Preference Heterogeneity and Portfolio Choices over the Wealth Distribution

Gualtiero Azzalini* Markus Kondziella † Zoltán Rácz ‡

SSE University of St. Gallen, SFI SSE

This version: May 26, 2025. First version: May 8, 2023

Abstract

Evidence from administrative data reveals increasing participation in risky assets, expected returns and idiosyncratic return risk over the wealth distribution. We explain these patterns quantitatively with an incomplete markets model with endogenous portfolio choice, cyclical labor income risk, and preference heterogeneity. As a by-product of this approach, our framework also generates wealth inequality, wealth mobility and marginal propensities to consume as a function of wealth closely in line with data. Simultaneously fitting wealth and portfolio choices delivers a larger and more persistent increase in wealth inequality following an aggregate return shock relative to counterfactual models lacking such a fit.

JEL codes: D31, E21, G11, G51.

We are grateful to Timo Boppart, John Hassler, Per Krusell, Kieran Larkin, Lars Ljungqvist, Kurt Mitman and Paolo Sodini for their advice and support. We also thank Adrien d'Avernas, Tobias Broer, David Domeij, Mehran Ebrahimian, Richard Foltyn, Mark Huggett, Karin Kinnerud, Paul Klein, Winfried Koeniger, Alexander Michaelides, Ofer Setty, Per Strömberg, Roine Vestman, Gianluca Violante and seminar participants at the Arne Ryde Workshop "Micro Data meets Macro Models", the IIES Macro Group, the Greater Stockholm Macro Group, the Midwest Macroeconomic meetings, EEA 2024, the Stockholm School of Economics and the University of St. Gallen for useful feedback and comments. We are grateful to Paolo Sodini and the Institute for Micro Data (MiDa) at the Stockholm School of Economics for sharing with us moments related to the Swedish wealth distribution and to Paula Roth for sharing moments of the distribution of log aggregate income computed from Swedish administrative data. We thank Francisco Tavares and Andreas Karlsson for excellent research assistance. Azzalini and Rácz thank the Jan Wallanders and Tom Hedelius and the Tore Browaldhs foundations for generous financial support. The computations were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

*Stockholm School of Economics, gualtiero.azzalini@hhs.se

†University of St. Gallen and Swiss Finance Institute, markus.kondziella@unisg.ch

‡Stockholm School of Economics, zoltan.racz@hhs.se

1 Introduction

The high level of inequality in advanced economies is one of the most discussed issues in economics of the last decades. Among the several possible explanations of the highly skewed distribution of wealth,[1] a particularly promising one emphasizes the role of capital income: if the rich experience systematically higher returns on their savings than the poor, wealth concentration increases. In line with this view, recent empirical literature leveraging high-quality administrative data (Bach et al., 2020; Fagereng et al., 2020) has documented that participation in risky assets, expected returns and idiosyncratic return risk increase over the wealth distribution.

The relationship between portfolio choices and wealth is also a key determinant of changes in inequality following economic shocks or policy changes. As an example, consider a sudden increase in stock market returns. How does such a shock impact the cross-sectional distribution of wealth? First, since capital gains equal the product of wealth and returns, the contemporaneous response in inequality critically depends on the joint distribution of the latter two. Second, the long-run response hinges on how individuals adjust their investment choices to the additional capital income. Hence, any attempt to evaluate distributional consequences must account for the heterogeneity in portfolio composition resulting from agents' optimal decisions. The same logic applies to a wide range of policy questions, such as the redistributive implications of capital income or wealth taxation.

Nevertheless, a model that endogenously generates a joint distribution of wealth and portfolio choices quantitatively in line with the data is, to the best of our knowledge, still missing. On the one hand, while the household finance literature starting with Cocco et al. (2005) has proposed several mechanisms to explain the low risk-taking of the poor,[2] as matching wealth inequality is outside the scope of papers in this field, it is generally unknown if these solutions can quantitatively fit the increasing expected returns and return risk schedules over the whole distribution. This is crucial because the overwhelming majority of wealth is held by the richer half of the population. Therefore, undershooting their risky investments could lead to substantial biases in both aggregate quantities and distributional changes, even if the portfolio choices of the poor are perfectly matched. On the other hand, the wealth inequality literature has avoided the notoriously difficult task of generating an increasing relationship between the risky share and wealth as a result of optimal choices by either hard-wiring differences in portfolio returns (Benhabib et al., 2019; Hubmer et al., 2021) or attributing them purely to luck (Benhabib et

[1] See De Nardi and Fella (2017) and Benhabib and Bisin (2018) for a review.

[2] Vissing-Jorgensen (2003) points out the role of participation costs, Wachter and Yogo (2010) highlight the role of non-homothetic preferences, while Chang et al. (2018) and Catherine (2021) emphasize the role of labor income risk. We discuss the related literature in detail in the following section.

al., 2015; Nirei and Aoki, 2016). As a consequence, these models are not immune to the Lucas critique and, therefore, are less suited to study changes affecting portfolio choices.

This paper aims to fill this gap with a model that explains the observed wealth and portfolio heterogeneity in the data at once. Besides incomplete markets and endogenous portfolio choice, our framework features cyclical skewness in labor income shocks – that is, disastrous labor income outcomes are more likely in times of stock market crashes – and ex-ante heterogeneity in preferences and idiosyncratic return risk across individuals. We estimate the model parameters by targeting the increasing schedules of participation, expected excess returns, and idiosyncratic return risk over the wealth distribution documented in Bach et al. (2020), and find that our framework matches these portfolio choice moments very well thanks to the latter two elements mentioned above, as we explain next.

As pointed out by Catherine (2021), cyclical skewness in labor income shocks introduces significant background risk for agents whose net worth mostly consists of human capital. We show that while this mechanism explains the depressed risk-taking at the bottom of the wealth distribution, it falls short in explaining portfolio choices at the top half of the wealth distribution. This is because when individuals accumulate enough savings to self-insure against labor income shocks, the importance of the latter for optimal portfolio allocations becomes negligible. As a result, in a model in which agents are ex-ante identical, the share of wealth invested in risky assets eventually declines over the distribution, as first described by Cocco et al. (2005).

This is where the ex-ante heterogeneity becomes relevant. In our framework, individuals are of two different types. As an outcome of the model estimation, one type of agents is characterized by high risk aversion, low patience and reduced idiosyncratic return risk. The other type is characterized by the opposite traits. Thus, mainly thanks to their higher patience, less risk-averse agents who choose a higher share invested in risky assets endogenously end up in the right part of the wealth distribution. The compositional effect resulting from the rising fraction of these individuals as a function of wealth thus delivers the increasing relationship between wealth and the risky share even for the highest quantiles of the distribution. Additionally, as these agents are characterized by higher idiosyncratic return risk, we also match the higher share of idiosyncratic variance at the top.

The estimation of ex-ante heterogeneity across individuals in our model also bridges the often inconsistent preference parameter values across the household finance and macroeconomics literature. Studies in household finance – typically explaining savings and portfolio choices of the average household – frequently find yearly discount rates around 8% and relative risk aversion coefficients above six (e.g., Catherine, 2021; Fagereng et al., 2017), while the macroeconomics literature – typically setting preference parameters to match aggregate quantities – usually operates with values less than half of these two figures. Consistent with these differences, in our benchmark specification 90% of the agents feature relatively low patience and high risk aversion in line with the household finance literature. Yet, these agents do not accurately represent the whole economy: even though only 10% of the population is characterized by relatively high patience and low risk aversion – typical in the macroeconomics literature – their impact on aggregates is crucial, as these individuals concentrate on the top of the wealth distribution and hold more than half of total wealth.

As a by-product of matching the portfolio schedules over wealth, our framework also delivers a close fit of three non-targeted key characteristics of the wealth distribution. First, the model replicates the degree of wealth inequality observed in administrative data. In particular, compared to the previous literature, the implied shares of wealth held by different segments of the wealth distribution are remarkably close to their counterpart in the data, even for the richest 1%. Together with the targeted portfolio choice patterns, this ensures that the joint distribution of wealth and risky assets generated by our framework is empirically sound. Second, mobility across wealth groups is also consistent with existing empirical evidence (Bach et al., 2020; Benhabib et al., 2019). This is relevant both for accurately measuring the welfare cost of inequality and for delivering a realistic dynamic evolution of the wealth distribution. Third, the marginal propensity to consume – a crucial determinant of responses to economic shocks – decreases over wealth in a way quantitatively in line with earlier literature (Carroll et al., 2017; Kaplan et al., 2018). While these quantities have been replicated before separately, our framework fits all of them at once without any additional ingredient besides those needed for generating empirically plausible portfolio compositions over wealth. We take this result as indirect evidence that our model mechanisms are consistent with how wealth inequality emerges in reality.

As a next step, we confirm the last statement with a counterfactual analysis. Specifically, we isolate the importance of matching portfolio schedules in delivering the joint match of the above features by shutting down key elements of our model one at a time. We solve a version with homogeneous preferences, one with heterogeneity only in the discount factor and one without idiosyncratic risk in individual portfolios, and find in all cases that either the match of the portfolio schedules, wealth inequality or mobility is worsened.[3] This exercise also highlights the importance of allowing for preference heterogeneity in several dimensions simultaneously. For example, while it is known at least since Krusell and Smith (1998) that differences in discount factors can generate realistic levels of inequality, in our counterfactual analysis we find that when only heterogeneity in this parameter is allowed for, the model indeed generates higher levels of wealth inequality but implies too low mobility at the top and unrealistic portfolio choice patterns.

[3]In Appendix D, we find similar results also when considering a counterfactual economy with heterogeneity only in risk aversion, one without cyclical skewness in labor income shocks and one with hard-wired portfolio choices.

Lastly, motivated by the example at the beginning of this introduction, we quantify the response of wealth inequality to a 10% aggregate return shock. While the shock has a positive and persistent effect on aggregate wealth, the impact is very heterogeneous across agents. On the one hand, due to their higher risky share, wealthier individuals experience larger capital gains. This translates into a 0.2 percentage points higher share of wealth held by the top 1% on impact.[4] On the other hand, as the rich are also more patient, they reinvest a larger fraction of the additional capital income, which, together with their higher propensity to invest in risky assets, makes the effect on inequality long lasting. Indeed, even though the model generates high top mobility levels as in the data, it takes about one hundred years for the initial effect to be halved. To highlight the importance of matching portfolio choice patterns over wealth also in this dynamic context, we show that the counterfactual model limiting the extent of ex-ante preference heterogeneity to only differences in patience results in a *decrease* in the share of wealth held by the top 1% and in a less persistent effect, following the positive return shock. The reason is that this alternative specification – despite replicating the degree of observed wealth inequality – fails to generate the large capital gains of the rich because it predicts a decreasing risky share as a function of assets in the right tail of the wealth distribution. Consistent with this explanation, we further document that this bias is magnified in a model with only one type of agents, which, in addition to producing such a decreasing relationship already from lower wealth levels, also implies much less wealth concentration at the top. We conclude, therefore, that not matching portfolio choice patterns over a realistic wealth distribution might lead to incorrect policy conclusions, even when investment decisions are not the main focus of the analysis.

Related literature. This paper contributes to and connects the household finance literature on portfolio choice and the macroeconomics literature on wealth inequality.

Starting from Cocco et al. (2005), a sizable literature in household finance has analyzed the portfolio choices of households in life-cycle models.[5] The focus of such papers has mostly been understanding the optimal portfolio allocation of the average investor, and how that is influenced by age. Our contribution to this literature is extending this analysis over the wealth distribution. We take a state-of-the-art explanation for portfolio choices over the life-cycle by assuming labor income shocks featuring cyclical skewness[6] and show that when wealth inequality is as high as

[4]To see that this effect is non-negligible, note that in the extreme scenario in which the top 1% holds only stocks and the bottom 99% only bonds, the wealth distribution generated by our model would imply a response of 1.7 percentage points.

[5]See Gomes (2020) for a review.

[6]This property of the income process was first documented by Guvenen et al. (2014) using U.S. data and has then been confirmed for other countries and different measures of earnings by Busch et al. (2022). Theoretically, Catherine

in the data, this mechanism is unable to rationalize the increasing risky shares over the top half of the wealth distribution. To rationalize the portfolio allocation decisions of the rich, we introduce preference and idiosyncratic return risk heterogeneity in the model, and find that one can obtain an accurate match for participation, risky share and idiosyncratic risk schedules already with two types of households. In addition to matching the risky shares of the rich, we also show how that helps to obtain realistic wealth inequality, marginal propensities to consume and invest, and mobility across the wealth distribution. Due to the use of different kinds of models, exploring these consequences has been, so far, outside of the scope of household finance papers.

Turning to preference heterogeneity, two related household finance papers on this topic are Gomes and Michaelides (2005) and Vestman (2018).[7] The latter investigates the effects of joint heterogeneity in risk aversion, elasticity of intertemporal substitution (EIS) and participation costs on stock market participation patterns and their connection with home-ownership. However, the author does not consider the risky share, which is the main focus of our paper. The former work, instead, utilizes heterogeneity in risk aversion and EIS to explain participation and risk-taking of the poor. Our model rationalizes the same targets via skewed labor income shocks and instead fits portfolio allocations on the top of the wealth distribution via preference heterogeneity.[8]

Several papers in the wealth inequality literature that explain moments of the wealth distribution and marginal propensities to consume apply preference heterogeneity. Krusell and Smith (1998), Krueger et al. (2016) and Hubmer et al. (2021) all use a stochastic process that generates heterogeneity in discount factors to improve the fit of the wealth distribution. Carroll et al. (2017) and Aguiar et al. (2024) show that preference heterogeneity also helps in bringing the distribution of marginal propensities to consume and consumption growth more in line with the data. These papers, however, usually calibrate the amount of heterogeneity directly by matching these moments of interest. We instead show that information contained in the three portfolio choice schedules previously described can identify a rich heterogeneity in preference parameters without targeting any moment related to the wealth distribution. Moreover, we document that a good fit of wealth inequality (even at the very top), marginal propensities to consume and invest and wealth mobility can be achieved as a by-product of this approach. By

(2021) illustrates that cyclical skewness of labor income shocks can generate the increasing risky share over the life-cycle found in the data with reasonable participation costs. Empirically, using Swedish registry data, Catherine et al. (2024) document that workers facing higher cyclical skewness display lower risky shares.

[7] In another recent contribution, Ebrahimian and Sodini (2025) document that individual preferences are correlated with parental background, which, in turn, limits intergenerational mobility.

[8] Our specification also allows for heterogeneity in the elasticity of intertemporal substitution (EIS). However, as discussed below in Section 3, in this paper we abstract from it because the lack of differences in liquidity between assets in our framework does not allow us to identify well this parameter (Aguiar et al., 2024). Nevertheless, we conduct a sensitivity analysis to see how our results change for different values of EIS.

doing so, we show that various earlier studies in household finance and macroeconomics that explain portfolio choices or wealth inequality, respectively, through preference heterogeneity, are consistent with each other and can be rationalized by the same degree of preference heterogeneity, which increases the plausibility of the latter as an underlying explanation.[9] Furthermore, we illustrate how heterogeneity along several preference parameters at the same time is crucial to achieve this result. Finally, as we use our framework to structurally estimate the parameters governing preference heterogeneity, we also relate to the emerging household finance literature in this area (Calvet et al., 2021).

Within the literature on wealth inequality, several papers theoretically investigate the implications of return heterogeneity. Benhabib et al. (2011, 2015) and Nirei and Aoki (2016) show in different settings that capital income risk generates a stationary wealth distribution with a fat tail. Gabaix et al. (2016) further illustrate that heterogeneous returns created by either *type*- or *scale*-dependence are also needed to generate the rapid dynamics in the right tail of the distribution observed in the data. Benhabib et al. (2019), Hubmer et al. (2021) and Gomez (2023) investigate the quantitative implications of return heterogeneity and find that it plays a large role for the shape and dynamics of the wealth distribution. Our main contribution with respect to this literature is endogenizing the portfolio allocation decision, as in these papers return heterogeneity is either hard-wired or purely due to luck. More specifically, we show how a non-standard income process and *type*-dependence through ex-ante differences in preference parameters across agents are enough to generate realistic portfolio schedules and, through that, also a wealth distribution with properties in line with the data. Additionally, we document that explaining both the consumption/saving and portfolio allocation decisions endogenously matters for the transmission of exogenous shocks.

There also exists a literature that attempts to rationalize the increasing risky share over the wealth distribution. This is achieved via non-homothetic preferences by Carroll (2000), Wachter and Yogo (2010) and Cioffi (2021): while the exact mechanisms differ, all these papers endogenously produce a lower risk aversion for richer agents, generating a positive correlation between wealth and the optimal risky share. Alternative explanations include experience-based learning of expected stock returns (Foltyn, 2020), and a crowding-out channel driven by the effect of human capital on optimal housing decisions (Rácz, 2024). We view our contributions to this literature as complementary: we explore preference heterogeneity as an alternative mechanism that is able to produce an increasing risky share while generating a wealth distribution consistent

[9] Support for preference heterogeneity is motivated not only by the fact that this feature allows models to match several important empirical moments, but also by substantial empirical micro evidence. For instance, Lawrance (1991) and Epper et al. (2020) empirically document heterogeneity in time discounting and von Gaudecker et al. (2011) find heterogeneity in risk preferences using an experiment in a representative sample of Dutch respondents.

with the data. Importantly, none of these papers generates increasing risk-taking over a realistic wealth distribution that also exhibits empirically plausible mobility patterns. Our model achieves the latter through idiosyncratic risk in returns, which has been shown to be crucial for obtaining realistic dynamics of wealth inequality by the aforementioned theoretical literature.

The paper is structured as follows. Section 2 outlines the model, Section 3 describes our estimation procedure and results, Section 4 presents the model-implied wealth distribution, wealth mobility and marginal propensities to consume and invest, Section 5 investigates counterfactual specifications, Section 6 reports the dynamic implications of an MIT shock to aggregate returns and Section 7 concludes.

2 Model

In this section we outline the model framework. Our setting is a partial equilibrium economy in the spirit of Bewley (1977), to which we add a non-normal return process, cyclical skewness in labor income shocks, ex-ante heterogeneity in individual-specific parameters and endogenous portfolio choice between a risky and a safe asset. The result is a hybrid between the Bewley-type models used in macroeconomics to study inequality and the household finance models used to study portfolio allocation (e.g., Cocco et al., 2005).

Agents and preferences. The economy is populated by a continuum of infinitely lived individuals deriving utility from consumption $c_{i,t}$ through Epstein-Zin preferences (Epstein and Zin, 1989). Agents are ex-ante different in terms of preference parameters: δ_i captures their impatience, γ_i their risk aversion and ψ_i the inverse of their elasticity of intertemporal substitution.[10] Preferences are, then, given by the following expression:

$$U_{i,t} = \left[(1-\delta_i) c_{i,t}^{1-\psi_i} + \delta_i \left(\mathbb{E}_t U_{i,t+1}^{1-\gamma_i} \right)^{\frac{1-\psi_i}{1-\gamma_i}} \right]^{\frac{1}{1-\psi_i}}.$$

We assume that preference parameters are fixed over time.[11] When parameterizing the model, we will allow for finitely many types of individuals, where a type is defined by a combination of preference parameters.

[10] The next paragraph provides a detailed explanation of another individual-specific parameter, ζ_i, which governs agents' heterogeneity in terms of idiosyncratic return risk.

[11] We experimented with a version of the model in which switches across types were governed by a slow moving Markov chain. As the key results were very similar, we decided to remove this feature in order to reduce the number of free parameters.

Financial assets. Agents can invest in two financial assets, one risky, $k_{i,t}$, with time-varying individual-specific gross return $R_{i,t+1}$, and one safe, $d_{i,t}$, with constant gross return R^f. Investing in the risky asset is subject to a participation cost f that is paid in every period the agent chooses to hold that asset. Apart from adding idiosyncratic return risk to the framework, we largely follow Catherine (2021) in modeling risk present in stock returns and labor income shocks and most importantly, the connection between the two. Letting small letters indicate log returns, $r_{i,t+1}$ is equal to:

$$r_{i,t+1} = r_{1,t+1} + r_{2,t+1} + \eta_{i,t+1} - m.$$

The effective return individual i gets by investing in the risky asset is the sum of a systematic component and an idiosyncratic component η_i, net of a management fee m. The systematic part consists, in turn, of a component co-varying with labor market conditions, r_1, and a component that is independent of labor market conditions, r_2. The former variable is distributed as follows:

$$r_{1,t+1} = \begin{cases} \underline{\mu}_r & \text{w.p.} \quad p_r \\ \overline{\mu}_r & \text{w.p.} \quad 1 - p_r \end{cases}.$$

Without loss of generality, we interpret p_r as the probability of stock market crashes and $\underline{\mu}_r$ the log return during these periods. Similarly, $1 - p_r$ is the probability of normal periods and $\overline{\mu}_r$ the corresponding log return. As explained in Section 3.1, the presence of stock market crashes and their effect on the distribution of labor income shocks play a key role in generating realistic portfolio shares for the poor. The other systematic component, r_2, is drawn from a Normal distribution:

$$r_{2,t+1} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_{r_2}^2\right).$$

We add an idiosyncratic component $\eta_{i,t+1}$ to the return process to understand the relative importance of systematic and idiosyncratic return shocks. While it has been established that under-diversification is a prominent feature of household portfolios, its causes are not well understood.[12] Therefore, instead of microfounding why some households diversify more than others, we take a shortcut and assume each agent is endowed with an exogenous level of idiosyncratic return risk ζ_i, which is determined by the type of the agent. We then model the idiosyncratic component, $\eta_{i,t+1}$, as follows:

$$\eta_{i,t+1} \overset{i.i.d.}{\sim} \mathcal{N}\left(-\frac{\sigma_{ir}^2}{2}, \sigma_{ir}^2\right)$$

where $\sigma_{ir} = \sigma_r \zeta_i$. The term σ_r denotes the standard deviation of the systematic part of the log return, and the individual (type-specific) parameter ζ_i governs the share of idiosyncratic risk in

[12]Possible explanations include differences across individuals in terms of preferences, financial illiteracy, overconfidence and reliance on private equity. See Chapter 4.2 in Guiso and Sodini (2013) for a survey of the related empirical and theoretical literature.

total portfolio volatility.

