Richer earnings dynamics, consumption and portfolio choice over the life cycle

Julio Gálvez and Gonzalo Paz-Pardo*

February 11, 2022

Abstract

Households face earnings risk which is non-normal and varies by age and over the income distribution. We show that allowing for the rich features of earnings dynamics that are present in the data, in the context of a structurally estimated life-cycle portfolio choice model, helps to better understand the limited participation of households in the stock market, the age profile of participation, and their low holdings of risky assets. Households are subject to more background risk than previously considered; as a result, the estimated model under the flexible earnings process implies a substantially lower coefficient of risk aversion.

Keywords: Portfolio choice, life cycle, earnings dynamics, household finance, simulated method of moments.
JEL Codes: G11, G12, D14, D91, J24, H06

*Julio Gálvez: Banco de España. Gonzalo Paz-Pardo: European Central Bank. We thank internal seminar participants at the Banco de España for their comments, and Manolo Arellano, Francisco Gomes, and Paolo Sodini for helpful feedback. The views expressed here are those of the authors and are not necessarily the views of the European Central Bank, the Banco de España or the Eurosystem.

1 Introduction

Many households do not participate in the stock market and, when they do, they invest a limited amount of their wealth into stocks. Standard household portfolio choice models, instead, predict that most households should invest in the stock market, given its high excess returns, even after adjusting for risk. This disconnect has been called the "stock market participation puzzle", and it is the counterpart at the micro level of the "equity premium puzzle".

Understanding why households choose to keep a large fraction of their wealth in assets with lower returns requires studying their portfolio choices in a broad context, taking into account their motivations to save in the first place, and the risks and frictions they face when they take their consumption and saving decisions. Amongst the risks that households face, labor earnings play a crucial role. They are the major source of income for most households, but their fluctuations are difficult to insure against. Thus, the features of earnings risk crucially determine both the ability and the willingness of households to save, and the risk they are ready to bear in their financial portfolio.

In this paper, we study the effect of labor income dynamics on household consumption, savings, and portfolio allocations. In particular, we use a flexible econometric framework (Arellano, Blundell and Bonhomme, 2017) that captures the rich dynamics of shocks to household earnings, including their age-dependence, non-normality, and non-linearity (Guvenen, Karahan, Ozkan and Song, 2021). The rich features of this process are in contrast with traditional, *canonical* earnings processes, which are frequently used in portfolio choice models, that assume that shocks to earnings are normal and that the distribution of earnings shocks does not vary over the age and the income distribution. We estimate both earnings processes on US data from the Panel Study of Income Dynamics (PSID) and use them as inputs to a life-cycle model of portfolio choice, based on Cocco, Gomes and Maenhout (2005), where households choose between saving in risk-free or risky assets, subject to potential entry and per-period participation costs to the stock market. We estimate our model via indirect inference to match a rich set of features from US data, including stock market participation, wealth to income ratios, and portfolio shares of stocks. We compare the implications of a linear, canonical earnings process and our non-linear earnings process on household portfolio choices, and look at the implied parameters both models need to match the data.

We find that the model with a nonlinear earnings process, compared to that with a canonical earnings process, can better explain the limited participation in the stock market with a much lower coefficient of risk aversion. In the presence of richer earnings risk, households increase their precautionary savings to be insured against potential drops to their earnings. Hence, the coefficient of risk aversion that is required to rationalize their portfolio decisions drops from 8.82, which is in the ballpark of standard portfolio choice models that match limited participation and low risky shares (e.g., Cocco et al. (2005), Fagereng, Gottlieb and Guiso (2017), etc.), to 5.64. This estimate is closer to microeconometric estimates that elicit the CRRA coefficient via survey data, which is around 4 (Guiso and Sodini (2013)).[1]

Two features of the earnings data are key to understand why the canonical earnings process can generate counterfactual implications for portfolio choices. First, earnings changes are *negatively skewed*, implying that large negative earnings shocks are more likely than large positive earnings shocks of the same magnitude. This feature, which is at odds with the canonical model with normal shocks, raises the need for precautionary saving and reduces the demand for risky assets, even if their returns are uncorrelated with earnings shocks. In other words, the intuition in Huggett and Kaplan (2016) that future labor market income behaves more like a risky asset and less like a bond than usually considered, and thus the optimal share of stocks with respect to total and financial wealth is lower, is even more relevant in the presence of richer, age-dependent earnings dynamics. Second, earnings risk is still substantial at later ages in the working life (50-60): a non-negligible number of individuals experience adverse events (e.g., unemployment, health) which produce lasting effects on their earnings up to retirement. The canonical process, which is age invariant, generates a large amount of stock market participation at later ages, where agents are subject to counterfactually little background risk. As a result of

[1] Additionally, the model under richer earnings dynamics implies a different structure of adjustment costs than the canonical process.

both factors, the nonlinear process also generates a life-cycle stock market participation profile that is closer to the data and steadily increases as people age even if we do not explicitly target it in our estimation.

We also assess the consumption implications of both earnings processes. Under the richer earnings process, the model closely replicates the consumption reaction to persistent income shocks, as measured by the partial insurance coefficients in Arellano et al. (2017) and Blundell, Pistaferri and Preston (2008) (BPP). Moreover, we find that stockholders are more insured from persistent income shocks than non-stockholders.

Our results are robust to a variety of estimation weights and modelling choices, which include studying housing, disaster risk in the stock market, an empirical distribution of initial wealth and initial stock market participation, different frameworks for the pension system, and different degrees of correlation between stock market returns and earnings shocks. In all of these cases, the nonlinear earnings process better matches household portfolio decisions with lower risk aversion, lower stock market participation costs, or both. Besides, the additional flexibility of the nonlinear process does not imply a sizeable computational cost, as it does not require to increase the state space of the model.

Related literature. This paper contributes to a broad literature in household finance that studies the causes of limited stock market participation. Several papers look at the roles of disaster risk (Fagereng et al. (2017)), housing (Cocco (2005)), trust (Guiso, Sapienza and Zingales (2008)), lack of investor sophistication (Haliassos and Bertaut (1995), Calvet, Campbell and Sodini (2007)), health risk (Rosen and Wu (2004)) and wealth (Calvet and Sodini (2014), Briggs, Cesarini, Lindqvist and Östling (2015)). We contribute to the literature by shedding new light on the effect of income risk on portfolio choice decisions, which has been well studied in the literature (see Guiso, Jappelli and Terlizzese (1996) for an early contribution). The earnings process we choose highlights the importance of age dependence, nonlinearity and non-normality in earnings risks.

Our analysis is focused around a life-cycle model of household portfolio choices, based on the seminal work of Cocco et al. (2005). Subsequent papers have looked at the roles of habit formation (Gomes and Michaelides (2003)), income volatility (Chang, Hong and Karabarbounis (2018)) and personal disaster risk (Nicodano, Bagliano and Fugazza (2021)). We show that the introduction of a richer earnings process yields more reasonable estimates of structural parameters in this class of models, while maintaining a relatively simple model structure.

