NBER WORKING PAPER SERIES

RETURNS HETEROGENEITY AND CONSUMPTION INEQUALITY OVER THE LIFE CYCLE

Claudio Daminato
Luigi Pistaferri

Working Paper 32490
http://www.nber.org/papers/w32490

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2024

We are grateful to various conference and seminar participants for comments and to Gabriela Rays Wahba for research assistance. All errors are ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Returns Heterogeneity and Consumption Inequality Over the Life Cycle
Claudio Daminato and Luigi Pistaferri
NBER Working Paper No. 32490
May 2024
JEL No. E21,G51,J24

ABSTRACT

A recent literature argues that persistent heterogeneity in wealth returns ("type dependence") as well as a positive association with wealth levels ("scale dependence") play an important role for explaining features of the wealth distribution, especially its extreme concentration at the top. In contrast, traditional models of wealth accumulation emphasize the role of persistent differences in labor earnings. Using panel data from the PSID, we first document that a common unobserved component (which we interpret as the endowment of cognitive and non-cognitive skills of an individual) drives persistent heterogeneity in both wealth returns and labor earnings. We embed these features of the joint wealth return-earnings process in a life-cycle model of consumer behavior and show that ignoring them would dramatically understate average returns for people at the top of the wealth distribution as well as the level and rise of consumption inequality over the life cycle.

Claudio Daminato
Lund University
Department of Economics, Scheelevägen 15B
223 63 Lund
Sweden
claudio.daminato@nek.lu.se

Luigi Pistaferri
Department of Economics
579 Jane Stanford Way
Stanford University
Stanford, CA 94305-6072
and NBER
pista@stanford.edu

1 Introduction

Most households finance their consumption expenditures through two channels: their labor earnings and their capital income. Traditionally, the focus in the consumption literature has been on the first channel (Deaton, 1991, Carroll, 1997). In recent years, however, evidence has been rapidly accumulating on the vast and persistent heterogeneity that exists on returns to capital (and, symmetrically, on the cost of debt). A series of papers have documented, in particular, that returns to wealth exhibit *type dependence* (returns are heterogeneous and persistently so) and *scale dependence* (returns increase with the level of wealth). See, for example, Fagereng et al. (2020b), Bach et al. (2020), and Bricker et al. (2018), among others. In theoretical contributions, these features of the returns distribution have been shown to have powerful implications, in particular to explain the rise in the level of wealth inequality and the increasing concentration at the very top (Saez and Zucman, 2016; Smith et al., 2022). See in particular Benhabib and Bisin (2018) and Gabaix et al. (2016).

While these new empirical and theoretical contributions have shifted the focus from heterogeneity in labor income to heterogeneity in capital income, there are still many open questions. The first is what economic mechanisms or features of the economic environment can explain type or scale dependence in returns. A traditional explanation in household finance is risk-taking: because of the insurance provided by their wealth or because they can more easily overcome the cost of participating in stock markets, people at the top of the wealth distribution have an asset portfolio that is more geared towards risky assets, which helps explaining the higher wealth returns they generate (i.e., scale dependence), and – given that risk tolerance is a trait of individual preferences that changes very slowly, if at all – their persistence as well (i.e., type dependence). A different, not necessarily alternative explanation, is financial literacy or sophistication: wealthy individuals may have better analytical skills or knowledge of financial products or have the wealth to acquire the services of financial advisors, which may produce higher returns on average. Like preferences, differences in financial literacy are likely to be very persistent.

A second important question is about the implications of type dependence in wealth returns for well-being. Does it exacerbate inequality in consumption, and hence living standards? The answer is not obvious. On the one hand, a greater accumulation of wealth induced by persistently higher returns may not necessarily translate into higher disparities in consumption. In some extreme examples, such as those considered in the "wealth in the utility function" literature (see for example Michaillat and Saez, 2021 and the references therein), people accumulate wealth to achieve "status", which in principle may not alter their consumption levels. On the other hand, persistent heterogeneity in returns produces, for consumption, the equivalent of HIP (heterogeneous-income-profile) trajectories observed in the literature on income dynamics (Guvenen, 2009). From an Euler equation perspective, persistent return heterogeneity induces a distribution of consumption growth rates in the cross-section: some households with high returns defer consumption and others do the opposite. The main implication for consumption inequality is that the dispersion in consumption profiles will grow over the life cycle because of persistent return heterogeneity. Moreover, the extra wealth generated by higher returns at the top probably feeds into higher bequests or *in vivo* transfers, which simply shifts the inequality concerns from intra-generational to inter-generational, with implications also for welfare and taxation. Finally, persistent return heterogeneity broadens the concept of permanent income to include persistent differences in capital income alongside differences in labor earnings that have been the focus of the literature. A positive correlation between type dependence in wealth returns and type dependence in labor earnings (which as we shall see is a robust feature of the data) will exacerbate the extent of consumption inequality.

This paper tackles these two broad, open questions. First, to be able to distinguish between scale dependence induced by higher risk-taking vs. that explained by "skills", it is key to allow for both features in a model. We thus include portfolio choice between safe and risky assets in a rich buffer stock model, as well as unobserved skills shifting average returns. Second, we study how the return heterogeneity mechanism alters consumption inequality over the life cycle. In traditional buffer stock models (Deaton and Paxson, 1994), the growth in consumption inequality over the life cycle can only be justified by uninsurable permanent (labor) income shocks that accumulate as people age, as initial conditions in the income process can only explain differences in consumption levels but not growth rate heterogeneity. In our setting, initial conditions matter because heterogeneity in wealth returns play a role through the intertemporal budget constraint. One important difference with previous work is thus that our focus is more on the role of "initial conditions" than on shocks that accumulate over the life cycle (although the latter are also allowed for).

To better quantify the importance of persistent return heterogeneity on consumption and saving choices, our model allows for all motives for saving emphasized in the literature. Precautionary saving in our model emerges in response to several risks faced by the household, such as shocks to labor earnings, survival uncertainty, and out-of-pocket medical care spending shocks. Bequest motives are accounted for by including a non-homothetic bequest function. Finally, we model the tax and transfer system in a realistic way, including a progressive tax and social security system and an explicit consumption floor (which incorporates the distortions on saving behavior emphasized by e.g., Hubbard et al., 1995).

To guide us towards a more realistic theoretical modeling of the nature of heterogeneity in wages and capital income, we start by looking at panel data on labor earnings and estimated wealth returns. The data show that a common unobserved component (which we interpret as the stock of endowed cognitive and non-cognitive skills) appears to drive persistent heterogeneity in both wealth returns and labor earnings: individuals who do persistently well in labor markets appear to do persistently well in asset markets too (and *vice versa*). While we are agnostic about the precise mechanism behind the common factors that drive this correlation, one can easily imagine that they derive from differences in preference traits driving both investment choices in human or financial capital, or differences in the efficiency of search for jobs or financial investment opportunities, among other things. For example, financially sophisticated individuals may also be the ones who adopt more effective job search strategies or savvier investments in their human capital.[1]

We estimate the parameters of interest (extent of persistent heterogeneity in wealth returns and earnings, as well as preference parameters) using panel data from the US Panel Study of Income Dynamics (PSID). Structural estimation uses moments of the joint distribution of risky asset market participation, consumption, earnings, assets, and wealth returns. The data confirm findings from other countries that both type and scale dependence characterize the behavior of returns from assets. The model replicates well the rise over the life cycle of the wealth-income ratio and the life cycle evolution of consumption inequality. In a counterfactual exercise, we show that risk-taking alone would be unable to replicate the extent of scale dependence in wealth returns we see in the data; unobserved skills are an important factor, and especially so at the very top of the distribution. In another counterfactual exercise, we document that eliminating persistent return heterogeneity or common unobserved factors affecting both labor earnings and wealth returns would dramatically understate the level and the growth of consumption inequality over the life cycle. These are novel findings that, as far as we know, have not been shown in existing work.

Our paper is related to some important contributions in the literature. Benhabib et al. (2019) investigate to what extent heterogeneity in saving rates, capital income risk, and skewed and persistent stochastic earnings can explain the cross-sectional distribution of wealth as well as the extent of social mobility observed in the data. Our perspective is more on the life cycle than on the cross-sectional features of the wealth distribution; partly as a consequence of this different focus, households in our model face a more complex environment than in Benhabib et al. (2019).

Lusardi et al. (2017) consider a rich model where heterogeneity in returns to wealth

[1]We want to stress that the persistent heterogeneity in wealth returns that we attribute to differences in "skills" or financial sophistication does not necessarily imply that "skilled" or financially literate individuals persistently generate *above market* returns. All individual returns may well be below the average market returns; what matters is that they are persistently different across investors.

derives endogeneously from a process of investment in financial knowledge (an approach similar to Jappelli and Padula, 2013; see also Fagereng et al., 2020a). Since they lack wealth returns data, the model is calibrated to match various features of the distribution of wealth. Their conclusion is that heterogeneity in returns to wealth explains 30%-40% of overall wealth inequality. Finally, Barth et al. (2023) uses gene endowments and gene-by-environment interactions to study how genes that predict education affect wealth accumulation over the life-cycle. In particular, the authors study how genetic endowments affect labor income and financial sophistication (which translates into heterogeneous returns) and in turn amplify the role of genetic endowment on wealth accumulation. Our focus is more on trying to understand how return heterogeneity affects consumption inequality over the life cycle and how to separate scale dependence due to "skill endowments" vs. portfolio choice.

The rest of the paper is as follows. In Section 2 we first describe the data and how we measure wealth, labor earnings, and returns to wealth, and then provide some key facts about type and scale dependence in returns. Section 3 describes the life cycle model, while Section 4 explains how we estimate the parameters of interest. Section 5 focuses on the implications of the model for the observed inequality in consumption and wealth, and considers a variety of counterfactual exercises to shed light on the forces behind them. Sections 6 concludes.

2 Data and Key Facts

We start by documenting some key facts about household wealth returns and household earnings using panel data from various waves (1997-2019) of the PSID. The PSID is a longitudinal survey of a representative sample of U.S. individuals and the families in which they reside. During the 1997-2019 period the survey was conducted bi-annually. Besides demographic information, the PSID routinely collects information on labor earnings of household members (husbands and wives in traditional couples), as well as detailed information on end-of-year household assets and liabilities, investment flows into different asset classes, as well as sources of capital income and debt payments. The key advantage of the PSID over, e.g., administrative data, is that it combines information on income and assets with detailed information on households' consumption. Moreover, survey data also allow to control for additional dimensions of heterogeneity when trying to explain individual differences in wealth returns. In particular, the PSID data allow researchers to construct a proxy for respondents' risk preferences and the relevance of intergenerational transfers.[2]

[2]The 1997 wave of the PSID includes a battery of questions on individuals' risk aversion, while the 2013 wave includes an additional module collecting information on whether the respondents received help from their parents (in various forms) since turning 18.

2.1 Sample and basic definitions

We restrict our analysis to the period 1999-2019 since it is only starting with the 1999 wave that information on asset holdings and investment flows is collected in each wave of the survey. Both in the empirical analysis and in the quantitative model, we consider a unitary framework of household decision making and therefore focus on household-level data. We select households whose head is neither studying nor retired and between 25 and 65 years of age. We also drop households where the head is mostly self-employed in the observation period, as our quantitative framework is better suited to study the decisions of employed workers. Finally, we drop the oversampled low-income households ("SEO" sample). The analysis sample consists of 19,399 household-year observations.

We define gross wealth as the sum of financial (safe and risky) assets and real assets: cash, bonds, stocks, pensions/IRA, private business wealth, housing and other real estate, and vehicles. Net wealth is defined as the difference between gross wealth and outstanding debt (the sum of mortgage debt, credit debt, student debt, medical debt, and legal debt). Our measure of total household consumption expenditure includes (a) food (food at home and food away from home, and the monetary value of food stamps), housing (actual and imputed rent for renters and homeowners, respectively), transportation, utilities, and expenditure on vehicles (such as maintenance, repairs, etc.), education, and insurance, and (b) spending on clothing, home repairs and furnishing, vacation, recreation and donations. The categories in group (a) are available for the entire sample period 1999-2019, while the categories in group (b) were added with the 2005 wave. Because these additional spending categories are quantitatively important ($\simeq$ 18% of total consumption), our baseline analysis uses consumption data for the sub-sample period 2005-2019. The survey also includes detailed information on out-of-pocket medical spending (prescriptions, in-home medical care, doctors, outpatient surgery, dental bills, hospital bills and nursing home). We use these data to model the risk of unexpected health care shocks but assume that medical expenses produce no utility benefits *per se*, and thus exclude medical spending from our consumption expenditure measure.

2.2 Returns to Wealth from the PSID

We construct three measures of returns to wealth from the PSID:

1. Returns to net wealth:
$$r_t = \frac{(y_t^c + cg_t - y_t^d)}{(A_{t-1} + 0.5F_t)} \tag{1}$$
where y_t^c are interest income and dividends from bonds, stocks, real estate (not includ-

ing own housing) and private business, cg_t the "capital gains/losses" from business, rents, stocks, real estate, pension/IRA,[3] y_t^d payments on debt, A_{t-1} is total household's net wealth at the beginning of the previous period, and F_t is an estimate of the net investment flows into businesses, stocks, real estate and pensions (which are directly reported in the PSID). This term accounts for the fact that asset yields are generated not only by beginning-of-period wealth but also by additions/subtractions of assets during the year. Without this adjustment estimates of returns would be biased. The bias is most obvious in the case in which beginning-of-period wealth is "small" but capital income is "large" due to positive net asset flows occurring during the period. Ignoring the adjustment would clearly overstate the return. The opposite problem occurs when assets are sold during the period.

2. Returns to gross wealth:
$$r_t^G = \frac{(y_t^c + cg_t)}{(A_{t-1}^G + 0.5F_t)}$$
where $A_{t-1}^G = A_{t-1} + D_{t-1}$ is gross household wealth and D household debt.

3. Returns on assets (ROA):
$$r_t^A = \frac{(y_t^c + cg_t - y_t^d)}{(A_{t-1}^G + 0.5F_t)}$$

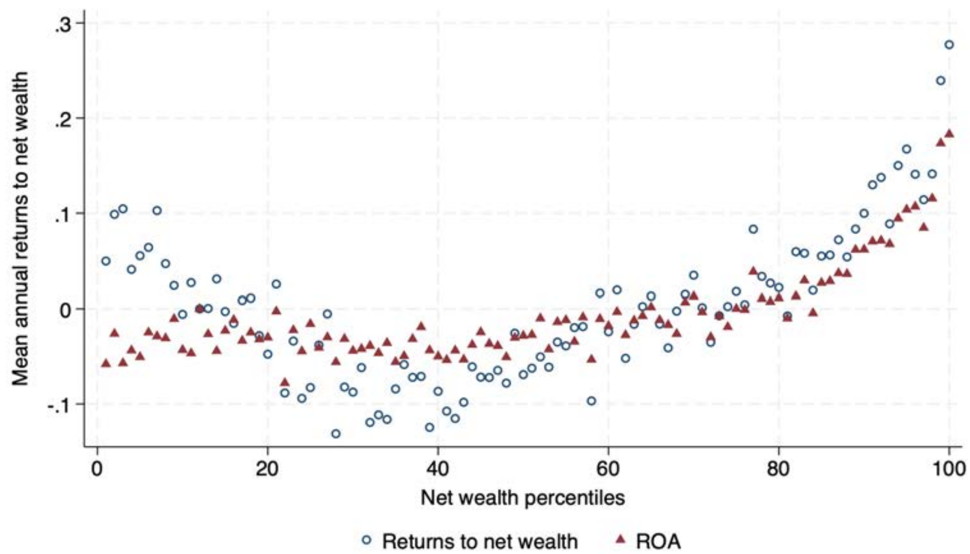
In Figure 1 we plot the estimated return to net wealth against the percentile of net wealth distribution. The figure shows strong evidence for *scale dependence*, at least above the 30-th percentile. Average returns to net wealth below the 30-th percentile are positive because in this region households have negative net wealth and payments on debt exceeding interests and dividends, so that both the numerator and denominator of equation (1) are negative. In the same figure we also plot the return on assets (ROA, the red dots), which eliminates this issue. The ROA measures how much net income is generated by one dollar worth of assets.

The degree of scale dependence is substantial: households in the top percentile have returns to net wealth (ROA) that are approximately 30 (20) percentage points higher than households with median wealth. There is also an increase in the slope of the relationship (i.e., a more convex relationship) as we move to the upper percentiles. In Figure 2 we replicate Figure 1 for the return to gross wealth and find again strong evidence for scale dependence, at least above the median.

To get a gauge of the importance of *type dependence* in wealth returns, in Table 1 we replicate the approach of Fagereng et al. (2020b) and regress wealth returns against various controls as well as household fixed effects. We do so for two samples: the whole sample

[3] We estimate capital gains using the difference in reported stocks across waves.