With this modeling choice, rather than having access to the same risky asset, each individual rationally invests in her own risky asset, which has identical expected excess return as the market, but additional type-specific idiosyncratic risk. While guaranteeing that idiosyncratic risk is not priced, this strategy captures – in line with the empirical findings in Calvet et al. (2007) – that agents worse at diversifying will, everything else equal, optimally choose a lower risky share and vice versa. In other words, this specification captures in reduced form that agents have a heterogeneous ability or desire to diversify, which they take as given when optimally deciding how to allocate their wealth.

Labor income. Individual log earnings, $y_{i,t}$, are the sum of an aggregate component, w_t, and two idiosyncratic components, one persistent, $z_{i,t}$, and one transitory, $\nu_{i,t}$:

$$y_{i,t} = w_t + z_{i,t} + \nu_{i,t}.$$

The aggregate component follows a random walk with drift, driven by shocks to the market return through a parameter λ_{rw}:

$$w_t = g + w_{t-1} + \lambda_{rw} r_{1,t} + \phi_t$$

where $\phi_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_\phi^2\right)$.

The persistent component is an AR(1) process:

$$z_{i,t} = \rho z_{i,t-1} + \varepsilon_{i,t}$$

with innovations drawn from a mixture of Normals:

$$\varepsilon_{i,t} = \begin{cases} \underline{\varepsilon}_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(\underline{\mu}_{\varepsilon,t}, \underline{\sigma}_\varepsilon^2\right) & \text{w.p.} \quad p_\varepsilon \\ \overline{\varepsilon}_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(\overline{\mu}_{\varepsilon,t}, \overline{\sigma}_\varepsilon^2\right) & \text{w.p.} \quad 1-p_\varepsilon \end{cases}.$$

Without loss of generality, we interpret p_ε as the probability of tail events and $\underline{\mu}_{\varepsilon,t}$, $\underline{\sigma}_{\varepsilon,t}$ the expected value and standard deviation of persistent income shocks during tail events, respectively. A similar interpretation holds for the parameters governing the distribution of normal events. To match the cyclicality of the skewness of labor income shocks, $\underline{\mu}_{\varepsilon,t}$ is defined as:

$$\underline{\mu}_{\varepsilon,t} = \mu_\varepsilon + \lambda_{\varepsilon w}(w_t - w_{t-1}).$$

Thus, tail events imply on average higher persistent shocks during expansions and vice versa during recessions. In addition, since these shocks have zero mean, it must hold:

$$p_\varepsilon \underline{\mu}_{\varepsilon,t} + (1-p_\varepsilon)\overline{\mu}_{\varepsilon,t} = 0.$$

Finally, the transitory shocks follow a Normal distribution:

$$\nu_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_\nu^2\right).$$

The framework for earnings and returns shares common features with Catherine (2021). As discussed in detail in his paper, including countercyclical income risk through cyclical skewness in the distribution of individual income shocks enables to obtain realistic portfolio choices over the life-cycle. Intuitively, if adverse income shocks occur with greater probability when the stock market crashes, agents with relatively high human capital – the young and the poor – will be more cautious when investing in risky financial instruments. We retain this feature in our framework to examine whether a mechanism allowing to match portfolio choices over age can also produce realistic risk-taking patterns over the wealth distribution. However, we also deviate from his setup in several ways. First, we use these stochastic processes in an infinite horizon rather than a life-cycle model. Second, as explained above, we add to the return of the risky asset an idiosyncratic component, $\eta_{i,t}$, to capture differences in idiosyncratic return risk. Third, to enable tight identification in Swedish data via reducing the number of free parameters, some elements of the model in Catherine (2021) were simplified, as detailed in Appendix C. These modifications do not change the main implications of the framework.

Safety net and taxes. In order not to overestimate the importance of persistent negative shocks, we introduce a safety net program that replicates in a parsimonious way the main characteristics of the Swedish social assistance system. The latter has been in place since the 1950s, and its main purpose is to provide a subsistence level of income to people in need.[13] Expressing such subsistence level as a fraction b_{rate} of the aggregate income, we model the total benefits received by individual i in period t, $b_{i,t}$, as follows:

$$b_{i,t} = \max\left\{0, b_{\text{rate}} \cdot \exp(w_t) - \exp(y_{i,t})\right\}.$$

Individuals pay three separate taxes in the model: a tax on labor income, capital income and wealth. For each tax, we assume that the same tax rate is applied to all individuals. We label such rate τ_{labor} for labor income, τ_{capinc} for capital income and τ_{wealth} for wealth.

The optimization problem. At the beginning of each period t, an individual enters with a given cash-on-hand $a_{i,t}$, persistent income $z_{i,t}$ and aggregate income w_t. She then chooses how much to consume in the current period, $c_{i,t}$, and how much to save through the risky asset, $k_{i,t}$, and through the safe asset, $d_{i,t}$. Conditional on investing in the risky asset – indicated by $F_{i,t}$, a dummy equal to one if the individual participates – she pays the fixed participation cost f. Let $\theta_i := (\delta_i, \gamma_i, \psi_i, \zeta_i)$ denote the collection of individual-specific parameters, $\Xi_{i,t} := (a_{i,t}, z_{i,t}, w_t; \theta_i)$ the state, $R^f := \exp(r^f)$ the gross risk free return and

[13] We refer the interested reader to Bastani and Lundberg (2017) for a more detailed description of the Swedish social assistance system.

$R_{i,t+1} := \exp(r_{i,t+1})$ the gross risky return. Then, the maximization problem of agent i is:

$$V(\Xi_{i,t}) = \max_{c_{i,t},k_{i,t},d_{i,t}} \left\{ (1-\delta_i) c_{i,t}^{1-\psi_i} + \delta_i \left(\mathbb{E}_t \left[V(\Xi_{i,t+1})^{1-\gamma_i} \right] \right)^{\frac{1-\psi_i}{1-\gamma_i}} \right\}^{\frac{1}{1-\psi_i}}$$

subject to

$$\begin{aligned} a_{i,t} &= c_{i,t} + k_{i,t} + d_{i,t} + F_{i,t} f \cdot \exp(w_t) \\ a_{i,t+1} &= (1-\tau_{\text{wealth}})(k_{i,t} + d_{i,t}) + (1-\tau_{\text{capinc}}) \Big[(R^f - 1) d_{i,t} + (R_{i,t+1} - 1) k_{i,t} \Big] + \\ &\quad + \left(\exp(y_{i,t+1}) + b_{i,t+1} \right) (1 - \tau_{\text{labor}}) \\ k_{i,t} &\geq 0 \\ d_{i,t} &\geq \bar{d} \cdot \exp(w_t). \end{aligned}$$

The last two inequalities capture that agents cannot short-sell the risky asset, and that the participation cost and the borrowing limit on the safe asset vary over time through the dependence on the aggregate part of labor income w_t.[14] We describe in detail in Appendix B how we solve the model numerically.

3 Estimation

One of the contributions of this paper is to structurally estimate the model presented in the previous section. In doing so, we target characteristics of aggregate income and returns, individual earnings, and portfolio choice patterns over the wealth distribution in Swedish data. In particular, we show that matching portfolio choice schedules over wealth is crucial to estimate the individual specific parameters included in θ_i.

Similarly to Catherine (2021), the estimation procedure follows a sequential structure. First, we estimate the processes governing aggregate income, returns, and individual earnings. Taking the results of the first step as given, we then estimate the participation cost, the borrowing limit, the share of each of the two types in the population and the individual-specific parameters in θ_i.

Each estimation step relies on a Simulated Method of Moments (SMM) routine. The SMM estimate is a vector of parameters that minimizes deviations between moments in the data and their respective counterparts generated by the model. Our estimation algorithm includes a global and a local stage and it is similar to the one outlined in Arnoud et al. (2019). In the global stage, we generate a large Sobol sequence of parameter vectors and compute the objective function for every vector of the sequence. In the local stage, we take the best candidates from the global stage

[14] This assumption makes the value function homogeneous with respect to w_t, which allows us to reduce the dimensionality of the problem by one.

as initial starting points to perform a local optimization for each of them using a local search algorithm (Nelder-Mead), again minimizing the objective function.

Estimating the income and return processes requires simulating them repeatedly. We closely match key moments in Swedish data and find that the effects of cyclical skewness are comparable to those found by Catherine (2021). Instead, the estimation of the remaining parameters requires solving the individual optimization problem and computing the resulting stationary distribution repeatedly. Since the main contribution of our estimation exercise is to obtain structural estimates of the parameters governing ex-ante individual heterogeneity, we outline below the details of the latter procedure. We report the full estimation results for the income and return processes in Appendix C, and refer the interested reader to Appendices A and C for, respectively, a description of the data and additional technical details related to the estimation technique and our mapping between the model variables and the data.

To ensure that the estimation of preference and idiosyncratic return risk parameters and of their heterogeneity from the observed portfolio choices over the wealth distribution remains feasible and identified, we proceed as follows. First, we assume that the economy is populated by two types of individuals. As shown later, two types turn out to be sufficient to match all targeted moments well. Second, although the outlined model features heterogeneity in patience (δ), risk aversion (γ), inverse EIS (ψ), and idiosyncratic return risk (ζ) simultaneously, in all the results reported from here onwards, we set ψ to unity for both types unless otherwise stated. Indeed, while the different role of ψ from that of risk aversion γ is captured through the adoption of Epstein-Zin preferences, joint identification of EIS and impatience is problematic in a model without liquidity differences across assets, as highlighted by Aguiar et al. (2024).[15] Lacking this dimension of heterogeneity in our model, we set ψ to one and investigate the sensitivity of our results with respect to different values of this parameter in Appendix F.2. With the EIS fixed at unity, the Epstein-Zin preferences allow us to isolate the role of heterogeneity in risk aversion, which turns out sufficient to explain the targets.

Summing up, we estimate a total of nine parameters: the preference parameters δ and γ, the idiosyncratic return risk parameter ζ (all by type), the share of individuals of each type in the population, the participation cost f and the borrowing limit $\bar{d}$.

Targets. Bach et al. (2020) report key portfolio choice characteristics over the Swedish wealth distribution, including (unconditional) expected excess returns, participation in risky assets and the share of idiosyncratic return risk. These portfolio choice patterns over the wealth distribution – covering both the intensive and extensive margin of the investment allocation choice – form the

[15]The main idea behind identification in their framework is that low EIS agents care more about consumption smoothing and, as a consequence, their saving decision is tilted more towards liquid assets.

basis of our estimation targets.

While a detailed description can be found in their paper, for our purposes it is worth reminding that the authors use administrative sources covering the wealth holdings all Swedish residents for the period 2000-2007, and that such holdings consist of cash, pension wealth, financial securities (including funds, stocks, derivatives, and bonds), private equity, real estate wealth and debt. The measure of wealth we will refer to throughout the paper is net wealth, defined – as they do – as the sum of all wealth holdings within the household minus debt.[16]

We rely on the micro data in Bach et al. (2020) to compute portfolio choice characteristics applicable to our model setting. When deciding how to allocate their savings – our measure of wealth in the model – individuals choose between a safe and a single composite risky asset. To map expected excess returns, participation and the share of idiosyncratic risk by asset type in the data into those of a composite risky asset, we proceed as follows. First, for every household we compute their wealth holdings in financial wealth, private equity and commercial real estate. Second, we classify the different assets as safe and risky: cash and money market funds belong to the former group, while all other securities, private equity and commercial real estate to the latter. As in Bach et al. (2020), the expected excess return and share of idiosyncratic variance are then computed for each household by weighting the asset-specific expected excess return and risk by the portfolio share of that asset. Participation is, instead, a binary variable equal to unity if the household holds any risky asset according to the above definition. Finally, an average value of these three quantities for different wealth groups is obtained by running a regression including indicator variables for the desired net wealth bins.

For the main estimation, we classify pension wealth and residential real estate as neither safe nor risky, and disregard these assets for the computation of expected returns, participation and idiosyncratic risk (but not for the relative wealth rank). The reason for this choice is that, as the aim is inferring households' preference parameters from observed portfolio choices, it is best to remain agnostic towards asset classes where the observed risky share is most likely determined by other factors. In particular, the asset composition of pension accounts is heavily affected by government regulations and nudging, while the housing share of one's portfolio might be more indicative of preferences in terms of where to live and frictions in the rental market than optimal exposure to real estate price fluctuations. Computing portfolio schedules over total wealth without these two assets implies that the risky part in each of them is equal to the one found for the restricted portfolio including only financial wealth, private equity and commercial

[16] The individual earnings process is instead estimated using individual-level data for males. While this is mainly due to data availability constraints, Busch et al. (2022) find that households do not seem to be able to use their spouse's earnings to insure against the higher downside risk in recessions, essentially because they both face the same labor market conditions.

real estate. We consider pension wealth and residential real estate as safe assets in Appendix F.[17]

Figure 1 displays the resulting schedules of the expected excess returns, participation and share of idiosyncratic risk over the wealth distribution. As in Bach et al. (2020), wealthier households are more likely to hold risky assets, invest a higher share of their wealth in those risky assets and load their portfolios with more idiosyncratic risk than poorer households. These three schedules (by wealth quantile) constitute our estimation targets, together with the ratio of aggregate wealth to income and the share of households with negative wealth – which in Sweden are, respectively, equal to four (as reported by Bach et al., 2018) and approximately 8.3% – for a total of 41 moments. We match the wealth-to-income ratio to ensure that the relative importance of labor income and wealth is comparable to that in the data. The share of households with negative wealth pins down the borrowing limit. We postpone the discussion of the identification of the preference parameters and the participation cost from the portfolio choice moments to Section 3.1 when inspecting the model mechanisms.

Externally calibrated parameters. Before presenting the estimation results, we describe how we choose the remaining parameters needed to solve the model. We set b_{rate} equal to the ratio of the subsistence income threshold and aggregate income.[18] Between 2000 and 2007, this ratio averaged to 0.184. In the same period, capital income was taxed at a flat rate of 30 percent and wealth was taxed at 1.5 percent (Henrekson and Stenkula, 2015). Thus, we set τ_{capinc} and τ_{wealth} equal to these two numbers, respectively. The average labor income tax – computed as the sum of the average municipal and central rates reported in Bastani and Lundberg (2017) – was about 32 percent, which is then the number we use for τ_{labor}. The risk-free rate, r^f, is set to be equal to the average yearly real yield of Swedish Treasury bills, which is 2.9 percent between 1984 and 2016. Following Catherine (2021), the portfolio management cost m is set to 1 percent. Table 1 lists all externally calibrated parameters. One period in the model corresponds to one year.

[17] Quantifying the exact share of risky and underdiversified pension and residential real estate wealth is an empirical task entailing phenomenal data requirements, which we leave to future research.

[18] We use monthly subsistence income thresholds for single individuals without children and annualize them. Tables with the subsistence levels can be found on the website of the Swedish National Board of Health and Welfare (Socialstyrelsen): https://www.socialstyrelsen.se/kunskapsstod-och-regler/omraden/ekonomiskt-bistand/riksnormen.

Parameter	Description	Value	Source
b_{rate}	subsistence income rate	0.184	computed from Socialstyrelsen and Hammar et al. (2022) data
τ_{capinc}	capital income tax	0.3	Henrekson and Stenkula (2015)
τ_{wealth}	wealth tax	0.015	Henrekson and Stenkula (2015)
τ_{labor}	labor income tax	0.32	Bastani and Lundberg (2017)
r^f	risk-free rate	0.029	computed from Sodini et al. (2020) data
m	management cost	0.01	Catherine (2021)

Table 1: Externally calibrated parameters.

Estimation results. The estimated parameters, reported in Table 2, imply a stark separation between the two types of agents. Type-two agents discount the future less strongly than type-one agents (δ of 0.966 vs. 0.924), are less risk averse (γ of 2.671 vs. 9.891), and feature a lower degree of idiosyncratic return risk. In particular, the estimated values for portfolio diversification imply that the share of idiosyncratic variance in total return variance for type-two agents is 46 percent ($\zeta = 0.825$), whereas it is 27 percent for type-one agents ($\zeta = 0.543$). It is noteworthy that while the discount factor and risk aversion of type-two agents are closer to those usually adopted in the macro literature on wealth inequality, type-one agents are characterized by higher values for these two parameters, a combination more often used in the household finance literature. Interestingly, only a relatively small fraction – approximately 10 percent – of type-two agents (but, as we will see below, holding most of the wealth in the economy) is needed for the model economy to replicate all targets.[19]

[19]The lower risk aversion and idiosyncratic return risk and the higher patience found for type-two agents resemble characteristics of entrepreneurs, which is a complementary explanation for the observed portfolio choice patterns (Cagetti and De Nardi, 2006; De Nardi and Fella, 2017). However, the evidence provided by Bach et al. (2020) shows that expected excess returns and the share of idiosyncratic variance are also increasing in wealth when considering only financial wealth portfolios, and that even among households at the top 1%-0.5% of the wealth distribution, the majority does not own any privately listed firms. This suggests that the portfolio choice patterns found in the data – Figure 1 shows that excess returns and idiosyncratic risk increase over the whole range of the wealth distribution – cannot be fully explained by the narrative of entrepreneurs investing in their own business.

Estimated parameters

	Type 1			Type 2					
	δ	γ	ζ	δ	γ	ζ	Share of Type 1	f	$\bar{d}$
Estimate	0.924	9.891	0.543	0.966	2.671	0.825	0.901	0.001	−0.835

Moments (excl. portfolio choice targets)

	Wealth/Income	Share with neg. wealth
Data	4.000	0.083
Model	3.996	0.083

Table 2: Estimated model parameters. We solve the model assuming that there are only two types of individuals and setting inverse EIS (ψ) equal to one. The estimation procedure targets the aggregate wealth to income ratio, the share of households with negative wealth and the schedules of expected excess returns, participation and share of idiosyncratic return variance over the wealth distribution computed from Swedish administrative data and reported in Figure 1.

Finally, the stock market participation cost f is estimated at 0.001. Using the average of our measure of aggregate income over the period 2000-2007, this corresponds to about 263 SEK in 2021 terms (about $30) per year.

Figure 1 shows the model fit of the targeted portfolio choice moments. Expected excess returns, participation and the share of idiosyncratic variance are increasing over the wealth distribution, quantitatively in line with the data. The following section inspects how the model (henceforth labelled benchmark) matches the portfolio choice schedules so well for the estimated parameters.

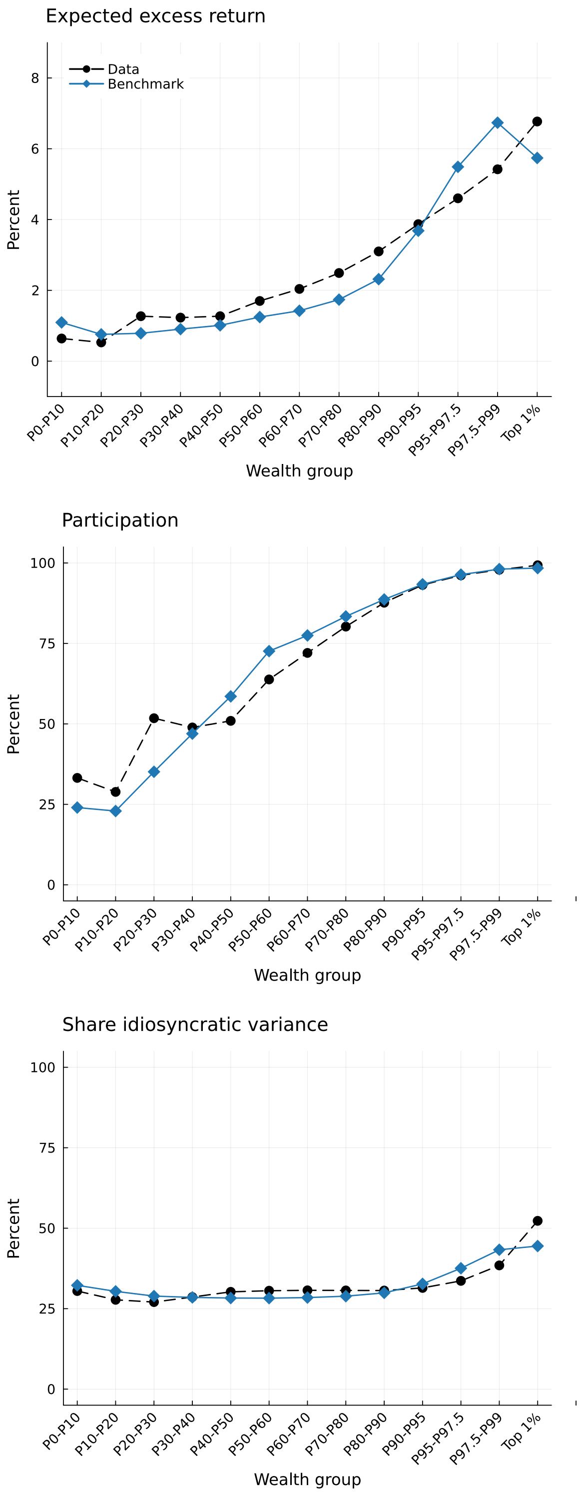

Figure 1: Fit of the estimated model. The figure shows the schedules of expected excess returns, participation and the share of idiosyncratic return variance over the wealth distribution from the model and their data equivalents computed from the Swedish administrative data used by Bach et al. (2020). Safe assets include cash and money market funds. Risky assets include all other securities, private equity and commercial real estate. The risky share within pension wealth and residential real estate is assumed to be the same as the one found for the restricted portfolio including only financial wealth, private equity and commercial real estate. The share of idiosyncratic return variance is computed on participants.

3.1 Inspecting the mechanism

In this section, we describe in detail how the two key ingredients in our model – cyclical skewness of labor income shocks and ex-ante heterogeneity in preferences and idiosyncratic return risk – help to explain the schedules of expected excess returns, participation and the share of idiosyncratic variance presented in Figure 1.