We also contribute to a literature that estimates stock market participation costs. Earlier papers obtain participation cost bounds via minimal assumptions on the structural model in the background (Vissing-Jorgensen (2002), Paiella (2007)). This was followed by a subsequent literature that calculates participation costs via structural models of portfolio choice. Most of these papers consider either a one-time fixed entry cost (see e.g., Alan (2006)) or a per-period participation cost (see, e.g., Khorunzhina (2013), Fagereng et al. (2017)), and infer the cost structure under a canonical earnings process. In contrast, the participation costs in this paper are closer to the one in Vissing-Jorgensen (2002), who proposes modelling both fixed and per-period costs to stock market participation. We show that the estimates of these costs are closely linked to the earnings process considered. More recently, Bonaparte, Korniotis and Kumar (2020) use the new PSID waves to study household stock market entry and exit in a life-cycle portfolio choice model with canonical earnings dynamics.

The effect of richer earnings risk on portfolio choices has received much less attention (Shen (2018), Catherine (2020), Catherine, Sodini and Zhang (2020)). We build on this recent literature in two ways. First, we show that richer earnings dynamics are relevant to understand portfolio decisions even in the absence of business cycle fluctuations or correlations of labor market income shocks with stock market returns. Second, our earnings process based on Arellano et al. (2017) is more general than those based on a mixture of normals and does not impose any parametric assumptions. Thus, it allows us to reproduce not only the negative skewness of earnings changes, but also the variation of earnings persistence and of the distribution of earnings shocks over the age and the income distributions. These are key to understand the consumption and savings motives of households (De Nardi, Fella and Paz-Pardo (2020)) and, as a result, their asset allocations.

The rest of the paper is organized as follows. Section 2 discusses the models of earnings dynamics, and shows the statistics on stock market participation that we target for our estimation. Section 3 presents the structural model that we estimate following the procedure described in Section 4. We present the estimation results in Section 5, and show that they are robust to modifications of the structural model in Section 6. Finally, Section 7 concludes. We provide further details and robustness checks in the Appendix.

2 Data: earnings dynamics and portfolio decisions

2.1 Data sources and sample selection

Our main data source is the PSID, which is a longitudinal household survey conducted by the University of Michigan since 1968. We use the 1999 to 2017 biennial waves, as these editions provide complete information on consumption, income, and wealth for a representative panel of US households. We complement the PSID information with those from the Survey of Consumer Finances (SCF), which is a cross-sectional survey conducted by the Federal Reserve Board every three years. The main advantage of the SCF is that it provides a more comprehensive picture of the wealth of US households; its main disadvantage is (with few exceptions) the lack of a panel component. To be close as possible to the PSID data, we work with the 1998 to the 2016 editions of the SCF. We detail the sample selection and construction of the datasets in Appendix A. In what follows, we define stockholders as households who directly or indirectly own stocks in non-retirement accounts.

2.2 Earnings dynamics

Earnings dynamics are key to understand household consumption, saving, and portfolio decisions, and are a crucial ingredient in the calibration and estimation of life-cycle models. Recent literature (e.g., Arellano et al. (2017), De Nardi et al. (2020), and Guvenen, Karahan, Ozkan and Song (2016)) highlights the role of nonlinearities in earnings dynamics and their implications for household consumption. In this subsection, we compare and contrast the *canonical* model of earnings dynamics and its nonlinear generalization as presented by Arellano et al. (2017), which will serve as inputs to the structural model.

Consider households indexed by $i = 1, \ldots, N$ that we observe from age $t = 1, \ldots, T$. We decompose log earnings y_{it} as the sum of a deterministic component ($f(X_{it}; \theta)$) and stochastic components:

$$y_{it} = f(X_{it}; \theta) + \eta_{it} + \varepsilon_{it},\ t = 1, \ldots, T. \tag{1}$$

The first component, η_{it}, is persistent and follows a first-order Markov process. The second component, ε_{it}, is transitory in nature, and has zero mean, independent of the persistent component, and independent over time.

The *canonical* model of earnings dynamics is described by the following process:

$$\eta_{it} = \rho \eta_{it-1} + u_{it} \tag{2}$$

$$\eta_{i0} \sim N(0, \sigma_z^2), u_{it} \sim N(0, \sigma_u^2), \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2). \tag{3}$$

As emphasized by Arellano et al. (2017) and De Nardi et al. (2020), among others, the canonical earnings process imposes the following restrictions:

1. *Linearity* of the process of the persistent earnings component. Linearity implies that the right hand side of equation (2) is additively separable to the conditional expectation and the innovation u_{it}. It also implies the linearity of the conditional expectation.

2. *Normality* of the shock distributions. Normality implies that the shock distributions are symmetric, and should not exhibit skewness.

3. *Age-independence* of the autoregressive component ρ and the moments of the shock distributions, which imply the age independence of second and higher-order moments of the conditional distributions of the persistent and transitory component.

Given that these assumptions are at odds with the empirical evidence, Arellano et al. (2017) propose a quantile-based panel data method that allows for nonlinearity, non-normality, and age-dependence. In particular, they model the persistent component of income via the following quantile model:

$$\eta_{it} = Q_t(\eta_{it-1}, u_{it}), \quad (u_{it} | \eta_{it-1}, \eta_{it-2}, \ldots) \sim U[0, 1], \quad t = 2, \ldots, T. \tag{4}$$

where $Q_t(\eta_{it-1}, \tau)$ is the τ-th conditional quantile function of η_{it} given η_{it-1} for a given τ. Intuitively, the quantile function maps random draws from the uniform distribution u_{it} (i.e., cumulative probabilities) into corresponding random draws (i.e., quantile) from the persistent component. The canonical earnings process is a special case: $\eta_{it} = \rho\eta_{it-1} + F^{-1}(u_{it})$.

One way to understand the role of nonlinearity is in terms of a generalized notion of persistence

$$\rho(\eta_{it-1}, \tau) = \frac{\partial Q_t(\eta_{it-1}, u_{it})}{\partial \eta} \tag{5}$$

which measures the persistence of η_{it-1} when it gets hit by a current shock u_{it} with rank τ. This quantity depends on the past persistent component η_{it-1} and the shock percentile τ. Note that while the shocks u_{it} are i.i.d. by construction, they may differ with respect to the persistence associated with them. One can then think of persistence in this context as *persistence of earnings histories*. Moreover, persistence is allowed to depend on the size and the direction of the shock u_{it}. As such, the persistence of η_{it-1} is dependent on the size and sign of current and future shocks $u_{it}, u_{it+1}, \ldots$ In particular, our model allows current shocks to wipe out the memory of past shocks. In contrast, in the canonical linear process, $\rho(\eta_{it-1}, \tau) = \rho$, independent of the realization of the past persistent component η_{it-1} or the shock u_{it}. The model also allows for general forms of conditional heteroscedasticity, as the conditional distribution of η_{it} given η_{it-1} is left unrestricted. More importantly, the model allows for conditional skewness and kurtosis in η_{it}.[2]

Arellano et al. (2017) model the initial distribution of the persistent component η and the transitory component ε via similar quantile representations. The main difference, of course, is that these are not persistent.

