The Distributional Effects of Asset Returns

Jesús Fernández-Villaverde

University of Pennsylvania, NBER, and CEPR

Oren Levintal*

Reichman University

February 21, 2024

Abstract

We study the distributional effects of asset returns using a heterogeneous-agent model estimated to match the joint distribution of wealth and returns. In the model, endogenous portfolio decisions play a key role through their impact on households' wealth accumulation. We find substantial welfare effects of changes in asset returns. A permanent decline of one percentage point in expected returns increases the consumption share of the top 10% by 6% permanently. Our findings suggest that lower returns increase inequality, which contradicts Piketty's (2014) $r - g$ formula. To resolve this contradiction, we derive a generalized formula that includes the consumption/wealth ratio and which is consistent with our empirical and theoretical findings. Nonetheless, wealth inequality within the Pareto tail is fairly insensitive to asset returns. Instead, inequality between the Pareto tail and the lower range of the distribution responds strongly to asset returns through their differential effects on active savings relative to wealth. Simulations suggest that asset price dynamics can explain the main variations in U.S. top wealth shares since the 1960s.

*Correspondence: jesusfv@econ.upenn.edu (Fernández-Villaverde) and oren.levintal@runi.ac.il (Oren Levintal). This research was supported by a grant from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel.

1 Introduction

How do asset returns affect inequality? This question has attracted growing attention, starting with the seminal paper by Piketty (2014). Recent contributions include Hubmer et al. (2020), Bach et al. (2020), Fagereng et al. (2020), Kuhn et al. (2020), Greenwald et al. (2021), Fagereng et al. (2022), Gomez (2022), and Gomez and Gouin-Bonenfant (2024) among many others. Changes in asset returns affect the market value of households' wealth and hence directly impact the wealth distribution. However, if households' consumption remains the same, these revaluation effects may have no distributional welfare consequences (Cochrane, 2020). Hence, a key question is whether asset returns induce distributional welfare effects through changes in consumption.

In this paper, we show that endogenous portfolio choices play a central role in determining the welfare effects of asset returns. As documented in various studies, e.g., Fagereng et al. (2020) and Bach et al. (2020), rich households hold high-yield assets and earn higher returns on wealth compared to poor households. This observation implies that rich households accumulate wealth mainly through returns on wealth, whereas poor households accumulate wealth through "active savings," defined as savings out of labor income (Bach et al., 2017). When expected returns fall, asset prices go up, and the value of aggregate wealth increases. As a result, it becomes harder for poor households to keep their wealth shares from falling because their active savings as a fraction of aggregate wealth become smaller. Hence, the wealth accumulation of poor households slows down compared to rich households, leading eventually to higher inequality in wealth, capital income, and consumption.

To quantify this effect, we study a heterogeneous-agent model with endogenous portfolio decisions. The model is set in a Lucas (1978) economy with uninsurable idiosyncratic shocks to labor and capital income, as well as aggregate disaster shocks as in Barro (2006). Agents have Epstein and Zin (1989) preferences with time-varying risk aversion (Barro et al., 2022). Low risk-aversion agents choose to save in riskier assets and earn higher returns on wealth, whereas high risk-aversion agents save in safer low-return assets. We estimate the model using a simulated method of moments to match key moments of the joint distribution of wealth and returns. We find that a permanent decline of 1 percentage point in expected returns increases consumption of the top 10% by 6% permanently. That is, the distributional welfare effects are large and permanent.

The fundamental property of the model is the positive correlation between wealth and returns (i.e., the rich earn higher returns). Since portfolios are endogenous, agents that choose riskier portfolios earn higher returns on wealth and accumulate wealth faster. A stationary equilibrium is attained when the active saving rates of the wealthy agents offset (roughly)

their higher returns on wealth so that the wealth distribution remains constant.[1] Hence, the positive correlation between wealth and returns implies a negative correlation between wealth and active saving rates. Put differently, in our model (as in the data), wealthy agents accumulate wealth mainly through higher returns, and poor agents accumulate wealth through active savings. The distributional effects of asset returns depend critically on this property.

To illustrate this point, we compare our model to an alternative version with heterogeneous time discount rates as in Krusell and Smith (1998). In this alternative version, poor agents are relatively less averse to stock market risk because their income is composed mostly of labor income. Hence, poor agents save in riskier assets and earn higher returns but have lower active saving rates compared to wealthy agents, which is in stark contrast with the data. We show that in this model, the distributional welfare effects are exactly the opposite of our baseline case: poor agents are better off when returns are lower (and asset prices higher). We conclude that the wealth-return correlation is key to understanding the distributional welfare effects of asset returns.

We also use our model to examine the Pareto tail of the wealth distribution. Focusing on a steady-state analysis, Piketty and Zucman (2015) show that the Pareto exponent of the right tail is decreasing in $r - g$.[2] Their result implies that top wealth inequality is increasing in $r - g$, as suggested by Piketty (2014). By contrast, our findings suggest that lower returns increase inequality, which contradicts the $r - g$ hypothesis. We revisit Piketty and Zucman (2015) and derive an "extended Piketty formula," whereby top wealth inequality is increasing in $r - g - c$, where c denotes the consumption/wealth ratio at the Pareto tail. The $r - g$ formula is correct as long as the consumption/wealth ratio is fixed, which is the case in Piketty and Zucman (2015). However, when the consumption/wealth ratio is endogenous (as in our model), the general $r - g - c$ formula applies.

More importantly, we find that changes in key parameters affect $r - g$ and c by roughly the same magnitude, so $r - g - c$ hardly changes. Therefore, in our model, wealth inequality **within** the Pareto tail is fairly insensitive to asset returns.[3] Nevertheless, inequality **between** the Pareto tail and the lower range of the distribution responds strongly to asset returns through their differential effects on active savings relative to wealth ("active saving rate").

We draw two main conclusions. First, focusing on inequality **within** the Pareto tail may miss the full impact of asset returns on wealth inequality since most of the effect takes place at the lower range of the distribution. Second, these effects work mainly through changes in

[1] The active saving rate is defined as the ratio between savings out of labor income and wealth.

[2] In this formula, r denotes the aggregate return on wealth and g is the economy growth rate.

[3] Gomez and Gouin-Bonenfant (2024) study a model with limited access to capital markets and show that low returns may increase inequality within the Pareto tail. Our model assumes free access to capital markets.

the active saving rate rather than the $r - g - c$ formula.

To complement the analysis, we examine transitional dynamics across different steady states. We find that variations in wealth inequality along a transition path are driven primarily by unanticipated capital gains, which are earned disproportionately by the rich, who own larger shares of equity. That is, along a transition path, inequality is increasing in realized returns (including unanticipated capital gains). To explore this channel empirically, we feed the model with a sequence of discount-rate shocks that replicate the dynamics of the U.S. aggregate price/earning ratio from 1960 to 2020. The model generates variations in wealth inequality that track the actual dynamics fairly well. We find that the decline in the top 10% wealth share in the 1970s and the rebound since the 1980s can be explained largely by changes in asset prices.

Finally, we contribute to the computational literature of heterogeneous-agent models by employing a novel computational approach that extends the method of Campbell (1998), Reiter (2009), Winberry (2018), and Ahn et al. (2018). These studies combine global approximations along the individual state variables with local approximations around the deterministic steady state. We use a similar idea, but our local approximation is different. Instead of standard perturbation, we use the Taylor projection method developed by Levintal (2018) and implemented in Fernández-Villaverde and Levintal (2018) and Barro et al. (2022). The main advantage of this method is that it is applicable for portfolio choice models because the approximation point incorporates the full volatility of the aggregate shocks. By comparison, perturbation is done around a point of no aggregate shocks (the steady state), where the portfolio decision is indeterminate. Hence, perturbation is usually inapplicable for portfolio-choice models.

Our paper is related to the broad literature on income and wealth inequality (Piketty and Saez, 2003; Piketty, 2014; Saez and Zucman, 2016; Piketty et al., 2018). More specifically, we build on a recent strand of the literature that studies the welfare implications of asset returns. Glover et al. (2020) explore a recessionary shock that triggers a decline in asset prices, similar to the magnitude of the Great Recession. They find that welfare losses are increasing with age. Greenwald et al. (2021) show that a lower discount rate shifts welfare from young to old. The effects of these two studies work through the life-cycle profile of financial transactions. Specifically, the young are the main buyers of assets, and the old are the sellers. Lower returns shift welfare from prospective buyers to sellers, thus making the old better off. Building on this insight, Fagereng et al. (2022) estimate the distributional effects of asset price dynamics in Norway since 1994 using portfolio and asset price data. They find a substantial welfare shift from young to old due to the overall increase in asset prices over this period.

We study a different channel that is independent of the life-cycle saving profile. Our channel works through heterogeneity in active saving rates induced by heterogeneous portfolio decisions. Thus, compared to the previous literature, our focus is on active savings rather than total savings. More specifically, we emphasize that rich and poor households use different sources of wealth accumulation. The rich save mainly through returns on wealth, and the poor save through active savings. Asset returns generate welfare effects because they impact these wealth processes differently.

A key feature of our model is heterogeneity in returns on wealth. Previous studies have acknowledged the importance of heterogeneous returns as a key determinant of wealth inequality (Benhabib et al., 2011, 2015, 2016; Atkeson and Irie, 2020; Hubmer et al., 2020; Xavier, 2021). These studies assumed exogenous heterogeneity in returns, whereas our model generates heterogeneous returns through endogenous portfolio choices. Related models with endogenous portfolios include Krusell and Smith (1997), Cagetti and De Nardi (2006), Favilukis (2013), Lei (2019), Kacperczyk et al. (2019), and Gomez (2022). Relative to these models, our main contribution is quantitative, as we are able to match the joint distribution of wealth and returns. Gomez (2022) also studies the joint distribution of wealth and returns but focuses only on the top tail of the distribution. Instead, we study the entire wealth distribution. Importantly, we find that most of the effects of asset returns work through saving decisions at the lower range of the distribution.

Our portfolio-choice model builds on the long literature of two-agent models with heterogeneous risk aversion, most notably Dumas (1989), Wang (1996), Longstaff and Wang (2012), Gârleanu and Panageas (2015), and Barro et al. (2022) among many others. These models tend to generate too much wealth inequality because the wealth and consumption shares of high risk-aversion agents tend to converge to zero. We avoid this problem by introducing a non-negativity constraint on financial wealth (which we consider empirically plausible given existing liquidity constraints), which is absent from previous models with heterogeneous risk aversion. This constraint serves as a reflecting barrier that prevents the wealth distribution from over-expanding (Gabaix, 2009). The drawback is that it comes at high computational costs. We address the computational problem with our novel solution method.

Our analysis of the dynamics of U.S. wealth inequality is related to a number of previous studies that explored the contribution of asset prices to changes in U.S. wealth inequality. In particular, our results are consistent with model-based estimates provided in Greenwald et al. (2021) and Gomez (2022) as well as model-free estimates presented in Kuhn et al. (2020), Greenwald et al. (2021), and Gomez (2022), as well as the partial-equilibrium analysis in Cioffi (2021). We extend this empirical literature in three ways. First, our model is matched to a different set of moments focused on the joint distribution of wealth and returns. Second,

we construct a new measure of the aggregate price/earning ratio and use it in our simulations. Third, we simulate the dynamics of U.S. wealth inequality over a long period (1960 to 2020).

Gomez and Gouin-Bonenfant (2024) document the rise in U.S. wealth inequality **within** the Pareto tail over the past several decades, and attribute it to the decline in asset returns. A key assumption of their results is limited access to capital markets. Agents with investment opportunities (entrepreneurs) borrow from other agents (rentiers) and gain an internal return that is higher than the external return. A decline in the external return increases the internal return, thus creating more heterogeneity in returns on wealth. Consequently, wealth inequality within the Pareto tail increases. By comparison, in our model, access to capital markets is free, so returns on all assets move together roughly similarly. We find that, in this environment, Pareto inequality is insensitive to asset returns, although total wealth inequality is strongly affected by asset returns.

The rest of the paper is organized as follows: Section 2 derives the model. Section 3 discusses our computational approach. Section 4 presents our estimation. Section 5 conducts a comparative statics analysis. Section 6 reports the main analysis of the distributional effects of asset returns. Section 7 studies the dynamics of U.S. wealth inequality. Section 8 concludes. A computational appendix provides technical details.

2 The model

We postulate an economy consisting of a continuum of agents indexed by i on the unit interval. Agents have Epstein and Zin (1989) preferences over sequences of consumption $C_{i,t}$:

$$U_{i,t}^{1-\theta} = (1-\beta)\, C_{i,t}^{1-\theta} + \beta \left(\mathbb{E}_{i,t} U_{i,t+1}^{1-\gamma_{i,t}} \right)^{\frac{1-\theta}{1-\gamma_{i,t}}}, \tag{1}$$

where β is the discount factor, and the conditional expectation operator $\mathbb{E}_{i,t}$ is taken with respect to future aggregate shocks and idiosyncratic shocks. For most of the analysis, we focus on the case of unit elasticity of intertemporal substitution attained at the limit as $\theta \to 1$. Throughout, we let ρ denote the time discount rate defined as $\beta^{-1} - 1$.

Our main departure from the standard Epstein-Zin parameterization is that the risk-aversion parameter $\gamma_{i,t}$ is an uninsurable idiosyncratic shock that follows a 2-state Markov process with values $\gamma_1 < \gamma_2$. Agents are aware of the possibility of future changes in their risk aversion and consider it in their economic decisions. The time variation in risk aversion is a parsimonious way to model more complex phenomena, such as life-cycle considerations or liquidity constraints.

Output Y_t is exogenous (Lucas, 1978) and follows the process:

$$\Delta \log Y_t = g + \epsilon_t^Y - bd_t,$$

where ϵ_t^Y is a standard business cycle shock, d_t is a low-probability disaster shock (taking the values 0 or 1), and b is the size of the disaster. The disaster shock, modeled as in Barro (2006), allows us to match the equity premium without resorting to unrealistic values of risk aversion. Throughout, we focus on samples without realized disasters corresponding to the U.S. post-WWII.

A share $1-\alpha$ of output is paid to agents as labor income and a share α is paid to stockholders as dividends. Agents can trade equity shares and short-term safe bonds. Therefore, the budget constraint for an agent is:

$$C_{i,t} + P_t E_{i,t} + Q_t B_{i,t} = \left(1+\epsilon_{i,t}^L\right)(1-\alpha)\, Y_t + \left(1+\epsilon_{i,t}^E\right)(\alpha Y_t + P_t)\, E_{i,t-1} + B_{i,t-1}. \quad (2)$$

Here, $E_{i,t}$ and $B_{i,t}$ denote equity and bonds acquired in period t at market prices P_t and Q_t, respectively, while $\epsilon_{i,t}^L$ and $\epsilon_{i,t}^E$ denote uninsurable idiosyncratic shocks to labor income and stock returns.

The labor-income shock $\epsilon_{i,t}^L$ follows a zero-mean 2-state Markov process calibrated to match the labor-income shares of the top 10% and bottom 90%. The stock-return shock $\epsilon_{i,t}^E$ is a zero-mean i.i.d. shock calibrated to match the cross-sectional variation in returns on wealth observed in the data. A law of large numbers ensures that the idiosyncratic shocks are averaged out at the aggregate:

$$\int \epsilon_{i,t}^L di = \int \epsilon_{i,t}^E di = 0.$$

The return on equity and bonds is:

$$R_{i,t}^e = \frac{\left(1+\epsilon_{i,t}^E\right)(\alpha Y_t + P_t)}{P_{t-1}}, \quad R_t^f = \frac{1}{Q_{t-1}}.$$

Thus, the total return on wealth is:

$$R_{i,t}^w = R_t^f + \lambda_{i,t-1}\left(R_{i,t}^e - R_t^f\right),$$

where $\lambda_{i,t}$ denotes the equity portfolio share:

$$\lambda_{i,t} = \frac{P_t E_{i,t}}{P_t E_{i,t} + Q_t B_{i,t}}.$$

Trade in financial assets is subject to three constraints:

$$P_t E_{i,t} + Q_t B_{i,t} \geq 0 \tag{3}$$

$$E_{i,t} \geq 0 \tag{4}$$

$$P_t E_{i,t} \leq \bar{\lambda} \left(P_t E_{i,t} + Q_t B_{i,t} \right). \tag{5}$$

Constraint (3) precludes agents from having a negative net worth, constraint (4) rules out short sale of equity, and constraint (5) imposes an upper bound on leverage. Constraint (3) ("net worth" constraint) is key because it imposes a lower bound on financial wealth, which is necessary to attain a stationary wealth distribution. In the unconstrained version of our model, studied in Barro et al. (2022), the equilibrium growth rate of agents' wealth (including human capital) is an exogenous random variable. Hence, wealth inequality grows without bound (see Gabaix, 2009, for a discussion of random growth models). The "net worth" constraint (3) pins down the wealth distribution by inducing higher wealth growth rates as wealth gets closer to the "net worth" constraint. Constraint (4) prohibits short sales of equity, and constraint (5) imposes an upper bar on leverage, denoted $\bar{\lambda} > 1$, thus limiting the amount of debt an agent can issue. These two constraints are empirically plausible, though they are not essential for the main results.