The role of cyclical skewness can be understood by comparing the policy functions for the risky share over cash-on-hand in the estimated model against a model where tail shocks in the persistent component of earnings are shut off.[20] As shown in the left panel of Figure 2, in contrast to the benchmark specification, the risky share in the model without tail risk is decreasing.

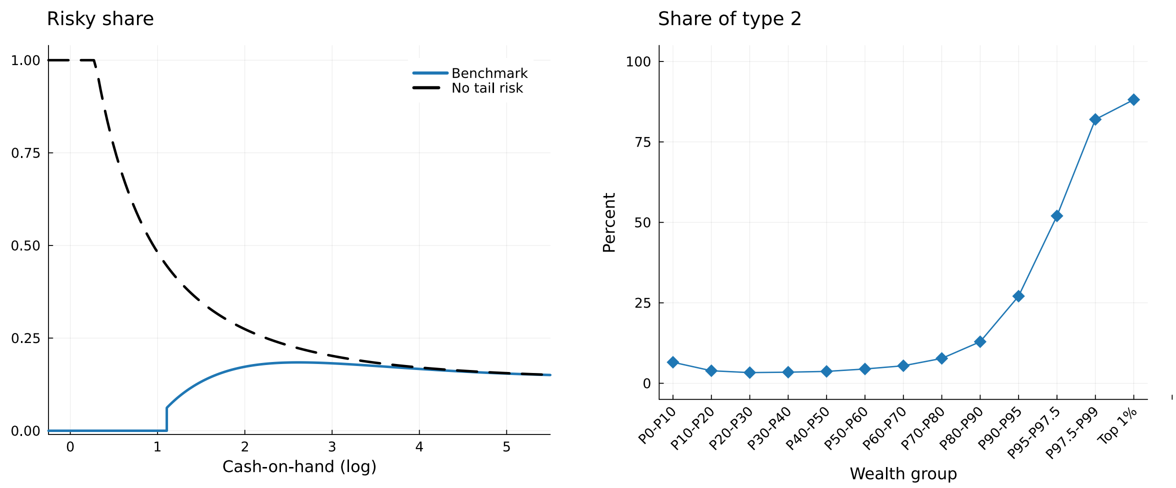

Figure 2: Key model mechanisms. The left panel compares the risky share in the benchmark model and in a model without tail risk in earnings shocks (i.e., we set $\mu_\varepsilon = \lambda_{\varepsilon w} = 0$ and the standard deviation of the persistent part of earnings shocks equal to the unconditional standard deviation of the mixture of Normals). The right panel shows the share of type-two individuals in a given wealth bracket.

There are two well known factors (see, e.g., Campbell and Viceira, 2002) shaping the optimal choice of the risky share: the human capital-wealth ratio and the extent to which human capital has bond-like properties (i.e., should human capital be considered more similar to a safe or risky asset). This follows from the fact that optimal consumption is a function of such ratio, so its level and riskiness matter for consumption smoothing. In a model where labor income is bond-like – such as the version of our framework without tail risk or the main specification in Cocco et al. (2005) – asset poor agents whose total wealth mainly consists of human capital invest fully in the risky asset as shown in the left panel of Figure 2. As first shown by Catherine (2021)

[20]To facilitate the comparison we use a log scale for the x-axis and we solve both models with the parameters estimated in the previous section, except for the borrowing limit, which we set to zero. Furthermore, in the model without cyclical skewness we set $\mu_\varepsilon = \lambda_{\varepsilon w} = 0$ and the standard deviation of the persistent part of earnings shocks equal to the unconditional standard deviation of the mixture of Normals governing $\varepsilon_{i,t}$.

in the context of a life-cycle model, cyclical skewness in labor income shocks – that is, a higher probability of receiving negative labor income shocks when the aggregate return component of the risky asset is low and vice versa – overturns this result. As labor income already features properties of the risky asset, young agents shy away from investing in the latter. Applied to our specification, this mechanism refrains asset poor agents from investing in the risky asset and makes them raise their risky share only gradually with wealth, which delivers a positive relation between these two quantities. Nevertheless, this channel alone has difficulties matching the increase in the empirical risk-taking patterns over the whole wealth distribution. Intuitively, as agents accumulate wealth to self-insure against shocks, their human capital-to-wealth ratio declines and the properties of their earnings process become irrelevant for their portfolio choice. Consequently – as clearly shown in the left panel of Figure 2 – in both models the risky share converges to Merton's constant (Merton, 1969).

Turning to ex-ante heterogeneity in preferences and idiosyncratic return risk across individuals, this feature is crucial to match portfolio choices at the top of the distribution. Indeed, as shown in the right panel of Figure 2, the lower degree of risk aversion and higher patience of type-two agents induce them to optimally choose higher savings and a higher risky share, which, as a result, endogenously lead them to the top of the wealth distribution. As the share of this type of agents increases gradually until the very top of the distribution – where these individuals constitute the vast majority of the population – the resulting compositional effects deliver increasing excess returns and participation even at the top of the wealth distribution.[21] Additionally, as these agents are characterized also by higher idiosyncratic risk, this mechanism also allows us to match the empirical schedule of the share of idiosyncratic return variance.

Finally, it is worth noting that the high-risk aversion of type-one individuals combined with the effect of countercyclical income risk implies that even a small value of the participation cost has a sufficient deterring effect on participation. As a result, the estimated participation cost is smaller than other values in the literature (e.g., Vissing-Jorgensen, 2003).

4 Implications for the wealth distribution

The endogenous sorting of the two types suggests that portfolio choice heterogeneity has implications for the wealth distribution. In this section, therefore, we assess how well our benchmark model matches wealth inequality and mobility in the data. To confirm that the model delivers reasonable predictions for other policy relevant measures, we also show that

[21] The reason why the excess return is slightly lower for the top 1% compared to the P97.5-P99 group is that when type-two individuals almost fully populate a certain part of the distribution, the compositional effect on the average risky share is weaker.

our framework generates realistic marginal propensities to consume (MPCs) and invest (MPIs). Importantly, none of these quantities has been targeted in the estimation.

Wealth inequality. Table 3 presents the model fit of the wealth distribution, contrasting the wealth held by different groups in the model and data. We compare our results with the corresponding values in the Swedish administrative data averaged over the period 2000-2007.[22]

	Model				Data
Share held by (%):	Benchmark	One type	Only δ	No idio. ret.	Sweden (2000-2007)
Q1	0.1	0.3	0.1	0.2	-1.1
Q2	4.6	6.6	2.5	5.8	2.8
Q3	8.8	12.8	5.0	11.2	8.7
Q4	15.8	22.3	10.7	19.4	19.4
Q5	70.6	58.0	81.7	63.4	70.2
90-95 %	11.9	12.8	19.5	12.1	13.4
95-99 %	22.5	16.4	29.5	17.4	17.9
Top 1 %	22.5	11.0	16.7	17.8	21.3
Wealth Gini	0.69	0.57	0.76	0.62	0.71

Table 3: Wealth inequality. The table reports the share of wealth held by households between different quantiles of the wealth distribution and the Gini coefficient. "One type" indicates the model with one type of agents, i.e., no heterogeneity in preference parameters and idiosyncratic return risk. "Only δ" the model with two types differing in δ but not in γ and ζ. "No idio. ret." the model without idiosyncratic risk in returns in which we set ζ equal to zero for both types. The last column shows the data equivalents computed in the Swedish administrative data compiled by Bach et al. (2020) – averaged over the 2000-2007 period.

The model closely matches overall wealth inequality, as reflected by the Gini coefficient (0.69 vs. 0.71 in the data) and the shares of wealth held by different groups. Remarkably, the fit is very good even at the top 1%, where standard models usually do not deliver values as high as in the data. We emphasize that although our model features preference heterogeneity, the parameter values for the more patient type are standard in the macroeconomics literature.

Wealth mobility. Table 4 reports the share of agents transitioning across different wealth groups over a period of twenty-five years. We compare these figures with the numbers from Table 3 in Benhabib et al. (2019), which report the probability that a child belongs to a wealth quintile given the wealth quintile of the parent, computed using PSID data.

[22]The wealth measure we use – which is the same as that adopted by Bach et al. (2020) – includes an imputed value for total pension wealth. Without this component, wealth inequality would be higher. This is in line with the findings in Catherine et al. (Forthcoming), where the authors show that properly accounting for social security wealth delivers much smaller changes in the top wealth shares over the last three decades in the United States. We think that using a measure of wealth including pension holdings is more appropriate to compare the output of an infinite horizon model with the data.

	Wealth rank in 25 years				
Wealth rank today	0%-20%	20%-40%	40%-60%	60%-80%	80%-100%
0%-20%	0.30	0.23	0.20	0.16	0.11
	[0.36]	[0.29]	[0.16]	[0.12]	[0.07]
20%-40%	0.24	0.23	0.21	0.18	0.13
	[0.26]	[0.24]	[0.24]	[0.15]	[0.12]
40%-60%	0.21	0.22	0.22	0.20	0.15
	[0.16]	[0.21]	[0.25]	[0.24]	[0.15]
60%-80%	0.17	0.20	0.22	0.22	0.19
	[0.15]	[0.13]	[0.20]	[0.26]	[0.26]
80%-100%	0.08	0.12	0.17	0.22	0.41
	[0.11]	[0.16]	[0.14]	[0.24]	[0.36]

Table 4: Wealth mobility. The table shows the share of agents moving across wealth quintiles of the wealth distribution over a period of 25 years under our benchmark specification. The numbers in square parentheses are taken from Table 3 in Benhabib et al. (2019), where the authors report parent-child intergenerational wealth mobility figures from Table 2 in Charles and Hurst (2003) computed with PSID data. Rows might not sum exactly to one due to rounding.

Our model compares relatively well to the data, with transition probabilities differing only by a few percentage points. Appendix Table 3 in Bach et al. (2020) reports wealth mobility in Swedish registry data, but focuses on top wealth mobility and spans only a period of seven years. Nevertheless, we also compare our model performance against these numbers in Appendix E and find a good match.

Marginal propensities to consume and invest. Figure 3 reports the yearly MPCs over the wealth distribution in our benchmark specification, unveiling a clear decreasing relationship between these two quantities. The average MPC in the model is 0.13, but there is substantial heterogeneity across the distribution, ranging from 0.19 for the poorest wealth group to 0.05 for the top 1%.

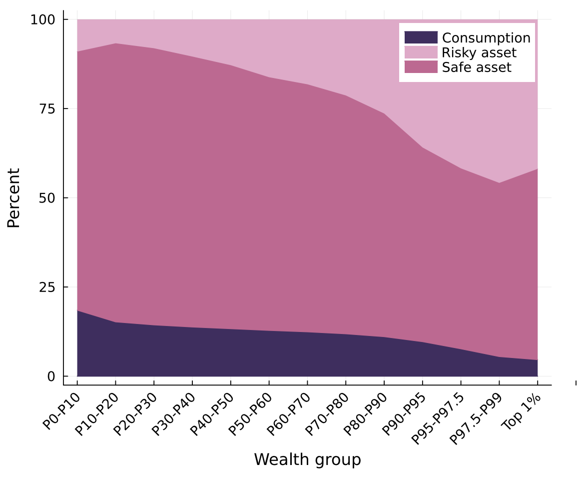

Figure 3: Marginal propensities to consume and to invest. The figure shows the schedules of the marginal propensities to consume, to invest in the safe asset and to invest in the risky asset over the wealth distribution obtained with our benchmark model.

While the presence of more patient type-two individuals mainly at the top of the distribution helps to obtain – in line with the theoretical findings in Kaplan et al. (2018) and Carroll et al. (2017) – a decreasing schedule of MPCs over wealth, our average estimate places itself in the lower end region of the values found in the literature.[23] The key reason behind this result is that in our model the MPCs of agents in the left part of the distribution are relatively low because assets are perfectly liquid. Despite this, we believe that the overall match is satisfying, considering our relatively parsimonious framework.

We report two other interesting patterns in Figure 3, namely the marginal propensities to invest in risky and safe assets over the wealth distribution. The MPI in the safe asset is always larger than that of the risky asset, and it is overall decreasing over the distribution (vice versa for the risky asset). This result is again due to the concentration of more patient and less risk averse type-two individuals at the top.

Taking stock, the previous paragraphs show that our benchmark specification generates wealth inequality, mobility and marginal propensities to consume and invest that match the data relatively well. We interpret the good fit of all these moments (despite not having targeted any of them in the estimation) as a validation of our model mechanisms. While intuition suggests that portfolio heterogeneity matters for wealth inequality, it is not obvious ex-ante that the key features of our model would quantitatively generate a good match also along the other dimensions just mentioned. In the next section, we investigate in detail the contribution of the different model components in jointly fitting these quantities.

5 Quantifying the contribution of model components

The key features of our framework are ex-ante heterogeneity across individuals in preference parameters and idiosyncratic return risk, endogenous portfolio choice between a safe and risky asset and cyclical skewness of labor income shocks. In this section, we quantify the contribution of these elements by estimating counterfactual model economies in which we turn off these additions one at a time. Specifically, we consider a model in which preference and idiosyncratic return risk parameters are the same across agents, a model in which the two types differ only in their discount factor and a model without idiosyncratic risk in returns. We relegate to Appendix D additional results for other counterfactual economies in which, respectively, the two types

[23]Table 1 in Carroll et al. (2017) summarizes empirical estimates of yearly MPCs. Even though there is large heterogeneity, the authors claim that the majority of estimates is between 0.2 and 0.6. On the modeling side, Kaplan et al. (2018) find an average yearly MPC of around 0.3, and Carroll et al. (2017) a range of values roughly between 0.2 and 0.4 depending on the model specification considered.

differ only in their risk aversion, cyclical skewness in labor income risk is turned off or portfolio choices are hard-wired.

5.1 One type of agents

One key component of the model is the rich ex-ante heterogeneity in agents' preferences and idiosyncratic return risk. In the following, we show that this feature is crucial in explaining the targeted moments of portfolio choice over the wealth distribution shown in Figure 1. To this end, we assume that there is just one type of agents, and re-estimate δ, γ, ζ, the participation cost f and the borrowing limit $\bar{d}$, while targeting the same moments as for the benchmark model. One can interpret this exercise as quantifying the importance of type dependence (Gabaix et al., 2016) for our main results.

	Type 1			Type 2					
Model	δ	γ	ζ	δ	γ	ζ	Share of Type 1	f	$\bar{d}$
Benchmark	0.924	9.891	0.543	0.966	2.671	0.825	0.901	0.001	−0.835
One type	0.950	9.229	0.613				1.0	0.006	−1.118
Only δ	0.884	7.629	0.615	0.981			0.859	0.000	−0.287
No idio. ret.	0.939	9.551		0.947	2.310		0.887	0.005	−0.891

Table 5: Estimated model parameters, benchmark model vs. alternative specifications. We solve each model setting inverse EIS (ψ) equal to one. "One type" indicates the model with one type of agents, i.e., no heterogeneity in preference parameters and idiosyncratic return risk. "Only δ" the model with two types differing in δ but not in γ and ζ. "No idio. ret." the model without idiosyncratic risk in returns in which we set ζ equal to zero for both types.

Table 5 reports the parameter estimates for this case. With only one type, the parameter values are in between the numbers obtained in the benchmark case, as the algorithm faces the tradeoff between type-one parameter values to match portfolio choices at the bottom and type-two parameter values to match portfolio choices at the top of the distribution. This is clearly visible from the top panel in Figure 4, which compares the expected excess returns over the wealth distribution in the benchmark model with this alternative specification.[24]

[24]Figure E.1 reports the schedule of participation over wealth for all counterfactuals. The wealth-to-income ratio and the share of agents with negative wealth are almost always perfectly matched when re-estimating the model under all the alternative specifications considered.

Figure 4: Fit of the estimated model, benchmark vs. alternative specifications. The figure shows the schedules of expected excess returns over the wealth distribution in the benchmark model, the different alternative specifications, and their data equivalents computed in the Swedish administrative data used by Bach et al. (2020). The top panel reports the results for the model with one type of agents, i.e., no heterogeneity in preference parameters and idiosyncratic return risk. The middle panel for the model with two types differing only in δ. The bottom panel for the model without idiosyncratic risk in returns in which we set ζ equal to zero for both types.

The reason why the model cannot match the empirical patterns with only one type is the lack of compositional effects at the top of the wealth distribution, as described in Section 3.1. While cyclical skewness – in addition to a higher fixed participation cost – still generates an increasing risky share in the lower part of the distribution, the absence of type-two agents and the ensuing lack of compositional effects prevent the model from generating the same pattern at the top. As a consequence, the risky share plateaus before converging back to Merton's constant. Furthermore, lacking heterogeneity in ζ, the share of idiosyncratic variance is identical for all agents and hence the schedule of this quantity over wealth is flat.

In addition to the effects on portfolio choice, removing preference heterogeneity has further implications. As shown by Krusell and Smith (1998), heterogeneity in patience across individuals increases wealth inequality, since more patient individuals with higher saving rates concentrate at the top of the wealth distribution in the long run. Combined with the lower savings and capital gains implied by the lack of less risk averse and less diversified wealthy agents, the wealth distribution resulting from the model with only one type undershoots the amount of wealth held at the top and overshoots it at the bottom, as shown in Table 3. As a result, the Gini coefficient declines from 0.69 to 0.57. The absence of type-two individuals also does not allow the model to generate an increasing marginal propensity to invest in risky assets at the top, and generates too much wealth mobility among the rich, as shown in the Appendix in Figures E.2 and E.4, respectively.

5.2 Heterogeneity only in patience

To isolate the contribution of heterogeneity in the discount factor, δ, we re-estimate the model, allowing for ex-ante heterogeneity only in this parameter while targeting the same moments as in the benchmark case.

The middle panel in Figure 4 reports the fit of the schedules of expected excess returns, while the effect on the wealth distribution and the estimated parameters are presented in Tables 3 and 5 under the label "Only δ". While the estimated risk aversion and level of idiosyncratic risk are in between the two values obtained in the benchmark specification – as in the case with only one type of agents – the discount factors are more extreme than in the benchmark, namely 0.884 for type-one and 0.981 for type-two. A low discount factor implies both a low savings rate and low participation, so in theory this setup can make the type more willing to invest into stocks concentrate on the top of the wealth distribution. Indeed, as shown by Figure 4, this alternative specification delivers very similar schedules to the benchmark until the P90-P95 group. However, it does a poor job in matching the expected excess returns data at the top: after the wealth group just mentioned, the predicted schedule is counterfactually decreasing in wealth.

Even though this alternative specification still predicts a strong separation between impatient and patient individuals, with the latter populating mostly the top of the wealth distribution, the lack of lower risk aversion for wealthy individuals prevents the model from generating an increasing excess return schedule until the top.

Looking at wealth inequality in Table 3, the results resemble the findings in Krusell and Smith (1998): due to the resulting powerful separation between the two types, the overall Gini coefficient increases – even surpassing the empirical value from Swedish data – but the model underpredicts the share of wealth held at the very top 1%. This highlights the importance of matching the portfolio allocation decisions of the very rich. Figure E.2 displays that heterogeneity in the discount factor also allows to generate a decreasing MPC in wealth, in line with Carroll et al. (2017), but not an increasing MPI in the risky asset for the same reasons as those outlined for the model with only one type. Another important dimension where the "Only δ" does not perform well is wealth mobility. As pointed out by Benhabib et al. (2019), idiosyncratic return risk is an important contributor of top wealth mobility: the fact that in this specification patient and impatient individuals face the same idiosyncratic return risk combined with the higher discount factor of the former type than in the benchmark delivers too low mobility in the right tail of the distribution, as Figures E.3 and E.4 clearly show.

5.3 No idiosyncratic risk in returns

For the purpose of understanding the role of idiosyncratic return risk in shaping the wealth distribution, we evaluate the performance of another counterfactual model in which we eliminate idiosyncratic returns by setting ζ equal to zero.[25] As depicted in the bottom panel of Figure 4, the model-generated policies match the data almost as accurately as the benchmark.

Despite doing a good job in matching the empirical schedule of expected excess returns, this counterfactual specification does not deliver an equally good result as the benchmark model in terms of wealth inequality. As Table 3 shows, too little wealth is concentrated at the top of the distribution compared to the data: the share of wealth held by the top 1% drops from 22.5% to 17.8% and the Gini coefficient from 0.69 to 0.62. Besides idiosyncratic returns directly generating dispersion in wealth, one reason for this result is that – as shown in Table 5 – when not constrained to match the higher share of idiosyncratic risk in the right tail of the distribution, the algorithm requires a less stark type-separation in the estimated discount factors. As a consequence of this, while our results qualitatively line up with the findings in Hubmer et

[25]As there is no idiosyncratic return risk, in addition to the wealth-to-income ratio and the share of agents with negative wealth, in this case we only target the schedules of expected excess returns and participation over the wealth distribution.

al. (2021) – who document a limited impact of idiosyncratic return risk, mainly clustered at the top – compared to them, the effect we find is quantitatively larger. Since the wealth distribution implied by the estimated parameters of this counterfactual is more equal than in the benchmark economy, we also find a limited effect on wealth mobility, as shown in Figures E.3 and E.4.[26]

An important conclusion from this counterfactual is that a good fit of the portfolio choice patterns does not necessarily imply an equally good fit of the wealth distribution.

6 Dynamic implications

Understanding the response of wealth inequality to shocks is crucial for the design of redistributive policies. Since our estimated model captures well key metrics relevant for the transmission of shocks – i.e., the shape of the wealth distribution, wealth mobility and marginal propensities to consume and invest – we present below the response of the share of wealth held by the top 10% and top 1% to an aggregate return shock in our benchmark specification and contrast it to counterfactual models.

We implement the shock as a one-time, unexpected ("MIT") ten percentage points increase in r_2, the component of the aggregate stock market return not correlated with the aggregate component of individuals' labor income.[27] We refer the interested reader to Appendix B for the technical details of the exercise.