Figure 1: Annualized returns to wealth across the net wealth distribution. The figure plots the mean annual returns to net wealth (the blue dots) and the return on assets (the red triangles), in each percentile of net wealth. Data is from the 1999-2019 waves of the PSID.



and those with non-missing information on intergenerational transfers and the degree of risk aversion (variables that we use for some of the exercises discussed below).[4] The relevance of *type dependence* can be assessed by looking at the increase in the adjusted R^2 when we move from a specification without to one with fixed effects. Notice that in a standard Merton-Samuelson portfolio choice model, controls for portfolio allocation shares, year effects, risk aversion as well as socio-demographic variables capturing differences in human capital should explain the entire variation in wealth returns. In contrast, we find that fixed effects increase the explained variation in wealth returns by 20 to 30 percent, controlling for such variables, consistent with the presence of non-negligible type dependence. The distribution of estimated fixed effects in returns to net wealth is plotted in Figure A1.

2.3 The Correlation between Persistent Heterogeneity in Wealth Returns and Persistent Heterogeneity in Earnings

Besides returns from capital, a key source of income for most households is labor earnings (income from employment). A large literature in labor economics suggests that the earnings process is best characterized by the sum of observable characteristics (such as education and labor market experience), as well as unobservable components capturing transitory shocks

[4]Estimates of risk aversion are based on PSID survey questions eliciting people's preferences for risk, see Kimball et al. (2009); however, these data are only available in two waves, 1996 and 2012.

Figure 2: Annualized gross returns to wealth across the gross wealth distribution. The figure plots the mean annual returns to gross wealth, in each percentile of gross wealth. Data is from the 1999-2019 waves of the PSID.

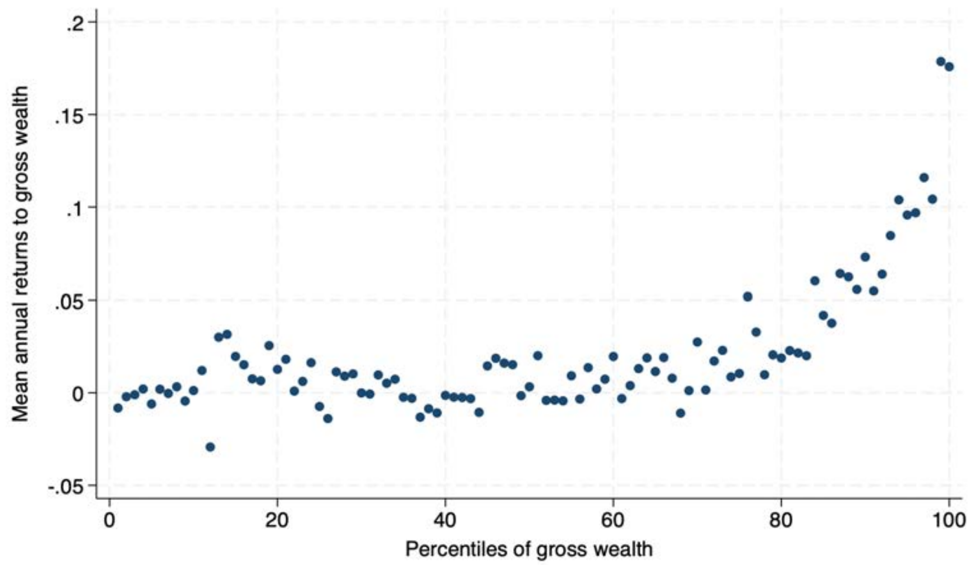


Table 1: Fixed effects in returns to wealth

	Whole sample		*Non-missing risk avs and transfers*			
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Shares*Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Risk aversion	No	No	No	Yes	Yes	No
Intergenerational transfers	No	No	No	No	Yes	No
Individual FE	No	Yes	No	No	No	Yes
Adj R squared	0.209	0.272	0.247	0.247	0.248	0.299
N	8274	8274	2566	2566	2566	2566

Notes: Dependent variable is net returns to wealth. Baseline controls include age, education, employment, year and state dummies, share of wealth allocated to different asset classes, leverage of mortgage and other debt, and wealth percentiles. Data is from the 1999-2019 waves of the PSID.

(due to movements in and out of work, periods of poor health, etc.) and more permanent traits of an individual (labor market skills, etc.). See for example, Gottschalk and Moffitt (2009) and Meghir and Pistaferri (2004).

Given the evidence of substantial persistent heterogeneity in returns to wealth documented in the previous section, it is important to understand whether it is associated with the widely documented persistent heterogeneity in earnings. To do so, we first extract the permanent component of earnings from a fixed effect regression of log earnings on observable characteristics, and then look at the correlation between the estimated fixed effect in earnings and the estimated fixed effect in household returns (as estimated in the previous section). Since wealth is measured at the household level, our measure of earnings is also at the household level (the sum of husband and wife earnings).[5]

Figure 3: Rank correlation between fixed effects in household wealth returns and fixed effects in log earnings. The figure plots the average percentile of net returns to wealth fixed effects by the percentile of the earnings fixed effects distribution. Data is from the 1999-2019 waves of the PSID.

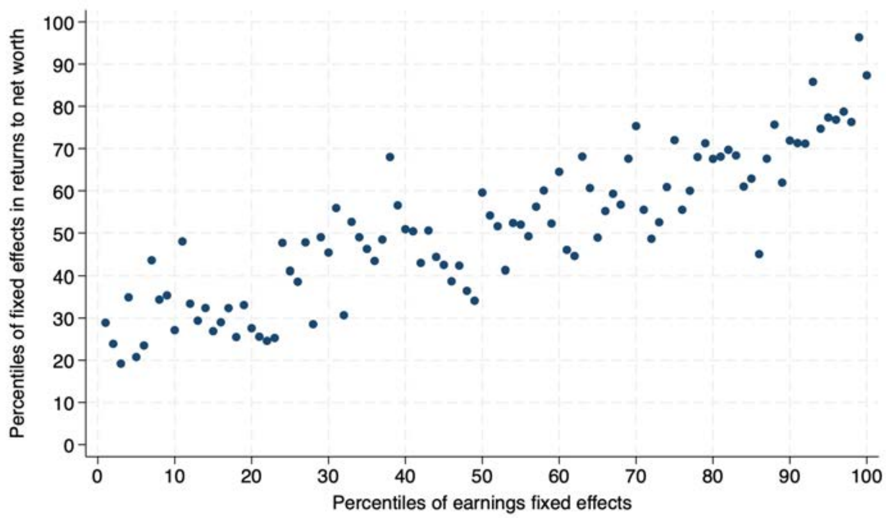


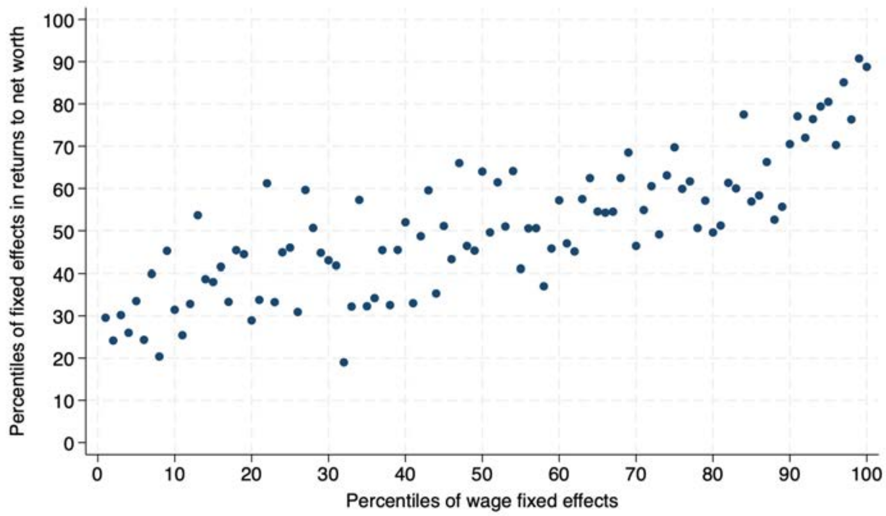
Figure 3 is a binscatter, plotting the average percentile of the distribution of wealth return fixed effects by percentile of the earnings fixed effect distribution. The Figure documents the existence of a positive rank correlation between fixed effects in household wealth returns and fixed effects in log household earnings.[6] To eliminate heterogeneity in household earnings coming from differences in the number of working members, in Figure 4 we replicate the same exercise for the log of hourly wages, defined as the ratio of household annual earnings and household annual hours.[7] The evidence from Figures 3 and 4 is similar both qualitatively and quantitatively. Figure 3, for example, shows that a move from the 10-th to the 90-th percentile of the earnings fixed effect distribution is associated with a fairly large increase in the average percentile of the wealth return fixed effect (a move from the 40-th to the 70-th percentile). The rank correlation is 0.502. The fact that the figures for earnings and hourly

[5]The permanent component in household earnings will also reflect any assortative mating that may characterize choices in the marriage market. The results are similar if we focus on the head's earnings. See the Online Appendix.

[6]When we conduct this exercise, we restrict the estimation sample for those individuals we observe for at least 5 periods in the sample period.

[7]This is equivalent to the weighted sum of individual hourly wages, with weights given by the share of total household hours worked by each member.

Figure 4: Rank correlation between fixed effects in household wealth returns and fixed effects in log wages. The figure plots the average percentile of net returns to wealth fixed effects by the percentile of the wages fixed effects distribution. Data is from the 1999-2019 waves of the PSID.



wages are quite similar suggest that labor supply choices are not driving the correlation. For simplicity, in the model below we assume that hours are chosen exogenously.

In Table 2 we present estimates of the rank correlation between persistent heterogeneity in returns and persistent heterogeneity in earnings for different subgroups. The correlation increases with education and is higher for households with a non-white head.

Table 2: Rank correlation between permanent component of wages and returns to wealth

	Total	Education		Race	
		High school or below	College	White	Other
	(1)	(2)	(3)	(4)	(5)
Total labor income					
$\rho(P_i, \psi_i)$	0.502	0.391	0.435	0.469	0.553
N	1002	310	692	807	194

Note: This table reports the rank correlation between fixed effects in returns to net wealth and fixed effects in total household labor income across sub-groups of the population. Data is from the 1999-2019 waves of the PSID.

What explains the correlation between earnings and returns fixed effects? The positive rank correlation between persistent heterogeneity in returns to wealth and persistent heterogeneity in earnings documented in Figure 4 and Table 2 could be induced by the presence of common cognitive and non-cognitive factors. Cognitive characteristics could be factors capturing the individual's ability or genetic endowment, driving both productivity on the job as well as more financially sophisticated portfolio choices. Non-cognitive factors potentially influencing both earnings and returns to wealth include, among other things, various personality traits, such as openness to experience, social skills, etc. An alternative explanation for the correlation that does not reflect cognitive or non-cognitive components is heterogeneity in risk tolerance, which naturally affects portfolio composition but may also lead workers to sort in industries or jobs that are safer (and hence command lower compensating wage differentials).

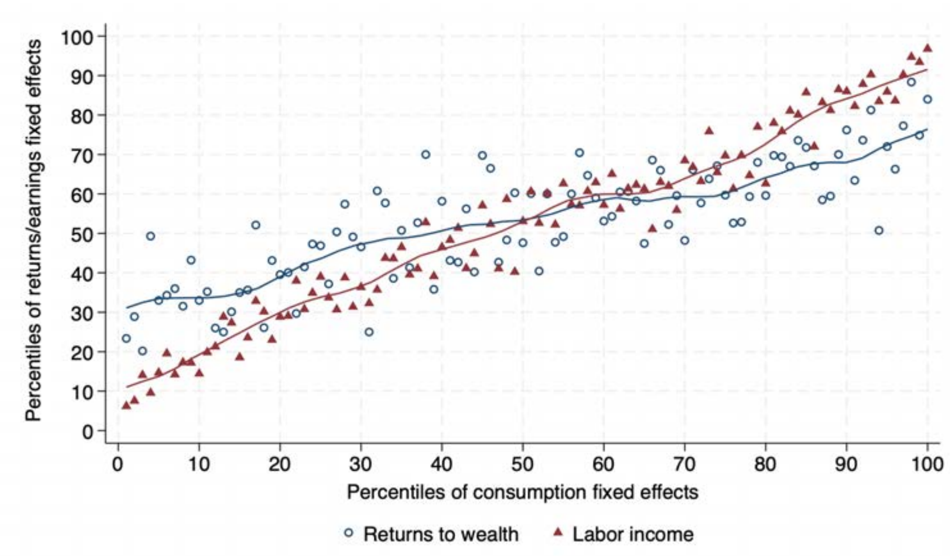
In Appendix B we use PSID data to estimate how much of the rank correlation between persistent heterogeneity in returns to wealth and persistent heterogeneity in earnings can be explained by heterogeneity in risk tolerance. A first measure of risk tolerance is based on questions asked in in the 1996 and 2012 waves that tried to directly elicit the respondents' degree of risk aversion. We complement this measure with two other variables that are often used as proxies for risk tolerance. The first is whether the individual belongs to a generation "scarred" by poorly performing stock markets. As argued by Malmendier and Nagel 2011, people who share a birth year may be subject, as children or adolescents, to epochal events (e.g., growing up during the Great Depression) that create "experience effects" and shape their beliefs and their risk-taking attitudes later in life. The second is whether people receive intergenerational transfers, as they (especially if received at an early stage of the life-cycle) may allow individuals to make different choices both in the labor market, financial markets, and potentially even in the marriage market. We show that these proxies explain roughly 15% of overall variation. We interpret this evidence as suggesting that the bulk of the correlation is driven by unobserved cognitive or non-cognitive skills rather than preference heterogeneity. Below we put some structure on the role of cognitive and non-cognitive skill endowment in the theoretical model.

2.4 Persistent Wealth Returns, Earnings, and Consumption

The final key fact we document in this Section is the relationship between fixed effects in labor income and wealth returns and persistent heterogeneity in household consumption. We first replicate the exercise described above for wealth returns and earnings and obtain an estimate of the fixed effects in household consumption; we then plot average fixed effects for earnings and wealth returns against the position in the distribution of consumption fixed effects. Figure 5 shows that households with persistently high consumption tend to be households with high permanent earnings or wealth returns, controlling for observables such as age and education.

The positive rank correlation between wealth returns and consumption is interesting because it indicates that persistent returns heterogeneity does not translate merely into heterogeneous gross saving rates, but it also matters for consumption decisions. However, the correlation could also be induced by the positive association between wealth returns and earnings documented in Figure 3. To explain the evidence, and explore the quantitative importance of wealth returns heterogeneity on consumption and wealth inequality, we rely on a quantitative life-cycle model of consumption and portfolio choice in which persistent returns heterogeneity is directly embedded.

Figure 5: Rank correlation between fixed effects in consumption and returns and labor income fixed effects. The figure plots the average percentile of net returns to wealth fixed effects (blue dots) and the average percentile of labor income fixed effects (red triangles) by the percentile of the consumption fixed effects distribution. Data is from the 1999-2019 waves of the PSID.



3 A life-cycle model with skills endowment

Motivated by the empirical evidence presented above, we develop a life-cycle model to explore the effects of persistent wealth returns (as well as their correlation with the permanent component of wages) on the level of consumption inequality. The stochastic life-cycle model builds on Deaton (1991), Carroll (1997), Attanasio et al. (1999) and Gourinchas and Parker (2002). We contribute to previous quantitative models that attempt to explain the extent of wealth inequality observed in the data (such as De Nardi and Fella, 2017) by introducing two sources of individual unobserved heterogeneity: (i) differences in the initial level of permanent labor income, and (ii) differences in skill endowments (the unobserved stock of cognitive and non-cognitive skills). In the model, individual unobserved skills affect both the household's returns to wealth as well as its labor income. Households face uncertainty with respect to financial assets and human capital returns, out-of-pocket medical expenditures, as well as length of life. The time unit is one year.

Heterogeneity by Education Besides the ex-post heterogeneity induced by the realization of income and asset returns shocks, we allow for (ex-ante) heterogeneity with respect to educational attainment. This is to rationalize the fact that different education groups face different labor income profiles over their working life cycle.

Length of life To model the uncertainty about the length of life, we denote as d_t the probability that the household is alive in period $t+1$, conditional on being alive in period t. Households live at most until age T.

Preferences The utility function is intertemporally separable. The period utility function is:

$$u(C_t; z_t) = q(z_t)\frac{\widetilde{C}_t^{1-\gamma}}{1-\gamma}$$

where C_t is consumption, $q(z_t)$ a function of demographic shifters z_t (number of adults and number of children in the household) to account for the evolution of households composition over the life-cycle, such that $\widetilde{C}_t = \frac{C_t}{q(z_t)}$ and C_t is total consumption.