Agents maximize their preferences (1) subject to the constraints (2)-(5). For convenience, let $\xi_{i,t+1}$ denote the stochastic discount factor:

$$\xi_{i,t+1} = \beta \frac{C_{i,t}^{\theta}}{C_{i,t+1}^{\theta}} \frac{U_{i,t+1}^{\theta-\gamma_{i,t}}}{\left(\mathbb{E}_{i,t} U_{i,t+1}^{1-\gamma_{i,t}} \right)^{\frac{\theta-\gamma_{i,t}}{1-\gamma_{i,t}}}}. \tag{6}$$

Then, the consumption-saving decision is determined by the first-order condition:

$$\left\{ \begin{array}{l|c} \mathbb{E}_{i,t} \xi_{i,t+1} R_{i,t+1}^{w} = 1 & P_t E_{i,t} + Q_t B_{i,t} > 0 \\ P_t E_{i,t} + Q_t B_{i,t} = 0 & \text{else} \end{array} \right., \tag{7}$$

and the portfolio decision is determined by the following condition:

$$\left\{ \begin{array}{l|c} \mathbb{E}_{i,t} \xi_{i,t+1} \left(R_{i,t+1}^{e} - R_{t+1}^{f} \right) \leq 0 & E_{i,t} = 0 \\ \mathbb{E}_{i,t} \xi_{i,t+1} \left(R_{i,t+1}^{e} - R_{t+1}^{f} \right) = 0 & 0 < P_t E_{i,t} < \bar{\lambda} \left(P_t E_{i,t} + Q_t B_{i,t} \right) \\ \mathbb{E}_{i,t} \xi_{i,t+1} \left(R_{i,t+1}^{e} - R_{t+1}^{f} \right) \geq 0 & P_t E_{i,t} = \bar{\lambda} \left(P_t E_{i,t} + Q_t B_{i,t} \right) \end{array} \right. . \tag{8}$$

Finally, we have market-clearing conditions in the goods market, the stock market, and

the bond market:

$$\int C_{i,t} di = Y_t \tag{9}$$

$$\int E_{i,t} di = 1 \tag{10}$$

$$\int B_{i,t} di = 0. \tag{11}$$

The cross-sectional distribution. Let $W_{i,t}$ denote the financial wealth of agent i in period t (including dividend and interest income):

$$W_{i,t} = \left(1 + \epsilon_{i,t}^{E}\right) \left(\alpha Y_t + P_t\right) E_{i,t-1} + B_{i,t-1}.$$

Aggregate wealth is given by $W_t = \alpha Y_t + P_t$.

Then, we can define the normalized wealth of agent i as a fraction of aggregate wealth:

$$\mu_{i,t} = \frac{W_{i,t}}{W_t},$$

and $\boldsymbol{\mathcal{F}}_t\left(\mu, \gamma, \epsilon^{L}\right)$ as the joint cumulative distribution function of normalized wealth, risk aversion, and the labor-income shock:

$$\boldsymbol{\mathcal{F}}_t\left(\mu, \gamma, \epsilon^{L}\right) = \int \mathbb{1}\left(\mu_{i,t} \leq \mu, \gamma_{i,t} \leq \gamma, \epsilon_{i,t}^{L} \leq \epsilon^{L}\right) di, \tag{12}$$

where $\mathbb{1}\left(\cdot\right)$ is the indicator function.

Recursive equilibrium. Let $\mathbf{x}_{i,t} = \left(\mu_{i,t}, \gamma_{i,t}, \epsilon_{i,t}^{L}\right)$ denote the vector of individual state variables of agent i in period t. Analogously, let $\mathbf{X}_t = \left(Y_t, \boldsymbol{\mathcal{F}}_t\right)$ denote the aggregate state variables in period t, which include output and the cross-sectional distribution. When we solve the model, we detrend all the variables by output and, therefore, we can exclude output from the state vector. However, for notational clarity, we define here the solution of the non-detrended model.

A recursive equilibrium is defined by the function $\mathbf{G}$:

$$\begin{pmatrix} P_t \\ Q_t \end{pmatrix} = \mathbf{G}\left(\mathbf{X}_t\right),$$

the function $\mathbf{g}$:

$$\begin{pmatrix} C_{i,t} \\ E_{i,t} \end{pmatrix} = \mathbf{g}\left(\mathbf{x}_{i,t}, \mathbf{X}_t\right),$$

and the function $\mathbf{H}$:

$$\boldsymbol{\mathcal{F}}_{t+1} = \mathbf{H}\left(\mathbf{X}_t, \Delta \log Y_{t+1}\right),$$

such that in each period t the optimality conditions (7)-(8) and the budget constraint (2) hold for each agent i, the market-clearing conditions (9)-(11) hold, and the distribution function satisfies (12), given the exogenous processes of $\gamma_{i,t}$, $\epsilon^L_{i,t}$, $\epsilon^E_{i,t}$ and Y_t.

3 Solution method

The introduction of heterogeneous risk aversion in an incomplete-markets heterogeneous-agent model involves computational challenges that have not been addressed before. Previous models with heterogeneous risk aversion have usually been embedded in complete markets, e.g., Gârleanu and Panageas (2015), or have facilitated simple aggregation, e.g., Barro et al. (2022).

Our computational approach combines global and local approximations. The effects of the individual state variables, potentially highly nonlinear, are approximated by a global method. In contrast, the effects of the aggregate state variables, which are locally close to linear, are approximated by a local method. A similar idea has been implemented previously in Campbell (1998), Reiter (2009), Winberry (2018), and Ahn et al. (2018).

We deviate from the previous literature in our choice of a local method. The studies cited above implement perturbation, a local solution method that is accurate near the steady state. However, in the context of our model, perturbation has two major limitations. First, the portfolio decision is indeterminate at the deterministic steady state because equity and bonds are identical in a world without aggregate uncertainty. Second, perturbation often fails to approximate the equity premium accurately, even when implemented to high orders (Fernández-Villaverde and Levintal, 2018). This is particularly relevant for us because our model incorporates disaster risk.

To resolve this problem, we implement the Taylor projection method proposed by Levintal (2018). This method has proved successful for models with rare disasters and a large equity premium, e.g., Fernández-Villaverde and Levintal (2018) and Barro et al. (2022). Taylor projection is a generalization of perturbation. It exploits information embedded in derivatives

of the equilibrium conditions in a way that is very similar to perturbation. Like perturbation, Taylor projection builds on Taylor theorem to compute a solution that is locally accurate in the neighborhood of the approximation point.

The advantages of Taylor projection over perturbation are twofold. First, the method can be applied at any arbitrary point of the state space. It is not restricted to the steady state and, hence, does not depend on the existence of a steady state. Second, Taylor projection does not have a certainty equivalence property. The approximation accounts for the full volatility of the model (unlike standard perturbation methods unless one includes higher-order terms). Consequently, it captures well asset price effects of risk and uncertainty, which are essential elements of our model.

The detailed implementation of our method is relegated to the computational appendix. Here, we highlight some key ideas. First, following Krusell and Smith (1998), we approximate $\boldsymbol{\mathcal{F}}_t\left(\mu, \gamma, \epsilon^L\right)$ by a finite set of M moments. We find that the first moments conditional on risk aversion and labor income are sufficient to approximate the local dynamics of the model accurately (although more moments could be used in other applications if needed).

Second, we approximate a distribution function that satisfies the M moments and use it to integrate across agents, as done in Winberry (2018). To improve the approximation, we introduce additional K moments so that, in practice, our distribution function satisfies $M + K$ moments. The additional K moments are not important quantitatively for the local dynamics of the model, so we exclude them from the state vector and treat them as locally constant. This approach substantially reduces computational costs.

Third, we approximate an auxiliary function $P_{t+1} = \mathbf{P}\left(\mathbf{X}_t, \Delta \log Y_{t+1}\right)$ that provides the future stock price as a function of current aggregate state and future aggregate shocks (Kubler and Schmedders, 2003; Cao et al., 2023), which is essential for deriving our recursive solution.

Fourth, for our simulations, we discretize the cross-sectional distribution over a highly dense grid. We start at an arbitrary initial distribution, solve the model, and simulate the distribution along a path of aggregate shocks that happen to be zero. We solve the model at several points along the simulation path (recall that Taylor projection can be implemented at any point) until convergence. This point is the stochastic steady state of the model.[4] For simplicity, when we report the results of the model below, we will refer to the stochastic steady state more simply as the "steady state."

[4]The deterministic steady state is a point where all aggregate variables are constant, and there are no aggregate shocks, but agents still face idiosyncratic risk. In contrast, the stochastic steady state is a point at which all aggregate variables are constant, and the realizations of the aggregate shocks are zero. However, agents are aware that non-zero realizations can come in the future. Also, agents still face idiosyncratic risk. The stochastic steady state is a useful concept because it often provides a better summary of the ergodic distribution of nonlinear models than the deterministic steady state.

4 Estimation

This section discusses our choice of calibrated parameters and estimation moments. We calibrate the model parameters that can be inferred directly from the data or that are conventional in the literature. The remaining parameters are estimated by the simulated method of moments (SMM). We evaluate the estimation results by examining the model fit for a set of untargeted moments. Table 1 summarizes the parameter values.

4.1 Calibrated parameters

As mentioned above, we set the inverse IES equal to $\theta = 1$ (below, we will report results for the model when we change many parameter values, including θ). Capital income share is calibrated to the conventional value of $\alpha = 1/3$. The maximum leverage (i.e., the equity portfolio share) is calibrated at $\bar{\lambda} = 2$, which implies a maximum return on wealth of 9.8% (given the targeted asset returns described below). This constraint rarely binds in our simulations, as only 6% of the population and 3% of total wealth are bound by this constraint.

The mean growth rate of U.S. GDP per capita post-WWII is $g = 0.02$ with a standard deviation $\sigma^Y = 0.023$ (Barro and Ursúa, 2008). The disaster parameterization follows Barro and Ursúa (2012). Specifically, a disaster is defined as a rare event with a GDP contraction of 10% or more (from peak to trough). Based on a sample of 40 countries observed from 1870-2009, Barro and Ursúa (2012) estimate the disaster probability at 3.7% per year and the average GDP contraction at 21%.

The labor-income shock $\epsilon^L_{i,t}$ takes two values, ϵ^L_1 and ϵ^L_2, which correspond to the bottom 90% and top 10% of the labor-income distribution, respectively. We calibrate these parameters based on the U.S. distribution over the period 1960-1970, which we consider as a long-run steady state. Over this period, the average labor income of individuals at the top 10% was roughly three times larger than that of the bottom 90% (Piketty et al., 2018), which implies the values $\epsilon^L_1 = -1/6$ and $\epsilon^L_2 = 3/2$ with the corresponding population shares $\theta^L_1 = 0.9$ and $\theta^L_2 = 0.1$, respectively. Since we are dealing with infinitely lived agents and large and persistent labor-income shocks, we use dynastic data to match the persistence of labor income. Based on data from Chetty et al. (2014), we compute a probability of 0.92 that a child whose parents are in the bottom 90% of the income distribution is still in the bottom 90% at the age of 30. This translates into an annual probability of 0.9972 of remaining at the bottom 90%.

4.2 Estimation moments

The remaining parameters are the 2-state Markov process of risk aversion, the discount factor β, and the volatility of the i.i.d. stock-return shock ϵ^E. We estimate these six parameters by SMM using the following six moments: the risk-free rate, the return on aggregate wealth, the wealth shares of the top 10% and top 0.1%, the excess return on wealth of the top 10%, and the cross-sectional variation in return on wealth.

Asset returns are taken from Jordà et al. (2019). Wealth shares are taken from Piketty et al. (2018). We follow the procedure of Hubmer et al. (2020) and focus on the period 1960-1970 when wealth shares were relatively stable. We interpret this period as a steady state. Excess returns by wealth percentiles are taken from Hubmer et al. (2020). Cross-sectional variation in returns on wealth is from Fagereng et al. (2020).

4.3 Estimation results

Table 2 compares simulated moments to data moments. The model fits exactly the estimation moments (marked in bold font). Importantly, the model accurately replicates the fat tail of the wealth distribution (top 10% and 1%), which has been the subject of a large literature. Our contribution is that we account for this fat tail through endogenous portfolio decisions that generate large heterogeneity in returns on wealth.

Figure 1 plots the optimal consumption/wealth ratio. This function is decreasing for all agent types and converges to $1-\beta$ within the top 1% of the wealth distribution. In contrast, Figure 2 shows that the portfolio decisions are starkly different for low risk- and high risk-aversion agents. Agents with low risk aversion hold riskier portfolios composed of equity holdings financed by debt (i.e., $\lambda_{i,t} > 1$). Their leverage is falling with wealth because they issue debt against their labor income, which is large relative to capital income at low levels of wealth. By comparison, agents with high risk aversion do not hold equity at low wealth levels. Hence, our model generates limited stock market participation endogenously. But even above the participation threshold, agents with high risk aversion hold less than 50% of their wealth in equity.

We now assess the model's performance by examining moments not used in our estimation procedure (i.e., the untargeted moments in normal font in Table 2). We start with the wealth shares of the bottom 50% and the middle 50-90%. The model replicates these moments fairly well, with a slightly over-estimated wealth for the bottom 50%. In this range, wealth inequality is largely determined by labor-income inequality. For computational reasons, we allow labor income to take only two values. This comes at the cost of insufficient labor-income inequality and, hence, insufficient wealth inequality at the bottom of the distribution.

Excess return is the return on wealth in excess of the risk-free rate. Hubmer et al. (2020) compute excess returns for different wealth percentiles in the U.S. economy. Their estimate for the excess return of the top 10% is targeted by our SMM and matched exactly by the model. Excess returns within the bottom 90% are not targeted but are still replicated fairly accurately by the model. In particular, the bottom 50% in the model save mainly in risk-free assets, so their excess return is zero, compared to 0.6% in the data. However, the model fails to generate a steep rise in excess returns within the top 10%. For instance, the excess return of the top 0.1% is 9.4% in the data but only 5.1% in the model. This difference might be due to tax and regulatory considerations that our model misses. For example, access to private equity funds is severely limited in the U.S.[5] Since much of the excess returns to the top 0.01% come from private equity, our model misses an important source of heterogeneity. We leave this issue for future research and focus in this paper on the top 10%, which the model captures fairly well.[6]

Interestingly, the amount of safe assets generated by the model (untargeted by the SMM) conforms reasonably well with the amount of safe assets in the U.S. estimated at 32% of total assets by Gorton et al. (2012). This result suggests that trade in safe assets is an important driving force of wealth inequality.

Turning to the income distribution presented in Table 3, the model replicates fairly accurately the income share of the top 10%, an untargeted moment. Income inequality within the bottom 90% is somewhat underestimated. As explained earlier, this can be improved by a finer approximation of the labor-income distribution, though at larger computational costs. Income inequality within the top 10% is also slightly underestimated. This is related to the underestimation of the return on wealth at the top of the wealth distribution, discussed earlier. For the top 10%, which will be the focus of our analysis, the simulated income share is 36.1% (untargeted), very close to the 35.4% in the data.

Krueger et al. (2016) use data from the Consumer Expenditure Survey (CEX) and the Panel Study of Income Dynamics (PSID) to compute the U.S. distribution of consumption in 2006. Table 3 compares their data to our model simulations.[7] The model captures the fat right tail, though the simulated distribution is more unequal than suggested by the data. Specifically, the consumption share of the top 10% implied by the model is 38.1%, while the

[5]The SEC defines "accredited investors," who can invest in some private equity as individuals that earn over $200,000 per year or who have a net worth of over $1 million dollars (excluding the value of their home). "Qualified purchasers," who can invest in a larger range of private equity, are required to hold a wealth of $5 million or more, not including their primary residence or property used in the normal conduct of business.

[6]Another obstacle for matching the excess returns within the top 10% is that the portfolio equity share of low risk-averse agents is falling with wealth as shown in Figure 2.

[7]The model is calibrated to the period 1960-1970, when consumption inequality was probably lower than in 2006 (Attanasio and Pistaferri, 2014). Unfortunately, consumption data from this early period are rather limited (Attanasio and Pistaferri, 2016).

data suggest a range between 26.4% (CEX) to 29.8% (PSID). Apart from the well-known difficulties in measuring consumption inequality (Attanasio and Pistaferri, 2016), part of the gap between the model and the data is due to the omission of progressive taxes. In the absence of a redistributive tax system, matching top wealth shares requires higher top consumption shares to slow down wealth accumulation at the top.

Next, we examine the "active saving rate" defined by Bach et al. (2017) as the ratio of savings out of labor income to wealth. As we show in the next sections, the active saving rate plays an important role in the dynamics of wealth distribution. In what follows, we use the subscript k to denote a certain group of agents, for example, k can refer to the top 10%. The active saving rate of group k is given by:

$$s_k = \frac{\int_{i \in k} \left[\left(1 + \epsilon^L_{i,t}\right) (1 - \alpha) Y_t - C_{i,t} \right] di}{\int_{i \in k} W_{i,t} di}. \tag{13}$$

Table 3 reports estimates of the active saving rate in Sweden obtained by Bach et al. (2017) and compares them to the simulated saving rates (as far as we know, there is no equivalent exercise using U.S. data). A key pattern in the data and the model is that the active saving rate is decreasing with wealth. The simulated ratios are fairly close to the data. For instance, the active saving rate of the top 10% is -4.6% in the model and -3.7% in the Swedish data. Within the top 10%, the active saving rates fall significantly in the data but slightly increase in the model. This is related to the previous discussion on the failure of the model to capture the observed heterogeneity in returns on wealth within the top 10%. For this reason, our focus below is on the top 10%.

5 Comparative statics

Our next step is to understand the forces behind the results in the previous section, in particular, the main forces that drive wealth inequality in the model. To do so, Table 4 explores the effects of the model parameters on key moments of the model. In each case, we change one or two parameters relative to the benchmark parameterization and compute the new stationary equilibrium. In the interest of space, we will focus our discussion on the three most interesting exercises: varying the idiosyncratic shocks, the capital income share, and the intertemporal elasticity of substitution.

5.1 The role of the idiosyncratic shocks

We start by shutting down all idiosyncratic shocks, except labor-income shocks (column (2) vs. (1)). Wealth inequality falls significantly, particularly at the right tail. For instance, the wealth share of the top 0.1% falls from 9.9% to 1.0%. Hence, our model shares the same property found in the literature: labor-income shocks alone cannot explain the fat tail of the wealth distribution (Benhabib and Bisin, 2018) unless a very high degree of skewness is assumed (Castañeda et al., 2003).