Figure 5 presents the results of this exercise. Considering first our benchmark economy, the solid lines in the picture show that the additional capital gains substantially boost inequality, as agents in the right tail of the distribution have a larger share of their wealth invested in the risky asset. The magnitude of the initial responses of about 0.55 and 0.2 percentage points can be compared to the theoretical maximal responses, obtained by assuming that the top 10% (or 1%) of the population holds only stocks while the rest only bonds. Using the wealth distribution generated by our benchmark specification, these upper bounds amount to 2.3 and 1.7 percentage points, respectively, implying that the model produces approximately 23.7% and 11.7% of these figures. The graph also shows that the effect of the shock is overall very long lasting, as it takes

[26] Atkeson and Irie (2022) show that the crucial determinant for high mobility at the top is the higher idiosyncratic return risk in the portfolios of wealthy individuals. However, their statement is conditional on keeping top wealth inequality constant.

[27] While we acknowledge the advantages of general equilibrium, we think that most of our analysis is better suited to a partial equilibrium framework. Matching portfolio choices over the wealth distribution is a relevant exercise only if the joint distribution of idiosyncratic labor income shocks and stock market returns is realistic, and producing the latter as a general equilibrium outcome is outside the scope of this paper. For the given exercise, as r_2 can be interpreted as an external shock to foreign stock markets, studying the effect of an unexpected change in this variable allows us to mitigate concerns related to using a partial equilibrium model to assess aggregate responses.

more than two hundred years for the economy to converge back to the steady state. The persistent response is driven by type-two individuals, whose higher patience and propensity to invest in non-safe holdings fosters additional savings and capital gains.

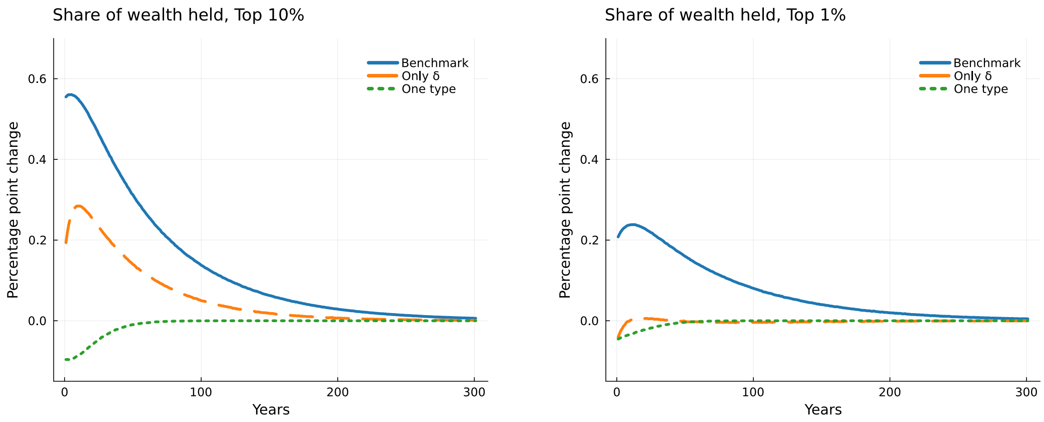

Figure 5: Response to an aggregate return shock, benchmark vs. alternative specifications. The figure shows the response of the share of wealth held by the top 10% and 1% to a ten percentage point increase in r_2. "One type" indicates the model with one type of agents, i.e., no heterogeneity in preference parameters and idiosyncratic return risk. "Only δ" the model with two types differing in δ but not in γ and ζ.

One of the key features of our benchmark model is the ability to generate realistic portfolio choices and wealth concentration levels in line with the data. An interesting exercise, therefore, is to compare the responses of wealth inequality obtained with two counterfactual specifications that do not match well these two quantities, namely the model with only one type of agents ("One type") and the model with heterogeneity only in discount factors ("Only δ").[28] The first case is interesting because – as highlighted in Section 5.1 – that counterfactual economy fails to match both portfolio choices and inequality. Comparing it with the benchmark, therefore, allows us to understand the role of achieving a good fit in these two dimensions at once. The "Only δ" case, instead, allows us to isolate the role of realistic portfolio choices, since that alternative specification delivers high wealth inequality (even higher than in the benchmark) but fails to replicate the investment allocation decisions of the rich – as discussed in Section 5.2.

The responses obtained from these two alternative models are presented in Figure 5 (the dotted lines correspond to "One type" and the dashed ones to "Only δ"). First, compared to the benchmark, the share of wealth held by the top 10% increases much less with heterogeneity only in patience (about 0.2 vs. 0.55 percentage points) and even decreases in the specification with only one type of agents. Second, only the benchmark model generates an increase in the share of wealth held by the top 1%, while both in the "One type" and the "Only δ" settings the response is small and negative on impact. In both these counterfactuals, the risky share starts to decrease in

[28]We compare in Appendix D.3 the responses between the benchmark economy and the model with hard-wired portfolio choices.

wealth at lower quantiles of the the wealth distribution (see Figure 4), so the reduced capital gains of the wealthiest translate into lower inequality compared to the benchmark. In particular, in the case of the "One type" specification the risky share is decreasing in wealth already from the 60th percentile of the wealth distribution, hence we detect a decrease in inequality with both the top 10% and the top 1% measure. The "Only δ" setting however, features a better match of top inequality combined with a decreasing risky share only after the 90th percentile, which enables the model to achieve a more comparable response to the benchmark for the top 10%. However, as at the top of the wealth distribution the portfolio choice patterns are worse matched than in the benchmark, the inequality response in the top 1% is even negative on impact. In addition, all counterfactual responses feature a much lower persistence (despite the higher δ estimated for type-two agents in "Only δ"), as in the benchmark specification the marginal propensity to invest in risky assets is significantly higher for the rich, and thus the additional wealth keeps on generating high returns.

Taking stock, both capturing the distribution of wealth and portfolio choices in the economy is crucial for the sign, magnitude and persistence of the response of wealth inequality to an aggregate return shock. Importantly, generating a high initial inequality in a model is not sufficient to capture these effects, if household investment allocations are not explained well.

7 Conclusion

This paper introduces a macroeconomic angle to recent empirical findings in the household finance literature by developing a quantitative theory that jointly explains key features of the wealth distribution and portfolio choices. Beyond the endogenous decision between a risky and a safe asset, two elements we add to a standard incomplete markets model are cyclical skewness in labor income shocks and ex-ante heterogeneity across individuals in terms of their discount factor, risk aversion and idiosyncratic return risk. The former ingredient rationalizes the increasing relationship between the risky share and wealth at the bottom of the wealth distribution. The latter extends this relationship also to the right tail, where agents are well insured from income shocks. Owing in large part to preference heterogeneity, the model also delivers a very good match of wealth inequality – remarkably even at the very top – wealth mobility, and marginal propensities to consume over the wealth distribution. To illustrate the importance of portfolio choices for macroeconomic outcomes, we analyze how wealth concentration evolves following an aggregate return shock. We find that counterfactual models not capturing the positive relation between risk-taking and wealth imply a less persistent and weaker response in inequality.

Given that our framework replicates well the joint distribution of wealth and portfolio choices and its main characteristics, we see our model as a solid basis for further extensions to address other interesting questions. For instance, while converting our setting to a fully general equilibrium model is a challenging task as it requires explaining the equity premium (Krusell and Smith, 1997), one could endogenize the risk-free rate as in Hubmer et al. (2021). In turn, this would allow quantifying the impact of tax policies – in the form of, for example, changes in wealth or capital income taxes – in a framework where agents optimize both their consumption/saving and investment allocation decisions. We leave these important questions to further research.

References

Aguiar, Mark, Mark Bils, and Corina Boar, "Who Are the Hand-to-Mouth?," *The Review of Economic Studies*, 05 2024.

Arnoud, Antoine, Fatih Guvenen, and Tatjana Kleineberg, "Benchmarking Global Optimizers," Working Paper 26340, National Bureau of Economic Research October 2019.

Atkeson, Andrew G. and Magnus Irie, "Rapid Dynamics of Top Wealth Shares and Self-Made Fortunes: What Is the Role of Family Firms?," *American Economic Review: Insights*, December 2022, *4* (4), 409–24.

Azzalini, Gualtiero and Zoltán Rácz, "MomentMatching.jl: Parameter Estimation of Economic Models via Moment Matching Methods in Julia," `https://github.com/ZoltanRacz/MomentMatching.jl` 2024.

Bach, Laurent, Laurent Calvet, and Paolo Sodini, "From Saving Comes Having? Disentangling the Impact of Savings on Inequality," *Available at SSRN*, 2018.

—, Laurent E. Calvet, and Paolo Sodini, "Rich Pickings? Risk, Return, and Skill in Household Wealth," *American Economic Review*, September 2020, *110* (9), 2703–47.

Bastani, Spencer and Jacob Lundberg, "Political preferences for redistribution in Sweden," *The Journal of Economic Inequality*, 2017, *15* (4), 345–367.

Benhabib, Jess, Alberto Bisin, and Mi Luo, "Wealth Distribution and Social Mobility in the US: A Quantitative Approach," *American Economic Review*, May 2019, *109* (5), 1623–47.

—, —, and Shenghao Zhu, "The Distribution of Wealth and Fiscal Policy in Economies With Finitely Lived Agents," *Econometrica*, 2011, *79* (1), 123–157.

—, —, and —, "The wealth distribution in Bewley economies with capital income risk," *Journal of Economic Theory*, 2015, *159*, 489–515.

— and —, "Skewed Wealth Distributions: Theory and Empirics," *Journal of Economic Literature*, December 2018, *56* (4), 1261–91.

Bewley, Truman, "The permanent income hypothesis: A theoretical formulation," *Journal of Economic Theory*, 1977, *16* (2), 252–292.

Busch, Christopher, David Domeij, Fatih Guvenen, and Rocio Madera, "Skewed Idiosyncratic Income Risk over the Business Cycle: Sources and Insurance," *American Economic Journal: Macroeconomics*, April 2022, *14* (2), 207–42.

Cagetti, Marco and Mariacristina De Nardi, "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, 2006, *114* (5), 835–870.

Calvet, Laurent E, John Y Campbell, Francisco Gomes, and Paolo Sodini, "The Cross-Section of Household Preferences," Working Paper 28788, National Bureau of Economic Research May 2021.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini, "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes," *Journal of Political Economy*, 2007, *115* (5), 707–747.

Campbell, John and Luis Viceira, *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*, Oxford University Press, 2002.

Carroll, Christopher D., "Portfolios of the Rich," NBER Working Papers 7826, National Bureau of Economic Research, Inc August 2000.

— , "The method of endogenous gridpoints for solving dynamic stochastic optimization problems," *Economics Letters*, 2006, *91* (3), 312–320.

Carroll, Christopher, Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White, "The distribution of wealth and the marginal propensity to consume," *Quantitative Economics*, 2017, *8* (3), 977–1020.

Catherine, Sylvain, "Countercyclical Labor Income Risk and Portfolio Choices over the Life Cycle," *The Review of Financial Studies*, 12 2021.

— **, Max Miller, and Natasha Sarin**, "Social Security and Trends in Wealth Inequality," *The Journal of Finance*, Forthcoming.

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

Chang, Yongsung, Jay H. Hong, and Marios Karabarbounis, "Labor Market Uncertainty and Portfolio Choice Puzzles," *American Economic Journal: Macroeconomics*, April 2018, *10* (2), 222–62.

Charles, Kerwin Kofi and Erik Hurst, "The Correlation of Wealth across Generations," *Journal of Political Economy*, 2003, *111* (6), 1155–1182.

Cioffi, Riccardo A., "Heterogeneous Risk Exposure and the Dynamics of Wealth Inequality," 2021. Working Paper.

Cocco, João F., Francisco J. Gomes, and Pascal J. Maenhout, "Consumption and Portfolio Choice over the Life Cycle," *The Review of Financial Studies*, 02 2005, *18* (2), 491–533.

De Nardi, Mariacristina and Giulio Fella, "Saving and wealth inequality," *Review of Economic Dynamics*, 2017, *26*, 280–300.

Ebrahimian, Mehran and Paolo Sodini, "More than Money: The Role of Preferences on Wealth Mobility," *Working Paper*, 2025.

Epper, Thomas, Ernst Fehr, Helga Fehr-Duda, Claus Thustrup Kreiner, David Dreyer Lassen, Søren Leth-Petersen, and Gregers Nytoft Rasmussen, "Time Discounting and Wealth Inequality," *American Economic Review*, April 2020, *110* (4), 1177–1205.

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

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

— **, Luigi Guiso, Davide Malacrino, and Luigi Pistaferri**, "Heterogeneity and Persistence in Returns to Wealth," *Econometrica*, 2020, *88* (1), 115–170.

Foltyn, Richard, "Experience-based Learning, Stock Market Participation and Portfolio Choice," *Available at SSRN*, 2020.

Gabaix, Xavier, Jean-Michel Lasry, Pierre-Louis Lions, and Benjamin Moll, "The Dynamics of Inequality," *Econometrica*, 2016, *84* (6), 2071–2111.

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

— **and Alexander Michaelides**, "Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence," *The Journal of Finance*, 2005, *60* (2), 869–904.

Gomez, Matthieu, "Decomposing the Growth of Top Wealth Shares," *Econometrica*, 2023, *91* (3), 979–1024.

Guiso, Luigi and Paolo Sodini, "Chapter 21 - Household Finance: An Emerging Field," in George M. Constantinides, Milton Harris, and Rene M. Stulz, eds., *Handbook of the Economics of Finance*, Vol. 2, Elsevier, 2013, pp. 1397–1532.

Guvenen, Fatih, "Reconciling conflicting evidence on the elasticity of intertemporal substitution: A macroeconomic perspective," *Journal of Monetary Economics*, 2006, *53* (7), 1451–1472.

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

—, **Luigi Pistaferri, and Giovanni L Violante**, "The Global Repository of Income Dynamics," 2022. https://www.grid-database.org/. Accessed 25.10.2023 and 28.11.2023.

—, **Serdar Ozkan, and Jae Song**, "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 2014, *122* (3), 621–660.

Hammar, Olle, Paula Roth, and Daniel Waldenström, "The Swedish Income Distribution, 1968–2016," *Working Paper*, 2022.

Henrekson, Magnus and Mikael Stenkula, *Swedish taxation: developments since 1862*, Palgrave Macmillan New York, 2015.

Hintermaier, Thomas and Winfried Koeniger, "Differences in euro-area household finances and their relevance for monetary-policy transmission," *Quantitative Economics*, 2024, *15* (4), 1249–1301.

Hubmer, Joachim, Per Krusell, and Anthony A. Smith., "Sources of US Wealth Inequality: Past, Present, and Future," *NBER Macroeconomics Annual*, 2021, *35*, 391–455.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante, "Monetary Policy According to HANK," *American Economic Review*, March 2018, *108* (3), 697–743.

Koop, Gary, M.Hashem Pesaran, and Simon M. Potter, "Impulse response analysis in nonlinear multivariate models," *Journal of Econometrics*, 1996, *74* (1), 119–147.

Krueger, D., K. Mitman, and F. Perri, "Chapter 11 - Macroeconomics and Household Heterogeneity," in John B. Taylor and Harald Uhlig, eds., *Handbook of Macroeconomics*, Vol. 2, Elsevier, 2016, pp. 843–921.

Krusell, Per and Anthony A. Smith, "Income and wealth heterogeneity, portfolio choice, and equilibrium asset returns," *Macroeconomic Dynamics*, 1997, *1* (2), 387–422.

— **and Anthony A. Smith Jr.**, "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 1998, *106* (5), 867–896.

Lawrance, Emily C., "Poverty and the Rate of Time Preference: Evidence from Panel Data," *Journal of Political Economy*, 1991, *99* (1), 54–77.

Merton, Robert, "Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case," *The Review of Economics and Statistics*, 02 1969, *51*, 247–57.

Nirei, Makoto and Shuhei Aoki, "Pareto distribution of income in neoclassical growth models," *Review of Economic Dynamics*, 2016, *20*, 25–42.

OECD, "OECD Economic Outlook No 113 (Edition 2023/1)," *OECD Economic Outlook: Statistics and Projections (database)*, 2023. https://doi.org/https://doi.org/10.1787/b27cc3a6-en (accessed on 9 November 2023).

Rácz, Zoltán., "Housing and Portfolio Choice over the Wealth Distribution," 2024. Working Paper.

Sodini, Paolo, Yao Fu, and Huseyin Aytug, "Construction of the Fama-French-Carhart Four Factors Model for the Swedish Stock Market Using the Finbas Data.," *Swedish House of Finance Research Data Center. Accessed*

September 7, 2022., 2020.

Vestman, Roine, “Limited Stock Market Participation Among Renters and Homeowners,” *The Review of Financial Studies*, 08 2018, *32* (4), 1494–1535.

Vissing-Jorgensen, Annette, “Perspectives on Behavioral Finance: Does "Irrationality" Disappear with Wealth? Evidence from Expectations and Actions,” *NBER Macroeconomics Annual*, 2003, *18*, 139–194.

von Gaudecker, Hans-Martin, Arthur van Soest, and Erik Wengstrom, “Heterogeneity in Risky Choice Behavior in a Broad Population,” *American Economic Review*, April 2011, *101* (2), 664–94.

Wachter, Jessica A. and Motohiro Yogo, “Why Do Household Portfolio Shares Rise in Wealth?,” *The Review of Financial Studies*, November 2010, *23* (11), 3929–3965.

A Data

Deflator. We transform variables into real terms using the Consumer Price Index (CPI) for Sweden recovered from the OECD Economic Outlook No. 113 (OECD, 2023). We use 2021 as the base year.

Risk-free rate and stock market returns. We recover our measures of the risk-free rate and the market return from Sodini et al. (2020). Their data contain monthly returns of the SIX index and the yields of the one-month Swedish Treasury bill from 1983 to 2019. We transform these two series in logs and sum up the resulting monthly values in each year to obtain an annual version (our annual series starts from 1984 since the monthly data are available from February 1983). We then obtain the final series by subtracting inflation.

Portfolio choice moments. We use the schedules of participation, expected excess returns and idiosyncratic variance of households’ gross wealth portfolios over the distribution of net wealth computed from the Swedish administrative data used in Bach et al. (2020). We refer to their paper for more details on the data and the definition of different asset classes. Note that, differently from them, we exclude pension and residential real estate wealth when computing portfolio choice moments. Participation is a dummy equal to one if the household has a positive amount of wealth invested in risky assets. As in Bach et al. (2020), the data cover the period 2000-2007.

Labor income. We recover the standard deviation and Kelly’s skewness of one- and five-year residual log labor income changes and the autocorrelation at the first and fifth lag of residual log labor income for Sweden from the Global Repository of Income Dynamics (Guvenen et al., 2022). The data are available for the period 1985-2016 and we restrict the sample to males between 25 and 55 years old. The series and more details on how the moments are computed can be found on their website: https://www.grid-database.org.

As the measure of the aggregate component of labor income, we use the average of the logarithm of individual income computed from the Swedish administrative data used in Hammar et al. (2022). The series is in real terms (deflated with CPI, with 2021 as base year), includes labor and entrepreneurial income and is net of taxable benefits. Because the data on the latter started to be recorded in 1974, the log growth rates of aggregate income used in this paper cover the period 1975-2016. To be consistent with GRID, the sample is restricted to males between 25 and 55 years old. In addition, observations with a value lower than 100 SEK are excluded.

B Numerical solution

B.1 Discretization and grids construction

Solving the model requires the computation of expectations of non trivial functions. In the most general case, we need to compute expectations with respect the shocks r_1, r_2, ϕ, ε, ν and η. Normally distributed random variables are discretized using Gaussian quadrature with N_q nodes. Furthermore, note that the distributions of r_1 is conditional on the realization of a stock market crash and, similarly, that of ε on the realization of a tail event. This is taken into account simply by scaling the probability of the discretized conditional distributions of these variables by the probability of these events. After discretizing all shocks, we can proceed as follows: (i) for all the possible combinations of grid values of these variables, we compute the value of the function (ii) we multiply it by the probability of that particular combination of values (iii) once we have done this for all the possible combinations we sum up all the function values obtained.

The solution procedure very often requires evaluating the value and policy functions at points off the grid. We do not discretize the persistent component of individual income, which implies that we need to interpolate these functions not only at points off the cash-on-hand grid, but also off the grid of persistent income. We achieve this with 2-dimensional linear interpolation over the (a, z) grid.

We set up a grid for $z_{i,t}$ by constructing an exponentially spaced grid of $(N_z + 1)/2$ points with minimum value equal to zero and maximum value equal to $z_{\max}$. This gives us the positive side of the grid plus the central point (which is, therefore, equal to zero). We then obtain the negative $(N_z - 1)/2$ values by taking the negative of the positive values just computed and obtain the full grid of N_z points by merging the two parts (after having excluded zero from one of them). For a given $z_{i,t}$ in the grid just constructed, we obtain its next period value using the AR(1) process specified in the main text and the discretized values of the conditional distribution of $\varepsilon_{i,t}$.

The advantage of this method is that it requires to discretize just the latter distribution, which is easier than discretizing the full process of $z_{i,t}$ given its non-standard features. In particular, the crucial connections between the higher moments of $z_{i,t}$ and other variables are preserved.

Turning to cash-on-hand, we need to keep track of the minimum value of cash-on-hand implied by each value in the grid of $z_{i,t}$. Thus, we construct N_z grids of cash-on-hand values – one for each grid value of $z_{i,t}$ – each of which is an exponentially spaced grid of N_a points with minimum value equal to the lowest possible realization of cash-on-hand implied by the specific grid value of $z_{i,t}$ under consideration, model parameters and the discretized values of the shocks and maximum value equal to $a_{\max}$. Finally, as we will see below, to solve the model we do not need to keep track of $k_{i,t}$ and $d_{i,t}$ separately but just of their sum, which we denote $s_{i,t}$ and refer to as savings. The grid for $s_{i,t}$ is an exponentially spaced grid of N_a points with minimum value equal to the borrowing limit and maximum value equal to $a_{\max}$.

