Preference heterogeneity and portfolio choices over the wealth distribution

Gualtiero Azzalini* Markus Kondziella † Zoltán Rácz ‡

IIES IIES SSE

Date: May 4, 2023

Abstract

What are the key elements to generate portfolio choices over the wealth distribution in line with the data? In this paper, we argue that capturing preference heterogeneity across individuals is one of them. Using a partial equilibrium Bewley-type model with endogenous portfolio choice and cyclical skewness in labor income shocks, we show that heterogeneity in risk aversion, impatience and portfolio diversification is crucial to match the empirical schedules of unconditional risky share, participation and share of idiosyncratic variance in individual portfolios. At the same time, these elements generate dispersion in wealth through their heterogeneous effects on individuals' investment decisions resulting in a cross-sectional wealth distribution that provides a close fit of the data, particularly at the very top.

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 Tobias Broer, Karin Kinnerud, Richard Foltyn and seminar participants at the IIES Macro Group for useful feedback and comments.

*Institute for International Economic Studies, Stockholm University, gualtiero.azzalini@iies.su.se

†Institute for International Economic Studies, Stockholm University, markus.kondziella@iies.su.se

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

1 Introduction

Increased availability of data on portfolio holdings and wealth at the individual level has spurred growth in the literature on portfolio choice and capital income returns. On the empirical side, Bach et al. (2020) and Fagereng et al. (2020) study how portfolio choice and return heterogeneity vary over the wealth distribution. On the theoretical side, Benhabib et al. (2011) and Hubmer et al. (2021) emphasize the importance of return heterogeneity in explaining wealth accumulation over time at the individual level and wealth inequality in the cross-section.

Despite these recent advancements, explaining individuals' portfolio choices and cross-sectional wealth inequality remain two challenging issues in household finance and macroeconomics, respectively. In this paper, we show that connecting the two literatures by introducing a macroeconomic angle to the recent empirical findings in the household finance literature can help to address both issues simultaneously. Specifically, we extend an otherwise standard incomplete markets model along several dimensions to generate portfolio choice patterns consistent with the empirical findings and show that such extensions improve the match of the cross-sectional wealth distribution in the data, particularly at the very top.

The core of our framework is a Bewley model, the workhorse for studying the interplay between the wealth distribution and macroeconomic aggregates. We add to the standard setting endogenous portfolio choice, a non-normal return process, cyclical skewness in labor income shocks, Epstein-Zin preferences and preference heterogeneity. The latter includes heterogeneity in individuals' time preference rate (TPR), risk aversion and ability or inclination towards portfolio diversification.[1] The result is a hybrid between Bewley-type and portfolio choice models in the household finance literature.

We estimate the unobserved model parameters governing the heterogeneity in preferences to match the increasing risky share, participation rate and share of idiosyncratic return risk over the wealth distribution documented in Bach et al. (2020). As an outcome of the estimation, the economy is populated by the following two types: one type features a relatively high risk aversion and impatience (parameter values commonly used in the household finance literature), whereas the other type is characterized by lower risk aversion and impatience (parameter values commonly used in the macroeconomics literature). The former type also features a higher preference for portfolio diversification than the latter.

The combination of individuals of both types delivers portfolio choice patterns that closely match the patterns in the data. The aforementioned introduction of preference heterogeneity and

[1] Our specification allows also for heterogeneity in the elasticity of intertemporal substitution (EIS). However, as discussed below in Section 3, in this paper we abstract from that because our framework does not allow to identify well this parameter.

the rich stochastic process governing income and returns - which we borrow from Catherine (2021)[2] - are crucial to generate the increasing relation between the risky share and wealth quantiles. In particular, the income process is relevant to explain the risky share towards the bottom of the wealth distribution and the preference heterogeneity towards the top. Intuitively, human capital is a higher share of net worth at the bottom, which makes the stochastic properties of income matter more than at the top, where, instead, preference parameters become the primary determinant of portfolio choices.

These two channels ensure a positive correlation between risky shares and wealth over the whole distribution. Towards the bottom, this trend follows from the hesitancy of asset-poor individuals to invest in the stock market because of the riskiness of their labor income. At the top, instead, such relationship is obtained only in combination with preference heterogeneity. Intuitively, the optimal risky share of rich agents converges to the constant in Merton (1969), as high wealth essentially protects the individual from non-linear features in our model. If all agents share identical preference parameters, generating the empirically substantial positive relationship between wealth and risky share is therefore impossible. However, preference heterogeneity causes less risk-averse agents with a higher risky share to endogenously end up on the top of the wealth distribution generating a positive relationship between risky shares and wealth in the cross-section. In addition, because these individuals are characterized by lower portfolio diversification, we also match the higher share of idiosyncratic variance at the top. Together with the fact that their higher degree of patience ensures that less risk-averse and less diversified individuals endogenously end up at the top of the distribution, these mechanisms also increase wealth inequality through higher expected returns among the richest.

Finally, to gauge the relative importance of the different elements in our framework, we compare the results in our benchmark model with counterfactuals in which we shut down different components one at a time. More in detail, we solve a version with homogeneous preferences, one with heterogeneity in just the time preference rate and another in just risk aversion, a version without endogenous portfolio choice, one without idiosyncratic returns and one without skewness in labor income and return shocks. Except for the case without idiosyncratic returns, in which (in line with Hubmer et al., 2021) we find relatively small changes, in all the other cases either the match of the portfolio schedules or of wealth inequality or of both is worsened.

[2]While we take from his paper most of the elements governing the joint stochastic process of income and returns, it is important to state that we use them for different purposes than his paper. Indeed, rather than explaining portfolio choices over the life-cycle dimension, we are interested in understanding them over the wealth distribution and how they affect inequality through their impact on the formation of individual returns.

Related literature. This paper contributes to both the household finance literature and to the macroeconomics literature on wealth inequality.

Our main contribution to the former is capturing endogenously realistic portfolio choices over the wealth distribution as documented in Swedish registry data by Bach et al. (2020). In doing that, our paper relates to the literature studying the interplay between income, preference heterogeneity and portfolio choices.

Both theoretical and empirical studies show that labor income is a determinant of portfolio choice. For example, the persistent component of labor income is linked to human capital, and from theory starting with Merton (1969) we know the latter influences participation and the risky share. Fagereng et al. (2017) and Chang et al. (2018) further emphasize that the riskiness of labor income influences portfolio choice (the riskier labor income, the lower the risky share). Using Swedish registry data Catherine et al. (2020) show that workers facing higher cyclical skewness display lower risky shares. In line with these findings - and those in Guvenen et al. (2014) - we follow Catherine (2021) and include an income process in our model that features skewness of idiosyncratic income shocks that is linked to movements in aggregate returns.

While including this central feature of Catherine (2021), we also extend his setting to allow a rich set of parameters governing preference heterogeneity.[3] Thus, our paper also relates to the literature studying the role of the latter in portfolio choice models (e.g. Vestman, 2018).[4] In addition, 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 studies emphasize that capturing return heterogeneity is a crucial component to match the shape of the wealth distribution. Benhabib et al. (2011) show analytically that the introduction of stochastic idiosyncratic returns implies a wealth distribution that is Pareto. In their model, wealth accumulation and decumulation occur randomly. Benhabib et al. (2015), Nirei and Aoki (2016) and Gabaix et al. (2016) are further examples that introduce heterogeneous returns to the consumption-savings decision. Benhabib et al. (2019) quantitatively show that heterogeneous returns jointly with savings and bequests behavior are crucial elements for top wealth inequality and to explain social mobility. Hubmer et al. (2021) study plausible explanations for the increase in wealth inequality in the U.S. Heterogeneity in asset returns turns out to be key to account for the dynamics in wealth inequality.

[3]Also differently from his framework, we allow for idiosyncratic return shocks and use an infinite horizon setup.

[4]Vestman (2018) investigates the effects of joint heterogeneity in risk aversion, EIS and participation cost on stock market participation patterns and their connection with home ownership. However, he does not consider the risky share, which is the main focus of this paper. Furthermore, contrary to Vestman (2018), instead of presetting preference parameters, we are estimating them.

The return on assets is modelled as an increasing function of wealth plus an idiosyncratic shock. Thus, individuals end up with different returns both because they have different wealth levels (which can, potentially, be controlled through the savings decision) and because of randomness. That returns on assets are increasing in wealth can be interpreted as reduced-form portfolio choice that is consistent with the results by Bach et al. (2020) found in Swedish registry data.

Despite the fact that portfolio choice is a crucial component to generating individual returns, the above papers take shortcuts in obtaining return heterogeneity. In order to take the driver of return heterogeneity into account in models analysing the wealth distribution, it is of first-order importance to endogenize an individual's investment decision between different kinds of assets. Our contribution to this literature is, therefore, adding realistic endogenous portfolio choices to this class of models, and showing that the latter is crucial to capture wealth inequality. In doing that, we also try to connect this research area with the household finance literature described above.

Finally, as one important element in our framework to achieve a good match of the wealth distribution in the data is preference heterogeneity, we also relate to the papers emphasizing the role of the latter for inequality (see De Nardi and Fella, 2017, for a review). The new element in our paper is considering a richer structure compared to what has been done so far. In particular, our paper explores the consequences of introducing heterogeneity in impatience, risk-aversion and lack of diversification[5] - and of allowing correlations between them - on the wealth distribution through their impact on both agents' optimal consumption-savings and portfolio choices. As we will see below, we find this to be an important element to fit the data and to investigate the relative importance of different channels.

The paper is structured as follows. Section 2 outlines the model, section 3 describes our calibration procedure, section 4 presents results on our benchmark specification, section 5 investigates counterfactuals and section 6 concludes.

2 Model

Agents and preferences. The economy is populated by a continuum of infinitely lived ex-ante identical individuals deriving utility from consumption ($c_{i,t}$) through Epstein-Zin preferences. Agents differ in terms of preference parameters. We capture this heterogeneity with an individual-specific preference state ($\theta_{i,t}$) which, in turn, determines impatience (δ), risk aversion (γ), the

[5]Including lack of diversification as a preference parameter allows us to match on the part of the wealth distribution in which idiosyncratic return variance is more important, and hence we can investigate the relative importance of idiosyncratic return shocks realistically.

inverse of the elasticity of intertemporal substitution (ψ) and the lack of diversification[6] (ζ). Preferences are, then, given by the following expression:

$$U_{i,t} = \left[(1-\delta(\theta_{i,t}))c_{i,t}^{1-\psi(\theta_{i,t})} + \delta(\theta_{i,t})\left(\mathbb{E}_t U_{i,t+1}^{1-\gamma(\theta_{i,t})}\right)^{\frac{1-\psi(\theta_{i,t})}{1-\gamma(\theta_{i,t})}}\right]^{\frac{1}{1-\psi(\theta_{i,t})}}$$

Financial assets. Agents can invest in two financial assets, one risky with time-varying individual-specific gross return $R_{i,t+1}$ and one safe with constant gross return R^f. Letting small letters indicate log returns, $r_{i,t+1}$ is equal to:

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

The effective return individual i gets by investing in the risky asset is the sum of two systematic components, one co-varying with labor market conditions (r_1) and one that does not (r_2), of an idiosyncratic component (η) and is net of management cost m, that is thus paid conditional on holding the risky asset. The systematic components are modeled as in Catherine (2021). Specifically, to take into account stock market crashes, r_1 is distributed as a mixture of Normals:

$$r_{1,t+1} = \begin{cases} \underline{r}_{1,t+1} \overset{i.i.d.}{\sim} \mathcal{N}\left(\underline{\mu}_r, \sigma_{r_1}^2\right) & \text{w.p.} \quad p_r \\ \overline{r}_{1,t+1} \overset{i.i.d.}{\sim} \mathcal{N}\left(\overline{\mu}_r, \sigma_{r_1}^2\right) & \text{w.p.} \quad 1-p_r \end{cases} \tag{2}$$

Without loss of generality, we interpret p_r as the probability of stock market crashes and $\underline{\mu}_r$ the expected log return during these periods. Similarly, $1-p_r$ is the probability of normal periods and $\overline{\mu}_r$ the corresponding average log return. The other systematic component, r_2, is drawn from a Normal distribution:

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

Finally, the idiosyncratic component, $\eta_{i,t+1}$, is modeled as follows:

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

where $\sigma_{ir} = \sigma_r\zeta(\theta_{i,t})$, with σ_r being the standard deviation of the systematic part of the log return and the preference parameter $\zeta(\theta_{i,t})$ governing the share of idiosyncratic risk in total portfolio volatility. Note that this specification ensures that idiosyncratic risk is not priced since the idiosyncratic part does not affect the mean return.[7]

Introducing the idiosyncratic component enables us to understand the relative importance of systematic and idiosyncratic return shocks. However, as standard portfolio choice models imply that everyone should invest in some efficient mixture of riskless and risky assets, it is far from straightforward to represent lack of diversification in a framework otherwise based on optimizing behavior.