Next, column (3) introduces the idiosyncratic stock-return shock (ϵ^E) in addition to the labor-income shock (ϵ^L) but still shuts down the risk-aversion shock. The right tail of the wealth distribution hardly changes (column (3) vs. (2)). Interestingly, the lower end of the wealth distribution changes quite dramatically. The wealth share of the bottom 50% increases from 0.1% to 12.6%. Moreover, poor agents save larger shares of their portfolios in risky equity compared to rich agents. This is because capital income is a relatively small source of income for poor agents. Hence, poor agents are willing to take on more stock market risk. This is at odds with the data (Bach et al., 2020), suggesting that heterogeneity in risk aversion is essential to fit the model to the data.

When we shut down the idiosyncratic stock market shock but keep the labor-income shock and the risk-aversion shock active, the top wealth shares fall quite significantly (compare column (4) to (1)). This effect is explained through precautionary savings. The decline in stock market risk mainly affects wealthy agents who save more in stocks. Their precautionary savings fall and their wealth shares decline consequently. By comparison, when we shut down the idiosyncratic labor-income shock only, precautionary savings of the poor fall, so top wealth shares rise (column (5) vs. (1)). We conclude that idiosyncratic shocks to capital income and labor income affect the wealth distribution mainly through precautionary savings. Labor-income risk affects the precautionary savings of the poor, whereas capital income risk affects the precautionary savings of the rich.

5.2 The role of the capital income share α

Column (6) in Table 4 changes the capital income share from $\alpha = 1/3$ to $\alpha = 1/2$. The effect on wealth inequality is substantial. The wealth share of the top 10% increases from 70.3% to 78.4%. Moreover, the effect on consumption inequality is even larger, e.g., the consumption share of the top 10% increases from 38.0% to 50.7%. The main channel at work is income redistribution. The rise in the capital income share reduces the income of poor agents, whose earnings come primarily from labor income. Thus, they save less and hence accumulate wealth at a slower rate.

5.3 The role of the intertemporal elasticity of substitution

The parameter θ controls the intertemporal elasticity of substitution (IES$= 1/\theta$). When θ falls, top wealth shares and consumption shares fall. See column (7) vs. (1) in Table 4.

To gain some insight into this effect, consider the optimal decision of an unconstrained agent with fixed preferences and no labor, studied in Epstein and Zin (1991). Under i.i.d. returns, the consumption/wealth ratio is:

$$\frac{C_t}{W_t} = 1 - \frac{\left(\beta \left(\mathbb{E}_t \left(R^w_{t+1}\right)^{1-\gamma}\right)^{\frac{1}{1-\gamma}}\right)^{\frac{1}{\theta}}}{\left(\mathbb{E}_t \left(R^w_{t+1}\right)^{1-\gamma}\right)^{\frac{1}{1-\gamma}}}.$$

If returns are sufficiently volatile, the term $\beta \left(\mathbb{E}_t \left(R^w_{t+1}\right)^{1-\gamma}\right)^{\frac{1}{1-\gamma}}$ is smaller than one, so that the consumption/wealth ratio is falling with θ. This behavior is a fair approximation to the behavior of the wealthy agents in the model since they are relatively unconstrained. The decline in θ increases their consumption/wealth ratio, thus reducing their savings and, eventually, their top wealth shares.

6 The distributional effects of asset returns

We turn now to study the distributional effects of asset returns. Let I_k, C_k, and W_k denote group k's labor income, consumption, and wealth, respectively. With this notation, the active saving rate s_k defined in equation (13) becomes:

$$s_k = \frac{I_k - C_k}{W_k},$$

which we can decompose as follows:

$$s_k = \frac{Y}{W} \cdot \frac{I_k/Y - C_k/C}{W_k/W}, \tag{14}$$

where C, Y, and W denote aggregate consumption, output, and wealth, respectively, and $C = Y$ through the market-clearing conditions. Holding constant the distributional term $\frac{I_k/Y - C_k/C}{W_k/W}$, asset prices affect the active saving rate through the aggregate term Y/W (the output/wealth ratio). Importantly, the effect is larger for poor households because the active saving rate is falling in wealth. Therefore, changes in asset prices impose a differential effect on the active saving rate. Specifically, higher asset prices reduce the active saving rate of the

poor relative to the rich. As we show below, this channel drives our main results.

Let g_k^w denote the growth rate of W_k, defined by:

$$g_k^w = s_k + r_k^w + \tau_k, \tag{15}$$

where r_k^w denotes the return on wealth of group k and τ_k denotes net wealth transfers into group k (as a share of W_k) from the rest of the population.[8] For the aggregate economy, we omit the k subscript and drop the net transfer term. Thus, the aggregate wealth growth rate g^w is:

$$g^w = s + r^e, \tag{16}$$

where s is the aggregate active saving rate, and r^e is the return on equity, which is just the return on aggregate wealth.

Let $\mu_k \equiv W_k/W$ denote the wealth share of group k, and let g_k^μ denote the growth rate of μ_k:

$$g_k^\mu = g_k^w - g^w = (s_k - s) + (r_k^w - r^e) + \tau_k. \tag{17}$$

Hence, the change in the wealth share of group k is determined by the three terms on the RHS of equation (17). We call these terms the "saving term," the "portfolio term," and the "between term," respectively. The saving term is the active saving rate of group k in excess of the aggregate saving rate. The portfolio term is the excess return on wealth of group k due to a larger portfolio share of high-yield assets. The last term τ_k is the net transfer of wealth into group k from other groups, which Gomez (2023) calls the "between term." In what follows, we use equation (17) to understand the dynamics of wealth inequality.

6.1 Steady state

We start by analyzing the (stochastic) steady state, where the distribution is constant, i.e., $g_k^\mu = 0$ for all k. Substituting in equation (17) yields:

$$(\bar{s}_k - \bar{s}) + (\bar{r}_k^w - \bar{r}^e) + \bar{\tau}_k = 0, \tag{18}$$

where upper bars denote the steady state. A key property of our model is that the return on wealth is rising with wealth, while the active saving rate is falling with wealth, as shown in Tables 2 and 3.[9] Hence, the saving term $\bar{s}_k - \bar{s}$ is falling with wealth and the portfolio term $\bar{r}_k^w - \bar{r}^e$ is rising with wealth.

[8] The exact equation is $1 + g_k^w = (1 + s_k)(1 + r_k^w)(1 + \tau_k)$, so equation (15) is a log approximation.

[9] Our model successfully replicates these empirical patterns across the bottom 50%, middle 50-90%, and top 10% of the wealth distribution.

Consider a permanent fall in the time discount rate ρ. Table 4 presents the effects of a lower ρ on the steady state. The discount rate is $\rho = 4.2\%$ in column (1) (the benchmark estimate), but it falls to $\rho = 3.2\%$ in column (10). Accordingly, the dividend yield r^d (the inverse of the price/dividend ratio) falls from 3.5% to 2.5%, which implies a rise of 40% in the price/dividend ratio. As a result, we find a substantial rise in wealth and consumption inequality. The wealth share of the top 10% increases from 70.3% to 71.3%, and their consumption share rises from 38.0% to 40.4%. This implies a permanent rise of 6% in consumption of the top 10%. Recall that in a representative-agent version of the model, variations in the discount rate affect asset prices but not consumption. This result does not carry over to heterogeneous agents.

To understand these numbers, we examine the three terms on the RHS of equation (17). Starting at the steady state, the sum of the three terms is zero, as stated in equation (18). Now, consider a fall in ρ, holding constant the joint distribution of wealth and consumption shares. The effect on the saving term $s_k - s$ can be deduced from the active saving rate s_k, defined in (14). A lower ρ implies a lower dividend yield and, hence, a lower output/wealth ratio Y/W. Since s_k is falling in wealth, a lower Y/W reduces the active saving rate of the poor more than the rich (holding the distributional term $\frac{I_k/Y - C_k/C}{W_k/W}$ constant). It follows that the saving term falls for the poor and rises for the rich. This effect widens the wealth distribution, as wealth accumulation of the poor slows down relative to that of the rich.

We turn now to the portfolio term $r_k^w - r^e$, which can be written as:

$$r_k^w - r^e = (\lambda_k - 1)\left(r^e - r^f\right),$$

where λ_k denotes the share of equity in the portfolio of group k. The portfolio term is composed of the equity premium $r^e - r^f$ and the portfolio share of equity λ_k. How does a lower ρ affect these two variables?

The simulation results presented in Table 4 show that the equity premium and portfolio decisions are fairly insensitive to permanent changes in the discount rate; compare columns (1) and (10). These findings are surprisingly similar to frictionless versions of our model. For instance, under complete markets, the equity premium is independent of the discount rate. Furthermore, in the classical unconstrained portfolio choice model with CRRA studied in Merton (1969), optimal portfolios depend on risk aversion and the equity premium, but not the discount rate. These models suggest that portfolio decisions and the equity premium are not sensitive to the discount rate. Interestingly, despite the frictions included in our model (i.e., market incompleteness and constraints on asset trade), the main insights coming from unconstrained asset price and portfolio choice models still hold. We conclude that the portfolio term $r_k^w - r^e$ is approximately insensitive to permanent changes in ρ.

The final term in (17) is the between term τ_k. Gomez (2023) derives this term analytically for a class of models that are similar (but not identical) to our model. He shows that the between term depends on the volatility of the wealth growth rate. In our model, this volatility is dominated primarily by idiosyncratic shocks to equity and labor income, which are both exogenous. Hence, changes in ρ are not likely to affect τ_k much. Indeed, columns (1) and (10) in Table 4 confirm numerically that τ_k is roughly independent of ρ.

Our analysis suggests that a decline in ρ affects the saving term in equation (17) (holding constant wealth and consumption shares) but leaves the other two terms roughly constant. The saving term falls for the poor and rises for the rich, leading to higher wealth growth rates for the rich relative to the poor. Thus, wealth inequality is growing. To restore a steady state, the rich must reduce their active saving rates relative to the poor (otherwise, the wealth distribution keeps spreading out). This is attained through higher consumption shares of the rich, as shown in Table 4, columns (1) vs. (10). The anticipation of higher future income induces them to increase their consumption to the point where their wealth share stabilizes and a new steady state is attained.

We complete the analysis with further experiments. First, we consider a change in the growth rate g. Under unit elasticity of intertemporal substitution ($\theta = 1$), the change in g generates a similar change in asset returns, but the price/dividend ratio and the output/wealth ratio remain constant. Thus, the saving term in equation (18) does not change, and the steady state remains the same. Table 4 confirms this result numerically. Column (8) changes the growth rate g from the benchmark 2% to 3%. Asset returns increase by one percentage point, but the dividend yield and the distribution of wealth, consumption, and labor income do not change (compare columns (8) and (1)).

Next, we turn to the case $\theta = 1/2$, which implies IES $= 2$. In this case, a permanent change Δg in the growth rate changes asset returns by roughly $\theta \Delta g$, as illustrated in Table 4, columns (7) and (9), for $\Delta g = 1\%$. The effect on the dividend yield is $\Delta r^d \approx (\theta - 1)\, \Delta g$.[10] Thus, under $\theta = 1/2$, a rise in g decreases the dividend yield and hence the output/wealth ratio. Consequently, the active saving rate of the poor falls more than that of the rich, generating inequality dynamics similar to the case of a lower discount rate.

Table 4 reports the effect of an increase in the growth rate g from 2% to 3% under $\theta = 1/2$; see columns (7) and (9). We find that the consumption share of the top 10% increases from 35.6% to 36.6%, and their wealth share increases from 64.1% to 65.2%. Portfolios remain roughly fixed, the equity premium hardly changes, and the between terms are roughly constant, which is consistent with our previous analysis.

[10] In a steady state, the return on equity is (up to first-order) $r^e = r^d + g$. Since $\Delta r^e \approx \theta \Delta g$, we get that $\Delta r^d \approx (\theta - 1)\, \Delta g$.

The results above demonstrate that changes in asset returns have large permanent welfare effects. A decline of one percentage point in the dividend yield (due to a lower discount rate) increases the long-run consumption of the top 10% by 6%. Similarly, a decline of 0.5 percentage points in the dividend yield (due to a higher growth rate under IES = 2) increases the consumption of the top 10% by 3%. Joining these two results, the elasticity of consumption of the top 10% with respect to the dividend yield is computed to be around -6.

6.2 $r - g - c$ and the Pareto tail

Piketty (2014) argues that wealth inequality is increasing in $r - g$, where r denotes the return on aggregate wealth (equivalent to r^e in our model). To formalize this idea, Piketty and Zucman (2015) show that the right tail of the stationary wealth distribution follows an approximate Pareto distribution, where the Pareto exponent is a decreasing function of $r - g$ (i.e., top wealth inequality is increasing in $r - g$).[11]

Table 4 reports the Pareto exponent for the top 0.1% and $r - g$. The results contradict Piketty and Zucman (2015). For instance, when $r - g$ falls from 3.6% to 2.6% due to a lower ρ, the Pareto exponent hardly changes; compare columns (1) and (10). Similarly, $r - g$ falls from 3.9% to 3.4% due to a higher growth rate (under IES=2), but the Pareto exponent does not change much; compare columns (7) and (9).

To explain these results, consider the law of motion of wealth of agent i:

$$W_{i,t+1} = (W_{i,t} + I_{i,t} - C_{i,t})\, R^w_{i,t+1},$$

where $I_{i,t}$ denotes labor income. We focus on the steady state, where aggregate wealth grows at the rate of the aggregate economy $e^g - 1$. Hence, normalized wealth $\mu_{i,t} = W_{i,t}/W_t$ follows:

$$\mu_{i,t+1} = \frac{Y}{W} \cdot \frac{I_{i,t}}{Y_t} \cdot \exp\left(r^w_{i,t+1} - g\right) + \exp\left(r^w_{i,t+1} - g - c_{i,t}\right) \cdot \mu_{i,t},$$

where Y/W is fixed at the steady state, $r^w_{i,t+1} \equiv \log R^w_{i,t+1}$, and $c_{i,t} \equiv -\log\left(1 - C_{i,t}/W_{i,t}\right)$. The classical result of Kesten (1973) implies that if $I_{i,t}/Y_t$, $c_{i,t}$, and $r^w_{i,t+1}$ are i.i.d., then $\mu_{i,t}$ converges to a stationary distribution with a right Pareto tail. The Pareto exponent is decreasing in $r - g - c$, which is the mean of $r^w_{i,t+1} - g - c_{i,t}$.[12] For the case of zero labor income ($I_{i,t} = 0$), Gomez (2022) derives the Pareto exponent analytically in continuous time.

Benhabib et al. (2011) extend Kesten (1973) to Markov processes satisfying certain regu-

[11] We define the "Pareto exponent" as the parameter ζ of the Pareto counter CDF $\Pr(\tilde{x} \geq x) = kx^{-\zeta}$.

[12] Formally, the Pareto exponent ζ satisfies $\mathbb{E}e^{\phi\zeta} = e^{-\zeta(r-g-c)}$, where $\phi \equiv \left(r^w_{i,t+1} - g - c_{i,t}\right) - (r - g - c)$ is a zero-mean i.i.d. variable; see Gabaix (2009). It follows that ζ is decreasing in $r - g - c$ (holding the distribution of ϕ constant).

larity conditions, building on de Saporta (2005). Our model does not fall exactly within the general conditions of Benhabib et al. (2011), but we can use their result as an approximation for the right tail of the wealth distribution. Along the right tail, the term $r^w_{i,t+1} - g - c_{i,t}$ behaves as an approximate Markov process.[13] It follows from Benhabib et al. (2011) and de Saporta (2005) that the stationary distribution of the right tail is approximately Pareto with a coefficient that is decreasing in $r - g - c$, where c denotes the consumption/wealth ratio at the Pareto tail. Put differently, top wealth inequality is increasing in $r - g - c$. If the top consumption/wealth ratio is fixed, we get the $r - g$ result of Piketty and Zucman (2015). However, the top consumption/wealth ratio is endogenous in our model, and the general $r - g - c$ formula applies.

Consider a decline in ρ of 1 percentage point reported in column (10) of Table 4. The term $r - g$ falls by one percentage point, but $r - g - c$ does not change. The Pareto exponent remains roughly unchanged, which is consistent with our $r - g - c$ formula. Similarly, a rise of one percentage point in the growth rate g (under IES>1) reduces $r - g$ by half a percentage point; see columns (7) and (9). However, $r - g - c$ does not change and so neither does the Pareto exponent. Hence, wealth inequality **within** the Pareto tail is fairly insensitive to asset returns.

However, these results do not imply that **overall** inequality does not respond to changes in asset returns. On the contrary, we find that wealth and consumption inequality **between** the Pareto tail and the lower range of the distribution change quite substantially. In this range, the $r - g - c$ formula does not apply. As we explained in the previous section, in this range, asset returns drive inequality primarily through differential effects on the active saving rate. Interestingly, these effects work in the opposite direction to Piketty's (2014) $r - g$ formula. Namely, lower (expected) returns increase inequality.

We conclude that changes in asset returns do not automatically translate into changes in wealth inequality **within** the Pareto tail because they may change the consumption/wealth ratio so that $r - g - c$ stays fixed. Moreover, we find that inequality **within** the Pareto tail is often a poor measure of overall wealth inequality. In particular, changes in asset returns may have substantial effects on **overall** wealth inequality without changing inequality **within** the Pareto tail. These effects work through the active saving rate rather than the $r - g - c$ formula.

[13] At the right tail, the consumption/wealth ratio $c_{i,t}$ converges to a constant under $\theta = 1$ or a Markov process depending on risk aversion under $\theta \neq 1$. The (log) return $r^w_{i,t+1}$ is determined by the risk aversion $\gamma_{i,t}$, which is a Markov process.

6.3 Transitional dynamics

Our next step is to study the effect of asset returns on the cross-sectional distribution along a transition path from one steady state to another. We show that wealth and consumption inequality are increasing along a transition path with the price/dividend ratio. The main driver is unanticipated capital gains, which are earned disproportionately by the rich since they own larger shares of equity.