Estimating the nonlinear process. Following Arellano et al. (2017), we specify the quantile functions for the persistent and transitory components as lower-order Hermite

[2]Specifically, a measure of period t uncertainty generated by shocks to the persistent component of productivity η_{it-1} is, for some $\tau \in (1/2, 1)$, $\sigma_t(\eta_{it-1}, \tau) = Q_t(\eta_{it-1}, \tau) - Q_t(\eta_{it-1}, 1-\tau)$. Meanwhile, a measure of skewness is $sk(\eta_{it-1}, \tau) = \frac{Q_t(\eta_{it-1},\tau)+Q_t(\eta_{it-1},1-\tau)-2Q_t(\eta_{it-1},\frac{1}{2})}{Q_t(\eta_{it-1},\tau)-Q_t(\eta_{it-1},1-\tau)}$ for some $\tau \in (1/2, 1)$.

polynomials:

$$Q_t(\eta_{it-1}, \tau) = \sum_{k=1}^{K} a_k^{\eta}(\tau) f_k(\eta_{it-1}, age_{it}) \tag{6}$$

$$Q_t(\eta_{i1}, \tau) = \sum_{k=1}^{K} a_k^{\eta 1}(\tau) \tilde{f}_k(age_{i1}) \tag{7}$$

$$Q_t(\varepsilon_{it}, \tau) = \sum_{k=1}^{K} a_k^{\varepsilon}(\tau) f_k^{\varepsilon}(age_{it}) \tag{8}$$

where $a_k^{\eta}(\tau)$, $a_k^{\eta 1}(\tau)$, and $a_k^{\varepsilon}(\tau)$ are modelled as piece-wise linear splines on a grid $[\tau_1, \tau_2]$, $\ldots$, $[\tau_{L-1}, \tau_L]$, which is contained in the unit interval. f_k, $\tilde{f}_k$, and f_k^{ε}, meanwhile, are the approximating functions. We then extend the specification for the intercept coefficients $a_0^{\eta}(\tau)$, $a_0^{\eta 1}(\tau)$, and $a_0^{\varepsilon}(\tau)$ to be the quantile of the exponential distribution on $(0, \tau_1]$ (with parameter λ_-^Q) and $[\tau_L, 1)$ (with parameter λ_+^Q).

If the stochastic earnings components are observed, we could estimate the parameters of the quantile models via ordinary quantile regression. However, as these are latent variables, we proceed with a simulation-based algorithm. Starting with an initial guess of the parameter coefficients, we iterate sequentially between draws from the posterior distribution of the latent earnings components and quantile regression estimation until convergence of the sequence of parameter estimates. Standard errors are computed via nonparametric bootstrap, with 500 replications.

Estimating the canonical process. We estimate the canonical earnings process via a quasi-maximum likelihood procedure, which we explain in Appendix B. The results of the estimation are in Table B.2 of Appendix B. As can be observed, we find that the persistent component is highly persistent, albeit less than a unit root.

Comparing the nonlinear and canonical earnings processes. To compare and contrast the implications of the nonlinear and canonical earnings process, we present results related to persistence and conditional skewness from our estimations. Compared to Arellano et al. (2017), who use pre-tax earnings data, we follow De Nardi et al. (2020), who use post-tax disposable income data.

Figure 1 presents pictures of persistence as a function of the household's position in the income distribution (τ_{init}) and the shock that it receives (τ_{shock}), computed for the average age of a household in the sample (47.5 years). The left graph shows the estimates of the average derivative of y_{it} given y_{it-1}, with respect to y_{it-1}. The figure suggests the presence of nonlinear persistence in the data. In contrast, simulated data from the canonical earnings process implies constant persistence, which is in the right panel of the figure. The nonlinear earnings process, meanwhile, is able to reproduce the empirical patterns quite well, which we show in Figure 8 of Appendix C. We also show in the same figure the persistence of the persistent component η_{it}. As we can observe, the estimates are higher than that observed in the data, which is consistent with the fact that the figure is net of transitory shocks. The associated standard errors are small (see Figure 9 of Appendix C).

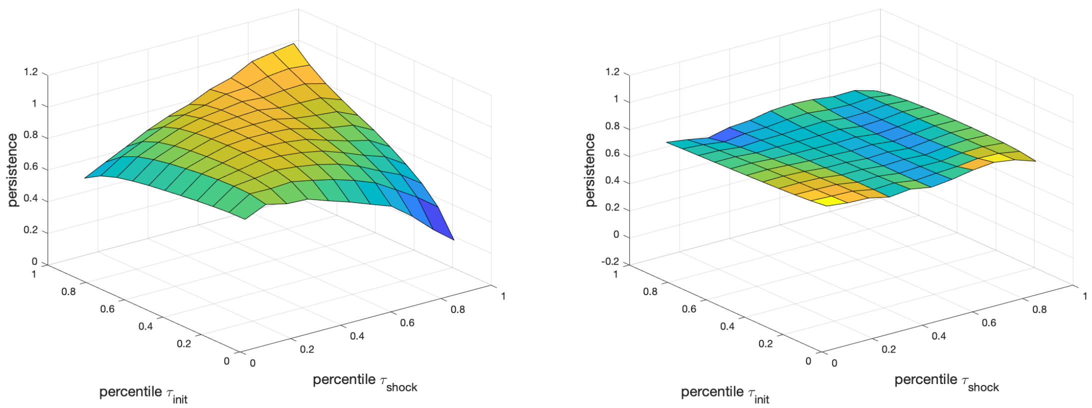

Figure 1: Persistence in the PSID. The left panel presents the graph of the average derivative of y_{it} given y_{it-1}, with respect to y_{it-1}, which was estimated from a quantile autoregression of y_{it} on a third-order Hermite polynomial on y_{it-1}. The right panel presents the same average derivative, but estimated on simulated data from the canonical earnings model.

Figure 2 shows the results with respect to conditional skewness. The left panel shows conditional skewness as a function of the household's position in the income distribution in the data (blue) and in simulated data (green) from the nonlinear earnings model. As the results indicate, we find some evidence of conditional skewness. Moreover, skewness is positive for households with low y_{it}, and negative for households with high y_{it}. The right panel shows the conditional skewness based on simulated data from the canonical earnings model. As the graph indicates, the canonical earnings model predicts symmetric shock distributions. We finally show in Figure 9 of Appendix C the conditional skewness estimates of η_{it}; we find the same patterns, but with a larger magnitude than those for y_{it}. We compute the standard errors and show the results in Figure 10 of Appendix C. The results, once again, are precisely estimated.