Bequests If households die at age t, the remaining wealth, A_t, is left to their heirs. As in De Nardi (2004), households value bequests according to the bequest function $b(A_t) = \theta\frac{(A_t+k)^{1-\gamma}}{1-\gamma}$, where θ is the intensity of the bequest motive and k the parameter controlling the curvature of the bequest function. If $k > 0$, there is a form of non-homotheticity implying that bequests are luxury goods. This is an alternative explanation for wealth concentration (besides persistent return heterogeneity) often considered in the literature (see De Nardi and Fella, 2017).

Skills endowment Household i starts life with an endowment of (cognitive and non-cognitive) skills $\Phi_{i,0}$ which (in the baseline) remains constant over the life-cycle. The endowment of skills is heterogeneous in the population and it has variance σ_{Φ}^2.

A higher skill endowment plays two roles. First, it increases returns to wealth (e.g., by reducing the learning costs of identifying better investment opportunities or the probability of making financial mistakes). Second, it enters the process of human capital formation and hence labor earnings, described next.

Earnings During their working life cycle, households receive (gross) real labor earnings $Y_{i,t}$. We write log earnings of household i at time t as:

$$\log Y_{i,t} = X_{i,t}^{'}\beta^y + p_i^y + u_{i,t} + z_{i,t} \tag{2}$$

where $X_{i,t}$ are observed characteristics affecting wages (which we subsume in an education-specific cubic polynomial in age), p_i^y is a permanent individual component, $u_{i,t}$ an idiosyncratic transitory shock, and $z_{i,t}$ an AR(1) component:

$$z_{i,t} = \rho z_{i,t-1} + v_{i,t} \tag{3}$$

with innovation $v_{i,t}$. The transitory shock $u_{i,t}$ and the innovation to the persistent component $v_{i,t}$ have constant variances σ_u^2 and σ_v^2, respectively. When $\rho = 1$ we have the traditional earnings process considered in most of the literature.

The permanent individual component p_i^y depends on on the individual's skill endowment as well as an individual fixed effect ω_i^y (assumed to be independent of the individual's skill endowment), with mean zero and variance $\sigma_{\omega^y}^2$. We assume that the permanent component of earnings can be written as:

$$p_i^y = \omega_i^y + \delta^y \Phi_{i,0} \tag{4}$$

Portfolio choice and Asset returns Households allocate their wealth, $A_{i,t}$, between a riskless asset $B_{i,t}$ and a risky asset $S_{i,t}$. We assume costly collection and processing of the financial information needed to access the return from the risky asset. Since the access decision is made on a period basis, households pay a per-period fixed cost, κ, to hold the risky assets (as in, e.g., Fagereng et al., 2017). Moreover, we assume borrowing constraints ($B_{i,t} \geq 0$) and short-sale constraints ($S_{i,t} \geq 0$): hence, the share of risky assets, $\alpha_{i,t}^s =$

$S_{i,t}/A_{i,t}$, lies between zero and one. It depends, among other things, on the degree of risk aversion.

The return from a household's portfolio can then be written as:

$$r_{i,t}^{p} = r^{b} + \alpha_{i,t-1}^{s}(r_{i,t}^{s} - r^{b}) \tag{5}$$

where r^b is a constant return on bonds, and $r_{i,t}^s$ the individual return on stocks. We assume that the individual excess return on risky asset $(r_{i,t}^s - r^b)$ is the sum of a common component (the average risk premium), a persistent individual component, plus an idiosyncratic component:

$$r_{i,t}^{s} - r^{b} = \mu_S + \psi_i + \xi_{i,t}^{s} \tag{6}$$

where the $\xi_{i,t}^s$ are independently and identically distributed according to $\mathcal{N}(0, \sigma_S^2)$, and $\mu_S > 0$. We assume that the persistent individual component ψ_i is a linear function of the stock of cognitive/non-cognitive skills:

$$\psi_i = \omega^r + \delta^r \Phi_{i,0} \tag{7}$$

Combining (6) and (7), the excess return on the risky asset evolves according to:

$$r_{i,t}^{s} - r^{b} = (\mu_S + \omega^r) + \delta^r \Phi_{i,0} + \xi_{i,t}^{s} \tag{8}$$

Finally, following Fagereng et al. (2017), we allow for tail risk in the risky assets return distribution. In particular, we assume that the excess return on the risky asset is drawn from (8) with probability $(1 - p_{tail})$, and equals $(r_{tail} - r^b)$ with probability p_{tail}. We obtain estimates of the tail risk parameters p_{tail} and r_{tail} from Barro (2009).

One of the goals of the empirical analysis is to disentangle the contribution of the skill endowment variable $\Phi_{i,0}$ on persistent heterogeneity in earnings (equation (4)) and on persistent heterogeneity in wealth returns (equation (8)) using a factor structure approach. As in all factor structure models with a common unobservable component (in this case the skill endowment), in the empirical analysis we need to impose a normalization for identification purposes, and hence we will assume that the loading factor $\delta^r = 1$.

Motivated by the empirical evidence documented above, we allow for correlation between household labor income and returns to wealth, but assume that the correlation is driven entirely by their permanent components: $(Cov(r_{i,t}^s, \log Y_{i,t}) = Cov(\psi_i, p_i^y))$. In principle, some of the correlation could be driven by "hedging" motives (i.e., an association between idiosyncratic labor income shocks and idiosyncratic shocks to risky returns). However, previous studies have found only weak evidence regarding the correlation between wage shocks and returns from stocks.[8] We hence assume that shocks to earnings and shocks to stock returns are independent.

Taxes and Transfers Households face a progressive labor income tax schedule. We write the log of after-tax labor income as:

$$\log Y_{i,t}^{n} = (1-\tau)\log Y_{i,t} + \log \nu, \tag{9}$$

where we use the approximation to the US tax schedule proposed by Heathcote et al. (2017). Returns to wealth are taxed at a constant rate τ_c.

Following Hubbard et al. (1995), we allow for the presence of a consumption floor $\underline{c}$ through means-tested transfers:

$$TR_{i,t} = \max(0, \underline{c} + m_{i,t} - (1+r^p)A_{i,t} - Y_{i,t}^{n}) \tag{10}$$

Finally, we assume that households retire exogenously at age 65 and start drawing public pension benefits Y^p (following legislation, we assume that only 85% of social security benefits are subject to income taxation). Public pension benefits are computed as a non-linear function of household's lifetime earnings H, which evolve according to $H_{i,t+1} = \frac{Y_{i,t+1}+tH_{i,t}}{1+t}$ if $t < 65$ and remain constant after retirement ($H_{i,t+1} = H_{i,65}$, if $t \geq 65$), as in O'Dea (2018).

Medical expenses Households face uncertainty with respect to their out-of-pocket medical expenses m_t. We assume that the log of medical expenses of household i at time t can be written as:

$$\log m_{i,t} = m(H_{i,t}, t) + \epsilon_{i,t} \tag{11}$$

where the idiosyncratic component $\epsilon_{i,t}$ is modeled using a persistent-transitory decomposition, as in French and Jones (2004); the persistent component follows an AR(1) process with coefficient ρ^m and shock $\zeta_{i,t}$; as done with earnings, we assume that the transitory component of medical expenditure represents measurement error. Medical expenses reduce income available for consumption but produce no utility benefits *per se*. Since medical expenses are generally negligible in the early stages of the life-cycle, we follow Hubbard et al. (1995) and Lusardi et al. (2017), among others, and allow for risk of out-of-pocket medical expenses only after retirement.

[8] In particular, using data from the PSID, Cocco et al. (2005) do not reject the null of no correlation between labor income shocks and stock returns.

3.1 The households' optimization problem

Households choose consumption and the portfolio share of risky assets to maximize:

$$E_0 \left\{ \sum_{t=0}^{T} \beta^t \left[d_t u(C_t; z_t) + (1 - d_t) b(A_t) \right] \right\}$$

where $\beta < 1$ is the subjective discount factor. Before retirement, the dynamic budget constraint reads as:

$$A_{t+1} = \left(1 + r^p_{t+1}\right) A_t + Y^n_t - C_t + TR_t - \kappa \times \mathbb{1}(\alpha^s_t > 0) \tag{12}$$

and after retirement as:

$$A_{t+1} = \left(1 + r^p_{t+1}\right) A_t + Y^p - m_t - C_t + TR_t - \kappa \times \mathbb{1}(\alpha^s_t > 0) \tag{13}$$

State variables The dynamic optimization problem of the household is characterized by seven state variables: age (t), assets (A), cognitive/non-cognitive skill endowment (Φ), initial labor income (ω^y), history of persistent income shocks ($\{v_j\}_{j=1}^{t-1}$), average lifetime earnings H and the permanent medical expense shock (ζ).

The recursive optimization problem of the household is solved using a modification of the solution algorithm in Iskhakov et al. (2017), which combines continuous and discrete choices. Compared to the standard EGM proposed by Carroll (2006), the employed upper envelope algorithm disregards non-optimal consumption emerging due to the presence of kinks in the value function at the points of the state space where the household is indifferent between alternatives in discrete choice space. Details on the solution of the model are reported in Appendix C.1.

4 Model estimation

To estimate the model we adopt a standard two-step strategy (Gourinchas and Parker, 2002). In the first step, we calibrate some of parameters outside the model, either by estimating them directly from the data or using established estimates from the literature. These parameters include those that characterize the earnings process and the out-of-pocket medical expenses, the distribution of risky asset returns, the survival probabilities, the demographic shifters, the curvature of the bequest function, the labor income tax schedule, and the pension rules. In the second step we use by Simulated Method of Moments (SMM) to estimate the remaining structural parameters $\Lambda = (\beta, \gamma, \kappa, \widetilde{\theta}, \sigma^2_{\omega^y}, \sigma^2_{\Phi}, \delta^y)$, taking the set of first step parameters as given.[9] We estimate the per-period risky assets participation cost κ and the parameters characterizing the structure of covariance between fixed effects in wealth returns and earnings $(\sigma^2_{\omega^y}, \sigma^2_{\Phi}, \delta^y)$ separately by education group. Table 3 provides a summary of the model parameters and what sources we use to estimate them.

Table 3: Model Parameters

Parameter	Explanation	Source
$q(z)$	Equivalence scale	Scholtz et al. (2006)
k	Curvature utility from bequests	French (2005)
δ_t	Conditional survival probability	2014 CDC Life Tables
T	Terminal age	Set to 90
r^b	Average return safe assets	Avg. return on 3-month T-Bill, 1998-2018
p_{tail}, r_{tail}	Tail risk	Barro (2019)
τ^{SS}	SS tax rate	Avg. rate 1998-2018
τ^t	Capital tax rate	Avg. rate 1998-2018
$\bar{c}$	Consumption floor	2014 maximum SSI+TANF benefit
Y^p	Public pension benefits	2014 SS rules
β_y	Earnings: Cubic polynomial in age (separately by education)	First-stage PSID
ρ	AR(1) coefficient persistent earnings component	First-stage PSID
σ^2_v	Variance persistent earnings shocks	First-stage PSID
σ^2_u	Variance transitory earnings shocks (set to 0 in the model)	First-stage PSID
μ_s	Average excess return risky assets	First-stage PSID
σ^2_S	Variance excess return risky assets	First-stage PSID
$m(H,t)$	Out-of-pocket medical spending (obs.)	First-stage PSID
ρ^m	AR(1) coefficient persistent component, OOP medical spending	First-stage PSID
σ^2_ζ	Variance shocks to out-of-pocket medical spending	First-stage PSID
ν	Approximation progressive tax system, constant	First-stage PSID
τ	Approximation progressive tax system, before-tax earnings	First-stage PSID
β	Intertemporal discount factor	Model
γ	Coeff. of relative risk aversion	Model
$\widetilde{\theta}$	Marginal propensity to bequeath	Model
κ	Cost of stock market participation	Model
δ^y	Permanent earnings (load on skill endowment)	Model
$\sigma^2_{\omega^y}$	Variance initial draw permanent earnings	Model
σ^2_Φ	Variance stock of cognitive/non-cognitive skills	Model

4.1 First stage estimation

Labor income process Our measure of labor income is the sum of both spouses' labor earnings. In equation (2) we first need to estimate the education-specific cubic polynomial in age. To do so, we start by estimating the following fixed effect regression on the entire

[9] The parameter $\widetilde{\theta}$ is the marginal propensity to bequeath and is related to the intensity of the the bequest motive parameter θ through the relationship: $\theta = \frac{1}{\beta(1+r^b)}\left(\left(\frac{1}{\widetilde{\theta}} - 1\right)\frac{1}{1+r^b}\right)^{-\gamma}$. Since θ depends also on the values of β and γ, to increase the efficiency of the estimation algorithm we estimate $\widetilde{\theta}$ instead of θ. For a given set of parameter values, we can then easily back out θ from the estimated $\widetilde{\theta}$.

sample:

$$\log Y_{i,t} = \beta_0 + \sum_{j=1}^{3} \left(\beta_j age_{i,t}^j + \beta_{j+3} age_{i,t}^j * college_i \right) + \delta_i + \eta_{i,t} \tag{14}$$

We also add control for household composition (dummies for number of adults and kids), whether head and spouse are unemployed or retired, and time fixed effects to account for sources of heterogeneity that are outside the model. In equation (14) the education-specific shifter cannot be separately identified from the fixed effect δ_i. To isolate the former, we regress the predicted combined residual from (14) ($e_{i,t} = \delta_i + \eta_{i,t}$) on a dummy for whether the head has some college education:

$$e_{i,t} = \alpha_0 + \alpha_1 college_i + \nu_{i,t}$$

We then use the estimated age coefficients ($\beta_0, \beta_1, ...\beta_6$) and education shifters (α_0, α_1) to construct the life-cycle profiles of earnings used in the model (the β^y parameters in equation (2) and Table 3).[10] The estimated age profiles of earnings (Figure A2) show that more educated households face steeper growth of earnings over their working life cycle.

Identification and estimation of the parameters characterizing the stochastic component of the earnings process (the persistence parameter ρ and the variances of persistent and transitory shocks, σ_v^2 and σ_u^2, respectively) use the moment conditions derived by the specification of the process governing earnings (equations (2) and (3)). We use a GMM estimator with the identity matrix as weighting matrix.[11] The link between the earnings fixed effects and the stock of cognitive/non-cognitive skills is estimated structurally (see Section 4.2 below).

Overall, our earnings parameters estimates (reported in Table A2) fall in the range of values reported in the literature (see, e.g., Kaplan 2012). We estimate earnings shocks to be slightly less persistent for less educated households ($\rho \simeq 0.957$) than for households whose head attained at least some college education ($\rho \simeq 0.977$). In line with results from previous studies (e.g., Hubbard et al. 1995; Carroll and Samwick 1997), the variance of persistent shocks is higher among less educated households. Following Hubbard et al. (1995) and Scholz et al. (2006), among others, we assume that transitory variation in earnings reflect measurement error; we thus set the variance of transitory shocks in the model to 0.

[10] The estimated age coefficients and education shifters are reported in full in Table A1 of the Online Appendix.

[11] We use the residuals $\eta_{i,t}$ from the fixed effects earnings regression in (14) to construct these moment conditions. Following Hubbard et al. (1995), we exclude observations with very low earnings realizations (1% bottom winsorizing) when we estimate these uncertainty parameters.

Asset returns We set the annual real return of the risk-free asset to 0%, which matches the average inflation-adjusted on the "3-month treasury bill" during the period covered by our data (1998-2018), which we take from FRED. Since our empirical measure of wealth returns includes interests, dividends, capital gains/losses from stocks, but also capital income from owning business, real estate or pension/IRA funds, the excess return from risky assets is set to the average empirical risky returns (by education) estimated from PSID data. As shown in Table A3, more educated households obtain on average a higher excess return from risky investments ($\mu_s = 0.040$) compared to households whose head is less educated ($\mu_s = 0.018$), perhaps reflecting their higher financial sophistication (Lusardi and Mitchell, 2014). To estimate the variance of the risky assets returns, we use second order moments of residuals obtained projecting wealth returns onto a third order polynomial in age, wealth shares in different asset classes, wealth percentiles, individual fixed effects, as well as time fixed effects. We find shocks to wealth returns to be more volatile among more educated households ($\sigma_S^2 = 0.048$) than among less educated households ($\sigma_S^2 = 0.035$), suggesting that the former choose, on average, a portfolio of risky assets with a higher risk-return profile. These values for the variances of risky asset returns are slightly higher but similar in magnitude to those used in the literature focusing on stock returns (Cocco et al., 2005). We set the parameters characterising tail risk following Barro (2009). In particular, we assume that in each period there is a 2% probability that households would face a "disaster" event in which the realization of the return to risky assets is -40%.