Figure 3 presents the impulse response function for an unanticipated decline of one percentage point in the discount rate. The decline in ρ generates an unexpected rise in the price/dividend ratio, which translates into capital gains for stockholders, who are relatively wealthier. Consequently, the wealth and consumption shares of the top 10% increase. The discount-rate effect on equity prices is due to their longer maturity relative to bonds.

Figure 3 also depicts the dynamics of $r^e - g$, which correspond to $r - g$ in Piketty (2014). A temporarily higher $r^e - g$ increases top wealth shares along the transition path. This finding is consistent with Piketty's (2014) explanation for the rise in wealth inequality in the last decades of the 20th century. However, this result holds only along the transition path, where capital gains are higher than initially anticipated. Once the economy stabilizes at the new steady state, $r^e - g$ falls below its initial level, yet wealth inequality remains higher. Namely, the correlation between $r^e - g$ and top wealth shares is positive along a transition path but negative at the steady state. Similar findings are obtained when there is an unanticipated shock to the growth rate g under IES>1; see Figure 4.

This analysis highlights two different channels through which asset returns affect wealth distribution. At the steady state, the "saving term" in equation (17) is more dominant. By comparison, the "portfolio term" prevails along a transition path. In Section 7, we show that this channel can explain a large part of the dynamics of U.S. wealth inequality.

These results build on two important properties of the model. First, the return on wealth is rising with wealth. Second, the active saving rate is falling with wealth. Both properties are supported by the data and generated endogenously by the agents' portfolio choices, not imposed ex-ante. In the next subsection, we present a model with heterogeneous saving rates that delivers different results because it violates these two properties.

6.4 Heterogeneous saving rates

Much of the literature has studied heterogeneous saving rates as a driver of wealth inequality (Krusell and Smith, 1998). We contrast the heterogeneous-saving model against our heterogeneous-risk-aversion model. The main difference between the two models is the distributional effects of asset returns.

We follow Krusell and Smith (1998), who generate heterogeneous savings rates by introducing idiosyncratic shocks to the time discount factor β but holding γ fixed. Specifically, β is replaced with a time-varying parameter $\beta_{i,t}$ that follows an idiosyncratic 2-state Markov process. Thus, the stochastic discount factor (6) becomes:

$$\xi_{i,t+1} = \beta_{i,t} \frac{1-\beta_{i,t+1}}{1-\beta_{i,t}} \frac{C_{i,t}^{\theta}}{C_{i,t+1}^{\theta}} \frac{U_{i,t+1}^{\theta-\gamma}}{\left(\mathbb{E}_t U_{i,t+1}^{1-\gamma}\right)^{\frac{\theta-\gamma}{1-\gamma}}}.$$

We estimate this "β-shock model" by SMM using the same moments we used in estimating our benchmark "γ-shock model." Since the β-shock model cannot explain the excess returns observed in the data, we drop the excess returns of the top 10% from the targeted moments to allow the SMM to match exactly the remaining moments. The estimation results are reported in Table 5, where targeted moments are marked in bold.

The two models can replicate asset returns and the wealth distribution. However, they differ substantially in the equilibrium portfolios and saving rates. In the γ-shock model, the rich earn higher returns on wealth since they save larger shares of their portfolios in equity, which is consistent with the data. By contrast, the β-shock model delivers exactly the opposite result. The poor save larger shares of their portfolios in equity and earn higher returns than the rich because they are relatively less averse to stock market risk since their income is mainly labor income. Furthermore, in the γ-shock model, the active saving rate falls with wealth, which is supported by the data (Bach et al., 2017). By contrast, in the β-shock model, the active saving rate rises with wealth since the rich tend to have higher values of β.

The steady-state correlation between wealth and returns is related to the correlation between wealth and the active saving rate through equation (18). If returns are rising with wealth, then the active saving rate is likely to fall with wealth, as long as the between term $\bar{\tau}_k$ is not falling too much with wealth. Indeed, in the γ-shock model, returns are rising with wealth and active saving rates are falling with wealth, whereas, in the β-shock model, these two patterns are reversed.

These reversed patterns generate a completely different response to asset pricing shocks. Figure 5 presents an impulse response function of the β-shock model to a permanent decline in ρ. The distributional effects significantly differ from the γ-shock model in Figure 3. On impact, top wealth shares and consumption shares fall in the β-shock model and rise in the γ-shock model. This is a direct result of the different portfolios in the two models. Specifically, in the β-shock model, the poor take on more stock market risk than the rich, earning higher returns when equity prices rise.

In the long run, the β-shock model converges to lower wealth and consumption inequality,

whereas the γ-shock model converges to higher inequality. Namely, the two models deliver opposite welfare effects of asset returns. This is because, in the β-shock model, the active saving rate of the rich is higher than that of the poor (contrary to the data). Thus, higher asset prices reduce the active saving rate of the rich more than that of the poor, which tends to contract the wealth distribution.

To sum up, the correlation between wealth and returns is a key determinant of the distributional effects of asset returns. When the correlation is positive (as in the data), the active saving rate is falling with wealth. In this case, lower returns (higher asset prices) make the rich better off in the short run and the long run. Conversely, when the wealth-return correlation is negative, these results are reversed.

7 The dynamics of U.S. wealth inequality

In our final exercise, we use the model to study the effects of asset prices on U.S. wealth inequality. Asset prices declined significantly in the 1970s and increased during the 1980s and 1990s. This is evident in the price/earning ratio of stocks and price/rent ratio of housing presented in Figure 6. The main driving force was the rise and fall in the real interest rate.

We take a neutral approach regarding the sources of U.S. asset price dynamics. Instead, we impose the dynamics on the model exogenously through discount rate shocks. We feed the model with unanticipated discount rate shocks that replicate the dynamics of the U.S. price/earning ratio. That is, we interpret changes in the price/earning ratio as unanticipated capital gains/losses triggered by changes in ρ.

We estimate the impact of these capital gains on the dynamics of U.S. wealth inequality through model simulations. To this end, we construct a measure of the aggregate price/earning ratio by focusing on three asset classes: stocks, housing, and bonds. Our measure of the aggregate asset price is the sum of the market values of Stocks + Housing + Bonds. The measure of aggregate earning is the sum of firms' profits, housing rents, and real interest payments. It follows that the aggregate price/earning ratio is:

$$\text{Price/Earning} = \frac{\text{Stocks} + \text{Housing} + \text{Bonds}}{\text{Profits} + \text{Rent} + \text{Bonds} \cdot (i - \pi)} = \frac{\frac{\text{Stocks}}{\text{Wealth}} + \frac{\text{Housing}}{\text{Wealth}} + \frac{\text{Bonds}}{\text{Wealth}}}{\frac{\text{Profits}}{\text{Stocks}} \cdot \frac{\text{Stocks}}{\text{Wealth}} + \frac{\text{Rent}}{\text{Housing}} \cdot \frac{\text{Housing}}{\text{Wealth}} + \frac{\text{Bonds}}{\text{Wealth}} \cdot (i - \pi)}$$

The ratios of stocks, housing, and bonds to aggregate wealth are taken from the U.S. Flow of Funds and depicted in Figure 7. We measure the profits/stocks ratio by the S&P earning/price ratio taken from Shiller (2015). The rent/housing ratio is the housing rent/price ratio taken from Jordà et al. (2019). We approximate the interest rate i by a 10-year moving average of the 10-year yield on government bonds.

Figure 8 presents our measure of the aggregate price/earning ratio. It decreased during the 1970s, mainly due to the drop in stock prices, and remained relatively flat until the mid-1980s as the real interest rate increased and house prices declined. From the late 1980s, the price/earning ratio has increased steadily, though with high volatility due to the stock market crash following 2000 and the real estate crisis after 2007. We are able to replicate the dynamics of the price/earning ratio through a sequence of unanticipated discount rate shocks. Figure 8 shows that the simulated series closely tracks the actual dynamics.

Figure 9 presents the wealth share of the top 10% simulated by the model. We compare the simulated wealth share to two estimates of the U.S. wealth distribution. The first is obtained by Piketty et al. (2018) from tax data (henceforth "PSZ"), and the second is from Kuhn et al. (2020) based on the Survey of Consumer Finances (henceforth "SCF"). We start the simulation in 1960, which we consider a steady state in terms of wealth distribution and asset prices.

The model successfully captures the general trends of the top 10% wealth share, namely, the fall in the 1970s and the rise since the 1980s. While the model overestimates the level of the top 10% wealth share since the late 1980s, the overall change from 1980 to 2019 is fairly consistent with the data. For 2019, the model predicts a top 10% wealth share of 73.5%, roughly within the PSZ estimate of 71.5% and SCF estimate of 76.5%.

Our model abstracts from housing, which plays a key role in the dynamics of wealth inequality (Kuhn et al., 2020). Consequently, the model predicts the wrong dynamics following the financial crisis of 2008. During this period, housing prices fell strongly, which inflicted significant losses, particularly among poorer households (Glover et al., 2020; Kuhn et al., 2020) and, thus, increased top wealth shares. The rebound in house prices since 2015 reversed this trend and brought the data closer to the model. Our model does not capture the large debt-financed holdings of housing at the bottom of the wealth distribution, which greatly exposes these households to house price shocks. This requires introducing additional heterogeneity such as age (Glover et al., 2020), further complicating our model. We leave these issues for future research.

8 Conclusions

We study the distributional effects of asset returns. We find that lower expected returns increase long-run inequality. To resolve the contradiction with Piketty's (2014) $r-g$ formula, we derive an extended formula of the form $r-g-c$, which is applicable when the consumption/wealth ratio c is endogenous. Importantly, we find that asset returns affect steady-state inequality mainly through their differential impact on the active saving rate rather than the

$r - g - c$ formula.

Our results suggest that portfolio decisions are a key determinant of the distributional effects of asset returns. Specifically, we find that models that do not correctly capture the positive correlation between wealth and returns deliver the wrong distributional effects, both in the short run and in the long run. Incorporating endogenous portfolio decisions in heterogeneous-agent models is computationally difficult. However, given the importance of portfolio decisions, we believe that tackling these computational challenges is worthwhile. We hope that our methodological contributions presented in this paper will help to promote future research on heterogeneous portfolios and inequality.

References

Ahn, S., Kaplan, G., Moll, B., Winberry, T., and Wolf, C. (2018). When inequality matters for macro and macro matters for inequality. *NBER Macroeconomics Annual*, 32(1):1–75.

Atkeson, A. and Irie, M. (2020). Understanding 100 years of the evolution of top wealth shares in the US: What is the role of family firms? Technical report, National Bureau of Economic Research.

Attanasio, O. and Pistaferri, L. (2014). Consumption inequality over the last half century: Some evidence using the new PSID consumption measure. *American Economic Review*, 104(5):122–26.

Attanasio, O. P. and Pistaferri, L. (2016). Consumption inequality. *Journal of Economic Perspectives*, 30(2):3–28.

Bach, L., Calvet, L. E., and Sodini, P. (2017). From saving comes having? Disentangling the impact of saving on wealth inequality. Technical Report 18-8, Swedish House of Finance Research Paper.

Bach, L., Calvet, L. E., and Sodini, P. (2020). Rich pickings? Risk, return, and skill in household wealth. *American Economic Review*, 110(9):2703–47.

Barro, R. J. (2006). Rare disasters and asset markets in the twentieth century. *Quarterly Journal of Economics*, 121(3):823–866.

Barro, R. J., Fernández-Villaverde, J., Levintal, O., and Mollerus, A. (2022). Safe assets. *Economic Journal*, 132(646):2075–2100.

Barro, R. J. and Ursúa, J. F. (2008). Macroeconomic crises since 1870. *Brookings Papers on Economic Activity*, 1:255–350.

Barro, R. J. and Ursúa, J. F. (2012). Rare macroeconomic disasters. *Annual Review of Economics*, 4(1):83–109.

Benhabib, J. and Bisin, A. (2018). Skewed wealth distributions: Theory and empirics. *Journal of Economic Literature*, 56(4):1–31.

Benhabib, J., Bisin, A., and Zhu, S. (2011). The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica*, 79(1):123–157.

Benhabib, J., Bisin, A., and Zhu, S. (2015). The wealth distribution in Bewley economies with capital income risk. *Journal of Economic Theory*, 159:489–515.

Benhabib, J., Bisin, A., and Zhu, S. (2016). The distribution of wealth in the Blanchard-Yaari model. *Macroeconomic Dynamics*, 20(2):466.

Cagetti, M. and De Nardi, M. (2006). Entrepreneurship, frictions, and wealth. *Journal of Political Economy*, 114(5):835–870.

Campbell, J. R. (1998). Entry, exit, embodied technology, and business cycles. *Review of Economic Dynamics*, 1(2):371–408.

Cao, D., Luo, W., and Nie, G. (2023). Global DSGE models. *Review of Economic Dynamics*, 51:199–225.

Castañeda, A., Díaz-Gímenez, J., and Ríos-Rull, J.-V. (2003). Accounting for the US earnings and wealth inequality. *Journal of Political Economy*, 111(4):818–857.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Quarterly Journal of Economics*, 129(4):1553–1623.

Cioffi, R. A. (2021). Heterogeneous risk exposure and the dynamics of wealth inequality. *URL: https://rcioffi. com/files/jmp/cioffi_jmp2021_princeton. pdf (cit. on p. 7)*.

Cochrane, J. H. (2020). Wealth and taxes. *Cato Institute, Tax and Budget Bulletin*, 86.

de Saporta, B. (2005). Tail of the stationary solution of the stochastic equation $Y_{n+1} = a_n Y_n + b_n$ with Markovian coefficients. *Stochastic Processes and Their Applications*, 115(12):1954–1978.

Dumas, B. (1989). Two-person dynamic equilibrium in the capital market. *Review of Financial Studies*, 2(2):157–188.

Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57:937–969.

Epstein, L. G. and Zin, S. E. (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy*, 99(2):263–286.

Fagereng, A., Gomez, M., Gouin-Bonenfant, E., Holm, M., Moll, B., and Natvik, G. (2022). Asset-price redistribution. Technical report, Working Paper.

Fagereng, A., Guiso, L., Malacrino, D., and Pistaferri, L. (2020). Heterogeneity and persistence in returns to wealth. *Econometrica*, 88(1):115–170.

Favilukis, J. (2013). Inequality, stock market participation, and the equity premium. *Journal of Financial Economics*, 107(3):740–759.

Fernández-Villaverde, J. and Levintal, O. (2018). Solution methods for models with rare disasters. *Quantitative Economics*, 9(2):903–944.

Gabaix, X. (2009). Power laws in economics and finance. *Annual Reviews of Economics*, 1(1):255–294.

Gârleanu, N. and Panageas, S. (2015). Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing. *Journal of Political Economy*, 123(3):670–685.

Glover, A., Heathcote, J., Krueger, D., and Ríos-Rull, J.-V. (2020). Intergenerational redistribution in the Great Recession. *Journal of Political Economy*, 128(10):3730–3778.

Gomez, M. (2022). Asset prices and wealth inequality. *Unpublished manuscript*.

Gomez, M. (2023). Decomposing the growth of top wealth shares. *Econometrica*, 91(3):979–1024.

Gomez, M. and Gouin-Bonenfant, É. (2024). Wealth inequality in a low rate environment. *Econometrica*, 92(1):201–246.

Gorton, G., Lewellen, S., and Metrick, A. (2012). The safe-asset share. *American Economic Review*, 102(3):101–06.

Greenwald, D. L., Leombroni, M., Lustig, H., and Van Nieuwerburgh, S. (2021). Financial and total wealth inequality with declining interest rates. Technical report, National Bureau of Economic Research.

Hubmer, J., Krusell, P., and Smith, A. A. (2020). Sources of US wealth inequality: Past, present, and future. *NBER Macroeconomics Annual*, 35:391–455.

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. M. (2019). The rate of return on everything, 1870–2015. *Quarterly Journal of Economics*, 134(3):1225–1298.

Kacperczyk, M., Nosal, J., and Stevens, L. (2019). Investor sophistication and capital income inequality. *Journal of Monetary Economics*, 107:18–31.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248.

Krueger, D., Mitman, K., and Perri, F. (2016). Macroeconomics and household heterogeneity. In Taylor, J. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 843–921. Elsevier.

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

Krusell, P. and Smith, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy*, 106(5):867–896.

Kubler, F. and Schmedders, K. (2003). Stationary equilibria in asset-pricing models with incomplete markets and collateral. *Econometrica*, 71(6):1767–1793.

Kuhn, M., Schularick, M., and Steins, U. I. (2020). Income and wealth inequality in America, 1949–2016. *Journal of Political Economy*, 128(9):3469–3519.

Lei, X. (2019). Information and inequality. *Journal of Economic Theory*, 184:104937.

Levintal, O. (2018). Taylor projection: A new solution method for dynamic general equilibrium models. *International Economic Review*, 59(3):1345–1373.

Longstaff, F. A. and Wang, J. (2012). Asset pricing and the credit market. *Review of Financial Studies*, 25(11):3169–3215.

Lucas, Jr, R. E. (1978). Asset prices in an exchange economy. *Econometrica*, 46(6):1429–1445.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, pages 247–257.

Piketty, T. (2014). *Capital in the Twenty-First Century*. Harvard University Press.

Piketty, T. and Saez, E. (2003). Income inequality in the United States, 1913–1998. *Quarterly Journal of Economics*, 118(1):1–41.

Piketty, T., Saez, E., and Zucman, G. (2018). Distributional national accounts: methods and estimates for the United States. *Quarterly Journal of Economics*, 133(2):553–609.

Piketty, T. and Zucman, G. (2015). Wealth and inheritance in the long run. In Atkinson, A. B. and Bourguignon, F., editors, *Handbook of Income Distribution*, volume 2, pages 1303–1368. Elsevier.

