage rates are normalized to 1, leaving wage equal to productivity. We follow the human capital investment model in Kapička and Neira (2019), and set the elasticity of investment in financial skills $\gamma_i = 0.5$. Lastly, we set the monetary refinancing cost c_{ref} to equal 5% of the mortgage size.

3.4.1.2 Financial skills parameters

We assume that financial skill accumulation satisfies the flow equation:

$$\dot{f} = \frac{\mu}{\eta}(if)^{\eta} - \delta f.$$

We follow seminal papers by Lusardi et al. (2017), Lusardi et al. (2020), and Browning et al. (1999), and fix $\eta = \frac{1}{2}$, and $\delta = 0.7$. Next, we estimate the slope parameter μ using the SCF data on financial skills. Parameters η and δ correspond to human capital elasticity and depreciation estimates, respectively.

3.4.1.3 Expense shock probability

In our model, individual expense shocks translate into delinquency and default. We assume that shock probability depends on homeowners' assets and financial skills, $p(f,a)$, and approximate the functional form as

$$p(f,a) = \frac{\exp(p_0 + p_f f + p_a a)}{1 + \exp(p_0 + p_f f + p_a a)}. \quad (12)$$

We estimate the coefficients using the SCF data on late payments among homeowners with a mortgage on their primary residences. Corresponding to the assets in the model, the assets in the SCF include only liquid assets[30]. We re-scale these assets to match the grid bounds in the numerical implementation. The dependent variable is an indicator of over 60 days debt delinquencies. Our estimates control for mortgage size and house value, and thus pertain to the model assumptions.

According to our estimates, financial skills and assets correlate negatively with the likelihood of late debt payments, and the coefficients estimates are $p_0 = -1.08$, $p_f = -1.016$, and $p_a = -7.649$.

3.4.1.4 Mortgage specifications

The mortgage amount follows the standard and corresponds to 4 times the average borrower's income. Mortgage rate offers are exogeneous and beta-distributed. In the equilibrium, the accepted mortgage rate distribution is endogenous and stochastically dominated by the exogeneous offer rate distribution (analytical proof (48) appears in the Appendix).

3.4.1.5 Internally calibrated parameters

The rest of the parameters are internally calibrated using the simulated method of moments with moment targets that are salient for model performance. Target moments are weighted equally and comprised of the share of homeowners, normalized average financial skills, standard deviation of financial skills, and NSMO-based sample mean and standard deviation of mortgage rates attained, separately for first origination and upon refinancing. Although all parameters are calibrated jointly, we discuss below which moment aims to pin down which specific parameter.

We assume that the offer rate is beta-distributed and calibrate the two shape parameters $\beta = 6.0411$ and $\alpha = 6.0805$ to match the moments of the (endogenous) locked-in mortgage rate. The rental cost $\kappa = 0.7340$ is informed by the homeownership rate in the SCF sample and yields higher monthly payments on housing for renters, which is consistent with the data[31]. The elasticity of search effort $\gamma_s = 1.7539$ and scaling parameter of the

[30] Specifically, we include cash and prepaid cards, checking and savings accounts, directly held money market funds and stocks, and the value of mutual funds investment.

[31] Using the SCF data, we compare monthly rent and mortgage payments as income shares across financial skills. Rent shares are twice as high on average. (Table 34 in the Appendix.)

investment cost function $i_0 = 434.2084$ are pinned down by the sample moments of financial skills in the SCF data. The difference between average rates under refinancing and first origination pins down the scaling and search friction parameter $c_0 = 152.9484$, and $\phi = 0.8062$, respectively. In the equilibrium, renters search less than homeowners, aligning with SCF credit search estimates.

We report the targeted moments and the parameter values that minimize the distance between the moments in the data and in the model in Panel C of Table 18.

Table 18: Model parameter values. Source: Model, SCF, and NSMO.

	Definition	Symbol	Estimate	Source/Target		
			Panel A. Externally set			
(dis-)utility	Discount factor	ρ	0.05	Standard		
	CRRA parameter	σ	2	Standard		
	Investment cost elasticity	γ_i	0.5	Kapička and Neira (2019)		
assets	Return	R	0.04	Standard		
	Refinancing Cost	c_{ref}	0.21	Freddie Mac (5% of the mortgage size)		
productivity	Intensities	ω_1, ω_2	$\frac{1}{3}, \frac{1}{3}$	Guerrieri and Lorenzoni (2017)		
skill accumulation	Curvature f	η	0.5	Browning et al. (1999)		
	Depreciation	δ	0.07	Lusardi et al. (2017)		
			Panel B. Externally estimated			
skill accumulation	Slope	μ	0.2	SCF, lifecycle profile		
housing shock	Parameters	p_0, p_f, p_a	-1.08,-1.02,-7.65	SCF, late payments		
			Panel C. Internally estimated		Model	Data
dis-utility	Search cost - skill parameter	γ_f	0.2977	Average financial skills - HO	0.7690	0.7654
	Investment cost scaling	i_0	434.2084	Average financial skills - R	0.6270	0.6499
	Renting cost	κ	0.7340	Homeownership rate	0.6432	0.64
	Search cost elasticity	γ_s	1.7539	Standard deviation fin. skills	0.1868	0.3041
	Search cost scaling	c_0	152.9484	Average mrt. rate all	0.0398	0.0400
	Search friction	ϕ	0.8062	Average mrt. rate f.o.	0.0415	0.0408
	Offer distribution parameter	β	6.0411	Average mrt. rate - ref.	0.0362	0.0386
	Offer distribution parameter	α	6.0805	Standard deviation mrt. rate	0.0087	0.0073

3.4.2 Model fit

We validate the model fit using graphic and qualitative measures of consumption inequality in the data. Specifically, we use the Gini coefficient and Lorenz Curve as two relevant measures for comparing model-implied consumption and housing expenditures to data counterparts. For this purpose, we use the 2019 Bureau of Labor Statistic report (Garner et al., 2022) on consumption disparity across different types of goods.

We compare our model-implied consumption net of housing costs to the non-durable goods consumption reported in Garner et al. (2022). The Gini coefficient from our model simulations $G_{\text{model}} = 0.2$ matches the data Gini coefficient $G_{\text{BLS}} = 0.18$. We also compare Lorenz Curves from model simulations to the data. The left panel in Figure 26 shows that our model performs well, not only in fitting the area below the perfect equality curve, but in fitting the curve itself.

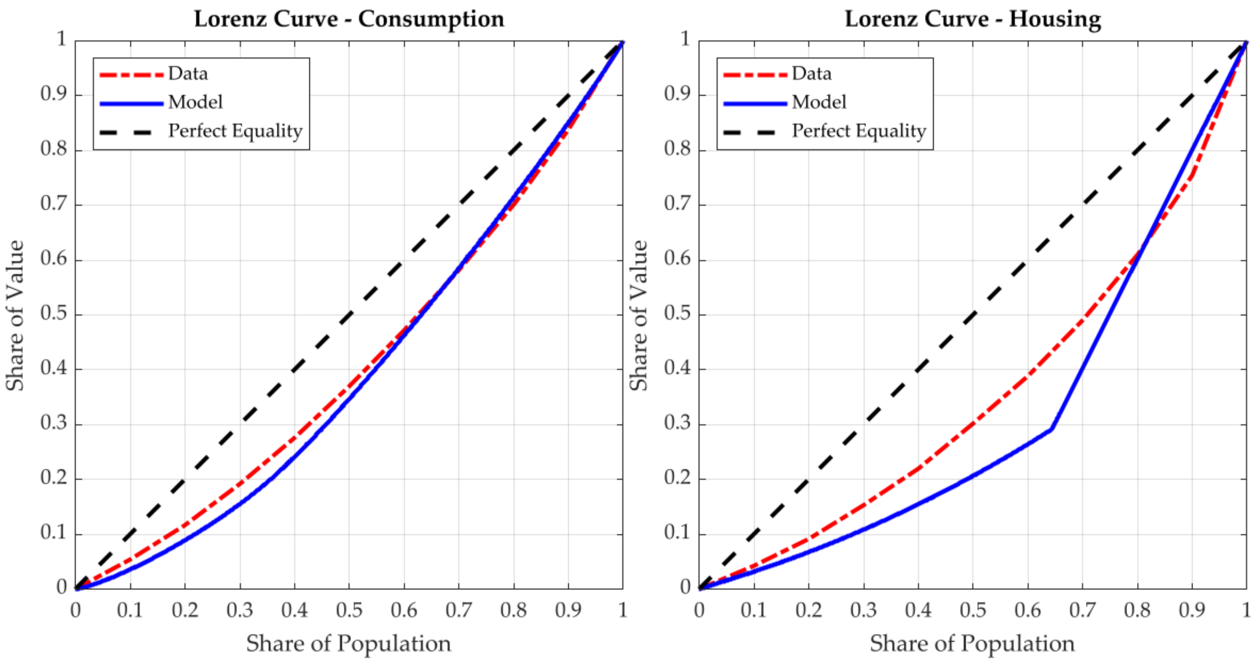


Figure 26: Lorenz Curves for consumption (left) and housing consumption (right) compared to the BLS data.

In addition, we compare total housing costs (including household expenditures on housing and utilities) from Garner et al. (2022) to the total individual housing cost in our model simulations. Depending on the homeownership state, housing costs in our model correspond to either rental rates or mortgage repayments.

In our simulation, the Gini coefficient of housing costs equals 0.37, slightly overstating the data counterpart value 0.29. The right panel in Figure 26 shows that our model understates housing costs for homeowners (the Lorenz Curve in this case does not overlap completely with data implied Lorenz Curve, though, they are close), potentially pertaining to fixed rental rates. Given that we test our model with policies that potentially reduce heterogeneity in individual liquidity (either through consumption or savings) we rely on the

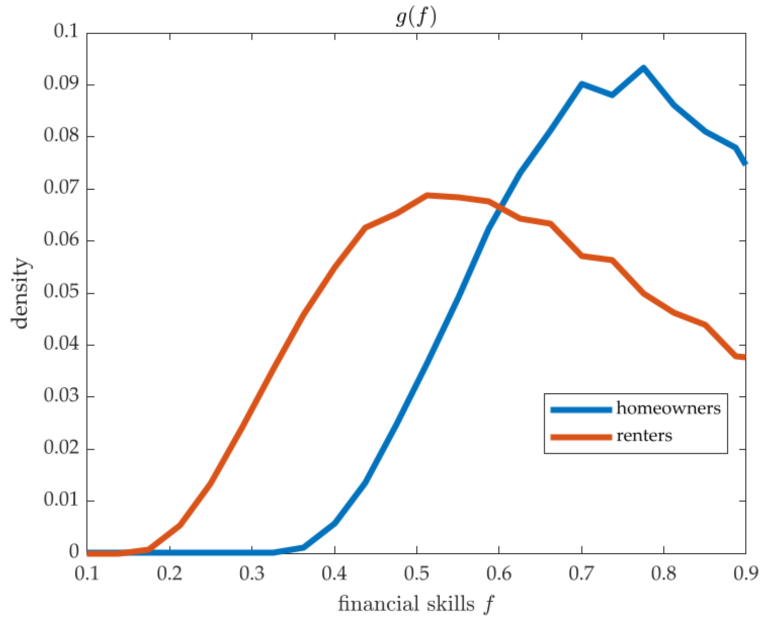


Figure 27: Differences in skill distribution between renters and homeowners.

good performance of our model in matching non-durable consumption inequality.

3.5 Skill-based consumption differences

Because our paper introduces a novel dimension in mortgage payment heterogeneity, our primary focus is to examine individual policy variations between low- and high-skilled borrowers. Incentives for mortgage take-up differ significantly across these two groups, contingent on their respective skill levels and asset holdings. Through the lens of our model, these differences translate to consumption disparities.

Our analysis of consumption inequality begins by aligning equilibrium patterns within our model with key data insights from the SCF and our new dataset. Firstly, our model highlights a correlation between skills and choices, and accurately predicts adjustments in housing costs at mortgage initiation and refinancing stages. Specifically, the model demonstrates a positive link between mortgage initiation and financial skills, leading to renters having lower average skill levels, as shown in Figure 27.

Skilled homeowners in the model are more likely to explore refinancing options and to attain lower mortgage rates on average, as depicted in Figure 31. We also identify a slight negative correlation between individual search behavior and asset holdings, as seen in Figures 32 and 29. Borrowers further from the liquidity constraint are more inclined to forego the advantages of lower mortgage payments.

The second set of model patterns directly relates to housing cost heterogeneity, leading to consumption differences across assets and skills. Skilled borrowers are more inclined to take up a mortgage and refinance soon, resulting in relatively lower shares of their savings being allocated to durable consumption. As a result, the model successfully delivers non-durable consumption inequality that aligns with the observed data, as already outlined in the Lorenz curve comparison in Figure 26.

The third set of model performance evidence pertains to our assumption on skill investment choice. In the equilibrium, the investment choice exhibits a hump-shaped pattern with respect to the skill level. As shown in Figure 28, an average homeowner invests in skills until a certain skill level. This behavior corresponds to another key pattern we observe in the SCF data (as depicted in Figure 18) - the hump-shaped life-cycle path of individual skills, evident in the SCF data analysis. When interpreted through the lens of our model, skill investment is not as prominent as the homeowner attains a lower interest rate.

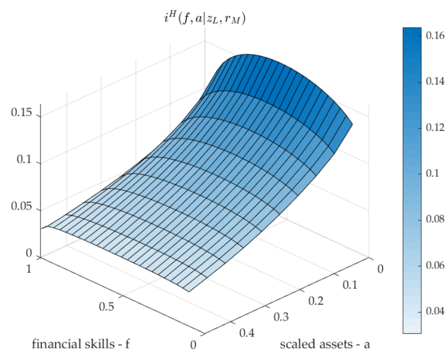


Figure 28: Investment in skill as a function of individual assets and skill level, for the low productive homeowner with an average mortgage rate.

3.5.1 Mortgage take-up across the skill distribution

Figure 29 presents the variation of search intensity among renters. The heatmap plot illustrates that as financial skills improve, there is an increase in search effort. This trend aligns with the SCF data findings regarding factors that influence homeownership (the estimates from the probabilistic regression model are provided in Table 35). Within the model, individuals within the second tercile of the skill distribution are 17% more likely to opt for a mortgage than are low-skilled renters. Moreover, those in the top tercile of the skill distribution have a 70% greater likelihood of transitioning into homeownership. In contrast, individuals with lower financial skills tend to continue as renters, which affects their available resources and results in reduced consumption.

Conversely, the search intensity plot shows that wealthier renters tend to search less and are more willing to forgo the benefits of lower mortgage payments[32].

[32] Comparing monthly housing ratios among similar households in the SCF data reveals higher rental rates in comparison to mortgage payments. Admittedly, our model may overstate the rent-to-mortgage payment ratio

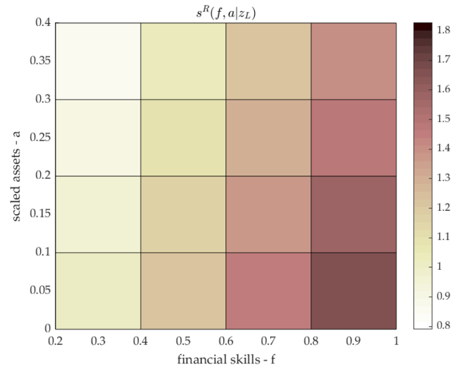


Figure 29: Likelihood to mortgage take-up across financial skills and current mortgage rates for low productive agents and average assets. Likelihood is depicted relative to the least financially skilled.

3.5.2 Mortgage rate differences among homeowners

In the model, housing cost heterogeneity among homeowners boils down to mortgage payment differences. Figure 30 depicts differences in locked-in mortgage rates between low- and high- skilled borrowers. Low-skilled borrowers search less and borrow at rates as-good-as random, pertaining to the exogenous random draw (represented with a purple histogram in Figure 30). On the other hand, high-skilled borrowers sample more from the offer rate distribution, ultimately landing at better rates (the green histogram in Figure 30). Our model successfully generates a mortgage rate dispersion that decreases as financial skills increase, which aligns with the findings from Chapter 2[33].

In our model, refinancing matches one-to-one with search intensity. That is, we evaluate the expression

$$\mathbb{P}_{\text{ref}}(s) = 1 - \exp(-\lambda s),$$

corresponding to individual refinancing probability. Among homeowners, refinancing activity depends on individual assets and current mortgage rates. Figure 31 shows that search intensity (i.e., refinancing activity) increases with financial skills and mortgage rates, consistent with the evidence from the SCF (regression Table 33 of the Appendix). Our model's predictions indicate that high-skilled borrowers are 30% more likely to search for and refinance their mortgages. Additionally, housing costs contribute to a 10% increase in the overall likelihood of refinancing, which aligns with the prediction differences observed in the SCF data (Figure 19).

Focusing on variation in refinancing activity across the asset-skill distribution, we observe a decrease in search for refinancing options as asset holdings increase 32. Wealthier homeowners are less constrained by housing costs, and are less willing to forego refinancing opportunities.

[33]The data shows consistently positive spread differences between the top and bottom skill terciles throughout the sample duration.

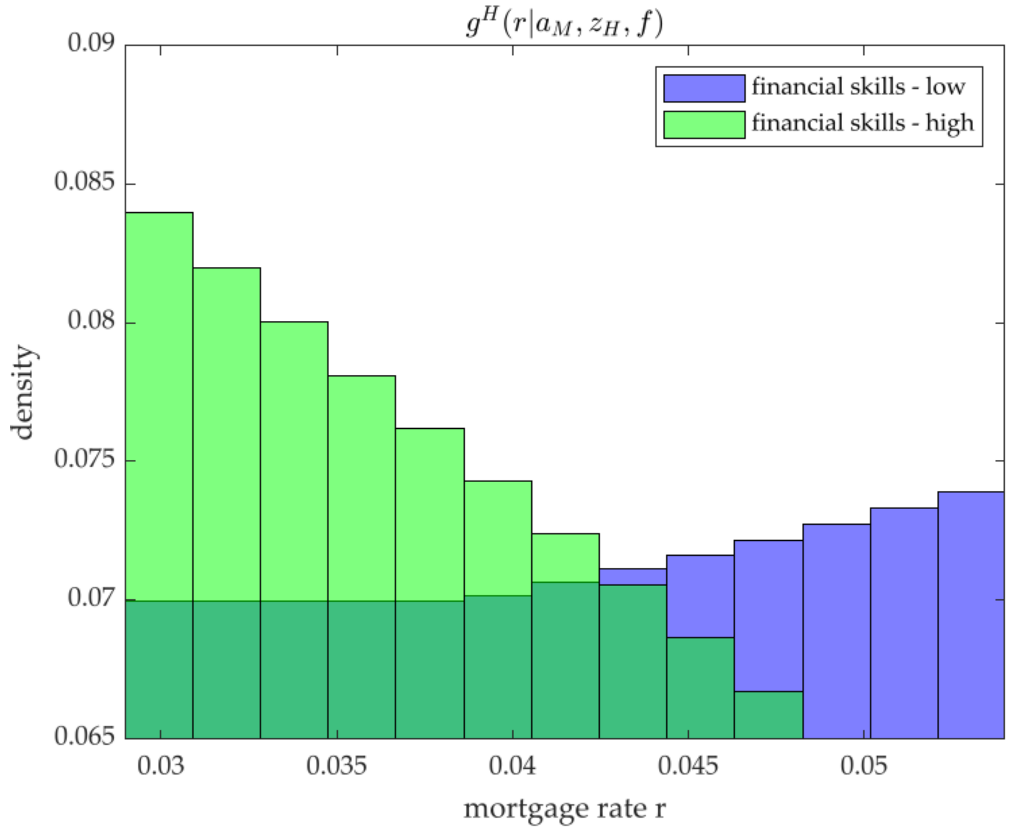


Figure 30: Distribution of low- and high-skilled homeowners across mortgage rates.

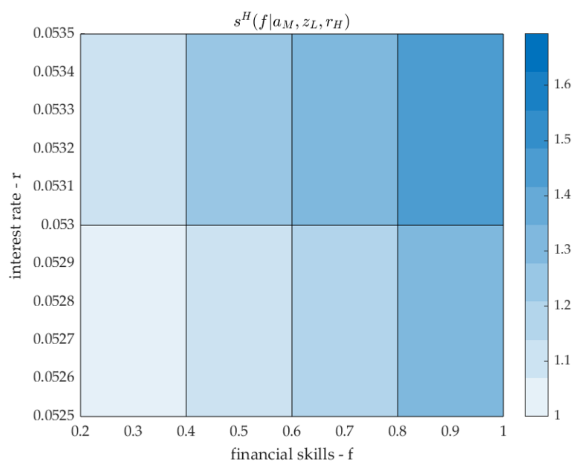


Figure 31: Likelihood to refinance across financial skills and current mortgage rates for low productive agents and average assets. Likelihood depicted relative to the least financially skilled.

3.5.3 Mortgage rate dispersion decomposed

To gain a more comprehensive understanding of mortgage rate dispersion in the model, we perform a variance decomposition across all dimensions of agent heterogeneity. First, we regress the mortgage rate on individual skills, assets, search intensity, and productivity to

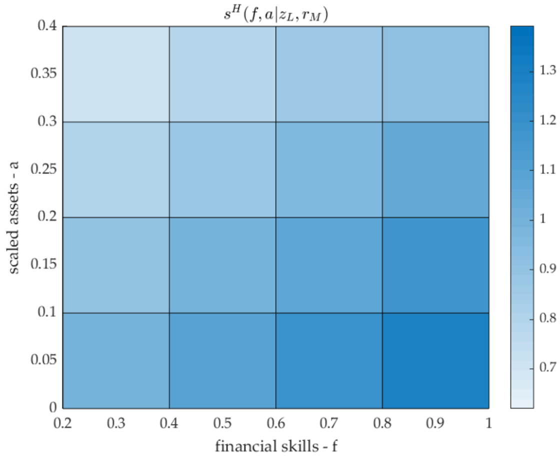


Figure 32: Likelihood to refinance across financial skills and assets for low productive borrowers with an average mortgage rate. Likelihood is depicted relative to the least financially skilled.

evaluate the model's performance recreating the correlation signs we observe in our novel dataset. Next, we decompose the variance in the mortgage rate to identify the contributions of each dimension of heterogeneity.

Rather than regressing the resulting interest rate, we regress a linear transformation $\log(1 + r)$ on sources of individual heterogeneity, namely assets a, productivity z, search intensity s and financial skills f. To make these estimates parallel to the merged data estimates, we account for the interaction term between search intensity and skill level. We estimate the regression equation

$$\log(1 + r) = \beta_0 + \beta_a a + \beta_z z + \beta_f f + \beta_s s + \beta_{f\times s}(f \times s) + \varepsilon$$

using the calibrated model, and we use weights derived from the steady-state distribution of the model. Keeping in mind that our model considers fairly similar borrowers in search of a basic mortgage product, the mortgage rate variation aligns with that in our data analysis, represented by yearly average rates for fixed loan amounts, as shown in Figure 22.

The linear regression estimates presented in Table 19 show that our model outcome aligns with the interest rate regression we obtain from the merged dataset, with coefficients that are consistent with those outlined in Table 16. Specifically, individual productivity (income) has a positive correlation with the mortgage rate, while skill and search show correlations with opposite signs: skills correlate negatively (regr. coefficient -0.0033), and search effort correlates positively (with a regr. coefficient 0.0884).

Given that financially skilled borrowers face lower search costs, they tend to refinance more frequently and to lock in at lower mortgage rates. This corresponds to the negative correlation we observe, and to a regression coefficient of -0.0033. Moreover, borrowers with greater wealth might be less motivated to refinance often, as indicated by the modest yet positive asset regression coefficient (0.0021). The main regression Table 16 supports this, showing a positive coefficient for total borrower income. Accordingly, in the model, individuals with higher productivity and wealth are less susceptible to mortgage repayment effects.

Table 19: The regression results for calibrated model weighted by the steady state distribution.

	Dependent variable:
	mortgage interest rate $\log(1+r)$
Financial skills (f)	-0.0033^{***}
	(0.00024)
Assets (a)	0.0021^{***}
	(0.00030)
Productivity:	
(z_H)	0.0002^{***}
	(0.00009)
Search intensity (s)	0.0884^{***}
	(0.00097)
Financial skills × search intensity ($f \times s$)	-0.0600^{***}
	(0.00156)
Constant	0.0434^{***}
	(0.00018)
Observations	15,000
R^2	0.554
Adjusted R^2	0.554
Residual Std. Error	0.0052 (df = 15,000)
F Statistic	3732.06^{***} (df = 6; 15,000)
Note:	*p<0.1; **p<0.05; ***p<0.01

Base category productivity is z_L.
Observations weighted by the equilibrium stationary distribution.
Continuous variables are normalized for better interpretability.

Most importantly, the regression estimates show that our model performs well in capturing the *effective search* phenomenon discussed in Chapter 2. Notably, there is a positive correlation between search intensity and individual lock-in rate (with a coefficient estimate 0.0884), indicative of the efforts made by individuals with lower skill levels to secure mortgages, even at the cost of accepting higher rates. In contrast, skilled borrowers search more effectively, meet more lenders, and tend to lock in lower mortgage rates. This distinction is quantified by the coefficient estimate of -0.0600. Consequently, our model effectively captures and reproduces the significant patterns governing individual mortgage rate attainment.

As mortgage rate dispersion in the model corresponds to the data-driven dispersion, which accounts for mortgage controls (represented with density plot in Figure 23, for example), model-based variance decomposition depicts the strength of each of the heterogeneity dimensions in explaining mortgage rate attainment.

Table 20 shows that most of the difference in rate attainment lies in search effort heterogeneity, at 55% and 10% of the variance, respectively. Search intensity accounts for the relatively higher rates among the financially unskilled who aim to secure any type of

Table 20: Mortgage rate variance decomposition across all sources of heterogeneity in the mortgage search model.

	explained variance share
Financial skills (f)	1.2877%
Assets (a)	0.3291%
Productivity: (z_H)	0.0480%
Search intensity (s)	55.2445%
Financial skills × search intensity ($f \times s$)	9.8865%

mortgage (positive slope β_s in regression Table 19) and lower rates among financially savvy borrowers who search for refinancing (negative interaction coefficient $\beta_{f\times s}$). Skills alone explain 1.3% of the variation, corresponding to small and significant regression estimates in the data. Overall, the model-implied link between search effort and levels of financial skill further reinforces the effective search mechanism, which plays a key role in explaining the residual dispersion of mortgage rates in our data analysis.