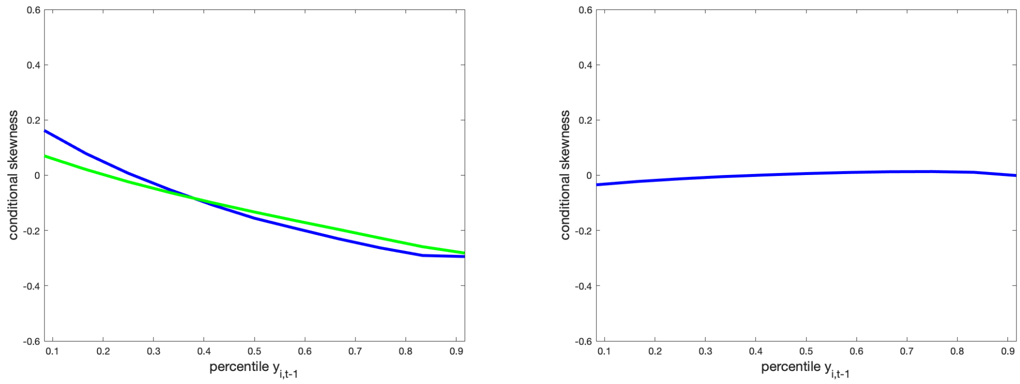

Figure 2: Conditional skewness in the PSID. The left panel presents the graph of the conditional skewness in the data (blue) and the conditional skewness of simulated data from the nonlinear earnings model (green). The right panel presents the conditional skewness based on simulated data from the canonical earnings model.

In sum, the results in this part show that earnings data showcase features that are at odds with the assumptions of previous literature on earnings dynamics, which can affect household portfolio decisions. Richer earnings dynamics, which have a relation to structural labor market models (e.g., job ladder models as in Lise (2013)), are able to capture richer features of the data.

2.3 Stock ownership statistics and life-cycle profiles

To gauge the role of these richer earnings dynamics on household portfolio choices, we begin by looking at the profiles of stock ownership over the life cycle. In Figure 3 we show the age profile for the PSID (left) and for the SCF (right), controlling alternatively for time and cohort effects (Ameriks and Zeldes, 2004). To calculate these profiles, we estimate linear probability models of a dummy of stock market participation on age and time/cohort dummies.[3] Stock ownership is limited at all ages: it starts around 15 percent

[3]We compare the results with a probit model, and results are similar. Results are available upon request.

for the young and increases to around 40 percent for those at age 60. Results are quite similar between the PSID and the SCF.

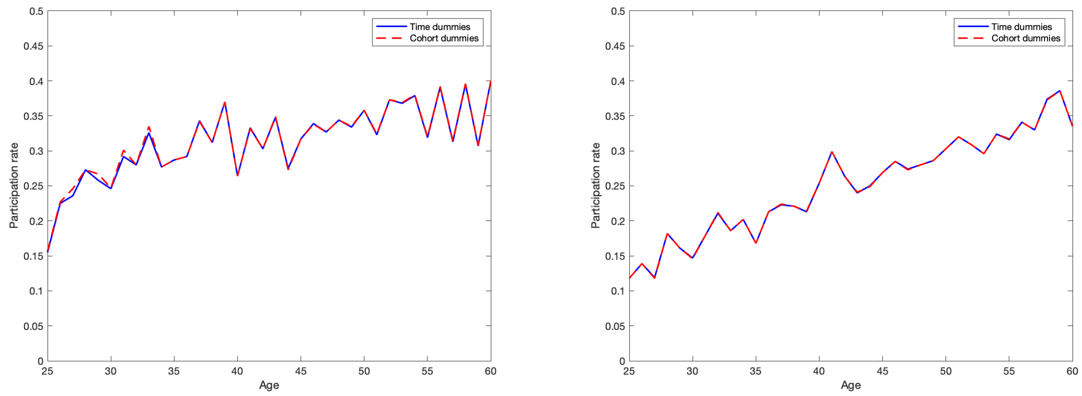

Figure 3: Age profile of stockownership, PSID and SCF. The graphs show the age profile of stockownership computed via linear regressions of a stockowner dummy on age dummies and time dummies (blue) and cohort dummies (red). The graph in the left is from the PSID and the graph in the right is from SCF. Data from the PSID from 1999-2017 waves. Data from the SCF is from 1998-2016.

Variable	N	Mean	SD	P25	Median	P75
Stock ownership	10655	0.324	0.468	0	0	1
Conditional risky share	3444	0.600	0.193	0.001	0.391	0.626
Financial wealth to income ratio	10655	1.549	6.224	0.067	0.354	1.333

Table 1: Summary statistics on stock ownership, conditional risky shares and financial wealth-to-income ratios, PSID, 1999-2017 waves.

We compute some relevant statistics on stock ownership, and moments related to income and wealth. Table 2.3 shows the proportion of stockholders in the data, the conditional risky share, and the financial wealth to income ratios. Around 32 percent of households own stocks; conditional on stock ownership, households allocate on average around 60 percent of their financial wealth on risky assets.[4] Finally, the mean financial wealth to income ratio stands at around 1.5, while the median is around 0.35.

[4]We compute the same statistics also using the SCF, where the patterns that we obtain are similar (see Appendix A).

2.4 Determinants of stockownership

Finally, we dig a bit deeper into the determinants of participation in the stock market by estimating the following model:

$$d_{it} = \alpha_0 + \alpha_1 d_{it-1} + \alpha_2 age_{it} + \alpha_3 age_{it}^2 + \alpha_4 age_{it}^3 + \alpha_5 age_{it}^4 + \alpha_6 w_{it} + \alpha_7 y_{it} + \mathbf{Z}_{it}' \gamma + e_{it}, \quad (9)$$

in which d_{it} is an indicator equal to one when households own stocks or not, age_{it} is household age, w_{it} is household wealth in logarithms, and y_{it} is household income in logarithms. The control variables that are in $\mathbf{Z}_{it}$ include education dummies, cohort dummies, and family size.

Notice that, in this specification, we allow for persistence in stock ownership via a past participation dummy, in line with the literature that studies the structure of participation costs in the stock market (see e.g., Alan (2006), Bonaparte et al. (2020)). As these papers underscore, a positive coefficient on the past participation dummy can be informative about entry costs. We estimate the model via OLS, probit, and a correlated random effects probit model[5].

For brevity, we report the results only for the past participation indicator and for income and wealth, which are presented in Table 2.4. Consistent with previous literature, we find that the past participation indicator is statistically significant and positive. Moreover, the fact that it is significant across all specifications shows that indeed, we are finding persistence in stock market participation; hence, these results suggest the presence of entry costs. Both wealth and income are statistically significant as well.

3 Model

We introduce both the canonical and nonlinear earnings process into a standard discrete time, life-cycle model of consumption and portfolio choices between risky and riskless assets that aims to replicate the facts on stock market participation and portfolio decisions that we have just described.

[5]The reason for considering such an approach is that it could be the case that persistence in stock market participation is actually driven by unobserved heterogeneity.