Out-of-pocket medical expenses Our measure of out-of-pocket medical expenditures is given by the sum of out-of-pocket expenses for prescriptions, in-home medical care, doctors, outpatient surgery, dental bills, hospital bills and nursing home. To estimate the deterministic component of out-of-pocket medical expenditure, we run a fixed effect regression for the log of total out-of-pocket medical expenditures on a third order polynomial in age, controlling for household composition, spouses' labor market status and year fixed effects. Following De Nardi et al. (2010), we also allow medical expenses to vary with household permanent income (which we measure as the average observed household earnings before the age of 65). To estimate the parameters of the stochastic component of the medical expenses process, we consider the second order moments of the first-differenced residuals and estimate the parameters by GMM. We find that the persistent component of medical expense risk has an autoregression coefficient of 0.96 (see Table A4), which closely mirror what found in De Nardi et al. (2010). We estimate, however, slightly higher innovation variance of the persistent component ($\sigma_\zeta^2 = 0.067$), perhaps because we are looking at a more recent period characterized by a general increase in the probability of extreme health care expenses. These estimates imply non-negligible medical expense risk.

Social security and taxation We compute public pension benefits using the 2014 Social Security rules (in particular, the Primary Insurance Amount (PIA) at full retirement age). To account for the progressivity in the Social Security formula, we compute the PIA separately for each spouse. Because we consider a unitary framework and household's average earnings, for the computation of the PIA we assign to each spouse average lifetime earnings in proportion to the average earnings shares observed in the PSID (0.7 and 0.3 for first and second earner, respectively). Each spouse j's monthly pension benefits in household i are then obtained using the following formula:

$$Y^p_{j,i} = \begin{cases} 0.90\ H_{j,i,65} & \text{if } H_{j,i,65} \leq \$816 \\ 0.90\ \times \$816 + 0.32\,(H_{j,i,65} - \$816) & \text{if } \$4917 \leq H_{j,i,65} \leq \$816 \\ 0.90\ \times \$816 + 0.32\ \times \$4917 + 0.15\,(H_{j,i,65} - \$4917) & \text{if } H_{j,i,65} > \$4917 \end{cases}$$

The value of the annual consumption floor ($\bar{c}$) is set to $\$12,492$. This value is the sum of the 2014 maximum federal payment for SSI and the maximum benefit payable to a couple with one child under the TANF program (the only welfare cash programs serving the low-income population).

We parametrize the earnings tax schedule, including any EITC payments and child tax credits the household is entitled to, using the federal income tax schedules for the period 1998-2018. After computing after-tax earnings $Y^n_t = Y_t - tax^y_t$, we approximate the mapping from before- to after-tax labor income using the approach of Heathcote et al. (2017) described in Section 3, in particular estimating the parameters τ and ν in equation (9) with a simple OLS regression. The parameter estimates ($\tau = 0.0935$ and $\nu = 0.8770$) are then used to compute the after-tax earnings households receive in the model. We follow tax legislation and assume that during retirement households pay taxes on 85 percent of their Social Security benefits. Social security benefits are partly financed by a proportional Social Security tax of $\tau^{ss} = 6.2\%$ on the first $\$234,000$ of earnings. Capital income is instead taxed at a constant rate $\tau^r = 15\%$.

Other external parameters We compute the conditional survival probabilities d_t using the 2014 U.S. Centers for Diseases Control and Prevention (CDC) Life Tables. We set the terminal age T to 90 (i.e., $d_{90} = 0$). We parameterize equivalence scales in consumption as in Scholz et al. (2006), i.e., use a modified OECD equivalence scale normalized to that of a household with two adults and one child. Therefore, we define $q(z_t) =$

$\left((z_t^j + 0.7 z_t^\chi)/(2+0.7)\right)$, where z_t^j and z_t^χ are, respectively, the average number of adults and kids in the household at age t. We compute z_t^j and z_t^χ over the life-cycle using PSID data. We set the parameter k, which captures the curvature of the bequest function, to $\$1,040,000$, which corresponds approximately to the value of $\$500,000$ set by French (2005a) (after converting 1987 values to 2014 prices).

Initial conditions We model the behavior of households from age 25 onward. Initial assets are drawn from the empirical distribution observed in the PSID (by education group) for households aged 24-28 (as in French 2005a). The distribution of skills endowments and initial earnings fixed effects are random draws from two independent normal distributions with mean 0 and variances σ_Φ^2 and $\sigma_{y_i}^2$, separately for the two education levels, respectively.[12] We set the share of households in each education group to replicate the composition of the PSID data.

4.2 Remaining structural parameters estimation

We estimate the remaining structural parameters $\Lambda = (\beta, \gamma, \kappa, \widetilde{\theta}, \sigma_{\omega^y}^2, \sigma_\Phi^2, \delta^y)$ by a Simulated Method of Moments (SMM) approach. The structural parameters are chosen to minimize the (weighted) distance between target moments estimated in the data and the corresponding moments simulated by the economic model. Appendix C.2 presents technical details about the SMM approach.

Moments We target two sets of moments. The first describes the aggregate behavior of households over the life cycle with respect to wealth accumulation and risky assets participation: median wealth-to-income ratios and average risky assets participation rates in the age groups 25-34, 35-44, 45-54 and 55-64, separately for households with at least some college education and households with high school degree or less.[13] The second set of target moments is given by the empirical pairwise correlations between the fixed effects of consumption, returns to wealth, and labor income, separately by education group. Overall, we

[12]Because the variances σ_Φ^2 and $\sigma_{y_i}^2$ are estimated within the model (as described in Section 4.2), we adopt a simple recursive procedure. We start with randomly drawn values from two standard normal distributions; we then retrieve the distributions of skills endowments and initial earnings fixed effects to use in the model simulation by multiplying the standard normal by the square root of the variances, at each iteration of the structural estimation exercise.

[13]We obtain these moments using the estimated coefficients from two regressions. The first regression is a quantile regression for the wealth-income ratio against a third-order polynomial in the head's age, interactions with a dummy for the head having attained a college degree, dummies for the number of adults and kids, education dummies, and year fixed effects. The second regression is a linear probability model for risky assets participation against the same right hand side variables. We then take the predicted conditional median wealth-income ratio and the predicted risky asset participation by age group and education level.

target 22 distinct moments. One aspect worth stressing is that we do not directly target moments describing the behavior we wish to explain (e.g., the degree of scale dependence in wealth returns or the extent of consumption inequality over the life-cycle), since we want to use these moments to validate the model predictions against the data.

We use the model to construct the simulated counterpart of the empirical target moments. We simulate the behavior of 10,000 households over the life cycle starting from age 25. To simulate the behavior over the life-cycle, we take random draws from the education-specific earnings process and risky assets returns, out-of-pocket medical expenses, and mortality distributions.

Parameters identification The parameters are jointly identified, in the sense that variation in different target moments contribute to different extents to the identification of each parameter. Hence, there is no one-to-one mapping from moments to structural parameters. To get a metric of how variation in a given moment induces a change in the estimated parameters, we use the sensitivity measure proposed by Andrews et al. (2017). This is shown in Figures A3 and A4 in the Appendix. Standard arguments apply for the identification of the preference parameters and the cost of risky assets participation. The discount factor β is mostly pinned down by variation in the overall level of assets that households accumulate over the life-cycle. Identification of the degree of relative risk aversion γ is mostly driven by the age profile of risky assets participation, with higher γ increasing the relative probability that older households exit the market for risky assets, relative to younger households. An additional source of variation for the identification of γ comes from the heterogeneity in the assets profiles of college graduates and lower educated households: because the former face higher human capital risks, higher γ induces college graduates to accumulate more assets for precautionary savings motives, relative to lower educated households. The risky assets participation cost κ is mostly identified by variation in risky assets participation rates, especially early in the life cycle. Finally, the identification of the marginal propensity to bequeath $\widetilde{\theta}$ relies on the trajectory of assets later in life, with higher marginal values of leaving a bequest generating a steeper age profile for assets.

The implications for wealth and consumption inequality of persistent heterogeneity in returns to wealth and returns to human capital critically depend on the extent of (co-)variation in these measures and, therefore, the values of $\sigma^2_{\omega^y}$, σ^2_{Φ} and δ^y. To achieve identification of these parameters, we use the covariance structure that the model imposes on the fixed effects in wealth returns, earnings, and consumption.

In principle, one could exactly identify the variance-covariance structure of persistent heterogeneity by ignoring altogether consumption data and simply targeting the empirical variance-covariance matrix of earnings and wealth returns fixed effects (given by, respectively, $p_i^y = \omega_i^y + \delta^y \Phi_{i,0}$ and $\psi_i = \omega^r + \Phi_{i,0}$):

$$
\begin{aligned}
Var(p_i^y) &= \sigma^2_{\omega^y} + (\delta^y)^2 \sigma^2_\Phi \\
Var(\psi_i) &= \sigma^2_\Phi \\
Cov(\psi_i, p_i^y) &= \delta^y \sigma^2_\Phi .
\end{aligned}
$$

However, the fixed effects are themselves estimated quantities, and hence are likely to contain measurement error. We thus allow for measurement error and rewrite the fixed effects as:

$$
\begin{aligned}
p_i^y &= \omega_i^y + \delta^y \Phi_{i,0} + e_i^y \\
\psi_i &= \omega^r + \Phi_{i,0} + e_i^r
\end{aligned}
$$

Assuming that the measurement errors in the two equations are uncorrelated ($Cov(e_i^y, e_i^r) = 0$), the variance-covariance structure from above becomes:

$$
\begin{aligned}
Var(p_i^y) &= \sigma^2_{\omega^y} + (\delta^y)^2 \sigma^2_\Phi + Var(e_i^y) \\
Var(\psi_i) &= \sigma^2_\Phi + Var(e_i^r) \\
Cov(\psi_i, p_i^y) &= \delta^y \sigma^2_\Phi
\end{aligned}
$$

Without further restrictions, we cannot separately identify $\sigma^2_{\omega^y}$, σ^2_Φ, and δ^y (the economically relevant parameters) from the measurement error variances. To achieve identification we use the implications of the economic model for the behavior of consumption, in particular the covariances between consumption fixed effects and the fixed effects in wealth returns and labor earnings.

Unlike fixed effects in earnings or wealth returns, fixed effects in consumption do not have a closed form expression. Nevertheless, we can provide an intuition for how we achieve identification using an "auxiliary" equation approach (in the parlance of Indirect Inference). This approach provides guidance as to which moments are most likely to help identifying the parameters of interest. Let the auxiliary equation for (residual) consumption fixed effects be a function of earnings and wealth returns fixed effects:

$$
\widetilde{c}_i = \lambda^r (\omega^r + \Phi_{i,0}) + \lambda^y (\omega_i^y + \delta^y \Phi_{i,0}) + e_i^c
$$

where e_i^c is a measurement error in consumption, and λ^r and λ^y are auxiliary parameters measuring the relationship between the average level of individual consumption and (the economically relevant component of) fixed effects in returns to wealth and fixed effects in labor income, respectively. These auxiliary parameters are complicated function of the structural parameters of the model, including e.g., the elasticity of intertemporal substitution in consumption $1/\gamma$. Assuming that measurement errors in wealth returns and earnings fixed effects do not co-vary with consumption, the covariances between the consumption fixed effects and the fixed effects in wealth returns and labor earnings are, respectively:

$$Cov(\widetilde{c}_i, \psi_i) = (\lambda^r + \lambda^y \delta^y)\sigma^2_\Phi$$
$$Cov(\widetilde{c}_i, p^y_i) = (\lambda^r \delta^y + \lambda^y (\delta^y)^2)\sigma^2_\Phi + \lambda^y \sigma^2_{\omega^y}$$

The parameters $\sigma^2_{\omega^y}$, σ^2_Φ and δ^y are therefore just identified by $Cov(\psi_i, p^y_i)$, $Cov(\widetilde{c}_i, \psi_i)$ and $Cov(\widetilde{c}_i, p^y_i)$. To estimate the parameters characterising the covariance structure of persistent heterogeneity by education level, we minimize the distance (by education group) between the empirical correlations and their model-predicted counterparts.

Weighting of the moments We use the inverse of the diagonal elements of the bootstrapped variance-covariance matrix of the target moments as weighting matrix. We do not employ the optimal weighting matrix to avoid the small sample bias issues discussed by Altonji and Segal (1996).

4.3 Estimation results

Parameter estimates Table 4 reports estimates of the structural parameters. Our estimates for the preference and participation costs parameters are all statistically significant at conventional levels and similar to those typically found in the literature. For example, our estimate of the discount factor (0.978) is in the ballpark of estimates from e.g., Gourinchas and Parker (2002) and French (2005a). The coefficient of relative risk aversion (2.63) implies an elasticity of intertemporal substitution (EIS) in consumption which is close to the average EIS from micro studies ($0.3 - 0.4$) reported by Havránek (2015). Our estimate for the marginal propensity to bequeath (0.868) is also very close to that estimated by De Nardi et al. (2010) (0.88). Moving to parameters estimated separately by education, we find that the cost of participation in risky assets is higher among low education households (\$1974) than among those with college education (\$1197). This heterogeneity may reflect lower financial sophistication among the less educated, making it more costly for these households to invest in risky assets (as suggested in Jappelli and Padula 2015). Our estimates are similar to those estimated by Vissing-Jorgensen (2004), who reports a per-period financial market participation cost in the \$830 – \$1800 range (in 2014 dollars).

In the remainder of Table 4 we report, separately by education, estimates of the parameters capturing variation in initial earnings, the stock of cognitive and non-cognitive skills endowment, and the effect of the latter on permanent income. Skills have a larger impact on the permanent component of the highly educated. Low-educated households have lower dispersion in initial earnings and a slightly lower dispersion in skill endowments. Interestingly, our structural estimates for the dispersion in wealth returns fixed effects ($\sigma^2_\Phi = 0.0026$ and $\sigma^2_\Phi = 0.0013$ for higher and lower educated households, respectively) are similar to the variance of fixed effects in returns to net wealth estimated by Fagereng et al. (2020b) using Norway's administrative tax records (0.0036). All in all, our estimates are remarkably close to those reported in the literature even though they are obtained in a single, unifying framework.

Table 4: Estimated structural parameters

Parameter		Value	Std. error
Time discount factor	β	0.9795	(0.0006)
Coefficient of relative risk aversion	γ	2.7140	(0.0113)
Marginal propensity to bequeath	$\widetilde{\theta}$	0.8688	(0.0237)
Upper secondary education or less			
Risky assets markets participation cost	κ	1971.33	(30.207)
Variance skills FEs	$\sigma^2_{\Phi_i}$	0.0013	(0.00002)
Variance initial earnings	$\sigma^2_{y_i}$	0.0792	(0.0027)
Effect FEs skills on earnings	δ^y	3.0853	(0.1453)
Some college degree			
Risky assets markets participation cost	κ	1257.0	(177.831)
Variance skills FEs	$\sigma^2_{\Phi_i}$	0.0026	(0.0001)
Variance initial earnings	$\sigma^2_{y_i}$	0.0956	(0.0078)
Effect FEs skills on earnings	δ^y	3.1853	(0.1619)

Notes: The estimates are obtained using a minimum distance approach. We minimize the weighted distance between moments of actual and simulated data using the inverse of the diagonal of the bootstrapped variance-covariance matrix of the moments as a weighting matrix. The costs of risky assets participation are expressed in 2014 dollars. Asymptotic standard errors are reported in parentheses.

Sensitivity To examine which sources of variation in the data help the most with the identification of each structural parameter, we compute the sensitivity of the parameter estimates to changes in the target moments as proposed by Andrews et al. (2017). Consistent with the discussion above, Figures A3 and A4 show that variation in covariances between consumption, wealth returns and earnings fixed effects is key for pinning down the values of the variances of skills and earnings, as well as the effect of skills on earnings.

Model fit The model does a good job in replicating the moments we target in the structural estimation (the model is overidentified, so the good fit is not a mechanical feature). Figures A5 and A6 report a comparison between the wealth-to-income ratio and the risky assets participation rate, respectively, by age groups and education of the head. Similarly to what observed in the data, the model predicts higher wealth accumulation (and risky assets participation) among the more educated. Note that this is achieved without resorting to risk or time preference heterogeneity. While the overall fit is good, some moments fall out of the 95% confidence interval of the empirical moments. For example, the model tends to slightly underpredict the median asset accumulated by college graduates early in the life-cycle, while it overpredicts risky assets participation rates among the less educated.

In the data, lifetime consumption is strongly correlated not only with the permanent earnings component, but also with persistent heterogeneity in wealth returns, and the covariances are higher for the more educated households. As shown in Table 5, our model matches the empirical covariance structure (by education group) between consumption, wealth returns and earnings fixed effects. All model-predicted pairwise covariances fall within the 95% confidence intervals of the corresponding empirical moments.