3.5.4 Delinquency rate

We assume that expense shock probability (i.e., the delinquency rate) decreases with individual skills and assets (depicted by equation (12) in the model setup). The average borrower in the economy faces an expense shock probability of 0.02, matching the low number of delinquencies we retrieve from the SCF data. Our model is successful in matching the elasticity of an expense shock to individual financial skills level. Specifically, the model solution suggests that, averaged across assets, getting one more question wrong in the financial literacy tests corresponds to being 39.5% more likely to face expense shocks. Overall, the model prediction of expense shock probability aligns well with our NSMO+ data estimates 25.

3.5.5 Consumption differences across skills

Our analytical findings reveal that homeowners who repay their mortgages at the best rates exhibit the strongest precautionary motive in response to the positive probability of an expense shock, thereby influencing their consumption growth (39). In line with this, individuals who search the most and secure the best mortgage rates (as is evident from the likelihood comparison in Figure 31) are highly skilled borrowers with strong precautionary motives due to possible expense shocks. Figure 34 showcases consumption differences across financial skills levels, within each asset quartile. Consistent with our analytical results, these consumption disparities decrease as assets increase, signaling the impact of precaution among wealthier borrowers.

The most significant skill-based difference in consumption appears at the bottom of the asset distribution, primarily due to the heterogeneity in housing costs (illustrated by the

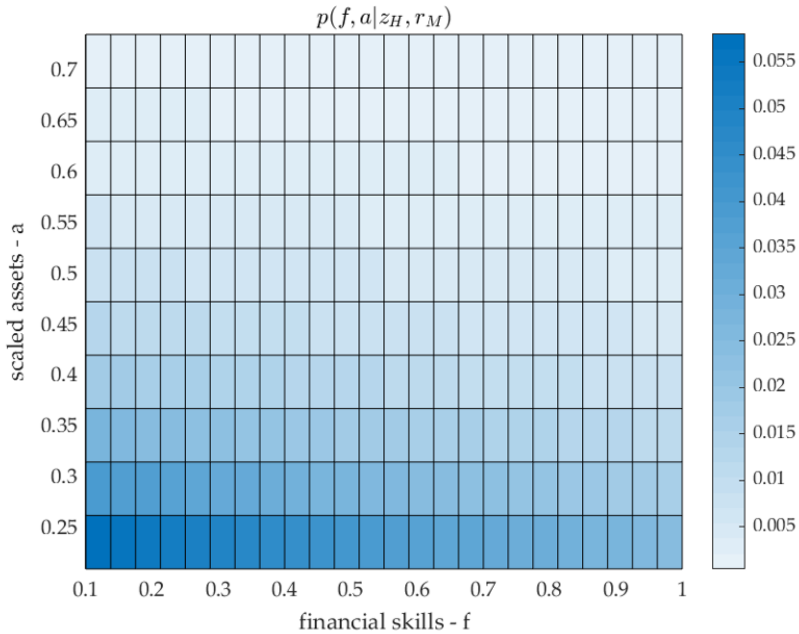


Figure 33: Probability of shock for high productive agents with med-sized mortgage payments.

leftmost bar plot in Figure 34). Specifically, financially savvy individuals at the lower end of the asset distribution are more likely to take up mortgages and face lower housing costs, leading to notable consumption disparities.

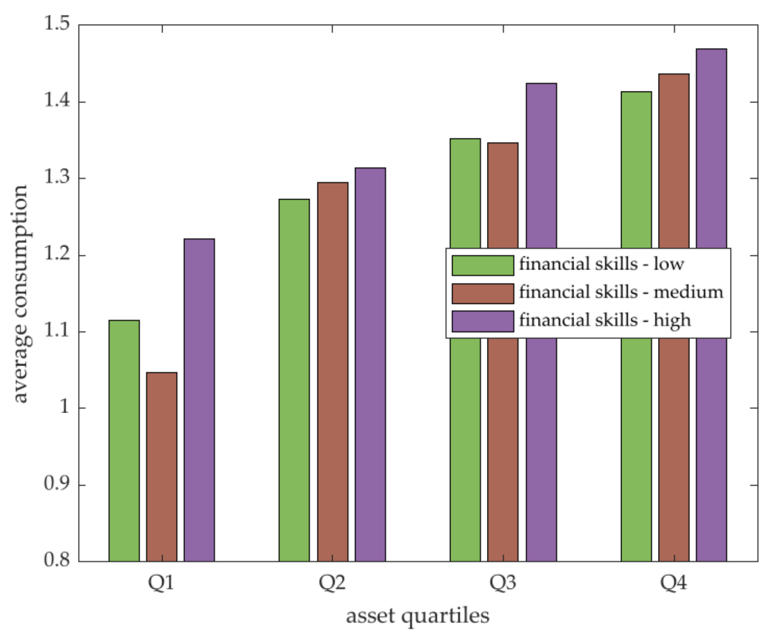


Figure 34: Skill-based consumption differences, within each asset quartile.

Ultimately, variation in liquidity among otherwise equal borrowers depends on their search and skill investment choices, speaking to a line of evidence in liquidity differences between financially skilled and unskilled borrowers (Bhutta et al., 2020, 2022; Agarwal et al., 2007). Because these agents are otherwise similar, their effort in adjusting housing costs directly translates to inequality in non-durable consumption and saving opportunities. To mitigate these differences, we render our model as a fitting laboratory for introducing financial education policy. We incorporate relevant changes in U.S. mortgage attainment over the past ten years, and observe differences in search incentives for different values of mortgage servicing costs.

3.6 Policy implications

In this section, we first tackle the adverse effects of skills on liquidity among similar borrowers, and introduce financial education as a potential remedy in our model. Second, we conduct a counterfactual exercise with more accessible mortgages, mirroring increasing entry of non-bank lenders into the market. When mortgages are easily attainable, search is cheaper, and the incentives to accumulate skills are relatively lower. Therefore, the skill-based gap between mortgage rates becomes even more prominent.

While both mortgage attainability and financial education policy enhance arrival rates of mortgages, financial education does so through stimulating individual search intensity. Moreover, education-driven increases in average skills implies a decrease in delinquency rates. On the other hand, the extension of our second experiment suggests that increased mortgage accessibility is more accommodating for financial education policy. With accessible mortgages, lower search and skill investment costs jointly reinforce skills accumulation incentives. Given that loan-level data studies indicate persistence of mortgage price differentiation based on search behavior and financial sophistication (Fuster et al., 2019; Bartlett et al., 2022), our policy tests call for an extension beyond a partial equilibrium setting.

In our last exercise, we investigate the difference in mortgage take-up and refinancing for two different interest rate levels. Our findings reinforce model validity, even when we move away from data used for calibration purposes. We find that the low mortgage rate environment benefits homeowners because they engage in refinancing more often. There is also a small increase in total homeownership, corresponding to the steeper increase in refinancing during the low interest rate period at the onset of the COVID pandemic in the U.S.[34]. In the high-rate environment, homeowners tend to remain at their initial lock-in rates, while renters do not exhibit a significant change in mortgage take-up. To that extent, the high-rate environment depicts lower consumption disparity due to decreases in the gap between rental and housing costs.

3.6.1 Introducing financial education

We introduce a financial education policy by changing the quality of financial education for renters, through a decrease in investment cost elasticity. This way, we introduce "a

[34] https://libertystreeteconomics.newyorkfed.org/2023/05/the-great-pandemic-mortgage-refinance-boom/

nudge" for renters to invest in skills a bit more if they find it optimal. Not all renters react to these incentives.

Reducing the cost elasticity by 5% effectively reduces the average work time by 0.01%. We interpret financial education as a course for renters that takes 90 minutes out of their working hours[35]. This policy targets young households before their home purchase, and we interpret it as more than mortgage undertaking courses. In the model, financial skills affect individual financial shock probability, so higher investments contribute to mortgage performance that is affected by financial distress not related to the mortgage. Therefore, our policy exercise introduces an increase in the incentive to invest in financial knowledge.

The red bars in the right graph of figure 35 outline the differences in financial skill investment across different asset quartiles. Wealthy renters react to these incentives more and invest up to 18% more relative to no education policy. More importantly, poor renters do not react to these incentives, and they still find mortgage take-up too costly.

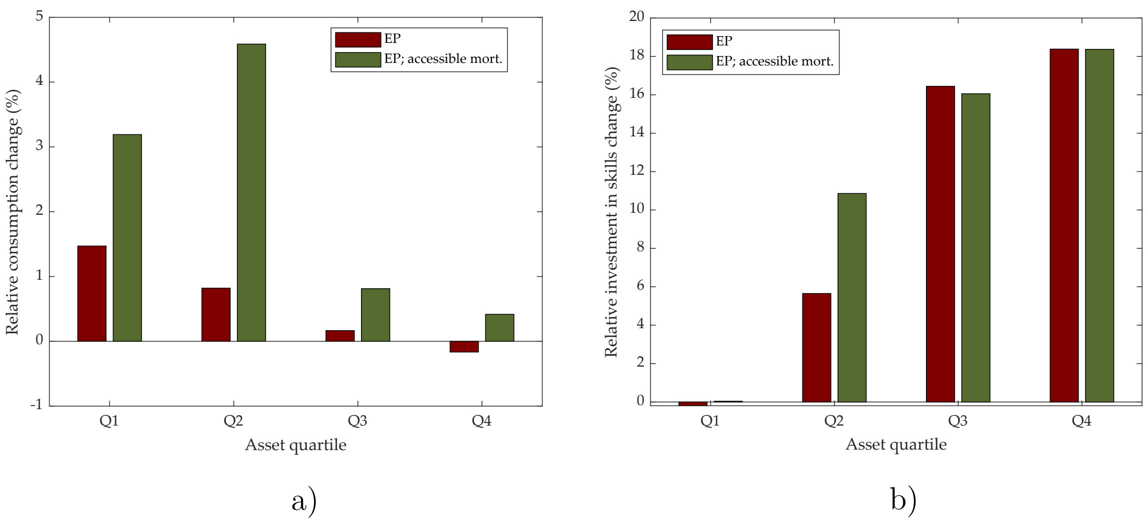


Figure 35: Relative change in a) consumption and b) financial skill investment with financial education. Source: model simulations.

On aggregate, the homeownership rate increases by 1.5%, owing to an increase in average search intensity among renters (by 0.4%, Table 21). New homeowners lock in at higher mortgage rates that still imply lower housing costs than the rental rate, thereby consuming and saving more, propagating a decrease in consumption and asset inequality. Moreover, higher financial skills reduce the average delinquency rate by 2.8%, substantiating the increase in welfare.

3.6.2 Increase in mortgage accessibility

Our second exercise mirrors the entrance of non-bank mortgage lenders in the U.S. mortgage market. In this counterfactual, we reduce the search cost elasticity parameter that

[35]We calculate this cost based on a standard working week of 40 hours without any time off.

Table 21: Introducing financial education with renters, source: model simulations.

Measure	relative change
average search renters	$\nearrow$ 0.4%
consumption Gini	$\searrow$ 1.4%
assets Gini	$\searrow$ 1.3%
share of homeowners	$\nearrow$ 1.6%
average financial skills	$\nearrow$ 9%
average delinquency rate	$\searrow$ 2.8

directly affects the search effort needed to obtain a larger sample from the pool of mortgage offers.

We reduce cost elasticity that represents a decrease in average search costs of 5% for renters and 10% for homeowners. In other words, renters spend 10 hours less per year on their mortgage search. Homeowners who opt to refinance spend hours less per year.

Table 22 outlines the relative difference between the benchmark and higher accessibility counterfactual. The first line in Table 22 shows that the average search of homeowners increases by 16.9%, whereas renters search 7.8% more intensively. Taking up a mortgage lowers housing costs, and together with intensified refinancing, decreases consumption inequality by 3% (Table 22). As less savvy borrowers become homeowners, they are exposed to financial shock, implying an increase in average delinquencies by 1.7%. Lower search costs do not incentivize skills accumulation, and skills increase only by 1.1% (Table 22).

Table 22: Market competition increase, source: model simulations.

Measure	relative change
average search renters	$\nearrow$ 7.8%
average search homeowners	$\nearrow$ 16.9%
consumption Gini	$\searrow$ 3%
assets Gini	$\searrow$ 2.4%
share of homeowners	$\nearrow$ 3.3%
average financial skills	$\nearrow$ 1.1%
average delinquency rate	$\nearrow$ 1.7%.

Search cost reduction does not instigate financial skill accumulation. We show that introducing financial education helps to mitigate the increase in delinquency rates that results from highly attainable mortgages. Green bars in figure 35 represent relative change in consumption (left) and financial skill investment (right) after we introduce financial education in the accessible mortgage setting. Figure 35 shows that, across all asset quartiles, incentives to invest in skills are stronger in the accessible mortgage setting, rendering financial education more effective. We observe a significant increase in consumption and skill investment among the bottom 50% of the asset distribution. Financial education is particularly targeted at those individuals whose mortgage repayments constitute a substantial portion of their monthly budgets.

Table 23 compares the relative change between the benchmark scenario with and without financial education. Financial education encourages skill accumulation, increasing average skills by 0.3 percentage points more than in the benchmark scenario, resulting in an average skill level of 9.3% (Table 23, right column). Additionally, delinquency rates decrease by 2.7%, reducing the impact of low-skilled homeowners entering the market (Table 23, right column, last row). Consequently, financial education reduces current and potential future consumption disparity among similar borrowers. Our analysis indicates that the current increase in mortgage availability in the U.S. provides a strong foundation for introducing financial education policies.

Table 23: Financial education policy and increase in competition. Source: model simulations.

Measure	Competition increase	Fin. education, competitive benchmark
average search renters	↗ 7.8%	↗ 0.4%
average search homeowners	↗ 16.9%	↗ 2.6%
consumption Gini	↘ 3%	↘ 1.5%
assets Gini	↘ 2.4%	↘ 1.3%
share of homeowners	↗ 3.3.%	↗ 1.5%
average financial skills	↗ 1.1%	↗ 9.3%
average delinquency rate	↗ 1.7%	↘ 2.7%

3.6.3 Exogenous change in the mortgage repayment level - implications for inequality

In our third experiment, we contrast the baseline steady-state with two distinct mortgage rate scenarios in terms of their effects on consumption inequality patterns. Without an explicit representation of the supply side (a topic we investigate in a separate paper), we introduce external shifts in the average mortgage repayment, representing mortgage policies that affect all borrowers equally. We leverage the adaptability inherent in the Beta distribution, assuming both a downward and an upward shift in the mean offer rate.

To derive parameters for the new Beta distribution, we maintain the calibrated model's offer distribution spread and compute parameters that align with the new mean. In the initial case, we compare the baseline with a lower average rate scenario, signifying a leftward shift (as shown in Figure 36). Quantitatively, this corresponds to a decrease of 20 basis points in the mean of the offer distribution. The parameters for the new offer distribution, characterized by the lowered mean when we retain the same spread as the baseline offer distribution, are $\alpha^{low-i.r.} = 5.1016$ and $\beta^{low-i.r.} = 6.7629$.

Table 24 showcases the relative differences in key model metrics, encompassing search intensity by homeowners and renters, Gini coefficients, and financial skills. In the low-rate scenario, a significant variance arises in homeowners' search intensity, reaching 60% higher than the benchmark value. Homeowners opt to forgo search costs and capitalize on the advantages of securing lower rates that lead to reduced housing expenses. In contrast, renters do not engage in skill accumulation due to the persistence of search costs. Ultimately, the

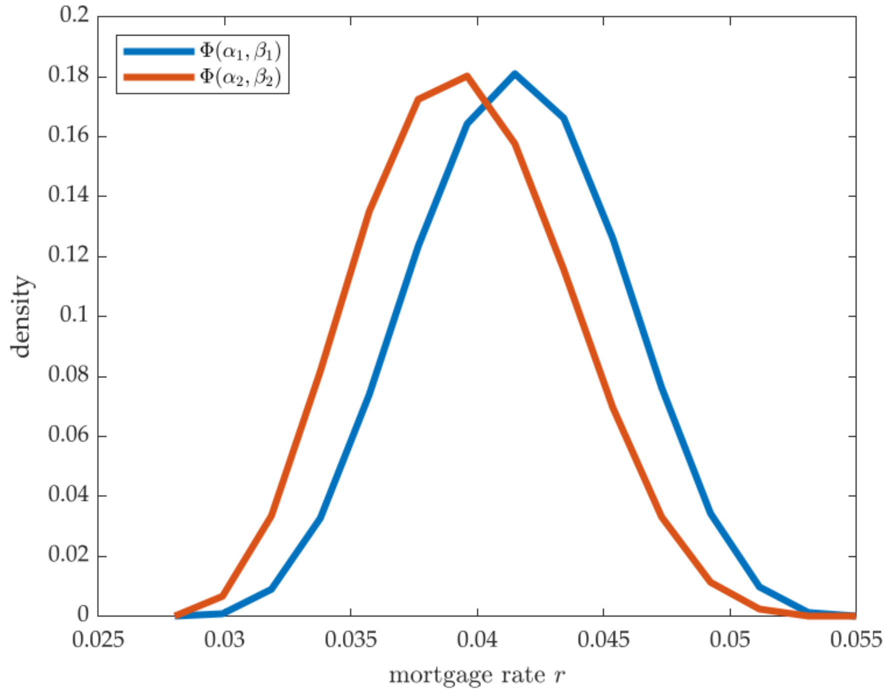


Figure 36: Original and left-shifted distribution of the offer rate.

Table 24: Comparison of the baseline with a downward shift in the offer rate. Source: model simulations.

Measure	relative change
average search renters	$\nearrow$ 1.4%
average search homeowners	$\nearrow$ 64.9%
consumption Gini	$\nearrow$ 1.4%
assets Gini	$\nearrow$ 1.1%
average financial skills	$\nearrow$ 0.1%

downward shift in mortgage payments is linked primarily to existing homeowners, exacerbating the disparity in consumption levels between renters and homeowners. This outcome is reflected in the relative difference of 1.4% and 1.1% in the consumption and asset Gini coefficients, respectively. Across all measures of comparison, lower mortgage payments perpetuate skill-based inequality.[36].

Next, we compare our benchmark with an exogenous upward shift, and implement an average offer rate 10 b.p higher than the baseline. Table 25 presents relative differences in our model metrics. As housing costs flatten out across the skill distribution, consumption inequality falls relatively lower in relation to the benchmark (−5.6% depicted in the Table 25). Search disincentives kick in mostly among homeowners, depicted by the relative fall in search intensity of more than 36%. Similar to the downward shift, renters' search effort does not change significantly (−0.7% depicted in Table 25).

[36]Admittedly, without modeling the general equilibrium effects, we keep the rental rate fixed.

Table 25: Comparison of the baseline equilibrium with the equilibrium with a higher mean of the offer distribution. Source: model simulations.

Measure	relative change
average search renters	$\searrow$ 0.7%
average search homeowners	$\searrow$ 36.5%
consumption Gini	$\searrow$ 5.6%
assets Gini	$\searrow$ 4.3%
average financial skills	$\searrow$ 0.6%

Figure 37 presents the difference in non-durable consumption inequality. We compare Lorenz Curves for non-durable consumption across baseline, upward, and downward shift scenarios. Higher mortgage payments bring skill-based housing cost differences closer together, flattening non-durable consumption across households (depicted with a red curve in Figure 37). On the other hand, a lower repayment scenario yields an outward shift from perfect equality, (depicted with a green curve in Figure 37), which reflects the advantages of current homeowners. In this scenario, the low incentives for financial skills accumulation speaks to low-skilled mortgage take-up.

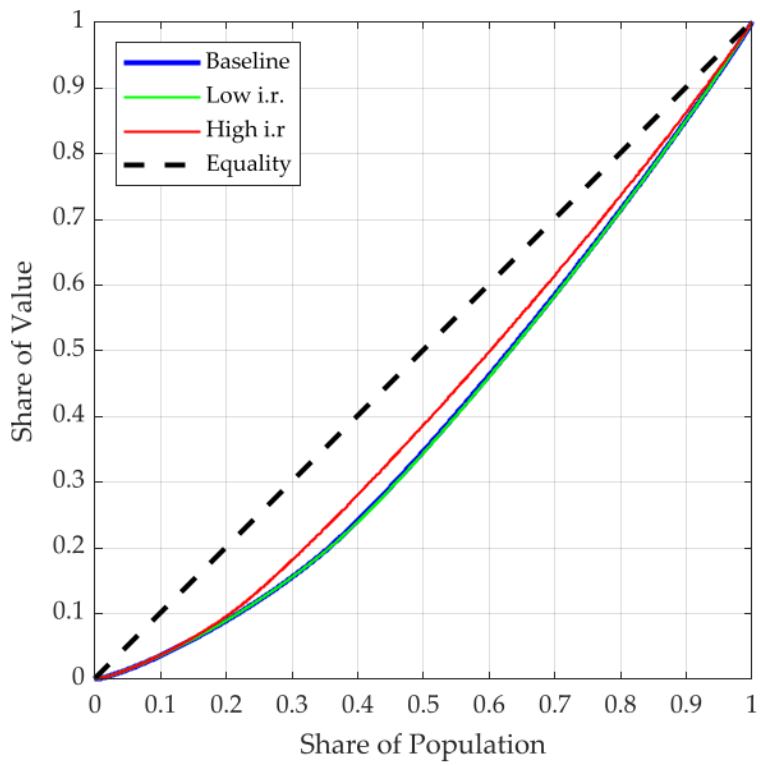


Figure 37: Lorenz Curves comparison between baseline and the exogenous upward and downward shift in the offer rate. Source: model simulations.

3.7 Conclusion

Over the past decade, the surge in mortgage accessibility driven by the entry of non-bank lenders has decreased mortgage rate dispersion among comparable borrowers (Bartlett et al., 2022). Nevertheless, disparities linked to individuals' loan shopping behavior and financial skill levels have remained persistent. Our paper quantifies the effect of individual financial skills and search effort in shaping mortgage repayment variation, giving rise to consumption inequality across households.

Backed by empirical findings and stylized facts from our previous work, we embed a micro-founded mortgage search framework within a continuous-time heterogeneous agents model. In our model, agents accumulate financial skills and exert search effort endogenously during the mortgage acquisition process. Exerting effort delivers cognitive costs that systematically vary with financial skill level.

Current borrowers can engage in a costly search and refinance their mortgages to lower their mortgage repayments. In the steady state, individual consumption growth depends on individual skills and assets, and is shaped by expected future mortgage rates and the need for precautionary savings for unexpected expenses. The steady state distinguishes between renters and homeowners, defining the joint distribution of mortgage rates, assets, skills, and productivity within the borrower group. Validity tests confirm the model's accuracy in reproducing consumption inequality using out-of-sample consumption data. The model indicates that search behavior and skills significantly contribute to determining mortgage rates among comparable borrowers, explaining 55% and 10% of the variation in rates.

We employ our model as a controlled environment and conduct a series of three experiments: introducing financial education, enhancing mortgage availability, and comparing various mean mortgage rate scenarios. Our first experiment uses a baseline framework, and shows that financial education stimulates skill accumulation and results in a modest relative increase in search effort among renters, ultimately yielding a higher homeownership rate. New homeowners exhibit higher financial proficiency; they allocate fewer resources to mortgage servicing and thereby contribute to decreased consumption inequality and a lower delinquency rate.

Our second test introduces increased mortgage availability, reflecting digital advancements in the U.S. mortgage market. We show that financial education reinforces skill accumulation and decreases the delinquency rate, mitigating the adverse effects of mortgage take-up among low-skilled borrowers. Mortgage accessibility effectively flattens out search costs across the skill distribution, showing negligible effects on skill accumulation. New mortgage owners are less financially savvy, and thus are more likely to face expense shocks, ultimately resulting in higher delinquency rates. We show that, with increased availability, financial education delivers a relatively higher skill level. In this regard, financial education accommodates growing trends in credit availability.

In our final experiment, we contrast two mean mortgage rate scenarios, holding the dispersion of mortgage offers constant. This approach accommodates the external shift in mean mortgage rates, encompassing mortgage relief policies that reduce payments for all existing borrowers. Lower rates benefit current homeowners, thereby intensifying the divergence in consumption levels between homeowners and renters. Within this context, search costs persistently lead to renters continuing to rent, forfeiting the advantages of comparatively lower mortgage payments. With lower mean rates, skill-based differences in mortgage take-up yield relatively higher consumption inequality. In contrast, the high-rate scenario closes the gap between rental and mortgage repayments, decreasing consumption inequality.

Backed by our current findings, our ongoing work includes the extension of the model to general equilibrium. A richer set of sources of household heterogeneity and a careful outline of the mortgage supply can yield insights into market responses to financial education and monetary policy.

Summary

This thesis comprises three chapters that investigate the underlying mechanisms in individual financial decision-making regarding retirement savings and mortgage choices. The first chapter examines differences in individual income expectations through the lens of extrapolative behavior and its impact on retirement savings. The study finds that pessimistic individuals prefer liquid accounts for immediate access, while rational and subjective workers' savings patterns vary over the life cycle. Despite incentives like employer matches and tax deferrals in 401(k) accounts, extrapolative workers initially delay their contributions, but gradually increase them, which reflects actual contribution data. My policy simulations show that mandatory automatic enrollment does not increase savings among extrapolative workers, because their expectations favor contributions to liquid accounts.

The second chapter, co-authored with Ante Šterc, investigates mortgage rate disparities among similar U.S. borrowers, highlighting the roles of financial literacy and loan shopping behavior. By merging two public U.S. datasets, the study finds that financial literacy follows a hump-shaped life cycle and significantly influences mortgage rates. Financially literate borrowers who compare multiple lenders secure lower mortgage rates, resulting in substantial savings over the loan term. The study also shows that financially unskilled borrowers have a greater likelihood of becoming delinquent within three years post-origination. These findings inform the mortgage search model developed in the subsequent chapter.