[6]The next paragraph provides a detailed explanation of lack of diversification.

[7]For the exact formulas of the statistical moments of the full return process see Appendix C.

Our modeling choice implies the following: 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 preference-dependent idiosyncratic risk. While guaranteeing that idiosyncratic risk is not priced, this strategy also implies - 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.

However, linking the share of idiosyncratic risk to a stable preference type also effectively restricts the domain of portfolio composition decisions. In other words, we do not model how lack of diversification arises (e.g. financial knowledge, overconfidence, reliance on private equity) but, rather, capture in reduced form that agents' ability or desire to diversify is limited and that they optimally decide how to allocate their wealth given this constraint. Thus, our approach lies between a completely micro-founded, realistic model of portfolio choice in the presence of a menu of different risky assets, and a framework in which the stochastic properties of returns over the wealth distribution are hard-wired (Hubmer et al., 2021).

Investing in the risky asset is subject to a fixed participation cost f that is paid in every period the agent chooses to hold that asset. Finally, individuals face a borrowing limit on their total savings proportional to the exogenously set parameter $\bar{s}$. The repayment rate per unit of borrowing is equal to the risk-free rate.

Labor income process. We follow Catherine (2021) for modeling the stochastic process governing labor income. Let $y_{i,t}$ denote the residual of log individual earnings. We assume that $y_{i,t}$ is the sum of an aggregate component (w_t) and of two idiosyncratic components, one persistent ($z_{i,t}$) and one transitory ($\nu_{i,t}$):

$$y_{i,t} = w_t + z_{i,t} + \nu_{i,t} \tag{3}$$

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

$$w_t = g + w_{t-1} + \lambda_{rw} r_{1,t} + \phi_t \tag{4}$$

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

The persistent component is an AR(1) process:

$$z_{i,t} = \rho z_{i,t-1} + \varepsilon_{i,t} \tag{5}$$

with innovations drawn from a mixture of Normals:

$$\varepsilon_{i,t} = \begin{cases} \underline{\varepsilon}_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(\underline{\mu}_{\varepsilon,t}, \underline{\sigma}_\varepsilon^2\right) & \text{w.p.} \quad p_\varepsilon \\ \overline{\varepsilon}_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(\overline{\mu}_{\varepsilon,t}, \overline{\sigma}_\varepsilon^2\right) & \text{w.p.} \quad 1 - p_\varepsilon \end{cases} \tag{6}$$

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 skewness, $\underline{\mu}_{\varepsilon,t}$ is defined as:

$$\underline{\mu}_{\varepsilon,t} = \mu_{\varepsilon} + \lambda_{\varepsilon w}(w_t - w_{t-1}) \tag{7}$$

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:

$$p_{\varepsilon}\underline{\mu}_{\varepsilon,t} + (1 - p_{\varepsilon})\overline{\mu}_{\varepsilon,t} = 0 \tag{8}$$

Finally, the transitory shock is Normally distributed, with variance depending on whether the persistent shock was drawn from the tail distribution or not:

$$\nu_{i,t} = \begin{cases} \underline{\nu}_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \underline{\sigma}_{\nu}^{2}\right) & \text{if} \quad \varepsilon_{i,t} = \underline{\varepsilon}_{i,t} \\ \overline{\nu}_{i,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \overline{\sigma}_{\nu}^{2}\right) & \text{if} \quad \varepsilon_{i,t} = \overline{\varepsilon}_{i,t} \end{cases} \tag{9}$$

As discussed in detail by Catherine (2021), including countercyclical income risk enables to obtain realistic portfolio choices over the life cycle. Intuitively, if adverse income shocks occur with greater probability when the stock market is low, agents with relatively high human capital (i.e. the young and the poor) will be more cautious when investing in risky financial instruments.

The reason for retaining this feature in our framework is two-fold. First, examining whether a mechanism allowing to match portfolio choices over age can also produce realistic risk-taking patterns over the wealth distribution is an interesting question. Second, by including a potential alternative channel of generating increasing risky shares over the wealth distribution relative to preference heterogeneity, we let the estimation decide whether the latter is crucial to match the desired patterns.

Safety net. To take into account how welfare programs potentially affect consumption, savings and portfolio choice, we follow Catherine (2021) who models the Supplemental Nutrition Assistance Program (sometimes called food stamp program). Specifically, working-age individuals with wealth holdings below 5% of the average national wage and earnings below 20% of the wage index receive 6% of the wage index minus 30% of their earnings as benefits. Mathematically:

$$b_{i,t} = \max\left\{0.06 \cdot \exp(w_t) - 0.3 \cdot \exp(y_{i,t}), 0\right\} \quad \text{if} \quad a_{i,t} < 0.05 \cdot \exp(w_t) \quad \text{and} \quad \exp(y_{i,t}) < 0.2 \cdot \exp(w_t) \tag{10}$$

where $b_{i,t}$ denotes the benefits and $a_{i,t}$ cash-on-hand, which is defined in the next paragraph.

The optimization problem. At the beginning of each period t, the agent enters with given cash-on-hand $a_{i,t}$, persistent income $z_{i,t}$, preference state $\theta_{i,t}$, and aggregate income w_t. She then chooses how much to consume in the current period $c_{i,t}$, how much to save for the next period $s_{i,t}$, whether to hold risky assets $F_{i,t}$ (dummy equal to one if she participates) and, conditional on participation, the share of savings invested in risky assets $\xi_{i,t}$.

Let $\Xi_{i,t} := (a_{i,t}, z_{i,t}, \theta_{i,t}, w_t)$ denote the state, $R^f := \exp(r^f)$ the gross risk free return and $R^e_{i,t+1} := \exp(r_{i,t+1}) - R^f$ the excess return. Then the maximization problem of agent i is:

$$V(\Xi_{i,t}) = \max_{\{c_{i,t}, s_{i,t}, \xi_{i,t}, F_{i,t}\}} \left\{ (1-\delta(\theta_{i,t}))c_{i,t}^{1-\psi(\theta_{i,t})} + \delta(\theta_{i,t}) \left(\mathbb{E}_t \left[V(\Xi_{i,t+1})^{1-\gamma(\theta_{i,t})} \right] \right)^{\frac{1-\psi(\theta_{i,t})}{1-\gamma(\theta_{i,t})}} \right\}^{\frac{1}{1-\psi(\theta_{i,t})}} \tag{11}$$

subject to

$$c_{i,t} + s_{i,t} + F_{i,t} f = a_{i,t} \tag{12}$$

$$a_{i,t+1} = \left[R^f + \xi_{i,t} R^e_{i,t+1} \right] s_{i,t} + \exp(y_{i,t+1}) + b_{i,t+1} \tag{13}$$

$$s_{i,t} \geq \bar{s} \cdot \exp(w_t) \tag{14}$$

The borrowing constraint (14) varies over time through the dependence on the aggregate part of labor income w_t, which can be interpreted as this constraint becoming tighter in recessions and looser in expansions.[8] Appendix B describes in detail how the model is solved numerically.

3 Estimation and calibration

The goal of the calibration is to deliver a parametrized model in line with novel empirical evidence on portfolio choice over the wealth distribution documented in Bach et al. (2020). To do so, we follow a two-step approach. First, we exogenously set the parameters governing the income and return processes. Second, we estimate the fixed participation cost and individuals' preference parameters. We describe in the following the details of the procedure adopted.

3.1 Exogenously set parameters

Since we model the stochastic processes governing income and returns as in Catherine (2021), we use the same parameter estimates reported in his paper. While an extensive explanation of the estimation procedure can be found there, we still provide a brief description of the approach. The parameters governing the aggregate processes $(\underline{\mu}_r, \overline{\mu}_r, \sigma_{r_1}, \sigma_{r_2}, p_r, \sigma_\phi, g, \lambda_{rw})$, are estimated by Simulated Method of Moments (SMM) to capture the joint dynamics of log yearly SP500 returns and aggregate wage log growth from US Social Security panel data on earnings by targeting mean, standard deviation, third and fourth standardized moments (skewness and kurtosis) and the correlation between these two series. Estimation of the stochastic process for individual income requires, instead, to find values for $(p_\varepsilon, \mu_\varepsilon, \lambda_{\varepsilon w}, \underline{\sigma}_\varepsilon, \overline{\sigma}_\varepsilon, \underline{\sigma}_\nu, \overline{\sigma}_\nu, \rho)$. To do so,

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

SMM is used again targeting the time series between 1978 and 2010 of the standard deviation of log earnings growth at the one- and five-year horizons, Kelly's skewness of log earnings growth at the one-, three- and five-year horizons (from Guvenen et al., 2014) and the within-cohort variance of log earnings for ages between 25 and 60 (from Guvenen et al., 2021).

The risk-free rate r^f is set as in Catherine (2021) to 0.02, which is a standard value in the literature (e.g. Cocco et al., 2005; Gomes and Michaelides, 2005). Finally, no borrowing is allowed so $\bar{s}$ is set to zero. Table 1 summarizes all these parameter choices.

3.2 Estimated parameters

The main contribution of our estimation exercise is to obtain structural estimates of individuals' preference parameters - and of their heterogeneity - from their observed portfolio choices over the wealth distribution. To achieve this, we assume that the economy is populated by two types (i.e. the support of $\theta_{i,t}$ has two states).[9]

It is worth noticing that, as θ is a vector including time preference rate (δ), risk aversion (γ), inverse EIS (ψ), and lack of diversification (ζ), our model enables the investigation of potential heterogeneities across all these dimensions simultaneously. Nevertheless, as it cannot be well identified in our framework, in all the results reported from here onwards, ψ is set to unity for all types. Indeed, while the different role of ψ from that of risk aversion γ is discerned through the adoption of Epstein-Zin preferences, as highlighted by Aguiar et al. (2021), joint identification of EIS and time preference rate is problematic in a model without liquidity differences across assets.[10] In any case, we show in the results section below that heterogeneity in ψ is not necessary to match our targets.

Summing up, eight parameters are estimated in total: six preference parameters (δ, γ, ζ, for each type), the share of individuals of the first type, and the fixed participation cost.

Targets. The main targets of our estimation are portfolio choice patterns over the wealth distribution. To compute them, we rely on the data from Bach et al. (2020), compiled from Swedish administrative sources covering earnings and wealth holdings of all Swedish residents between 1999 and 2007, and on the figures already available in their appendix. While a detailed

[9]In our current specification, agents' preference types are assumed to remain fixed over time, as this reduces considerably the computational time. We have experimented with an extension of the model allowing transitions between types according to a Markov process in which the transition probabilities are jointly estimated with the other parameters and did not find not significant differences.