Table 5: Goodness of fit - FE covariances

Target moments	Education level	Model	Data	[95% C.I. Diff.]	
$Cov(\psi_i, P_i^y)$	HS degree or less	0.3614*	0.4226	-0.099	0.221
	Some college +	0.4734*	0.5216	-0.036	0.133
$Cov(\widetilde{c}_i, \psi_i)$	HS degree or less	0.3823*	0.4755	-0.046	0.232
	Some college +	0.5660*	0.4973	-0.158	0.021
$Cov(\widetilde{c}_i, P_i^y)$	HS degree or less	0.7657*	0.7515	-0.081	0.053
	Some college +	0.7669*	0.7893	-0.021	0.066

Notes: The table reports a comparison between the covariances of fixed effects estimated in the PSID data and the corresponding moments estimated in the data simulated by the economic model. The 95% confidence intervals for the difference between data and simulations is also reported. *indicates simulated moment falls within the 95% confidence interval of the empirical moment.

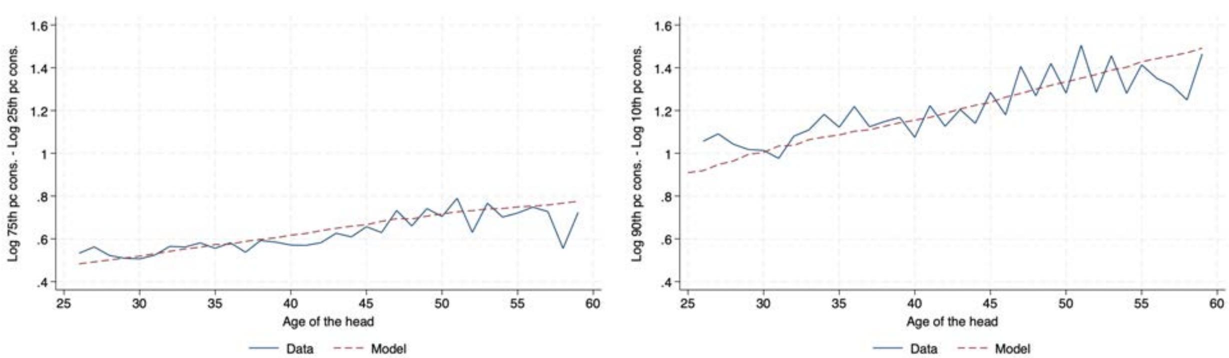
Model Validation We conduct several model validation exercises using moments that are not directly targeted in the structural estimation. Figure A7 is a first validation of the model. We plot median log consumption over the life cycle in the data and as predicted by the model. The model replicates extremely well both the level and the concave shape by age.

Figure 6 focuses on inequality in log consumption over the life cycle. We measure inequality with the interquartile range in panel (a) and with the 90th-10th percentile difference in panel (b). Relative to the variance (a popular measure of inequality used in the consumption literature, see Deaton and Paxson, 1994), these measures have the advantage of being more robust to the influence of observations in the tails of the distribution. In the data, inequality increases over most of the working life cycle. The model does a good job of replicating both the level and the growth of consumption inequality observed in the data (regardless of the inequality measure used).

Figure 6: Untargeted moments: Consumption inequality. The figure plots the life-cycle profiles of consumption inequality in the data (solid lines) and simulated by the economic model (dashed line). Panel (a) reports the 75-th/25-th percentile difference, while panel (b) reports the 90-th/10-th percentile difference in log consumption.

(a) 75th-25th percentile

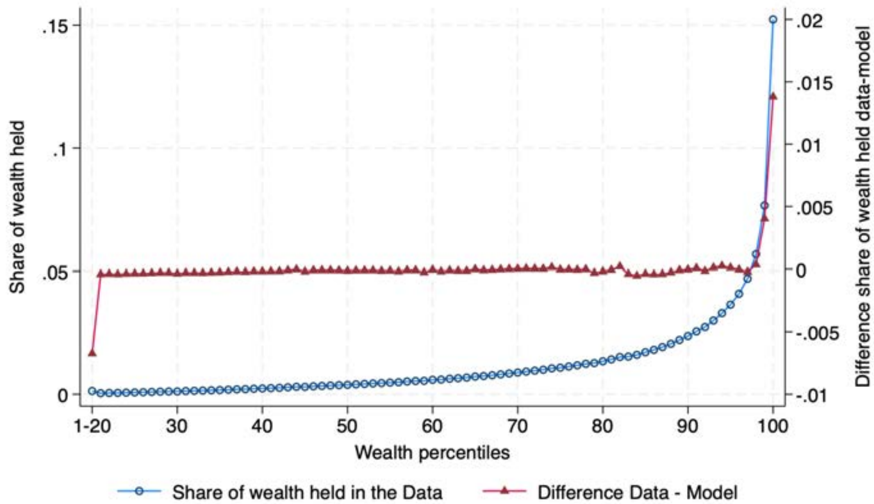
(b) 90th-10th percentile



Besides consumption inequality, the model replicates quite well the evolution of wealth inequality over the life-cycle (as measured again by the interquartile range and the 90th-10th percentile difference, see Figure A9 in the Appendix). Even more remarkably, but perhaps not surprisingly given the importance of return heterogeneity stressed by Benhabib et al. (2019), Figure 7 shows that the model is able to replicate very accurately the cross-sectional wealth shares at different percentiles of the wealth distribution, only slightly under-predicting the share of wealth held at the very top (by less than 1.5 percentage points) and at the very bottom (by roughly 0.5 percentage points).

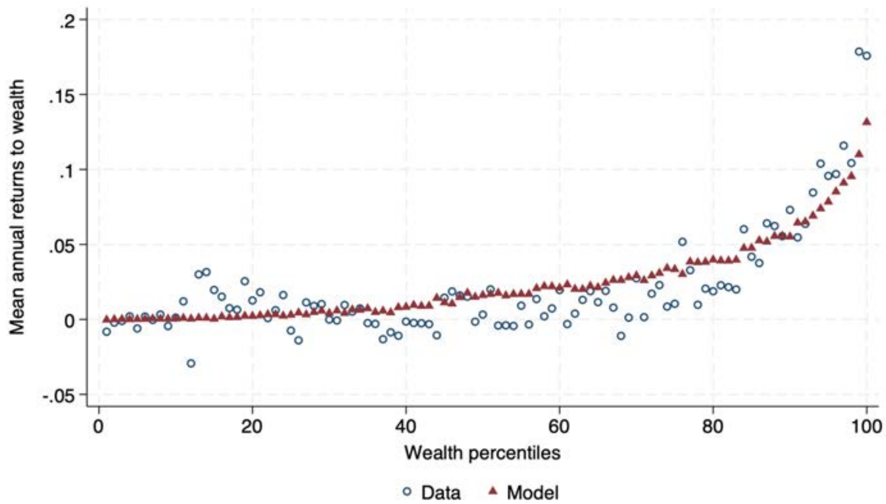
Finally, we investigate the model's ability to replicate an additional key feature of stochas-

Figure 7: Untargeted moments: Share of wealth held by households over the wealth distribution - Model vs. Data. The figure plots the average share of wealth held by households in each percentile of the wealth distribution in the 1999-2019 PSID data (blue dots and left axis). It also plots the difference between data and model that allows for skills heterogeneity (red triangles and right axis).



tic wealth returns remarked in theoretical work and empirical analyses: the degree of scale dependence in wealth returns. In our model, any degree of scale dependence will be endogenously generated by the portfolio and saving choices of the agents, rather than being an assumed feature of the stochastic process for wealth returns (as in Benhabib et al., 2019). Figure 8 shows that the model matches remarkably well some key features of the relation between wealth returns and a household's position in the wealth distribution. Indeed, the model predicts a similar degree of convexity in the relation between wealth returns and wealth percentile, while predicting wealth returns at the top of the wealth distribution that are similar to those observed in the data.

Figure 8: Untargeted moments: Wealth returns across the wealth distribution - Model vs. Data. The figure plots the average percentile of wealth returns fixed effects by the percentile of wealth distribution in the 1999-2019 PSID data (blue dots) and predicted by the model which allows for skills heterogeneity (red triangles).



5 Implications

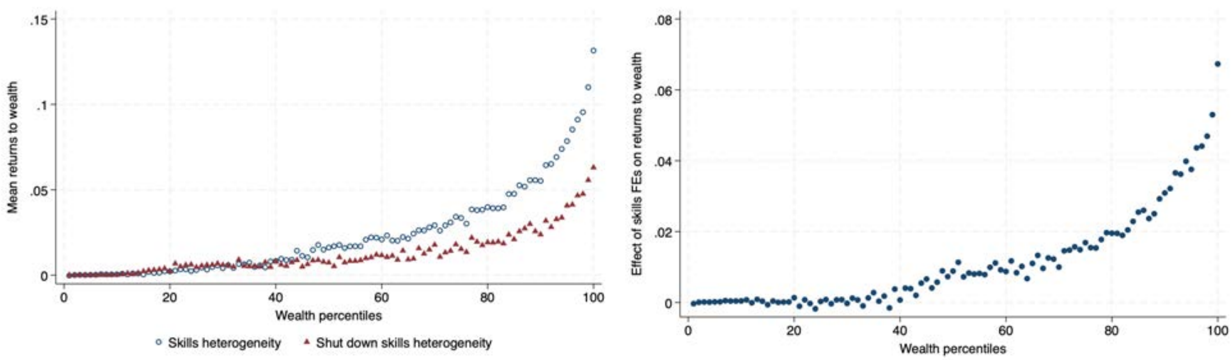
5.1 Scale dependence in wealth returns: risk-taking or skills?

One of the most intriguing questions in the recent literature on return heterogeneity is what mechanism is capable of generating the extent of type and scale dependence in wealth returns visible in the data. Our model assumes that type dependence arises from the presence of a persistent component capturing the endowment of cognitive and non-cognitive skills of an individual; this component appears statistically and economically significant. Scale dependence, however, is an endogeneous by-product of the saving and portfolio choices made by agents in our model. In particular, a traditional household finance explanation for scale dependence is risk-taking: because of the insurance provided by their wealth or because they can more easily overcome the cost of participating in stock markets, people at the top of the wealth distribution have an asset portfolio that is more geared towards risky assets, which helps explaining the higher wealth returns they generate. A different, not necessarily alternative explanation, is financial education or sophistication: wealthy individuals may have better analytical skills or knowledge of financial products (here captured by the stock of endowed cognitive and non-cognitive skills) or have the wealth to acquire the services of financial advisors, which may produce higher returns on average.[14]

Figure 9: The effect of type dependence in explaining scale dependence in returns to wealth. Panel (a) compares the actual return to wealth (blue dots) and the counterfactual return to wealth generated when we shut down the contribution of the stock of skills endowment (red triangles), over the actual wealth distribution. Panel (b) plots the difference between actual and counterfactual returns.

(a) Actual returns vs. shut down skills

(b) The effect of skills heterogeneity



How important are these two distinct forces? In Figure 9.a we plot the actual return to wealth observed in the data and a counterfactual return to wealth that would be generated in a setting in which the contribution of the stock of skills endowment to returns was eliminated (i.e., imposing $\delta^r = 0$ in equation (7)), against wealth distribution percentiles. Figure 9.b plots the implied difference between actual and counterfactual returns. If scale dependence was entirely attributable to risk-taking, the line in Figure 9.b would be flat at 0. Clearly, that is not the case; more importantly, cognitive and non-cognitive skill factors appear more important at the top of the wealth distribution. People in the top percentile of the wealth distribution have wealth returns that are almost 8 percentage points higher than those that would be predicted by a model in which portfolio choices were driven by standard portfolio allocation rules with homogeneous returns across households. The estimated model predicts that about half the wealth returns at the top of the wealth distributions are explained by skill heterogeneity (see Figure A8).

[14]One point to stress is that the "skills" we refer to here have little to do with those studied in finance, such as the ability of finance professionals to "beat" or "time" the market. More simply, in a population of heterogeneous, non-professional investors, our notion of "skills" may reflect some minimal financial education which helps people avoiding costly mistakes such as failing to diversify a portfolio between bonds and stocks, investing in specific securities as opposed to index funds, or selling "low" and buying "high" in response to short-term market gyrations; it may also reflect a denser network of friends or relatives providing advice on financial matters. Even the best investor in a cross-section of non-professional investors can earn lower-than-market average returns; the point is that she still does (persistently) better than the average non-professional investor.

5.2 The quantitative importance of skills on wealth and consumption inequality

We next use the estimated model to study the quantitative importance of persistent wealth returns heterogeneity – and its correlation with earnings – on wealth and consumption inequality. In Figure 10.a we plot consumption inequality (measured by the 90th-10th percentile difference of the log consumption distribution) over the life cycle under four different scenarios: (a) the baseline (solid blue line), which allows for the stock of skills endowment to affect both the permanent components of wealth returns and wages, and – as shown in Figure 6 – replicates the data well; (b) a counterfactual scenario in which we let the permanent component of wealth returns and the permanent component of earnings to be uncorrelated (the green dashed line) – a scenario obtained setting $\delta^y = 0$; (c) a counterfactual case in which we shut down persistent returns to wealth heterogeneity (the red dashed line) – $\delta^r = 0$ in equation (8); and (d) a final counterfactual scenario (the orange dotted line) in which we shut down skills heterogeneity altogether (i.e., we impose $\sigma_\Phi = 0$). These counterfactual exercises highlight the importance of heterogeneity in wealth returns for consumption inequality dynamics. The figure shows that without correlation in return and earnings FE's, there will be too little consumption inequality at the start of the life cycle, and its growth over the life cycle will be more moderate.

Figure 10: The role of skills heterogeneity on consumption inequality over the life-cycle. Panel (a) of the figure plots the life-cycle profiles of consumption inequality simulated by the full economic model (solid blue line), the restricted model where we set the permanent component of wealth returns and the permanent component of earnings to be uncorrelated (green dashed line), the model in which we shut down persistent returns to wealth heterogeneity (the red dashed line), and the model in which we shut down skill heterogeneity (orange dotted line). Panel (b) reports the model-implied share of consumption inequality explained by correlated fixed effects in wealth returns and earnings (green dashed line), wealth returns heterogeneity (red dashed line), and skill heterogeneity (orange dotted line).

(a) Consumption inequality

(b) Explained share of consumption inequality

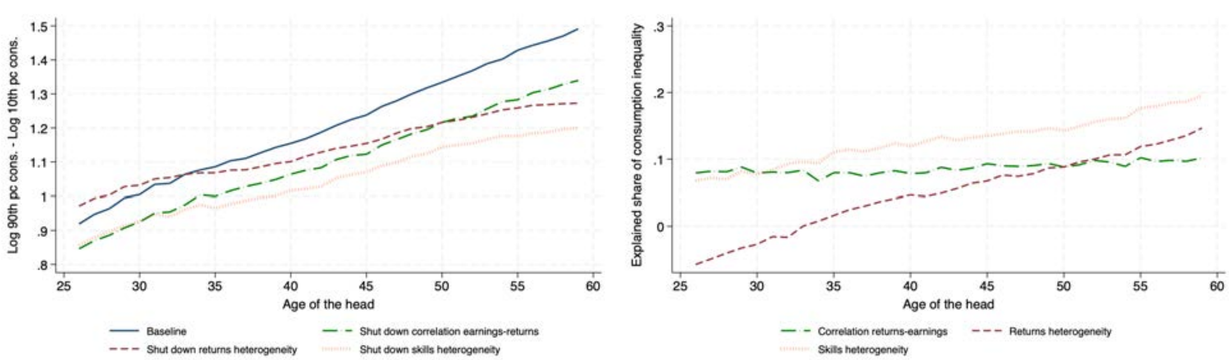


Figure 11: Persistent wealth returns heterogeneity and consumption age profile. The figure plots the age profiles of 90th and 10th percentiles of log consumption simulated by the estimated model (blue and red solid lines for 90th and 10th percentiles, respectively), and the restricted model where we shut down returns heterogeneity (blue and red dashed lines for 90th and 10th percentiles, respectively).