The third chapter, also co-authored with Ante Šterc, integrates a micro-founded mortgage search framework into a standard heterogeneous agents model of consumption and saving. The model, calibrated using data from Chapter 2, examines the effects of financial education, mortgage accessibility, and rate changes on financially unskilled households. Our key findings include that increased mortgage accessibility raises delinquency risks among low-skilled households, financial education mitigates these risks, and lower mortgage rates benefit high-skilled homeowners, exacerbating consumption disparities due to enhanced refinancing activities. Collectively, these chapters offer valuable insights into how individual behaviors and financial literacy affect retirement and mortgage decisions and have significant implications for policy design.

References

Achdou, Y., Han, J., Lasry, J.-M., Lions, P.-L., and Moll, B. (2022). Income and Wealth Distribution in Macroeconomics: A Continuous-time Approach. *The Review of Economic Studies*, 89(1):45–86.

Agarwal, S., Ben-David, I., and Yao, V. (2017). Systematic Mistakes in the Mortgage Market and Lack of Financial Sophistication. *Journal of Financial Economics*, 123(1):42–58.

Agarwal, S., Driscoll, J. C., Gabaix, X., and Laibson, D. (2007). The Age of Reason: Financial Decisions over the Lifecycle. Working Paper 13191, National Bureau of Economic Research Cambridge, Mass., USA.

Agarwal, S., Grigsby, J., Hortaçsu, A., Matvos, G., Seru, A., and Yao, V. (2020). Searching for approval. Working Paper 27341, National Bureau of Economic Research.

Agarwal, S., Rosen, R. J., and Yao, V. (2016). Why do Borrowers make Mortgage Refinancing Mistakes? *Management Science*, 62(12):3494–3509.

Alexandrov, A. and Koulayev, S. (2018). No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information. Working Paper 2017-01, Consumer Financial Protection Bureau Office.

Ameriks, J., Caplin, A., Leahy, J., and Tyler, T. (2007). Measuring Self-Control Problems. *American Economic Review*, 97(3):966–972.

Andersen, S., Campbell, J. Y., Nielsen, K. M., and Ramadorai, T. (2020). Sources of Inaction in Household Finance: Evidence from the Danish Mortgage Market. *American Economic Review*, 110(10):3184–3230.

Auclert, A., Rognlie, M., and Straub, L. (2020). Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model. Working Paper 26647, National Bureau of Economic Research.

Bailey, M., Dávila, E., Kuchler, T., and Stroebel, J. (2018). House Price Beliefs And Mortgage Leverage Choice. *The Review of Economic Studies*, 86(6):2403–2452.

Bartlett, R., Morse, A., Stanton, R., and Wallace, N. (2022). Consumer-Lending Discrimination in the FinTech Era. *Journal of Financial Economics*, 143(1):30–56.

Benartzi, S. and Thaler, R. (2007). Heuristics and Biases in Retirement Savings Behavior. *Journal of Economic Perspectives*, 21(3):81–104.

Bernheim, B. D., Fradkin, A., and Popov, I. (2015). The Welfare Economics of Default Options in 401(k) Plans. *American Economic Review*, 105(9):2798–2837.

Beshears, J., Choi, J. J., Iwry, J. M., John, D. C., Laibson, D., and Madrian, B. C. (2020). Building Emergency Savings through Employer-Sponsored Rainy-Day Savings Accounts. *Tax Policy and the Economy*, 34(1):43–90.

Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C., and Skimmyhorn, W. L. (2022). Borrowing to Save? The Impact of Automatic Enrollment on Debt. *The Journal of Finance*, 77(1):403–447.

Bhandari, A., Borovička, J., and Ho, P. (2016). Identifying Ambiguity Shocks in Business Cycle Models Using Survey Data. Working Paper 22225, National Bureau of Economic Research.

Bhattacharya, K., Kapoor, A., and Madan, A. (2021). Five Trends Reshaping the US Home Mortgage Industry. Technical report, McKinsey & Company.

Bhutta, N., Blair, J., and Dettling, L. (2022). The Smart Money is in Cash? Financial Literacy and Liquid Savings Among U.S. Families. *Journal of Accounting and Public Policy*, page 107000.

Bhutta, N., Fuster, A., and Hizmo, A. (2020). Paying Too Much? Price Dispersion in the U.S. Mortgage Market. Finance and Economics Discussion Series 2020-062, Board of Governors of the Federal Reserve System.

Blanchett, D., Finke, M. S., and Liu, Z. (2021). The Impact of Employer Defaults and Match Rates on Retirement Saving. Research Paper 3992899, Social Science Research Network.

Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption Inequality and Partial Insurance. *American Economic Review*, 98(5):1887–1921.

Bordalo, P., Gennaioli, N., Porta, R. L., and Shleifer, A. (2019). Diagnostic Expectations and Stock Returns. *The Journal of Finance*, 74(6):2839–2874.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2018). Diagnostic Expectations and Credit Cycles. *The Journal of Finance*, 73(1):199–227.

Browning, M., Hansen, L. P., and Heckman, J. J. (1999). Micro Data and General Equilibrium Models. *Handbook of macroeconomics*, 1:543–633.

Choi, J. J., Laibson, D., Madrian, B. C., and Metrick, A. (2003). Optimal Defaults. *American Economic Review*, 93(2):180–185.

Choi, J. J. and Mertens, T. M. (2019). Extrapolative Expectations and the Equity Premium. Technical Report 3462056, Social Science Research Network.

Choukhmane, T. (2021). Default Options and Retirement Saving Dynamics. In *Working Paper*.

Cocco, J. F., Gomes, F., and Lopes, P. (2022). Evidence on Expectations of Household Finances. Technical Report DP17447, Centre for Economic Policy Research.

Cocco, J. F., Gomes, F. J., and Maenhout, P. J. (2005). Consumption and portfolio choice over the life cycle. *The Review of Financial Studies*, 18(2).

Coen, J., Kashyap, A. K., and Rostom, M. (2023). Price discrimination and mortgage choice. Working Paper 31652, National Bureau of Economic Research.

Cornelli, G., Frost, J., Gambacorta, L., and Jagtiani, J. (2022). The Impact of Fintech Lending on Credit Access for U.S. Small Businesses.

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

Damen, S. and Buyst, E. (2017). Mortgage Shoppers: How Much do They Save? *Real Estate Economics*, 45(4):898–929.

Das, M. and Van Soest, A. (1999). A Panel Data Model for Subjective Information on Household Income Growth. *Journal of Economic Behavior & Organization*, 40(4):409–426.

De Nardi, M. and Fella, G. (2017). Saving and Wealth Inequality. *Review of Economic Dynamics*, 26:280–300.

De Nardi, M., French, E., and Jones, J. B. (2009). Life Expectancy and Old Age Savings. *American Economic Review*, 99(2):110–15.

Degerli, A. and Wang, J. (2022). The Rise of Nonbanks and the Quality of Financial Services: Evidence from Consumer Complaints. *Finance and Economics Discussion Series*, (2022-059).

DellaVigna, S. (2018). Chapter 7 - structural behavioral economics. In Bernheim, B. D., DellaVigna, S., and Laibson, D., editors, *Handbook of Behavioral Economics - Foundations and Applications 1*, volume 1 of *Handbook of Behavioral Economics: Applications and Foundations 1*, pages 613–723. North-Holland.

Devlin-Foltz, S., Henriques, A., and Sabelhaus, J. (2015). The Evolution of Retirement Wealth. *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, 108:1–34.

dHaultfoeuille, X., Gaillac, C., and Maurel, A. (2021). Rationalizing Rational Expectations: Characterizations and Tests. *Quantitative Economics*, 12(3):817–842.

Druedahl, J. (2020). A Guide On Solving Non-Convex Consumption-Saving Models. *Computational Economics*, pages 1–29.

Druedahl, J. and Jørgensen, T. H. (2020). Can Consumers Distinguish Persistent from Transitory Income Shocks? *The Economic Journal*, 130(632):2410–2437.

Duarte, V., Fonseca, J., Goodman, A. S., and Parker, J. A. (2021). Simple Allocation Rules and Optimal Portfolio Choice over the Lifecycle. Technical Report 29559, National Bureau of Economic Research.

Enamorado, T., Fifield, B., and Imai, K. (2019). Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records. *American Political Science Review*, 113(2):353–371.

Exler, F., Livshits, I., MacGee, J., and Tertilt, M. (2021). Consumer Credit with Overoptimistic Borrowers.

Fagereng, A., Holm, M. B., Moll, B., and Natvik, G. (2019). Saving Behavior Across the Wealth Distribution: The Importance of Capital Gains. Working Paper 26588, National Bureau of Economic Research.

Farber, H. S. et al. (2005). What do We Know About Job Loss in the United States? Evidence from the Displaced Workers Survey, 1984–2004. *Economic Perspectives*, 29(2):13–28.

Fellegi, I. P. and Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*, 64(328):1183–1210.

Fuster, A., Plosser, M., Schnabl, P., and Vickery, J. (2019). The Role of Technology in Mortgage Lending. *The Review of Financial Studies*, 32(5):1854–1899.

Garner, T. I., Martin, R. S., Matsumoto, B., and Curtin, S. (2022). Distribution of US Personal Consumption Expenditures for 2019: A Prototype Based on Consumer Expenditure Survey Data. Technical report, BLS working paper 557.

Gelman, A. (2008). Scaling Regression Inputs By Dividing By Two Standard Deviations. *Statistics in Medicine*, 27(15):2865–2873.

Gerardi, K., Willen, P. S., and Zhang, D. H. (2023). Mortgage Prepayment, Race, and Monetary Policy. *Journal of Financial Economics*, 147(3):498–524.

Ghilarducci, T., Saad-Lessler, J., and Reznik, G. (2018). Earnings Volatility and 401(k) Contributions. *Journal of Pension Economics and Finance*, 17(4):554–575.

Goda, G. S., Levy, M. R., Manchester, C. F., Sojourner, A., and Tasoff, J. (2020). Who is a Passive Saver Under Opt-in and Auto-enrollment? *Journal of Economic Behavior & Organization*, 173:301–321.

Grevenbrock, N., Groneck, M., Ludwig, A., and Zimper, A. (2021). Cognition, Optimism, and the Formation of Age-dependent Survival Beliefs. *International Economic Review*, 62(2):887–918.

Guerrieri, V. and Lorenzoni, G. (2017). Credit Crises, Precautionary Savings, and the Liquidity Trap. *The Quarterly Journal of Economics*, 132(3):1427–1467.

Jappelli, T. and Padula, M. (2017). Consumption Growth, the Interest Rate, and Financial Sophistication. *Journal of Pension Economics and Finance*, 16(3):348–370.

Kapička, M. and Neira, J. (2019). Optimal Taxation with Risky Human Capital. *American Economic Journal: Macroeconomics*, 11(4):271–309.

Keys, B. J., Pope, D. G., and Pope, J. C. (2016). Failure to Refinance. *Journal of Financial Economics*, 122(3):482–499.

Krijnen, J. M., Zeelenberg, M., Breugelmans, S. M., and Van Der Schors, A. (2022). Intention and Action in Retirement Preparation. *Behavioural Public Policy*, 6(2):191–212.

Love, D. A. (2006). Buffer Stock Saving in Retirement Accounts. *Journal of Monetary Economics*, 53:1473–1492.

Lusardi, A. (2019). Financial Literacy and the need for Financial Education: Evidence and Implications. *Swiss Journal of Economics and Statistics*, 155(1):1–8.

Lusardi, A., Michaud, P.-C., and Mitchell, O. S. (2017). Optimal Financial Knowledge and Wealth Inequality. *Journal of Political Economy*, 125(2):431–477.

Lusardi, A., Michaud, P.-C., and Mitchell, O. S. (2020). Assessing the Impact of Financial Education Programs: A Quantitative Model. *Economics of Education Review*, 78:101899.

Lusardi, A. and Mitchell, O. S. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature*, 52(1):5–44.

Lusardi, A., Mitchell, O. S., and Curto, V. (2010). Financial Literacy Among the Young. *Journal of consumer affairs*, 44(2):358–380.

Massenot, B. and Pettinicchi, Y. (2019). Can Households See Into the Future? Survey Evidence from the Netherlands. *Journal of Economic Behavior & Organization*, 164:77–90.

Mazzonna, F. and Peracchi, F. (2023). Are Older People Aware of Their Cognitive Decline? Misperception and Financial Decision Making. *Journal of Political Economy; Forthcoming*.

McCafrey, O. (2021). Nonbank Lenders Are Dominating the Mortgage Market: Nonbanks issued more than two-thirds of mortgages in 2020, their highest market share on record.

Mian, A. R., Straub, L., and Sufi, A. (2020). The saving glut of the rich. Working Paper 26941, National Bureau of Economic Research.

Mitchell, O. S., Mottola, G. R., Utkus, S. P., and Yamaguchi, T. (2006). The Inattentive Participant: Portfolio Trading Behavior in 401 (k) Plans. *Michigan Retirement Research Center Research Paper No. WP*, 115.

Olafsson, A. and Pagel, M. (2018). The Retirement-Consumption Puzzle: New Evidence from Personal Finances. Working Paper w24405, National Bureau of Economic Research Paper.

Parker, J. A., Schoar, A., Cole, A. T., and Simester, D. (2022). Household Portfolios and Retirement Saving over the Life Cycle. Working Paper 29881, National Bureau of Economic Research.

Paz-Pardo, G. (2024). Homeownership and Portfolio Choice over the Generations. *American Economic Journal: Macroeconomics*, 16(1):207–237.

Schlafmann, K. (2020). Housing, Mortgages, and Self-Control. *The Review of Financial Studies*, 34(5):2648–2687.

Schlafmann, K. and Rozsypal, F. (2023). Overpersistence Bias in Individual Income Expectations and its Aggregate Implications. *American Economic Journal: Macroeconomics*.

Storesletten, K., Telmer, C. I., and Yaron, A. (2004). Consumption and Risk Sharing Over the Life cycle. *Journal of Monetary Economics*, 51(3):609–633.

Zhou, X. (2022). FinTech Lending, Social Networks and the Transmission of Monetary Policy. Working Paper 2203, FRB of Dallas.

Annexes

A Extrapolative Expectations and Retirement Savings

A.1 Three-period model, proof

The maximization problem for the worker in the stylized model is

$$\max_{s_1, s_1^R \geq 0} u(c_1) + \mathbb{E}[u(c_2) + u(c_3)] \text{ such that } c_1 + s_1 + s_1^R = y_1,$$
$$c_2 + s_2 = y_2,$$
$$c_3 = Rs_1^R + s_2,$$

where $u(c) = \dfrac{c^{1-\gamma}}{1-\gamma}$ and 2^{nd}-period income is stochastic and distributed as

$$y_2 \sim \begin{pmatrix} y_L & y_H \\ p & 1-p \end{pmatrix}.$$

The proof builds on two lemmas that ensure that the savings constraint in the second period does not bind in the case of high income realization. This is true for reasonable conditions on the retirement savings account return:

Lemma 1. *If $R < (1 + R^{\frac{\gamma-1}{\gamma}})$ and $y_2 = y_H \implies$ borrowing constraint does not bind (i.e., $s_2 > 0$, in the high income state).*

Proof. Suppose that the claim is not true. In this case, since the high-income agent is binding ($s_2(y_H) = 0$), then the same must hold for the low-income recipient $\implies s_2(y_L) = 0$. This means that $c_3(y_H) = c_3(y_L) = Rs_1^R$. The optimality of a binding constraint implies that $c_2(y_L)$ and $c_2(y_H)$ are both strictly lower than the 3^{rd}-period consumption[37]. However, the 1^{st}-period FOC implies $u'(c_1) = Ru'(c_3) \implies c_3 = R^{\frac{1}{\gamma}} c_1 \implies c_1 = \dfrac{y_1 - s_1}{R^{\frac{1}{\gamma}-1}}$ and $c_3 = \dfrac{y_1 - s_1}{1 + R^{\frac{1}{\gamma}-1}} R^{\frac{1}{\gamma}}$. But then $c_3 > c_2^H \iff \dfrac{R^{\frac{1}{\gamma}}}{1 + R^{\frac{1}{\gamma}-1}}(y - s_1) > y_H + s_1$, which yields contradiction, since $R < (1 + R^{\frac{\gamma-1}{\gamma}})$. □

Lemma 2. *If the agent chooses to allocate to both liquid and illiquid savings and $R \neq 1$, the borrowing constraint binds in the low income state y_L.*

Proof. Suppose that the claim is not true, i.e. that the low-income agents do not bind. The first period optimality condition states

$$u'(c_1) = R\mathbb{E}u'(c_3),$$

[37]Otherwise, both agents would be able to smooth their consumption across periods.

for both high-income and low-income agents. Also, optimality in the second period (taking the first-order condition with respect to liquid savings) yields $u'(c_1) = \mathbb{E}u'(c_2)$. Taking the assumption into account, if the savings constraint does not bind for both types of agents then the consumption smoothing ($c_2^{L,H} = c_3^{L,H}$) implies

$$\mathbb{E}u'(c_3) = R\mathbb{E}u'(c_3),$$

which cannot be true since $R \neq 1$. $\square$

Ultimately, the effect of pessimism regarding illiquid to liquid savings accounts reallocation is implied by the two lemmas and the Implicit Function Theorem.

Proposition 1. *Suppose that retirement savings exhibit greater returns than liquid assets, $R > 1$, but are not too large, satisfying $R < (1 + R^{\frac{\gamma-1}{\gamma}})$. Define $s_1(p)$ an $s_1^R(p)$ as optimal liquid and illiquid savings. Given that the uncertainty in second period income is large enough, $s_1(p) > 0$, the following inequality holds:*

$$\frac{\partial s_1}{\partial p} > \frac{\partial s_1^R}{\partial p}.$$

That is, an increase in pessimism (assigning $\tilde{p} > p$) implies an increase liquid asset holdings, and a decrease retirement savings.

Proof. Optimal assets allocations are pinned down by the Euler equations

$$u'(y - s_1 - s_1^R) - \frac{1-p}{2}u'\left(\frac{y_H + s_1 + Rs_1^R}{2}\right) - pu'\left(y_L + s_1\right) = 0 \quad F_1$$

$$u'(y - s_1 - s_1^R) - \frac{R(1-p)}{2}u'\left(\frac{y_H + s_1 + Rs_1^R}{2}\right) - Rpu'(Rs_1^R) = 0 \quad F_2,$$

which implicitly define s_1 and s_1^R as functions of the probability parameter p. The Implicit Function Theorem for $F = (F_1, F_2)$ implies the existence of a function $f(p) = (s_1(p), s_1^R(p))$ such that in the optimum

$$F(p, s_1, s_1^R) = 0 \quad \Longrightarrow \quad F(p, s_1(p), s_1^R(p)) = 0.$$

Now, to determine the effect of a change in the percieved probability of a low-income realization I derive

$$\underbrace{\begin{pmatrix} \partial_2 F_1 & \partial_3 F_1 \\ \partial_2 F_2 & \partial_3 F_2 \end{pmatrix}}_{\mathbf{F}} \begin{pmatrix} \partial_1 f \\ \partial_2 f \end{pmatrix} = \begin{pmatrix} -\partial_1 F_1 \\ -\partial_1 F_2 \end{pmatrix},$$

where

$$\begin{pmatrix} \partial_1 f \\ \partial_2 f \end{pmatrix} = \begin{pmatrix} \frac{\partial s_1}{\partial p} \\ \frac{\partial s_1^R}{\partial p} \end{pmatrix}.$$

The inverse of a 4-dimensional matrix is given with

$$\begin{pmatrix} \partial_2 F_1 & \partial_3 F_1 \\ \partial_2 F_2 & \partial_3 F_2 \end{pmatrix}^{-1} = \frac{1}{\det \mathbf{F}} \begin{pmatrix} \partial_3 F_2 & -\partial_2 F_2 \\ -\partial_3 F_1 & \partial_2 F_1 \end{pmatrix}$$

used to obtain $\frac{\partial s_1}{\partial p}$ and $\frac{\partial s_1^R}{\partial p}$ at the solution.

Since

$$\begin{pmatrix}\partial_1 F_1\\ \partial_1 F_2\end{pmatrix} = \begin{pmatrix}\frac{u'(c_2^H)}{2} - u'(c_2^L)\\ \frac{Ru'(c_3^H)}{2} - Ru'(c_3^L)\end{pmatrix}$$

and

$$\begin{aligned}\det \mathbf{F} &= \det \begin{pmatrix}-u''(c_1) - \frac{1-p}{4}u''(c_2^H) - pu''(c_2^L) & -u''(c_1) - \frac{R(1-p)}{4}u''(c_2^H)\\ -u''(c_1) - \frac{R(1-p)}{4}u''(c_3^H) & -u''(c_1) - \frac{R^2(1-p)}{4}u''(c_3^H) - R^2pu''(c_3^L)\end{pmatrix}\\
&\overset{c_3^H=c_2^H}{=} \left(u''(c_1) + \frac{1-p}{4}u''(c_2^H) + pu''(c_2^L)\right)\left(u''(c_1) + \frac{R^2(1-p)}{4}u''(c_2^H) + R^2pu''(c_3^L)\right)\\
&\quad - \left(u''(c_1) + \frac{R(1-p)}{4}u''(c_3^H)\right)^2.\end{aligned}$$

Simpifying the expression yields

$$\begin{aligned}\det \mathbf{F} = {}& \cancel{u''(c_1^2)} + \left(\frac{R^2(1-p)}{4} + \frac{1-p}{4}\right)u''(c_1)u''(c_2^H) + R^2pu''(c_1)u''(c_3^L)\\
&+ \cancel{\frac{(1-p)^2R^2}{16}\left[u''(c_2^H)\right]^2} + \frac{R^2p(1-p)}{4}u''(c_2^H)u''(c_3^L) + pu''(c_2^L)u''(c_1)\\
&+ \frac{R^2(1-p)p}{4}u''(c_2^L)u''(c_2^H) + R^2p^2u''(c_2^L)u''(c_3^L) - \cancel{u''(c_1^2)} - \frac{R(1-p)}{2}u''(c_1)u''(c_2^H)\\
&- \cancel{\frac{R^2(1-p)^2}{16}\left[u''(c_2^H)\right]^2},\end{aligned}$$

so

$$\begin{aligned}\det \mathbf{F} = {}& \left(\frac{R^2(1-p)}{4} + \frac{1-p}{4} - \frac{R(1-p)}{2}\right)u''(c_1)u''(c_2^H) + R^2pu''(c_1)u''(c_3^L)\\
&+ \frac{R^2p(1-p)}{4}u''(c_2^H)u''(c_3^L) + pu''(c_2^L)u''(c_1) + \frac{R^2(1-p)p}{4}u''(c_2^L)u''(c_2^H)\\
&+ R^2p^2u''(c_2^L)u''(c_3^L) > 0,\end{aligned}$$

since $p \in \langle 0, 1\rangle$, i.e. the Implicit Function Theorem is applicable in this setting.

Altogether

$$\begin{pmatrix}\frac{\partial s_1}{\partial p}\\ \frac{\partial s_1^R}{\partial p}\end{pmatrix} = \frac{1}{\det \mathbf{F}}\begin{pmatrix}\partial_3 F_2(\frac{u'(c_2^H)}{2} - u'(c_2^L)) - R\partial_2 F_2(\frac{u'(c_2^H)}{2} - u'(c_3^L))\\ -\partial_3 F_1(\frac{u'(c_2^H)}{2} - u'(c_2^L)) + R\partial_2 F_1(\frac{u'(c_2^H)}{2} - u'(c_3^L))\end{pmatrix}.$$

Simultaneously

$$\frac{\partial s_1^*}{\partial p} > 0 \quad \text{and} \quad \frac{\partial s_1^R*}{\partial p} < 0$$

hold under two conditions.

First, the two lemmas imply

$$c_2^L < c_3^L \overset{u''<0}{\Longrightarrow} u'(c_3^L) < u'(c_2^L),$$

so

$$-\partial_3 F_1(\frac{u'(c_2^H)}{2} - u'(c_2^L)) + R\partial_2 F_1(\frac{u'(c_2^H)}{2} - u'(c_3^L)) < \underbrace{(-\partial_3 F_1 + R\partial_2 F_1)}_{<0}(\frac{u'(c_2^H)}{2} - u'(c_3^L)).$$

Under the assumption that "the agent is not too hungry in the low-income case" $u'(c_3^L) < \frac{u'(c_3^H)}{2}$, the retirement savings decrease once the probability of the low-income realization increases. That is, pessimistic expectations $\tilde{p} > p$ the retirement savings are decreased.

Under the same assumption it has to hold

$$\partial_3 F_2(\frac{u'(c_2^H)}{2} - u'(c_2^L)) - R\partial_2 F_2(\frac{u'(c_2^H)}{2} - u'(c_3^L)) > (\partial_3 F_2 - R\partial_2 F_2)\underbrace{(\frac{u'(c_2^H)}{2} - u'(c_3^L))}_{<0} > 0.$$

$$\partial_3 F_2 - R\partial_2 F_2 < 0 \Leftrightarrow (R-1)(u''(c_1) - R^2 p u''(c_3^L)) \overset{c_3^L > c_1}{<} (R - 1 - R^2 p)u''(c_3^L) \overset{p\in\langle 0,\frac{1}{4}\rangle}{<} 0.$$

Altogether, we have

$$\frac{\partial s_1^*}{\partial p} > 0 > \frac{\partial s_1^R *}{\partial p}.$$

□

A.2 Additional estimates

Age coefficients

Forecast error density estimates in the text reveal quantile-based differences and the transition from pessimism to optimism. Given that the regression coefficients with age polynomial are significant, albeit of different signs, this serves as another argument that age does affect the income growth bias.

Kernel density estimates consider only the working-age population and reveal that error distributions differ across age groups. During work life, the mode of the error distribution is positive. This finding indicates that even experienced workers do not completely correct their forecasts. The distribution changes shape over age groups, owing to income volatility for younger cohorts. Correspondingly, in the model, agents start to add to their retirement accounts as the effect of misperception in income volatility decreases.