[10]The main idea behind identification in their framework is that low EIS agents care more about consumption smoothing and so invest more in liquid assets. We plan to include the joint estimation of the EIS with the remaining preference parameters in future versions of this paper.

description can be found in their paper, for our purposes it is worth reminding that wealth holdings cover cash, pension wealth, financial securities (including funds, stocks, derivatives, and bonds), private equity, real estate wealth, and debt. These data are then aggregated at the household level using household identifiers. 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.

When deciding how to allocate their savings, in our model individuals choose between a safe and a single "composite" risky asset. To map excess returns, participation and the share of idiosyncratic risk by asset type in Bach et al. (2020), into those of a composite risky asset, we proceed as follows. First, we classify the different assets into safe and risky: cash, money market funds, pension wealth and residential real estate belong to the former group, while all other securities, private equity and commercial real estate to the latter. We then define the participation rate for the composite asset as the share of people in each wealth quantile holding any risky asset classified as such according to the method just described.[11]. To obtain the excess return and the portfolio share of idiosyncratic risk in each wealth quantile, we multiply, instead, the reported share of wealth invested in each asset type by, respectively, the expected excess returns and the share of idiosyncratic risk for that particular asset. We further rescale these excess returns by the average yearly excess return of the SIXRX Swedish equity index, that is is 8.7% over the 1983-2016 period (Bach et al., 2020). This transforms excess returns into the implicit unconditional risky share invested in the risky asset and eases comparison with other studies in the literature.

Figure 1 displays the resulting schedules of the unconditional risky share, 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 constitute our calibration targets, together with the ratio of aggregate wealth to income - which in Sweden is equal to four as reported by Bach et al. (2018) - for a total of 49 moments.

Estimation results. The SMM estimation procedure comprises a global and a local stage.[12] In the global stage, we generate 1,000 parameter vectors from a Sobol sequence.[13] For every

[11] We kindly thank Paolo Sodini for sharing the moments on participation as well as the moments on the Swedish wealth distribution that we compare our estimated model to in section 4.2

[12] To preserve computational feasibility, for the estimation procedure we adopt smaller grids than those reported in Appendix B.1, which we use in the results section. Specifically, we use 3 quadrature nodes, 75 points for the cash-on-hand grid, and 15 points for the grid of income's persistent component.

[13] Again for computational feasibility, guided by the arguments that will be outlined below when describing the results, we restrict the global stage of the estimation to search in regions of the parameter space in which the predominant

parameter vector Φ, we solve and simulate the model and then evaluate:

$$d(\Phi)'\Omega d(\Phi), \tag{15}$$

where $d(\Phi)$ is a vector containing the implied deviations of the model moments from their targets in the data and Ω is a diagonal weighting matrix. The deviations from the targets in $d(\Phi)$ are computed as a percentage deviation for the wealth-to-income ratio and relative to the average over the wealth distribution for the remaining targets. The weighting matrix puts 50% of the weight on the wealth-to-income ratio and splits the remaining 50% equally between the schedules of the unconditional risky share, participation, and share of idiosyncratic variance over the wealth distribution. We choose the parameter vector from the Sobol sequence that minimizes equation (15) and proceed to the local step. At this stage, we take the candidate from the first step as an initial starting point and perform a local optimization using the Nelder-Mead algorithm to minimize again equation (15).

The estimated parameters are presented in Table 1. Type-two individuals discount the future less strongly than type-one agents (δ of 0.96 vs. 0.87), are less risk averse (γ of 1.37 vs. 10.31), and seek a lower degree of portfolio diversification. In particular, the estimated values for portfolio diversification imply that the share of idiosyncratic variance in return variance for type-two individuals is 57% ($\zeta = 1.08$), whereas it is 28% for type-one agents ($\zeta = 0.59$).[14] As illustrated by the figures just reported, the results imply a very stark separation across the two types in terms of preferences. In particular, the lower risk aversion and diversification found for type-two agents resemble common anecdotal traits among entrepreneurs.[15]

Type-one individuals are predominant in this economy, as they make up 96% of the total population. This is an interesting result for two reasons. First, it is surprising that, while the preference parameters of type-two agents are more in line with those usually adopted in the macro literature on wealth inequality, the majority of individuals are of type-one, which is characterized by lower time preference rate and high risk aversion, a combination more often used in the household finance literature. Second, despite what just noted, it indicates that only a small fraction of type-two agents is needed to replicate portfolio choice patterns over wealth in the data (and, as we will see below, also wealth inequality).

---

type has jointly lower time preference rate, higher risk aversion, and higher portfolio diversification. The local stage, instead, is not bounded by this constraint.

[14] The share of idiosyncratic variance in total return variance can be computed using the formula for return variance reported in Appendix C.

[15] There is a vast literature on entrepreneurship and wealth inequality (see De Nardi and Fella, 2017, for a review) emphasizing the tension between individual ability and borrowing frictions and its impact on agents' savings behavior, an element which is not present in our model but that might have been captured by our estimation procedure through preference parameters.

Finally, the stock market participation cost f is estimated at 0.001, which is roughly 0.05% of the average yearly income and is smaller than the values typically found in the literature (e.g. Vissing-Jorgensen, 2003).

Preference parameters				
		Type 1	*Type 2*	
δ	time preference rate	0.87	0.96	estimated
γ	risk aversion	10.31	1.37	estimated
ψ	inverse EIS	1.0	1.0	preset
ζ	lack of diversification	0.59	1.08	estimated
	share of individuals	0.96	0.04	estimated
Participation and management costs, borrowing limit				
f	fixed participation cost		0.001	estimated
m	management fee		0.01	Catherine (2021)
$\overline{s}$	borrowing limit		0	preset
Returns				
r^f	risk-free rate		0.02	
$\underline{\mu}_r$	mean syst. return crashes		−0.245	
$\overline{\mu}_r$	mean syst. return normal times		0.115	
σ_{r_1}	cond. st. dev. syst. return, part linked to w		0.077	Catherine (2021)
p_r	probability crashes		0.146	
σ_{r_2}	st. dev. syst. return, part not linked to w		0.114	
Income				
g	drift aggregate wage growth		0.008	
λ_{rw}	sensitivity aggregate wage growth to return		0.161	
σ_ϕ	st.dev. aggregate wage growth shock		0.017	
ρ	autocorrelation persistent component		0.967	
μ_ε	constant mean persistent shock, tail		−0.086	
$\lambda_{\varepsilon w}$	sensitivity mean perm. shock to Δw, tail		4.291	Catherine (2021)
p_ε	probability tail events		0.136	
$\underline{\sigma}_\varepsilon$	st.dev. persistent shock, tail		0.562	
$\overline{\sigma}_\varepsilon$	st.dev. persistent shock, non-tail		0.037	
$\underline{\sigma}_\nu$	st.dev. transitory shock, tail		0.895	
$\overline{\sigma}_\nu$	st.dev. transitory shock, non-tail		0.089	

Table 1: Model parameters values.

4 Model fit

4.1 Targeted moments

We begin by assessing how well the model, which we refer to as our *benchmark model* in the following sections, matches the targeted moments. The wealth-to-income ratio is 4.09, which is close to 4, the value in the data. The unconditional risky share (defined as the individual portfolio's expected excess return over the market's), the share of idiosyncratic risk in the total variance of the individual portfolio, and the participation rate over the wealth distribution are reported in Figure 1 against their empirical counterparts.

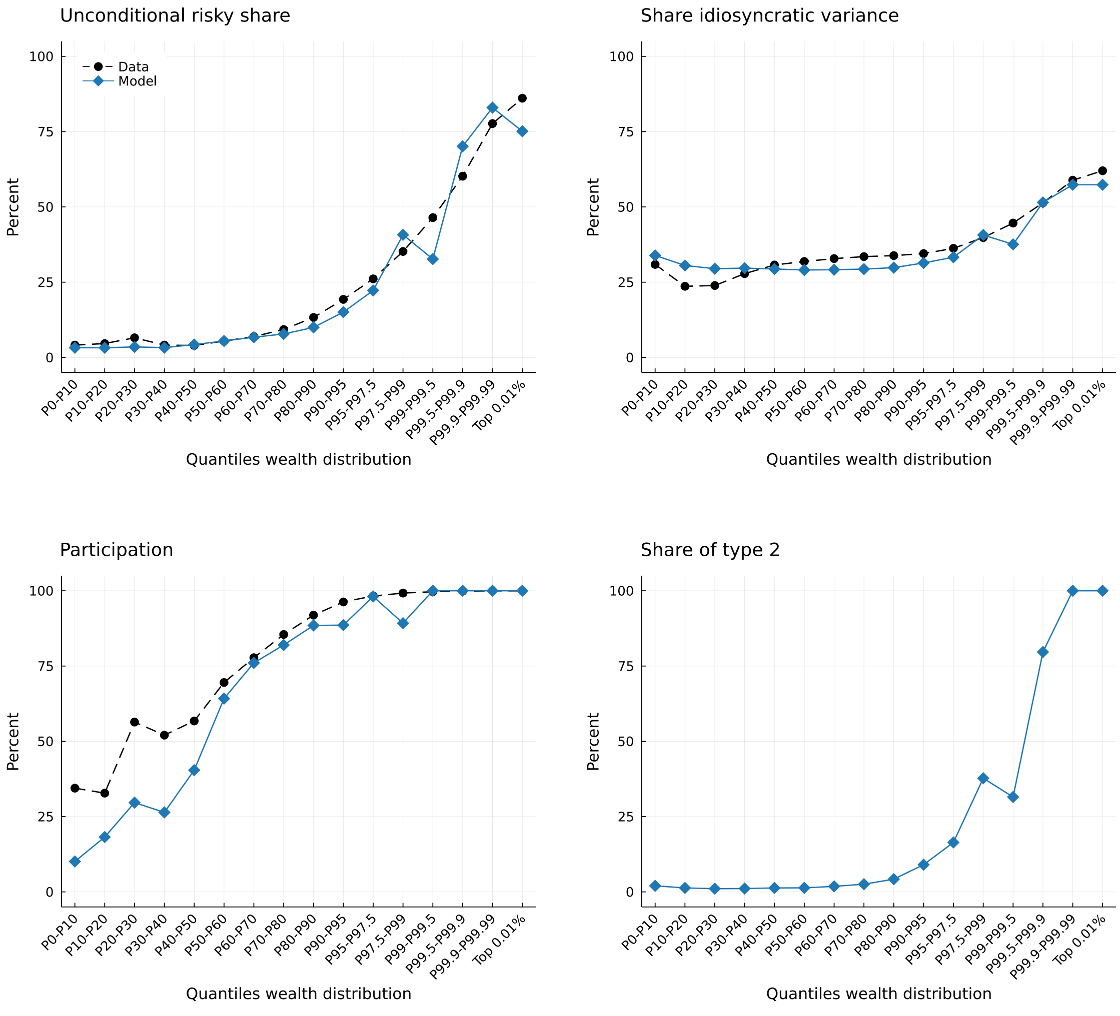


Figure 1: Policies (model vs. data) and share of Type 2 individuals over the wealth distribution.

Overall, the calibration procedure matches well the portfolio choices over the wealth distribution, even though it slightly undershoots the participation rate in the bottom quantiles.

Why does preference heterogeneity enable us to closely replicate the empirical patterns? The outcome of the calibration implies that the economy is predominantly populated by type-one individuals, distinguished by higher impatience, higher risk aversion, and lower share of idiosyncratic risk in their portfolios. Only 4% are type-two agents, who feature opposite characteristics. However, as shown in the bottom right panel of Figure 1, mainly thanks to their higher δ parameter, in equilibrium individuals of the latter type endogenously concentrate at the top of the wealth distribution.[16] As in the model portfolio choice patterns over the wealth distribution are largely determined by the relative share of types, the increasing number of type-two individuals - whose estimated preference parameters imply a high risky share and high idiosyncratic risk at the same time - over wealth allows the framework to reproduce the trends in the data.

