Expectations on Wealth Returns: Implications for Labor Supply During the Retirement Boom*

Serdar Birinci†
FRB St. Louis

Miguel Faria-e-Castro
FRB St. Louis

Kurt See
National University of Singapore

November 2025

Abstract

We use an overlapping-generations model with incomplete markets and a frictional labor market to study how assumptions about agents' expectations of changes in returns to wealth affect labor supply and retirement decisions. Focusing on 2020–23, when returns fluctuated sharply and retirements rose above trend, we find that when individuals internalize the dependence of returns on wealth and view changes in returns as persistent, the model generates counterfactual labor-market outcomes. Retirements fall because expectations of persistently high returns boost labor supply, outweighing wealth effects, and the model predicts retirements concentrated among the very wealthy, contrary to the microdata.

Keywords: Labor Supply, Retirement, Heterogeneous Returns, Incomplete Markets
JEL Classification: E24, G11, J21, J22, J26

*This research was supported through computational resources provided by the Big-Tex High Performance Computing Group at the Federal Reserve Bank of Dallas. The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of St. Louis or the Federal Reserve System. Birinci: serdar.birinci@stls.frb.org, Faria-e-Castro: miguel.fariaecastro@stls.frb.org, See: kurtgerrardsee@gmail.com.

†Corresponding author. Address: Broadway and Locust, 1 Federal Reserve Bank Plaza, St. Louis, MO 63102.

1 Introduction

A growing literature in macroeconomics and household finance documents that returns on wealth vary considerably across households (Bach et al., 2020; Fagereng et al., 2020; Smith et al., 2022; Ozkan et al., 2023), and that differences in returns to wealth are an important feature that can help heterogeneous-agent models match empirically observed levels of wealth inequality (Benhabib and Bisin, 2018; Benhabib et al., 2019; Xavier, 2021). However, the implications of heterogeneous asset returns for labor supply and retirement decisions, and how agents form expectations about changes in those returns, have not been widely studied. The post-pandemic period from 2020 to 2023, marked by volatile and heterogeneous returns to wealth alongside a surge in retirements, provides a unique setting to study these questions.

This paper makes a concise theoretical and quantitative contribution. First, we study how expectations about the future path of asset returns affect labor supply, with a focus on retirement decisions. Second, we quantitatively assess alternative assumptions about these expectations and determine which best account for the observed aggregate and distributional changes in retirement patterns during the period of interest. To do so, we employ a heterogeneous-agent overlapping-generations (OLG) model with incomplete markets, in which asset returns depend on wealth and age, and a frictional labor market. We show that expectations about the persistence of return fluctuations are central to labor supply responses. When agents view higher returns as *transitory*, wealth effects dominate and labor supply contracts. In contrast, if higher returns are seen as *permanent*, agents expand labor supply to accumulate more wealth. Thus, alternative expectations about changes in asset returns lead to opposite labor supply outcomes.

Our model incorporates realistic life-cycle income dynamics, unemployment insurance (UI), and social security (SS) benefits. We calibrate the model's steady state to the U.S. in 2019 and study transitions for 2020–23—a period of great interest given the large observed changes in both wealth returns and aggregate labor supply, especially among older workers. Our analysis feeds in exogenous shocks that could help explain the observed changes in retirement. These shocks capture (i) the heterogeneous movements in returns to wealth, (ii) the heterogeneous rise in job-separation rates across the labor income distribution, (iii) economic impact payment programs, (iv) the expansion of UI, and (v) the increase in mortality risk, which was steeper for older people. Birinci et al. (2025) show that when agents expect changes in returns along the transition to be transitory, the model captures the observed changes in aggregate labor market moments during 2020–2023 and that its predictions are consistent with observed patterns of retirement across wealth and income levels in the microdata. Birinci et al. (2025) also find that changes in asset returns explain one-fifth of excess retirements in 2022–23.