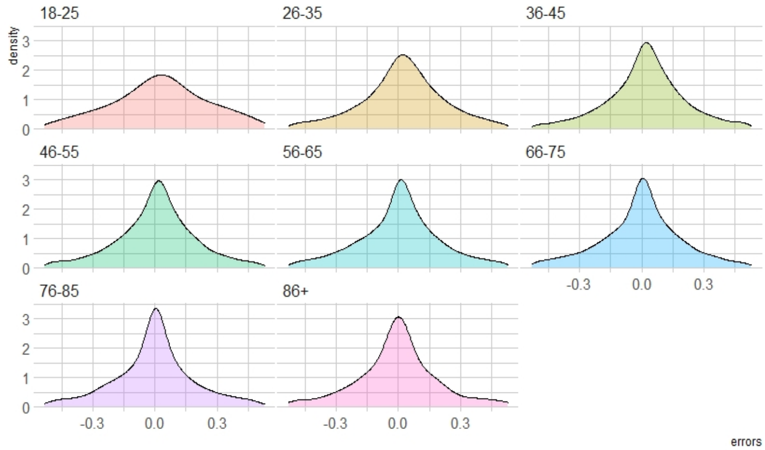


Figure 38: Income growth bias density by age group, MSC data

A.2.1 Regression checks

In addition to the linear regression model in the text, I estimate the model with HH who had their first interview in the second half of the year. Their responses are not sensitive to the imperfect time overlap between the period of expectations and realizations. Again, standard errors are clustered at the region level. Signs of all coefficients remain the same, while the size of coefficient with the income quantile input increases (Table 26). Again, results indicate that household tend to be overly pessimistic at the left end of the income distribution while their right-end counterparts tend to be overoptimistic.

Table 26: Linear Regression Results

	Dependent variable:
	Income Growth Forecast Errors
q_2	0.225***
	(0.008)
q_3	0.306***
	(0.008)
q_4	0.356***
	(0.009)
q_5	0.430***
	(0.014)
male	−0.015*
	(0.006)
no HS	0.039***
	(0.006)
college	−0.054***
	(0.003)
age	−0.114***
	(0.026)
age^2	0.155***
	(0.034)
1 *adult*	0.106***
	(0.005)
>2 *adults*	−0.029
	(0.008)
Constant	−0.381***
	(0.012)
Observations	29,414
Note:	*p<0.1; **p<0.05; ***p<0.01

A.2.2 Housing as a mean of saving for retirement

Finally, retirement savings may not include only private retirement accounts, as people may be saving in other illiquid savings such as housing. I address the issue of saving for retirement through real estate by checking to what extent home ownership affects retirement confidence. I use the survey question that asks to assign the probability of having a comfortable retirement **only from social security and job pensions**.

I binned subjective probabilities into four separate groups ($<25\%$, $25-50\%$, $50-75\%$ and $75-100\%$) that translate into groupings from harsh pessimists to enthusiastic optimists. The estimates imply that retirement confidence rises with age and income, whereas owning a home does not have a significant effect. Since the recent income growth forecast error is not significant, I conclude that retirement confidence is based on individual attitudes (persistent pessimism or optimism).

	Dependent variable:
	P(comfortable retirement)
male	0.215***
	(0.026)
no HS	0.093
	(0.074)
college	0.035
	(0.028)
age	0.507***
	(0.029)
1 *adult*	0.055
	(0.034)
> 2 *adults*	0.081*
	(0.041)
q_2	0.122***
	(0.058)
q_3	0.259***
	(0.056)
q_4	0.416***
	(0.058)
q_5	0.482***
	(0.060)
homeowner	0.060
	(0.037)
Observations	20,743
Note: Controlled for year effects, age is standardized.	*p<0.1; **p<0.05; ***p<0.01

Table 27: Ordered logistic regression results

In addition to retirement confidence, I check to what extent homeownership affects future income growth forecast. Since my model incorporates illiquid savings in the form of the retirement account, I check if housing assets position affect income growth forecasts. Including homeownership in the regression analysis reduces number of observations to 37,000. The income quantile coefficients remain similar. Moreover, among homeowners, home value has a significant,albeit small, effect (Table 28).

Job loss predictions

In the main text I argue that the income quintile is the significant predictor for pessimistic job loss predictions. Consequently, once I compare empirical job separation rates to the predicted values I argue that overstating these probabilities remains consistent with

Table 28: Linear regression results

	Dependent variable:	
	Income growth forecast errors	
	Homeowners only	All
Income quantile:		
q_2	0.176***	0.189***
	(0.018)	(0.014)
q_3	0.248***	0.273***
	(0.008)	(0.007)
q_4	0.272***	0.316***
	(0.010)	(0.009)
q_5	0.355***	0.381***
	(0.018)	(0.015)
male	−0.008	−0.014***
	(0.007)	(0.006)
HS	0.030*	0.050***
	(0.014)	(0.008)
College	−0.031***	−0.049***
	(0.008)	(0.004)
age	−0.123	0.016
	(0.058)	(0.043)
age^2	0.095	0.048
	(0.052)	(0.043)
1 adult	0.068***	0.081***
	(0.011)	(0.005)
$>$ 2 adults	−0.022*	−0.039***
	(0.009)	(0.011)
Home value, quantiles:		
h_2	−0.037***	–
	(0.008)	
h_3	−0.031***	–
	(0.012)	
h_4	−0.067***	–
	(0.016)	
h_5	−0.088***	–
	(0.013)	
Renter	–	0.068***
		(0.003)
Constant	−0.080	−0.334
	(0.362)	(0.330)
Observations	11,992	36,932
Note:	*p<0.1; **p<0.05; ***p<0.01	

how the income growth forecast bias is implemented in the life cycle model. Thus, the fact that the mispercieved persistence parameter implies the mispercieved volatility remains consistent with empirical estimates. The only age group that is significant are workers closer to retirement age.

Table 29: Ordered logistic regression results

	Dependent variable:
	P(job loss within 5 years)
male	0.081***
	(0.029)
no HS	0.288***
	(0.080)
college	$-$0.123***
	(0.032)
age 25-34	$-$0.112
	(0.073)
age 35-44	$-$0.050
	(0.068)
age 45-54	$-$0.023
	(0.067)
age 55-66	$-$0.550***
	(0.069)
1 *adult*	$-$0.049
	(0.038)
>2 *adults*	0.226***
	(0.047)
q_2	$-$0.001
	(0.062)
q_3	$-$0.115*
	(0.060)
q_4	$-$0.208***
	(0.062)
q_5	$-$0.324***
	(0.064)
Observations	20,395
Note: Year effects are not reported.	*p<0.1; **p<0.05; ***p<0.01

A.3 Model equations and numerical implementation

The agent's problem can be formulated as the dynamic programming problem, for the state variables mentioned in this paper. The model for subjective expectations satisfies the same equation with different expectations, so the derivations hold for both subjective and RE agents. The model given in the paper satisfies the Bellman equation:

$$v(1, m_t, p_t, \xi_t, \zeta_t, n_t) = \max_{0.03 \leq d_t \leq 1, c_t \geq 0} u(c_t) + \beta \mathbb{E}_t[v(1, m_{t+1}, p_{t+1}, \xi_{t+1}, \zeta_{t+1}, n_{t+1})] \tag{13}$$
$$\text{such that all the equations hold}$$

The value function V_t from the paper is not necessarily concave because of the discrete opting-in decision as an absorbing state. The *Nested endogenous grid method* uses the FOC for consumption[38]

$$c_t \ldots u'(c_t) = \beta R_a v_{m,t+1}(1, m_{t+1}, p_{t+1}, \xi_{t+1}, \zeta_{t+1}, n_{t+1})$$

and the standard approach of the EGM in general - computing the continuation value beforehand. The continuation value is obtained with functions of *post-decision* variables that map the solution into *pre-decision* variables. Following Druedahl (2020)

$$w_t(a_t, b_t, p_t) = \beta \mathbb{E}[v_t(1, m_{t+1}, p_{t+1}, \xi_{t+1}, \zeta_{t+1}, n_{t+1})]$$

for end-of period assets a_t and b_t and the persistent component p_t. The agent who does not contribute to DC account in time t, faces the standard consumption-savings problem, which is easily solved using EGM. Fixing $d_t \implies b_t = n_t + d_t y_t + \chi \log(1 + d_t y_t)$ and $n_t t + 1 = R_b b_t$, (13) boils down to

$$v(1, m_t, p_t, \xi_t, \zeta_t, n_t | d_t) = \max_{c_t \geq 0} u(c_t) + w_t(p_t, a_t, b_t)]$$
$$\text{such that} \quad c_t \leq m_t - y_t d_t - a_t$$
$$m_{t+1} = R_a a_t + y_{t+1},$$

which fixes the idea of interpolation of the continuation value for a_t, b_t and consequently using the FOC

$$u_c(c_t | d_t) = w_{a,t}(p_t, a_t, b_t) \implies \quad m_t = a_t + c_t, \tag{14}$$

and using the upper envelope method to interpolate values for all d_t on the exogeneous grid for cash-on-hand. Once consumption is calculated from (14) for each d_t, I use the grid search as in Druedahl (2020) and evaluate the optimal d_t. Even though interpolating the continuation value via post-decision variables seems cumbersome, I use the modified version of the interpolation, which shortens the grid-search procedure, due to monotonicity of the value function with respect to cash-on-hand generated by the previous period monotonically increasing assets.

[38] Interior solution

If the agent does not contribute, the problem is then similar to the contributor's, i.e. it can be nested into the special case for $d_t = 0$. I solve for the problem using the same "inner" function as above, using the post-decision value function. However, the post-decision function is corrected for $b_t = 0$. Once both consumption choices are computed, I use the upper envelope method that combines the solutions to the common grid for cash-on-hand, so that the values are comparable and defined on a regular grid.

Once retired, each agent gets the annuity payment out of their retirement account, so the retirement consumption depends on both liquid account and the amount saved for retirement through the DC account. Agents who did not contribute to the account get the minimum yearly income[39]. I build on Druedahl (2020) and extend the retirement consumption function solution method to a two-dimensional space.

I build on Druedahl (2020) and implement the solution method for a persistent ($AR(1)$) process. That is, I extend the method to track all the combinations of shocks to both the persistent and transitory components (altogether 30 combinations).

The grids are finer at low values of both savings accounts to take a closer look at the behavior of the poor low-income workers (or workers with low retirement savings). Utility function is a standard CRRA function

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma},$$

where $\gamma = 2$. I did not resort to high γs to isolate the effect of subjective expectations on the perceived variance of the future income.

Consumption and savings paths

Both correct and biased income forecasts imply consumption and savings paths that follow the usual patterns in the data. For example, both consumption paths exhibit a decrease in consumption towards the end of life, commonly stated as the *retirement consumption puzzle* in retirement studies (De Nardi et al., 2009; Olafsson and Pagel, 2018).

The effect of misperceived volatility acts across the income distribution is shown in the policy function plots (Figure 40). The shares of DC contributors are lower for all income quantiles. Ultimately, all workers start adding and catching up. Finally, there is a point in the work-life where workers forgo their liquid assets and start adding to illiquid ones.

Policy functions, 3-d planes

The paper states that differences in life-cycle paths come from differences in period-by-period consumption and savings allocations. Period-policy functions reflect additional precaution with young subjective workers and a slow increase in retirement contributions

[39] I computed solutions for various cases of minimum retirement subsidy. In every scenario upper quintiles end up contributing at some point in their worklife, so the subsidy is set at the lowest value of income grid.

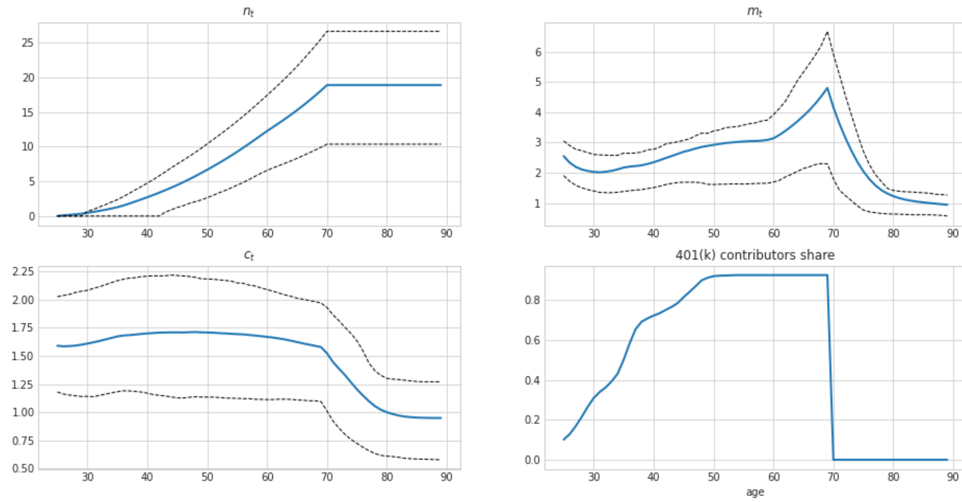


Figure 39: Lifecycle paths - RE

later on in life. All of the policy functions are depicted as functions of liquid and illiquid savings accounts. I denote the median within the income distribution with a red dot.

Consumption policy differs across income quantiles and, of course, expectations about future income. Figure (40) shows consumption as a function of savings levels for workers in the top 50% of the income distribution at the age of 30. At all savings levels, subjective workers consume less than their rational counterparts (Figure 40). Moreover, having a large amount of retirement savings does not imply higher consumption in the subjective expectation solution (Figure (40), right). Therefore, precautionary motives generate saving at the beginning of the work life.

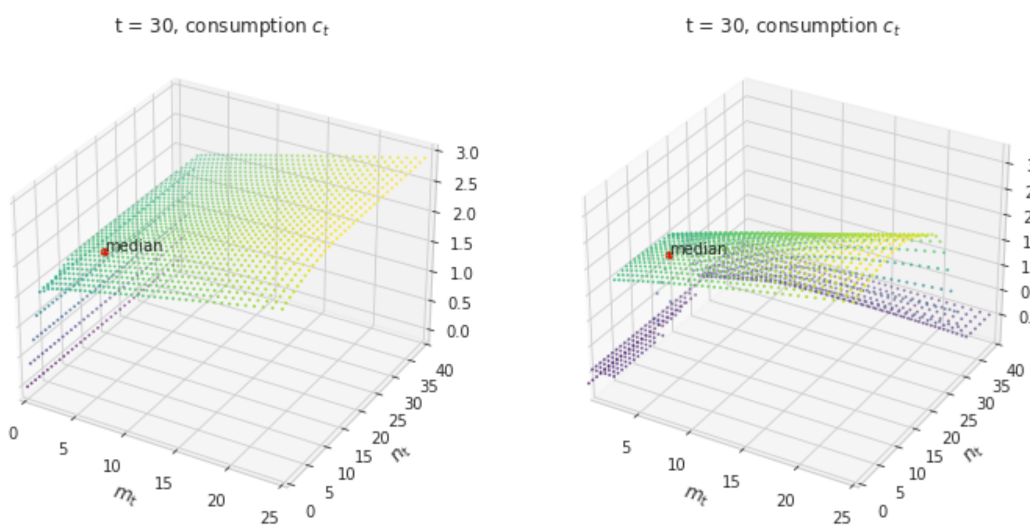


Figure 40: Consumption function, RE (left) and subjective expectations (right) solution.

Corresponding to everything presented in the main part of the paper, contribution rates switch from being lower for subjective workers at the beginning of work life. Figure 41, depicts contribution rates for top 50% of the income distribution - rational contributor adds 12%, whereas subjective one adds 6% of her current income. Later on in the work life, subjective workers start adding more and catch up (Figure 42, depicted for the top 50% of the income distribution). At the median, rational contributor adds 9% out of their wage, whereas subjective contributors add 12% (Figures 42, red dots in respective graphs). As the effects of income volatility overstating fade, workers with significantly low liquid savings contribute at the highest rates possible. Overall, contribution rates switch places for **all income quantiles**, as functions of liquid savings m_t and illiquid retirement account amounts n_t. Therefore, the effect of extrapolation does not depend on the amount of workers' savings, owing to the misperception of future income realizations only.

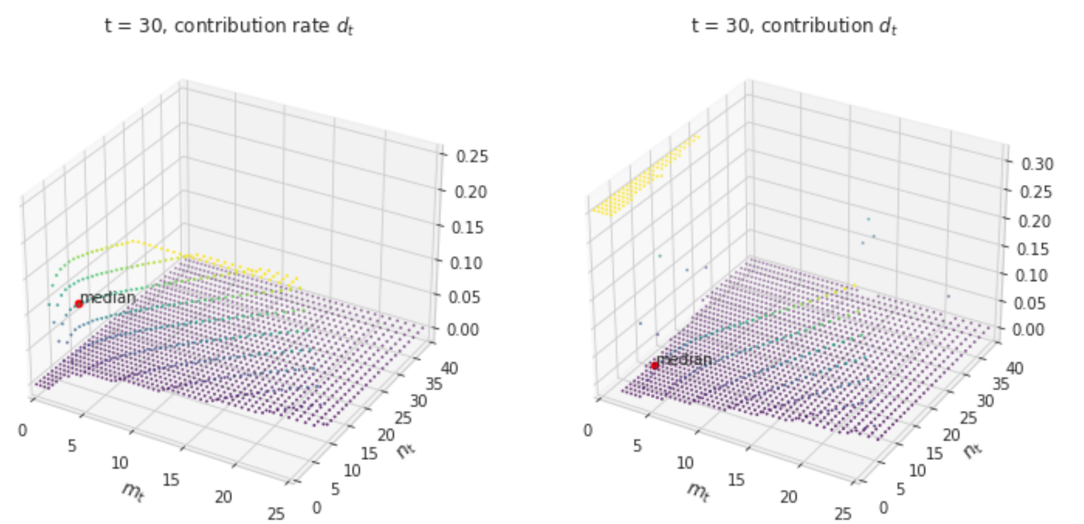


Figure 41: Median contributor at age 30, RE (left) and subjective expectation solution (right).

Rational expectations solutions overstate the share of contributors in the economy, when compared to empirical studies. On the other hand, contributor share is increasing for the extrapolative solution; as agents age they decide to participate in the DC account (Figure 43).

Quantile based comparisons show the presence of both optimism and pessimism on each side of the income distribution. In contrast to the rational expectations solution (Figure 44), workers who extrapolate start participating later (Figure 45).

Contribution rates differences vary by income quantile. In each part of the income distribution, subjective expectations capture the slow increase in contributions over the tenure, whereas the rational solution fails in this respect. This is not the case with rational workers - low income workers even decrease their contribution rates (Figure 46). In this regard, including extrapolative expectations shows that eligibility may be enough for low income workers to contribute in an auto-escalating manner. On the other hand, top-income workers who extrapolate contribute at significantly higher rates later on in work life, matching empirical patterns (Figure 47).

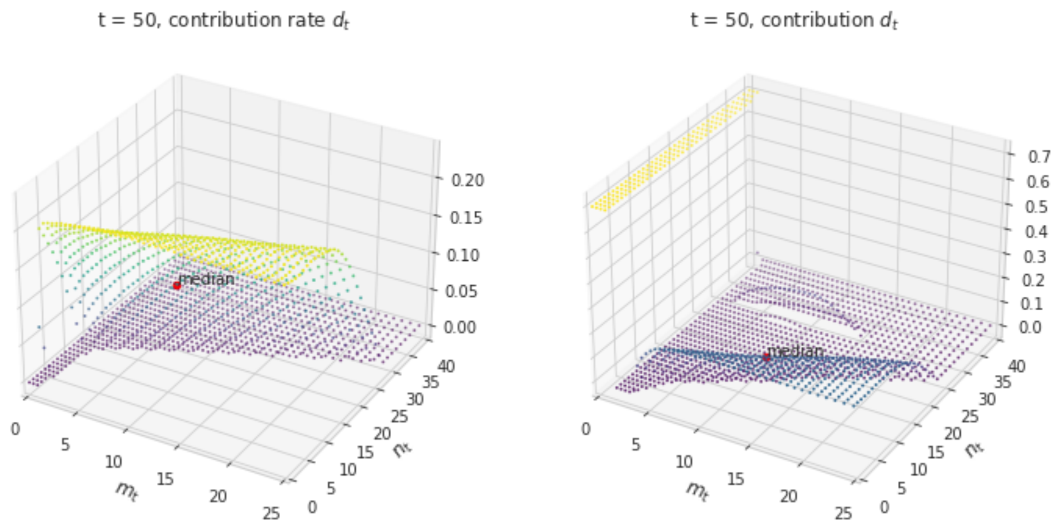


Figure 42: Median contributor at age 50, RE (left) and subjective expectation solution (right).

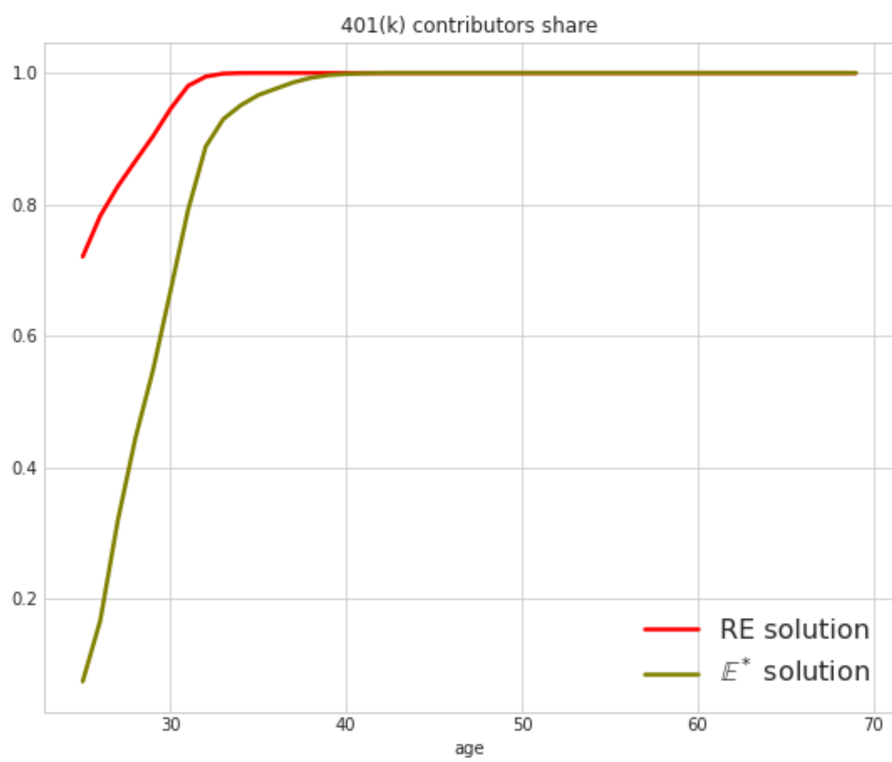


Figure 43: DC contributors share, rational (red) and extrapolative (green) expectations solution.

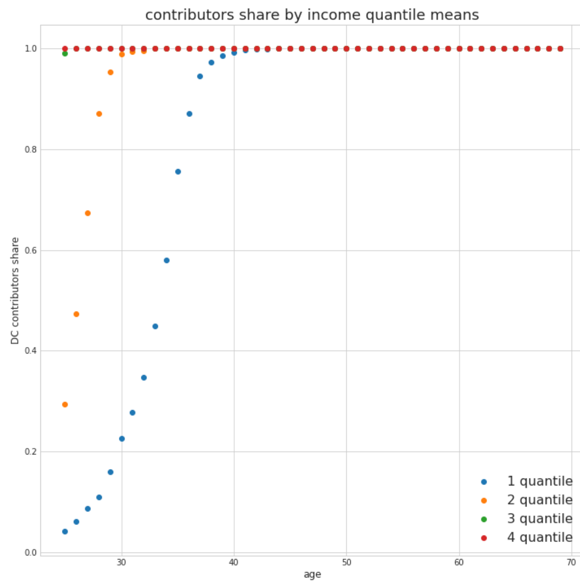


Figure 44: Rational workers, DC contributors' share by income quantile over the work life.

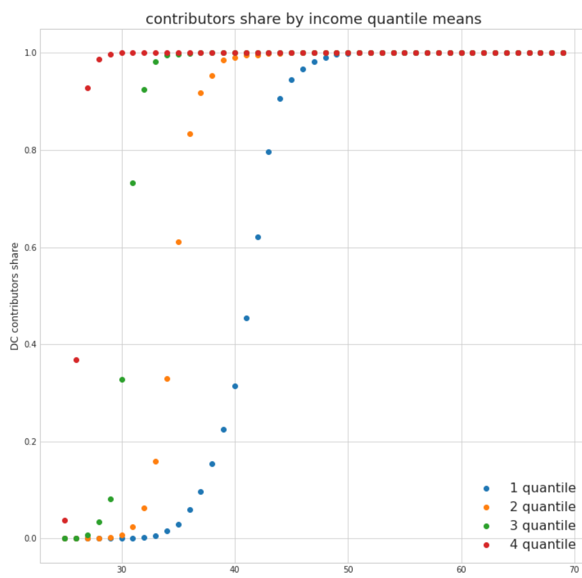


Figure 45: Biased workers, DC share lifts off gradually over the work life.

A.4 Savings ratios for the youngest and oldest workers

As previously mentioned, even though subjective expectations solution overstates liquid-to-illiquid savings ratios for the youngest cohort (Figure 48, right graph), model simulations show that the shape of the savings ratio across wage percentiles matches the empirical estimates from the SCF data. In contrast, rational expectations do not match the shape or the size (Figure 48, left graph). Moreover, just before retirement, the savings ratios of

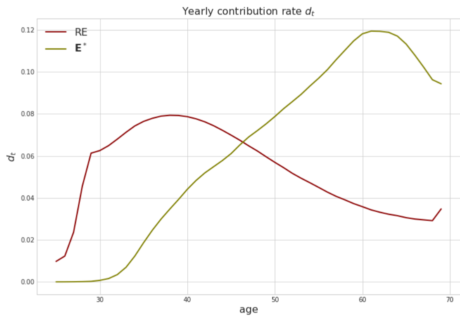


Figure 46: Contribution rates over the tenure, bottom 25%.