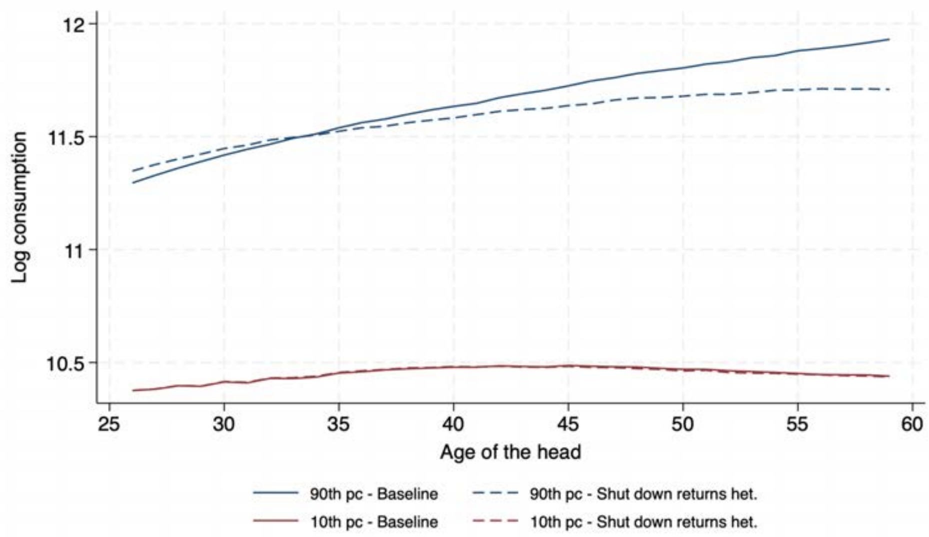
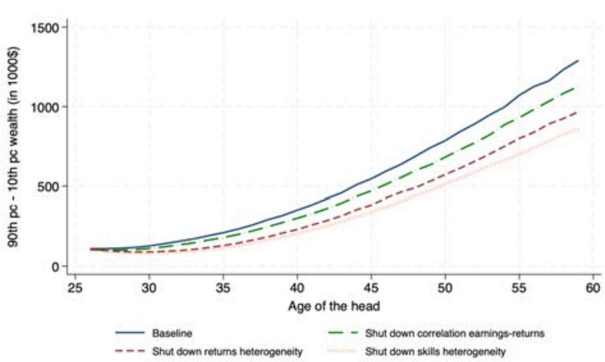


Figure 12: The role of skills heterogeneity on wealth inequality over the life-cycle. Panel (a) of the figure plots the life-cycle profiles of wealth inequality simulated by the full economic model (blue solid line), the restricted model where we set the permanent component of wealth returns and the permanent component of earnings to be uncorrelated (green dashed line), the model in which we shut down persistent returns to wealth heterogeneity (red dashed line), and the model in which we shut down skill heterogeneity (orange dotted line). Panel (b) reports the model-implied share of wealth inequality explained by correlated fixed effects in wealth returns and earnings (green dashed line), wealth returns heterogeneity (red dashed line), and skill heterogeneity (orange dotted line).

(a) Wealth inequality



(b) Explained share of wealth inequality

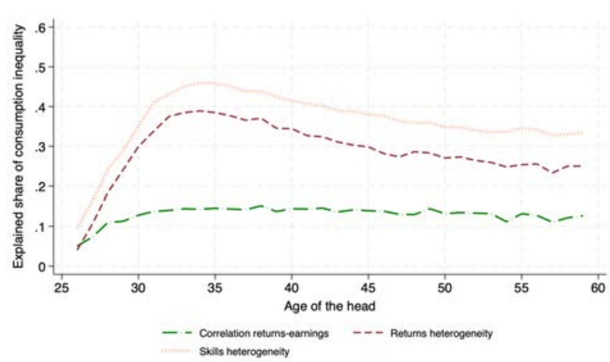


Figure 10.b shows that the correlation between wealth returns and earnings explain between 8% and 10% of consumption inequality over the life-cycle (as shown by the green

dashed line). Heterogeneity in returns is key for capturing the growth of consumption inequality over the life-cycle (as depicted by the red dashed line in Figure 10.b). Indeed, while wealth returns heterogeneity accounts for about one-sixth of consumption inequality before retirement, it reduces the level of consumption inequality early in the life-cycle. As shown in Figure 11, this is explained by persistent heterogeneity in returns inducing an upward tilting of the wealthy's consumption lifecycle profile, while the poorer households' consumption age profile is largely unaffected by persistent returns heterogeneity. Hence, since the Euler equation governs intertemporal allocation of consumption and savings, the model shows that persistent return heterogeneity induces a distribution of expected consumption growth rates in the cross-section: some households with high returns defer consumption to the later stage of their life cycle while others do the opposite. Through its joint contribution to generating heterogeneous wealth returns and additional heterogeneity in earnings, skills endowment explains between 8% and 20% of the observed consumption inequality. Figure 12 shows that heterogeneity in wealth returns explains around one-third of the observed degree of wealth inequality.

5.3 Understanding the mechanism: Wealth returns heterogeneity and saving behavior

We have shown that persistent wealth returns heterogeneity explains a substantial share of both wealth and consumption inequality over the life-cycle. However, the results we have presented so far are not necessarily informative about how heterogeneous wealth returns affect these measures of inequality. On the one hand, for given level of wealth, higher (lower) capital income leads households to accumulate more (less) wealth relative to the homogenous returns case, while keeping constant consumption and portfolio allocation decisions ("capital income effect"). On the other hand, households facing persistently different wealth returns can choose to make systematically different consumption and portfolio allocation choices ("policy functions effect", or behavioral responses).

To quantify the relative importance of these two mechanisms, we proceed in two steps. First, we consider the overall effect of heterogeneous wealth returns on the gross saving rate by computing the difference between actual and counterfactual saving rates in which we shut down return heterogeneity:[15]

$$SR^A\left(\phi(\delta^r=1), r^s_{i,t}(\delta^r=1)\right) - SR^C\left(\phi(\delta^r=0), r^s_{i,t}(\delta^r=0)\right)$$

[15]We define the gross saving rate as: $SR_t = \frac{Y^n_t + r^p_t A_{t-1} - c_t}{Y^n_t + r^p_t A_{t-1}}$.

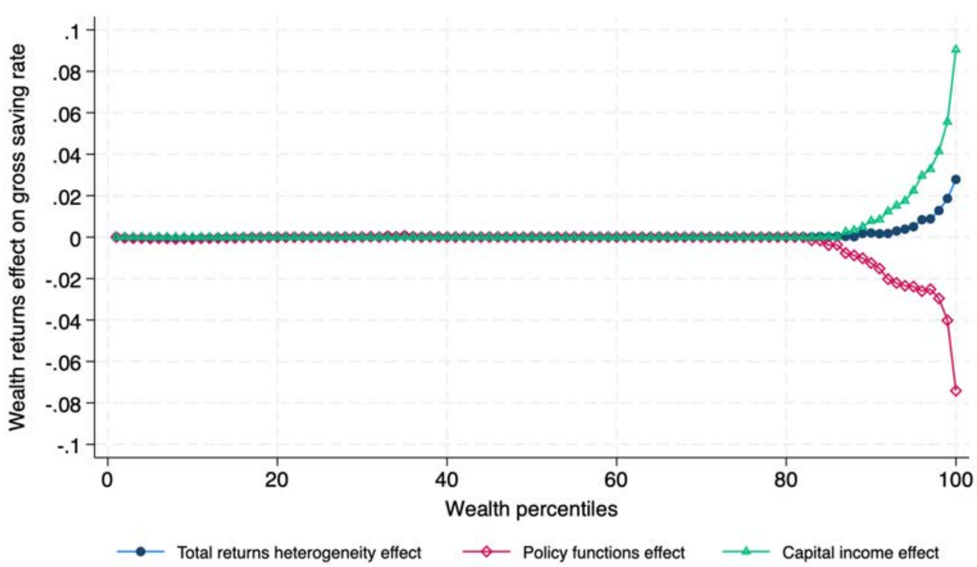
where $SR^A(.)$ is the actual saving rate as a function of the policy rules $\phi(\delta^r = 1)$ obtained in the presence of heterogeneous returns and the process of wealth returns that allows for individual fixed effects $r^s_{i,t}(\delta^r = 1)$, while $SR^C(.)$ is the counterfactual saving rate households choose when they do not anticipate heterogeneous persistent returns ($\phi(\delta^r = 0)$ and they face the standard process for wealth returns (i.e., $r^s_{i,t}(\delta^r = 0)$).[16]

Second, we decompose the overall effect of heterogeneous wealth returns on the saving rate between "capital income effect" and "policy functions effect" as follows:

$$SR^A\left(\phi(1), r^s_{i,t}(1)\right) - SR^C\left(\phi(0), r^s_{i,t}(0)\right) =$$
$$\underbrace{\left[SR^A\left(\phi(1), r^s_{i,t}(1)\right) - SR^C\left(\phi(0), r^s_{i,t}(1)\right)\right]}_{\text{Policy functions effect}} + \underbrace{\left[(SR^C\left(\phi(0), r^s_{i,t}(1)\right) - SR^C\left(\phi(0), r^s_{i,t}(0)\right)\right]}_{\text{Capital income effect}}$$

where $SR^C\left(\phi(0), r^s_{i,t}(1)\right)$ is the counterfactual saving rate obtained with persistently heterogeneous wealth returns but using policy rules obtained under common excess returns from risky assets.

Figure 13: Decomposition of the effect of wealth returns heterogeneity on the gross saving rate across the wealth distribution. The blue dots show the total returns heterogeneity effect on the gross saving rates, that is the difference between the actual saving rate and the counterfactual saving rate obtained shutting down wealth returns heterogeneity. The green triangle show the capital income effect, that is the effect on saving rates explained by differential capital incomes, conditional on policy functions, while the red diamonds show the policy functions effect that is the effect on saving rates due to the behavioral response of households to heterogeneous returns, conditional on actual heterogeneous wealth returns.



[16]The policy rules are the optimal solution to the dynamic optimization problem of the consumer.

Figure 13 depicts the overall effect (blue dots), the capital income effect (green triangles) and the policy functions effect (red diamonds) of returns heterogeneity on the gross saving rate, over the wealth distribution. First, the model shows that wealth returns heterogeneity affects gross saving rates mostly at the top of the wealth distribution (especially above the 80-th wealth percentile). Heterogeneous wealth returns induce households who end up at the very top of the wealth distribution to save annually about 3% more of their total income than in the case of homogeneous excess returns from risky assets. This overall effect of wealth returns on savings is the result of two opposite forces. Assuming all households expect to face common excess returns from risky assets, saving rates at the top of the wealth distribution are about 9 percentage points higher because of wealth returns heterogeneity. However, because households who end up at the top of the wealth distribution do anticipate to earn persistently higher returns from their investments, they choose to consume substantially more than they would were they facing the common excess return from risky assets. The model-predicted policy functions effect is important: at the top of the wealth distribution, households save about 7.5% less because of persistent wealth returns heterogeneity.

A first remark is that, absent behavioral responses, consumption inequality would have been substantially understated: the extra saving of people at the top of the wealth distribution due to return heterogeneity would feed primarily into higher bequests. A second interesting observation related to this analysis is that most of the behavioral responses are concentrated among the top 20% of the wealth distribution.

6 Conclusions

Most papers in the consumption literature that take a life-cycle perspective focus on the importance of *shocks* to permanent labor income and to capital income, rather than on heterogeneity in initial conditions that remain fixed over the life cycle. Persistent heterogeneity in returns matter for consumption inequality because it has dynamic effects playing through the intertemporal budget constraint.

In this paper we modify a standard life cycle model to allow for the possibility that an initial endowment of individual cognitive/non-cognitive skills (shaped by unobserved genetic components, human or perhaps social capital accumulation) drive both persistent differences in labor earnings as well as persistent differences in return to capital (a feature recently highlighted by several theoretical and empirical contributions). Thus, the persistent component of wealth returns and the persistent component of labor earnings are potentially correlated. We find that these variants are important. First, consumption inequality over the life cycle would be lower and grow more slowly if persistent heterogeneity in returns was ignored or assumed independent of persistent heterogeneity in earnings. This is because persistent return heterogeneity tilts the consumption profile of wealthy individuals by inducing more deferral of consumption relative to a world where returns are homogeneous. Second, the model would have a hard time explaining excess wealth returns at the top of the wealth distribution. Finally, we show that scale dependence in returns (a positive correlation between wealth and returns to wealth) is partly due to the risk composition of one's portfolio and partly to unobserved skills (reflecting, for example, financial sophistication or social network effects reducing the cost of collecting and processing financial information). The latter grows in importance with the position in the wealth distribution.

The findings of this paper have implications for the debate on the optimal design of capital and income taxation when policy-makers have to consider the trade-off between distributional issues and growth. Future research may be directed at evaluating the welfare implications of changes in capital income taxation in general equilibrium, something that is beyond the scope of this paper.

References

Altonji, J. G. and L. M. Segal (1996): "Small-sample bias in GMM estimation of covariance structures," *Journal of Business & Economic Statistics*, 14, 353–366.

Andrews, I., M. Gentzkow, and J. M. Shapiro (2017): "Measuring the sensitivity of parameter estimates to estimation moments," *The Quarterly Journal of Economics*, 132, 1553–1592.

Attanasio, O. P., J. Banks, C. Meghir, and G. Weber (1999): "Humps and Bumps in Lifetime Consumption," *Journal of Business & Economic Statistics*, 17, 22–35.

Bach, L., L. E. Calvet, and P. Sodini (2020): "Rich Pickings? Risk, Return, and Skill in Household Wealth," *American Economic Review*, 110, 2703–47.

Barro, R. J. (2009): "Rare disasters, asset prices, and welfare costs," *American Economic Review*, 99, 243–264.

Barth, D., N. W. Papageorge, K. Thom, and M. Velásquez-Giraldo (2023): "Genetic Endowments, Income Dynamics, and Wealth Accumulation Over the Lifecycle," *Unpublished manuscript.*

Benhabib, J. and A. Bisin (2018): "Skewed Wealth Distributions: Theory and Empirics," *Journal of Economic Literature*, 56, 1261–91.

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

Bricker, J., A. Henriques, and P. Hansen (2018): "How Much has Wealth Concentration Grown in the United States? A Re-Examination of Data from 2001-2013," *Finance and Economics Discussion Series.*

Carroll, C. D. (1997): "Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis," *The Quarterly Journal of Economics*, 112, 1–55.

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

Carroll, C. D. and A. A. Samwick (1997): "The nature of precautionary wealth," *Journal of monetary Economics*, 40, 41–71.

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

De Nardi, M. (2004): "Wealth Inequality and Intergenerational Links," *Review of Economic Studies*, 71, 743–768.

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

De Nardi, M., E. French, and J. B. Jones (2010): "Why do the elderly save? The role of medical expenses," *Journal of political economy*, 118, 39–75.

Deaton, A. (1991): "Saving and Liquidity Constraints," *Econometrica*, 59, 1221–48.

Deaton, A. and C. Paxson (1994): "Intertemporal Choice and Inequality," *Journal of Political Economy*, 102, 437–467.

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

Druedahl, J. and T. H. Jørgensen (2017): "A general Endogenous Grid Method for Multi-dimensional Models with Non-convexities and Constraints," *Journal of Economic Dynamics and Control*, 74, 87–107.

Fagereng, A., C. Gottlieb, and L. Guiso (2017): "Asset market participation and portfolio choice over the life-cycle," *The Journal of Finance*, 72, 705–750.

Fagereng, A., L. Guiso, M. B. Holm, and L. Pistaferri (2020a): "K-Returns to Education," Tech. rep.

Fagereng, A., L. Guiso, D. Malacrino, and L. Pistaferri (2020b): "Heterogeneity and Persistence in Returns to Wealth," *Econometrica*, 88, 115–170.

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

——— (2005b): "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies*, 72, 395–427.

French, E. and J. B. Jones (2004): "On the distribution and dynamics of health care costs," *Journal of Applied Econometrics*, 19, 705–721.

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016): "The Dynamics of Inequality," *Econometrica*, 84, 2071–2111.

Gottschalk, P. and R. Moffitt (2009): "The Rising Instability of U.S. Earnings," *Journal of Economic Perspectives*, 23, 3–24.

Gourinchas, P.-O. and J. A. Parker (2002): "Consumption Over the Life Cycle," *Econometrica*, 70, 47–89.

Guvenen, F. (2009): "An empirical investigation of labor income processes," *Review of Economic Dynamics*, 12, 58–79.

Havránek, T. (2015): "Measuring intertemporal substitution: The importance of method choices and selective reporting," *Journal of the European Economic Association*, 13, 1180–1204.

Heathcote, J., K. Storesletten, and G. Violante (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132, 1693—1754.

Hubbard, R. G., J. Skinner, and S. P. Zeldes (1995): "Precautionary saving and social insurance," *Journal of political Economy*, 103, 360–399.

Iskhakov, F., T. H. Jørgensen, J. Rust, and B. Schjerning (2017): "The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks," *Quantitative Economics*, 8, 317–365.

Jappelli, T. and M. Padula (2013): "Investment in financial literacy and saving decisions," *Journal of Banking & Finance*, 37, 2779–2792.

——— (2015): "Investment in financial literacy, social security, and portfolio choice," *Journal of Pension Economics & Finance*, 14, 369–411.

Kaplan, G. (2012): "Inequality and the life cycle," *Quantitative Economics*, 3, 471–525.

Kimball, M. S., C. R. Sahm, and M. D. Shapiro (2009): "Risk Preferences in the PSID: Individual Imputations and Family Covariation," *American Economic Review*, 99, 363–68.

Kirpatrick, S., C. D. Gelet, and M. P. Vecchi (1983): "Optimization by Simulated Annealing," *Science*, 220, 621–630.

Lusardi, A., P.-C. Michaud, and O. S. Mitchell (2017): "Optimal financial knowledge and wealth inequality," *Journal of Political Economy*, 125, 431–477.

LUSARDI, A. AND O. S. MITCHELL (2014): "The economic importance of financial literacy: Theory and evidence," *American Economic Journal: Journal of Economic Literature*, 52, 5–44.