	(1)	(2)	(3)
VARIABLES	OLS	Probit	CRE Probit
Past participation dummy	0.432***	1.201***	0.898***
	(0.0111)	(0.0362)	(0.0660)
Financial wealth (in logs)	0.0458***	0.331***	0.400***
	(0.00166)	(0.0137)	(0.0201)
Income (logs)	0.0286***	0.0404	0.0693*
	(0.00702)	(0.0314)	(0.0399)
Cohort dummies	x	x	x
Other demographics	x	x	x
Constant	-0.492	-9.088	-8.055
	(2.474)	(10.84)	(11.44)
Observations	8,569	8,569	8,545
R-squared	0.396		

Table 2: Determinants of household stock ownership. This table presents the estimation results of stock market participation at the household level. The key independent variables are past stock market participation, financial wealth, and income (both in logarithms). Other demographic characteristics include a quartic polynomial in age and family size. Robust standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Demographics Agents start working life at 25, face age-dependent positive death probabilities, and die with certainty at age 100. The model period is two years and our unit of observation is the household.

Preferences Households maximize:

$$\max E_t \left[\sum_{t=0}^{t=T} \beta^t \mathcal{S}_t \frac{c_t^{1-\gamma}}{1-\gamma} \right] \tag{10}$$

where c is nondurable consumption, γ is the coefficient of relative risk aversion, β is the discount factor, and $\mathcal{S}_t$ is the probability of arriving alive at time t.

Earnings process As described in Section 2.2, we assume that log earnings can be decomposed in a persistent and a transitory component (Equation 1), and we use alternatively a canonical linear specification and a non-linear, non-normal specification for both components of the earnings process. There is no earnings risk after retirement (age 65), from which households get a public pension.

Budget constraint Households can save in two types of financial assets:

$$c_{t+1} + s_{t+1} + a_{t+1} + \kappa^f(I_{t+1}, I_t) = (1 + r^s_{t+1})s_t + (1 + r)a_t + y_{t+1} \tag{11}$$

where s_t is the amount of wealth invested in the risky asset and a_t the amount of wealth invested in the risk-free asset at time t. r^s_{t+1} represents the risky return of stocks (which is i.i.d.), while r is the risk-free rate. κ^f represents potential costs of participation in the stock market, which depend on the households' stock market participation status I_t. We define

$$I_t = (s_t > 0). \tag{12}$$

Following Vissing-Jorgensen (2002), these may be per-period participation costs (just dependent on I_{t+1}), fixed participation costs (only paid if $I_t = 0$ and $I_{t+1} = 1$, and zero if $I_t = 1$) or a combination of both. The fixed cost of stock market participation can be understood as an entry cost to stock market participation, related to the time spent understanding the risks and returns associated with stocks. The per-period participation cost, meanwhile, can be understood as either the time spent in determining whether portfolio rebalancing is optimal (if the household actively manages its portfolio) or the cost of delegating the investment decisions to a fund manager (if the household indirectly holds stocks via mutual funds).[6]

We define

$$x_t = (1 + r^s_{t+1})s_t + (1 + r)a_t \tag{13}$$

as the amount of cash-on-hand that an individual has at the beginning of period t.

[6] There a third cost of stock market participation in Vissing-Jorgensen (2002), which is a proportional trading cost. As neither the PSID nor the SCF provides information on trading costs, we do not explicitly model this cost.

Finally, as in Cocco et al. (2005), we assume that the household faces borrowing and short-sale constraints:

$$a_t \geq 0,\ s_t \geq 0. \tag{14}$$

The borrowing constraint prevents the household from capitalizing against future labor income or retirement wealth. Meanwhile, the short-sales constraint ensures that the allocation to equities is non-negative.

Households' problem Households thus solve the following problem:

$$V_t(x_t, y_t, I_t) = \max_{c_t, a_{t+1}, s_{t+1}} \left\{ \frac{c_t^{1-\gamma}}{1-\gamma} + \beta \frac{\mathcal{S}_t}{\mathcal{S}_{t-1}} E_t V_{t+1}(x_{t+1}, y_{t+1}, I_{t+1}) \right\} \tag{15}$$

subject to the budget constraint (11) and the borrowing and short-sale constraints (14), and where the expectation E_t is taken with respect to future realizations of persistent income, transitory income, and stock market returns.

4 Structural Estimation

We estimate our structural model via the simulated method of moments (SMM), conditional on the pre-estimated household labor income process and some externally set parameters.

4.1 External parameters

Public pensions are 70% of the average realization of earnings at retirement age (i..e, 35% of average income of workers in the economy). Meanwhile, we set the risk-free rate to 2%, the equity premium to 4%, and the standard deviation of stock market returns to 0.157, following Cocco et al. (2005).[7]

We obtain survival probabilities from Bell, Wade and Goss (1992).

4.2 Estimated parameters

We estimate γ, β, and the stock market participation costs within the model. We assume that stock market participation costs have the following structure:

[7]In the version of the model with disaster risk, there is an ex-ante probability of 2% of stock returns being -48.5%, as in Fagereng et al. (2017).

$$\kappa^f(I_{t+1}, I_t) = \begin{cases} 0 & \text{if } I_{t+1} = 0 \\ \kappa^{FC} + \kappa^{PP} & \text{if } I_{t+1} = 1 \text{ and } I_t = 0 \\ \kappa^{PP} & \text{if } I_{t+1} = 1 \text{ and } I_t = 1 \end{cases} \tag{16}$$

where κ^{FC} and κ^{PP} represent fixed and per-period participation costs, respectively. We experiment with different combinations of both types of costs.

4.3 Targeted moments

We use 11 data moments for our estimation. The first three moments are related to wealth, which we earlier reported in Table 2.3. We target the percentage of people that own stocks directly in the PSID (0.33), median financial wealth-to-income ratios (0.35), and the conditional risky share (0.60). The last eight moments are the parameters of the OLS regression in Table 2.4, which provides information on the age profile, and the effect of state variables on stock market participation.[8]

4.4 Estimation method

We estimate the model via SMM. The SMM finds the values of the parameters γ, β, and κ's that minimize the following quadratic form:

$$\Pi = \min_{(\gamma, \beta, \kappa^{FC}, \kappa^{PP})} (M^s - M^d)' W (M^s - M^d). \tag{17}$$

Here, M^d is the data moments, M^s is the simulated moments from the structural model, and W is a weighting matrix. We report standard errors for our coefficient estimates that take into account our weighting choice, the variability of the moments in the data (through a bootstrap) and the responsiveness of our parameter estimates to the moments in the data. More details are in Appendix D.

5 Results

5.1 Estimated parameters and model fit

Table 3 shows the fit of the model by comparing our targets in the data (left column) with the model implications under a non-linear earnings process (NL, central column)

[8] In Appendix G we show that our conclusions are unchanged if we include housing in the model and target total wealth instead.