It is also worth noticing that the estimated participation fixed cost, 0.05% of average yearly income, is very small compared to other (often unrealistically large) values used in the literature. The fact that more risk-averse individuals mainly populate the bottom of the wealth distribution - where participation is not an obviously optimal choice - is the reason why our model can match participation and portfolio choices with a low fixed cost. Indeed, high-risk aversion combined with the effect of countercyclical income risk, implies that even a small value of this parameter has a sufficient deterring effect on stock market entry.[17]

4.2 Untargeted moments: wealth distribution

Table 2 presents the model's performance in matching the share of total wealth held by different quantiles of the wealth distribution. We compare our results with empirical values computed by Krueger et al. (2016) using PSID (2006) and SCF (2007) data and, since we use portfolio choice moments from Swedish administrative data to estimate the model, with corresponding measures of the Swedish wealth distribution. Furthermore, to check how our framework compares to a state-of-the-art model of wealth inequality without portfolio choice, we add a column with the values generated by the benchmark model in Krueger et al. (2016).

[16] Figure A.1 reports the mass of the two types in levels.

[17] This mechanism is missing in a representative agent model in which, by construction, a high fixed cost is required to discourage the average agent from participating.

Share held by (%):	Benchmark	Krueger et al. (2016)	PSID (2006)	SCF (2007)	Sweden (2000-2007)
Q1	2.2	0.3	-0.9	-0.2	-1.1
Q2	5.6	1.2	0.8	1.2	2.8
Q3	9.2	4.7	4.4	4.6	8.7
Q4	15.5	16.0	13.0	11.9	19.4
Q5	67.5	77.8	82.7	82.5	70.2
90-95 %	11.5	17.9	13.7	11.1	13.4
95-99 %	22.5	26.0	22.8	25.3	17.9
Top 1 %	20.1	14.2	30.9	33.5	21.3
Wealth Gini	0.64	0.77	0.77	0.78	0.69

Table 2: Share of wealth held by people in different quantiles of the wealth distribution: benchmark model vs. data and a state-of-the-art model of wealth inequality.

The model matches well the share of wealth held by the 90-95%, 95-99%, and top 1% groups in the US. In the first two cases, it gets closer than Krueger et al. (2016) who, instead, overshoot the actual values. In the last case, despite being still very far from the corresponding figure in the data, compared to theirs, our model is able to generate a six percentage points higher share of wealth held. Remarkably, our framework delivers an even better match of top wealth inequality in Sweden[18], except for the share held by the 95-99% group, which is slightly higher in the model.

When looking at the distribution as a whole, instead, the performance is less satisfactory. In particular, the first three quintiles hold too much wealth compared to the data, and, as a consequence, the last quintile holds too little. This translates into lower Gini coefficients for wealth inequality than the actual values. Allowing for borrowing - and thus for agents to have negative wealth - might attenuate this issue, which is particularly relevant for the first quintile.[19]

5 Counterfactuals

After having presented the fit of our benchmark specification, in this section we investigate the role of different model components in matching the targeted moments and generating a realistic wealth distribution. To this end, we shut down different features of our benchmark model and quantify the counterfactual predictions.

[18]This is likely related to the fact that we estimate the preference parameters using portfolio choices over wealth from Swedish data.

[19]We plan to extend the model towards this direction in future research.

5.1 Homogeneous preferences

One of the main novelties of this paper is introducing rich heterogeneity in agents' preferences. In the following, we argue that this feature of our framework is crucial for explaining the targeted moments on portfolio choice over the wealth distribution shown in Figure 1. To this end, we re-estimate the model restricting preference parameters to be identical for both types.[20] Table 3 reports the parameter estimates for this case. With only one type, the values are in between the figures obtained in the benchmark case, as this minimizes the differences at the extrema of the schedules. This is also clearly visible from Figure 2, which shows the unconditional risky share (left panel) and the share of idiosyncratic variance (right panel) over the wealth distribution for this specification and the benchmark model. Notably, the estimated fixed cost with is higher than in the benchmark case: as there is just one value for risk aversion, agents at the bottom need to be discouraged more from participating.[21]

The first argument for why the model without preference heterogeneity cannot deliver the empirical patterns concerns the schedule of the unconditional risky share over the wealth distribution. In the data - see Figure 1 - the risky share is increasing throughout the distribution, a feature which is captured by our model with preference heterogeneity. With only one type, instead, the unconditional risky share is flat at zero for the first four deciles and increases up to roughly 25% between the fourth and seventh decile. In contrast to the empirical patterns, risky shares are constant or even decline in wealth for the top three deciles of the distribution.

Apart from the participation fixed cost, that risky shares initially increase in wealth is due to the forces highlighted in Catherine (2021). Despite their high human capital-to-wealth ratio, asset-poor individuals choose not to invest in the stock market due to the riskiness of their labor income. The strength of this effect vanishes as individuals become richer, but, at the same time, their human capital-to-wealth ratio declines. The risky share, therefore, plateaus before converging back to Merton's constant (Merton, 1969).

Despite the mechanism just described, the benchmark model with preference heterogeneity delivers increasing risky shares even at the top of the wealth distribution thanks to compositional effects. Indeed, due to their lower risk aversion, the risky share of type-two individuals converges to a higher constant than for type-one agents. As the share of type two individuals rises at the top, the risky share increases.

[20] To be precise, we re-estimate δ, γ, ζ and the participation cost f while targeting the same moments as for the benchmark model.

[21] Figure A.2 reports the schedule of participation over wealth.

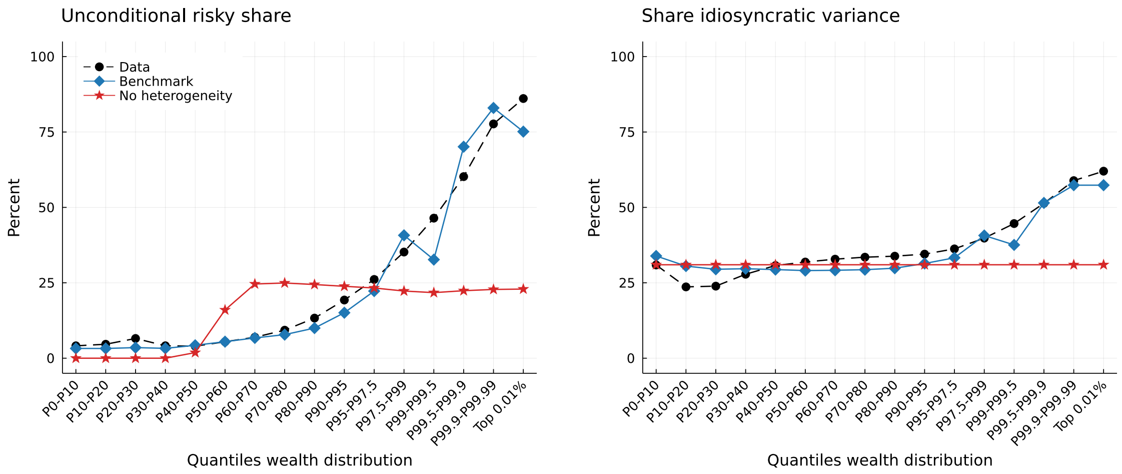


Figure 2: Policies over the wealth distribution in the model without preference heterogeneity, compared to the benchmark and data.

The second argument is related to the schedule of the share of idiosyncratic risk in total return risk over the wealth distribution. Figure 1 shows that (as in Bach et al. (2020)) the share of idiosyncratic risk is increasing in wealth, i.e., wealthier households hold relatively more idiosyncratic risk in their risky portfolios.

In our setting, the variance of the idiosyncratic return component is governed by ζ, i.e. individuals with the same ζ face the same idiosyncratic risk. Heterogeneity in the share of idiosyncratic return variance, therefore, arises only through differences in ζ across agents.

Compositional effects are again crucial for our benchmark model to replicate the empirical patterns for this schedule. Type-one individuals with relatively low ζ (and thus low idiosyncratic risk) mostly populate the bottom of the wealth distribution, whereas type-two agents with relatively high ζ (and thus high idiosyncratic risk) endogenously end up at the top. As a result, the share of idiosyncratic risk is increasing over the wealth distribution. As illustrated in the right panel of Figure 2, without ζ heterogeneity, the same quantity is constant.

The previous two points highlight the role of heterogeneity in risk aversion γ and portfolio diversification ζ. In both cases, we described that compositional effects due to the endogenous sorting of the two types over the wealth distribution were key to generating the increasing schedules of the risky share and of the share of idiosyncratic risk. To reinforce the argument, therefore, it is also important to highlight that attributing a higher degree of patience δ to type-two individuals (less risk averse and less diversified) and a lower one to type-one agents (risk averse and diversified) ensures that the former endogenously end up at the top of the wealth distribution.

Estimated value:		Benchmark	No het.	Only δ	Only γ	No idio. ret.	No skew.
Time preference rate, δ	*Type 1*	0.87	0.91	0.75	0.92	0.88	0.88
	Type 2	0.96		0.96	0.92	0.96	0.96
Risk aversion, γ	*Type 1*	10.31	6.54	5.42	4.34	9.22	18.42
	Type 2	1.37		5.42	6.31	1.20	1.28
Diversification, ζ	*Type 1*	0.59	0.63	0.67	0.76	0	0.64
	Type 2	1.08		0.67	0.76	0	1.09
Fixed cost, f		0.001	0.020	0.007	0.079	0.001	0.005
Share of Type 1		0.96	1	0.74	0.69	0.96	0.97

Table 3: Parameter estimates: benchmark model vs. alternative specifications. *"No het."* indicates the model without preference heterogeneity, *"Only δ"* the model with only heterogeneity in δ, *"Only γ"* the model with only heterogeneity in γ, *"No idio. ret."* the model without idiosyncratic returns and *"No skew."* the model without tail income shocks, stock market crashes and correlation between the income and return processes.

In addition to the effects on portfolio choice, shutting down preference heterogeneity has further implications for wealth inequality. It is known at least since Krusell and Smith (1998) that heterogeneity in patience (δ in our model) across individuals, can generate higher wealth inequality than the restricted case. The reason is that more patient individuals with higher saving rates are concentrated at the top of the wealth distribution in equilibrium, generating a longer right tail. The quantitative impact of preference heterogeneity on the wealth distribution is shown in Table 4. The Gini coefficient declines from 0.64 in the benchmark model to 0.61 in the model without preference heterogeneity, mainly due to a lower share of wealth held by the top quantile.

Share held by (%):	Benchmark	No het.	Only δ	Only γ	No port.	No idio. ret.	No skew.
Q1	2.2	1.9	0.5	1.2	2.4	2.2	2.3
Q2	5.6	5.4	1.5	4.0	6.5	5.7	5.9
Q3	9.2	10.0	2.9	7.7	10.8	9.6	9.7
Q4	15.5	17.9	10.2	16.1	18.0	16.2	16.2
Q5	67.5	64.8	85.0	71.0	62.3	66.3	65.8
90-95 %	11.5	12.8	17.6	14.0	12.3	12.0	11.2
95-99 %	22.5	20.2	28.4	22.9	18.8	21.6	22.2
Top 1 %	20.1	16.0	21.1	17.9	16.1	18.3	19.0
Wealth Gini	0.64	0.61	0.79	0.68	0.58	0.62	0.62

Table 4: Share of wealth held by people in different quantiles of the wealth distribution: benchmark model vs. alternative specifications. *"No het."* indicates the model without preference heterogeneity, *"Only δ"* the model with only heterogeneity in δ, *"Only γ"* the model with only heterogeneity in γ, *"No port."* the model without endogenous portfolio choice, *"No idio. ret."* the model without idiosyncratic returns and *"No skew."* the model without tail income shocks, stock market crashes and correlation between the income and return processes.

Understanding policies. In order to shed some light on the mechanisms driving the results for the model without preference heterogeneity, it is useful to examine the policy functions.