Our choices for the numerical parameters mentioned above are $N_q = 3$, $N_z = 13$, $N_a = 200$, $z_{\max} = 3.5$ and $a_{\max} = 4 \cdot 10^6$.[29]

B.2 Solving the optimization problem

To ease the exposition, in the following we will drop time (and indicate next period quantities with the prime symbol) and individual specific indices and the dependence of the value and policy functions on θ. Let $s = k + d$, so that, given the two inequality constraints of the problem $s \geq \bar{d} \exp(w)$. Define $\xi = k/(s - \bar{d} \exp(w))$ and note that, since $k \geq 0$ and $k \leq s - \bar{d} \exp(w)$, ξ can be seen as a risky share. Specifically, it is the share of savings plus borrowing capacity invested in the risky asset.[30] Furthermore, let $R^e := R - R^f$ denote the excess return. Then, the original problem can be rewritten as:

$$V(a, z, w) = \max_{c,s,\xi} \left\{ (1-\delta) c^{1-\psi} + \delta \left(\mathbb{E}\left[V(a', z', w')^{1-\gamma} \right] \right)^{\frac{1-\psi}{1-\gamma}} \right\}^{\frac{1}{1-\psi}}$$

[29]As described below, the model can be rescaled with aggregate income w_t. Thus, in practice, we use cash-on-hand and savings scaled by w_t and the values reported refer to these two variables rescaled.

[30]The usual definition of the risky share does not include the borrowing capacity. Nevertheless, we think this is most sensible quantity to use within our framework. Indeed, as in the model agents cannot have both cash and debt at the same time, using $k/(k+d)$ leads to values outside the unit interval and using $k/(k + \max\{0, d\})$ would imply a unitary risky share for everyone with debt.

subject to

$$
\begin{aligned}
a &= c + s + Ff \cdot \exp(w) \\
a' &= (1-\tau_{\text{wealth}})s + (1-\tau_{\text{capinc}})\Big[(R^f - 1)s + R^{e\prime}\xi(s - \bar{d}\exp(w))\Big] + \\
&\quad + (\exp(y') + b')\,(1-\tau_{\text{labor}}) \\
s &\geq \bar{d} \cdot \exp(w)
\end{aligned}
$$

Now, note that as w follows a random walk and utility is homogeneous, we can reduce the state-space by one dimension by scaling the problem with the aggregate income level. Let us define $\hat{x} = x/\exp(w)$ for a generic variable x representing c, a, s and b. Equivalently, for log income y, we define $\exp(\hat{y}) = \exp(y)/\exp(w)$. Also define:

$$
\widehat{V}(a,z) = V(a,z,0)
$$

so that we can write:

$$
V(a,z,w) = \exp(w)V\Big(\frac{a}{\exp(w)}, z, 0\Big) = \exp(w)\widehat{V}(\hat{a},z)
$$

The optimization problem for a participating agent ($F = 1$)[31] can be written as:

$$
\widehat{V}(\hat{a},z) = \max_{\hat{c},\hat{s},\xi}\left\{(1-\delta)\hat{c}^{1-\psi} + \delta\Big[\mathbb{E}\big[e^{(w'-w)(1-\gamma)}\widehat{V}(\hat{a}',z')^{1-\gamma}\big]\Big]^{\frac{1-\psi}{1-\gamma}}\right\}^{\frac{1}{1-\psi}} \tag{B.1}
$$

subject to

$$
\begin{aligned}
\hat{a} &= \hat{c} + \hat{s} + f \\
\hat{a}' &= \Big\{(1-\tau_{\text{wealth}})\hat{s} + (1-\tau_{\text{capinc}})\Big[(R^f-1)\hat{s} + R^{e\prime}\xi(\hat{s}-\bar{d})\Big]\Big\}\, e^{w-w'} + \\
&\quad + \max\big\{\exp(\hat{y}'), \hat{b}'\big\}\,(1-\tau_{\text{labor}}) \\
\hat{s} &\geq \bar{d}
\end{aligned}
$$

To simplify ideas and notation, let us introduce:

$$
\widetilde{V}(\hat{s},\xi,z) = \Big[\mathbb{E}\big[e^{(w'-w)(1-\gamma)}\widehat{V}(\hat{a}',z')^{1-\gamma}\big]\Big]^{\frac{1-\psi}{1-\gamma}} \tag{B.2}
$$

When the optimal risky share is an interior one, i.e., $\xi \in (0,1)$, it satisfies the following first order condition:

$$
\begin{aligned}
0 =& \frac{\partial\widetilde{V}(\hat{s},\xi,z)}{\partial\xi} = \frac{1-\psi}{1-\gamma}\Big[\widetilde{V}(\hat{s},\xi,z)\Big]^{\frac{\gamma-\psi}{1-\psi}}\mathbb{E}\Big[(1-\gamma)e^{(w'-w)(1-\gamma)}\widehat{V}(\hat{a}',z')^{-\gamma}\frac{\partial\widehat{V}(\hat{a}',z')}{\partial\hat{a}'}\frac{d\hat{a}'}{d\xi}\Big] \\
0 =& \mathbb{E}\Big[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}',z')^{-\gamma}\frac{\partial\widehat{V}(\hat{a}',z')}{\partial\hat{a}'}R^{e\prime}\Big]
\end{aligned}
$$

where for the last equation we used that $\widetilde{V}(\hat{s},\xi,z) \neq 0$. The first order condition for the

[31] Non-participants solve the same problem with F and ξ equal to zero. We will describe below the optimal participation choice.

consumption/saving decision when the borrowing constraint is not binding reads:

$$(1-\delta)(1-\psi)\hat{c}^{-\psi} = \delta\frac{\partial\widetilde{V}(\hat{s},\xi,z)}{\partial\hat{s}}$$

$$(1-\delta)(1-\psi)\hat{c}^{-\psi} = \delta\frac{1-\psi}{1-\gamma}\left[\widetilde{V}(\hat{s},\xi,z)\right]^{\frac{\gamma-\psi}{1-\psi}}\mathbb{E}\left[(1-\gamma)e^{(w'-w)(1-\gamma)}\widehat{V}(\hat{a}',z')^{-\gamma}\frac{\partial\widehat{V}(\hat{a}',z')}{\partial\hat{a}'}\frac{d\hat{a}'}{d\hat{s}}\right]$$

$$(1-\delta)\hat{c}^{-\psi} = \delta\left[\widetilde{V}(\hat{s},\xi,z)\right]^{\frac{\gamma-\psi}{1-\psi}}\times$$

$$\times\,\mathbb{E}\left[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}',z')^{-\gamma}\frac{\partial\widehat{V}(\hat{a}',z')}{\partial\hat{a}'}\Big((1-\tau_{\text{wealth}})+(R^f+\xi R^{e\prime}-1)(1-\tau_{\text{capinc}})\Big)\right]$$

Finally, the envelope condition is:

$$\frac{\partial\widehat{V}(\hat{a},z)}{\partial\hat{a}} = \frac{1}{1-\psi}\left[\widehat{V}(\hat{a},z)\right]^{\psi}\left[(1-\delta)(1-\psi)\hat{c}^{-\psi}\frac{d\hat{c}}{d\hat{a}}+\delta\left[\frac{\partial\widetilde{V}(\hat{s},\xi,z)}{\partial\hat{s}}\frac{d\hat{s}}{d\hat{a}}+\frac{\partial\widetilde{V}(\hat{s},\xi,z)}{\partial\xi}\frac{d\xi}{d\hat{a}}\right]\right]$$

$$\frac{\partial\widehat{V}(\hat{a},z)}{\partial\hat{a}} = (1-\delta)\left[\widehat{V}(\hat{a},z)\right]^{\psi}\hat{c}^{-\psi}$$

where the last steps uses $\frac{\partial\hat{c}}{\partial\hat{a}}+\frac{\partial\hat{s}}{\partial\hat{a}}=1$. After simplifying, the two first order conditions read:

$$0 = \mathbb{E}\left[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}',z')^{\psi-\gamma}(\hat{c}')^{-\psi}R^{e\prime}\right] \tag{B.3}$$

$$\hat{c}^{-\psi} = \delta\left[\widetilde{V}(\hat{s},\xi,z)\right]^{\frac{\gamma-\psi}{1-\psi}}\mathbb{E}\left[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}',z')^{\psi-\gamma}(\hat{c}')^{-\psi}\Big((1-\tau_{\text{wealth}})+(R^f+\xi R^{e\prime}-1)(1-\tau_{\text{capinc}})\Big)\right] \tag{B.4}$$

These two optimality conditions can be used to solve the model numerically with the endogenous gridpoint method (Carroll, 2006), extended with a an additional root solving step to compute the optimal risky share. We describe below the steps of the solution algorithm.

1. Let $\widehat{V}$, $\hat{c}$ and ξ be the current guesses for the value function and for the policy functions of consumption and risky share.[32] For all types of agents and all persistent income states:
   (a) Compute for all the values in the savings grid:
      i. The corresponding optimal risky share ξ relying on the first order condition (B.3) as follows:
         - If $\partial\widetilde{V}(\hat{s},0,z)/\partial\xi > 0$, set $\xi = 0$, if $\partial\widetilde{V}(\hat{s},1,z)/\partial\xi < 0$, then set $\xi = 1$;
         - Otherwise use the secant method to find ξ such that $\partial\widetilde{V}(\hat{s},\xi,z)/\partial\xi = 0$.

         This approach finds the optimum as the minimization problem is convex.
      ii. $\widetilde{V}$ using the value of ξ just computed in (B.2).
      iii. Optimal consumption $\hat{c}$ by solving (B.4) using the values of ξ and $\widetilde{V}$ just computed.
      iv. The value function $\widehat{V}$ by inserting $\hat{c}$ and $\widetilde{V}$ just computed into (B.1) and cash-on-hand as $\hat{a} = \hat{c} + \hat{s} + Ff$.
      v. Repeat steps i-iv for non-participants, that is, assuming $\xi = F = 0$.
      vi. Take care of the borrowing limit:

[32] As a starting point we take $\hat{c}(\hat{a},z) = \widehat{V}(\hat{a},z) = \hat{a} - \bar{d}$ and ξ equal to zero.

- If the first value of cash-on-hand (corresponding to savings equal to the borrowing limit) obtained for non-participants is higher than the corresponding value for participants, then the borrowing constraint binds for the latter and we add this point to the values obtained for non-participants (as agents who are at the borrowing limit do not participate). This increases the accuracy of determining the optimal threshold for participation.
- If the first value of cash-on-hand for non-participants is higher than the lowest value of the cash-on-hand grid for the current persistent income state, then the borrowing constraint binds also for them and we add such worst possible realization point to the values obtained for non-participants. This makes sure that optimal policies of constrained agents are accurately captured.

(b) Compare the value of participating with that of not participating to find the level of cash-on-hand at which the former becomes higher than the latter. The final value and policy functions correspond to those of non-participants for values of cash-on-hand below the threshold and to those of participants for values above.

(c) Evaluate the relative change between the old and new consumption policy functions and of the difference between the old and new risky share policy functions at each point of the cash-on-hand grid and compute the L^{∞} norm of the resulting vectors.

2. If the maximum of the array containing the two values computed in the last step for each type of agent and persistent income state is smaller than 10^{-3}, then convergence is achieved.[33] Otherwise, go back to step 1 using as new guesses the value and policy functions just obtained.

B.3 Simulation and stationary distribution

In this model the unique aggregate state is the distribution of agents across individual states which are characterized by the triple $(\theta_i, z_{i,t}, a_{i,t})$. As described before, the values of all these states are approximated by a finite grid, therefore in the numerical setting the distribution object can be described as a vector of length $N_t \cdot N_z \cdot N_a$ – with N_t being the number of agents' types – containing the probability weights corresponding to each individual state. To simulate the economy we need to compute the transition probabilities of moving from one individual state to another, which naturally depend on the actual value of the aggregate shocks. Thus, while –

[33]Using a smaller value would significantly increase the time to perform the estimation of preference parameters. We have tried setting a lower tolerance to solve the model with the estimated parameters and found no significant changes in the results.

strictly speaking – the model has no steady state, one could examine the dynamic properties of the distribution by characterizing a quasi steady state distribution computed from aggregating the transition matrices corresponding to all values of the aggregate shocks we want to simulate into a quasi steady state transition matrix, as described below. Intuitively, the quasi steady state distribution is generated by the unconditional transition probabilities from one state to another, keeping the aggregate shocks unknown. Due to ergodicity, the resulting distribution coincides with the limiting time average of the distributions obtained from a long enough simulation of distributions generated by an actual series of aggregate shocks. Accordingly, whenever referring to steady state objects, we always mean quasi steady state objects.

Conditional transition matrices. Given a realization of the aggregate shocks (r_1, r_2, ϕ) for each individual state we can compute future cash-on-hand and future persistent income corresponding to each realization of idiosyncratic shocks, where the latter are simulated from the grids described in B.1. Since in the generic case the simulated cash-on-hand and z values do not fall on grid points, the conditional probability weight corresponding to each simulated (a, z) pair is distributed between the neighboring four points proportionally to their relative distance. To avoid extrapolation errors, z is truncated if it falls below $-z_{max}$ or above z_{max}. From the transition probabilities of moving from one individual state to another we can build up the transition matrices conditional on any realization of the aggregate shocks.

Unconditional transition matrix. To obtain the steady state distribution we need an "average transition matrix". One way of defining it would be taking the conditional transition matrix corresponding to the average values of all aggregate shocks. However, a steady state computed from such a matrix would miss all the consequences of cyclical movements in moments of idiosyncratic shocks, central to our analysis. Therefore, we compute our steady state transition matrix as a weighted average of the conditional transition matrices corresponding to shock values used in the policy iteration, where weights are the probabilities that the given combination of shocks takes place. Hence, all entries in the steady state matrix are the true unconditional transition probabilities of moving from one individual state to another.

Stationary distribution. The steady state is found by iteration, i.e., by multiplying an arbitrary vector with the unconditional transition matrix until convergence. Note that the aggregate state object is a distribution over agent type, persistent income and cash-on-hand. Using the optimal policy function for savings at each state of the resulting stationary distribution we can compute implied distribution of savings, which is what we refer to as wealth distribution.

B.4 MIT shock

Let $(r_1^{\text{sh}}, r_2^{\text{sh}}, \phi^{\text{sh}})$ denote the values of the MIT shocks to the aggregate variables,[34] T^{sh} the time period of the shock and T the total number of time periods to be simulated. The dynamic response of the distribution of agents across individual states following the MIT shock is computed using the generalized impulse response approach proposed by Koop et al. (1996). Specifically, we proceed as follows:

1. Generate a time series of aggregate shocks $\{(r_{1,t}, r_{2,t}, \phi_t)\}_{t=1}^{T}$ by drawing values from the discretized distributions of the three aggregate shocks.
2. Choose a starting distribution and multiply it with the transition matrix conditional on the triplet $(r_{1,1}, r_{2,1}, \phi_1)$ realized in the first period to obtain the next period distribution. Repeat the procedure until T to obtain the evolution of the distribution without the shock.
3. Repeat the procedure at the previous point but at time T^{sh} use instead the transition matrix conditional on $(r_{1,T^{\text{sh}}} + r_1^{\text{sh}}, r_{2,T^{\text{sh}}} + r_2^{\text{sh}}, \phi_{T^{\text{sh}}} + \phi^{\text{sh}})$ which can be computed using the same procedure outlined in Section B.3 for conditional transition matrices. This delivers the evolution of the distribution with the shock.
4. Use the two evolutions to compute the time series of functions of the distribution with and without the shock.
5. Repeat the above steps for $n = 2, \ldots, N$ times keeping the same starting distribution.
6. Compute at each (t, n) pair the difference between the time series of the desired quantity with the shock and without the shock. Take the average over the N simulations to obtain the impulse response.

For the simulations reported in the paper we use $N = 30000$, $T = 500$ and $T^{\text{sh}} = 100$ and we use the stationary distribution implied by each model as starting point.

C Estimation

As explained in the main text, the goal of the SMM estimation routine is to find a vector of parameters Φ that minimizes the objective function defined by:

$$\sqrt{\frac{d(\Phi)'\Omega d(\Phi)}{\text{trace}(\Omega)}}$$

where $d(\Phi)$ is a vector containing deviations of model moments from their targets in the data and Ω is a weighting matrix. We rescale the weighted moment deviations by the trace of the weighting matrix and take the square root of the resulting quantity. In this way, when Ω is diagonal, the resulting value can be interpreted as the weighted quadratic mean of the moment

[34] In the paper we consider a shock to r_2, i.e., $r_1^{\text{sh}} = \phi^{\text{sh}} = 0$.

deviations.

The estimation procedure includes a global and a local stage. In the global stage, we generate a Sobol sequence of length N_{glo} over the parameter space and compute the objective function for each vector of parameters in the sequence. In the local stage, we use the best N_{loc} points from the global stage – i.e. the points achieving the lowest objective function values in the global stage – as starting guesses for running a local search with the Nelder-Mead algorithm. The global stage allows us to identify the most promising regions of the parameter space to find the optimum, thus minimizing the resources allocated for the computationally intensive local search algorithm.[35] In the following, we describe the specific settings adopted in each estimation step. In all cases, the implied objective functions are not well-behaved and finding the optimizer is a complex task. In particular, estimation of the participation cost, the borrowing limit, the share of types and the preference and idiosyncratic return risk parameters can take several days, even when using high-performance computing clusters, as solving the optimization problem and computing the stationary distribution is time intensive. The code implementing the estimation routine just described can be found in a dedicated Julia package (Azzalini and Rácz, 2024).

C.1 Aggregate return and income process

The seven parameters $(\underline{\mu}_r, \overline{\mu}_r, p_r, \sigma_{r_2}, g, \lambda_{rw}, \sigma_\phi)$ governing the stochastic processes of aggregate income and stock market returns are estimated with SMM to capture the joint dynamics of yearly stock market returns and aggregate income growth in Sweden. More specifically, we target the mean, standard deviation and skewness of the return of the Swedish stock market index and aggregate income growth in Sweden, as well as the (one-year lagged) correlation between the two series.[36] This results in a total of seven moments to identify the seven parameters of interest. The outcome of the SMM estimation is reported in Table C.1.

[35]Note that while the global stage searches in a bounded region of the parameter space decided by the researcher – the one for which the Sobol sequence is created – the local search step is not bounded and can potentially deliver any point.

[36]The sample used for log returns is 1984-2016 and for the log growth of aggregate income is 1975-2016.

Estimated parameters							
	Stock market returns				Aggregate income		
	$\underline{\mu}_r$	$\overline{\mu}_r$	p_r	σ_{r_2}	g	λ_{rw}	σ_ϕ
Estimate	−0.343	0.181	0.168	0.173	−0.001	0.108	0.020
SE	0.127	0.042	0.084	0.050	0.011	0.077	0.005

Moments							
	Log returns			Agg. income growth			
	Mean	SD	Skew	Mean	SD	Skew	Corr
Data	0.093	0.261	−0.752	0.010	0.029	−0.660	0.537
Model	0.093	0.261	−0.752	0.010	0.029	−0.660	0.537

Table C.1: Aggregate return and income processes estimation results. The stock market log return in year t is the sum of $r_{1,t}$ and $r_{2,t}$. With probability p_r there is a stock market crash and $r_{1,t} = \underline{\mu}_r$, otherwise $r_{1,t} = \overline{\mu}_r$. $r_{2,t}$ is normally distributed with zero mean and standard deviation σ_{r_2}. The logarithm of aggregate income is modelled as $w_t = g + w_{t-1} + \lambda_{rw} r_{1,t} + \phi_t$, with ϕ_t normally distributed with mean zero and standard deviation σ_ϕ. The estimation procedure targets mean, standard deviation, skewness of log stock market returns and aggregate income growth and their (one-year lagged) correlation. Standard errors are computed using parametric bootstrap.

Overall, the estimated model replicates the desired features of the data. The probability of a stock market crash is about 17%. In our sample, this corresponds to five and a half years, roughly matching the crises at the beginning of the 90s, the dot-com bubble and the Great Recession. The predicted average stock market drop during a crash is about 34% and the average return in normal times is about 18%. As for aggregate income dynamics, the aggregate component of individuals' log earnings displays virtually zero drift and is positively correlated with the stock market ($\lambda_{rw} = 0.108$). In line with the data, the implied predicted average aggregate income growth during crashes and normal times is, respectively, about -3.8% and 1.9%.

Differences from Catherine (2021). In Catherine (2021), r_1 is assumed to follow a mixed Normal distribution and when estimating the parameter values, the kurtosis of stock market returns and aggregate income changes are included among the targeted moments. We found that the empirical values of kurtosis in Swedish aggregate data are very sensitive to the exact choice of the time period and variable definitions. Furthermore, his framework cannot match the data moments well when the kurtosis values of stock market and aggregate labor income shocks differ substantially. Therefore, we exclude kurtosis from the set of targeted moments. As a consequence, we have to reduce the number of free parameters, which is achieved via assuming that r_1 follows a Bernoulli distribution instead of a mixed Normal.

Technical details. To estimate the stochastic processes governing the aggregate market return and income we set N_{glo} and N_{loc} equal to 10000 and 1000, respectively. We then simulate the processes for 100000 periods and three economies and compute the moments averaging across the economies and discarding the first 100 periods. As Ω is set equal to a diagonal unitary matrix all the targeted moments have the same weight. Letting $\mathrm{mom}^{\mathrm{data}}$ and $\mathrm{mom}^{\mathrm{model}}$ the vectors containing moments in the data and in the model, respectively, for each moment j the deviation is computed as follows:

$$d_j(\Phi) = \frac{\mathrm{mom}_j^{\mathrm{data}} - \mathrm{mom}_j^{\mathrm{model}}(\Phi)}{\mathrm{mom}_j^{\mathrm{data}}}$$

C.2 Individual earnings process

We estimate the parameters of the individual earnings process $(\rho, p_\varepsilon, \mu_\varepsilon, \lambda_{\varepsilon w}, \underline{\sigma}_\varepsilon, \overline{\sigma}_\varepsilon, \sigma_\nu)$ targeting three sets of moments: the cross-sectional standard deviation and Kelly's skewness of log earnings growth at the one- and five-year horizons and the first and fifth-order autocorrelation of log earnings. We use the actual time series of aggregate income shocks realized over the period 1975-2016 to simulate the individual income process and target the evolution of Kelly skewness from 1985 to 2016 and the time-series average of standard deviation and autocorrelation over the same period. This is motivated by the fact that in our data the latter two quantities are relatively constant over time.[37] We obtain these moments for earnings net of age effects from the Global Repository of Income Dynamics (GRID database, Guvenen et al., 2022), which, for Sweden, provides such quantities over the period 1985-2016. We restrict our analysis to males between 25 and 55 years old and we also remove a linear time trend from the time series of the three moments. Table C.2 and Figure C.1 report the results.