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control*, 33(3):649–665.

Saez, E. and Zucman, G. (2016). Wealth inequality in the United States since 1913: Evidence from capitalized income tax data. *Quarterly Journal of Economics*, 131(2):519–578.

Shiller, R. J. (2015). *Irrational Exuberance*. Princeton University Press.

Wang, J. (1996). The term structure of interest rates in a pure exchange economy with heterogeneous investors. *Journal of Financial Economics*, 41(1):75–110.

Winberry, T. (2018). A method for solving and estimating heterogeneous agent macro models. *Quantitative Economics*, 9(3):1123–1151.

Xavier, I. (2021). Wealth inequality in the US: The role of heterogeneous returns. *Available at SSRN 3915439*.

Table 1: Parameter values: Benchmark

Parameter	Value	Source
A. Calibrated parameters		
Inverse IES θ	1	Conventional
Capital income share α	1/3	Conventional
Maximum leverage $\bar{\lambda}$	2	Rarely binding
Output growth		
g (mean in normal periods)	0.02	Barro and Ursúa (2008)
σ^Y (Gaussian volatility)	0.023	Barro and Ursúa (2008)
p^{disaster} (disaster probability)	0.037	Barro and Ursúa (2012)
b (disaster size)	$-\log(1-0.21)$	Barro and Ursúa (2012)
Labor-income shock (ϵ_{it}^L)		
values		Match the bottom 90% and top 10% of the U.S. labor-income distribution in 1960-1970 from Piketty et al. (2018)
ϵ_1^L (low income)	-1/6	
ϵ_2^L (high income)	3/2	
population shares		
θ_1^L (high income)	0.9	
θ_2^L (low income)	0.1	
probability of remaining at the low-income state		
π_{11}^L	0.9972	Chetty et al. (2014)
B. Estimated parameters		
Markov process of risk aversion (γ_{it})		SMM estimates to match six moments: aggregate return on wealth, risk-free rate, wealth shares of top 10% and top 0.1%, excess return of top 10%, and standard deviation of returns on wealth across households.
values		
γ_1	3.7	
γ_2	10.0	
transition probabilities		
π_{11}^γ	0.993	
π_{22}^γ	0.999	
Time discount factor (β)	0.959	
Exogenous return volatility (σ^E)	0.053	

Table 2: Wealth and returns: Model vs. data

	Model	Data
Returns (%)		Jordà et al. (2019) (1950-2017)
return on equity	**5.6**	5.6
return on bonds	**1.4**	1.4
Wealth shares (%)		Piketty et al. (2018) (1960-1970)
bottom 50%	5.7	1.1
50-90%	24.0	28.6
top 10%	**70.3**	70.3
top 1%	29.3	28.1
top 0.1%	**9.9**	9.9
Excess returns (%)		Hubmer et al. (2020) (1967)
bottom 50%	0.0	0.6
50-90%	2.6	2.2
top 10%	**5.1**	5.1
top 1%	4.9	7.0
top 0.1%	4.8	8.8
		Fagereng et al. (2020) (Norway 2005-2015)
Return volatility (%)	**7.8**	7.8
		Gorton et al. (2012) (1952-2011)
Safe to total assets (%)	28.1	32.0

Targeted moments in bold. Data samples are in parentheses. All data are for the U.S. except return volatility, which is based on Norwegian data. The returns on (unleveraged) equity and bonds are the total return on wealth and the bill rate, respectively, from Jordà et al. (2019). Excess return is the portfolio return in excess of the risk-free rate, estimated by Hubmer et al. (2020). Return volatility refers to volatility across households and over time.

Table 3: Income, consumption, and savings: Model vs. data

	model	data
Income shares[1] **(%)**		Piketty et al. (2018) (1960-1970)
bottom 50%	28.5	18.5
50-90%	35.4	46.2
top 10%	36.1	35.4
top 1%	11.7	12.2
top 0.1%	3.7	4.2
Consumption shares[2] **(%)**		Krueger et al. (2016) (PSID/CEX 2006)
bottom 60%	33.0	31.9-34.3
60-90%	29.0	38.2-39.3
top 10%	38.1	29.8 - 26.4
top 1%	13.0	8.2 - 5.1
Active saving rate[3] **(%)**	%	Bach et al. (2017) (Sweden)
bottom 50%	0.6	14.3
50-90%	-1.7	0.0
top 10%	-4.2	-3.7
top 1%	-4.2	-5.2
top 0.1%	-4.1	-6.5

[1] Sorted by income.
[2] Sorted by consumption.
[3] Sorted by wealth. Active saving rate is saving out of labor income as a ratio of wealth.

Table 4: Comparative statics

	benchmark	$\gamma_1 = \gamma_2$, $\epsilon^E = 0$	$\gamma_1 = \gamma_2$	$\epsilon^E = 0$	$\epsilon^L = 0$	$\alpha = 0.5$	$\theta = 1/2$	$g + 1\%$	$\theta = 1/2$, $g + 1\%$	$\rho - 1\%$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Wealth share (%)										
bottom 50%	5.7	0.1	12.6	11.4	11.1	4.2	7.4	5.7	7.5	7.0
50-90%	24.0	49.3	41.9	26.6	16.3	17.4	28.6	24.0	27.3	21.6
top 10%	70.3	50.6	45.5	62.0	72.6	78.4	64.1	70.3	65.2	71.3
top 1%	29.3	7.9	7.9	21.3	32.9	43.5	20.8	29.3	21.3	30.0
top 0.1%	9.9	1.0	1.1	5.2	13.0	21.5	5.3	9.9	5.4	10.3
Consumption share (%)										
bottom 50%	27.5	27.9	33.9	27.4	33.2	20.7	27.6	27.5	27.3	26.4
50-90%	34.6	44.9	41.4	35.4	30.9	28.3	36.9	34.6	36.3	33.3
top 10%	38.0	27.1	24.9	37.4	35.9	50.7	35.6	38.0	36.6	40.4
top 1%	13.0	3.9	3.8	10.9	13.2	25.1	9.8	13.0	10.1	14.1
top 0.1%	4.1	0.5	0.5	2.5	4.9	12.1	2.3	4.1	2.4	4.5
Capital income share (%)										
bottom 50%	1.5	0.1	16.8	3.7	3.6	1.3	2.2	2.1	2.7	0.7
50-90%	17.0	49.3	41.9	17.8	14.6	12.8	21.7	18.0	21.4	13.3
top 10%	81.6	50.6	41.3	78.5	81.8	85.9	76.1	79.9	75.9	86.0
top 1%	33.2	7.9	7.1	26.8	35.3	46.7	23.8	32.6	23.9	34.5
top 0.1%	11.0	1.0	1.0	6.5	13.7	22.9	5.9	10.8	6.0	11.6
Portfolio share of equity (%)										
bottom 50%	1	100	140	0	1	1	0	1	0	1
50-90%	61	100	100	51	85	61	65	61	66	57
top 10%	121	100	89	139	119	114	127	121	126	123
top 1%	118	100	87	138	111	111	120	118	119	117
top 0.1%	115	100	86	137	108	110	117	115	116	114
Between term (%)										
bottom 50%	0.0	9.7	-0.2	0.0	0.0	-0.1	0.0	0.0	0.0	0.0
50-90%	-0.2	0.0	0.0	0.0	-0.2	-0.2	-0.2	-0.2	-0.2	-0.2
top 10%	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.1	0.1	0.1
top 1%	0.2	0.0	0.4	0.0	0.1	0.1	0.3	0.2	0.3	0.2
top 0.1%	0.2	0.0	0.6	0.0	0.1	0.1	0.3	0.2	0.3	0.2
Aggregate variables (%)										
r^e	5.6	5.8	6.3	5.1	5.9	5.9	5.9	6.6	6.4	4.6
r^f	1.4	1.5	1.1	1.7	1.9	1.8	1.8	2.4	2.3	0.4
r^d	3.5	3.7	4.1	3.0	3.8	3.8	3.8	3.5	3.3	2.5
g	2.0	2.0	2.0	2.0	2.0	2.0	2.0	3.0	3.0	2.0
c	4.2	5.2	5.3	4.2	4.2	4.1	4.7	4.2	4.2	3.2
$r^e - g$	3.6	3.8	4.3	3.1	3.9	3.9	3.9	3.6	3.4	2.6
$r^e - g - c$	-0.6	-1.4	-1.0	-1.1	-0.3	-0.2	-0.8	-0.6	-0.8	-0.6
Pareto	1.939	0.685	7.998	2.767	1.696	1.437	2.616	1.939	2.582	1.920

Steady state for different parameter values. In each column, the parameter values are identical to the benchmark estimation, except for the parameters listed in the header. Agents are sorted by wealth. In columns (2) and (3), risk aversion equals the level that matches the equity premium in a representative agent model. r^e, r^f, and r^d denote the equity return, bond return, and dividend yield, c is the consumption/wealth ratio of the top 0.1%. Pareto is the Pareto exponent of the top 0.1%.

Table 5: Heterogeneous risk aversion vs. heterogeneous saving rates

	γ shocks	β shocks	data
Returns (%)			
equity	**5.6**	**5.6**	5.6
bonds	**1.4**	**1.4**	1.4
Wealth shares (%)			
bottom 50%	5.7	3.4	1.1
50-90%	24.0	26.3	28.6
top 10%	**70.3**	**70.3**	70.3
top 1%	29.3	30.1	28.1
top 0.1%	**9.9**	**9.9**	9.9
Excess returns (%)			
bottom 50%	0.0	8.4	0.6
50-90%	2.6	5.3	2.2
top 10%	**5.1**	3.6	5.1
top 1%	4.9	3.3	7.0
top 0.1%	4.8	3.2	8.8
Active saving rate (%)			
bottom 50%	0.6	-6.5	14.3
50-90%	-1.7	-4.2	0.0
top 10%	-4.2	-2.9	-3.7
top 1%	-4.2	-2.7	-5.2
top 0.1%	-4.1	-2.6	-6.5
Return volatility (%)	**7.8**	**7.8**	7.8
Safe to total assets (%)	28.1	10.4	32.0

Targeted moments in bold.

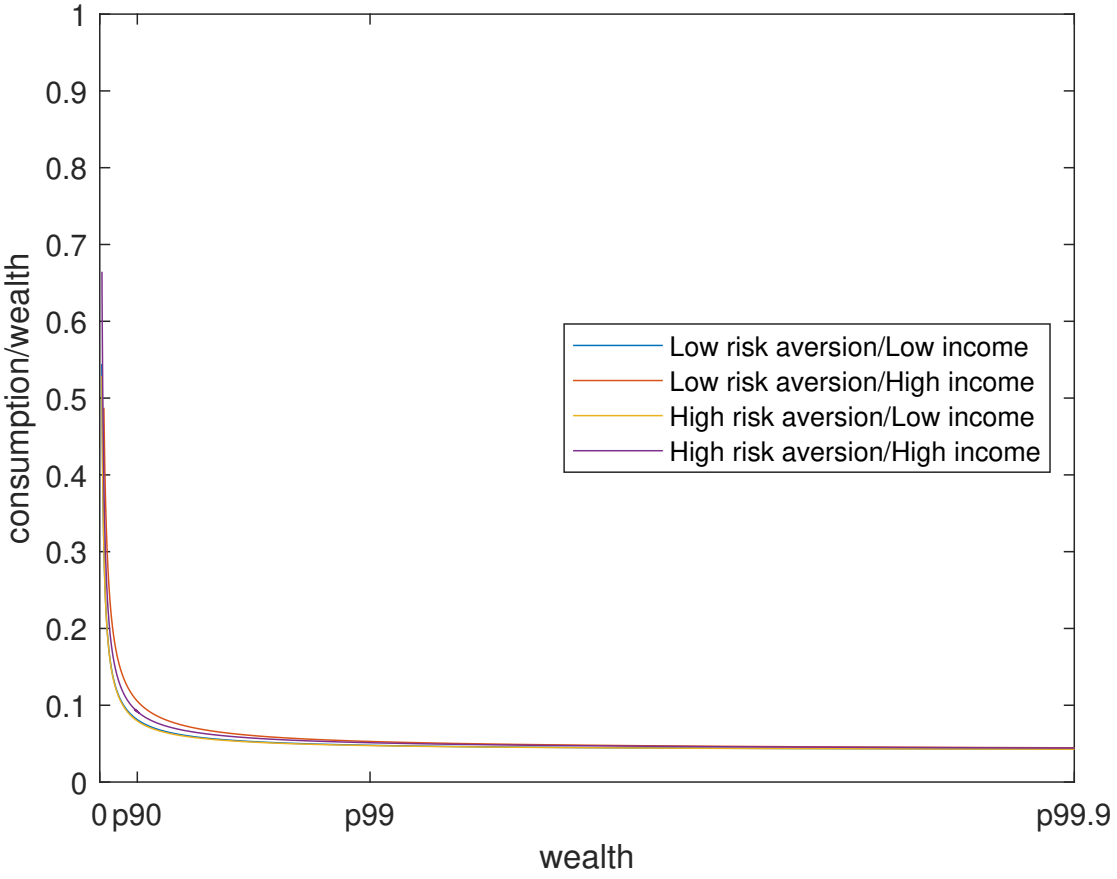

Figure 1: Consumption/wealth ratio

The optimal consumption/wealth ratio ($C_{i,t}/W_{i,t}$) for different wealth values and agent types evaluated at the steady state. The horizontal axis presents the agent's wealth and marks the 90th, 99th, and 99.9th percentiles.

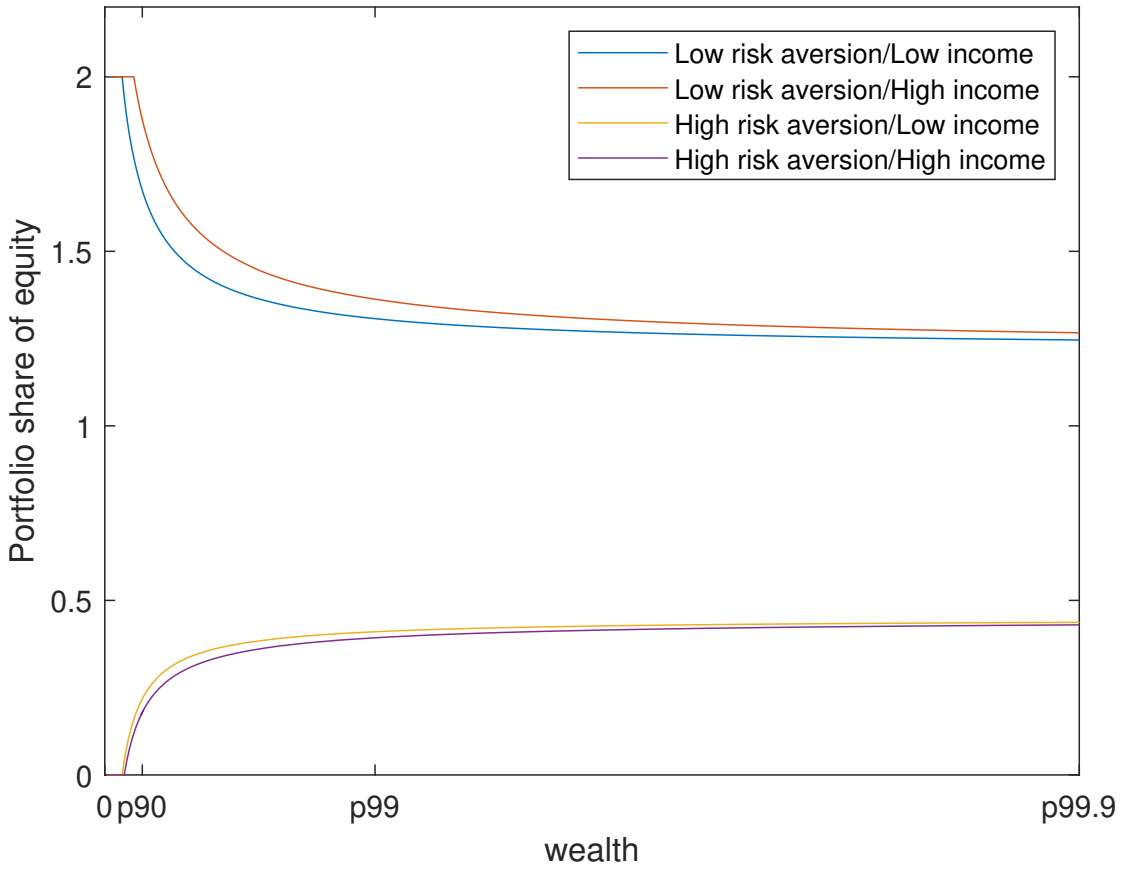

Figure 2: Portfolio share of equity

The optimal portfolio share of equity ($\lambda_{i,t}$) for different wealth values and agent types evaluated at the steady state. The horizontal axis presents the agent's wealth and marks the 90th, 99th, and 99.9th percentiles.