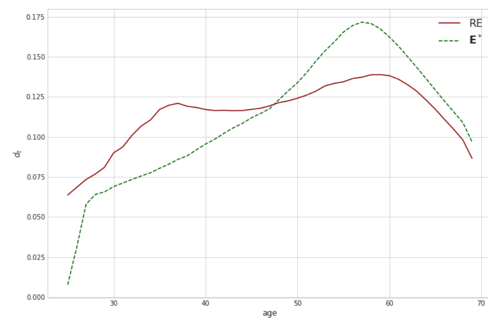


Figure 47: Contribution rates over the tenure, top 25%.

subjective workers are matched in shape, but slightly understated when compared to SCF workers (Figure 49, right graph and Figure 3 from the main part of the paper).

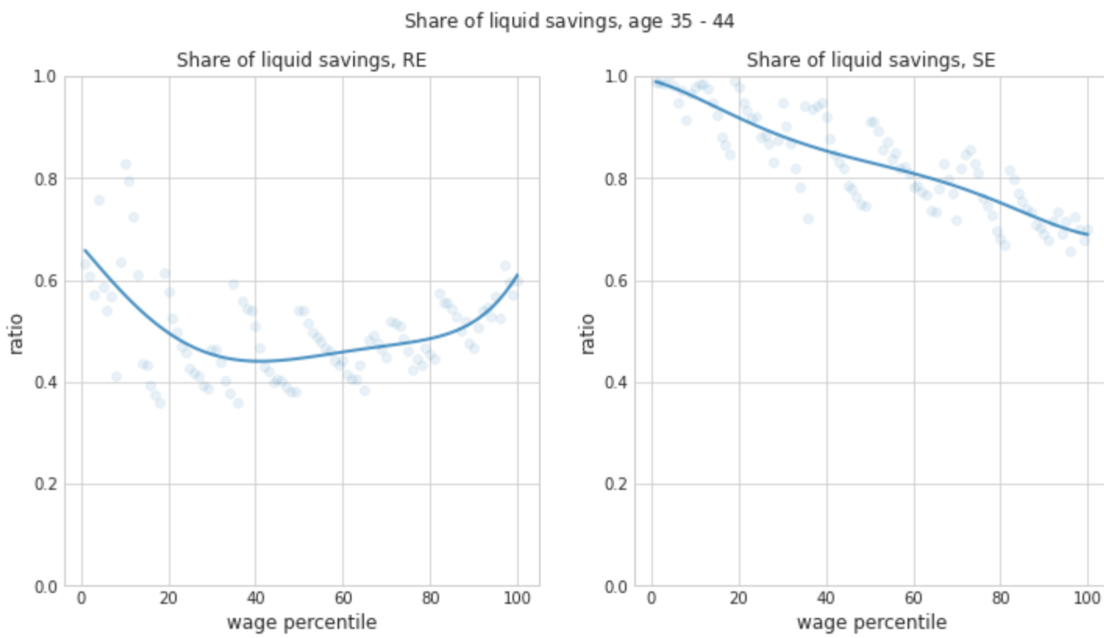


Figure 48: Savings ratios across wage percentiles, model simulations for workers aged 35-44. Rational expectations; left and subjective expectations; right.

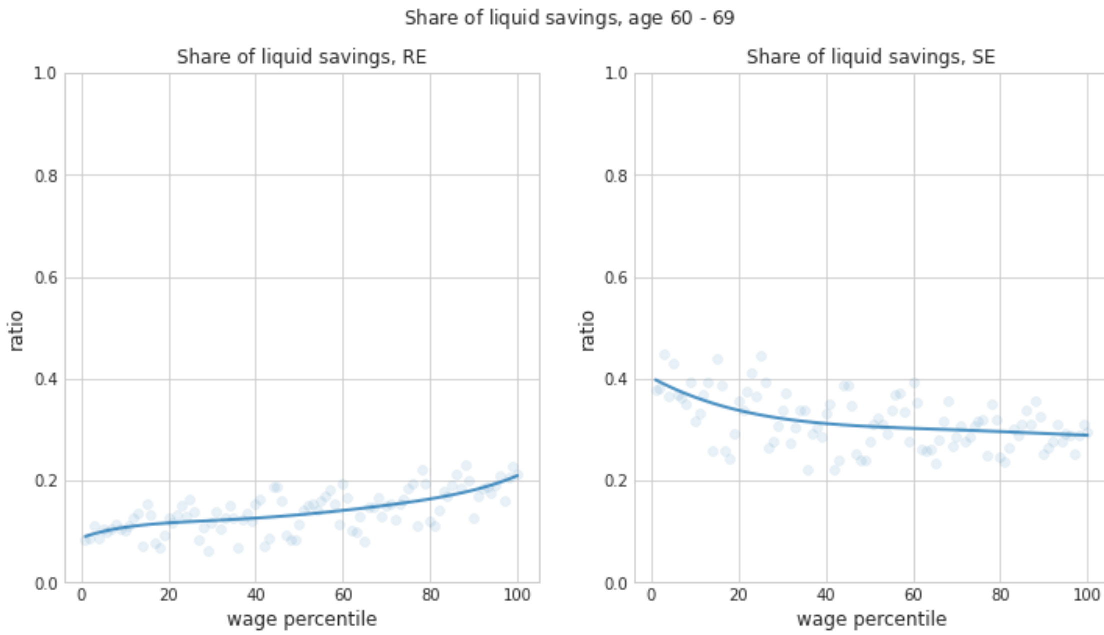


Figure 49: Savings ratios across wage percentiles, model simulations for workers aged 60-70. Rational expectations; left and subjective expectations; right.

B Mortgage Shopping Behavior in the U.S. - Stochastic Record Linkage

co-authored with Ante Šterc

B.1 Motivating Findings From SCE

Motivating findings based on the data from the U.S. Survey of Consumer Expectations. Figure 50 shows that the largest mass of non-informed households is from the lowest income group. Moreover, the figure shows that the mass of non-informed households decreases with higher income. Figure 51 shows that households from the lowest income group have the highest debt-to-income ratios. In addition, Figure 52 shows that the largest shares of the highest debt-to-income ratios are in the lowest part of the income distribution. The findings from these figures imply that most exposed households are those that are the least informed about credit possibilities.

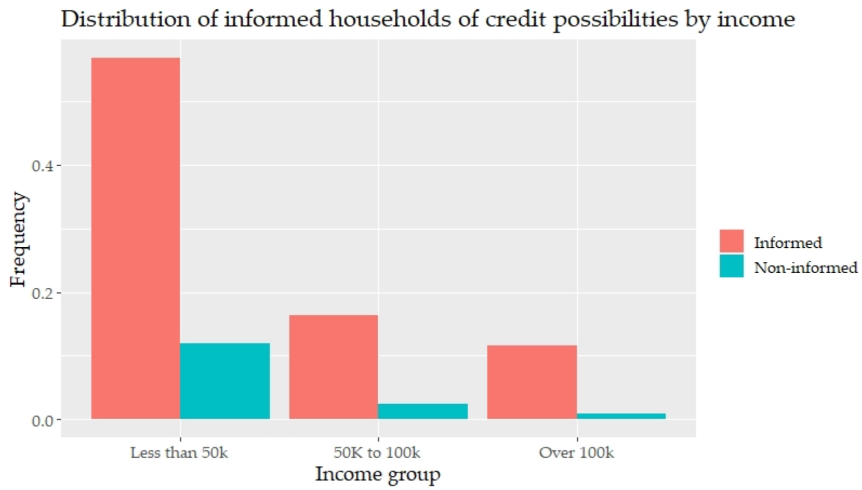


Figure 50: Share of non-informed households by income group. Source: SCE, authors' calculations.

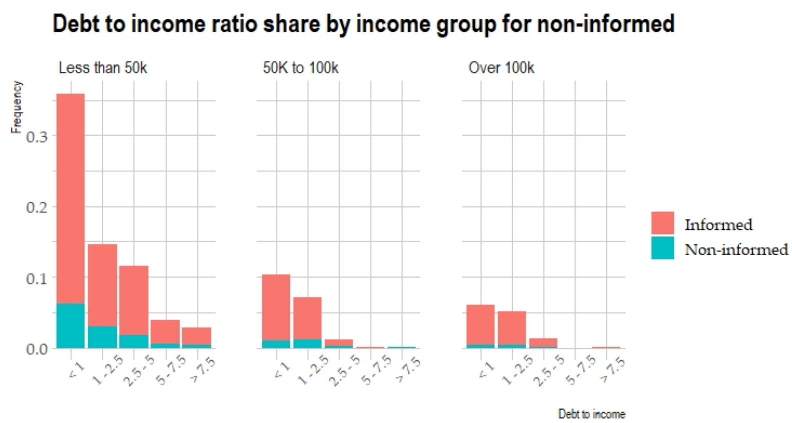


Figure 51: Share of non-informed households for each debt to income level over the income distribution. Source: SCE, authors' calculation.

B.2 The NSMO (2013-2020) analysis

The data on mortgages in the NSMO data range from 2013 to 2021, and tracks mortgages originated during the 2013-2020 period. Households were chosen at random to report the specifics of their mortgage contracts, reasons, and experiences. Details about mortgage origination, combined with demographic characteristics, allow us to estimate the effect of borrowers' characteristics on the acquired mortgage interest rate, controlling for mortgage specifics. First, we consider respondents' attitudes toward the mortgage market and their beliefs about the appropriateness of their lender selection. Second, we quantify the correlation between education and search effort variation and the mortgage rate attained at origination. Third, we extrapolate financial literacy from the Survey of Consumer Finances

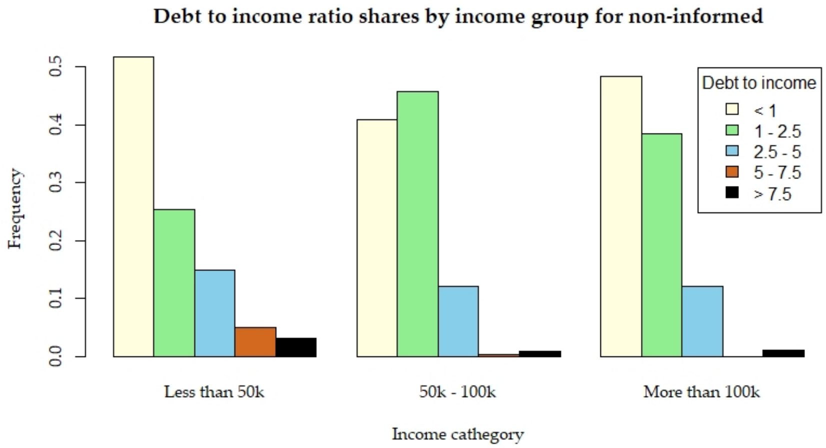


Figure 52: Debt to income ratio distributions for each income group. Source: SCE, authors' calculation.

to find a link between financial skills and the interest rate obtained after the mortgage is locked in.[40]

Interestingly, almost 70% of the borrowers believe that they would be getting the same interest rate regardless of their choice of lender. 86% initiated the contact with the lender themselves. While searching for options, 48% consider only one lender/mortgage broker. Consequently, 77% applies to only one lender. However, the number of lenders considered varies with education level (Figure 53). Borrowers who apply to multiple lenders usually do so in search of better contract terms.

When refinancing, 88% of borrowers found lower interest rates as an important reason to start the process. Moreover, 75% of these borrowers rendered lower monthly payments as equally important. In our paper, the search model conforms to the trade-offs of a homeowner and assigns lower repayments as the benefit. Figure 54 shows that almost 60 percent of high-skilled borrowers consider two or more lenders (the right histogram), which holds for the lower percentage of low-skilled borrowers (the left histogram). In the paper, we show that financial skills remain significant for search effort and that one standard deviation increase in skill leads to a four percent increase in the probability of considering more lenders.

Our latter findings suggest that education and effort simultaneously affect the mortgage interest rate. Using NSMO data only, we control for individual and loan characteristics to support our findings in the merged data set, as financial literacy exhibits a strong, but not perfect, correlation with education.

[40]Because we are the first to match the NSMO and the SCF to impute financial literacy scores in the NSMO, the imputation details are in the main part of the paper.

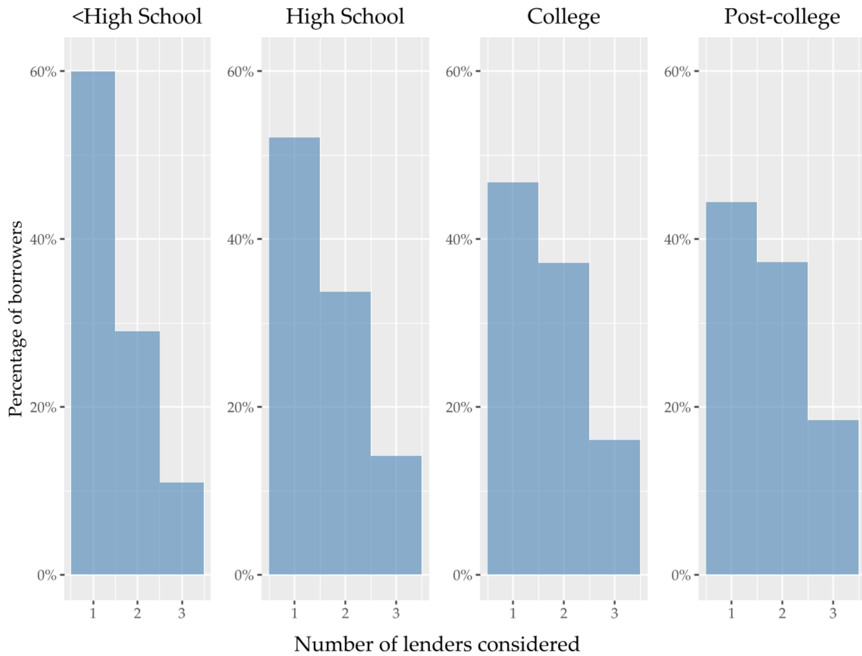


Figure 53: Number of lenders considered by education level. Source: NSMO data set, authors' calculations.

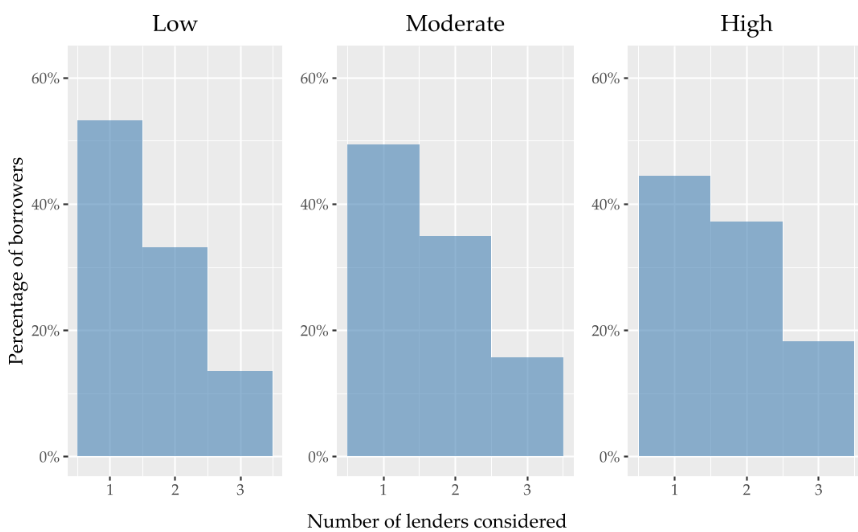


Figure 54: Number of lenders considered by financial skills tercile. Source: merged data set, authors' calculations.

B.2.1 Mortgage rate regressions

Mortgage interest rates are comprised of two components: PMMS determined by the borrower's characteristics[41] and the rate spread assigned to each borrower at origination.

[41] Freddie Mac's Primary Mortgage Market Survey (PMMS) surveys lenders each week on rates and points for their most popular 30-year fixed-rate, 15-year fixed-rate and other mortgage products.

Combining the two yields the mortgage interest rate, which is the dependent variable in the analysis.

Because nearly half of all reported mortgages are for refinancing, we estimate the linear regression separately. Both estimations control for loan-sponsorship types, guarantor enterprises (Fannie Mae, Freddie Mac, or Federal Home Loan Bank), loan amount, metropolitan (low-to-moderate) area, time effects, and the number of borrowers. The rate under refinance estimates control for non cash-out loans.

The variation in search efficacy with education is represented by interaction coefficients. Controlling for other demographic factors, we find that highly educated borrowers who shop around for loans get significantly lower interest rates. Given that we employ a novel measure that includes both cognitive and effort costs, our estimates account for an unprecedented part of the interest rate dispersion (Table 30, highlighted). All interaction coefficients are statistically significant and pass difference tests.

Model predictions allow us to calculate the present value of the difference in mortgage payments over the duration of a mortgage. We think of the payment difference as the additional costs low-educated and low-shopping behavior borrowers pay. For a 30-year loan at $200,000, high-school graduates pay on average at the 4.43% rate, whereas post-college graduates get 4.26%. The mortgage spread implies a $9900 mortgage payment difference over the duration of the mortgage. Keeping education fixed, search effort induces the mortgage spread of 8 b.p. and implies an additional $7500 in mortgage payments, on top of education differences. These estimates serve as a lower bound for mortgage payment losses in the market, as they abstract from additional correlations that substantiate search effort or mortgage process knowledge.

Our predicted rate plots (Figure 55) show that searches are most effective for highly educated borrowers as the predicted interest rate density moves to the left. On the other hand, those low-educated borrowers who search more do so due to the fear of rejection. All plots show that controlling for other characteristics still leaves the residual spread that borrowers face based on their education.

B.2.2 Education effects in mortgage search

Because the mortgage interest rate varies with search effort, we investigate borrower characteristics that affect the amount of search borrowers are willing to take on. Controlling for loan characteristics, ordered logistic model estimates show that college and post-college graduates are 50% and 65% more likely to search more (Table 31). On the other hand, women and financially inexperienced search less. Both of these characteristics are highly correlated with financial literacy in the SCF data and this strand of literature (Lusardi et al., 2010; Lusardi and Mitchell, 2014; Lusardi, 2019).

B.2.3 What agents are most likely to default on mortgage

The NSMO dataset allows us to track mortgage performance after origination. In the main part of the paper, we show that financially skilled borrowers are 50% more likely to

Table 30: Interest rate upon origination and under refinancing, explanatory characteristics, NSMO data.

	mortgage rate	
	(first origination)	(under refinancing)
Age	0.043^{***}	0.076^{***}
	(0.010)	(0.010)
Female	0.033^{***}	0.033^{***}
	(0.009)	(0.008)
Race: African-American	−0.005	0.026
	(0.019)	(0.018)
Asian	−0.020	-0.049^{***}
	(0.020)	(0.017)
Other	0.068^{***}	0.012
	(0.025)	(0.023)
Income: $30,000 - $50,000	0.008	-0.107^{***}
	(0.024)	(0.024)
$50,000 - $75,000	0.034	-0.082^{***}
	(0.023)	(0.022)
$75,000 - $100,000	0.031	-0.064^{***}
	(0.024)	(0.023)
$100,000 - $175,000	0.061^{**}	-0.063^{***}
	(0.024)	(0.023)
$175,000 or more	0.050^{*}	-0.063^{**}
	(0.026)	(0.025)
Credit Score	-0.264^{***}	-0.218^{***}
	(0.010)	(0.009)
Loan term	0.024^{***}	0.036^{***}
	(0.001)	(0.001)
Loan-to-Value ratio	0.004^{***}	0.004^{***}
	(0.0004)	(0.0003)
Number of lenders considered: 2 lenders	0.038	−0.014
	(0.030)	(0.027)
3 lenders or more	0.115^{**}	0.053
	(0.047)	(0.038)
Education: Some college	-0.037^{*}	−0.001
	(0.022)	(0.019)
college degree	-0.066^{***}	−0.024
	(0.021)	(0.019)
post-college degree	-0.079^{***}	−0.011
	(0.023)	(0.020)
Interaction: some college; considered 2	−0.028	0.005
	(0.036)	(0.033)
some college; considered 3 or more	-0.130^{**}	-0.102^{**}
	(0.055)	(0.045)
college degree; considered 2	-0.076^{**}	−0.011
	(0.034)	(0.031)
college degree; considered 3 or more	-0.177^{***}	-0.088^{**}
	(0.051)	(0.042)
post-college degree; considered 2	-0.085^{**}	-0.053^{*}
	(0.035)	(0.032)
post-college degree;considered 3 or more	-0.234^{***}	-0.131^{***}
	(0.052)	(0.043)
Constant	5.256^{***}	4.578^{***}
	(0.081)	(0.070)
Observations	21,469	21,625
R^2	0.370	0.466
Residual Std. Error	23.650 (df = 21417)	20.678 (df = 21572)
F Statistic	246.159^{***} (df = 51; 21417)	362.082^{***} (df = 52; 21572)

Note: Other regressors are stated in the text. *p<0.1; **p<0.05; ***p<0.01

Table 31: Ordered logistic regression results

	Dependent variable:	
	Number of lenders considered	
	(all originations)	(under refinancing)
Income: \$35,000-\$50,000	-0.018	-0.013
	(0.053)	(0.077)
\$50,000-\$75,0000	-0.024	-0.034
	(0.050)	(0.071)
\$75,000-\$100,000	-0.024	-0.070
	(0.051)	(0.073)
\$100,000-\$175,000	-0.054	-0.157^{**}
	(0.051)	(0.074)
\$175,000 or more	-0.090	-0.162^{**}
	(0.056)	(0.081)
Education: some college	0.267^{***}	0.263^{***}
	(0.035)	(0.049)
college degree	0.408^{***}	0.383^{***}
	(0.035)	(0.048)
post-college degree	0.501^{***}	0.431^{***}
	(0.036)	(0.051)
Female	-0.279^{***}	-0.336^{***}
	(0.019)	(0.027)
Age	-0.177^{***}	-0.040
	(0.019)	(0.030)
Have stocks	-0.097^{***}	-0.103^{***}
	(0.020)	(0.029)
Metro area, low-to-moderate income tract	0.007	-0.036
	(0.029)	(0.041)
Non-metro area	-0.053^{*}	-0.071
	(0.032)	(0.046)
Observations	43,094	21,625
Note: Controlled for time and loan amount effects.		*p<0.1; **p<0.05; ***p<0.01

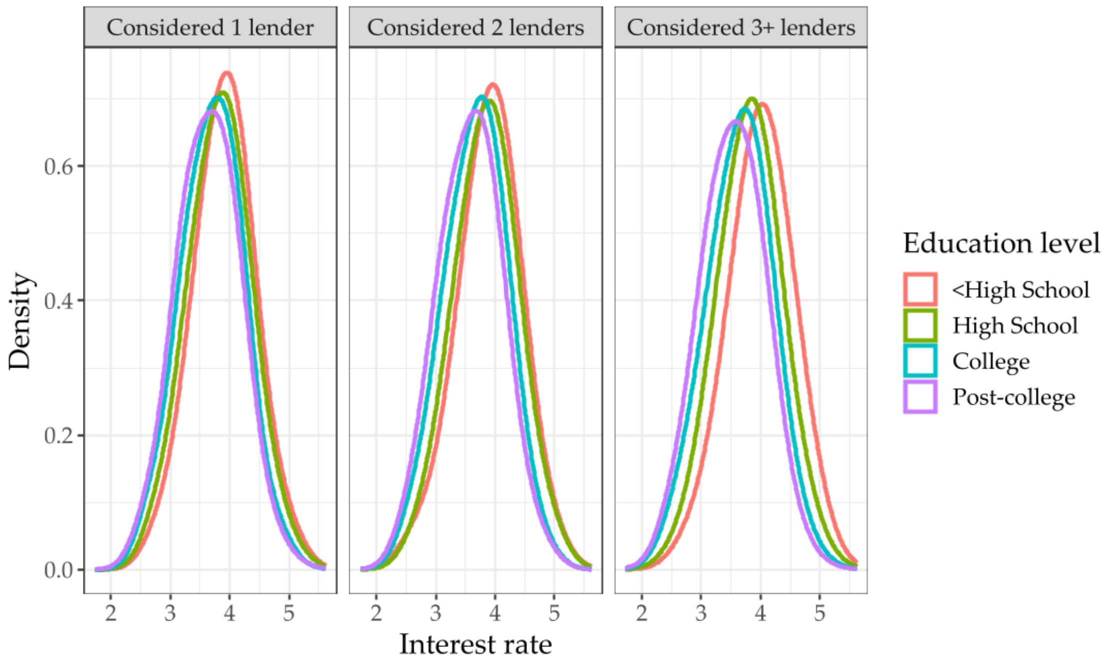
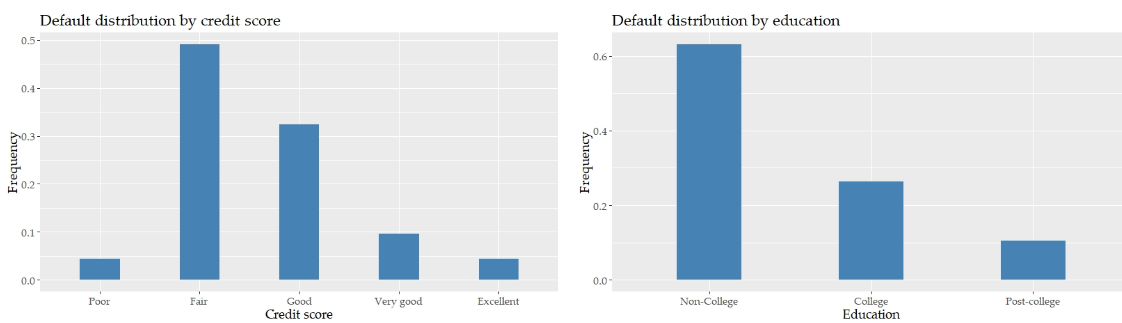


Figure 55: Predicted interest rate by education type. Each plot represents a separate case for the number of lenders considered in the mortgage process. Regression predictions, NSMO.

meet the due date of their mortgage payments. Here, we show that low-educated borrowers default more often (Figure 56b).



(a) Default by credit score.

(b) Default by education.

Figure 56: Share of households that default by credit score and education. Source: NSMO, authors' calculation.

The distributions in Figure 56 shows that households that default on a mortgage and face bankruptcy are associated with lower credit scores and lower education. The only exception is those with the lowest credit scores, but household mortgage requests with "Poor" credit scores are usually denied.