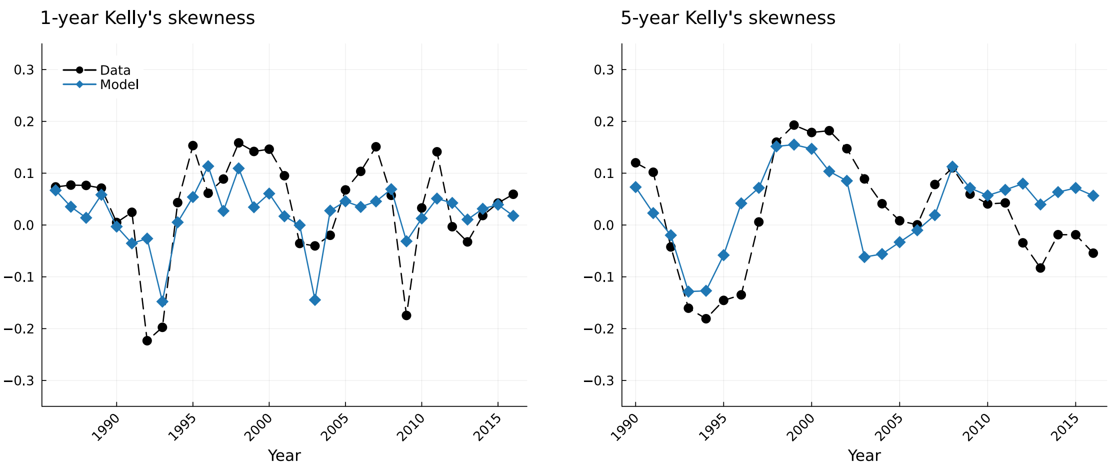

Figure C.1: Fit of the estimated individual earnings process. The figure shows the time series of Kelly's skewness of one- and five-year log earnings growth implied by the model against their data equivalent obtained from the GRID database (Guvenen et al., 2022).

[37] Busch et al. (2022) also find that the variance of earnings growth is flat and acyclical in Sweden, Germany and the United States.

Estimated parameters

	Persistent						Transitory
	ρ	p_ε	μ_ε	$\lambda_{\varepsilon w}$	$\underline{\sigma}_\varepsilon$	$\overline{\sigma}_\varepsilon$	σ_ν
Estimate	0.907	0.139	−0.044	12.215	0.434	0.021	0.266
SE	0.004	0.010	0.002	1.051	0.038	0.019	0.004

Moments

	St. Dev.		Autocorr.	
	1y	5y	1st	5th
Data	0.425	0.605	0.712	0.485
Model	0.436	0.590	0.713	0.484

Table C.2: Individual earnings process estimation results. The individual-specific part of log earnings for individual i at time t is the sum of a persistent and a transitory component, respectively, $z_{i,t}$ and $\nu_{i,t}$. The latter is normally distributed with zero mean and standard deviation σ_ν. The persistent component follows an AR(1) process, $z_{i,t} = \rho z_{i,t-1} + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is a mixture of Normals. That is, with probability p_ε a tail event is realized and $\varepsilon_{i,t}$ is drawn from a Normal distribution with mean $\underline{\mu}_{\varepsilon,t} = \mu_\varepsilon + \lambda_{\varepsilon w}(w_t - w_{t-1})$ and standard deviation $\underline{\sigma}^2_\varepsilon$. Otherwise, $\varepsilon_{i,t}$ is drawn from a Normal distribution with mean $\overline{\mu}_{\varepsilon,t}$ and standard deviation $\overline{\sigma}^2_\varepsilon$. These shocks are also assumed to have zero mean, so that $p_\varepsilon \underline{\mu}_{\varepsilon,t} + (1 - p_\varepsilon)\overline{\mu}_{\varepsilon,t} = 0$. The SMM estimation procedure targets the average standard deviation of one- and five-year residual log earnings growth and the average first- and fifth-order autocorrelation of residual earnings in the period 1985-2016 and the time series of Kelly's skewness of one- and five-year residual log earnings growth over the same period. Standard errors are computed using parametric bootstrap.

The model performs relatively well in matching the data. Tail events happen with a probability of about 14%, which implies that most of the workers receive persistent shocks from a Normal distribution with relatively low standard deviation (0.021). The variance of the transitory shocks is 0.266, which indicates that a relevant part of income shocks is not very persistent.

Differences from Catherine (2021). We deviate from Catherine (2021) in two distinct points. First, to ensure tight identification of all parameters, we model the transitory part of log earnings, $\nu_{i,t}$, as a simple Normal shock instead of Normal distributions conditional on the realization of $\varepsilon_{i,t}$. Indeed, in our setting we found that while the total variance of the transitory component was well identified, the variance of the transitory component conditional on $\varepsilon_{i,t}$ was not. Second, to make sure that the persistence of income shocks is accurately captured, we added the autocorrelation of income changes to the targeted moments. This choice resulted in a lower estimate of ρ, i.e., the autoregressive coefficient of the persistent component of individuals' log earnings, relative to studies featuring a similar income process in which autocovarices were not targeted, such as (Catherine, 2021; Guvenen et al., 2021, 2014).

Technical details. We estimate the stochastic process governing individual labor income setting N_{glo} to 10^6 and N_{loc} to 5000. We simulate the labor income histories for 10000 individuals in five economies over the period 1975-2016. We then compute the moments averaging across the economies. We set Ω equal to a diagonal unitary matrix also in this case and compute the deviation for each moment j as follows:

$$d_j(\Phi) = \frac{\text{mom}_j^{\text{data}} - \text{mom}_j^{\text{model}}(\Phi)}{\text{mom}_j^{\text{data, norm}}}$$

where $\text{mom}_j^{\text{data, norm}}$ is a normalization factor which we set equal to the time series standard deviation in the data of the moment considered.[38]

C.3 Preferences, idiosyncratic return risk and other model parameters

We estimate the model parameters related to agents' preferences and idiosyncratic return risk, the share of each type, the borrowing limit and the fixed participation cost setting N_{glo} equal to 10000 and N_{loc} to 20.[39] For every parameter vector we solve the model, compute the stationary distribution and obtain the desired moments described from the latter. The weighting matrix puts 25% of the weight on the wealth-to-income ratio, 25% on the share of agents with negative net wealth and splits the remaining 50% equally between the portfolio schedules moments. For the wealth-to-income ratio and the share of agents with negative net wealth, the deviation $d(\Phi)$ is computed using equation (C.1), while for each moment of the portfolio schedules using equation (C.2) where the normalizing factor is the standard deviation of all the moments comprised in each schedule.

The model equivalents of the portfolio choice quantities targeted in the estimation are defined as follows. Participation is equal to one if $\xi > 0$. Expected excess returns are computed as the product between ξ and the expected excess return of the systematic part of returns, that is, using the properties of mixed and log-normal distributions:

$$\xi \mathbb{E}\left[\exp\left(r_1 + r_2\right) - R^f\right] = \xi \left[p_r \exp\left(\underline{\mu}_r + \sigma_{r_2}^2/2\right) + (1 - p_r)\exp\left(\overline{\mu}_r + \sigma_{r_2}^2/2\right)\right]$$

Finally, the share of idiosyncratic variance is equal to the variance in individual portfolios in addition to the systematic return variance.[40] Given our modeling assumptions, the former (as

---

[38]We repeated this procedure three times until the local optimizations converged to the same maximizer. In each of the two additional estimations we restricted the search space of the global step to the most promising area found in the previous iteration.

[39]For computational feasibility, we restrict the global stage of the estimation to search in regions of the parameter space in which the predominant type has a higher probability of having jointly lower time preference rate, higher risk aversion, and lower idiosyncratic return risk. That is, the searching region for the global stage includes (but it is not limited) more points where the prevalent type has these characteristics. In an earlier version we did not apply this restriction and found similar, but less tightly identified results. The local stage, as previously described, is instead not bounded by this constraint.

[40]The share of idiosyncratic variance for a wealth group is computed conditional on participation. Since in the data both variables are increasing in wealth, if participation for a wealth group is zero we assign as share of idiosyncratic

a function of ζ) is equal to:

$$\begin{aligned}\mathbb{V}\left[\exp\left(r_1+r_2+\eta\right)\right](\zeta) = & \; p_r \exp\left(2\underline{\mu}_r + 2\sigma_{r2}^2 + \zeta^2\sigma_r^2\right) + (1-p_r)\exp\left(2\overline{\mu}_r + 2\sigma_{r2}^2 + \zeta^2\sigma_r^2\right) + \\ & - p_r^2 \exp\left(2\underline{\mu}_r + \sigma_{r2}^2\right) - (1-p_r)^2 \exp\left(2\overline{\mu}_r + \sigma_{r2}^2\right) + \\ & - 2p_r(1-p_r)\exp\left(\underline{\mu}_r + \overline{\mu}_r + \sigma_{r2}^2\right)\end{aligned}$$

where $\sigma_r^2 = p_r\underline{\mu}_r^2 + (1-p_r)\overline{\mu}_r^2 - \mu_r^2 + \sigma_{r_2}^2$ and $\mu_r = p_r\underline{\mu}_r + (1-p_r)\overline{\mu}_r$. Thus, the share of idiosyncratic variance (as function of ζ) is:

$$1 - \frac{\mathbb{V}\left[\exp\left(r_1+r_2+\eta\right)\right](0)}{\mathbb{V}\left[\exp\left(r_1+r_2+\eta\right)\right](\zeta)}$$

D Additional model specifications

D.1 Heterogeneity only in risk aversion

In order to understand the role of heterogeneity in risk aversion, we re-estimate the model allowing for ex-ante heterogeneity only in this parameter while targeting the same moments as in the benchmark case.

As high risk aversion implies low stock holdings but also high savings, generating a stock holder type at the top of the wealth distribution is not straightforward when heterogeneity in δ is missing. As shown by the estimated values under this specification, the model partially achieves this result by compensating the higher discount factor of type-one agents with a lower attitude towards risk and the lower discount factor of type-two individuals with the higher capital gains generated by their lower estimated γ. The drawback of this outcome is that, as shown in the left panel of Figure D.1, the expected excess returns overshoot the empirical values until the eight decile, especially at the bottom. On the other hand, the model matches relatively well the pattern at the top but – as shown in Table D.2 – the higher capital gains of the individuals at the top are not enough to generate the wealth concentration found in the data, resulting in a lower Gini coefficient. Interestingly, as shown in Figures E.2, E.3 and E.4 the model-implied marginal propensities and wealth mobility are not too far away from the benchmark, which highlights the importance of heterogeneity in risk aversion for these two dimensions.

variance the one implied by lowest ζ across the agents' types.

	Type 1			Type 2					
Model	δ	γ	ζ	δ	γ	ζ	Share of Type 1	f	$\bar{d}$
Benchmark	0.924	9.891	0.543	0.966	2.671	0.825	0.901	0.001	−0.835
Only γ	0.946	9.173	0.618		1.964		0.841	0.008	−0.659
No cy. skew.	0.946	15.462	0.721	0.938	5.453	0.602	0.838	0.032	−0.377
Hard-wired	0.902	10.334		0.975			0.871		−0.718

Table D.1: Estimated model parameters, benchmark model vs. alternative specifications. We solve each model setting inverse EIS (ψ) equal to one. "Only γ" the model with two types differing in γ but not in δ and ζ. "No cy. skew." the model without cyclical skewness in labor income shocks. "Hard-wired" the model with risky share and share of idiosyncratic return variance exogenously set.

	Model				Data
Share held by (%):	Benchmark	Only γ	No cy. skew.	Hard-wired	Sweden (2000-2007)
Q1	0.1	0.3	0.5	0.0	-1.1
Q2	4.6	5.9	6.0	3.8	2.8
Q3	8.8	11.6	12.0	7.9	8.7
Q4	15.8	20.3	22.0	14.9	19.4
Q5	70.6	61.9	59.5	73.3	70.2
90-95 %	11.9	12.6	12.9	13.5	13.4
95-99 %	22.5	17.3	16.6	24.2	17.9
Top 1 %	22.5	15.6	12.2	20.8	21.3
Wealth Gini	0.69	0.61	0.58	0.71	0.71

Table D.2: Wealth inequality, benchmark model vs. alternative specifications. The table reports the share of wealth held between different quantiles of the wealth distribution and the Gini coefficient. "Only γ" the model with two types differing in γ but not in δ and ζ. "No cy. skew." the model without cyclical skewness in labor income shocks. "Hard-wired" the model with risky share and share of idiosyncratic return variance exogenously set (in this case the eight shares are targeted in the estimation). The last column shows the data equivalents computed in the Swedish administrative data compiled by Bach et al. (2020) – averaged over the 2000-2007 period.

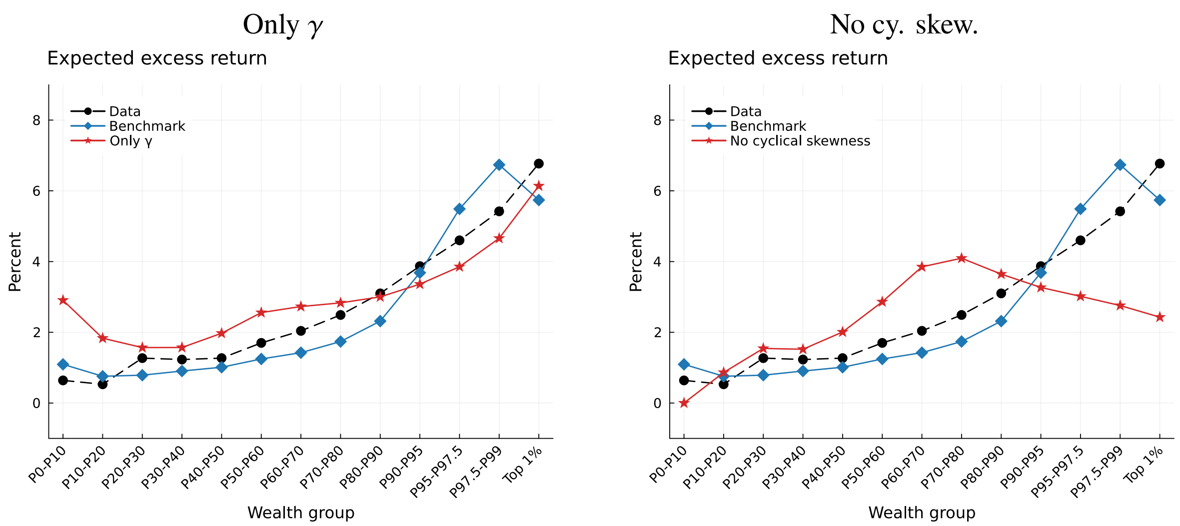

Figure D.1: Fit of the estimated model, benchmark vs. alternative specifications. The figure shows the schedules of expected excess returns over the wealth distribution from the benchmark model, from different alternative specifications, and their data equivalents computed in the Swedish administrative data used by Bach et al. (2020). The left panel shows the results for the model with two types differing only in γ. The right panel for the model without cyclical skewness in labor income shocks.

D.2 No cyclical skewness

We have discussed in Section 3.1 that modeling cyclical skewness of labor income shocks is crucial to generate an increasing risky share at the bottom of the wealth distribution. In order to understand how important this feature is for our results, we analyze a counterfactual economy in which this channel is turned off. Specifically, we remove stock market crashes and make aggregate income growth (and consequently also individual labor income) independent of aggregate returns. As a result, aggregate income and returns are not correlated anymore and the skewness of labor income shocks is fixed rather than cyclical.

When imposing these restrictions we make sure that the aggregate processes still match their unconditional means and standard deviations. Specifically, to eliminate stock market crashes while keeping the mean and variance of aggregate returns we set $r_{1,t+1} \overset{i.i.d.}{\sim} \mathcal{N}\left(\mu_r, \tilde{\sigma}^2_{r_1}\right)$ with $\mu_r = p_r \underline{\mu}_r + (1-p_r)\overline{\mu}_r$ and $\tilde{\sigma}^2_{r_1} = p_r \underline{\mu}^2_r + (1-p_r)\overline{\mu}^2_r - \mu^2_r$. Similarly, shutting down the connection between aggregate income and returns while keeping the mean and variance of the former is done by setting $w_t = g + w_{t-1} + \lambda_{rw}\mu_r + \phi_t$ and $\phi_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \tilde{\sigma}^2_{\phi}\right)$ with $\tilde{\sigma}^2_{\phi} = \sigma^2_{\phi} + \lambda^2_{rw}\tilde{\sigma}^2_{r_1}$. Finally, turning off cyclical skewness in labor income shocks is done by using the expression for Δw_t just specified in the process for the persistent component of labor income. Note that this still allows skewness in the distribution of ε, but not cyclical variation. Therefore, we re-estimate this restricted version of the individual labor income process by targeting the same moments as in the baseline and the average Kelly's skewness of log earnings growth at the one- and five-year horizons rather than its evolution from 1985 to 2016.[41] As in the new specification $\lambda_{\varepsilon w}$ is not identified separately from μ_{ε}, we set the former parameter equal to its value in the baseline. Table D.3 reports the estimation results which show that the model captures well the targeted moments.

[41] See Appendix C for the baseline estimation of the income process and the details of the estimation procedure.

Estimated parameters						
	Persistent					Transitory
	ρ	p_ε	μ_ε	$\underline{\sigma}_\varepsilon$	$\overline{\sigma}_\varepsilon$	σ_ν
Estimate	0.909	0.269	0.238	0.032	0.052	0.266
SE	0.001	0.004	0.004	0.008	0.004	0.000

Moments						
	Kelly's Skew.		St. Dev.		Autocorr.	
	1y	5y	1y	5y	1st	5th
Data	0.038	0.025	0.425	0.605	0.712	0.485
Model	0.034	0.031	0.436	0.587	0.714	0.483

Table D.3: Individual earnings process estimation results, no cyclical skewness. The individual-specific part of log earnings for individual i at time t is the sum of a persistent and a transitory component, respectively, $z_{i,t}$ and $\nu_{i,t}$. The latter is normally distributed with zero mean and standard deviation σ_ν. The persistent component follows an AR(1) process, $z_{i,t} = \rho z_{i,t-1} + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ is a mixture of Normals. That is, with probability p_ε a tail event is realized and $\varepsilon_{i,t}$ is drawn from a Normal distribution with mean $\underline{\mu}_{\varepsilon,t} = \mu_\varepsilon + \lambda_{\varepsilon w}\left[g + \lambda_{rw}\left(p_r\underline{\mu}_r + (1-p_r)\overline{\mu}_r\right)\right]$ and standard deviation $\underline{\sigma}^2_\varepsilon$. Otherwise, $\varepsilon_{i,t}$ is drawn from a Normal distribution with mean $\overline{\mu}_{\varepsilon,t}$ and standard deviation $\overline{\sigma}^2_\varepsilon$. These shocks are also assumed to have zero mean, so that $p_\varepsilon\underline{\mu}_{\varepsilon,t} + (1-p_\varepsilon)\overline{\mu}_{\varepsilon,t} = 0$. As in this specification $\lambda_{\varepsilon w}$ is not identified separately from μ_ε, we set the former parameter equal to its value in the baseline. The SMM estimation procedure targets the average Kelly's skewness and standard deviation of one- and five-year residual log earnings growth and the average first- and fifth-order autocorrelation of residual earnings in the period 1985-2016. Standard errors are computed using parametric bootstrap.

The right panel in Figure D.1 shows the schedule of expected excess returns obtained with this alternative specification. While the increasing relationship between this variable and wealth is preserved until the 80th percentile, the direction flips after that point. The reason for this worse match has to do with the inability of the new model to identify the same correlations between preference parameters as in the benchmark: as shown in Table D.1, the type of agents with higher patience and higher idiosyncratic return risk also have higher risk aversion.

Intuitively, without the additional risk in human capital generated by cyclical skewness, risk aversion and/or the fixed participation cost need to increase to match a lower participation at the bottom of the wealth distribution. This is exactly what the parameter estimates point to: the values of γ and f are higher than in the benchmark (the latter corresponds to a cost of about 8410 vs. 263 SEK per year, in 2021 terms). However, in addition to the unrealistically high estimated participation cost, matching the extensive margin with this approach also delivers – as illustrated by Figure E.1 – an almost binary distribution of participation, with too many wealth-rich individuals participating and too few wealth-poor agents doing so.

To highlight the crucial role of this model feature in correctly identifying the parameters governing ex-ante individual heterogeneities in the model, note that while in the benchmark the cyclical skewness channel was enough to prevent poor individuals from investing in the risky asset, now this result has to be achieved through a compositional effect between a highly risk averse type, ending up on the bottom of the wealth distribution, and a less risk averse majority. Consequently, no compositional effect is active at the top of the wealth distribution, where the increasing pattern in the excess return and share of idiosyncratic risk is completely missed. The lower expected excess returns in the right part of the distribution also imply (together with the lower discount factors) less wealth concentration in the right tail compared to the data, as reported in Table D.2.

While, as illustrated by Figure E.3, mobility is overall still well matched (except at the very top, where it is slightly higher, see Figure E.4), the lower capital gains in the right part of the wealth distribution suggest that the mechanisms underlying the flows are at odds with patterns in the data. Finally, as seen in Figure E.2, the lack of individuals with low enough risk aversion at the top does not deliver the increasing marginal propensity to invest in the risky asset as function of wealth obtained in the benchmark.

D.3 Hard-wired portfolio choice

To relate our paper to the literature hardwiring portfolio allocations as a function of wealth, in this section we investigate the role of endogenous portfolio choice between a safe and risky asset by studying the predictions of a counterfactual model without this feature. Specifically, we compute the average risky share and share of idiosyncratic return risk between 0.1 percent spaced quantiles of the wealth distribution generated in our benchmark specification, fit the two relationships with seventh-order polynomials of the logarithm of wealth, and use the resulting functions as model inputs instead of optimally solving for ξ and using the type-dependent value of ζ.