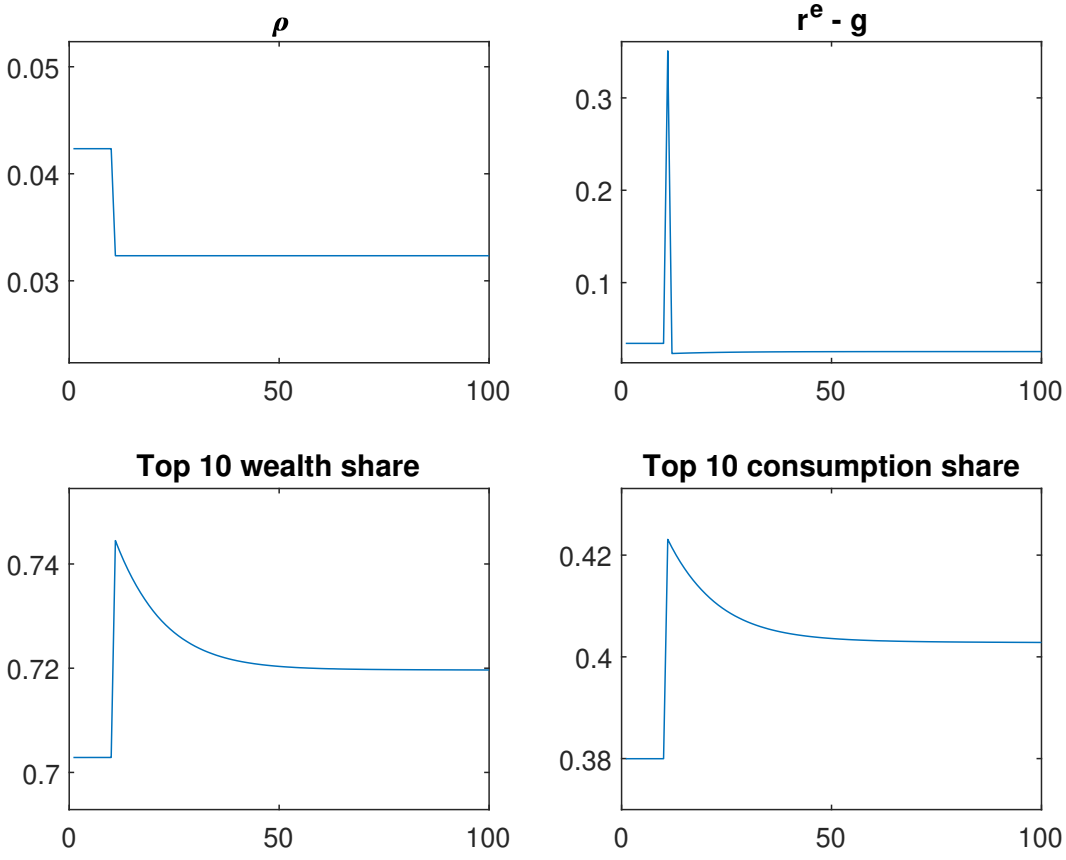

Figure 3: Impulse response: Permanent change in ρ (IES = 1)

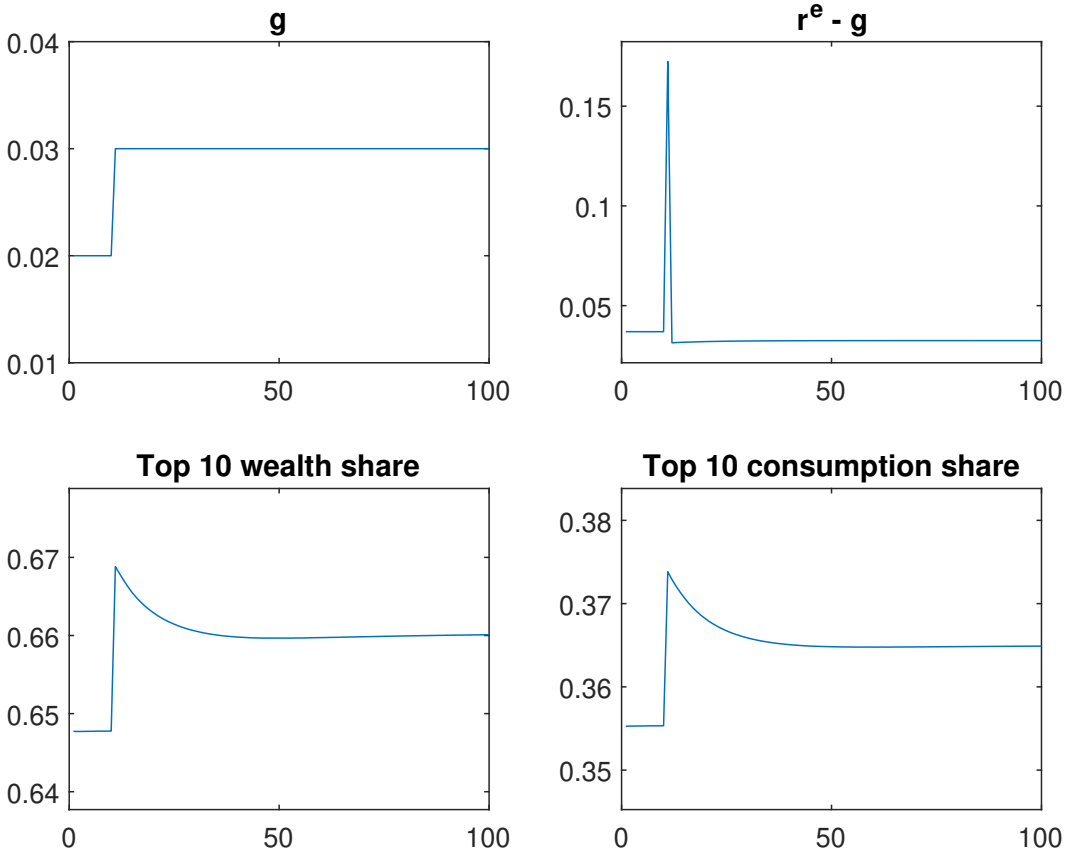

Figure 4: Impulse response: Permanent change in g (IES = 2)

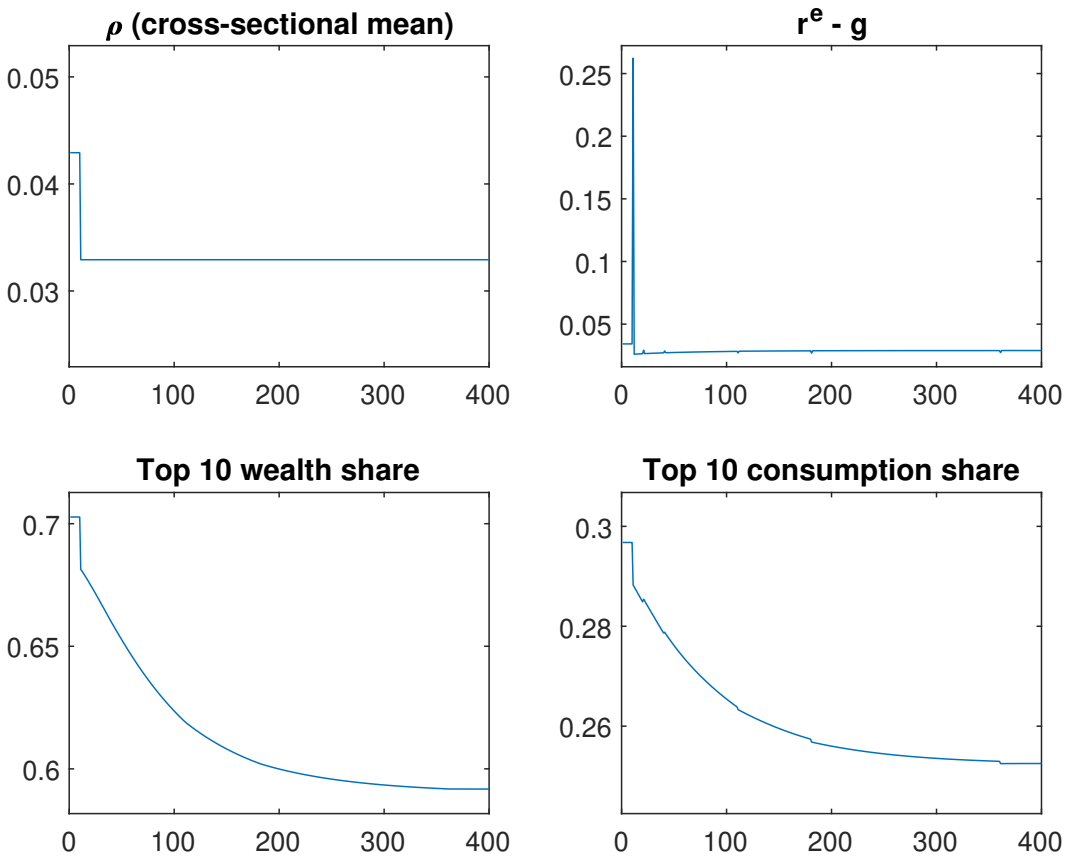

Figure 5: β-shock model: Impulse response to a permanent change in ρ (IES = 1)

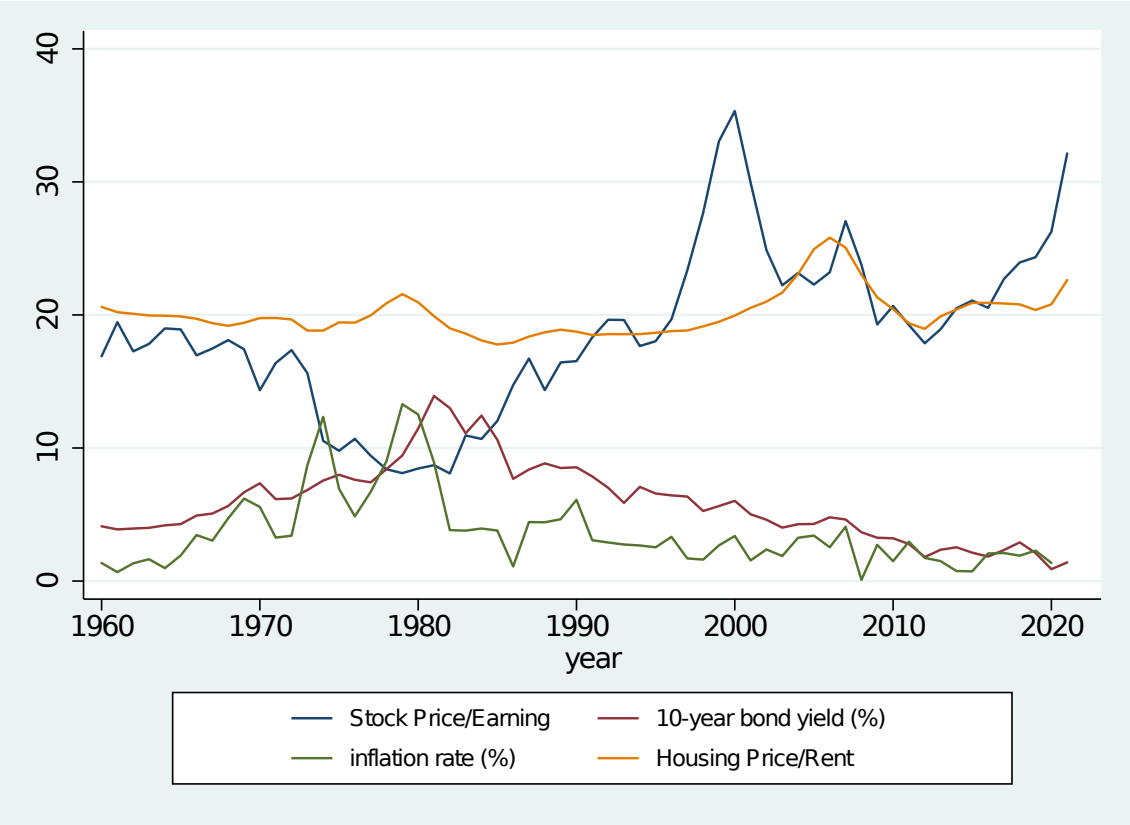

Figure 6: Stock prices, house prices, interest rate, and inflation

Stock and bond data are from Shiller (2015) and housing data from Jordà et al. (2019).

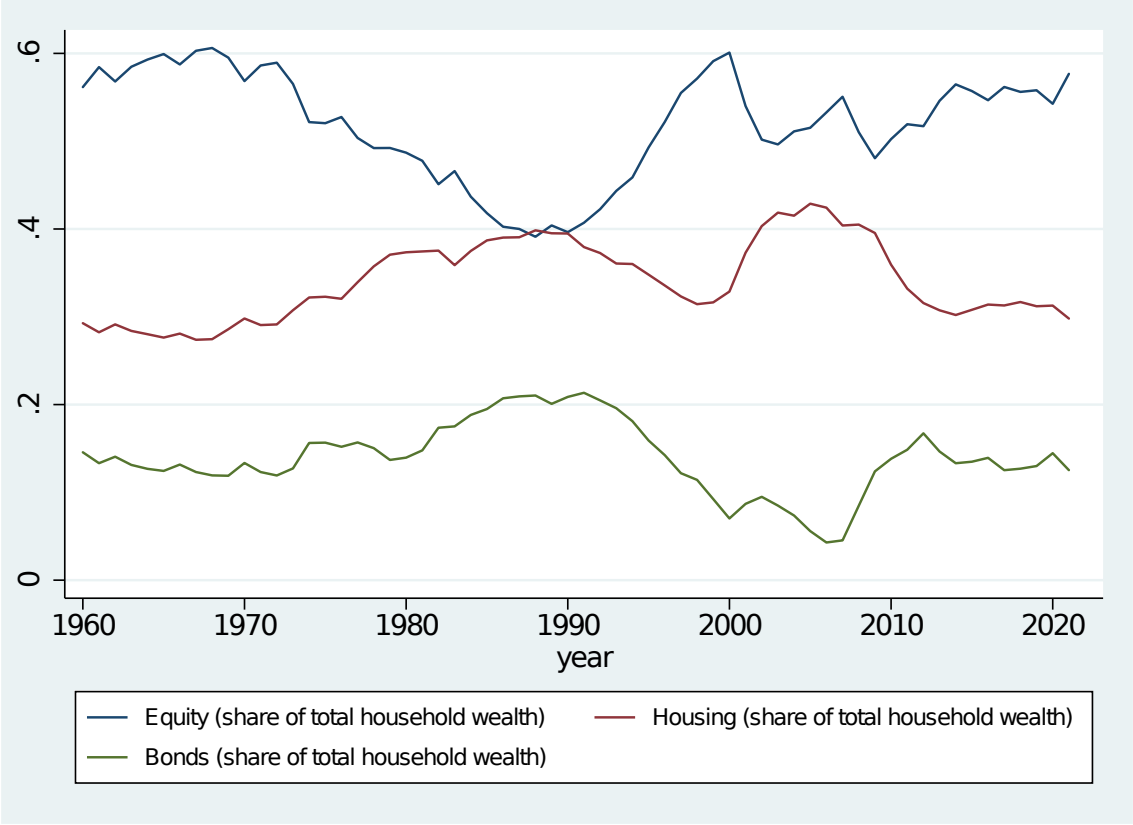

Figure 7: Aggregate portfolio shares of stocks, housing and bonds

Source: U.S. Flow of Funds. Equity includes direct holdings of corporate and noncorporate equity and indirect holdings through pension, insurance, and mutual funds. Bonds include direct and indirect holdings of bonds, deposits, and money market funds net of loans to households.

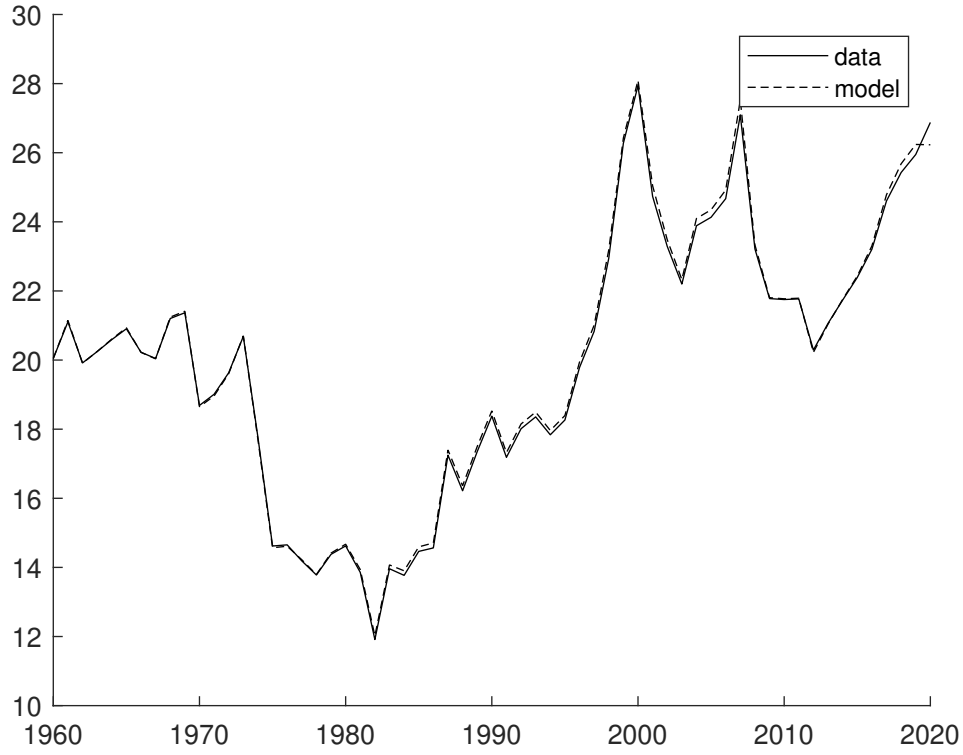

Figure 8: Aggregate price/earning ratio: Model vs. data

The solid line depicts our estimate for the U.S. aggregate price/earning ratio. In our simulation, we use a sequence of discount rate shocks that replicate the dynamics of the price/earning ratio observed in the data. The dashed line depicts the simulated price/earning ratio (normalized to coincide with the actual price/earning ratio in 1960).