In this paper, we contrast this result and show that when agents expect changes in asset returns to be permanent, the model predicts the opposite for retirement decisions. Instead

of increasing the fraction of retirees in the economy, elevated returns encourage greater labor supply, as individuals work more to accumulate wealth faster. This "substitution effect" more than offsets the wealth effects of elevated returns on labor supply. We show that this prediction is counterfactual, at least during the period of analysis. Beyond the opposite aggregate response, the predicted distribution of new retirees—contrary to the observed patterns in the microdata—is composed almost exclusively of individuals in the top quintile of the asset distribution, as other agents find it worthwhile to keep working and accumulate wealth. Overall, our results serve as a cautionary note for models incorporating heterogeneous and time-varying returns on wealth, since the assumption about expectations regarding asset returns greatly matters for labor supply.

2 The Analysis in Birinci et al. (2025)

Model. We combine a partial-equilibrium heterogeneous-agent incomplete-markets OLG model with a frictional labor market to quantify the contributions of various factors to changes in retirement outcomes in 2020–23.[1] The model is described in Birinci et al. (2025), and we provide a brief summary here. Agents are indexed by five states: age $j \in \{25, \ldots, 90\}$; wealth $a \in [-\underline{a}, \infty)$; employment status $\ell \in \{E, U, N\}$ (employed, unemployed, non-participant); wage $w \in \mathbb{R}^+$ if employed or last wage if not employed; and retirement age $k \in \{62, \ldots, 70\}$. They face an age- and employment-status-dependent death probability, $1 - \pi(j, \ell)$, and die with certainty at age 91.

Agents have CRRA preferences over consumption and incur age-dependent disutility from employment and unemployment. They can save in a risk-free asset that pays return $r(a, j)$ on savings ($a \geq 0$) and a constant rate r^b on borrowings ($a < 0$). A single-asset model with returns linked to individual wealth and age offers a tractable way to capture portfolio heterogeneity.

Labor income depends on a stochastic wage w' that evolves according to a persistent process $F(w' \mid w)$ and an age-specific profile $\psi(j)$, as in French (2005) and Blandin et al. (2023). Agents choose how much to consume and save. Employed agents face an exogenous probability $\delta(w, j)$ of job loss, which is wage- and age-dependent, and may also choose to quit their job and move to either U or N. Unemployed agents receive job offers at rate f, collect UI benefits $b(w, j)$ based on their last wage while employed, and have income from home production $h(j)$, which depends on age. Finally, non-participant agents receive job offers at a lower rate $\gamma \times f$ with $\gamma < 1$, have income from home production, and do not incur disutility from work or job search. All agents may receive fiscal transfers $T(y, j, a)$, where y is their total income. We classify non-participants as retired if they are 62 or older; any agent 62 or older may receive SS benefits that depend on their last or current wage, age, age of retirement, and employment status, $\bar{y}^{SS}(w, j, k, \ell)$.

[1] While a simpler model (e.g., a Bewley model) can generate the same labor supply contraction with persistently elevated returns, it cannot quantitatively evaluate alternative expectation assumptions against the observed labor market dynamics in this episode.

Calibration and validation. We calibrate the model to match relevant moments of the U.S. economy in 2019, which we consider to be the model's steady state. In particular, we focus on moments that discipline the relationship between labor supply, wealth, and income. We incorporate realistic UI benefits, calibrated to match both the average replacement rate and its variation with previous earnings in the data. We also consider a realistic SS income function, which accounts for the history of labor earnings and age. It features realistic penalties for those who work while claiming SS benefits, as well as for those who retire early. We estimate the return function $r(a, j)$ by imputing individual portfolio returns to the 2019 Survey of Consumer Finances, using aggregate returns across different asset classes and individual portfolio exposures to those asset classes.[2] Birinci et al. (2025) present the details of the calibration, along with several important model validation results. We show that the model does a good job of matching untargeted data moments, such as the wealth distribution, as well as how retirement outcomes change across quintiles of the wealth and income distributions of older employed workers.