As key determinants of the three portfolio choice schedules are exogenous, re-estimation of the hard-wired specification is, however, less straightforward than for the previous counterfactuals. We follow Hubmer et al. (2021) and estimate the two types' discount factor, a common value for their risk aversion, the borrowing limit and the share of each type by directly targeting wealth inequality moments.[42] More specifically, in addition to the aggregate wealth-to-income ratio and the share of agents with negative wealth, we match the share of wealth held by the eight wealth groups reported in Table 3 for Sweden.

Table D.2 shows that the targeted wealth inequality moments are relatively well matched and Table D.1 that to achieve this result a starker separation in discount factors is needed to compensate the reduced compositional effects in portfolio choices.[43] Interestingly, Figures E.2 and E.3 reveal key differences in marginal propensities to consume and invest and wealth

[42] We set the fixed participation cost equal to zero to avoid forcing poor agents to pay it in every period even if it is suboptimal.

[43] We do not report the portfolio choice schedules as, by construction, they are almost identical to those obtained with the benchmark.

mobility between this counterfactual specification and the benchmark. Similarly to the "Only δ" case, the higher discount factor of type-two individuals and the lack of type-dependent idiosyncratic risk generate lower mobility in the top quintile.[44] Turning to marginal propensities, the hard-wired model delivers decreasing but higher MPCs until about the 90th percentile, mainly as a consequence of the lower estimated discount factor for type-one individuals. Additionally, the MPI in the risky asset in the benchmark between the 80th and 99th percentiles is lower as, conditional on type, in the hard-wired case the risky share is increasing in wealth by construction rather than decreasing as in the benchmark.[45]

The impact of these differences can be seen in Figure D.2 which reports the response of the share of wealth held by the wealthiest 10% and 1% and aggregate consumption after a ten percentage point positive MIT shock in r_2.[46] On the one hand, the impact on wealth inequality is practically the same as in the benchmark. The high persistence achieved via type dependence in the benchmark specification is replicated by the increasing risk-taking of the rich after a return shock in the hardwired case. This suggests that for the response of wealth inequality, matching the share of risky assets over the wealth distribution in addition to the overall shape of the latter is more important than how this fit is achieved (i.e., with endogenous portfolio choice or not). However, the response in aggregate consumption is markedly different: it is about 25% larger on impact, as a consequence of the higher average MPC in this alternative specification.[47]

[44] Benhabib et al. (2019), instead, document reasonable levels of mobility in an extension of their model hard-wiring the relationship between wealth and returns. While it is hard to tell apart the cause of this difference since several features of their model are different from ours, a relevant factor might be the less steep relationship between these two variables derived from PSID data adopted in their paper.

[45] The MPI in the risky asset in the hard-wired case decreases again in the top two wealth groups because the goal of the counterfactual exercise requires us to target the risky share generated in the benchmark, which slightly underestimates the expected excess return level in the data for those two wealth groups.

[46] In a similar vein, Hintermaier and Koeniger (2024) quantify the importance of portfolio choices for the response of aggregate consumption to interest rate changes. Their focus is on the portfolio decision to own vs. rent housing.

[47] It is important to note that the above findings apply to the effect of a transitory shock. For permanent ones, we expect more substantial differences between the hardwired and benchmark specifications.

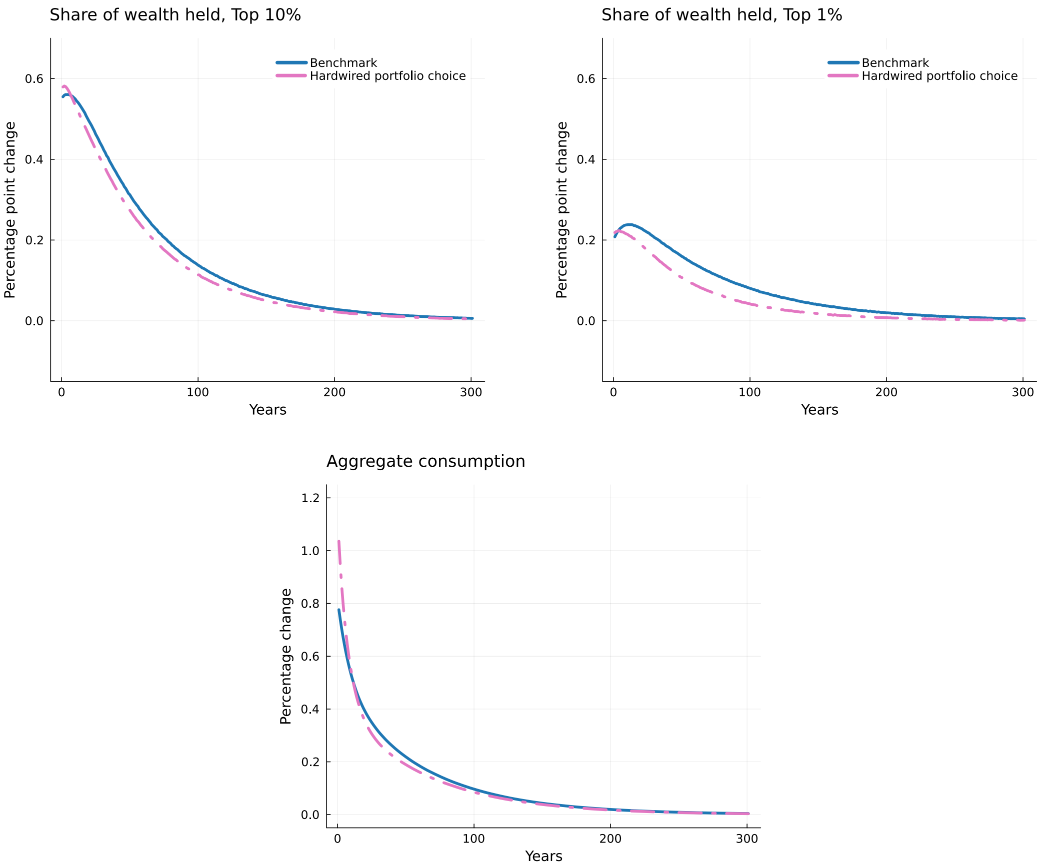

Figure D.2: Response to an aggregate return shock, benchmark vs. hard-wired portfolio choice. The figure shows the response of the share of wealth held by the top 10% and top 1%, and of aggregate consumption to a ten percentage point increase in r_2. "Hardwired portfolio choice" indicates the model with risky share and share of idiosyncratic return variance exogenously set.

E Supplementary tables and figures

	Wealth rank in 7 years					
Wealth rank today	Bottom 90%	Top 10%-5%	Top 5%-1%	Top 1%-0.1%	Top 0.1%-0.01%	Top 0.01%
Bottom 90%	0.97	0.03	0.01	0.00	0.00	0.00
	[0.96]	[0.03]	[0.01]	[0.00]	[0.00]	[0.00]
Top 10%-5%	0.56	0.29	0.14	0.00	0.00	0.00
	[0.33]	[0.44]	[0.22]	[0.01]	[0.00]	[0.00]
Top 5%-1%	0.09	0.28	0.56	0.07	0.00	0.00
	[0.10]	[0.21]	[0.61]	[0.08]	[0.00]	[0.00]
Top 1%-0.1%	0.00	0.00	0.35	0.61	0.04	0.00
	[0.03]	[0.04]	[0.33]	[0.56]	[0.04]	[0.00]
Top 0.1%-0.01%	0.00	0.00	0.00	0.38	0.58	0.04
	[0.02]	[0.01]	[0.05]	[0.39]	[0.50]	[0.03]
Top 0.01%	0.00	0.00	0.00	0.00	0.37	0.63
	[0.02]	[0.01]	[0.02]	[0.09]	[0.34]	[0.53]

Table E.1: Top wealth mobility. The table shows the share of agents moving across different groups of the wealth distribution over a period of 25 years under our benchmark specification. The numbers in square parentheses are taken from Appendix Table 3 in Bach et al. (2020), where the authors report wealth mobility figures computed with Swedish administrative data between the years 2000 and 2007. Rows might not sum exactly to one due to rounding.

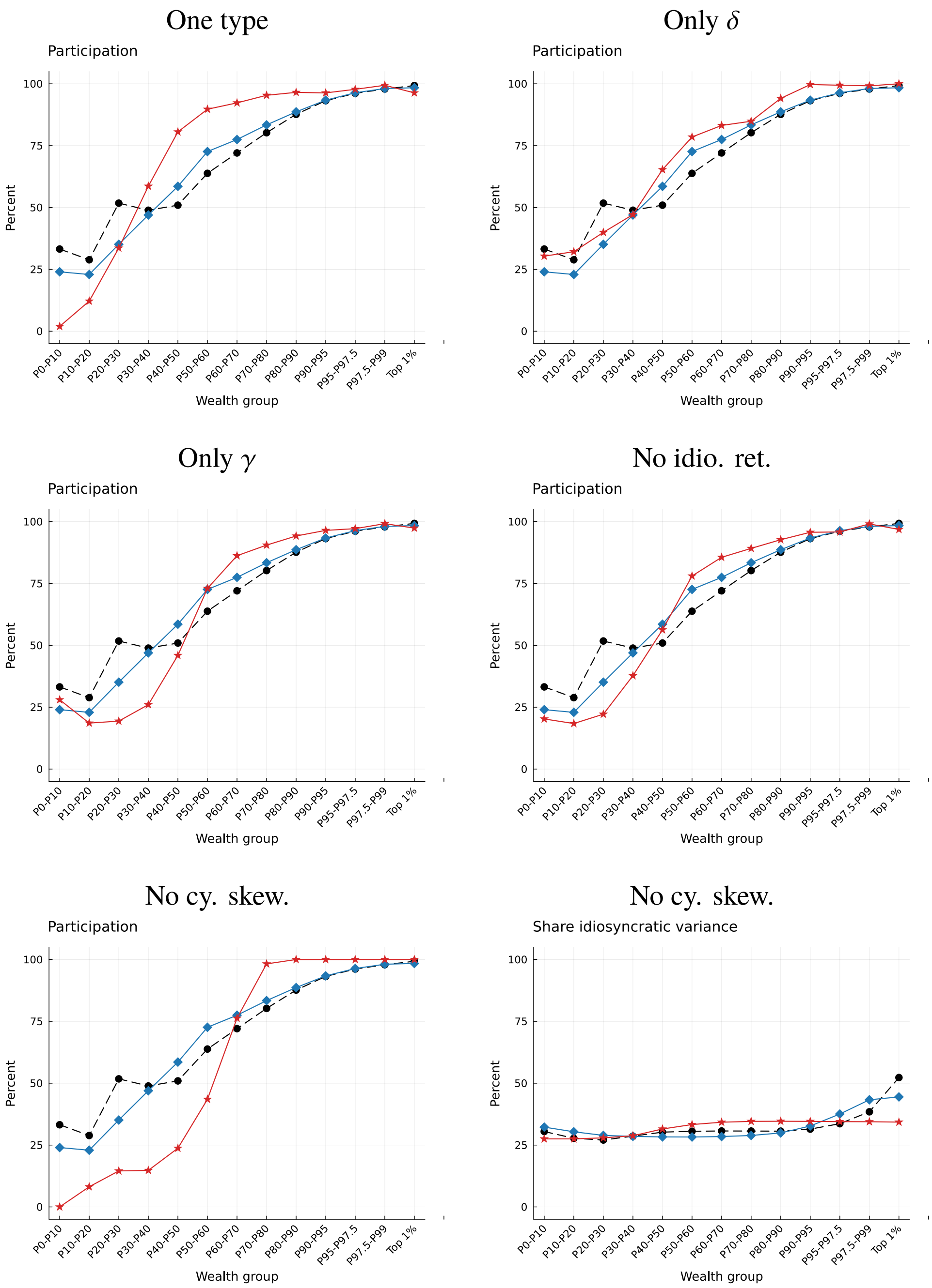

Figure E.1: Fit of the estimated model, benchmark vs. alternative specifications. The figure shows in the first five panels the schedules of participation over the wealth distribution from the benchmark model (diamond markers, solid line), from different alternative specifications (starred markers, solid line), and their data equivalents computed from the Swedish administrative data used by Bach et al. (2020) (circular markers, dashed line). We solve each model setting inverse EIS (ψ) equal to one. "One type" reports the results for the model with one type of agents, i.e., no heterogeneity in preference parameters and idiosyncratic return risk. "Only δ" for the model with two types differing only in δ. "Only γ" for the model with two types differing only in γ. "No idio. ret." for the model without idiosyncratic risk in returns in which we set ζ equal to zero for both types. "No cy. skew." for the model without cyclical skewness in labor income shocks. The last panel compares the schedules of the share of idiosyncratic return risk in the benchmark and in the model without cyclical skewness in labor income shocks (the different line patterns have the same meaning as in the fifth panel).

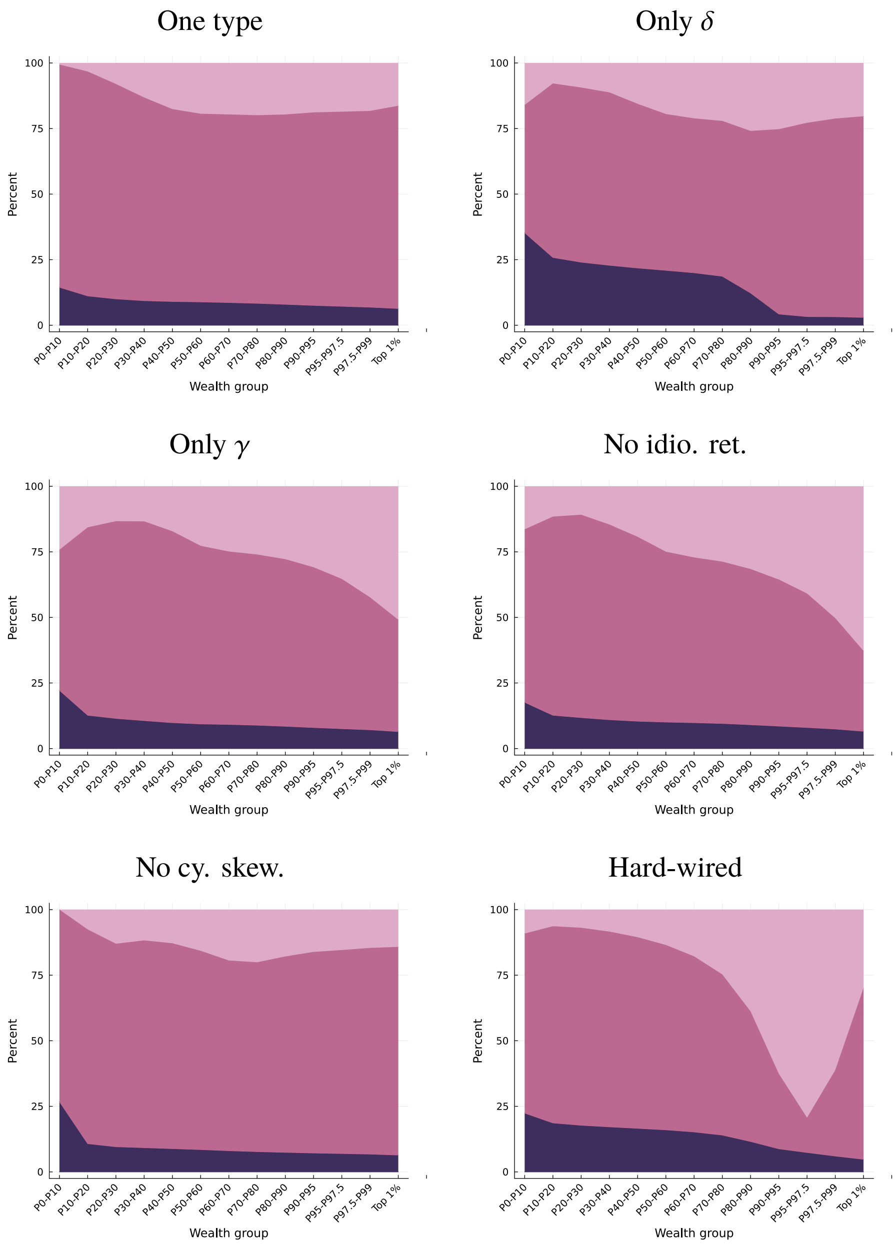

Figure E.2: Marginal propensities to consume and to invest over the wealth distribution, alternative specifications. The figure shows the schedules of the marginal propensities to consume (darker color), to invest in the safe asset and to invest in the risky asset (lighter color) over the wealth distribution obtained with alternative model specifications. "One type" indicates the model with one type of agents, i.e., no heterogeneity in preference parameters and idiosyncratic return risk. "Only δ" the model with two types differing in δ but not in γ and ζ. "Only γ" the model with two types differing in γ but not in δ and ζ. "No idio. ret." the model without idiosyncratic risk in returns in which we set ζ equal to zero for both types. "No cy. skew." the model without cyclical skewness in labor income shocks. For convenience we also report again the picture for the benchmark model. "Hard-wired" the model with risky share and share of idiosyncratic return variance exogenously set.

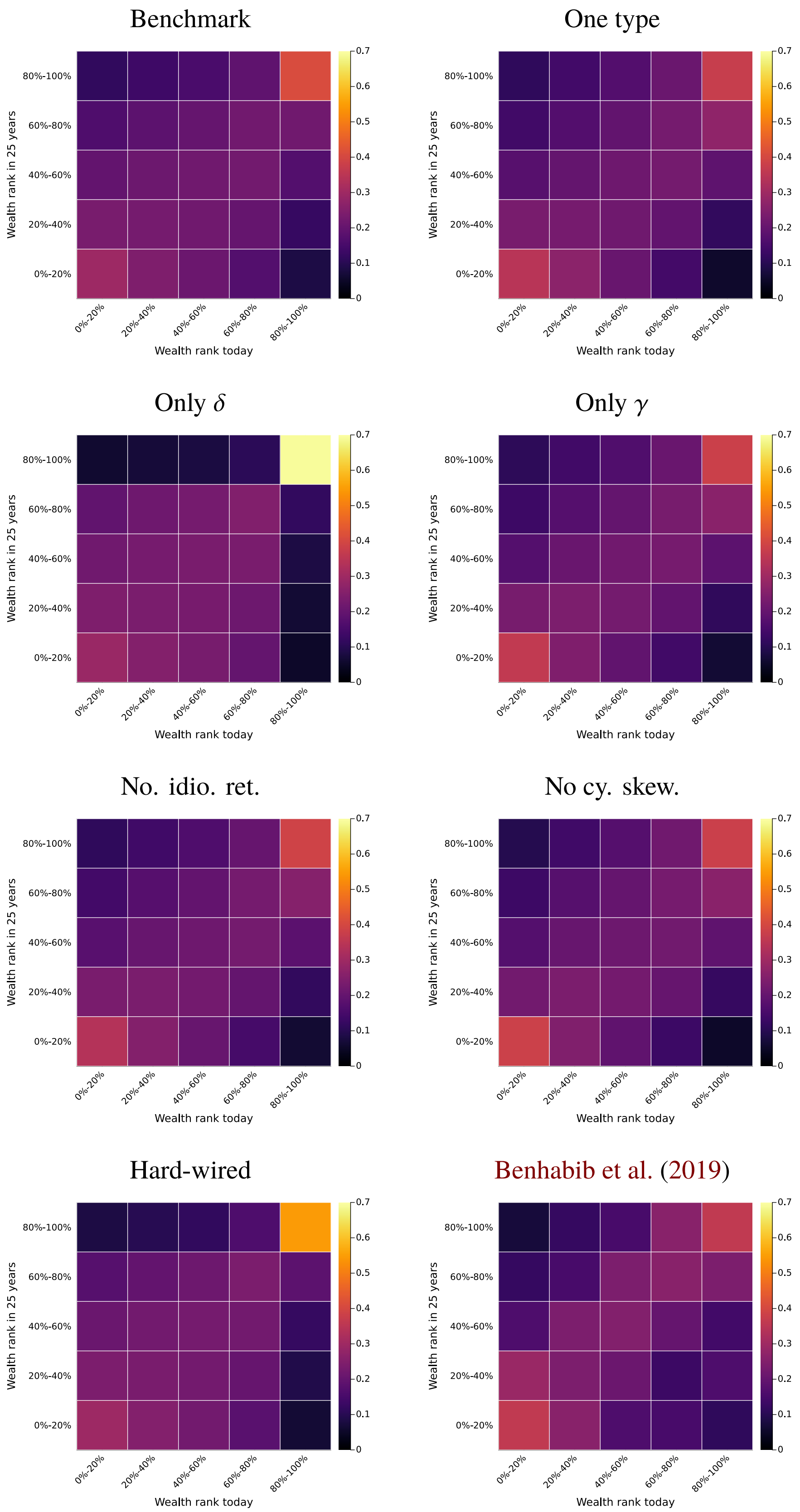

Figure E.3: Wealth mobility, benchmark vs. alternative specifications. The heatmaps show the share of agents moving across the quintiles of the wealth distribution over a period of 25 years for different model specifications. "One type" indicates the model with one type of agents, i.e., no heterogeneity in preference parameters and idiosyncratic return risk. "Only δ" the model with two types differing in δ but not in γ and ζ. "Only γ" the model with two types differing in γ but not in δ and ζ. "No idio. ret." the model without idiosyncratic risk in returns in which we set ζ equal to zero for both types. "No cy. skew." the model without cyclical skewness in labor income shocks. "Hard-wired" the model with risky share and share of idiosyncratic return variance exogenously set. For convenience we also present the parent-child intergenerational wealth mobility figures computed by Charles and Hurst (2003) with PSID data and reported by Benhabib et al. (2019).