B.3 SCF data analysis

We use the Bayesian Record Linkage algorithm to impute the financial literacy score from the SCF data into the NSMO data. To begin, we examine the average financial literacy score over the lifecycle to motivate investment in, and accumulation of financial skills in the model. Figure 18 shows increasing average financial literacy scores by age groups.

The first model estimates outline correlations between financial literacy and household characteristics. Our predicted probabilities of the ordered logistic model (Table 32) suggest that high-income level households are 12% more likely to be fully financially skilled, keeping other characteristics fixed. Though education explains the largest part of financial literacy, income-based differences relate to financial skills needed to understand the mortgage refinancing process.

Next, we restrict the SCF sample to borrowers who hold a mortgage on their primary residence and estimate a binary regression model to evaluate their likelihood of refinancing. The estimates pinpoint vital characteristics that explain a household's effort in shopping for credit.

Controlling for income and mortgage size, we find significant and large effects of financial literacy - a high financial literacy score relates to a 60% greater likelihood of refinancing. In contrast, education effects are insignificant (Table 33). Our analysis supports Lusardi (2019) and highlights the relevance of the financial knowledge margin in the decision to refinance.

Using the question about the amount of shopping time allocated to borrowing options, we proxy borrower's search effort and find a 12% greater likelihood of refinancing by borrowers who allocate time to exploring borrowing options (Table 33). Further, keeping other characteristics fixed, financial knowledge, and search effort positively correlate with the decision to refinance. As a result, the mortgage search model with financial skills investment and search effort disentangles the two dimensions relevant to the decision to refinance.

Our estimates on credit shopping behavior emphasize financial skills as an important dimension of heterogeneity (Table 12). While mortgage owners shop more on average, separate analyses for mortgage owners and renters reach the same conclusion: controlling for individual characteristics, including age, income, and education, financially savvy borrowers spend more time searching for credit.

Keeping other characteristics fixed at the mean of each subsample, we plot the likelihood change over financial literacy level and monthly housing expenses. Homeowners are more likely to spend a lot more time shopping for credit than renters. Specifically, financially savvy homeowners are up to 15 p.p. more likely to allocate more time to credit shopping than low-skilled homeowners (Figure 57, left). The difference in likelihood decreases with the size of their mortgage payment. In contrast, renters allocate their time to credit shopping independently of their rent amount, and financially skilled are 10 p.p. more likely to spend a great deal of time in searching for credit (Figure 57, right).

Table 32: Financial Literacy Score, relation to observables. Source: SCF data.

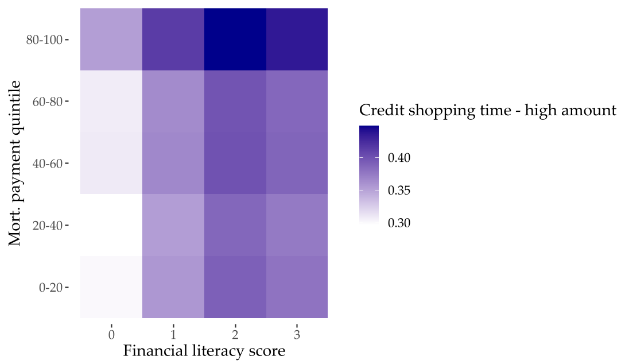
	Dependent variable: Financial literacy score
Worker	0.041^{*}
	(0.025)
Married	0.111^{***}
	(0.024)
Non-white	-0.392^{***}
	(0.019)
Female	-0.474^{***}
	(0.025)
Education: High-school	0.211^{***}
	(0.031)
Some college	0.599^{***}
	(0.031)
College degree	1.123^{***}
	(0.033)
Income percentile: 20^{th} - 40^{th}	0.049^{*}
	(0.028)
40^{th} - 60^{th} 3	0.073^{**}
	(0.031)
60^{th} - 80^{th}	0.179^{***}
	(0.035)
80^{th} - 90^{th}	0.349^{***}
	(0.043)
90^{th} - 100^{th}	0.649^{***}
	(0.048)
Pseudo R^2	0.134
Observations	60,125
Note: Controlling for age and asset amount.	*p<0.1; **p<0.05; ***p<0.01

B.3.1 Rent and mortgage payments as shares of labor income

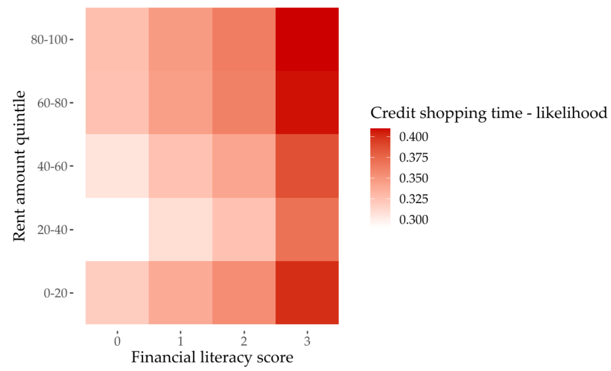
In the model calibration, we inform the rental rate κ with the share of homeowners in the SCF. When compared to an average mortgage monthly payment, rental payments are twice as high. The averages from the SCF data are computed for the subsample of workers up to age 55 with wage income higher than the yearly amount of retirement benefits. Sample averages show that monthly rental payments are up to two times higher than monthly mortgage payments.

Table 33: Binary regression estimates, likelihood to refinance, SCF data.

	Dependent variable:
	Ever refinanced their mortgage
Financial literacy score: low	0.099
	(0.104)
medium	0.252***
	(0.098)
high	0.400***
	(0.098)
Search effort, borrowing: medium	0.055
	(0.050)
high	0.110**
	(0.052)
Female	0.075
	(0.049)
non-white	−0.247***
	(0.034)
Mortgage size: \$83,000 - \$159,000	−0.148***
	(0.042)
\$159,001 - \$ 297,000	−0.285***
	(0.044)
\$ 297,001 - \$ 1,450,000	−0.304***
	(0.050)
Liquid savings: $\leq$ \$4,500	0.145***
	(0.049)
\$4,500 - \$21,000	- 0.045
	(0.050)
$\geq$\$21,000	-0.017
	(0.051)
Income percentile group: 20^{th}-40^{th}	0.242***
	(0.083)
40^{th}-60^{th}	0.260***
	(0.079)
60^{th}-80^{th}	0.482***
	(0.079)
80^{th}-90^{th}	0.874***
	(0.084)
top 10	1.047***
	(0.085)
Constant	−0.961***
	(0.145)
Pseudo R^2	0.083
Observations	22,178
Note: Controlled for age, family structure, education, and survey wave effects.	*p<0.1; **p<0.05; ***p<0.01



(a) Likelihood variation, homeowners.



(b) Likelihood variation, renters.

Figure 57: Great deal of time spent shopping for credit, SCF data. Ord. logit predictions.

Living arrangement	Financial literacy score			
	0	1	2	3
Homeowner	0.140	0.139	0.142	0.129
Renter	0.257	0.241	0.233	0.222

Table 34: First row: monthly mortgage payment as a share of income - homeowners, second row: monthly rent as a share of income; renters. SCF data, worker subsample.

B.3.2 Homeownership choice and financial literacy

Our model assumes that the homeownership choice depends on individual assets, financial skills, and productivity. As a result, the model's equilibrium generates a positive correlation between mortgage take-up and financial skills, which aligns with the similar positive association we observe in the SCF data. Table 35 presents estimates from the logistic regression, where we regress the choice to rent or own against a set of observable characteristics, including skills, assets, and wage income. To maintain consistency with our model, the estimates are derived from a subsample of workers. The first two rows in the coefficient table 35 show that the likelihood of owning a home increases with skills, with age and wage income showing the same direction. Importantly, education is non-significant and varies in the direction of the correlation. The SCF data reinstate the salience of individual skills in financial behavior and choice.

B.4 Bayesian Record Linkage method (BRL)

Recently developed in Enamorado et al. (2019), Bayesian Record Linkage (BRL) is a probabilistic approach designed to match census data. Unlike deterministic methods such as mean-imputation and cluster-based algorithms commonly used in standard imputation, BRL leverages probabilistic techniques to account for the uncertainty inherent in the merging process. The advantages of employing BRL in this context include its scalability to handle large datasets and its ability to facilitate post-merge analyses through the utilization of

Table 35: Binary regression estimates, homeownership choice, SCF data.

	Dependent variable:
	Owns a house or an apartment
Financial literacy score: medium	0.170^{***}
	(0.038)
high	0.146^{***}
	(0.039)
Education: high-school	0.067
	(0.052)
some college	-0.051
	(0.052)
college	-0.039
	(0.056)
Married	-0.852^{***}
	(0.042)
Female	0.176^{***}
	(0.044)
non-white	-0.536^{***}
	(0.029)
Leverage ratio	-0.029^{***}
	(0.003)
Willing to take risk	0.009
	(0.063)
Wage income quartile: \$ 25,800 -\$58,200	0.235^{***}
	(0.041)
\$58,200 - \$117,000	0.778^{***}
	(0.047)
≥\$117,000	1.143^{***}
	(0.061)
Constant	-1.112^{***}
	(0.064)
Observations	40,071
Note: Controlled for age, family structure, occupation category, liquid savings amount, and survey wave effects.	*p<0.1; **p<0.05; ***p<0.01

match-specific posterior weights.

In the context of Bayesian Record Linkage (BRL), the matching process assigns posterior probabilities of a match for each record pair (i, j), where i represents the records from the NSMO data ($i \in \mathcal{A}$), and j corresponds to the SCF dataset ($j \in \mathcal{B}$).

The BRL method employs pairwise comparisons for each distinct record pair (i, j) and computes the probability of a match based on the presence of a specific set of common observables denoted as K. The key assumption of the BRL method is that the set of common observables represents a relevant set of characteristics for assessing individual financial skills, including income, education, gender, age, race, occupation, family characteristics, presence of the retirement plan, and asset holdings. These are commonly found to be significant in explaining financial skills variation in the literature (Lusardi et al., 2010; Jappelli and Padula, 2017; Lusardi et al., 2017), and contribute to explaining the variance depicted in table 36. In line with other studies, table 36 highlights the importance of a deeper exploration of individual differences in financial literacy, for instance using panel data.

Table 36: Financial skill, variance decomposition across common observables, SCF data. Source: author's calculations.

	Decomposition of R^2:	
	Financial literacy	
	All households	Homeowners
Have financial assets	0.0215	0.0202
Income	0.0308	0.0289
Race	0.0160	0.0172
Sex	0.0124	0.0123
Age group	0.0062	0.0071
Employment	0.0021	0.0019
Education	0.0522	0.0568
Have retirement plan	0.0088	0.0061
Have kids	0.0032	0.0026
Asset group	0.0420	0.0421
R^2	0.1952	0.1952

Table 37 shows the population shares in SCF and NSMO for every common observable used in the matching process. To ensure consistency in the matching procedure, we impose certain restrictions on the SCF sample. Specifically, we only include homeowners who hold a first-lien mortgage, while we make no restrictions to the NSMO sample.

Table 37: Population shares in the respective samples. Source: NSMO 2013-2022 and SCF 2016-2019, authors' calculations.

	Data set	
	NSMO	SCF
income brackets	[6%, 9% , 18%, 19%, 30%, 18%]	[13%, 8%, 13% ,11%,20%, 35%]
education brackets	[1%, 10%, 5%, 20%, 35%, 29%]	[6%, 18%, 9%, 15%, 27%, 25%]
gender (Female,Male)	[44%, 55%]	[17%,83%]
age (<35,35-44,45-54,55-64,65-74,>=75)	[18%, 22%, 22%, 21%, 14% ,3%]	[8%, 14%, 20%, 26% , 20%, 12%]
race (Caucasian, African-American, other)	[84%, 6%, 10%]	[82%, 7%, 11%]
occupation (Employed, Self-employed, Retired/Student, Other)	[68%, 10%, 19% ,2%]	[47%, 26%, 25%, 2%]
has kids (Yes, No)	[64%, 36%]	[60% , 40%]
owns financial assets (Yes, No)	[57%, 43%]	[58% 42%]
retirement plan participation (Yes, No)	[86%, 14%]	[62%, 38%]
Number of observations	43,094	40,515

For each of $card(\mathcal{A})\times card(\mathcal{B})$ distinct observations, BRL defines an agreement vector $\gamma(i,j)$ of length K. The k-th element $\gamma_k(i,j)$ represents the degree of agreement corresponding to the k-th observable in the set of mutual observables[42]. Following Enamorado et al. (2019), for a given observable k, we assume the agreement degree to be discrete, with a maximum $L_k - 1$.

Based on variable k (for example, income category), $\gamma_k(i,j) = 0$ represents a no-match, whereas agreement level $\gamma_k(i,j) = L_k - 1$ corresponds to a perfect match for a pair of records (i,j). Therefore, two records from SCF and NSMO may be matching in education brackets but may differ in income levels, leading to a lower degree of agreement. The BRL takes every agreement degree into account and evaluates the posterior probability conditional on all agreement degrees for the pair. For each observation in the NSMO, we obtain the distribution of matches across the SCF sample.

BRL builds on the Fellegi-Sunter model (Fellegi and Sunter, 1969): $M_{i,j}$ denotes a

[42] Income brackets are not listed for compactness; we group income in the SCF according to brackets in the NSMO data: (<\$35,000,\$35,000-\$50,000,\$50,000-\$75,000,\$75,000-\$100,000,\$100,000-\$175,000, >\$175,000). Similarly, we take the highest education grade data in the SCF and group them according to education brackets in the NSMO: (Some schooling, High-School graduate, Technical School, Some College, College degree, Post-college degree).

latent mixing variable that shows whether distinct records pair (i, j) form a match or not. That is, M_{ij} is Bernoulli-distributed

$$M_{i,j} \overset{i.i.d.}{\sim} \mathrm{B}(\lambda),$$

and k-based agreement level $\gamma_k(i,j)$ has a discrete distribution

$$\gamma_k(i,j)|M_{i,j} \sim \begin{pmatrix} 0 & 1 & \dots & L_k - 1 \\ \pi_{k0} & \pi_{k1} & \dots & \pi_{kL_k-1} \end{pmatrix},$$

where $\pi_{kl}, \quad l \in \{0, \dots, L_k - 1\}$ represents the probability of each agreement degree for the pair (i, j). The vector of probabilities is denoted with $\pi\mathbf{0}_{km}$.

The BRL method relies on two key independence assumptions. First, every record i can be matched to multiple records j, and second, conditional on the match $M_{i,j}$, the method assumes conditional independence of the characteristics in the set of common observables that define the match probability.

Record matching probabilities imply the observed-data likelihood $\mathcal{L}_{\text{obs}}$, that we estimate later using the Expectation-Maximization algorithm (suggested by Enamorado et al. (2019)). Using the matched records from the NSMO and SCF data, we apply the Bayesian posteriors $\epsilon_{i,j} = \mathbb{P}(M_{ij} = 1|\gamma(i,j))$ as weights for statistical inference when we use the (imputed) financial literacy score. This way, we incorporate the match procedure uncertainty and avoid biases that emerge in standard deterministic methods.

Bayes rule implies the probability of a match which defines the post-merge weight

$$\begin{aligned}
\varepsilon_{ij} &= \mathbb{P}(M_{ij} = 1 \mid \gamma(i,j)) \\
&= \frac{\lambda \prod_{k=1}^{K} (\prod_{l=0}^{L_k-1} \pi\mathbf{0}_{k1l}^{\mathbf{1}_{\{\gamma_k(i,j)=l\}}})}{\sum_{m=0}^{1} \lambda^m (1-\lambda)^{1-m} \prod_{k=0}^{K} (\prod_{l=0}^{L_k-1} \pi\mathbf{0}_{kml}^{\mathbf{1}_{\{\gamma_k(i,j)=l\}}})},
\end{aligned}$$

that we use later for statistical inference. Financial literacy for the borrower i, $\bar{Z}_i$ is the sum of literacy scores of the respective record matches in the SCF Z_j, with corresponding weights ε_{ij}[43]:

$$\bar{Z}_i = \frac{\sum_{j=1}^{N_{\mathcal{B}}} \varepsilon_{ij} Z_j}{\sum_{j=1}^{N_{\mathcal{B}}} \varepsilon_{ij}}.$$

The post-merge analysis includes $\bar{Z}_i$ as the independent variable in linear model estimates.

Non-linear models, such as the ordered logistic and binary regression models we use for inference, need to be adjusted with the posterior weight. Therefore, the maximum likelihood function includes all the record pair matches with the corresponding Bayesian weight. With the assumption $Y_i|X_i, Z_i^* \overset{indep.}{\sim} P_\theta(Y_i|X_i, Z_i^*)$, the ML estimator

$$\hat{\theta} = \sum_{i=1}^{\mathcal{N}_A} \sum_{j=1}^{\mathcal{N}_B} \varepsilon_{ij}^* \log P_\theta(Y_i|X_i, Z = Z_j^*), \quad \varepsilon_{ij}^* = \frac{\varepsilon_{ij}}{\sum_{j=1}^{\mathcal{N}_B} \varepsilon_{ij}}$$

[43]Our merging procedure uses the standardized literacy score.

is consistent and asymptotically normal and hence follows standard rules of significance tests. We use these theoretical results derived in Enamorado et al. (2019) and implement our estimators that ensure solid statistical properties.

B.4.1 Number of lenders considered

For every record pair (i, j) with a corresponding match weight ε^*_{ij}, the likelihood of number of lenders considered num_cons is characterized using the borrower's observables $(X_i, \text{fin\_skills}_i)$

$$\mathbb{P}(\text{num\_cons}_{ij} = k) = p_{ij,k} = \mathbb{P}(-\kappa_{k-1} < \beta X_i + \beta^f \text{fin\_skills}_j + u_{ij,k} < \kappa_k), \quad k \in \{1, 2, 3+\},$$

with κ_{k-1} and κ_k representing latent thresholds that define the search effort level. The logistic model assumes

$$p_{ij,k} = \frac{1}{1 + \exp\big(- \kappa_k + \beta X_i + \beta^f \text{fin\_skills}_i\big)} - \frac{1}{1 + \exp\big(- \kappa_{k-1} + \beta X_i + \beta^f \text{fin\_skills}_i\big)},$$

which pins down the log-likelihood adjusted by the posterior match weight

$$\ln L = \sum_{i=1}^{\mathcal{N}_A} \sum_{j=1}^{\mathcal{N}_B} \varepsilon^*_{ij} \sum_{k=1}^{3+} \mathbf{1}_{\{\text{num\_cons}_{ij} = k\}} \ln(p_{ij,k} | X_i, \text{fin\_skills}_j).$$

B.4.2 Additional NSMO+ estimates

As an additional counterfactual exercise, we estimate the linear probability model where the dependent variable is the number of lenders considered with our new NSMO+ dataset. We estimate the model when the number of lenders considered equals one versus more than one. Estimates are presented in Table (38). The results imply a strong positive correlation between higher financial skills and the probability of considering more than one lender when searching for a mortgage. In particular, the model predicts that an average borrower who answered zero questions correctly has a probability of considering more than one lender equal to 0.381. On the other hand, for an average financially savvy borrower who answered all questions correctly, our linear probability model predicts a 0.546 probability of considering more than one lender. The model predicts similar probabilities of considering more than one lender for average borrowers upon refinancing the mortgage.

Table 38: Linear probability model for the number of lenders considered one vs. more. Source: NSMO+, own calculation.

	Lenders considered	
	All origination	Refinancing
Age	-0.042^{***}	-0.019^{**}
	(0.005)	(0.008)
Credit Score	0.009^{*}	0.005
	(0.005)	(0.007)
Married	0.020^{***}	0.014
	(0.006)	(0.010)
Female	-0.058^{***}	-0.076^{***}
	(0.005)	(0.007)
Race: Black or African-American	0.055^{***}	0.038^{**}
	(0.011)	(0.015)
Asian	0.055^{***}	0.055^{***}
	(0.010)	(0.014)
other (including hispanic)	0.059^{***}	0.083^{***}
	(0.014)	(0.020)
Financial Literacy	0.164^{***}	0.166^{***}
	(0.038)	(0.056)
Education: high school	0.056^{***}	0.052^{***}
	(0.009)	(0.013)
college graduate	0.090^{***}	0.075^{***}
	(0.009)	(0.013)
post-college graduate	0.107^{***}	0.086^{***}
	(0.010)	(0.014)
Loan amount: \$50,000 - \$99,999	0.019	0.066^{**}
	(0.019)	(0.029)
\$100,000 - \$149,999	0.037^{*}	0.130^{***}
	(0.019)	(0.029)
\$150,000 -\$199,999	0.047^{**}	0.154^{***}
	(0.020)	(0.029)
\$200,000 - \$249,999	0.066^{***}	0.152^{***}
	(0.020)	(0.030)
\$250,000 to \$299,999	0.071^{***}	0.167^{***}
	(0.021)	(0.031)
\$300,000 -\$349,999	0.071^{***}	0.180^{***}
	(0.021)	(0.032)
\$350,000 - \$399,999	0.088^{***}	0.182^{***}
	(0.022)	(0.033)
≥\$400,000	0.099^{***}	0.176^{***}
	(0.021)	(0.031)
Constant	0.271^{***}	0.246^{***}
	(0.047)	(0.068)
Observations	43,084	21,623
R^2	0.024	0.025
Adjusted R^2	0.023	0.023
Residual Std. Error	17.837 (df = 43039)	17.676 (df = 21578)
F Statistic	23.681^{***} (df = 44; 43039)	12.666^{***} (df = 44; 21578)

Note: *p<0.1; **p<0.05; ***p<0.01
Controlled for: Loan type, Year, Government Sponsored Enterprise, Term, LTV, Number of borrowers, and Income.

C Financial Skills and Search in the Mortgage Market

C.1 Bellman Equation Derivation

This section outlines the agent's problem in discrete time with period size equal to Δt, generalizes into continuous time, and derives first order conditions of the homeowner's and renter's problem.

Let V^R and V^H represent the renter's and homeowner's values, respectively. Each period, the renter faces a productivity shock, invests in financial skills, accumulates assets, and may choose to take up a mortgage and, if yes, decides how much to search. If the renter chooses to take up a mortgage, they become homeowners, obtaining value V^H. State variables of the renter's problem are financial skills f, liquid assets a, and productivity z:

$$V^R(f,a,z) = \max_{\{i,s,c\}} \left\{ [u(c) - c^f(i,z) - c^s(s,f)]\Delta t + \frac{1}{1+\rho \Delta t}\mathbb{E}V^{R+} \right\}, \tag{15}$$

where $c^s(s,f)$ represents the cost of searching for a mortgage and $\mathbb{E}V^{R+}$ is the expected next period value, comprised of three transitions:

$$\begin{aligned}
\mathbb{E}V^{R+} = & \underbrace{\left(1 - \lambda s \phi \Delta t - \Delta t \sum_{z'} \omega(z,z')\right) V^R(f+\Delta f, a+\Delta a, z)}_{\text{no change in } z\text{, no mortgage offers}} \\
& + \underbrace{\phi \lambda s \Delta t \int_{\underline{r}}^{\bar{r}} \max\left\{ V^H(f+\Delta f, a+\Delta a, z, r'), V^R(f+\Delta f, a+\Delta a, z) \right\} d\Phi(r')}_{\text{renter searches for an offer, decides based on the interest rate offered}} \\
& + \underbrace{\Delta t \sum_{z'} \omega(z,z') V^R(f+\Delta f, a+\delta a, z') + o(t)}_{\text{gets a productivity shock, does not search}},
\end{aligned}$$

where the decision to become a homeowner depends on the search intensity s, the mortgage interest rate r, accumulated assets $a + \Delta t$, and skills $f + \Delta f$.

Using

$$-\lambda s \phi \Delta t V^R(f+\Delta f, a+\Delta a, z) = -\lambda s \phi \Delta t V^R(f+\Delta f, a+\Delta a, z) \underbrace{\int_{\underline{r}}^{\bar{r}} d\Phi(r')}_{=1}$$

and

$$\max\left\{V^H, V^R\right\} - V^R = \max\left\{V^H - V^R, 0\right\}$$

and rearranging yields

$$\begin{aligned}
\mathbb{E}V^{R+} &= V^R(f+\Delta f, a+\Delta a, z)+ \\
&+ \phi\lambda s\Delta t\int_{\underline{r}}^{\bar{r}} \max\Big\{V^H(f+\Delta f, a+\Delta a, r') - V^R(f+\Delta f, a+\Delta a, z), 0\Big\} d\Phi(r') \\
&= \Delta t\sum_{z'}\omega(z,z')[V^R(f+\Delta f, a+\Delta a, z') - V^R(f+\Delta f, a+\Delta a, z)] + o(t).
\end{aligned}$$

Multiplying the value function 15 with $(1+\rho\Delta t)$ and denoting $u_R = u(c) - c^f(i,f) - c^s(s,f)$ yields

$$V^R(f,a,z)(1+\rho\Delta t) = \max_{\{i,s,c\}}\Big\{u_R\Delta t + \mathbb{E}V^{R+}\Big\}.$$

Plugging in for $\mathbb{E}V^{R+}$ and rearranging yields

$$\begin{aligned}
\rho\Delta t V^R(f,a,z) &= \max_{\{i,s,c\}}\Big\{u_R(1+\rho\Delta t)\Delta t + V^R(f+\Delta f, a+\Delta a, z) \\
&+ \phi\lambda s\Delta t\int_{\underline{r}}^{\bar{r}} \max\Big\{V^H(f+\Delta f, a+\Delta a, z, r') - V^R(f+\Delta f, a+\Delta a, z), 0\Big\} d\Phi(r') \\
&+ \Delta t\sum_{z'}\omega(z,z')[V^R(f+\Delta f, a+\Delta a, z') - V^R(f+\Delta f, a+\Delta a, z)] + o(t)\Big\},
\end{aligned}$$

and dividing by Δt to derive the limit:

$$\begin{aligned}
\rho V^R(f,a,z) = \max_{\{c,s,i\}}\Big\{&u_R(1+\rho\Delta t) + \frac{V^R(f+\Delta f, a+\Delta a, z) - V^R(f,a,z)}{\Delta t} \\
&+ \phi\lambda s\int_{\underline{r}}^{\bar{r}} \max\{V^H(f,a,z,r') - V^R(f,a,z), 0\} d\Phi(r') \\
&+ \sum_{z'}\omega(z,z')[V^R(f,a,z') - V^R(f,a,z)] + \frac{o(t)}{\Delta t}\Big\}.
\end{aligned}$$

Finally, we let $\Delta t \to 0$ and obtain the continuous version of the renter's Bellman equation:

$$\begin{aligned}
\rho V^R(f,a,z) = \max_{\{c,s,i\}}\Big\{&u_R + \frac{\partial V^R(f,a,z)}{\partial f}\dot{f} + \frac{\partial V^R(f,a,z)}{\partial a}\dot{a} && (16) \\
&+ \phi\lambda s\int_{\underline{r}}^{\bar{r}} \max\{V^H(f,a,z,r') - V^R(f,a,z), 0\} d\Phi(r') && (17) \\
&+ \sum_{z'}\omega(z,z')[V^R(f,a,z') - V^R(f,a,z)]\Big\}. && (18)
\end{aligned}$$

Deriving the continuous version of the Bellman equation for the homeowner follows the same approach. However, initial (discrete) value functions are different because:

1. Homeowners may search for refinancing options to ensure their liquidity
2. Homeowners may face financial shocks, after which they lose their house and become renters.