Model			Baseline	
Moment	Data	NL	CA	CA,bis
Participation	0.32	0.52259	0.54915	0.59642
Risky share	0.60	0.60701	0.56683	0.45098
Med W/I	0.35	0.99485	1.3731	1.1228
OLS constant	-0.492 (2.474)	-7.4301	-42.543	-45.754
OLS partic	0.432 (0.011)	0.77569	0.76944	0.73909
OLS age	0.0288 (0.234)	0.64526	3.5563	3.8489
OLS age2	-0.00105 (0.008)	-0.02049	-0.10927	-0.11918
OLS age3	9.96e-6 (1.28e-4)	0.00028562	0.0014606	0.0016093
OLS age4	-2.06e-8 (7.25e-7)	-1.4723e-06	-7.1668e-06	-7.9962e-06
OLS log income	0.0286	0.15717	0.54154	0.54854
OLS log wealth	0.0458	0.0093104	0.022713	0.010105

Table 3: Targeted parameters

and those under a linear, canonical earnings process (CA, right column). In both cases, the model fits its targets remarkably well given how parsimoniously parameterized it is (we estimate 4 parameters to fit 11 targets). In particular, the model closely replicates the limited level of stock market participation that we observe in the data (35% and 34% for the non-linear and canonical process, respectively, compared with an average of 32% in our data) and the conditional risky share of stockholders (60% and 66%, respectively, in contrast with 60% in the data), two crucial moments to understand the savings and portfolio decisions of US households (Alan (2012) and Bonaparte et al. (2020)).[9]

However, the model fits the data under remarkably different estimated parameters when we equip it, alternatively, with each of the earnings processes we consider (Table 4). Most notably, the implied CRRA risk aversion parameter is substantially lower (5.64) under the non-linear process than it is under the canonical earnings process (8.82). In order to derive intuition for this result, it is important to understand why households are

[9] Whilst our baseline estimations display different median financial wealth-to-income ratios, we show in Appendix F.9 that an alternative estimation that better matches this moment delivers similar implications.

Model	Baseline				
Parameter	NL		CA		CA, bis
γ	5.3967	(0.0573)	6.4167	(0.2875)	9.8833
β	0.9175	(0.0148)	0.99554	(0.0126)	0.96379
κ^{FC}	0.0103	(0.0054)	0.2635	(0.0047)	0.0928
κ^{PP}	0.0788	(0.0046)	0.0966	(0.0029)	0.069

Table 4: Parameter estimates (SD in parentheses). β is expressed in annual terms. The participation costs are expressed as fractions of average household income (the numeraire in the model)

saving and, when applicable, investing in the stock market under both income processes.

5.2 Policy functions

Before discussing the simulated life-cycle profiles for consumption and portfolio choice, we first analyze the policy functions that underlie these results, which we show in Figure 4, to gain intuition on the determinants of consumption and portfolio choice. The top panel shows the optimal risky shares for households of different ages under the non-linear process (left) and the canonical process (right). Low-wealth households do not invest at all in the stock market: given their low savings, they find it optimal not to pay the fixed participation costs. Starting from the threshold at which it becomes optimal to buy risky assets, the policy functions are decreasing in financial wealth. The key driver is the importance of human capital (discounted stream of future labor income) relative to accumulated wealth. During working age, since shocks to households' labor income are uncorrelated with stock returns, the *deterministic* component of labor income mimics the payoff of a riskless asset. Hence, for a given level of human capital, households with low financial wealth will tend to invest more aggressively in stocks than wealthier households. Higher financial wealth reduces the relative importance of "bond-like" human wealth, leading households to rebalance their portfolios by investing less in stocks.

There are relevant differences by age between the non-linear and the canonical earnings processes. Under the nonlinear earnings process, there is an outward shift of the portfolio rules from ages 20 to 30, followed by a inward shift at age 55 until retirement. Under the canonical earnings process, there is also an outward shift from ages 20 to 30, followed by a further outward shift at age 55. The differences can be linked to the types of income risks that these households face. The age-dependence of income shocks and its non-normality in the non-linear earnings process imply that households' income becomes riskier throughout the working life, even at older ages. Thus, households are conservative in their investments over their working life. Meanwhile, in the canonical earnings process, uncertainty with respect to labor income gets resolved much earlier, which implies that older households who still have low financial wealth with respect to their human capital invest aggressively in stocks. This result is emphasized in the middle and bottom panels of Figure 4. At young ages (middle), only the high earning households have the incentives to invest in the stock market, as the participation threshold that they have to overcome is relatively lower. This is true in both the cases of the nonlinear and canonical earnings processes. However, as the households reach middle age (bottom), households under the nonlinear earnings process still find that the costs outweigh the benefits of stock market participation, and, even if they do invest, they do not start investing aggressively into stocks. In contrast, under the canonical process, for many households at age 55, the benefits make it worthwhile to load their money into stocks before retirement.

As the policy functions illustrate, the participation costs induce a wealth threshold for participation in the stock market, which we show in Figure 5 for different percentiles of the income distribution. The left panel shows that, under the nonlinear earnings process, the wealth thresholds do not have a clear age profile, and depend strongly on the position of a household in the income distribution. This is in contrast to the wealth thresholds under the canonical earnings process, which exhibit a clear U-shape that reaches its trough at age 50.

5.3 Life-cycle profiles

These different policy functions lead to contrasting life-cycle profiles for both models. The left panel of Figure 6 shows average wealth accumulation, by age, in the two economies. In