Main transition experiment. The main quantitative experiment we conduct involves introducing five series of exogenous shocks into the model's steady state, each representing a potential driver of changes in labor force participation during the 2020-23 period. These five shocks are: (i) changes in the return function $r(a, j)$ by wealth and age, which we estimate for each month of the 2020-23 period using realized returns for aggregate asset classes; (ii) changes in job-separation rates by earnings, $\delta(w, j)$, to capture labor market disruptions; (iii) changes in fiscal transfers, $T(y, a, j)$, to capture federal economic impact payments; (iv) changes in UI benefits, $b(w, j)$; and (v) changes in mortality rates by age and employment status, $\pi(j, \ell)$, reflecting concerns about workplace contagion influencing decisions to stay out of the labor force.

The model successfully captures *untargeted* aggregate labor market dynamics, such as (i) the rise in excess retirements—which we define as the increase in the fraction of retirees in the population (retired share) beyond the trend predicted by the aging population; (ii) the rise in the unemployment rate; and (iii) the decline in the employment-to-population ratio. Importantly, the model captures the magnitude and persistence of the movements in these variables. In Birinci et al. (2025), we use the model to decompose which shocks were most important in explaining the aggregate dynamics. We find that elevated job separations and economic impact payments explain the bulk of excess retirements during 2020–2023, while changes in asset returns account for one-fifth of excess retirements during 2022–2023. Importantly, we show that the model matches retirement patterns observed in the data during this period, both by wealth and income. Specifically, most new retirements were undertaken by low-income individuals who were more exposed to rises in job separations and fiscal transfers. This result explains why labor market conditions and fiscal transfers were more important in explaining the rise in excess retirements.

[2]See Faria-e-Castro and Jordan-Wood (2024) for a description of this method.

3 Temporary vs. permanent changes in asset returns

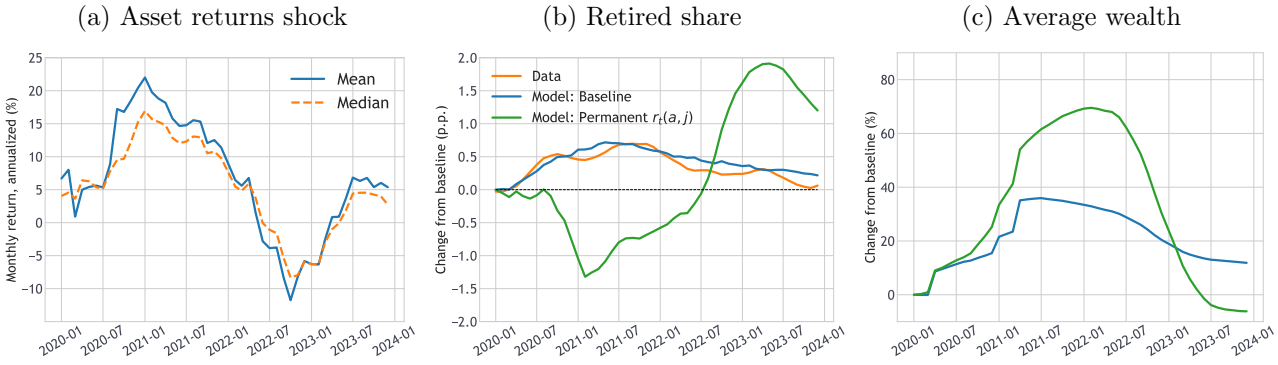
In Birinci et al. (2025), agents expect future returns to follow the steady-state function $\bar{r}(a, j)$. Thus, changes to this function are viewed as temporary surprises, with elevated returns acting as windfalls that generate wealth effects without distorting labor supply or savings incentives.

We now consider an alternative experiment in which agents perceive the changes in the return function to be permanent along the transition. That is, in each transition period t, agents observe the realized function $r_t(a, j)$ and expect the change in period t to be permanent. Panel (a) of Figure 1 shows the mean and median paths of the estimated return function $r_t(a, j)$.