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 income is bond-like (see for example the benchmark specification in Cocco et al., 2005), the optimal risky share is 100% for the wealth-poor, and gradually declines as human capital gets smaller relative to wealth. The specification we use, however, features cyclically skewed income shocks, which helps to generate realistic portfolio choices as first shown in Catherine (2021). Intuitively, if the most adverse income shocks are more likely to happen in times of bad stock market performance, agents with low wealth must be more careful in their portfolio decisions.

To visualize this effect and to show that we replicate the findings of Catherine (2021), Figure 3 presents the policy functions for the risky share (ξ) over cash-on-hand for three different values - low, medium and high - of the persistent component of idiosyncratic income. For the sake of the argument, note that the x-axis is in log scale and that the participation cost is temporarily set to a very low value to make patterns more visible.

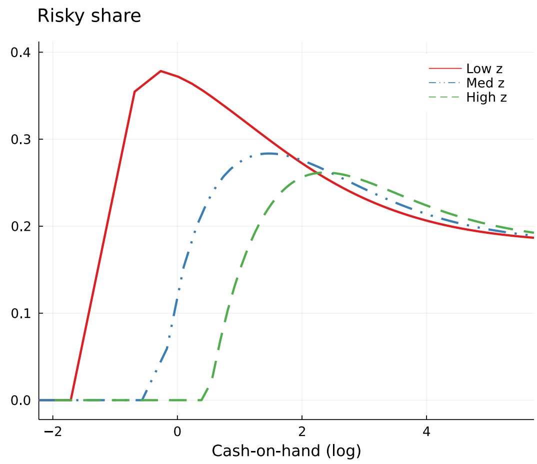


Figure 3: Policy functions for the risky share (ξ) for different states of the persistent component of idiosyncratic income. The three z values approximately correspond to the 25th, 50th and 75th percentiles of the steady state persistent income distribution.

In contrast to the bond-like human capital case, the optimal risky share for all three income states starts at zero, then gradually increases and finally decreases in cash-on-hand. The strongest effect is for agents with a higher z state and, for those among them with low levels of cash-on-hand, the standard finding of a positive relationship between human capital and the risky share is even reversed. This is because negatively skewed income shocks are especially severe for agents with higher persistent income as a fraction of total wealth.

There are two conclusions to draw from this discussion. Firstly, even lacking preference heterogeneity, the cyclical skewness channel helps to match portfolio patterns for the wealth-poor, as the extreme risk-taking of agents with a high human capital-to-wealth ratio implied by more traditional income processes is reduced. However, for most persistent income states, the increase in optimal risky share happens over a relatively narrow range of wealth (note the log scale in Figure 3). Thus - secondly - even if participation costs are absent, this mechanism alone has difficulties matching the gradual increase in the empirical risk-taking patterns over the whole wealth distribution. In particular, since any channel relying on the nature of human capital is by construction weak for agents with a low human capital-to-wealth ratio, without preference heterogeneity the model cannot match any increase at the top of the wealth distribution, which highlights the importance of this latter element for our results.

5.2 Fixed portfolio choice

To what extent does optimal portfolio choice amplify the effect of preference heterogeneity on wealth inequality? In this section, we study the counterfactual predictions of the benchmark model without endogenous portfolio choice. Specifically, we solve the model fixing for all individuals the share invested in the risky asset such that the ratio of risky assets held in the economy to total net wealth equals that of the benchmark specification.[22]

The impact on wealth inequality of optimal portfolio choice is significant, particularly at the top. Table 4 shows that under fixed portfolio choices, the wealth Gini decreases from 0.64 in the benchmark model to 0.58 and that the share of wealth held by the top quintile and by the top 1% decreases, respectively, from 67.5% to 62.5% and from 20.1% to 16.1%.[23]

The explanation for this finding is the following. Type-two individuals have a higher optimal risky share, which results in higher average returns than the rest of the population and amplifies the impact of larger saving rate due to lower impatience, as well as the large idiosyncratic shocks they are exposed to. Thus, when they are not forced to choose the same portfolio composition as the rest of the population, they are more likely to land on the top of the wealth distribution. The quantitative importance of these channels is illustrated in Figure 4, which shows that the concentration of type-two individuals among the wealthy is strongest in the benchmark model with endogenous portfolio choice.

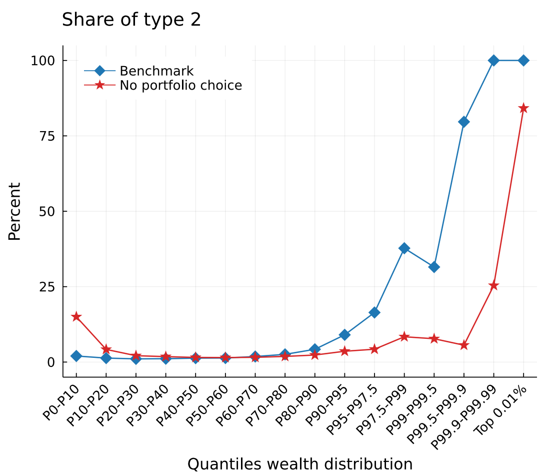


Figure 4: Share of type 2 individuals, benchmark model vs. model without portfolio choice.

As a result (see Figure 1), in line with the empirical patterns, in the model with preference heterogeneity and free portfolio choice, the average unconditional risky share is an increasing function of wealth, even at the very top of the distribution.

[22]Since all the targeted moments (except for the wealth-to-income ratio) are related to portfolio choice, we do not re-estimate the model parameters for this counterfactual, but use the same values of the benchmark case.

[23]Note that in an alternative counterfactual with the share of risky assets fixed at zero, the effect on wealth inequality would be even larger.

5.3 No idiosyncratic returns

To understand the role of idiosyncratic returns in shaping the wealth distribution, we evaluate the performance of another counterfactual model without idiosyncratic returns. Specifically, we set ζ equal to zero, which, in turn, implies zero mean and variance of log idiosyncratic returns $\eta_{i,t}$ for all individuals i and periods t.[24] As depicted in Figure 5, the model-generated policies match the data almost as accurately as the benchmark. The estimates of the common parameters across the two specifications are also similar, as reported in Table 3.

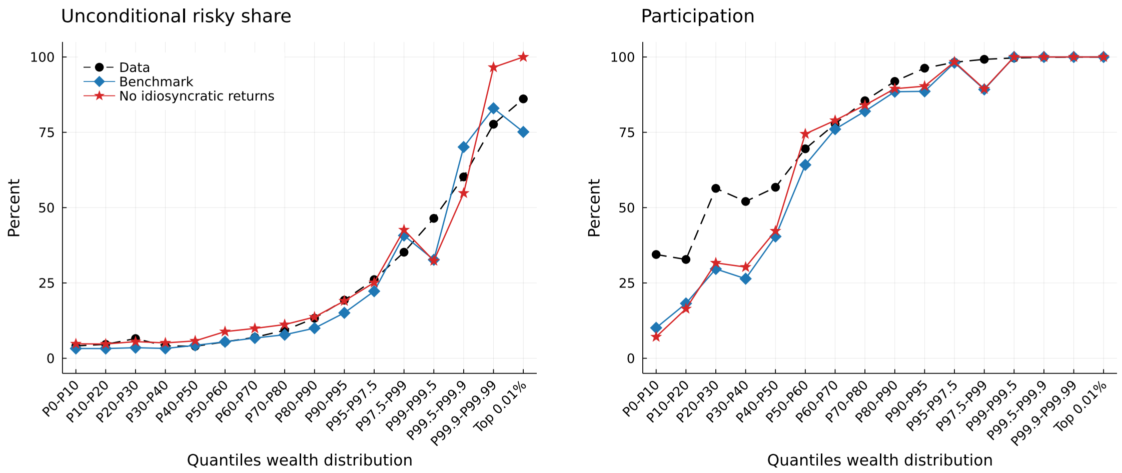


Figure 5: Policies over the wealth distribution in the model without idiosyncratic return risk, compared to the benchmark and data.

Table 4 shows the effect on the wealth distribution. As participation is low for the bottom quantiles in the benchmark and the counterfactual model, idiosyncratic returns barely affect the share of wealth held by the lowest three quintiles. The effect is largest for the top quantiles where agents participate more. For the top 1% of the wealth distribution, the share of wealth held declines from 20.1% in the benchmark to 18.3% in the model without idiosyncratic returns. When looking at total wealth inequality as measured by the Gini coefficient, the decrease in inequality is not too pronounced: 0.64 in the benchmark vs. 0.62 in the counterfactual. The limited impact of the idiosyncratic component of returns is in line with the results in Hubmer et al. (2021), who also find a small effect and mainly clustered at the top.

5.4 Heterogeneity in one preference parameter

To highlight the effects coming from impatience and risk aversion in isolation, we solve for counterfactuals where we restrict the model to heterogeneity in a single preference parameter

[24] As there is no idiosyncratic return risk, when we estimate this restricted model we only target the schedules of risky share and participation over the wealth distribution (and wealth-to-income ratio).

(δ, γ). In particular, we re-estimate the model parameters two times and in each estimation we allow for heterogeneity in just one of them. We target the same moments as in the benchmark model.

The estimated parameters and the effect on the wealth distribution are presented in Tables 3 and 4 in the columns "Only δ" and "Only γ". We will analyze more in detail below the results from the two specifications, but it is already worth noticing that both these counterfactual models generate higher wealth inequality (especially the former) without providing a good fit to empirical portfolio choice patterns, as illustrated in Figure 6.

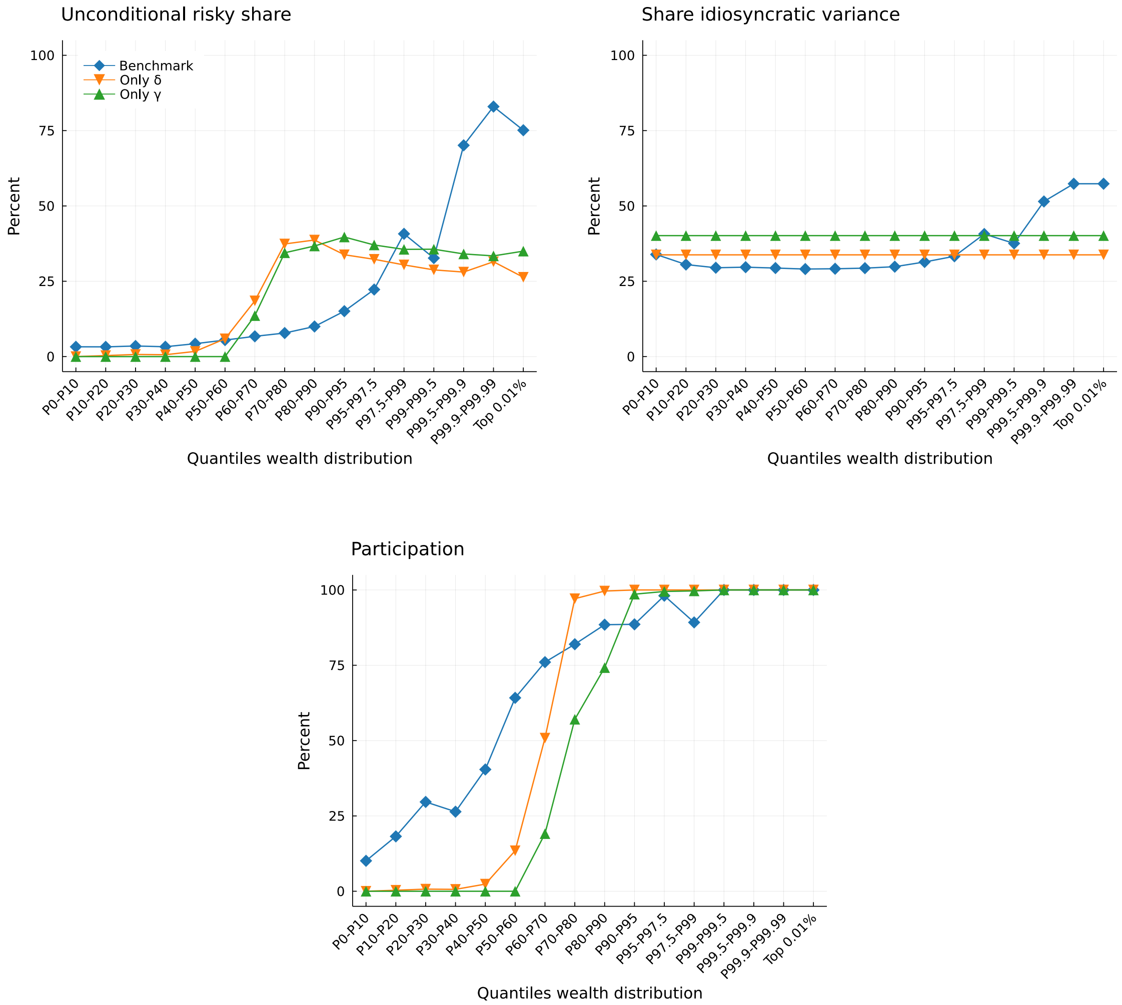


Figure 6: Policies over the wealth distribution. Specifications allowing for heterogeneity in one preference parameter at a time, compared to the benchmark and data.