MALMENDIER, U. AND S. NAGEL (2011): "Depression babies: Do macroeconomic experiences affect risk taking?" *The quarterly journal of economics*, 126, 373–416.

MEGHIR, C. AND L. PISTAFERRI (2004): "Income variance dynamics and heterogeneity," *Econometrica*, 72, 1–32.

MICHAILLAT, P. AND E. SAEZ (2021): "Resolving New Keynesian Anomalies with Wealth in the Utility Function," *The Review of Economics and Statistics*, 103, 197–215.

O'DEA, C. (2018): "Insurance, Efficiency and the Design of Public Pensions," .

SAEZ, E. AND G. ZUCMAN (2016): "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," *The Quarterly Journal of Economics*, 131, 519–578.

SCHOLZ, J. K., A. SESHADRI, AND S. KHITATRAKUN (2006): "Are Americans Saving "Optimally" for Retirement?" *Journal of Political Economy*, 114, 607–643.

SMITH, M., O. ZIDAR, AND E. ZWICK (2022): "Top Wealth in America: New Estimates Under Heterogeneous Returns*," *The Quarterly Journal of Economics*, 138, 515–573.

TAUCHEN, G. (1986): "Finite State Markov-chain Approximations to Univariate and Vector Autoregressions," *Economics Letters*, 20, 177–181.

VISSING-JORGENSEN, A. (2004): "Perspectives on Behavioral Finance: Does Irrationality Disappear with Wealth? Evidence from Expectations and Actions," in *NBER Macroeconomics Annual 2003, Volume 18*, National Bureau of Economic Research, Inc, NBER Chapters, 139–208.

Online Appendix

Appendix A reports the results of additional analyses complementing those in the main text. Appendix B describes how we provide descriptive evidence for the importance of heterogeneity in risk tolerance to explain the correlation between earnings and returns fixed effects. Additional details on the model solution and estimation are described in Appendix C.

Appendix A Additional results

Figure A1: The distribution of estimated fixed effects in returns to net wealth. The Figure reports the histogram of the estimated fixed effects in the net returns to wealth regressions (Column 2 of Table 1). Data is from the 1999-2019 waves of the PSID.

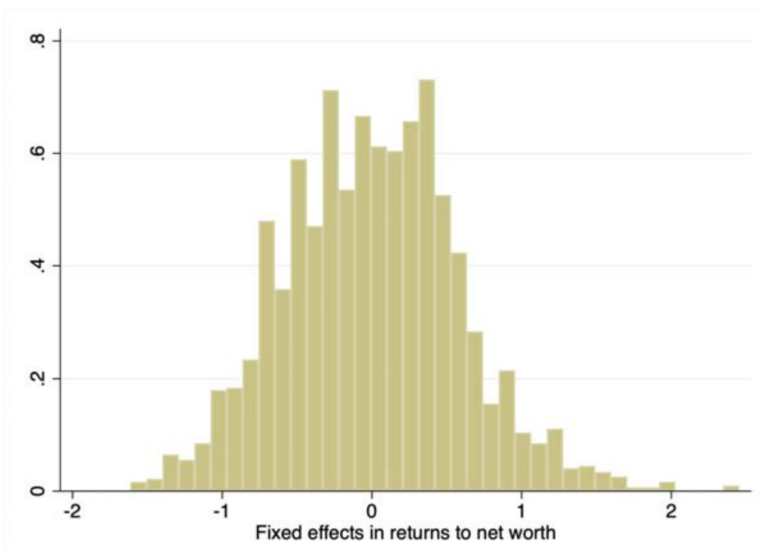


Figure A2: Estimated life-cycle profiles of earnings by education. The figure reports the age profiles of earnings by level of education of the head, obtained using the estimates in Table A1.

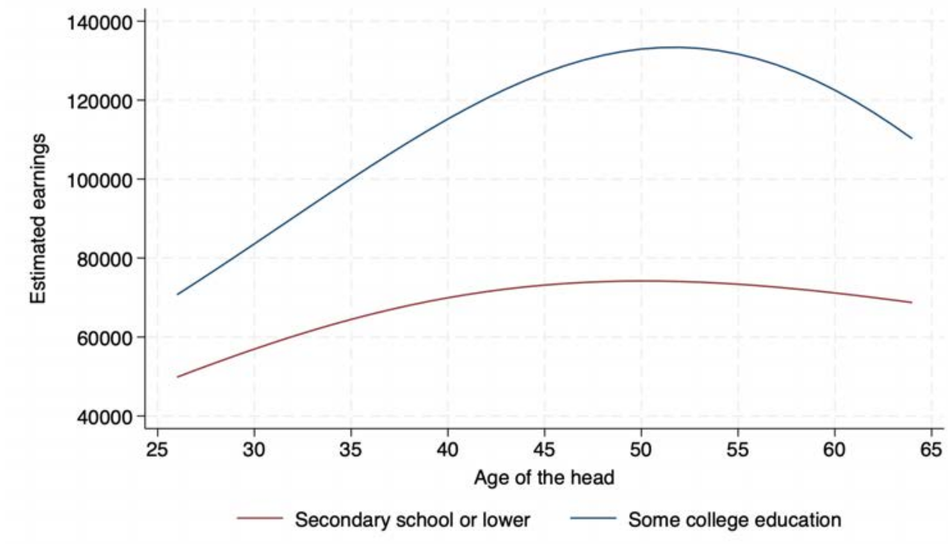
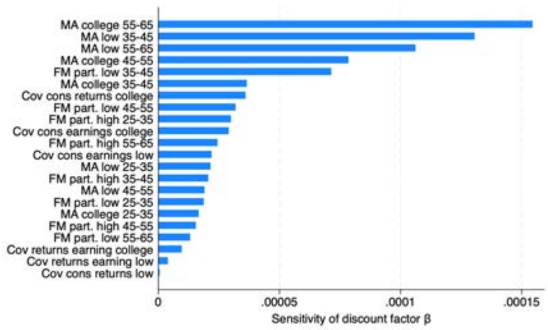
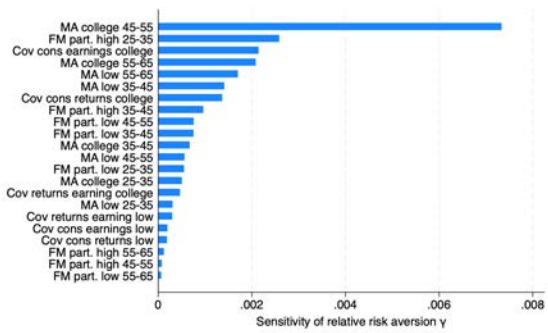


Figure A3: Sensitivity measure for common parameters and for Low Education-specific parameters. This Figure reports the absolute value of the scaled sensitivity matrix as defined in Andrews et al. (2017). The sensitivity measure has been rescaled to indicate the effect of a 1% increase in the moments on the parameters. “FM Part.” is financial market participation; “MA” is median assets. “College” denotes people with at least some college; “Low” denotes people with at most high school education.

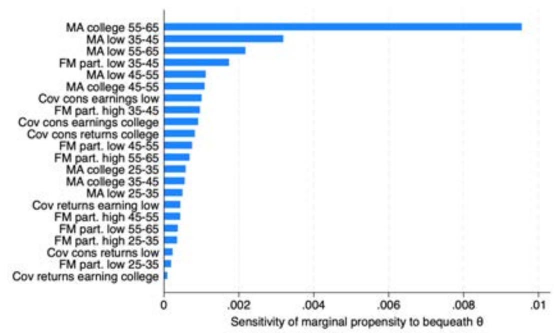
(a) β



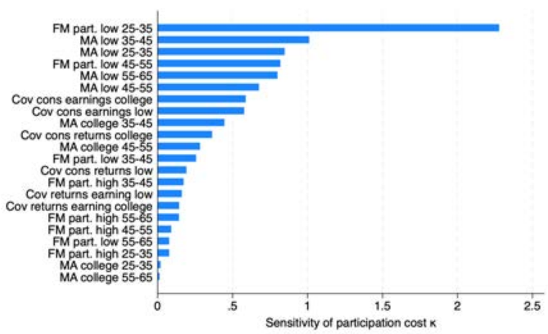
(b) γ



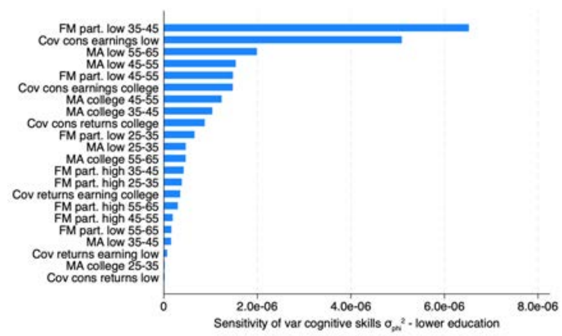
(c) $\widetilde{\theta}$



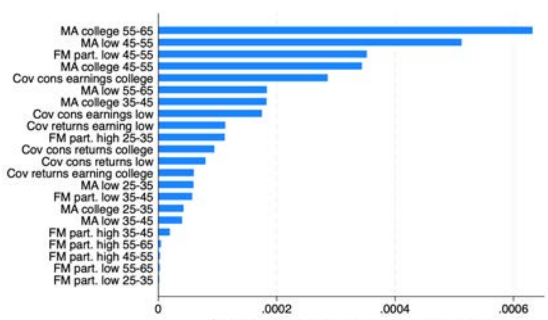
(d) κ - lower education



(e) σ^2_Φ - lower education



(f) $\sigma^2_{\omega^y}$ - lower education



(g) δ^y - lower education

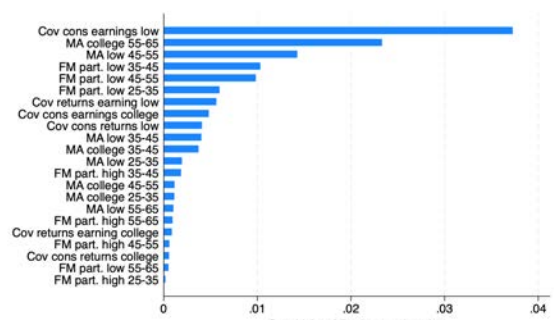
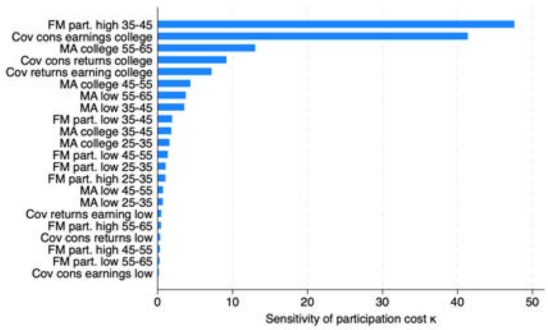
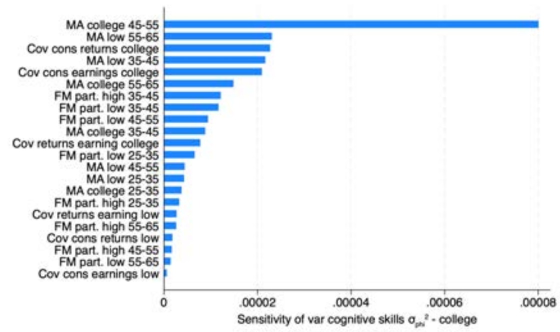


Figure A4: Sensitivity measure for High Education-specific parameters. This Figure reports the absolute value of the scaled sensitivity matrix as defined in Andrews et al. (2017). The sensitivity measure has been rescaled to indicate the effect of a 1% increase in the moments on the parameters. "FM Part." is financial market participation; "MA" is median assets. "College" denotes people with at least some college; "Low" denotes people with at most high school education.

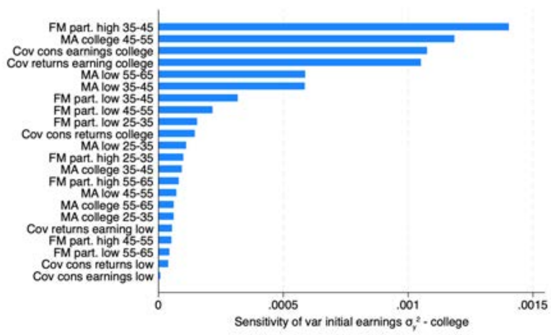
(a) κ - college



(b) σ^2_Φ - college



(c) $\sigma^2_{\omega^y}$ - college



(d) δ^y - college

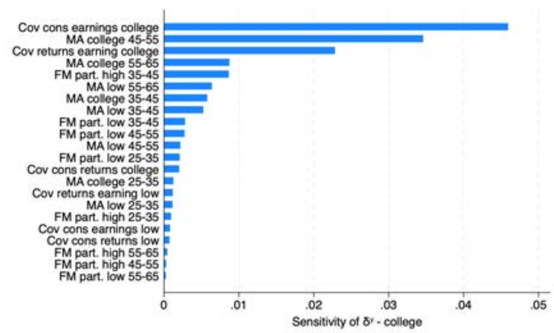


Figure A5: Model fit of assets. The figure reports the median wealth-to-income ratio by age group in the data (point estimates and 95% confidence intervals) and simulated by the economic model (red triangles) for households whose head attained upper secondary education degree or lower (panel a) and attained some college education (panel b). Empirical moments and bootstrapped standard errors are obtained as described in Section 4.2.

(a) Upper secondary or lower (b) Some college education

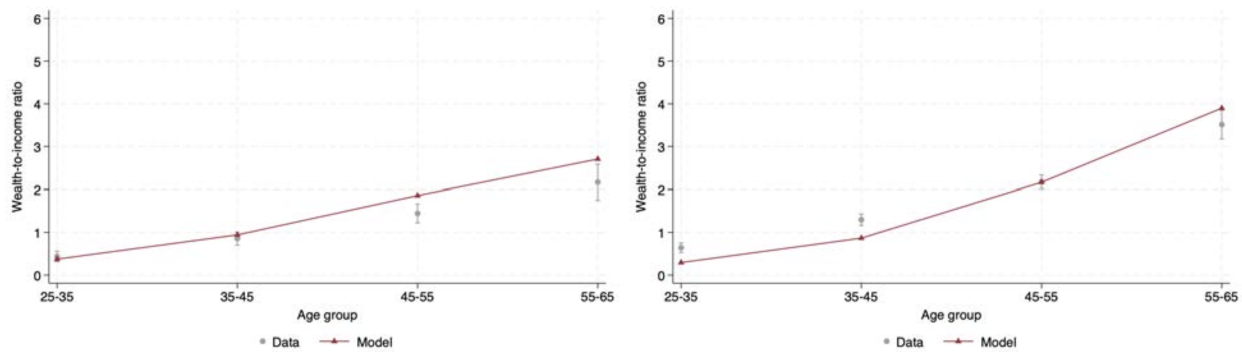


Figure A6: Model fit of risky assets participation. The figure reports the risky assets participation rates by age group in the data (point estimates and 95% confidence intervals) and simulated by the economic model (red triangles) for households whose head attained upper secondary education degree or lower (panel a) and attained some college education (panel b). Empirical moments and bootstrapped standard errors are obtained as described in Section 4.2

(a) Upper secondary or lower (b) Some college education

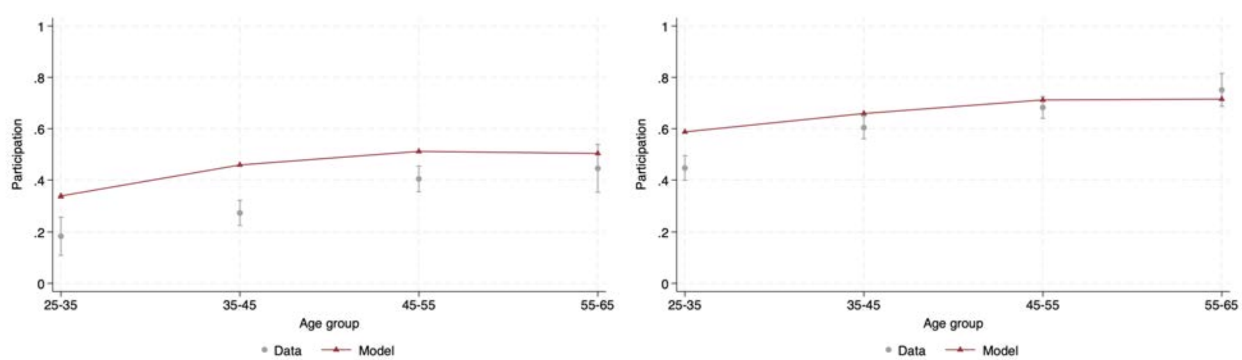


Figure A7: Untargeted moments: Median consumption by age - Model vs. Data. The figure reports median consumption by age group in the data (point estimates and 95% confidence intervals) and simulated by the economic model (red triangles).