Panel (b) compares excess retirements in the data (orange line, as defined above) with changes in the retired share from our baseline exercise (blue line) and from the alternative exercise assuming permanent changes in returns (green line). While our baseline exercise matches the data very well, the alternative exercise generates counterfactual predictions for the retired share. At its peak in late 2021, the excess retired share (i.e., the rise in the retired share from its steady-state level) was close to 0.7 pp, whereas the alternative exercise predicts -1.2 pp. Furthermore, while the excess retired share starts falling in 2022, the alternative exercise predicts a counterfactual boom in retirements during this time. Importantly, the counterfactual patterns in retirement outcomes in the alternative exercise mirror the path of returns in Panel (a). As agents perceive increased returns in 2020–21 to be permanent, they choose to work more to exploit those elevated returns. These incentives reverse in 2022, when returns fall below the steady state, which explains the protracted retirement boom during this period. Panel (c) shows that the increase in average net worth is much larger in the alternative exercise, as agents decide to work and save more to exploit what they perceive as permanently elevated returns. In particular, average net worth rises by around 70% relative to the steady state, double that of the baseline exercise. This is purely the result of different behavior, as the same return shocks are fed to both experiments.

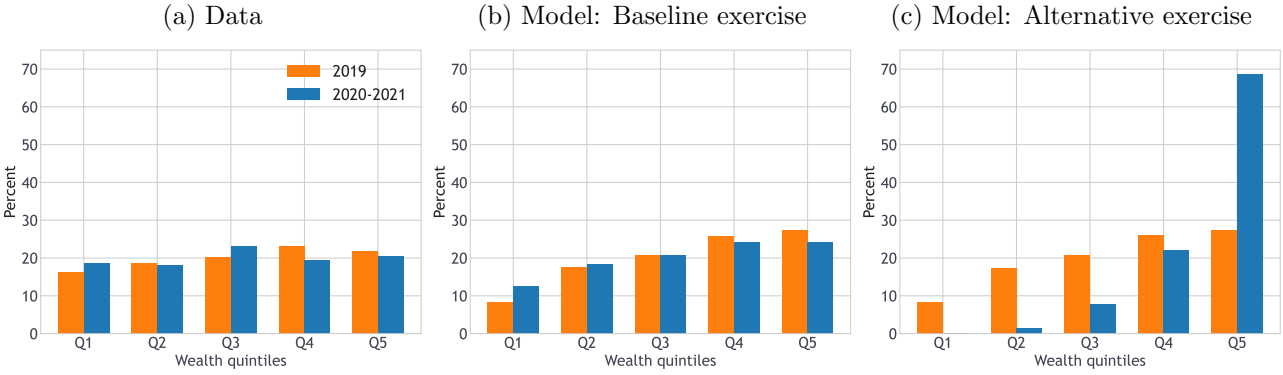
Turning to a distributional comparison, we examine the predictions of these two models in terms of retirement patterns across the wealth distribution using microdata from the Survey of Income and Program Participation (SIPP) 2020, 2021, and 2022 panels (covering monthly data between 2019 and 2021). In the data, for 2019, we define new retirees as individuals who report being labor force participants in one month and then indicate their retirement for the first time in the following month. We then categorize each new retiree into quintiles of the wealth distribution of employed individuals aged 62 to 72. This allows us to determine where each new retiree in 2019 falls within the wealth distribution of older employed workers eligible for retirement benefits, which is the key demographic for our analysis. We repeat this process for 2020 and 2021 to examine how retirement patterns based on wealth holdings changed during the pandemic. Panel (a) of Figure 2 plots the fractions of new retirees during each period (2019 or 2020–21) who are in each wealth quintile. It shows that the distribution was relatively flat,

Figure 1: Temporary vs. permanent changes in asset returns and retirement outcomes



Note: Panel (a) plots the mean and median paths of the estimated monthly return (annualized) function $r_t(a, j)$. We show only the mean and median values for expositional purposes. Panel (b) shows percentage-point deviations in the retired share (i.e., the fraction of retirees in the population) from the 2019 average in the data and the stationary state of the model. Panel (c) plots percent changes in average net worth in the model relative to the stationary state. Panels (b) and (c) present results from two model simulations: the baseline, where agents view changes in asset returns as temporary (blue lines), and the alternative, where changes in returns are perceived as permanent (green lines).