As for the renter's value function derivation, we start from the value function in discrete time:

$$V^H(f,a,z,r) = \max_{\{i,s,c\}} \left\{ \big[u(c) - c^f(i,z) - c^s(s,f)\big]\Delta t + \frac{1}{1+\rho\Delta t}\mathbb{E}V^{H+} \right\}, \tag{19}$$

where $c^s(s,f)$ represent the cost of searching for refinancing opportunities. Similar to the renter's case, the continuation value $\mathbb{E}V^{H+}$ for the homeowner $V^H(f,a,z,r)$ is comprised of disjoint transition possibilities

$$\begin{aligned}
\mathbb{E}V^{H+} &= \underbrace{(1-\lambda s\Delta t - p(f,a)\Delta t - \sum_{z'}\omega(z,z'))V^H(f+\Delta f, a+\Delta a, z, r)+}_{\text{no refinancing, no change in productivity}} \\
&\quad \underbrace{+\lambda s\Delta t \int_{\underline{r}}^{\bar{r}} \max\Big\{V^H(f+\Delta f, a+\Delta a - c_{\text{ref}}, z, r'), V^H(f+\Delta f, a+\Delta a, z, r)\Big\} d\Phi(r')}_{\text{searches for refinancing options and refinances if it yields higher value}} \\
&\quad \underbrace{+p(f,a)\Delta t V^R(f+\Delta f, a+\Delta a, z)}_{\text{loses the house, goes back to renting}} \\
&\quad + \Delta t \sum_{z'} \omega(z,z') V^H(f+\Delta f, a+\Delta a, z) + o(t).
\end{aligned}$$

Rearranging the expression implies

$$\begin{aligned}
\mathbb{E}V^{H+} &= V^H(f+\Delta f, a+\Delta a, z, r)+ \\
&+ \lambda s\Delta t \int_{\underline{r}}^{\bar{r}} \max\Big\{V^H(f+\Delta f, a+\Delta a - c_{\text{ref}}, z, r') - V^H(f+\Delta f, a+\Delta a, z, r), 0\Big\} d\Phi(r') \\
&+ \Delta t p(f,a)\big[V^R(f+\Delta f, a+\Delta a, z) - V^H(f+\Delta f, a+\Delta a, z, r)\big] \\
&+ \Delta t \sum_{z'} \omega(z,z')\big[V^H(f+\Delta f, a+\Delta a, z', r) - V^H(f+\Delta f, a+\Delta a, z, r)\big] + o(t),
\end{aligned}$$

and if we go back to the discrete value function (19) and multiply it by $(1+\rho\Delta t)$ and substitute for $\mathbb{E}V^{H+}$ and $u_H = u(c) - c^s(s,f) - c^f(i,f)$, (19) boils down to

$$\begin{aligned}
\rho\Delta t V^H(f,a,z,r) = \max_{\{c,s,i\}} \Big\{ & u_H(1+\rho\Delta t) + V^H(f+\Delta f, a+\Delta a, z, r) - V^H(f,a,z,r) \\
&+ \lambda s\Delta t \int_{\underline{r}}^{\bar{r}} \max\Big\{V^H(f+\Delta f, a+\Delta a - c_{\text{ref}}, z, r') - V^H(f+\Delta f, a+\Delta a, z, r), 0\Big\} d\Phi(r') \\
&+ \Delta t p(f,a)\big[V^R(f+\Delta f, a+\Delta a, z) - V^H(f+\Delta f, a+\Delta a, z, r)\big] \\
&+ \Delta t \sum_{z'} \omega(z,z')\big[V^H(f+\Delta f, a+\Delta a, z', r) - V^H(f+\Delta f, a+\Delta a, z, r)\big] + o(t) \Big\},
\end{aligned}$$

dividing by $\varDelta t$ and letting $\varDelta t \to 0$ yields

$$\rho V^H(f,a,z,r) = \max_{\{c,s,i\}}\Bigg\{u_H(1+\rho\varDelta t) \xrightarrow{\varDelta t\to 0} u_H \tag{20}$$

$$+ \frac{V^H(f+\varDelta f, a+\varDelta a, z, r) - V^H(f,a,z,r)}{\varDelta t} \xrightarrow{\varDelta t\to 0} \frac{\partial V^H(f,a,z,r)}{\partial f}\dot{f} + \frac{\partial V^H(f,a,z,r)}{\partial a}\dot{a} \tag{21}$$

$$+ \lambda s \int_{\underline{r}}^{\bar{r}} \max\Bigg\{V(f+\varDelta f \xrightarrow{\varDelta t\to 0} f, a+\varDelta a - c_{\text{ref}} \xrightarrow{\varDelta t\to 0} a - c_{\text{ref}}, z, r'), \tag{22}$$

$$V^H(f+\varDelta f \xrightarrow{\varDelta t\to 0} f, a+\varDelta a \xrightarrow{\varDelta t\to 0} a, z, r)\Bigg\} d\varPhi(r') \tag{23}$$

$$+ p(f,a)[V^R(f+\varDelta f \xrightarrow{\varDelta t\to 0} f, a+\varDelta a \xrightarrow{\varDelta t\to 0} a, z) - \tag{24}$$

$$V^H(f+\varDelta f \xrightarrow{\varDelta t\to 0} f, a+\varDelta a \xrightarrow{\varDelta t\to 0} a, z, r)] \tag{25}$$

$$+ \sum_{z'} \omega(z,z')[V^H(f+\varDelta f \xrightarrow{\varDelta t\to 0} f, a+\varDelta a \xrightarrow{\varDelta t\to 0} a, z', r) - \tag{26}$$

$$V^H(f+\varDelta f \xrightarrow{\varDelta t\to 0} f, a+\varDelta a \xrightarrow{\varDelta t\to 0} a, z, r)] \tag{27}$$

$$+ \frac{\mathcal{O}(t)}{\varDelta t} \xrightarrow{\varDelta t\to 0} 0\Bigg\}. \tag{28}$$

In the final step, we derive the continuous version of the budget constraint and financial skill accumulation. Again, we start from the discrete version and build up towards the expression suitable for division by $\varDelta t$ and letting $\varDelta t \to 0$. The renter's budget constraint translates to

$$\begin{aligned} a_{t+\varDelta t} &= (1+R\varDelta t)a_t + [wz_t - \kappa - c_t]\varDelta t \\ a_{t+\varDelta t} - a_t &= \varDelta t[Ra_t + wz_t - \kappa - c_t]/\varDelta t \\ \frac{a_{t+\varDelta t} - a_t}{\varDelta t} &= Ra_t + wz_t - \kappa - c_t \xrightarrow{\varDelta t\to 0} \dot{a} = Ra_t + wz_t - \kappa - c_t. \end{aligned} \tag{29}$$

The homeowner's budget constraint differs due to the mortgage repayment and boils down to

$$\dot{a}^H = Ra_t^H + wz_t - rM - c_t^H. \tag{30}$$

The financial skill accumulation process satisfies

$$\begin{aligned} f_{t+\varDelta t} &= (1-\delta\varDelta t)f_t + \frac{\mu}{\eta}(i_t f_t)^\eta \varDelta t \\ f_{t+\varDelta t} - f_t &= [\frac{\mu}{\eta}(i_t f_t)^\eta - \delta f_t]\varDelta t / : \varDelta t \\ \frac{f_{t+\varDelta t} - f_t}{\varDelta t} &= [\frac{\mu}{\eta}(i_t f_t)^\eta - \delta f_t] \\ \dot{f} &= \frac{\mu}{\eta}(i_t f_t)^\eta - \delta f_t. \end{aligned} \tag{31}$$

First order conditions

The full version of the continuous time problem permits us to take first order conditions to infer more about the search intensity and consumption elasticity in the model.

The renter's problem (18) under the budget constraint (29) and financial skill accumulation (31) satisfies

$$[i] \quad \frac{\partial c^f(i,f)}{\partial i} = \frac{\partial V^R(f,a,z)}{\partial f}\mu(if)^{\eta-1}f = \frac{V^R(f,a,z)}{\partial f}\mu f^{\eta} i^{\eta-1},$$

which after plugging in for $c^f(i,f) = i_0 i^{\overline{1+\gamma_i}} \frac{1}{1+f}$ yields

$$i^{\frac{1}{\gamma_i}-(\eta-1)} = \frac{\partial V^R}{\partial f}\mu f^{\eta}\frac{1+z}{i_0} \Big/ ()^{\overline{\frac{1}{\gamma_i}-(\eta-1)}}$$

$$i^* = \left[\frac{\partial V^R(f,a,z)}{\partial f}\frac{\mu f^{\eta}(1+z)}{i_0}\right]^{\overline{\frac{1}{\gamma_i}-(\eta-1)}}$$

and

$$[c] \quad u'(c) = \frac{\partial V^R(f,a,z)}{\partial a}$$

$$[s] \quad \frac{c^s(s,f)}{\partial s} = \phi\lambda \int_{\underline{r}}^{\bar{r}} \max\Big\{V^H(f,a,z,r') - V^R(f,a,z), 0\Big\} d\Phi(r') \tag{32}$$

where substituting for $c^s(s,f) = s_0 s^{1+\frac{1}{\gamma_s}} \frac{1}{1+f}$ yields

$$s^* = \left[\phi\lambda \int_{\underline{r}}^{\bar{r}} \max\Big\{V^H(f,a,z,r') - V^R(f,a,z), 0\Big\} d\Phi(r') \frac{1+f}{s_0}\right]^{\gamma_s}. \tag{33}$$

Under (31) and the budget constraint (30), first order conditions for the homeowner's problem 28 include comparing values between staying at the current mortgage rate or refinancing

$$i_H^* = \left[\frac{\partial V^H(f,a,z,r)}{\partial f}\frac{\mu f^{\eta}(1+z)}{i_0}\right]^{\overline{\frac{1}{\gamma_i}-(\eta-1)}} \tag{34}$$

$$u'(c_H^*) = \frac{\partial V^H(f,a,z,r)}{\partial a} \tag{35}$$

$$s_H^* = \left[\lambda \int_{\underline{r}}^{\bar{r}} \max\Big\{V^H(f,a-c_{\text{ref}},z,r') - V^H(f,a,z,r), 0\Big\} d\Phi(r') \frac{1+f}{s_0}\right]^{\gamma_s}. \tag{36}$$

$$
\begin{aligned}
\text{[renter]} \quad \frac{\partial c(s,f)^m}{\partial s} &= \lambda\phi \int_{\underline{r}}^{\overline{r}} \max\{V^H(f,a,z,r') - V^R(f,a,z), 0\} d\Phi(r'), \\
\frac{\partial c(i,z)^f}{\partial i} &= \frac{\partial V^R(f,a,z)}{\partial f} \frac{\partial \dot{f}}{\partial i}, \\
u'(c) &= \frac{\partial V^R(f,a,z)}{\partial a}, \\
\text{[homeowner]} \quad \frac{\partial c(s,f)^m}{\partial s} &= \lambda \int_{\underline{r}}^{\overline{r}} \max\{V^H(f,a-c_{\text{ref}},z,r') - V^H(f,a,z,r), 0\} d\Phi(r'), \\
\frac{\partial c(i,z)^f}{\partial i} &= \frac{\partial V^H(f,a,z,r)}{\partial f} \frac{\partial \dot{f}}{\partial i}, \\
u'(c) &= \frac{\partial V^H(f,a,z,r)}{\partial a}
\end{aligned}
$$

which ultimately yields

$$
\begin{aligned}
\text{[renter]} \quad s &= \left(\frac{1+f}{c_0} \lambda\phi \int_{\underline{r}}^{\overline{r}} \max\{V^H(f,a,z,r') - V^R(f,a,z), 0\} d\Phi(r') \right)^{\gamma_s}, \\
i &= \left(\frac{1+z}{i_0} \frac{\partial V^R(f,a,z)}{\partial f} \mu f^\eta \right)^{\frac{1}{\frac{1}{\gamma_i} - (\eta - 1)}} \\
c &= \left(\frac{\partial V^R(f,a,z)}{\partial a} \right)^{-\frac{1}{\sigma}}, \\
\text{[homeowner]} \quad s &= \left(\frac{1+f}{c_0} \lambda \int_{\underline{r}}^{\overline{r}} \max\{V^H(f,a-c_{\text{ref}},z,r') - V^H(f,a,z,r), 0\} d\Phi(r') \right)^{\gamma_s}, \\
i &= \left(\frac{1+z}{i_0} \frac{\partial V^H(f,a,z,r)}{\partial f} \mu f^\eta \right)^{\frac{1}{\frac{1}{\gamma_i} - (\eta - 1)}} \\
c &= \left(\frac{\partial V^H(f,a,z,r)}{\partial a} \right)^{-\frac{1}{\sigma}}.
\end{aligned}
$$

Boundary Conditions

Both the renter's and homeowner's problems are subject to the budget constraint $a_t \geq 0$. The constraint, using the derived first order conditions, translates to boundary conditions

$$
u'(c) \leq \frac{\partial V^H(f,0,z,r)}{\partial a} \tag{37}
$$

and

$$u'(c) \leq \frac{\partial V^R(f, 0, z, r)}{\partial a} \tag{38}$$

for homeowners and renters, respectively.

C.2 Analytical results from the model

Similar to standard search models, we characterize the reservation wage across assets and financial skills. The reservation mortgage rate is either constant across assets and skills or is implicitly given as a function of these two. Throughout this section, we assume deterministic productivity and no monetary refinancing costs (i.e., $c_{\text{ref}} = 0$).

Interest rate strategy

In a frictionless model, the arrival rate of mortgage offers is the same across homeownership rates. In this instance, the reservation rate does not depend on assets or financial skills and always corresponds to the mortgage payment.

Fixing productivity and denoting the reservation interest rate with $\tilde{r}$, the characterizing equality is

$$V^H(f, a, \tilde{r}) = V^R(f, a, \kappa).$$

Theorem C.1. *If the mortgage market does not differentiate between first time home-buyers and homeowners ($\phi = 1$), the reservation interest rate does not depend on assets or financial skills and corresponds to the costs of renting $\tilde{r}(f, a)M = \kappa$.*

Proof. The reservation mortgage rate $\tilde{r}(f, a)$ satisfies $V^H(f, a, \tilde{r}(f, a)) = V^R(f, a, \kappa)$. Because the value function V^H strictly decreases with the interest rate (budget constraint effect), the equation for $V^H(f, a, r)$ simplifies to

$$\begin{aligned}
\rho V^H(f, a, \tilde{r}(f, a)) = {} & u(c(f, a, \tilde{r}(f, a))) - c^m(s(f, a, \tilde{r}(f, a))) - c^f(i(f, a, \tilde{r}(f, a))) \\
& + \frac{\partial V^H(f, a, \tilde{r}(f, a))}{\partial f}\left[\frac{\mu}{\eta}(i(f, a, \tilde{r}(f, a)))^\eta - \delta f\right] \\
& + \frac{\partial V^H(f, a, \tilde{r}(f, a))}{\partial a}\left[Ra + w - \tilde{r}(f, a)M - c(f, a, \tilde{r}(f, a))\right] \\
& + \lambda s(f, a, \tilde{r}(f, a)) \int_{\underline{r}}^{\bar{r}} \max\left\{V^H(f, a, r') - V^H(f, a, \tilde{r}(f, a)), 0\right\} d\Phi(r') \\
& + p\left[V^R(f, a, \kappa) - V^H(f, a, \tilde{r}(f, a))\right],
\end{aligned}$$

while the renters value is

$$\begin{aligned}\rho V^R(f,a,\kappa) &= u(c(f,a,\kappa)) - c^m(s(f,a,\kappa)) - c^f(i(f,a,\kappa)) \\ &+ \frac{\partial V^R(f,a,\kappa)}{\partial f}\left[\frac{\mu}{\eta}(i(f,a,\kappa))^\eta - \delta f\right] \\ &+ \frac{\partial V^R(f,a,\kappa)}{\partial a}\left[Ra + w - \kappa - c(f,a,\kappa)\right] \\ &+ \lambda s(f,a,\kappa)\int_{\underline{r}}^{\bar{r}} \max\left\{V^H(f,a,r') - V^R(f,a,\kappa), 0\right\} d\Phi(r')\end{aligned}$$

Using the characterizing equation for the reservation rate and going into FOCs

$$\frac{\partial V^H(f,a,\tilde{r}(f,a))}{\partial a} = \frac{\partial V^R(f,a,\kappa)}{\partial a} = u'(c)$$

implies equal policy functions. Therefore, subtracting one value from the other yields

$$u'(c(f,a,\tilde{r}(f,a)))\left[\kappa - \tilde{r}(f,a)M\right] = 0 \implies \tilde{r}(f,a)M = \kappa.$$

□

The assumption $\phi = 1$ has a bite when used to infer policy function equalities. Plugging back $\phi < 1$ yields heterogeneity in reservation rates across financial skills and assets. In that case, subtracting $\rho V^R(f,a,\kappa)$ from $\rho V^H(f,a,\tilde{r}(f,a))$ yields

$$\begin{aligned}&- c^m(s(f,a,\tilde{r}(f,a))) + c^m(f,a,\kappa) + \\ &+ \lambda\left[s(f,a,\tilde{r}(f,a)) - s(f,a,\kappa)\right]\int_{\bar{r}}^{\tilde{r}(f,a)} (V^H(f,a,r') - V^R(f,a,\kappa))d\Phi(r') \\ &+ u'(c(f,a,\tilde{r}(f,a)))\left[\kappa - \tilde{r}(f,a)M\right] = 0,\end{aligned}$$

the implicit equation that characterizes the reservation rate $\tilde{r} = r(f,a)$.

Without assuming initial frictions at the mortgage market ($\phi = 1$) significantly reduces complexity when we infer the effect of mortgage performance on consumption growth. Optimal consumption growth in periods between mortgage refinancing elicits the effect of mortgage offer arrival rates and expense shocks. Whereas the expected change in debt repayment has the greatest effect on the highest-paying mortgage borrowers, the possibility of an expense shock has the strongest effect on borrowers with the lowest interest rates. For them, the next period value decreases all the way to the rent payment.

Corollary C.1.1. *Excluding external search frictions, variations in consumption growth can be attributed to three factors: patience, expected future mortgage rates, and precautionary measures in response to expense shocks.*

$$\frac{\dot{c}}{c} = \frac{1}{\sigma}\left[R - \rho - \lambda s\left(\int_{\underline{r}}^{r}\left(1 - \frac{u'(c(f,a,r'))}{u'(c(f,a,r))}\right)d\Phi(r')\right) + p\left(\frac{u'(c(f,a,\kappa))}{u'(c(f,a,r))} - 1\right)\right] \tag{39}$$

Proof. When we exclude external frictions, using the fact that value is decreasing in r, homeowners' problem simplifies to

$$\rho V^H(f,a,r) = \max_{\{c,s,i\}} \Bigg\{ u(c) - c^f(i) - c^m(s,f) + \frac{\partial V^H(f,a,r)}{\partial f}\dot{f} + \frac{\partial V^H(f,a,r)}{\partial a}\dot{a} + \\ + \lambda s \int_{\underline{r}}^{r} V^H(f,a,r') - V^H(f,a,r) d\Phi(r') + p\Big(V^R(f,a,\kappa) - V^H(f,a,r)\Big)\Bigg\}.$$

In the first step of the proof, we apply the envelope theorem to homeowners' problem with respect to assets (a) and obtain

$$\begin{aligned}\rho\frac{\partial V^H(f,a,r)}{\partial a} &= \frac{\partial^2 V^H(f,a,r)}{\partial f \partial a}\dot{f} + \frac{\partial^2 V^H(f,a,r)}{\partial a^2}\dot{a} + \frac{\partial V^H(f,a,r)}{\partial a}R \\ &+ \lambda s \int_{\underline{r}}^{r} \frac{\partial V^H(f,a,r')}{\partial a} - \frac{\partial V^H(f,a,r)}{\partial a} d\Phi(r') + \\ &+ p\bigg(\frac{\partial V^R(f,a,\kappa)}{\partial a} - \frac{\partial V^H(f,a,r)}{\partial a}\bigg).\end{aligned} \tag{40}$$

In the second step of the proof, we derive total derivative of $\frac{\partial V^H(f,a,r)}{\partial a}$. Possible changes can come from changes in assets da, changes in financial skills df, and changes in housing costs dr. Housing costs can change either due to refinancing $dq_{\lambda s}$ or due to financial shock, and transition to renting dq_p. Thus, we can summarize these changes with

$$dr = \min\{\tilde{r} - r, 0\} dq_{\lambda s} + \left(\frac{\kappa}{M} - r\right) dq_p.$$

Using changes mentioned above, total derivative of $\frac{\partial V^H(f,a,r)}{\partial a}$ satisfies

$$\begin{aligned} d\frac{\partial V^H(f,a,r)}{\partial a} &= \frac{\partial^2 V^H(f,a,r)}{\partial a^2}da + \frac{\partial^2 V^H(f,a,r)}{\partial f \partial a}df + \\ &+ \left[\frac{\partial V^H(f,a,\min\{\tilde{r}-r\})}{\partial a} - \frac{\partial V^H(f,a,r)}{\partial a}\right] dq_{\lambda s} + \\ &+ \left[\frac{\partial V^H(f,a,\frac{\kappa}{M})}{\partial a} - \frac{\partial V^H(f,a,r)}{\partial a}\right] dq_p.\end{aligned} \tag{41}$$

We focus on the case in which homeowners do not receive good enough refinancing offer and do not face financial shock, thus $dq_{\lambda s} = dq_p = 0$.

Next, we multiply equation (40) with dt and substitute for $\frac{\partial^2 V^H(f,a,r)}{\partial a^2}da + \frac{\partial^2 V^H(f,a,r)}{\partial f \partial a}df$ with the expression from equation (41). Thereby, equation (40) simplifies to

$$\begin{aligned}\rho\frac{\partial V^H(f,a,r)}{\partial a}dt &= d\frac{\partial V^H(f,a,r)}{\partial a} + \frac{\partial V^H(f,a,r)}{\partial a}Rdt \\ &+ \lambda s \int_{\underline{r}}^{r} \frac{\partial V^H(f,a,r')}{\partial a} - \frac{\partial V^H(f,a,r)}{\partial a} d\Phi(r')dt \\ &+ p\bigg(\frac{\partial V^R(f,a,\kappa)}{\partial a} - \frac{\partial V^H(f,a,r)}{\partial a}\bigg)dt.\end{aligned}$$

To further simplify, we use first order condition identities:

$$\frac{\partial V^H(f,a,r)}{\partial a} = u'(c(f,a,r)), \quad d\frac{\partial V^H(f,a,r)}{\partial a} = u''(c(f,a,r))dc$$

for homeowners, and

$$\frac{\partial V^R(f,a,\kappa)}{\partial a} = u(c(f,a,\kappa)),$$

for renters. Applying these identities yields the following expression

$$\rho u'(c(f,a,r))dt = u''(c(f,a,r))dc + u'(c(f,a,r))Rdt +$$
$$+ \lambda s \int_{\underline{r}}^{r} (u'(c(f,a,r')) - u'(c(f,a,r)))d\Phi(r')dt + p\Big(u'(c(f,a,\kappa)) - u'(c(f,a,r))\Big).$$

We divide this expression by $u'(c(f,a,r))dt$, we use CRRA property $\sigma = -\frac{u''(c)c}{u'(c)}$ and derivative notation $\frac{dc}{dt} = \dot{c}$, to obtain

$$\rho = -\sigma\frac{\dot{c}}{c} + R - \lambda s \int_{\underline{r}}^{r} \left(1 - \frac{u'(c(f,a,r'))}{u'(c(f,a,r))}\right) d\Phi(r') + p\left(\frac{u'(c(f,a,\kappa))}{u'(c(f,a,r))} - 1\right).$$

Finally, dividing by σ and rearranging yields final expression

$$\frac{\dot{c}}{c} = \frac{1}{\sigma}\left[R - \rho - \lambda s \int_{\underline{r}}^{r} \left(1 - \frac{u'(c(f,a,r'))}{u'(f,a,r)}\right) d\Phi(r') + p\left(\frac{u'(c(f,a,\kappa))}{u'(c(f,a,r))} - 1\right)\right].$$

□

C.2.1 Mortgage rate distributions

The *ad hoc* assumption on mortgage rate offer distribution $\Phi(r)$ dictates a structure for the endogenous accepted rate distribution. Utilizing the equilibrium flows between mortgage and rental markets, we derive the expression for the accepted rate distribution $G(r)$. Let h denotes the measure of homeowners in the equilibrium.