Only δ. A low discount factor implies both low saving rates 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. However, since this margin does not affect substantially the conditional risky share, the increasing risky share pattern is matched only through the participation channel and hence it is not sufficiently gradual. Furthermore, after full participation the risky share cannot increase anymore in wealth. The attempt to match portfolio choice only through heterogeneity in the discount factor results in a rather extreme value for δ of the majority type, namely 0.75. Due to the resulting powerful separation between the two types over the wealth distribution, wealth inequality becomes large, even surpassing the empirical benchmark from Swedish data.[25]

Only γ. As high risk aversion implies low stock holdings, but high savings, generating a stock holder type on the top of the wealth distribution is less straightforward by heterogeneity in γ. Therefore portfolio choice patterns are matched again mostly through the participation margin, and the calibrated risk aversion parameters are estimated to match the average stock holdings over the region where participation occurs, achieving a slightly less pronounced decrease in risky share for the richest than in the "Only δ" case.

5.5 No skewness in labor income and return

Labor income and returns in the benchmark model are skewed. We have discussed the implications of these properties for the policy functions in section 5.1. In this section, we turn off skewness in labor income and returns and assess how well this restricted model performs. To be precise, we turn off stock market crashes, tail income shocks and the correlation between the income and return process.[26]

As for the previous counterfactuals, the parameter estimates are shown in Table 3. Intuitively, to match a given schedule of participation, risk aversion or the fixed portfolio cost need to increase as skewness risk disappears. This is exactly what the parameter estimates point to. Risk aversion of type-one individuals increases to roughly 18, while the fixed cost of participation increases roughly five times relative to the benchmark case. Figure 7 shows the fit of the targeted portfolio choice patterns. The share of idiosyncratic variance in total return variance is well matched, however the model overshoots the unconditional risky share and participation for most percentiles of the wealth distribution.

[25] That δ heterogeneity can generate high wealth inequality is known at least from Krusell and Smith (1998).

[26] While turning off skewness, we ensure that returns and labor income have the same mean and variance as in the benchmark model. For a detailed outline see Appendix D.

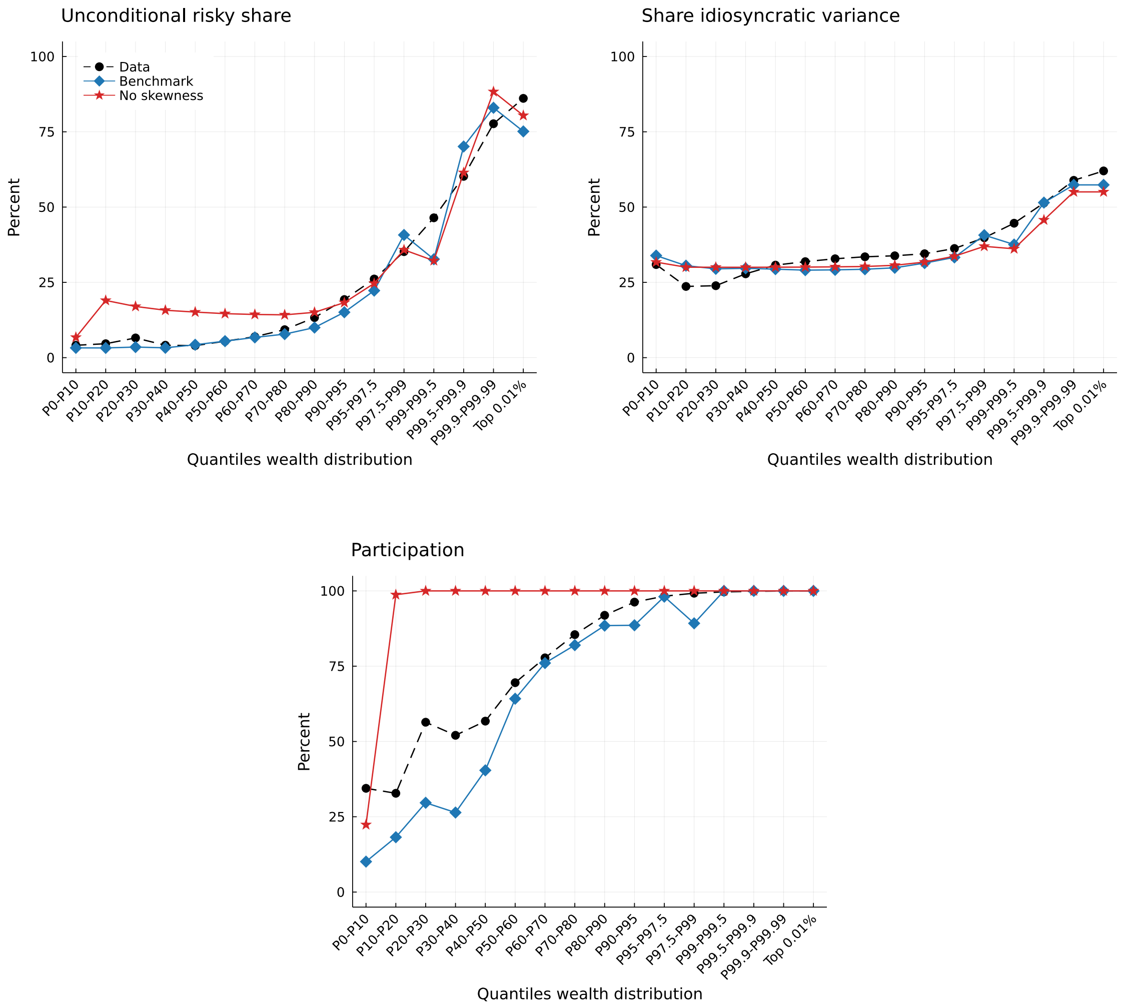


Figure 7: Policies over the wealth distribution when skewness and correlations are turned off, compared to the benchmark and data.

Even though preference heterogeneity still enables to accurately match inequality (see Table 4) and all portfolio patterns for the top 10% of the wealth distribution, without skewness and correlation in income and return risk the model struggles to provide a good fit on the lower portion of the wealth distribution. In particular, in line with the predictions of standard portfolio choice models (as discussed in section 5.1), between the initial jump due to increasing participation and the final hike due to the compositional effect, the average risky share is a counterfactually decreasing function of wealth.

6 Conclusion

This paper introduces a macroeconomic angle to the recent empirical findings on portfolio choice by answering the following questions. First, which additional model ingredients to an otherwise standard incomplete-markets model suffice to generate portfolio choice characteristics consistent with the data? Second, how do those different model ingredients help to generate a realistic wealth distribution?

We include heterogeneity in individual preferences and a rich process that features cyclical skewness in earnings and idiosyncratic returns à la Catherine (2021) to a Bewley model with endogenous portfolio choice. We estimate the parameters governing preference heterogeneity to match the portfolio choice patterns documented in Bach et al. (2020). Heterogeneity in patience, risk aversion and the desire to diversify idiosyncratic return risk are three examples of preference heterogeneity that *jointly* generate realistic portfolio choice patterns over the wealth distribution. Alternative model specifications that abstract from preference heterogeneity, endogenous portfolio choice and non-normalities in the shocks' distributions worsen the fit of the portfolio choice patterns considerably. The combination of preference heterogeneity and endogenous portfolio choice further yields a close match of the wealth distribution, particularly at the top.

This paper attempts to connect the household finance literature on portfolio choice and the macroeconomics literature on wealth inequality. We find that a key element to do that is preference heterogeneity: one type of individuals is characterized by a relatively high risk aversion and a low time preference rate - commonly used in the household finance literature - whereas the other, much less numerous type of individuals features preference parameters commonly used in the macroeconomics literature. Interestingly, this result is an outcome of our parameter estimation rather than an exogenous assumption.

The model presented in this paper captures salient features of portfolio choice in the data endogenously and delivers quantitative predictions on the distribution of wealth. In ongoing work, we quantify dynamic effects of changes in the environment, e.g. aggregate return shocks or modifications in the tax schedule, on the distribution of wealth along the transition, accounting for the optimal response in individuals' portfolio choices.

References

Aguiar, Mark, Mark Bils, and Corina Boar, "Who Are the Hand-to-Mouth?," *Working Paper*, 2021.

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

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.

Calvet, Laurent E., John Y. Campbell, Francisco J. Gomes, and Paolo Sodini, "The Cross-Section of Household Preferences," *Working Paper*, 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.

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

—, Paolo Sodini, and Yapei Zhang, "Countercyclical Income Risk and Portfolio Choices: Evidence from Sweden," *Working Paper*, 2020.

Chang, Yongsung, Jay Hong, Marios Karabarbounis, and Yicheng Wang, "Income Volatility and Portfolio Choices," 2018 Meeting Papers 412, Society for Economic Dynamics 2018.

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.

Fagereng, Andreas, Luigi Guiso, and Luigi Pistaferri, "Portfolio Choices, Firm Shocks, and Uninsurable Wage Risk," *The Review of Economic Studies*, 04 2017, *85* (1), 437–474.

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

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

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

Guvenen, Fatih, 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.

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

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

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

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.

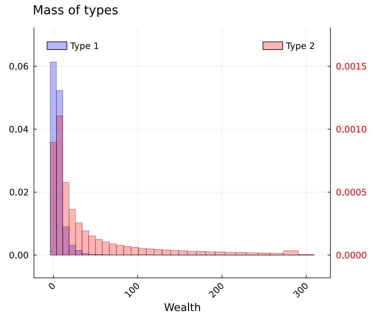
Vestman, Roine, "Limited Stock Market Participation Among Renters and Homeowners," *The Review of Financial Studies*, 08 2018, *32* (4), 1494–1535.

Vissing-Jorgensen, Annette, "Perspectives on Behavioral Finance: Does "Irrationality" Disappear with Wealth?

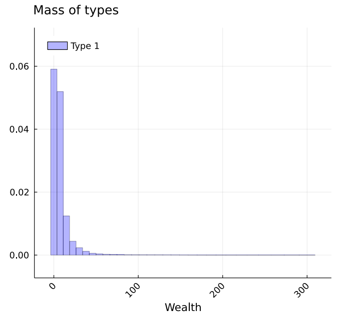
Evidence from Expectations and Actions," *NBER Macroeconomics Annual*, 2003, *18*, 139–194.