Figure 2: New retirees by wealth: Data vs model



Note: Panel (a) shows the fraction of new retirees across wealth quintiles using data from the SIPP, separately for those retiring in 2019 and 2020–2021. Panels (b) and (c) repeat the same calculations in the model, using results from the baseline and alternative exercises, respectively. The baseline exercise assumes that agents view changes in asset returns as temporary, whereas the alternative exercise assumes that changes in returns are perceived as permanent.

slightly increasing with wealth in 2019. Importantly, this pattern remained similar in 2020–21.

Panel (b) plots the same moments in the model for the baseline exercise, showing that it does well in matching these two facts: (i) a relatively flat distribution that is slightly increasing with wealth in 2019, and (ii) no significant changes in this pattern along the transition. Finally, Panel (c) plots the same results for the alternative exercise. It matches the first fact by construction, as both models have the same steady state, but fails to match the second fact. In particular, it produces a steep relationship between wealth and retirement along the transition. It predicts no retirements in the bottom quintile of the wealth distribution, and nearly 70% of all new retirees originate from the top quintile. Low-wealth individuals find it worthwhile to remain in the labor force and accumulate wealth, leveraging the high returns. Only sufficiently wealthy agents find it worthwhile to retire during this period—at odds with the empirical evidence in Panel (a).

4 Conclusion

We use an OLG, incomplete-markets model with a frictional labor market and labor force participation decisions to analyze the effects of expectations over changes in asset returns on labor supply, with an emphasis on retirement. In the baseline exercise, where return fluctuations are seen as temporary, labor supply follows traditional wealth effects, decreasing when returns are high. Conversely, in an alternative exercise where return fluctuations are expected to be permanent, a substitution effect outweighs this wealth effect: individuals increase labor supply when returns are high, prioritizing earnings and wealth accumulation to capitalize on higher returns.

We find that the alternative exercise generates predictions that are at odds with both the aggregate data and microdata for the 2020–23 period. It predicts a 1.2 pp decline in the retired share—while in reality it increased by 0.7 pp—and would have concentrated the distribution of new retirees among high-wealth individuals, which is at odds with the micro evidence. We conclude that expectations about returns are crucial in models with endogenous labor supply.

References

Bach, L., L. E. Calvet, and P. Sodini (2020): "Rich pickings? Risk, return, and skill in household wealth," *American Economic Review*, 110, 2703–47.

Benhabib, J. and A. Bisin (2018): "Skewed wealth distributions: Theory and empirics," *Journal of Economic Literature*, 56, 1261–91.

Benhabib, J., A. Bisin, and M. Luo (2019): "Wealth distribution and social mobility in the US: A quantitative approach," *American Economic Review*, 109, 1623–1647.

Birinci, S., M. Faria-e-Castro, and K. See (2025): "Dissecting the great retirement boom," *Journal of Monetary Economics*, forthcoming.

Blandin, A., J. B. Jones, and F. Yang (2023): "Marriage and work among prime-age men," Working Papers 2313, Federal Reserve Bank of Dallas.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020): "Heterogeneity and persistence in returns to wealth," *Econometrica*, 88, 115–170.

Faria-e-Castro, M. and S. Jordan-Wood (2024): "Pandemic labor force participation and net worth fluctuations," *Federal Reserve Bank of St. Louis Review*, 106, 40–58.

French, E. (2005): "The effects of health, wealth, and wages on labour supply and retirement behaviour," *The Review of Economic Studies*, 72, 395–427.

Ozkan, S., J. Hubmer, S. Salgado, and E. Halvorsen (2023): "Why are the wealthiest so wealthy? A longitudinal empirical investigation," Working Papers 2023-004, Federal Reserve Bank of St. Louis.

Smith, M., O. Zidar, and E. Zwick (2022): "Top wealth in America: New estimates under heterogeneous returns," *The Quarterly Journal of Economics*, 138, 515–573.

Xavier, I. (2021): "Wealth inequality in the US: The role of heterogeneous returns," Working paper, Federal Reserve Board of Governors.