The flow of renters becoming homeowners is given with

$$(1-h)\lambda\Phi(\bar{r})\phi \sum_z \int_a \int_f s^R(f,a,z,\kappa)g(f,a,z,\kappa), \tag{42}$$

whereas homeowners return to renting in case of an expense shock

$$h \sum_z \int_r \int_f \int_a p(f,a)g(f,a,z,x)dadfdx. \tag{43}$$

Equalizing the two yields

$$(1-h)\lambda\Phi(\bar{r})\phi \sum_z \int_a \int_f s^R(f,a,z,\kappa)g(f,a,z,\kappa) = h \sum_z \int_r \int_f \int_a p(f,a)g(f,a,z,x)dadfdx \tag{44}$$

For a mortgage rate r or higher, the flow to homeownership is governed only by renters, as the homeowner's utility decreases with higher interest rates:

$$(1-h)(1-\Phi(r))\lambda\phi\sum_z\int_a\int_f s^R(f,a,z,\kappa)g(f,a,z,\kappa)dfda, \tag{45}$$

whereas the outflow of homeowners occurs exogenously due to the expense shock or endogenously through mortgage refinancing

$$h(1-G(r))\sum_z\int_a\int_f\int_r p(f,a)g(f,a,z,x)dxdfda+h\Phi(r)\lambda\int_a\int_f\int_r^{\bar{r}} s^H(f,a,z,x)g(f,a,z,x)dxdfda. \tag{46}$$

Equalizing the flows at mortgage rate r (expressions (45) and (46)), dividing with total outflow of homeowners (43), and using the expression (44), we get

$$1-\Phi(r)=1-G(r)+\frac{\Phi(r)\lambda\sum_z\int_r^{\bar{r}}\int_a\int_f s^H(f,a,z,x)g(f,a,z,x)dfdadx}{\sum_z\int_r\int_f\int_a p(f,a)g(f,a,z,x)dadfdx},$$

which implies

$$\frac{G(r)-\Phi(r)}{\Phi(r)}=\frac{\lambda\sum_z\int_r^{\bar{r}}\int_a\int_f s^H(f,a,z,x)g(f,a,z,x)dfdadx}{\sum_z\int_r\int_f\int_a p(f,a)g(f,a,z,x)dadfdx}>0, \tag{47}$$

which implies that Φ first-order stochastically dominates G. Moreover, we rearrange

$$G(r)=\Phi(r)\left[1+\frac{\lambda\sum_z\int_r\int_a\int_f s^H(f,a,z,x)g(f,a,z,x)dfdadx}{\sum_z\int_r\int_f\int_a p(f,a)g(f,a,z,x)dadfdx}\right], \tag{48}$$

which yields

$$\int_{\underline{r}}^{\bar{r}}(G(r)-\Phi(r))dr\geq 0,$$

so Φ second-order stochastically dominates G. The mean of Φ is as least as high as the mean of G, eliciting positive effects of the search effort.

C.3 Numerical solution method

Our numerical computation of the continuous time problem follows the method in Achdou et al. (2022). Individuals' decisions define a joint distribution of wealth, individual productivity, and housing type choice (represented by mortgage repayments). The exogenous grid for the mortgage rate, HJB equations (7) and (8) with corresponding first order conditions characterize agent's choice, conditional on owning a home. For a given productivity level, individual choices aggregate to a distribution of homeowners and renters that satisfy the Kolmogorov Forward Equation (9) and (10).

C.3.1 Homeowner's and renter's problem

As in Achdou et al. (2022), solving the (7) and (8) includes using the finite difference method for a joint grid on assets, financial skill level, productivity, and mortgage rates. The finite difference method includes assigning grids $[a_1, a_2, \dots, a_n]$ and $[f_1, f_2, \dots, f_m]$ with respective steps Δa and Δf to solve the discretized homeowner's and renter's problem. The grid is four-dimensional: - i runs through the asset grid, j denotes the financial knowledge grid point, k separates between two productivity states, and r denotes the mortgage rate grid element $[r_1, \dots, r_s]$. At each point in the grid, the discretized HJB equation is:

$$
\begin{aligned}
\rho V^H_{i,j,k,r} = {} & u(c^H_{i,j,k}) - c^f_{i,j,k,r} - c^m_{i,j,k,r} + \frac{V^H_{i+1,j,k,r} - V^H_{i,j,k,r}}{\Delta a}[\dot{a}_{i,j,k,r}]^+ + \frac{V^H_{i,j,k,r} - V^H_{i-1,j,k,r}}{\Delta a}[\dot{a}_{i,j,k,r}]^- \\
& + \frac{V^H_{i,j+1,k,r} - V^H_{i,j,k,r}}{\Delta f}[\dot{f}_{i,j,k,r}]^+ + \frac{V^H_{i,j,k,r} - V^H_{i-1,j,k,r}}{\Delta f}[\dot{f}_{i,j,k,r}]^- \\
& + \lambda s^H_{i,j,k,r} \sum_{r'=r_1}^{r_s} \max\{V^H_{i,j,k,r'} - V^H_{i,j,k,r}, 0\} d\Delta_r \\
& + \omega(k,k')[V^H_{i,j,k',r} - V^H_{i,j,k,r}] + p[V^R_{i,j,k} - V^H_{i,j,k,r}],
\end{aligned}
$$

where the step differences $i+1$, i and $i-1$ approximate derivatives of the value function $\dfrac{\partial V}{\partial a}$ and $\dfrac{\partial V}{\partial f}$. Choosing between the forward and backward differencing ensures convergence to the unique HJB solution (Achdou et al., 2022). The individual choice of mortgage refinancing necessitates going through all possible mortgage options (i.e., mortgage rates). Numerically, the integral breaks down to the average value over the mortgage rate grid $[r_1, \dots, r_s]$, at every iteration.

$\dot{a}_{i,j,k,r}$ are calculated using the upwind scheme described in Achdou et al. (2022) and separate two cases - whenever the corresponding state variable (assets) exhibits a positive or negative drift.

That is, using the FOC for the homeowner, we separate consumption for a positive or a negative drift in assets. Denote the consumption with respective difference as

$$
\begin{aligned}
u'(c^{Hb}_{i,j,k,r}) &= {}_a V^{Hb}_{i,j,k,r} \\
u'(c^{Hf}_{i,j,k,r}) &= {}_a V^{Hf}_{i,j,k,r}.
\end{aligned}
$$

Plugging into the budget constraint of the homeowner yields

$$
\begin{aligned}
\dot{a}^{Hb}_{i,j,k,r} &= Ra^{i,j,k,r} + wz_{i,j,k,r} - Mr^m_{i,j,k,r} - c^{Hb}_{i,j,k,r} \\
\dot{a}^{Hf}_{i,j,k,r} &= Ra^{i,j,k,r} + wz_{i,j,k,r} - Mr^m_{i,j,k,r} - c^{Hf}_{i,j,k,r}.
\end{aligned}
$$

Now, setting

$$
c^H_{i,j,k,r} = \mathbf{1}_{\{\dot{a}^{Hf}_{i,j,k,r}>0\}} c^{Hf}_{i,j,k,r} + \mathbf{1}_{\{\dot{a}^{Hb}_{i,j,k,r}<0\}} c^{Hb}_{i,j,k,r} + \mathbf{1}_{\{\dot{a}^{Hf}_{i,j,k,r}<0<\dot{a}^{Hb}_{i,j,k,r}\}} c^0_{i,j,k,r}
$$

and denoting corresponding assets as $\dot{a}^H_{i,j,k,r} = Ra^{i,j,k,r} + wz_{i,j,k,r} - Mr^m_{i,j,k,r} - c^H_{i,j,k,r}$ ensures convergence to the unique solution of the HJB equation. Moreover, the boundary condition given with the equation (37), corresponding to $a \geq 0$ constraint is enforced by setting

$$_{a}V^{H,b}_{1,j,k,r} = u'(wz_{1,j,k,r} - Mr^m_{1,j,k,r}).$$

Similarly, the upwind scheme and FOC with respect to financial skills investment separate between two types of investment, depending on a drift in financial skills

$$i^{Hb}_{i,j,k,r} = \left(\frac{1+z_{i,j,k,r}}{i_0} V^{Hb}_{i,j,k,r} \mu f^{\eta}_{i,j,k,r}\right)^{\dfrac{1}{\frac{1}{\gamma_1} - (\eta - 1)}}$$

$$i^{Hf}_{i,j,k,r} = \left(\frac{1+z_{i,j,k,r}}{i_0} V^{Hf}_{i,j,k,r} \mu f^{\eta}_{i,j,k,r}\right)^{\dfrac{1}{\frac{1}{\gamma_1} - (\eta - 1)}},$$

which then imply the corresponding financial skill investment costs $c^f(i^{Hf}_{i,j,k,r}, z_{i,j,k,r})$. The solution defines the grid for financial skills between 0 and 1, and the bounds are enforced with reflections in the corners

$$_{f}V^{H,b}_{1,j,k,r} = {}_{f}V^{H,b}_{2,j,k,r} \quad \text{and} \quad {}_{f}V^{H,f}_{i,m,k,r} = {}_{f}V^{H,f}_{i,m-1,k,r}.$$

Lastly, the HJB solution satisfies the FOC for search intensity that

$$s^H_{i,j,k,r} = \left(\frac{1+f_{i,j,k,r}}{c_0} \lambda \sum_{r'=r_1}^{r_s} \max\left\{V^H_{i,j,k,r'} - V^H_{i,j,k,r}, 0\right\} d\Delta_r\right)^{\gamma_s},$$

which defines the costs endured when searching for better mortgage options $c^m(s^H_{i,j,k,r}, f^H_{i,j,k,r})$.

Thus, our algorithm uses the current value function iteration to compute the integral over possible mortgage offers, simply by averaging out over all grid points for the mortgage interest rate.

The value function iteration generated by the upwind scheme is

$$\begin{aligned}
\frac{V^{H,l+1}_{i,j,r,k} - V^{H,l}_{i,j,r,k}}{\Delta} + \rho V^{H,l+1}_{i,j,k,r} &= U(c^H_{i,j,k,r}) + \frac{_{a}V^{Hb,l+1}}{\Delta a}[\dot{a}^{Hb}_{i,j,k,r}] + \frac{_{a}V^{Hf,l+1}}{\Delta a}[\dot{a}^{Hf}_{i,j,k,r}] \\
&+ \frac{_{f}V^{Hb,l+1}}{\Delta f}[\dot{f}^{Hb}_{i,j,k,r}] \frac{_{f}V^{Hf,l+1}}{\Delta f}[\dot{f}^{Hf,l+1}_{i,j,k,r}] \\
&+ \lambda s^H_{i,j,k,r} \sum_{r'=r_1}^{r_s} \max\{V^{H,l}_{i,j,k,r'} - V^{H,l}_{i,j,k,r}, 0\} d\Delta_r \\
&+ \omega(k,k)'[V^{H,l+1}_{i,j,k',r} - V^{H,l+1}_{i,j,k,r}] + p[V^{R,l+1}_{i,j,k} - V^{H,l+1}_{i,j,k,r}],
\end{aligned}$$

and due to the finite sum calculation in each iteration, it does not allow for a compact expression. However, the value function update $V^{H,l+1}$ boils down to solving a linear system of equations, similar to (Achdou et al., 2022). The value function for the renter V^R is discretized analogously.

Value functions for the homeowner and the renter V^H and V^R are four-dimensional matrices. When stacked together, V satisfies the set of equations, written compactly as

$$\frac{V^{l+1} - V^l}{\Delta} + \rho V^{l+1} = U^l + (A^l + B^l + \Lambda + P)V^{l+1} + \Omega^l(V^l), \tag{49}$$

where dimensions correspond to joint grid points in a column vector $\dim V^l = \dim V^{l+1} = \dim U^l = N_a \times N_f \times N_z \times N_r$. Matrix A^l contains asset changes $\dot{a}^{Hb,Hf}$ and $\dot{a}^{Rb,Rf}$, whereas changes in financial skills comprise B^l. Analogously to the literature, Λ depicts productivity changes and P the stochastic transition from homeownership to renting.

Lastly, Ω^l is a max function that takes the current value and compares it to the new value along the r dimension. Our algorithm pre-computes $\Omega_l = \Omega^l(V^l)$ and and transforms (49) into a linear system that has a solution

$$V^{l+1} = \underbrace{\left((\frac{1}{\Delta} + \rho)I - A^l - B^l - \Lambda - P \right)^{-1}}_{\mathbb{C}} (U^l + \frac{1}{\Delta}V^l + \Omega_l), \tag{50}$$

given that the matrix $\mathbb{C}$ is not ill-conditioned.

C.3.2 Stationary distributions

The second part of the algorithm iterates on the discretized version of the KFE for homeowners (9) and renters (10), respectively. As KFEs include integration, we use Kronecker product matrix multiplication to include the integrals and ultimately obtain a linear system of equations. Our discretized version of the KFE for homeowners states

$$\begin{aligned}
0 = &-\frac{{}_f g^{Hb}_{i,j,k,r}}{\Delta_f}[\dot{f}_{i,j,k,r}]^- - \frac{{}_f g^{Hf}_{i,j,k,r}}{\Delta_f}[\dot{f}_{i,j,k,r}]^+ \\
&- \frac{{}_a g^{Hb}_{i,j,k,r}}{\Delta_a}[\dot{a}_{i,j,k,r}]^- - \frac{{}_a g^{Hf}_{i,j,k,r}}{\Delta_a}[\dot{a}_{i,j,k,r}]^+ \\
&- (p + \lambda s^H_{i,j,k,r}\Phi(r))g^H_{i,j,k,r} + \lambda \sum_{r'=r_1}^{r_s} s^H_{i,j,k,r'} g^H_{i,j,k,r'} d\Delta_r \\
&+ \lambda \phi s^R_{i,j,k} g^R_{i,j,k} + \omega(k,k')(g^H_{i,j,k',r} - g^H_{i,j,k,r}),
\end{aligned}$$

and for the renter takes the form

$$
\begin{aligned}
0 = &- \frac{{}_a g^{Rb}_{i,j,k}}{\varDelta_a}[\dot{a}_{i,j,k,r}]^- - \frac{{}_a g^{Rf}_{i,j,k}}{\varDelta_a}[\dot{a}_{i,j,k,r}]^+ \\
&- \frac{{}_f g^{Rb}_{i,j,k}}{\varDelta_f}[\dot{f}_{i,j,k}]^- - \frac{{}_f g^{Rf}_{i,j,k}}{\varDelta_f}[\dot{f}_{i,j,k}]^+ \\
&+ p \sum_{r'=r_1}^{r_s} g^H_{i,j,k,r'} \varDelta_r - \lambda \phi s^R_{i,j,k} g^R_{i,j,k} \\
&+ \omega(k,k')(g^R_{i,j,k'} - g^R_{i,j,k}).
\end{aligned}
$$

The two equations together can be denoted in a more compact way, stacking two distributions (homeowners and renters) on top of each other. Compact notation reduces the system to a homogeneous linear system of equations.

While other components are simple to denote as linear operators, we construct the operator that produces the sum over all mortgage rates for each of the state variables as a Kronecker product of a sparse matrix τ that contains ones along the corresponding dimension and a matrix $\mathbb{S}^H$, which is constructed from a vectorized policy matrix $\text{vec}(s^H)$. That is, we obtain $\sum_{r'=r_1}^{r_s} s^H_{i,j,k,r'} g^H_{i,j,k,r'}$ with $\tau \mathbb{S}^H g$. In the discretized version of KFE for the renter, we do the same thing, and define a matrix that extracts the distribution of homeowners along the mortgage rate dimension, $g^H_{i,j,k,r}$, for $r_1, \ldots, r_s$ and multiply the matrix τ with the vectorized policy matrix $\text{vec}(s^H)$.

Using a similar argument for the renter's KFE equation, the stacked distribution g satisfies:

$$(A + B - P + \varLambda + \mathbb{S}^H)g = 0,$$

with an additional equation that ensures that stacked g is, in fact, a distribution and integrates into one.

C.3.3 Individual decisions in the equilibrium

In Figures 58 and 59, we present the individual policy functions as 3-dimensional surfaces. These policy functions capture the behavior and characteristics of homeowners and renters. Notably, both the slopes and relative relationships depicted in the figures closely align with our empirical data findings in Chapter 2.

In line with the non-monotonic age averages observed in financial literacy scores from the SCF (Figure 18 in the main text) and the panel data findings presented in Agarwal et al. (2007), the left panel of Figure 58 displays a non-monotonic pattern of investment in skills concerning current skill levels. Over time, as the level of skills increases, investment in skills begins to decrease. The right panel of Figure 58 matches the variation in search effort shown in our data findings - search effort increases with skill level and mortgage repayment amount, conforming to our estimates from the NSMO (Table 33) and SCF data (Figure 57).

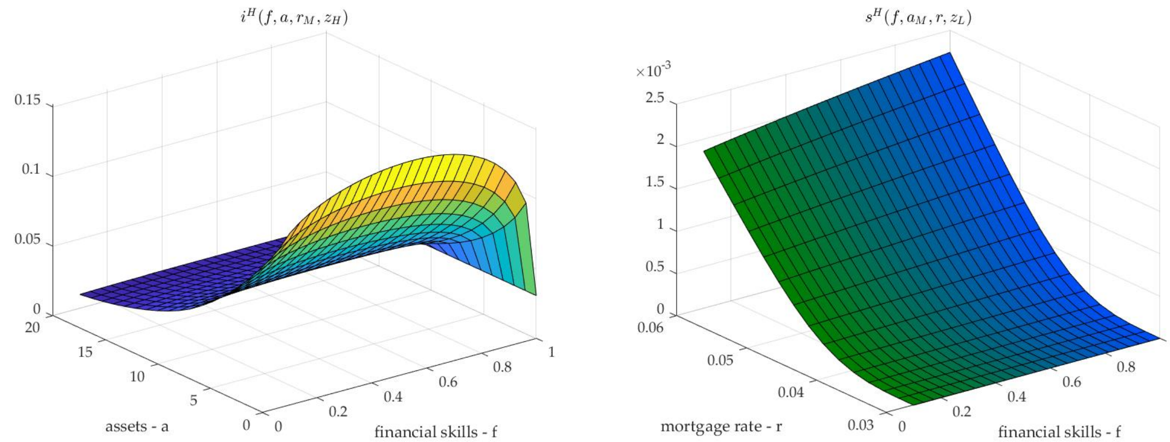


Figure 58: Investment (left) and search effort (right) policy functions for low productive homeowners over mortgage rates and financial skills, averaged over assets.

Skills investment policy for renters (left panel in Figure 59) resembles that of homeowners in term of its shape, but is lower in the level pertaining to the lack of learning-by-doing effects (Agarwal et al., 2007). Moreover, renter search intensity increases with skills, conforming to our SCF data findings (see Figure 20). Individual search and investment policies generate housing cost heterogeneity through mortgage take-up and refinancing, both of which we analyze in depth in our main text.

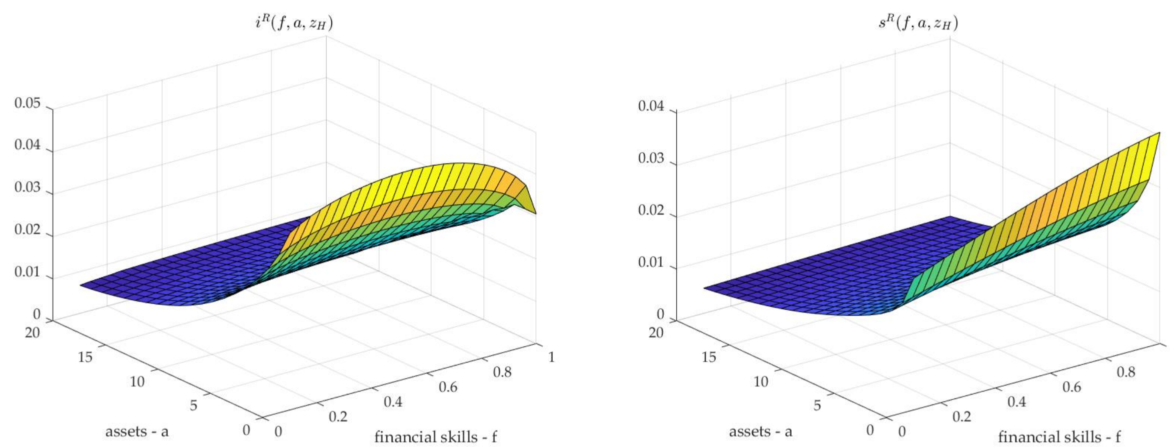


Figure 59: Renter's policy function for investment (left) and search effort (right) over assets and financial skills.

C.3.4 Locked-in rate in the equilibrium

In the equilibrium, the heterogeneity in locked in interest rates dictates consumption disparity net of housing costs. Figure 60 compares conditional densities of locked in mortgage rates between high and low skilled borrowers. In the equilibrium, unskilled borrowers lock-in at almost random rates, due to less sampling from the mortgage offer distribution (Figure 60). In the data, this translates to considering only one lender. Financially savvy borrowers exert more search effort, draw from a larger sample of mortgage offers (which we interpret as considering more lenders), and ultimately achieve better rates. In comparison with unskilled borrowers, the financially savvy end up with more resources net of mortgage repayment.

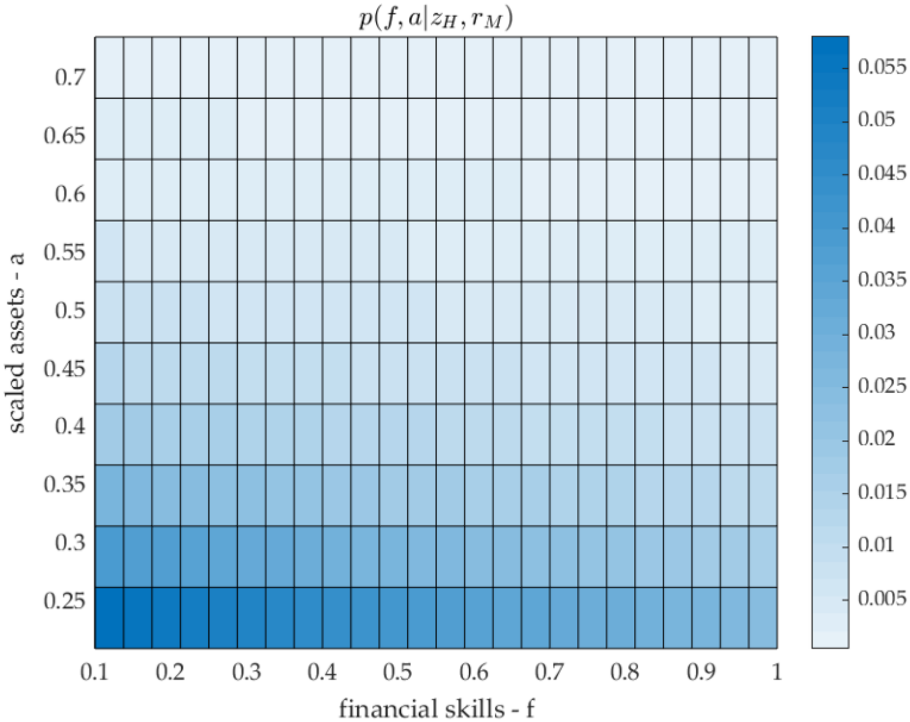
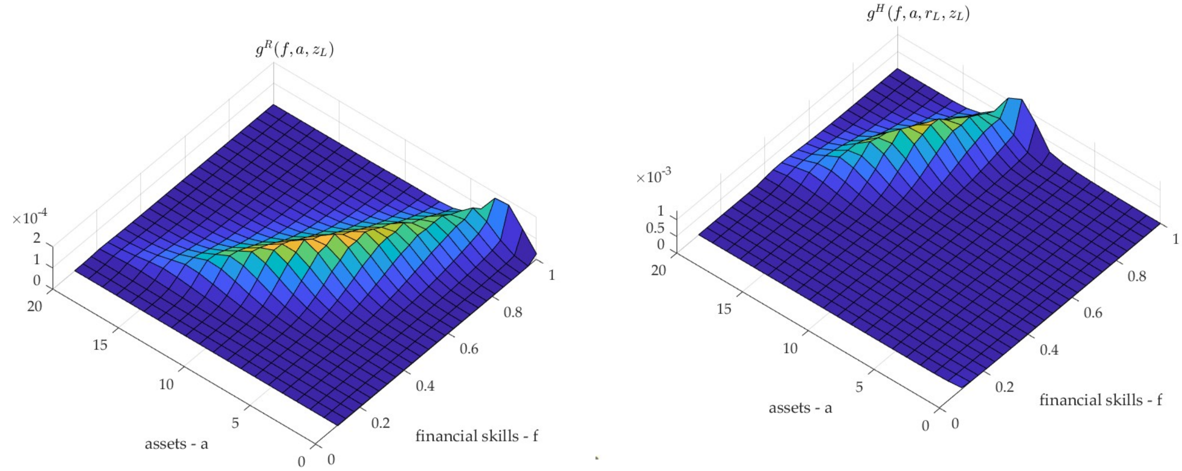


Figure 60: Mortgage rate density, for low- and high- skilled borrowers, averaged over assets and productivity.

The green histogram in Figure 61 highlights the search intensity among savvy homeowners. Due to their search efforts, high-skilled borrowers bunch at the lowest rates, and adjust their consumption due to precautionary motives caused by the possibility of an expense shock (states in the consumption growth equation (39)). Additional precautions induces saving among the best-performing mortgage owners (shown in the right panel of Figure 61). In this regard, saving policy conforms to higher savings rates among wealthy homeowners found in Mian et al. (2020).

On the other hand, high-rate payers continue to invest in their financial skills so as to reduce their future mortgage rates, exhibiting the dissaving effect of the expected mortgage rate change channel outlined in the equation (39). Because returning to renting is not as costly when compared to large mortgage payments, lower precautionary motives propagate savings inequality among homeowners. To this extent, model simulations show that the equilibrium consumption growth exhibits the precautionary channel in the less simple setting with endogenous default rates. In addition, the model generates stationary distributions that capture the disincentive for homeowners with substantial assets from refinancing simply because the mortgage payment does not affect their liquidity.

On the renters' side, assets are more dispersed, as a majority of renters accumulate skills to enter the mortgage market and face lower housing costs (the joint assets-skill density for renters is shown in the left panel in Figure 61). The model suggests that the wealthy renter exhibits low incentives to accumulate additional skills and prefers to remain a renter, regardless of paying higher housing costs. For wealthy renters, a costly search has a significant effect.



(b) marginal density over assets and skills; home-
(a) marginal density over assets and skills; renters.owners.

Figure 61: Density over assets and skills, low productive renters (left) and low productive homeowners with low mortgage payments (right).