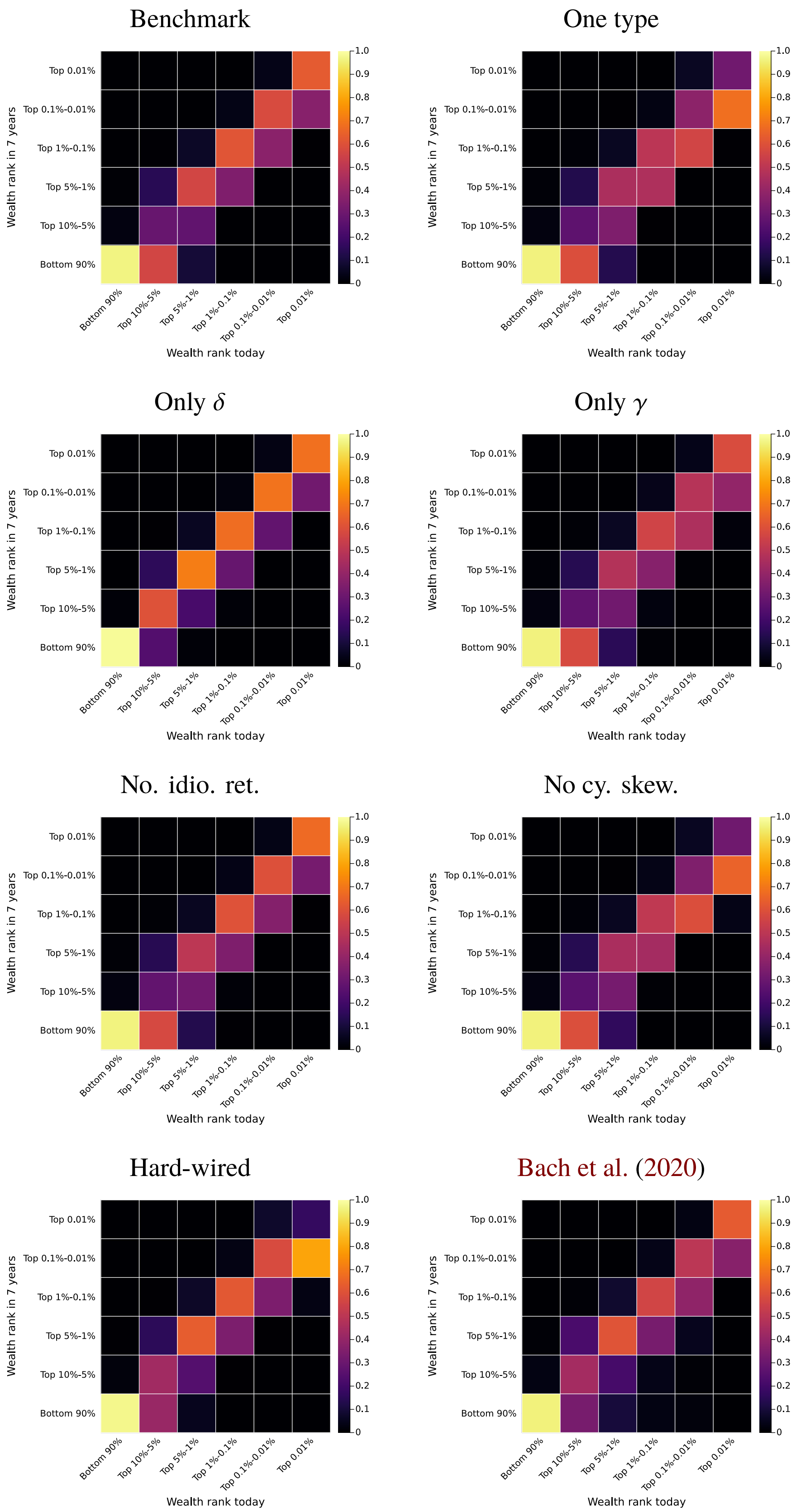

Figure E.4: Top wealth mobility, benchmark vs. alternative specifications. The heatmaps show the share of agents moving across different groups of the wealth distribution over a period of 7 years under different model specifications. "One type" indicates the model with one type of agents, i.e., no heterogeneity in preference parameters and idiosyncratic return risk. "Only δ" the model with two types differing in δ but not in γ and ζ. "Only γ" the model with two types differing in γ but not in δ and ζ. "No idio. ret." the model without idiosyncratic risk in returns in which we set ζ equal to zero for both types. "No cy. skew." the model without cyclical skewness in labor income shocks. "Hard-wired" the model with risky share and share of idiosyncratic return variance exogenously set. For convenience we also present the wealth mobility figures reported in Appendix Table 3 in Bach et al. (2020) computed with Swedish administrative data between the years 2000 and 2007.

F Robustness checks

F.1 Residential real estate and pension wealth as safe assets

When estimating the benchmark specification, we assume that the risky share within residential real estate and pension wealth is equal to the one found for the restricted portfolio including only financial wealth, private equity and commercial real estate. The key reason for this choice is that assuming these two assets to be fully safe could possibly underestimate the amount of risk borne by agents in the middle and left part of the distribution for whom these two categories constitute a substantial part of their wealth – see Figure 2 in Bach et al. (2020) – and, as a consequence, impact the parameter estimates. In this section, therefore, we investigate how our results change when these two asset types are classified as safe. Specifically, we recompute from the Swedish registry data the portfolio schedules of expected excess returns, participation and share of idiosyncratic return variance assuming that the risky share of residential real estate and pension wealth is zero and re-estimate our benchmark specification.

As expected, Figure F.1 shows that the empirical portfolio schedules increase more markedly at the top than before. As a consequence of the less smooth schedules, the algorithm matches them through a starker separation of types, as shown by the values of the new parameter estimates, which, in turn, generates too much concentration at the very top.[48]

Summing up, the key message is that when residential real estate and pension wealth are classified as safe, the mechanisms and implications of our model are overall the same.

	Type 1			Type 2					
Model	δ	γ	ζ	δ	γ	ζ	Share of Type 1	f	$\bar{d}$
Benchmark	0.924	9.891	0.543	0.966	2.671	0.825	0.901	0.001	−0.835
Res. & pens. safe	0.924	10.141	0.583	0.980	2.864	1.036	0.983	0.000	−0.863
EIS = 0.5	0.893	9.182	0.554	0.965	3.031	0.851	0.917	0.000	−0.804
EIS = 1.5	0.922	8.387	0.571	0.976	3.242	0.797	0.945	0.001	−0.367
EIS = {0.1, 1}	0.699	8.901	0.538	0.978	6.232	0.944	0.906	0.000	−0.763

Table F.1: Estimated model parameters, robustness. We solve our benchmark model setting inverse EIS (ψ) equal to one. "Res. & pens. safe" indicates a model with all heterogeneities active as in the benchmark but targeting portfolio choice schedules with residential real estate and pension wealth considered safe. "EIS = 0.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2$ for both types. "EIS = 1.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2/3$ for both types. "EIS = $\{0.1, 1\}$" indicates a model with all heterogeneities active as in the benchmark but $\psi = 10$ for type-one and $\psi = 1$ for type-two, in the spirit of Guvenen (2006).

[48] The wealth-to-income ratio and the share of agents with negative wealth are always almost perfectly matched when re-estimating the model under all the alternative specifications considered in this robustness section.

	Model					Data
Share held by (%):	Bench.	Res. & pens. safe	EIS = 0.5	EIS = 1.5	EIS = {0.1, 1}	Sweden (2000-2007)
Q1	0.1	0.2	0.1	0.2	0.3	-1.1
Q2	4.6	4.7	4.4	3.1	4.9	2.8
Q3	8.8	8.8	8.8	6.0	10.0	8.7
Q4	15.8	15.1	15.8	10.7	18.9	19.4
Q5	70.6	71.2	70.9	80.0	65.8	70.2
90-95 %	11.9	9.5	11.9	9.4	14.3	13.4
95-99 %	22.5	15.5	21.6	28.8	20.4	17.9
Top 1 %	22.5	33.6	23.4	31.9	13.7	21.3
Wealth Gini	0.69	0.70	0.69	0.78	0.64	0.71

Table F.2: Wealth inequality, robustness. The table reports the share of wealth held between different quantiles of the wealth distribution and the Gini coefficient. We solve our benchmark model setting inverse EIS (ψ) equal to one. "Res. & pens. safe" indicates a model with all heterogeneities active as in the benchmark but targeting portfolio choice schedules with residential real estate and pension wealth considered safe. "EIS = 0.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2$ for both types. "EIS = 1.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2/3$ for both types. "EIS = $\{0.1, 1\}$" indicates a model with all heterogeneities active as in the benchmark but $\psi = 10$ for type-one and $\psi = 1$ for type-two, in the spirit of Guvenen (2006). The last column present the data equivalents computed, with the Swedish administrative data compiled by Bach et al. (2020) – averaged over the 2000-2007 period.

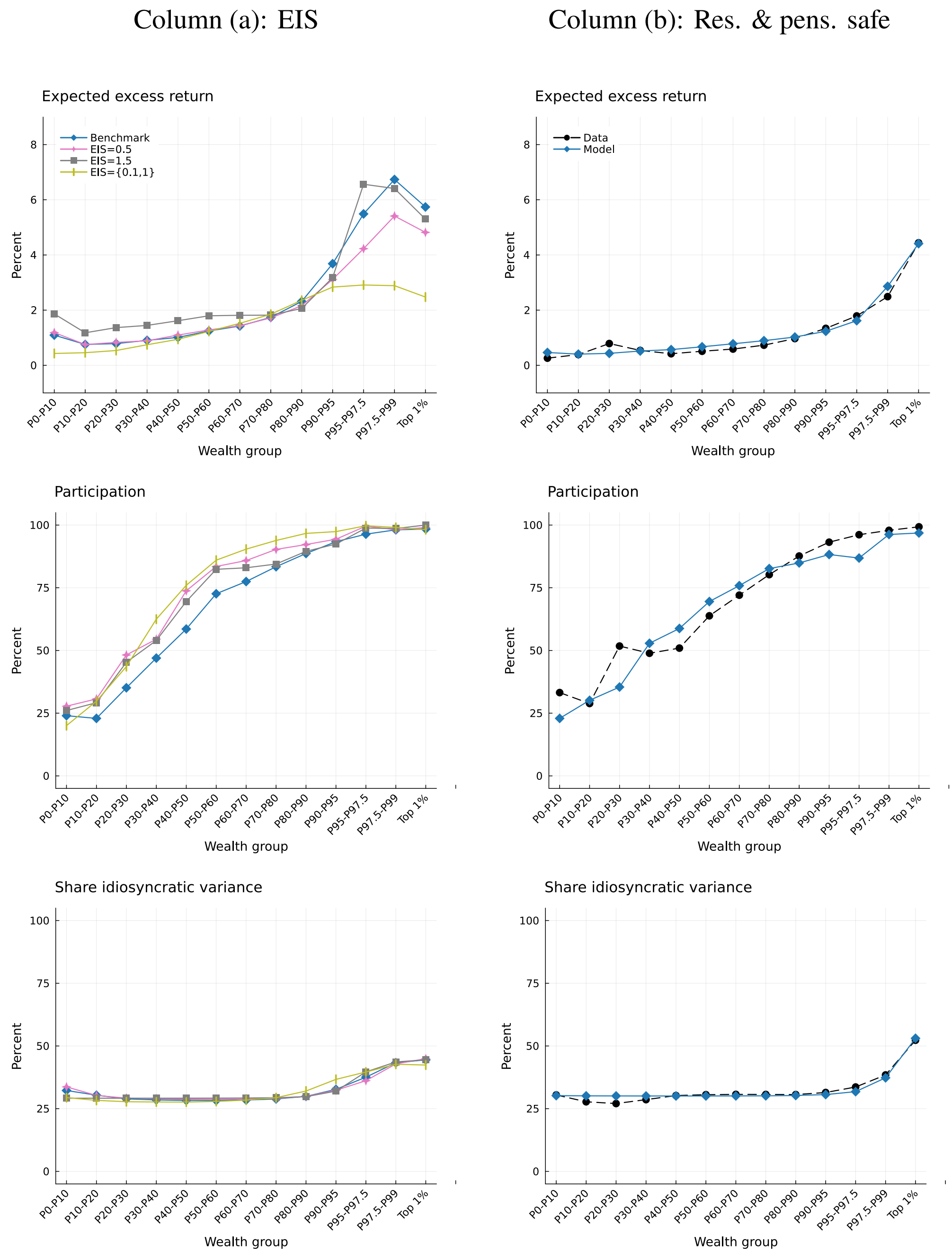

Figure F.1: Fit of the estimated model, robustness. The figure compares the schedules of expected excess returns, participation and share of idiosyncratic return variance over the wealth distribution from the benchmark model (diamond markers, solid line) with those obtained with different EIS values in Column (a) and when targeting portfolio choice schedules with residential real estate and pension wealth considered safe in Column (b) (starred markers, solid line). "EIS = 0.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2$ for both types (starred markers with four edges, solid line). "EIS = 1.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2/3$ for both types (squared markers, solid line). "EIS = $\{0.1, 1\}$" indicates a model with all heterogeneities active as in the benchmark but $\psi = 10$ for type-one and $\psi = 1$ for type-two, in the spirit of Guvenen (2006) (vertical markers, solid line). Data schedules in column (b) (circular markers, solid line) were computed from the Swedish administrative data used by Bach et al. (2020).

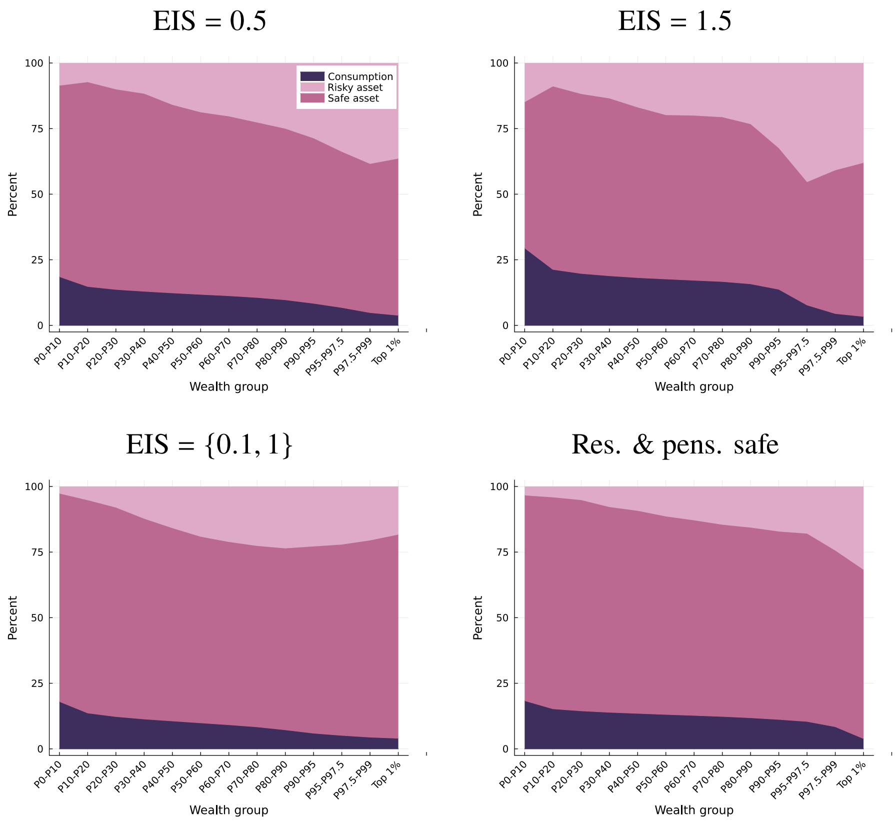

Figure F.2: Marginal propensities to consume and to invest over the wealth distribution, robustness. The figure shows the schedules of the marginal propensities to consume (darker color), to invest in the safe asset and to invest in the risky asset (lighter color) over the wealth distribution obtained with alternative model specifications. "EIS = 0.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2$ for both types. "EIS = 1.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2/3$ for both types. "EIS = $\{0.1, 1\}$" indicates a model with all heterogeneities active as in the benchmark but $\psi = 10$ for type-one and $\psi = 1$ for type-two, in the spirit of Guvenen (2006). "Res. & pens. safe" indicates a model with all heterogeneities active as in the benchmark but targeting portfolio choice schedules with residential real estate and pension wealth considered safe.

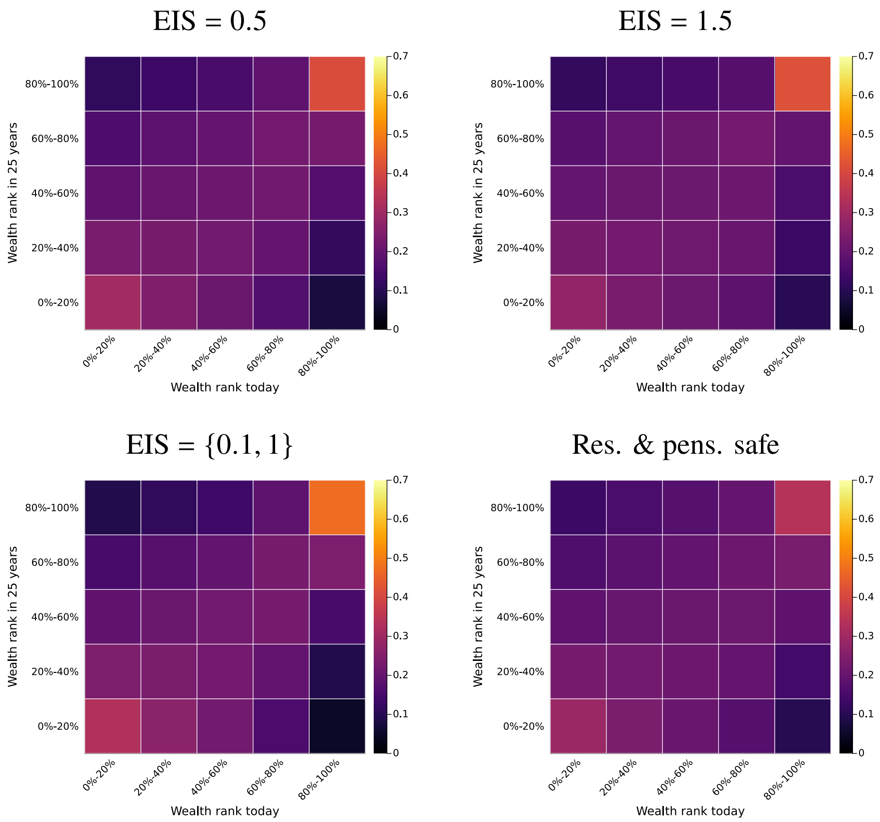

Figure F.3: Wealth mobility, robustness. The heatmaps show the share of agents moving across the quintiles of the wealth distribution over a period of 25 years for different model specifications. "EIS = 0.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2$ for both types. "EIS = 1.5" indicates a model with all heterogeneities active as in the benchmark but $\psi = 2/3$ for both types. "EIS = $\{0.1, 1\}$" indicates a model with all heterogeneities active as in the benchmark but $\psi = 10$ for type-one and $\psi = 1$ for type-two, in the spirit of Guvenen (2006). "Res. & pens. safe" indicates a model with all heterogeneities active as in the benchmark but targeting portfolio choice schedules with residential real estate and pension wealth considered safe.

F.2 Different EIS values

As discussed in Section 3, the lack of assets with different liquidity in our model prevents us from jointly identifying the EIS and discount factor. For this reason, in our benchmark specification we set ψ equal to one for both types. In this section, as a robustness, we check how our results change when using different values for this parameter. Specifically, we re-estimate three versions of the model with all heterogeneities active as in the benchmark: in the first we assume that both agents have EIS equal to 0.5 ($\psi = 2$), in the second that they both have EIS equal to 1.5 ($\psi = 2/3$) and in the third that type-one agents have EIS equal to 0.1 ($\psi = 10$) while type-two agents equal to 1.[49]

[49] The third parametrization is in the spirit of Guvenen (2006), where heterogeneity in EIS generates wealth dispersion as low EIS agents are exogenously restricted from investing in the risky asset while high EIS agents are not. In his model there are also two types of agents, non-stockholders with EIS equal to 0.1 and stockholders with EIS equal to 1.

As shown in Figure F.1, the overall match of these alternative specifications replicates quite similarly the patterns in the benchmark. Table F.1 reports the estimated parameters. A lower EIS value means a stronger consumption smoothing motive, which implies higher willingness to insure against shocks especially for agents close to the borrowing constraint. To counteract this force, the estimated discount factor and risk aversion for type-one agents are lower. Yet, the changes in parameters are not sizable and this alternative model also delivers a good match of wealth inequality and similar patterns for mobility and marginal propensities (see Figures F.2 and F.3). On the other hand, while generating higher MPCs at the bottom of the wealth distribution, the higher δ for type-two obtained when EIS is equal to 1.5 delivers too much concentration at the top. Turning to the case in which EIS is 0.1 for type-one and 1 for type-two agents, the estimation balances the lower EIS of the former individuals by substantially reducing their discount factor and risk aversion. The opposite happens for the latter type, whose idiosyncratic return risk is also increased. However, to compensate for the reduction in total savings from the new parameters of the other type, more patient agents start to populate the right tail of the wealth distribution at lower quantiles compared to the benchmark, which, together with their higher risk aversion, results in a too mild increase in expected excess returns at the top, in a lower share of wealth held by the top 1% and in lower top mobility.

In conclusion, there are two key lessons from this exercise. First, in the EIS = 0.5 case we get similar results as in our benchmark as the estimation procedure adjusts the other parameters – especially the discount factor – accordingly. This shows the robustness of one of the key mechanisms behind our results, namely the compositional effects ensuing from the separation of the two types over the distribution, but also confirms our discussion in Section 3 on the difficulty of identifying both δ and ψ without heterogeneity in assets' liquidity. Nevertheless, some EIS values (i.e., EIS = $\{0.1, 1\}$) do not even allow a satisfactory match of the targeted schedules. Second, too high EIS values for both individuals would generate counterfactually too large wealth holdings at the top. This is because the estimation cannot compensate by raising the discount factor of type-one agents, as doing that would require also an increase in the patience of the other type to keep enough separation to match the increasing portfolio schedules which, in turn, would generate a too high wealth-to-income ratio. As a consequence, only δ of type-two individuals is increased, which allows to get a good match of the targeted moments but a worse fit of (untargeted) wealth inequality moments.