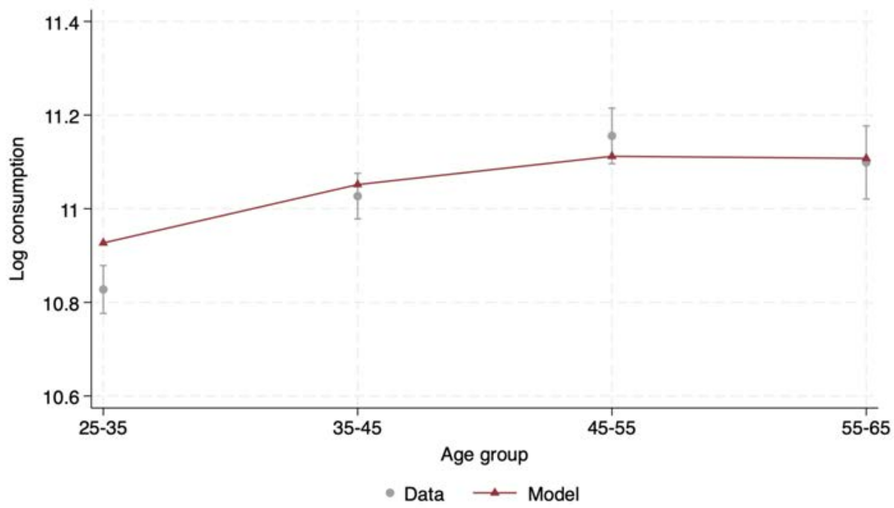


Figure A8: Explained share of scale dependence by type dependence

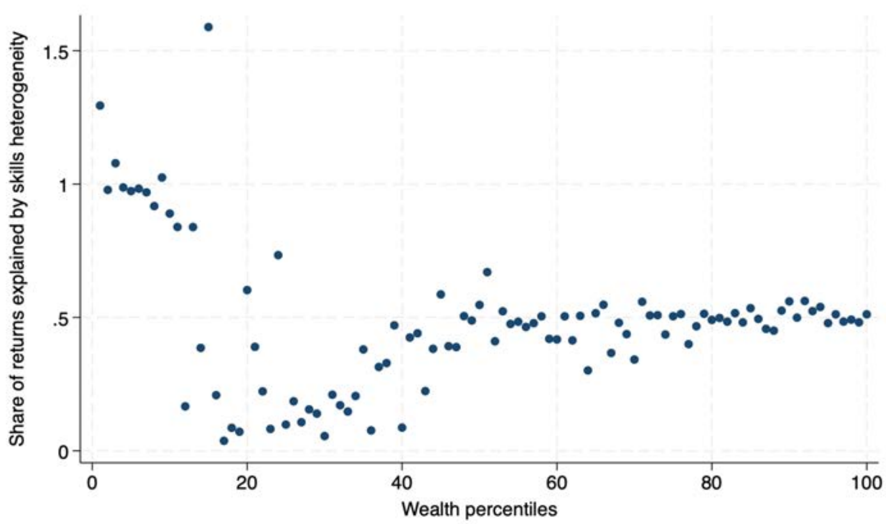
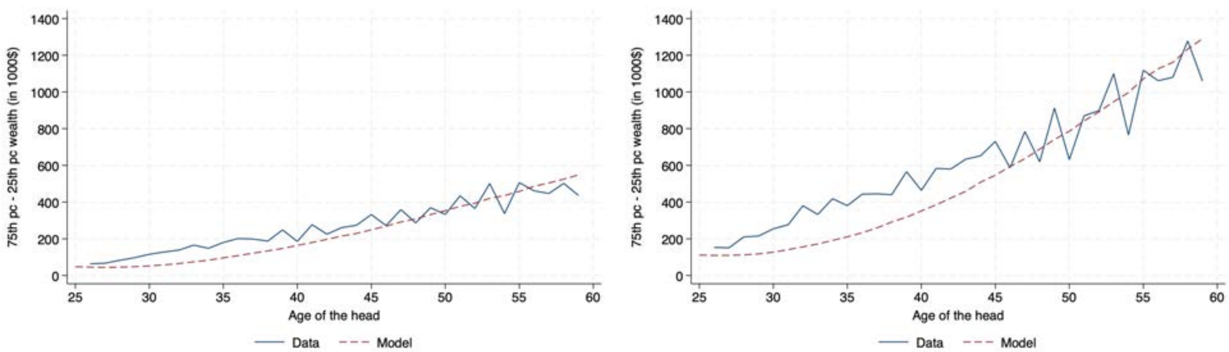


Figure A9: Untargeted moments: Wealth inequality. The figure plots the life-cycle profiles of wealth inequality in the data (solid lines) and simulated by the economic model (dashed line). Panel (a) reports the 75-th/25-th percentile difference, while panel (b) reports the 90-th/10-th percentile difference in assets (in 1000$).

(a) 75th-25th percentile

(b) 90th-10th percentile



Appendix B Using proxies for risk tolerance

What explains the correlation between earnings and returns fixed effects? In the main text, we argued that it may reflect observed and unobserved cognitive and non-cognitive components. Alternatively, it may reflect preference heterogeneity, in particular differences in risk tolerance. Risk aversion naturally affects portfolio composition but may also lead workers to sort in industries or jobs that are safer (and hence command lower compensating wage differentials). Measuring risk aversion is notoriously difficult. In the 1996 and 2012 waves, PSID respondents were asked questions aimed at eliciting their degree of risk aversion.[17] We complement the PSID variable directly measuring risk tolerance with two other variables that are often used as proxies for risk tolerance. The first is whether the individual belongs to a generation "scarred" by poorly performing stock markets. As argued by Malmendier and Nagel 2011, people who share a birth year may be subject, as children or adolescents, to epochal events (e.g., growing up during the Great Depression) that create "experience effects" and shape their beliefs and their risk-taking attitudes later in life. The second variable is whether people receive intergenerational transfers, as they (especially if received at an early stage of the life-cycle) may allow individuals to make different choices both in the labor market, financial markets, and potentially even in the marriage market. The PSID includes information on year of birth and whether people have received private transfers (in the form of an inheritance) in each wave.

[17]Information on other non-cognitive traits were included in the early waves of the PSID, but discontinued in the 1970's.

Table A1: Estimated coefficients deterministic component of earnings

	(1) log Earnings	(2) Residual earnings
Age*college	-0.0499	
	(0.0400)	
$\text{Age}^2/10^2$ * college	0.1690*	
	(0.0921)	
$\text{Age}^3/10^3$ * college	-0.0154**	
	(0.0069)	
Age	0.108***	
	(0.0356)	
$\text{Age}^2/10^2$	-0.1650**	
	(0.0802)	
$\text{Age}^3/10^3$	0.0076	
	(0.0060)	
College education		0.775***
		(0.0109)
Controls	Yes	No
Individual FE	Yes	No
Year FE	Yes	No
Constant	9.545***	-0.546***
	(0.424)	(0.00913)
Observations	13436	13436

Notes: The estimation of the parameters of the income process uses data from the 1999-2019 waves of the PSID. We apply the sample selection described in section 2.1. Controls include dummies for household composition (number of adults and kids), education, unemployment and retirement status of both spouses. Three stars indicate statistical significance at the 1 percent confidence level, two stars at the 5 percent level and one star at the 10 percent level.

Table A2: Estimated parameters stochastic component of earnings

Parameter		HS degree or less	Some college
ρ	Autocorrelation	0.9572***	0.9779***
		(0.0407)	(0.0655)
σ_u^2	Variance transitory shock to earnings	0.0072	0.0305
		(0.0394)	(0.0436)
σ_v^2	Variance permanent shock to earnings	0.0239	0.0185
		(0.0216)	(0.0235)

Notes: The estimation of the parameters characterising the stochastic component of the earnings process uses data from the 1999-2019 waves of the PSID and a GMM estimator with the identity matrix as weighting matrix. The sample selection described in section 2.1 applies. Clustered standard error are reported in parentheses. Three stars indicate statistical significance at the 1 percent confidence level, two stars at the 5 percent level and one star at the 10 percent level.

Table A3: Parameters of the risky assets returns distribution

Parameter		HS degree or less	Some college
μ_s	Excess returns from risky assets	0.0175	0.0400
σ_s^2	Variance shocks to risky assets returns	0.0350	0.0484
p_{tail}	Tail event probability	0.02	
r_{tail}	Return tail event	-0.40	

Notes: The Table reports sample means of the returns from risky assets by level of education of the head. The estimation of the variance of the shocks to risky asset returns uses the residuals of a fixed effects regression of wealth returns on a third-order polynomial in age, time-varying portfolio shares, dummies for lagged wealth percentiles and time fixed effects and a GMM estimator. Data is from the 1999-2019 waves of the PSID. The parameters characterising tail risk are borrowed from Barro (2009).

Table A4: Estimated parameters stochastic component of medical expenses

Parameter		Estimate
ρ^m	Autocorrelation	0.9557***
		(0.1122)
σ^2_ξ	Variance transitory shock to medical expenses	0.4948***
		(0.1264)
σ^2_ζ	Variance permanent shock to medical expenses	0.0672
		(0.0742)

Notes: The estimation of the parameters characterising the stochastic component of the earnings process uses data from the 1999-2019 waves of the PSID and a GMM estimator with the identity matrix as weighting matrix. The sample is restricted to households whose head is between 65 and 90 years of age. Consistent with the sample restriction applied to obtain the main estimation sample, we also drop households where the head is mostly self-employed in the observation period. Finally, we drop the oversampled low-income households ("SEO" sample). Clustered standard error are reported in parentheses. Three stars indicate statistical significance at the 1 percent confidence level, two stars at the 5 percent level and one star at the 10 percent level.

One way to test whether these proxies for risk tolerance matter for explaining the observed correlation between earnings and returns fixed effects is to check how the rank correlation changes after we partial them out of the estimated fixed effects. To do so, we estimate the following regression in the cross-section:

$$\psi_i = \alpha_\psi + \gamma\tilde{\Gamma}_i + \eta_i^\psi$$
$$P_i = \alpha_P + \delta\tilde{\Gamma}_i + \eta_i^P$$

where $\tilde{\Gamma}_i$ includes our proxies for risk tolerance, and the residuals η_i^ψ and η_i^P can be interpreted as the components of persistence in returns and wages due to individual's unobserved cognitive and non-cognitive characteristics. We can then use η_i^ψ and η_i^P to test for the degree of rank correlation between wages and returns to wealth fixed effects that is driven by unobserved cognitive and non-cognitive factors. Results are reported in Table B1. They show that while our proxies for risk tolerance (at least in the richest specification) explain approximately 15% of variation in the rank correlation of the fixed effects, the bulk of variation remains unexplained. We interpret this as evidence that unobserved cognitive or non-cognitive skills components are the most relevant explanation for the correlation. We put some structure on the role of these components in the theoretical model in the main text.

Table B1: Residual rank correlation after partialling out proxies for non-cognitive skills

	No controls	cohort effects	cohort effects and transfers	cohort effects risk avs and transf.
	(1)	(2)	(3)	(4)
$\rho(\eta^P, \eta^\psi)$	0.502	0.493	0.418	0.463
$\rho(P, \psi)$ subsample	0.502	0.502	0.490	0.544
Explained share by obs.		0.018	0.146	0.149
N	1,002	1,002	908	355

Notes: The table reports the residual rank correlation $\rho(\eta^P, \eta^\psi)$ between wealth return fixed effects and the earnings fixed effects, after partialling out various proxies for non-cognitive skills. The rank correlation between wealth return fixed effects and the earnings fixed effects $\rho(P, \psi)$ computed over the sample for which we have full data on the variables used in each column is also reported.

Appendix C Details on model solution and estimation

C.1 Model solution

The dynamic programming problem is solved numerically by backward induction. The solution starts at the end of life (time T) building on the Endogenous Grid Method (EGM) developed by Carroll (2006). Because the probability of being alive at age $T+1$, d_T, is equal to zero, (i.e., agents die with certainty at T), the household is retired and the functional form of the utility $u(c_T)$ and bequest $b(a_T)$ functions are known, we can compute the optimal consumption decision c_T as a function of end-of-period wealth $a_T = m_T - c_T$, for each point of the state space. The endogenous grid for cash-on-hand is then obtained as $m_T = a_T + c_T$. Starting from $T-1$, policy functions and value functions are obtained in two steps. In a first step, we use the EGM conditional on a given level of skills endowment, initial labor income, history of permanent income shocks, average lifetime earnings and permanent medical expense shock, along the discrete portfolio choice dimension (i.e., different shares of wealth allocated to risky assets). In the second step, to ensure the uniqueness of the Euler equation, we apply an upper envelope algorithm as suggested by Iskhakov et al. (2017). More specifically, we disregard non-optimal points of the conditional consumption functions obtained in step 1 applying the upper envelope algorithm as in Druedahl and Jørgensen (2017) and Druedahl (2020). We thus obtain consumption and value functions evaluated on an *exogenous* grid for cash-on-hand for the different alternatives of portfolio allocation. Finally, by comparing the conditional value-of-choice associated with different alternatives along the portfolio dimension, we compute the optimal policy functions and the associated value function at time t. The solution of the problem requires discretizing the six continuous state variables (cash-on-hand, skills endowment, initial labor income, history of permanent income shocks, average lifetime earnings and permanent medical expense shock). Further, we discretize (using an exponential grid) the set of admissible values for the exogenous end-of period wealth. Because consumption is obtained with the EGM, its admissible values are not chosen ex-ante while we use a grid for the admissible values of the share of wealth households can allocate to risky assets. Finally, the density functions for permanent shocks to earnings, out-of-pocket medical expenses and returns from risky assets are approximated using a gaussian quadrature method (Tauchen, 1986). We use linear interpolation in multiple dimensions to evaluate the next-period consumption and value functions associated to values of the states that do not lie on predefined grid points.

C.2 Model estimation

The estimation of the model via SMM proceeds as follows. After having obtained the policy functions for a given set of structural parameters, we use them to simulate the life-cycle profiles of households' decisions and states, starting from age 25. We endow each of these households with skills (Φ_i) and initial labor income (ω_i^y) as drawn from two independent normal distributions with zero mean and variances $\sigma^2_{\Phi_i}$ and $\sigma^2_{\omega^y}$, respectively.[18] Each household is then assigned an initial wealth-to-income ratio as drawn from the empirical distribution of wealth-to-income ratio (conditional on education level) in the PSID data for individuals aged 24-28, as in French (2005b).

We take random draws from the permanent income shocks, medical expenses shocks, risky asset returns and mortality distributions. We use the life-cycle histories of shocks for each household and the policy functions to simulate households' behavior over the life-cycle, starting from the initial conditions. We pool the simulated series for a number of households in each education group that is proportional to that observed in the PSID data. We then use the simulated data to construct moments that correspond to the empirical target moments. Denote with $\Lambda = (\beta, \gamma, \kappa, \widetilde{\theta}, \sigma^2_{\omega^y}, \sigma^2_{\Phi}, \delta^y)$ the collection of model parameters, $\hat{M}^d$ as the vector of moments estimated in the data and $\hat{M}^s(\Lambda)$ as the corresponding simulated moments obtained for a given set of structural parameter values Λ. The structural estimation exercise is to choose Λ to minimize the following SMM criterion:

$$\widehat{\Lambda} = \arg\min_{\Lambda} \left(\widehat{M^d} - \hat{M}^s(\Lambda)\right)' W \left(\widehat{M^d} - \widehat{M^s}(\Lambda)\right) \tag{15}$$

[18]Note that while the population variances $\sigma^2_{\Phi_i}$ and $\sigma^2_{\omega^y}$ are structural parameters and their value changes at each iteration of the estimation algorithm, we fix the rank of the households in the skills endowment and initial labor income distributions ex-ante.

where W is a weighting matrix equal to the diagonal of the variance-covariance matrix of the sample moments (obtained using the bootstrap). We search for the global minimizer of (15) by employing a simulated annealing algorithm (Kirpatrick et al., 1983). The algorithm moves probabilistically in the parameter space given initial values and bounds for the parameter values. To improve the model fit, we adjust the bounds for the parameter values to search within an increasingly narrower parameter space. To ensure that the algorithm can move away from local minima, the "temperature" of the system is reduced by 0.01% at each iteration. In the last step, we report the converged estimates after 1,000 iterations.

The standard errors of the structural parameters reported in Table 4 are computed as the square roots of the diagonal elements of:

$$var(\widehat{\Lambda}) = \left(1 + \frac{1}{J}\right) \frac{\partial \widehat{M^s}(\Lambda)}{\partial \Lambda}' W \frac{\partial \widehat{M^s}(\Lambda)}{\partial \Lambda}$$

where $(1 + \frac{1}{J})$ is the adjustment for simulation error, with J the ratio of the number of observations in the simulated dataset to the number of observation in the PSID data ($J \simeq 27$). We compute the numerical derivatives by finite difference.