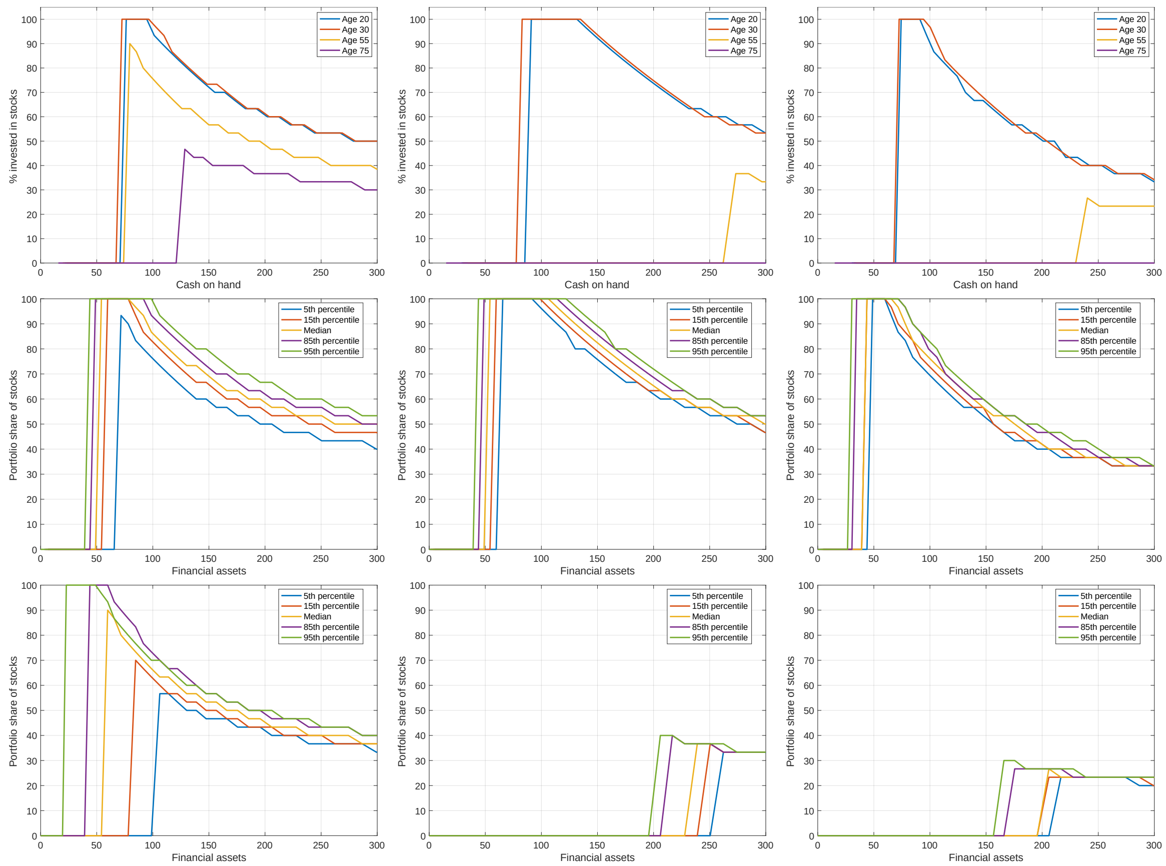

Figure 4: Portfolio share of stocks, by cash on hand and age for the median earner (top), by cash on hand and income for the 30-year-old group (middle), by cash on hand and income for the 55 year-old group (bottom). Left: NL; right: canonical.

the case of the canonical process, households barely save up until age 45, from which they start accumulating wealth very fast, and then decumulate it just as fast after retirement. This effectively illustrates that the key reason why households save at all in this economy is related to the life-cycle: they want to smooth their consumption between their pre- and post-retirement periods. In the non-linear process, while this motive for saving is also there, households also save for precautionary reasons: they want to insure themselves against potentially large income shocks that are a result of the non-linear and non-normal labor income risks they face. Therefore, they start accumulating wealth much earlier in life. These differences are not driven by the average profile of labor market income in the two economies, which is kept constant.

As a result, the saver households in the canonical model are relatively old and have a strong incentive to invest a large part of their wealth in stocks; in contrast, the saver

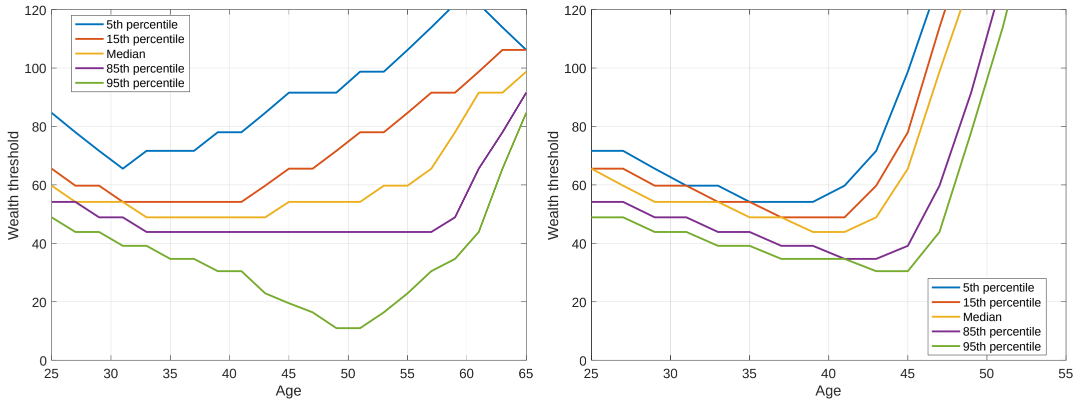

Figure 5: Wealth thresholds (in units of biennial average income, at beginning of period) for participation in the stock market, by age and income level. Left: NL process, right: canonical process.

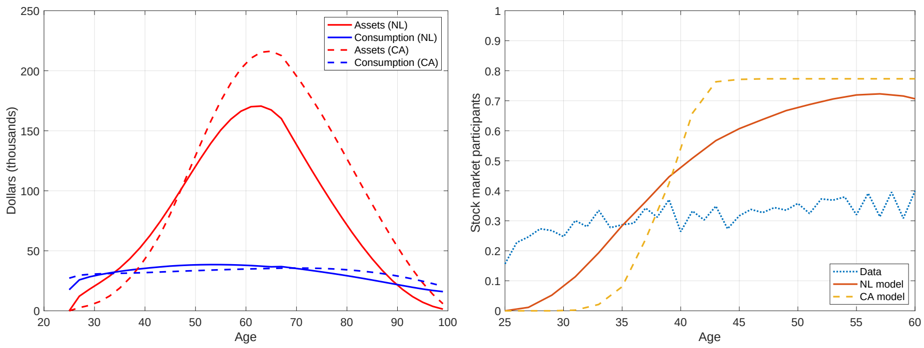

Figure 6: Average savings and consumption (left) and stock market participation (right) over the life-cycle, NL vs canonical process

households under the non-linear process have a wider age distribution, and for many of them it is not optimal to invest all of their wealth into stocks. Even at older ages, the negative skewness they face is enough to keep many households off stocks, and those that enter the stock market do not invest heavily into it.

Thus, in order to rationalize the limited participation in the stock market and the relatively low stock share in financial assets, the model under the canonical process implies that households must be very risk averse. In contrast, the model under the non-linear process can explain low stock market participation and risky shares with much lower risk aversion. As Table 5 shows, the model under the canonical process would generate a counterfactually high risky share (0.87) if equipped with the parameter estimates for the non-linear process.

Process	NL	CA	NL	CA
γ	5.3967	6.4167	6.4167	5.3967
β	0.9175	0.99554	0.99554	0.9175
κ^{FC}	0.0103	0.2635	0.2635	0.0103
κ^{PP}	0.0788	0.0966	0.0966	0.0788
Participation	0.52259	0.54915	0.34748	0.43392
Risky share	0.60701	0.56683	0.39815	0.82219

Table 5: Parameters and targets: NL vs canonical. NL needs a much lower CRRA coefficient to get a correct risky share given participation.

Additionally, as the right panel of Figure 6 shows, these dynamics also imply that the NL process replicates better, without explicitly targeting it, the flatness of the life-cycle profile of stock market participation. With the NL process, young rich households have incentives to save, and once they go over a certain cash-on-hand threshold they start investing in the stock market, although not much (see Figure 5). In the canonical process, these households do not have incentives to save, so they do not have enough wealth to justify paying the participation cost in the stock market (which is similar across both economies) - which is why they mostly stay non-participants.

5.4 Consumption implications

The effect of an earnings shock. To gain further intuition about the differences between processes, Figure 7 studies the effects of earnings shocks in the consumption, savings, and portfolio allocation decisions of households. The panels on the left hand side look at the case of a 35-year-old, with median earnings and accumulated wealth equivalent to 1.5 years of earnings, that suffers, alternatively, a very good earnings realization (a shock at the 90th percentile of the possible earnings shocks for this individual) and a very bad earnings realization (a shock at the 10th percentile of the shock distribution). For both cases and all variables, we represent deviations from the average path this individual would have followed if he or she had received a shock of rank 0.5 (i.e., very close to no changes in earnings).