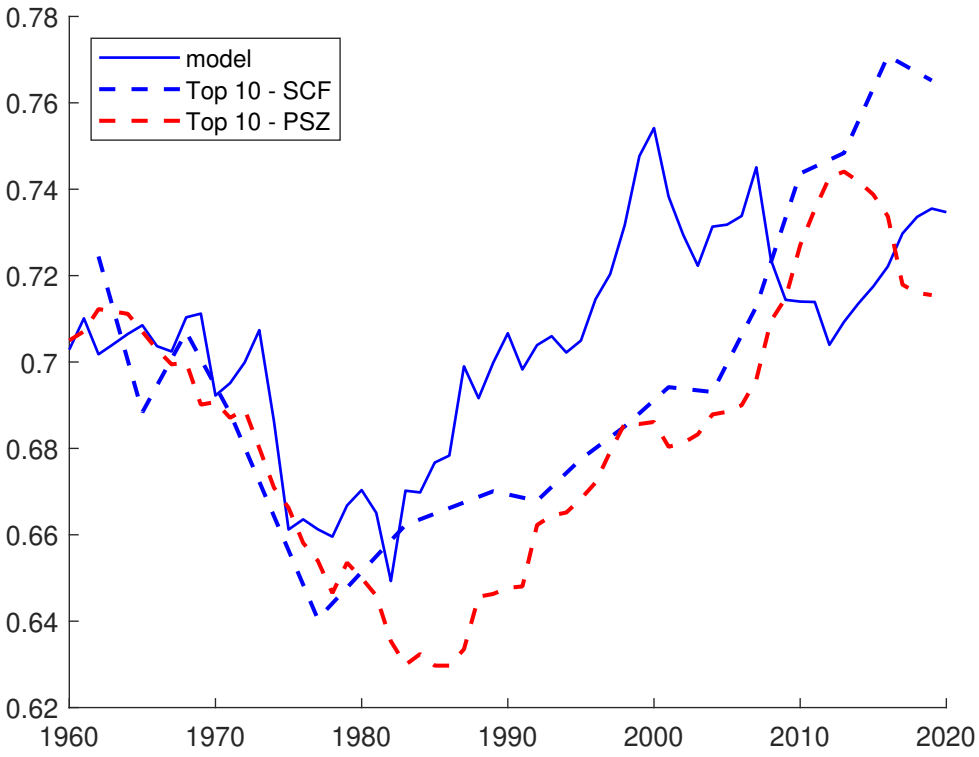

Figure 9: Top 10% wealth share: Model vs. data

The solid blue line is the simulated wealth share of the top 10%. The dashed blue line is actual data from the SCF compiled by Kuhn et al. (2020). The dashed red line is data compiled by Piketty et al. (2018).

Computational Appendix

This appendix describes our computational procedure in detail. First, we rewrite the model in a form similar to Winberry (2018), where the individual variables are approximated globally over a grid, the cross-sectional distribution is approximated by a finite number of moments, and the aggregate variables are approximated locally. Then, we apply the Taylor projection method developed by Levintal (2018) to approximate the local effects of the aggregate variables. Since Taylor projection can be implemented around any point in the state space, we start at an arbitrary initial distribution and simulate the model from that state on. We solve the model several times along the simulation path until the distribution converges.

The appendix is organized as follows. Section A detrends the model and simplifies the notation. Section B approximates the individual decisions over a grid of wealth nodes. Section C approximates the cross-sectional distribution. Section D derives the market-clearing conditions. Section E introduces an auxiliary variable that forecasts the future price/dividend ratio. This variable is used in Section F to define the dynamics of the wealth distribution in a recursive form. Section G summarizes the model conditions in a recursive form. Section H discusses the solution and simulation methods.

A The detrended model

Define the variables $y_{t+1} = \frac{Y_{t+1}}{Y_t}$, $p_t = \frac{P_t}{\alpha Y_t}$, $q_t = Q_t$, $u_{i,t} = \frac{U_{i,t}}{Y_t}$, $c_{i,t} = \frac{C_{i,t}}{Y_t}$, $e_{i,t} = E_{i,t}$, and $b_{i,t} = \frac{B_{i,t}}{Y_t}$. Then, we can rewrite the Epstein-Zin utility function (1) as follows:

$$u_{i,t}^{1-\theta} = (1-\beta_{i,t})\, c_{i,t}^{1-\theta} + \beta_{i,t} z_{i,t}^{1-\theta}, \tag{19}$$

where $z_{i,t}$ is the (detrended) Epstein-Zin certainty-equivalence variable:

$$1 = \mathbb{E}_t \left(\frac{u_{i,t+1} y_{t+1}}{z_{i,t}} \right)^{1-\gamma_{i,t}}, \tag{20}$$

and $\beta_{i,t}$ and $\gamma_{i,t}$ are assumed to be time-varying.

The stochastic discount factor for this general case is:

$$\xi_{i,t+1} = \beta_{i,t} \frac{1-\beta_{i,t+1}}{1-\beta_{i,t}} \frac{c_{i,t}^{\theta}}{c_{i,t+1}^{\theta}} \frac{u_{i,t+1}^{\theta-\gamma_{i,t}}}{z_{i,t}^{\theta-\gamma_{i,t}}} y_{t+1}^{-\gamma_{i,t}}. \tag{21}$$

Define the (detrended) wealth (including labor income):

$$w_{i,t} = \left(1+\epsilon^L_{i,t}\right)(1-\alpha) + \alpha\left(1+p_t\right)\left(1+\epsilon^E_{i,t}\right)e_{i,t-1} + \frac{b_{i,t-1}}{y_t}. \tag{22}$$

Note that $w_{i,t}$ is equal to $\frac{\left(1+\epsilon^L_{i,t}\right)(1-\alpha)Y_t+W_{i,t}}{Y_t}$ in the main text.

The portfolio share of equity is given by:

$$\lambda_{i,t} = \frac{\alpha p_t e_{i,t}}{w_{i,t} - c_{i,t}}. \tag{23}$$

Thus, the budget constraint (2) can be rewritten as:

$$c_{i,t} + \alpha p_t e_{i,t} + q_t b_{i,t} = w_{i,t}, \tag{24}$$

and the future returns on equity and bonds are:

$$R^e_{i,t+1} = \left(1+\epsilon^E_{i,t+1}\right)\frac{1+p_{t+1}}{p_t}y_{t+1} \tag{25}$$

$$R^f_{t+1} = 1/q_t. \tag{26}$$

In addition, define the total portfolio return at the maximum portfolio share $\bar{\lambda}$ by:

$$R^B_{i,t+1} = R^f_{t+1} + \bar{\lambda}\left(R^e_{i,t+1} - R^f_{t+1}\right). \tag{27}$$

The optimality conditions are given by equations (7) and (8) in the main text. Equation (7) becomes:

$$\begin{cases} 0 = \mathbb{E}_t\left(\xi_{i,t+1}R^f_{t+1} - 1\right) & \left| \begin{array}{c} w_{i,t} - c_{i,t} > 0\,,\ \lambda_{i,t} < \bar{\lambda} \\ w_{i,t} - c_{i,t} > 0\,,\ \lambda_{i,t} = \bar{\lambda} \\ \text{else} \end{array} \right. \\ 0 = \mathbb{E}_t\left(\xi_{i,t+1}R^B_{i,t+1} - 1\right) & \\ c_{i,t} = w_{i,t} & \end{cases}, \tag{28}$$

while equation (8) becomes:

$$\begin{cases} 0 = \mathbb{E}_t\left(\xi_{i,t+1}R^e_{t+1} - 1\right) & \left| \begin{array}{c} w_{i,t} - c_{i,t} > 0\,,\ 0 < \lambda_{i,t} < \bar{\lambda} \\ w_{i,t} - c_{i,t} > 0\,,\ 1 > \mathbb{E}_t\xi_{i,t+1}R^e_{t+1} \\ w_{i,t} - c_{i,t} > 0\,,\ 1 > \mathbb{E}_t\xi_{i,t+1}R^f_{t+1} \\ \text{else} \end{array} \right. \\ e_{i,t} = 0 & \\ \alpha p_t e_{i,t} = \bar{\lambda}\left(w_{i,t} - c_{i,t}\right) & \\ e_{i,t} = 0 & \end{cases}. \tag{29}$$

Finally, the market-clearing conditions (9)-(11) are now:

$$\int c_{i,t} di = 1 \tag{30}$$

$$\int e_{i,t} di = 1, \tag{31}$$

where the bond market clears by Walras' law.

Individual wealth $w_{i,t}$ follows the law of motion in equation (22). The idiosyncratic shocks $\gamma_{i,t}$, $\beta_{i,t}$, and $\epsilon^L_{i,t}$ follow Markov processes. The idiosyncratic shock $\epsilon^E_{i,t}$ is an i.i.d. discrete shock, and y_t is an aggregate i.i.d. discrete shock.

B Approximating individual decisions over a grid

Next, we approximate the individual decisions over a grid and derive the optimality conditions at the grid nodes. The general approach is similar to Winberry (2018), with some differences in the details. We classify all agents into types according to the values of the idiosyncratic Markov shocks γ, β, and ϵ^L. The total number of types is denoted I. From now on, we use subscript $i = 1, \ldots, I$ to denote the agent type in period t and subscript $j = 1, \ldots, I$ to denote the agent's type in period $t+1$.[14] The transition probability from i to j is denoted π_{ij}. We drop time subscripts and denote the next period by $'$.

The optimality conditions depend on two constraints. The "portfolio constraint" imposes the condition that $0 \leq \lambda \leq \bar{\lambda}$, where λ is the portfolio share of equity defined in equation (23). The "net-worth constraint" imposes the condition that $w - c \geq 0$.

The portfolio constraint is difficult to address computationally because we do not know a priori if the portfolio decision hits the bounds of this constraint. To resolve this problem, we approximate the portfolio decision under three different scenarios, henceforth "portfolio states," which are as follows: i) The portfolio decision is unconstrained; ii) The portfolio decision hits the lower bound; iii) The portfolio decision hits the upper bound. We denote the portfolio state by superscript index S that takes the values U (Unconstrained), E (constrained to sell more Equity), and B (constrained to sell more Bonds), respectively. The actual portfolio state is endogenous. If the portfolio decision under state U does not violate the bounds of the portfolio constraint, then the endogenous portfolio state is U. If state U violates the lower-bound constraint, then the endogenous portfolio state is E. Otherwise, the portfolio state is B.

[14] Note that we change the meaning of subscript i. In the main text and in Section A, subscript i is an index of the individuals over the unit interval. Here and in the following sections, subscript i is an index of agent type, which determines the values of risk aversion, the time discount rate, and labor income.

B.1 The grid

Fix a grid of N nodes over the interval $[0, \bar{x}]$, where $\bar{x} > 0$ is a fixed parameter. Denote the vector of nodes by:

$$\mathbf{x} = (x_1, \ldots, x_N), \tag{32}$$

where the first node is $x_1 = 0$ and the last node is $x_N = \bar{x}$. The nodes are indexed by subscript $n = 1, \ldots, N$. For instance, c_{in} denotes the consumption of agent type i at node n.

For each agent type $i = 1, \ldots, I$ and portfolio state $S \in \{U, E, B\}$, define the vectors of individual variables over the grid:

$$\mathbf{C}_i^S = \left(C_{i1}^S, \ldots, C_{iN}^S\right) \tag{33}$$

$$\mathbf{E}_i^S = \left(E_{i1}^S, \ldots, E_{iN}^S\right) \tag{34}$$

$$\mathbf{Z}_i^S = \left(Z_{i1}^S, \ldots, Z_{iN}^S\right), \tag{35}$$

where C_{in}^S, E_{in}^S, and Z_{in}^S correspond to endogenous variables determined in period t by agent type i at node n in portfolio state S.[15] Since $S \in \{U, E, B\}$ and i takes I values, the total number of variables in equations (33)-(35) is $3 \times 3 \times I \times N$.

B.2 The wealth cutoff of the net-worth constraint

Now, we derive the wealth cutoff at which the net-worth constraint becomes binding (for each agent type i and portfolio state S). The cutoff is endogenous. We will use it below to determine if the net-worth constraint binds or not.

Let c_{i1}^S and w_{i1}^S denote consumption and wealth at the first node ($n = 1$) for agent type i in portfolio state S. Define consumption at the first node by:

$$c_{i1}^S = \left(w_{i1}^S + h\right) \exp\left(C_{i1}^S\right),$$

where $h = (1 - \alpha)\, p > 0$ is a measure of per capita human capital. This ensures (numerically) that c_{i1}^S is strictly positive (provided that h is sufficiently larger than w_{i1}^S).

Define the first node ($n = 1$) as the wealth cutoff. Namely, the net-worth constraint is slack for any wealth level above w_{i1}^S and binds otherwise. Thus, since $w_{i1}^S = c_{i1}^S$, we obtain the wealth cutoff:

$$w_{i1}^S = \frac{h \exp\left(C_{i1}^S\right)}{1 - \exp\left(C_{i1}^S\right)}. \tag{36}$$

[15] In practice, we eliminate vectors $\mathbf{E}_i^S$ in portfolio states $S = E, B$, because they can be derived analytically by the corner solution. For notational clarity, we treat these vectors here as unknown endogenous variables.

B.3 The interpolation algorithm

We use the following algorithm to interpolate (or extrapolate) the individual decisions for a given level of wealth:

Algorithm 1 *For a given wealth w of agent type i, we compute consumption, $c_i(w)$, Epstein-Zin variable, $z_i(w)$, and equity decision, $e_i(w)$. To do so, we first compute these variables for a given portfolio state $S \in \{U, E, B\}$, and then derive the portfolio state endogenously.*

Given w, we compute the distance of w from the wealth cutoff w_{i1}^S, defined in (36) for agent type i in portfolio state S:

$$x_i^S(w) = \log(w+h) - \log\left(w_{i1}^S + h\right).$$

Use vectors (32)-(35) to interpolate:

$$\begin{aligned} C_i^S(w) &= interpolate\left(x_i^S(w), \boldsymbol{x}, \boldsymbol{C}_i^S\right) \\ E_i^S(w) &= interpolate\left(x_i^S(w), \boldsymbol{x}, \boldsymbol{E}_i^S\right) \\ Z_i^S(w) &= interpolate\left(x_i^S(w), \boldsymbol{x}, \boldsymbol{Z}_i^S\right). \end{aligned}$$

The notation $y = interpolate(x, \boldsymbol{x}, \boldsymbol{y})$ denotes the value of y given x, interpolated through the vectors of nodes and values denoted $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. We use piecewise interpolation.

For agent type i in portfolio state S, compute:

$$c_i^S(w) = \begin{cases} (w+h)\exp\left(C_i^S(w)\right) & \text{if } x_i^S(w) > 0 \\ w & \text{else} \end{cases} \tag{37}$$

$$e_i^S(w) = \frac{1}{\alpha p} \times \begin{cases} -h + (w+h)\exp\left(E_i^S(w)\right) & \text{if } x_i^S(w) > 0 \\ -h + \left(w_{i1}^S + h\right)\exp\left(E_{i1}^S\right) & \text{else} \end{cases} \tag{38}$$

$$z_i^S(w) = \begin{cases} (w+h)\exp\left(Z_i^S(w)\right) & \text{if } x_i^S(w) > 0 \\ \left(w_{i1}^S + h\right)\exp\left(Z_{i1}^S\right) & \text{else} \end{cases}. \tag{39}$$

When $x_i^S(w) \leq 0$, the net-worth constraint binds. Hence, consumption is equal to wealth, as defined in equation (37). Moreover, when the net-worth constraint binds, the equity decision and the Epstein-Zin variable are independent of wealth w. Hence, they are equal to their values at the first node of the grid, as defined in equations (38)-(39).

Finally, derive the portfolio state and the corresponding endogenous variables, depending

on the portfolio decision in the unconstrained state ($S = U$):

$$c_i(w), e_i(w), z_i(w) = \begin{cases} c_i^U(w), e_i^U(w), z_i^U(w) \\ c_i^E(w), e_i^E(w), z_i^E(w) \\ c_i^B(w), e_i^B(w), z_i^B(w) \end{cases} \begin{array}{c|c} 0 < \alpha p e_i^U(w) < \bar{\lambda}\left(w - c_i^U(w)\right) \\ \alpha p e_i^U(w) \leq 0 \\ else \end{array} . \quad (40)$$

B.4 The optimality conditions

For each agent type i in portfolio state S, define wealth at node n by:

$$w_{in}^S = -h + \left(w_{i1}^S + h\right) \exp(x_n),$$

where w_{i1}^S is the wealth cutoff defined in equation (36).

Use Algorithm 1 to compute $c_i^S\left(w_{in}^S\right)$, $e_i^S\left(w_{in}^S\right)$, and $z_i^S\left(w_{in}^S\right)$, through equations (37)-(39). To simplify notation, denote these variables c_{in}^S, e_{in}^S, and z_{in}^S, respectively. The bond decision comes from the budget constraint (24):

$$b_{in}^S = \frac{w_{in}^S - c_{in}^S - \alpha p e_{in}^S}{q}.$$

Next period wealth follows from equation (22):

$$w_{ijn}^{\prime S} = \left(1 + \epsilon_j^{\prime L}\right)(1 - \alpha) + \alpha\left(1 + p^\prime\right)\left(1 + \epsilon^{\prime E}\right) e_{in}^S + \frac{b_{in}^S}{y^\prime}. \quad (41)$$

Here, $w_{ijn}^{\prime S}$ denotes the future wealth of an agent at node n that is currently in portfolio state S and switches from type i to type j, conditional on the future realizations of the stock-return idiosyncratic shock $\epsilon^{\prime E}$, the aggregate shock $y^\prime$, and the price/dividend ratio $p^\prime$.