A Additional figures

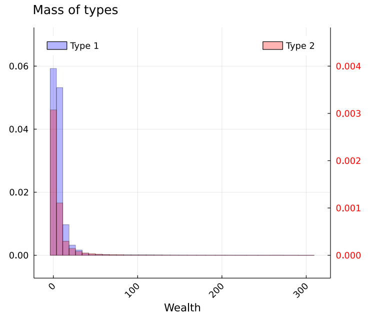
Panel (a): Benchmark



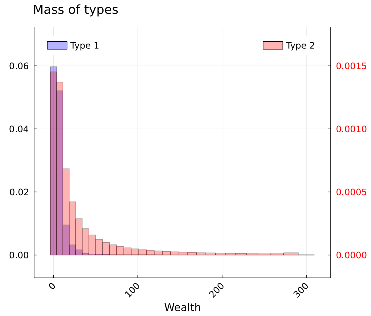
Panel (b): Homogeneous preferences



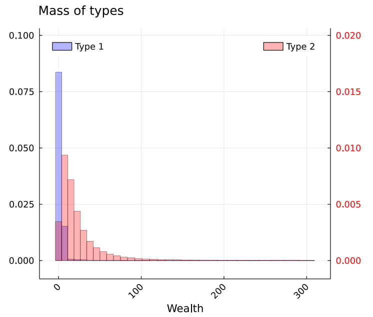
Panel (c): Fixed portfolio choice



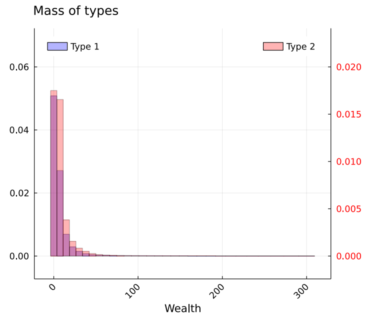
Panel (d): No idiosyncratic returns



Panel (e): Only δ heterogeneity



Panel (f): Only γ heterogeneity



Panel (g): No skewness

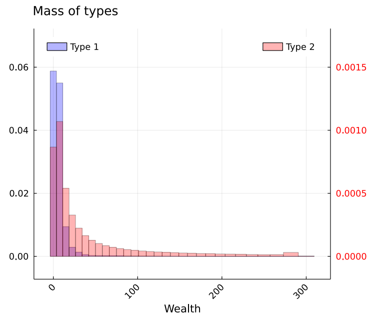


Figure A.1: Mass of types over wealth (y-axis for Type 2 on the right) in different model specifications.

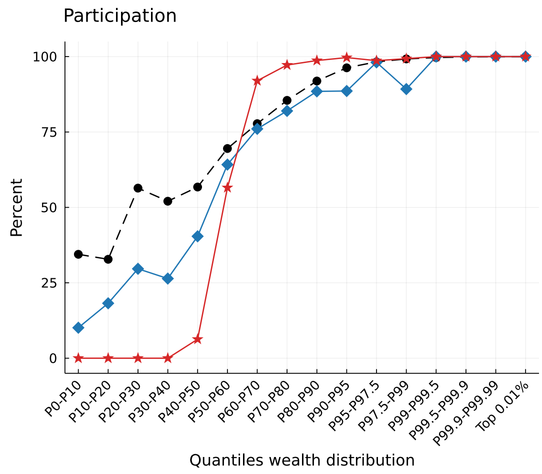


Figure A.2: Participation over the wealth distribution in the model with homogeneous preferences (stars), compared to benchmark (diamonds) and data (circles).

B Numerical solution

B.1 Discretization and grids construction

Normally distributed random variables. Let X be an i.i.d. Normally distributed random variable with mean μ_x and standard deviation σ_x. We discretize X using Gaussian quadrature. Specifically, the support of X is approximated with a finite grid of values $x_1, \ldots, x_{N_q}$ computed as follows:

$$x_j = \mu_x + \sqrt{2}\sigma_x Z_j, \quad j = 1, \ldots, N_q$$

where the Z_j's are Gauss-Hermite nodes and the probability mass of each point of the discretized support is computed as:

$$p(x_j) = \omega_j / \sqrt{\pi}, \quad j = 1, \ldots, N_q$$

where the ω_j's are Gauss-Hermite weights. This procedure is used to discretize $r_{2,t}$, ϕ_t, $\eta_{i,t}$ and the distributions of $\nu_{i,t}$ and $r_{1,t}$ conditional, respectively, on tail/non-tail event and on stock market crash/normal period. N_q is the same for all shocks.

Persistent component of idiosyncratic income. We approximate the process governing the evolution of $z_{i,t}$ as follows: (i) we discretize the conditional distribution of $\varepsilon_{i,t}$, (ii) we compute the evolution of the persistent component according to equation (5) and (iii) we evaluate the model functions at the resulting value of $z_{i,t}$. The advantage of this method is that it requires to discretize just the conditional distribution of $\varepsilon_{i,t}$, which is easier than discretizing the full process of $z_{i,t}$. In particular, the crucial connections between the higher moments of $z_{i,t}$ and other variables are preserved. The disadvantage is that the resulting values of $z_{i,t}$ will very often be off grid, so we need a grid of values that captures well the behavior of the model at such points given our interpolation procedure.[27]

Given the above discussion, the grids for $\varepsilon_{i,t}$ and $z_{i,t}$ are constructed as follows. The conditional distribution of $\varepsilon_{i,t}$ is discretized using the procedure described above for Normally distributed shocks with N_q points. To set up the grid for $z_{i,t}$, instead, we first construct an exponentially spaced grid of $(N_z - 1)/2 + 1$ points with minimum value equal to zero, maximum value equal to $z_{\max}$ and spacing parameter equal to spacing_z. This gives us the positive side of the grid plus the central point (which is therefore equal to zero). Then, we add 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.

Cash-on-hand and savings. For reasons that will be explained when describing the interpolation procedure, 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 - computed from equation (13) - implied by the specific grid value of $z_{i,t}$ under consideration, the borrowing limit and the discretized values of the shocks, maximum value equal to $a_{\max}$ and spacing parameter equal to spacing_a.

The grid for savings $s_{i,t}$ is an exponentially spaced grid of N_s points with minimum value equal to $\bar{s}$, maximum value equal to $s_{\max}$ and spacing parameter equal to spacing_s.

Table 5 summarizes our choices for the numerical parameters.

N_q	N_z	N_a	N_s	$z_{\max}$	$a_{\max}$	$s_{\max}$	spacing_z	spacing_a	spacing_s
5	15	350	N_a	3.5	300.0	300.0	1.6	1.25	1.25

Table 5: Numerical parameters.

B.2 Solving the optimization problem

Whenever it does not lead to confusion, we are dropping time and individual specific indices. To ease up exposition, we also drop the dependence of the value and policy functions (and thus of the preference parameters) on θ. First of all note that as w_t follows a random walk and utility is homogeneous, we can scale with the wage level and reduce the state-space by one dimension. Let us define $\hat{x} = x/\exp(w)$ for a generic variable x representing c, a, s and equivalently for log

[27]See below for more details on the interpolation method.

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

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

Now the optimization problem can be written as

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

subject to

$$\hat{c} + \hat{s} + F_{i,t}\hat{f} = \hat{a}$$

$$\hat{a}' = \Big[R^f + \xi R^{e\prime}\Big]\hat{s}e^{w-w'} + \exp(\hat{y}') + \hat{b}'$$

$$\hat{s} \geq \bar{s}.$$

To simplify ideas and notation, let us introduce

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

The first order condition with respect to the risky share is then:

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

$$0 = \mathbb{E}\Big[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}',z')^{-\gamma}\frac{\partial\widehat{V}(\hat{a}',z)}{\partial\hat{a}'}R^{e\prime}\Big],$$

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

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

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

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

and the envelope condition is:

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

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

After simplifying the two first order conditions read

$$0 = \mathbb{E}\Big[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}',z')^{\psi-\gamma}\hat{c}'(\hat{a}',z')^{-\psi}R^{e\prime}\Big] \tag{B.1}$$

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

Algorithm to solve for value and policy functions To solve equations (B.1) and (B.2) we need to evaluate expectations over policy and value function both on and off the grid points for cash-on-hand and persistent income. Sections B.3 and B.4 provide further details on how we compute expectations and interpolate.

1. Assume we have a guess for $\widehat{V}$, $\hat{c}$ and ξ. For starting one can simply take $\hat{c}'(\hat{a}, z) = \hat{a}$, $\widehat{V}'(\hat{a}, z) = (1-\delta)^{\frac{1}{1-\psi}}\hat{a}$ with an arbitrary ξ' function. Fix a grid $\{\theta_1, \ldots, \theta_k, \ldots, \theta_K\}$ for the preference states, $\{z_1, \ldots, z_j, \ldots, z_M\}$ for the possible values of persistent income and $\{\hat{s}_1 = \bar{s}, \hat{s}_2, \ldots, \hat{s}_i, \ldots, \hat{s}_N\}$ for savings.
2. For all i, j (i.e. for any preference and persistent income state)
   (a) For all i (savings values) compute
      i. the optimal risky share ξ (under participation only, for non-participation set $\xi = 0$ and go to ii.). Recall that ξ is chosen to maximize $\widetilde{V}(\hat{s}, \xi, z)$ and that (ignoring constants)
      $$\frac{\partial \widetilde{V}(\hat{s}, \xi, z)}{\partial \xi} = \mathbb{E}\Big[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}', z')^{\psi-\gamma}\hat{c}'(\hat{a}', z')^{-\psi}R^{e\prime}\Big].$$
      From the second order condition it follows that there is a unique local maximum. Optimal risky share is computed as follows:
         A. if a risky share of 1 was optimal in the previous iteration, check whether this is still true. This is the case if
         $$\frac{\partial \widetilde{V}(\hat{s}, \xi, z)}{\partial \xi} > 0.$$
         If not, save the information that $\xi < 1$.
         B. if a risky share of 0 was optimal in the previous iteration, check whether this is still true. This is the case if
         $$\frac{\partial \widetilde{V}(\hat{s}, \xi, z)}{\partial \xi} < 0.$$
         If not, save the information that $\xi > 0$.
         C. if in the previous iteration neither $\xi = 0$ nor $\xi = 1$ was optimal, use the secant method (combined with the information from i. and ii.) to find ξ such that
         $$\frac{\partial \widetilde{V}(\hat{s}, \xi, z)}{\partial \xi} = 0.$$
         For the secant method two starting points are needed. As the first point use the previous iteration's optimal risky share. The second point is found by moving slightly to the left or right of the first point (depending on the sign of $\frac{\partial \widetilde{V}(\hat{s},\xi,z)}{\partial \xi}$).
      ii. optimal consumption $\hat{c}$ by solving
      $$\hat{c}^{-\psi} = \delta\Big[\widetilde{V}(\hat{s}, \xi, z)\Big]^{\frac{\gamma-\psi}{1-\psi}}\mathbb{E}\Big[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}', z')^{\psi-\gamma}\hat{c}'(\hat{a}', z')^{-\psi}(R^f + \xi R^{e\prime})\Big]$$

With this at hand we can compute the value function as

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

and cash-on-hand as

$$\hat{a} = \hat{c} + \hat{s} + \hat{F}f$$

Note that for computing cash-on-hand the budget constraint for participation and non-participation differ wrt. to the participation costs. Due to the discrete choice of whether to participate or not, computation of consumption, value and cash-on-hand has to be done separately assuming using the optimal risky share computed in (i) and assuming non-participation.

(b) Interpolate the value function and the policy functions for consumption and the risky share (separately for participation and non-participation) over cash-on-hand. By comparing value functions over the cash-on-hand grid we obtain at which part of the grid the agent is participating. We connect value functions and policy functions over the cash-on-hand grid at an ϵ environment around the participation threshold.

3. Convergence is declared when the absolute change in the optimal risky share and the relative change of optimal consumption policies are both smaller than a pre-specified tolerance at every grid point.