Use Algorithm 1 to compute $c_j\left(w_{ijn}^{\prime S}\right)$, $e_j\left(w_{ijn}^{\prime S}\right)$, and $z_j\left(w_{ijn}^{\prime S}\right)$ through equation (40). To simplify notation, denote these variables $c_{ijn}^{\prime S}$, $e_{ijn}^{\prime S}$ and $z_{ijn}^{\prime S}$, respectively. Evaluate equations (19) and (21) as follows:

$$u_{ijn}^{\prime S} = \left[(1 - \beta_j)\left(c_{ijn}^{\prime S}\right)^{1-\theta} + \beta_j\left(z_{ijn}^{\prime S}\right)^{1-\theta}\right]^{\frac{1}{1-\theta}}$$

$$\log \xi_{ijn}^{\prime S} = \log \beta_i + \log(1 - \beta_j) - \log(1 - \beta_i) + \theta \log c_{in}^S - \theta \log c_{ijn}^{\prime S} + (\theta - \gamma_i)\left[\log u_{ijn}^{\prime S} - \log z_{in}^S\right] - \gamma_i \log y^\prime.$$

Evaluate equation (20) for each agent type $i = 1, \ldots, I$, node $n = 1, \ldots, N$, and portfolio

states $S = \{U, E, B\}$:

$$0 = \mathbb{E} \sum_{\epsilon^E} \pi^E \sum_j \pi_{ij} \left[\left(\frac{u'^S_{ijn} y'}{z^S_{in}} \right)^{1-\gamma_i} - 1 \right]. \tag{42}$$

The net-worth constraint is (weakly) slack over the grid. Hence, optimality conditions (28)-(29) hold in the following form:

$$\begin{cases} 0 = \mathbb{E} \sum_{\epsilon^E} \pi^E \sum_j \pi_{ij} \left[\xi'^S_{ijn} R'^f - 1 \right] & S = \{U, E\} \\ 0 = \mathbb{E} \sum_{\epsilon^E} \pi^E \sum_j \pi_{ij} \left[\xi'^S_{ijn} R'^B - 1 \right] & S = B \end{cases} \tag{43}$$

$$\begin{cases} 0 = \mathbb{E} \sum_{\epsilon^E} \pi^E \sum_j \pi_{ij} \left[\xi'^S_{ijn} R'^e - 1 \right] & S = U \\ e^S_{in} = 0 & S = E \\ \alpha p e^S_{in} = \bar{\lambda} \left(w^S_{in} - c^S_{in} \right) & S = B. \end{cases} \tag{44}$$

The returns R'^e, R'^f, and R'^B are defined in equations (25)-(27). The expectation operator $\mathbb{E}$ is with respect to the distribution of the aggregate shock y'. The first sum operator runs over the values of the idiosyncratic stock-return shock ϵ^E with corresponding probabilities π^E. The second sum operator runs over future types j with transition probabilities π_{ij}.

C The cross-sectional distribution

Given wealth w, define a normalized wealth measure μ:

$$\mu = \frac{w + h}{1 + \alpha p + h}. \tag{45}$$

Here, wealth is augmented by the measure of per capita human capital h to ensure that μ is strictly positive. In the main text, μ is defined in terms of financial wealth only, which is clearer analytically but computationally less convenient because of negative or zero values.

Let $\mathcal{F}_i(\mu)$ denote the CDF of μ across type i agents, where the mean is:

$$\boldsymbol{\mu}_i = \int \mu d\mathcal{F}_i(\mu). \tag{46}$$

Let θ_i denote the type i population share. Then, at the aggregate:

$$\sum_i \theta_i \boldsymbol{\mu}_i = 1. \tag{47}$$

Define a measure of wealth relative to type i mean $\zeta_i = \log \mu - \log \boldsymbol{\mu}_i$ and the cumulative wealth-share function relative to type i wealth:

$$\Phi_i(\zeta) = \frac{1}{\boldsymbol{\mu}_i} \int_0^{\boldsymbol{\mu}_i \exp(\zeta)} \mu d\mathcal{F}_i(\mu).$$

We approximate the distribution by discretizing the cumulative wealth-share function $\Phi_i(\zeta)$. Specifically, discretize Φ by K fixed points over the interval $[0, 1-\tau]$ (for some $\tau > 0$), denoted $\boldsymbol{\Phi} = (\Phi_1, \ldots, \Phi_K)$. Find K values:

$$\boldsymbol{\zeta}_i = (\zeta_{i1}, \ldots, \zeta_{iK}), \tag{48}$$

that satisfy $\Phi_i(\zeta_{ik}) = \Phi_k \ \forall k = 1, \ldots, K$. The vectors $\boldsymbol{\zeta}_i$ and $\boldsymbol{\Phi}$ provide a discrete approximation of $\Phi_i(\zeta)$, where $\boldsymbol{\zeta}_i$ changes over time and $\boldsymbol{\Phi}$ is held fixed.

The complete distribution is represented by $\boldsymbol{\mu}_i$ and $\boldsymbol{\zeta}_i$ for all types $i = 1, \ldots, I$. We find that the information in $\boldsymbol{\mu}_i$ (the first moment of type i normalized wealth) is sufficient to approximate the local dynamics of asset prices successfully. Hence, we include $\boldsymbol{\mu}_i$ in the vector of state variables. Since all $\boldsymbol{\mu}_i$'s aggregate to 1 through condition (47), we drop $\boldsymbol{\mu}_1$ and define the aggregate state variables as follows:

$$\mathbf{X} = [\boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_I]. \tag{49}$$

The information in $\boldsymbol{\zeta}_i$ captures higher moments. It is important for aggregating accurately across all agents, but its local dynamics are less important. Hence, we treat $\boldsymbol{\zeta}_i$ as locally fixed. This approach simplifies the computations considerably.

D Market-clearing conditions

Consider a certain cumulative wealth share $\Phi_q \in [0, 1]$. Compute the associated wealth measure $\zeta_{iq} = interpolate(\Phi_q, \boldsymbol{\Phi}, \boldsymbol{\zeta}_i)$ and use ζ_{iq} and equation (45) to compute the corresponding wealth $w_{iq} = \boldsymbol{\mu}_i (1 + \alpha p + h) \exp(\zeta_{iq}) - h$. Then, use Algorithm 1 to compute $c_i(w_{iq})$ and $e_i(w_{iq})$ through equation (40).

The market-clearing conditions (30)-(31) are obtained by integrating with respect to

wealth shares:

$$1 = \sum_i \left\{ (1 + \alpha p + h)\, \theta_i \boldsymbol{\mu} \left(\int_0^1 \frac{c_i(w_{iq})}{w_{iq} + h} d\Phi_q \right) \right\}$$
$$1 = \sum_i \left\{ (1 + \alpha p + h)\, \theta_i \boldsymbol{\mu} \left(\int_0^1 \frac{e_i(w_{iq})}{w_{iq} + h} d\Phi_q \right) \right\}.$$

At high wealth levels, individual decisions are proportional to wealth. We exploit this property to approximate the integral over the top tail. For each agent type i, the top tail is defined as the top τ wealth share of type i. Let $\bar{w}_i$ denote an arbitrary wealth level within tail i, and use Algorithm 1 to compute $c_i(\bar{w})$ and $e_i(\bar{w})$. Use these variables to compute the consumption/wealth ratio and the equity/wealth ratio at the top tail.

A quadrature approximates the integral over the region below the top tail. Define a set of Q quadrature nodes over the interval $[0, 1-\tau]$: $\boldsymbol{\Phi} = (\Phi_1, \ldots, \Phi_Q)$. The corresponding weights are $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_Q)$.

Approximate the market-clearing conditions as follows:

$$1 = \sum_i \left\{ (1 + \alpha p + h)\, \theta_i \boldsymbol{\mu}_i \left(\sum_{q=1}^{Q} \frac{c_i(w_{iq})}{w_{iq} + h} \omega_q + \frac{c_i(\bar{w}_i)}{\bar{w}_i + h} \tau \right) \right\} \tag{50}$$
$$1 = \sum_i \left\{ (1 + \alpha p + h)\, \theta_i \boldsymbol{\mu}_i \left(\sum_{q=1}^{Q} \frac{e_i(w_{iq})}{w_{iq} + h} \omega_q + \frac{e_i(\bar{w}_i)}{\bar{w}_i + h} \tau \right) \right\}. \tag{51}$$

E Future price/dividend ratio

Let $\mathbf{H}$ denote the law of motion of the state vector $\mathbf{X}$:

$$\mathbf{X}' = \mathbf{H}(\mathbf{X}, y'), \tag{52}$$

where y' is the only aggregate shock in the model. Let the function $\mathbf{p}$ denote the solution of the price/dividend ratio $p = \mathbf{p}(\mathbf{X})$.

Let $p'^f(y')$ be the future price/dividend ratio forecasted in period t conditional on the future realization of y'. Namely, $p'^f(y')$ is defined by:

$$p'^f(y') = \mathbf{p}(\mathbf{X}') = \mathbf{p}(\mathbf{H}(\mathbf{X}, y')). \tag{53}$$

Note that $p'^f(y')$ is determined in period $t+1$, but it is known in period t, because it is conditional on the future realization of y', which is the only unknown future variable. This

will enable us below to define the evolution of wealth as a function of current state variables and future aggregate shocks.

F Law of motion of the moments

This section computes the law of motion of $\boldsymbol{\mu}_i$, defined in equation (46). The normalized wealth of agent type i with wealth w_{iq} is $\mu_{iq} = \frac{w_{iq}+h}{1+\alpha p+h}$. Compute the bond decision through the budget constraint (24): $b_{iq} = \frac{w_{iq}-c_{iq}-\alpha p e_{iq}}{q}$. Here, c_{iq} and e_{iq} denote $c_i(w_{iq})$ and $e_i(w_{iq})$ from Section D, respectively. Future wealth is derived from equation (22):

$$w'_{ijq} = \left(1+\epsilon'^L_j\right)(1-\alpha) + \alpha\left(1+p'^f(y')\right)\left(1+\epsilon'^E\right)e_{iq} + \frac{b_{iq}}{y'}. \tag{54}$$

Here, w'_{ijq} denotes the future wealth of agent type i with current wealth w_{iq} that switches in period $t+1$ to type j, conditional on the future realizations of y' and ϵ'^E. We use the future price/dividend ratio $p'^f(y')$ defined in equation (53), which is a function of the current state $\mathbf{X}$ and the future realization y'. Hence, equation (54) describes future wealth as a function of variables known in period t and future shocks only. This allows us to define the dynamics of the wealth distribution recursively.[16]

Define $h' = (1-\alpha)\,p'^f(y')$ and use equation (45) to define future normalized wealth:

$$\mu'_{ijq} = \frac{w'_{ijq}+h'}{1+\alpha p'^f(y')+h'},$$

where h' ensures that μ'_{ijq} depends on variables known in period t and future shocks only.

It follows that the mean future normalized wealth of agent type j conditional on future realization y' is:

$$\boldsymbol{\mu}'_j(y') = \frac{1}{\theta_j}\sum_{\epsilon^E}\pi^E\sum_i \theta_i\boldsymbol{\mu}_i\pi_{ij}\int_0^1 \frac{\mu'_{ijq}}{\mu_{iq}}d\Phi_q.$$

Here, $\theta_i\boldsymbol{\mu}_i$ is the current share of type i agents in aggregate wealth, and $\theta_i\boldsymbol{\mu}_i\pi_{ij}d\Phi_q$ is the current wealth share associated with type i agents that switch to type j, whose wealth-share growth is $\frac{\mu'_{ijq}}{\mu_{iq}}$. It follows that $\theta_i\boldsymbol{\mu}_i\pi_{ij}\int_0^1\frac{\mu'_{ijq}}{\mu_{iq}}d\Phi_q$ is the future wealth share of agents that switch from i to j. Integrating across all types $i=1,\ldots,I$ and future realizations of the idiosyncratic stock-return shock ϵ^E yields the future wealth share of type j agents (conditional on y'), namely $\theta_j\boldsymbol{\mu}'_j$. Multiplying by $\frac{1}{\theta_j}$ yields $\boldsymbol{\mu}'_j$.

[16] By comparison, future wealth in equation (41) is defined as a function of p', which is an endogenous variable determined in $t+1$. Hence, equation (41) does not admit a recursive form for wealth.

To approximate the integral accurately, we follow the procedure used in equations (50)-(51):

$$\boldsymbol{\mu}_j'\left(y'\right)=\frac{1}{\theta_j}\sum_{\epsilon^E}\pi^E\sum_i\theta_i\boldsymbol{\mu}_i\pi_{ij}\left\{\sum_{q-1}^{Q}\frac{\mu_{ijq}'}{\mu_{iq}}\omega_q+\frac{\mu_{ij}'\left(\bar{w}_i\right)}{\mu_i\left(\bar{w}_i\right)}\tau\right\}. \tag{55}$$

G Summary of the model conditions

The state vector is defined in condition (49). The control variables include the variables over the grid, defined in equations (33)-(35), the price/dividend ratio p, and the price of a risk-free bond q.

We define additional control variables as follows. Let $y_1',\ldots,y_M'$ denote the future realizations of y'. Define a new variable denoted $p_m'^f$, which is the future price/dividend ratio, defined in equation (53), conditional on realization y_m':

$$p_m'^f=p'^f\left(y_m'\right). \tag{56}$$

Similarly, let $\boldsymbol{\mu}_{im}'$ denote future mean normalized wealth conditional on realization y_m' determined by equation (55):

$$\boldsymbol{\mu}_{im}'=\boldsymbol{\mu}_i'\left(y_m'\right). \tag{57}$$

Importantly, $p_m'^f$ and $\boldsymbol{\mu}_{im}'$ depend only on variables known in period t. Hence, we include them in the vector of control variables together with the other variables that are determined in period t: $\mathbf{Y}=\left[\left\{\mathbf{C}_i^S,\mathbf{E}_i^S,\mathbf{Z}_i^S,p_m'^f,\boldsymbol{\mu}_{im}'|i=1,\ldots,I,S=\{U,E,B\},m=1,\ldots,M\right\},p,q\right]$. The solution of $\mathbf{Y}$ is denoted:

$$\mathbf{Y}=\mathbf{G}\left(\mathbf{X}\right). \tag{58}$$

We can define the future mean normalized wealth as $\boldsymbol{\mu}_i'=\sum_m\mathbb{1}\left(y'=y_m'\right)\boldsymbol{\mu}_{im}'$, where $\boldsymbol{\mu}_{im}'$ is defined by equation (57) and known in period t. It follows that the future vector of aggregate state variables (49) can be written as a function of variables known in period t and the future aggregate shock, denoted:

$$\mathbf{X}'=\Theta\left(\mathbf{X},\mathbf{Y},y'\right), \tag{59}$$

where Θ is a known function. Substituting equation (58) yields the function $\mathbf{H}$ in equation (52):

$$\mathbf{H}\left(\mathbf{X},y'\right)=\Theta\left(\mathbf{X},\mathbf{G}\left(\mathbf{X}\right),y'\right).$$

The solution is identified by the optimality conditions (42)-(44), the market-clearing con-

ditions (50)-(51), and conditions (56)-(57). Together, these conditions can be written as:

$$0 = \mathbb{E}\mathbf{F}\left(\mathbf{Y}', \mathbf{Y}, \mathbf{X}', \mathbf{X}, y'\right), \tag{60}$$

where $\mathbb{E}$ denotes expectations with respect to the distribution of the aggregate shock y'.[17] Equations (58), (59), and (60) complete the definition of the model.

H Solving and simulating the model

The model solution (58) provides the effects of the cross-sectional distribution, approximated by the moments in $\mathbf{X}$, on the control variables $\mathbf{Y}$. We approximate this solution by the Taylor projection method developed by Levintal (2018). First, postulate a linear solution:[18]

$$\mathbf{Y} = \mathbf{X}\mathbf{b},$$

and substitute in equations (59) and (60) to get:

$$0 = \mathbb{E}\mathbf{F}\left(\Theta\left(\mathbf{X}, \mathbf{X}\mathbf{b}, y'\right)\mathbf{b}, \mathbf{X}\mathbf{b}, \Theta\left(\mathbf{X}, \mathbf{X}\mathbf{b}, y'\right), \mathbf{X}, y'\right). \tag{61}$$

Then, differentiate (61) with respect to $\mathbf{X}$:

$$0 = \mathbb{E}\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\left(\Theta\left(\mathbf{X}, \mathbf{X}\mathbf{b}, y'\right)\mathbf{b}, \mathbf{X}\mathbf{b}, \Theta\left(\mathbf{X}, \mathbf{X}\mathbf{b}, y'\right), \mathbf{X}, y'\right). \tag{62}$$

We evaluate equations (61)-(62) at a given state $\mathbf{X}_0$ and find $\mathbf{b}$ that satisfies these conditions. As shown in Levintal (2018), the solution is accurate locally around $\mathbf{X}_0$.

To simulate the model, we start at an arbitrary distribution of normalized wealth μ, measured by the CDF $\mathcal{F}_i(\mu)$ $\forall i$, and discretized over a dense grid with 10^4 nodes. Denote the simulated distribution in period t by $\boldsymbol{\mathcal{F}}_t$. Starting at $t = 0$, we use $\boldsymbol{\mathcal{F}}_t$ to compute the first moments $\boldsymbol{\mu}_i$ $(\forall i)$ defined in equation (46), and store them in the aggregate state $\mathbf{X}_t$, defined in equation (49). In addition, we compute the higher moments $\boldsymbol{\zeta}_i$ $(\forall i)$ defined in equation (48), and pass them into equation (60) as fixed parameters. As explained in the main text and in Section C, the local dynamics of the higher moments are not essential for forecasting asset prices. Namely, forecasts based on first moments only (included in $\mathbf{X}_t$) are sufficiently accurate. However, the higher moments are important for aggregating accurately across the distribution. Hence, they are included in the model as fixed parameters. This

[17] To define conditions (56)-(57) in the form of equation (60), we can use indicator functions that select that relevant realization y'_m.

[18] The method can be implemented at higher orders, but this is not necessary for our analysis.

approach substantially reduces computational costs.

We solve the model at $\mathbf{X}_t$ and obtain the solution $\mathbf{Y}_t = \mathbf{G}(\mathbf{X}_t)$. We use this solution to compute asset prices and individual decisions across the wealth distribution, from which we compute the next-period distribution $\boldsymbol{\mathcal{F}}_{t+1}$. Repeating the same procedure, we compute $\mathbf{X}_{t+1}$ and evaluate the model conditions (60) under the previous solution. If the residuals are too large, we solve the model again at the new state. We continue with the simulation until the distribution converges.