B.2.1 Changes when $\psi = 1$

The value function satisfies

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

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

The first order condition with respect to the risky share is:

$$0 = \frac{\partial\widetilde{V}(\hat{s},\xi,z)}{\partial\xi} = \hat{c}^{1-\delta}\frac{\delta}{1-\gamma}\Big[\widetilde{V}(\hat{s},\xi,z)\Big]^{\frac{\gamma-1+\delta}{\delta}}\mathbb{E}\Big[(1-\gamma)e^{(w'-w)(1-\gamma)}\widehat{V}(\hat{a}',z')^{-\gamma}\frac{\partial\widehat{V}(\hat{a}',z')}{\partial\hat{a}'}\frac{d\hat{a}'}{d\xi}\Big]$$

$$0 = \mathbb{E}\Big[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}',z')^{-\gamma}\frac{\partial\widehat{V}(\hat{a}',z')}{\partial\hat{a}'}R^{e\prime}\Big],$$

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

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

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

$$(1-\delta)\hat{c}^{-1} = \delta\Big[\widetilde{V}(\hat{s},\xi,z)\Big]^{\frac{\gamma-1}{\delta}}\mathbb{E}\Big[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}',z')^{-\gamma}\frac{\partial\widehat{V}(\hat{a}',z')}{\partial\hat{a}'}(R^f+\xi R^{e\prime})\Big]$$

and the envelope condition is:

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

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

After simplifying the two first order conditions read

$$0 = \mathbb{E}\left[e^{-\gamma(w'-w)}\widehat{V}(\hat{a}',z')^{1-\gamma}\hat{c}'(\hat{a}',z')^{-1}R^{e\prime}\right] \tag{B.3}$$

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

B.3 Interpolation

The solution procedure outlined in section B.2 will very often require to evaluate the value function and the consumption policy at points off the grid. As explained in section B.1, we do not discretize the persistent component of idiosyncratic 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. In other words, we need a 2-dimensional interpolation procedure over the (a,z) grid. This is achieved by 2-dimensional linear interpolation.

B.4 Computing expectations

In order to solve the model, it is necessary to compute expectations of some 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.[28] To do that, we 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. Note that except for transitions in the preference state, the grid values and probabilities of the other shocks coincide with Gaussian quadrature nodes and weights[29], which enables us to compute expectations very accurately. Finally, remember that the distributions of r_1 is conditional on the realization or not of a stock market crash and, similarly, those of ε and ν on the realization of a tail event or not. This is taken into account simply by scaling the probability of the discretized conditional distributions of these variables by the probability of these events.

[28]Cases in which we do not need to take expectations with respect to one or more of these variables can be handled by the same procedure outlined here with straightforward modifications.

[29]See section B.1.

B.5 Simulation and stationary distributions

In this model the unique aggregate state is the distribution of agents across individual states which are characterized by the triple $(\theta_{i,t}, z_{i,t}, a_{i,t})$. As described before, 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 $N_t \cdot N_z \cdot N_a$ 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. Therefore to examine the dynamic properties of the distribution and to characterize the quasi steady state distribution we need to construct transition matrices corresponding to all values of the aggregate shocks we want to simulate and then aggregate them into a quasi steady state transition matrix.

Conditional transition matrices Taken a realization of the aggregate shocks (r_1, r_2, ϕ) given for each individual state we can compute future cash-on-hand and future persistent income corresponding to each realization of idiosyncratic shocks, where these 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 point, the conditional probability weight corresponding to each simulated (a, z) pair is distributed between the neighboring 4 (or on edges of the grid 2) points proportionally to their relative distance. To avoid extrapolation errors, z is truncated between $-z_{max}$ and z_{max}. From the transition probabilities of moving from one individual state to another we can build up the transition matrices conditional on any realisation of the aggregate shocks.

Unconditional transition matrix To obtain a steady state distribution we need an "average transition matrix". One way of defining one 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 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 (before knowing the shock values).

Steady state distribution The steady state is found by iteration, i.e. multiplying an arbitrary vector with the unconditional transition matrix until convergence. Note that the aggregate state object is a distribution over preference type, persistent income and cash-on-hand. Using the optimal saving policy function, we can compute steady state distribution over preference type, persistent income and end-of-period assets, which is what we refer to as wealth distribution.

C Distribution of total return

Denote the pdf of a normal distribution with mean μ and variance σ^2 with $f_{N(\mu,\sigma^2)}$ and the pdfs of r_1, r_2 and η with f_{r_1}, f_{r_2} and f_η. Since the latter three random variables are independent, we can write the joint pdf of (r_1, r_2, η) as

$$
\begin{aligned}
f_{(r_1,r_2,\eta)}(r_1, r_2, \eta) =& f_{r_1}(r_1) f_{r_2}(r_2) f_\eta(\eta) = \\
=& p_r f_{N(\underline{\mu}_r, \sigma^2_{r1})} f_{N(0,\sigma^2_{r2})} f_{N(-\zeta^2\sigma^2_r/2, \zeta^2\sigma^2_r)} + \\
&+ (1-p_r) f_{N(\overline{\mu}_r, \sigma^2_{r1})} f_{N(0,\sigma^2_{r2})} f_{N(-\zeta^2\sigma^2_r/2, \zeta^2\sigma^2_r)}
\end{aligned}
$$

where $\sigma^2_r = \sigma^2_{r_1} + p_r \underline{\mu}^2_r + (1-p_r)\overline{\mu}^2_r - \mu^2_r + \sigma^2_{r_2}$ and $\mu_r = p_r \underline{\mu}_r + (1-p_r)\overline{\mu}_r$. By properties of normal distirbutions, it follows that $r_1 + r_2 + \eta$ follows a mixed normal distribution with pdf

$$
f_{(r_1+r_2+\eta)}(r) = p_r f_{N(\underline{\mu}_r - \zeta^2\sigma^2_r/2, \sigma^2_{r1}+\sigma^2_{r2}+\zeta^2\sigma^2_r)}(r) + (1-p_r) f_{N(\overline{\mu}_r - \zeta^2\sigma^2_r/2, \sigma^2_{r1}+\sigma^2_{r2}+\zeta^2\sigma^2_r)}(r)
$$

and total risky return $R_r = \exp(r_1 + r_2 + \eta)$ follows a corresponding mixed log-normal distribution. Therefore:

$$
\begin{aligned}
\mathbb{E}[R_r] =& p_r \exp\left(\underline{\mu}_r - \zeta^2\sigma^2_r/2 + (\sigma^2_{r1} + \sigma^2_{r2} + \zeta^2\sigma^2_r)/2\right) + \\
&+ (1-p_r) \exp\left(\overline{\mu}_r - \zeta^2\sigma^2_r/2 + (\sigma^2_{r1} + \sigma^2_{r2} + \zeta^2\sigma^2_r)/2\right) = \\
=& p_r \exp\left(\underline{\mu}_r + (\sigma^2_{r1} + \sigma^2_{r2})/2\right) + (1-p_r) \exp\left(\overline{\mu}_r + (\sigma^2_{r1} + \sigma^2_{r2})/2\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\left[R^2_r\right] =& p_r \exp\left(2\underline{\mu}_r - \zeta^2\sigma^2_r + 2(\sigma^2_{r1} + \sigma^2_{r2} + \zeta^2\sigma^2_r)\right) + \\
&+ (1-p_r) \exp\left(2\overline{\mu}_r - \zeta^2\sigma^2_r + 2(\sigma^2_{r1} + \sigma^2_{r2} + \zeta^2\sigma^2_r)\right) = \\
=& p_r \exp\left(2\underline{\mu}_r + 2\sigma^2_{r1} + 2\sigma^2_{r2} + \zeta^2\sigma^2_r\right) + (1-p_r) \exp\left(2\overline{\mu}_r + 2\sigma^2_{r1} + 2\sigma^2_{r2} + \zeta^2\sigma^2_r\right)
\end{aligned}
$$

so that:

$$
\begin{aligned}
\mathbb{V}[R_r] =& \mathbb{E}\left[R^2_r\right] - (\mathbb{E}[R_r])^2 = p_r \exp\left(2\underline{\mu}_r + 2\sigma^2_{r1} + 2\sigma^2_{r2} + \zeta^2\sigma^2_r\right) + \\
&+ (1-p_r) \exp\left(2\overline{\mu}_r + 2\sigma^2_{r1} + 2\sigma^2_{r2} + \zeta^2\sigma^2_r\right) - \\
&- p^2_r \exp\left(2\underline{\mu}_r + \sigma^2_{r1} + \sigma^2_{r2}\right) - (1-p_r)^2 \exp\left(2\overline{\mu}_r + \sigma^2_{r1} + \sigma^2_{r2}\right) - \\
&- 2p_r(1-p_r) \exp\left(\underline{\mu}_r + \overline{\mu}_r + \sigma^2_{r1} + \sigma^2_{r2}\right)
\end{aligned}
$$

Assuming $p_r = 1$ we would have:

$$\begin{aligned}\mathbb{V}[R_r] &= \exp\left(2\mu_r + \sigma_{r1}^2 + \sigma_{r2}^2\right)\left(\exp\left(\sigma_{r1}^2 + \sigma_{r2}^2 + \zeta^2\sigma_r^2\right) - 1\right) \approx \\ &\approx \exp\left(2\mu_r + \sigma_{r1}^2 + \sigma_{r2}^2\right)\left(\sigma_{r1}^2 + \sigma_{r2}^2 + \zeta^2\sigma_r^2\right) = \\ &= \left(1 + \zeta^2\right)\exp\left(2\mu_r + \sigma_r^2\right)\sigma_r^2\end{aligned}$$

which is intuitive.

D Counterfactuals

In this appendix we describe how we run the counterfactual experiments listed in section 5.

D.1 No idiosyncratic returns

To shut down idiosyncratic return shocks, we simply set $\zeta(\theta_{i,t}) = 0$ for all i and t.

D.2 No skewness in labor income and return

Stock market crashes. To eliminate stock market crashes, the distribution of r_1 should be non-skewed while retaining its mean and variance from the benchmark specification with no preference heterogeneity. This is done by setting:

$$r_{1,t} \overset{i.i.d.}{\sim} \mathcal{N}\left(\mu_r, \tilde{\sigma}_{r_1}^2\right)$$

where $\mu_r = p_r\underline{\mu}_r + (1 - p_r)\overline{\mu}_r$ and $\tilde{\sigma}_{r_1}^2 = \sigma_{r_1}^2 + p_r\underline{\mu}_r^2 + (1 - p_r)\overline{\mu}_r^2 - \mu_r^2$.

Connection between wages and returns. Shutting down the connection between wages and returns implies that equation (4) should be replaced with:

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

Skewness in idiosyncratic shocks. To turn off the skewness of idiosyncratic shocks, the distributions of ε and ν and should be replaced by non-skewed distributions having identical first and second moments as in the benchmark parametrization with no preference heterogeneity. Therefore:

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

where

$$\sigma_\nu^2 = p_\varepsilon\underline{\sigma}_\nu^2 + (1 - p_\varepsilon)\overline{\sigma}_\nu^2$$

and

$$\begin{aligned}
\sigma_\varepsilon^2 &= p_\varepsilon \underline{\sigma}_\varepsilon^2 \\
&\quad + (1 - p_\varepsilon)\overline{\sigma}_\varepsilon^2 + p_\varepsilon(\mu_\varepsilon + \lambda_{\varepsilon w}(g + \lambda_{rw}\mu_r))^2 \\
&\quad + (1 - p_\varepsilon)\Big(- \frac{p_\varepsilon}{1 - p_\varepsilon}(\mu_\varepsilon + \lambda_{\varepsilon w}(g + \lambda_{rw}\mu_r))\Big)^2 \\
&\quad = p_\varepsilon \underline{\sigma}_\varepsilon^2 + (1 - p_\varepsilon)\overline{\sigma}_\varepsilon^2 + \frac{p_\varepsilon}{1 - p_\varepsilon}(\mu_\varepsilon + \lambda_{\varepsilon w}(g + \lambda_{rw}\mu_r))^2
\end{aligned}$$