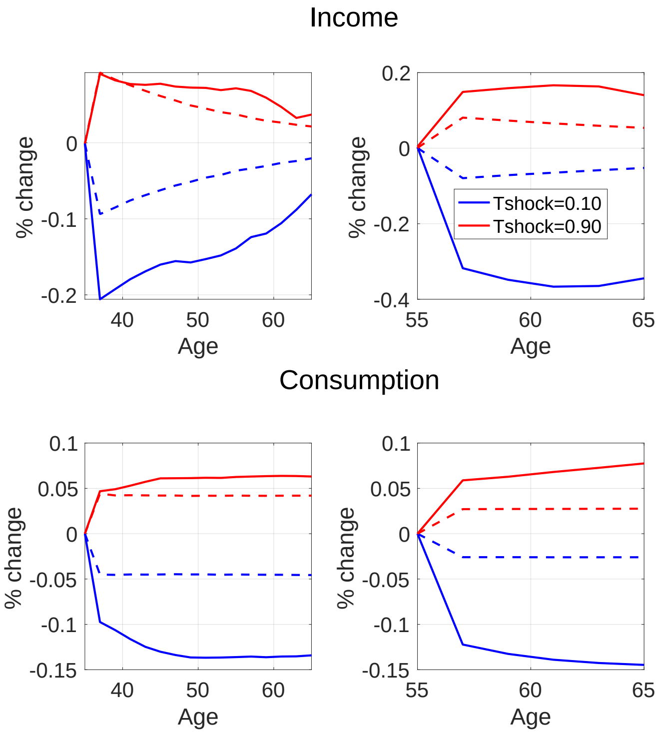

Figure 7: Reactions of a median earner with wealth equal to 1.5 years of average earnings at age 35 (left) and a median earner with wealth equal to 2.5 years of average earnings at age 55 (right) to shocks of rank 0.10 and 0.90, respectively, for the nonlinear (solid line) and canonical (dashed line) processes.

In the top left graph, we observe that 35-year olds are subject to substantial earnings risk, and that the shocks they suffer can be relatively persistent in the case of both earnings processes. For instance, looking at the canonical process (dashed lines), a negative shock of -10% of earnings at age 35 still translates into lower earnings (around -4% with respect to the counterfactual) at age 50. However, there are also relevant differences across processes. By construction, shocks in the canonical process are symmetric: a positive shock and a negative shock of the same magnitude are equally likely, and have the same persistence. The situation is different in the non-linear process (solid lines): because there is negative skewness of earnings shocks, the 10th-percentile worst shock is larger (almost double) than the 10th-percentile best shock, and has much longer persistence over time.

Because of this, the impacts on consumption (second panel on the left) of such a shock are larger and much more persistent under the NL process. This explains why the precautionary savings motive is stronger under this earnings process, and why people accumulate more wealth when they are young.

On the right hand side, instead, we represent the same magnitude of shocks, also for the median earner, but now of age 55. The most salient fact is that, under the canonical process, this individual is subject to much smaller earnings risk than under the NL process. In the NL process, instead, there is still a reasonable chance of a large earnings shock, that can also significantly impact consumption as this household approaches retirement. This explains why, at this age, most households in the canonical world participate in the stock market.

Consumption pass-throughs. An alternative to understand the consumption implications of the two earnings process is to estimate the degree to which households can self-insure against income shocks. Specifically, we simulate data from both economies and compute consumption insurance coefficients via the procedure proposed by Arellano et al. (2017), which we describe fully in Appendix E of the paper. The results that we obtain imply that the nonlinear earnings process implies estimates of consumption insurance which are in line with the empirical results in Arellano et al. (2017). Expressed in terms of Blundell et al. (2008) coefficients, 44% of persistent earnings shocks do not translate into consumption changes, which compares with 36% in the data and only 18% in the canonical process. We also find that stockholders appear to self-insure their consumption against income shocks better than non-stockholders, which suggests the benefits of diversification.

5.5 Other implications: entry and exit dynamics

The non-linear process also generates more realistic dynamics of stock market participation (Table 6). In the model under the canonical process, almost all working-age households enter the stock market once, and exit it after retirement; very few stay non-participants throughout. In the non-linear process, even though the cross-sectional share of stock market participation is very similar, there is much more heterogeneity: around 40% of households never participate in the stock market (30% in the data), 18% always

Process	Never	Always	Single entry	Other cases
Data	0.3028	0.0478	0.2111	0.4383
NL	0.23489	0.34424	0.36375	0.057123
Canonical	0.2268	0.2904	0.4828	0

Table 6: Entry and exit from stock market over a 10-period (20-year) window. The data corresponds to a balanced panel in the PSID (1999-2017).

do (4% in the data), 21% only enter the stock market once in the sample (31% in the data), and the remainder are multiple entrants or exiters, some of which alternate between both statuses multiple times. As Figure 5 shows, the wealth thresholds to enter the stock market are relatively stable over the life cycle under the canonical process, but have a strong age component in the nonlinear case.

6 Robustness

In this section we show that the main implications of our model, most relevantly in terms of the estimated coefficient of risk aversion under both earnings processes, are unchanged under a set of additional elements that our baseline model does not consider.

6.1 Alternative assumptions on the model without housing

We first study the changes in the parameter estimates when we introduce additional elements to the baseline Cocco et al. (2005) model. These include the possibility of a "rare disaster" in the stock market that implies large losses for stockholders, replicating the empirical distribution of wealth at age 25, replicating the empirical initial stock market participation at age 25, an alternative pension system in which old age benefits depend on the last realization of earnings, correlation between labor market income shocks and stock market returns, and an alternative calibration in which both processes more closely replicate the median wealth to income ratio that we observe in the PSID. For all of these, the canonical process implies larger coefficients of risk aversion, stock market participation costs, or both. Table 7 offers a quick summary of the parameter estimates for each of these cases, and Appendix F contains detailed descriptions of each of the experiments.

Model	NL process				Canonical process			
Baseline	5.3967	0.9175	0.0103	0.0788	6.4167	0.99554	0.2635	0.0966
Disaster risk	4.981	0.83193	0	0.05	9.2759	0.78473	0.0125	0.025
Initial wealth	5.5411	0.79668	0.05	0.0375	8.1628	0.79668	0.0625	0.025
Initial stock part.	4.981	0.80784	0.07	0.025	7.2741	0.79668	0	0.0125
Alt. pension	4.9	0.74162	0	0.05	10.362	0.74733	0.05	0.025
Correlation, pos.	5.4666	0.83193	0	0.0625	9.3507	0.78307	0.04	0.025
Correlation, neg.	5.7093	0.79969	0	0.0625	9.5934	0.7746	0.04	0.025
Same w to inc	4.6642	0.74675	0.01162	0.03	5.5526	0.91652	0.01071	0.10286

Table 7: Estimated parameters under variations of the baseline model

6.2 Portfolio choice under the presence of housing

Houses are often the most significant component of household balance sheets, and the decision to own or rent a home has implications for household wealth accumulation (Paz-Pardo, 2021) and financial portfolio decisions (Cocco, 2005). In this subsection, we show that our main conclusions are unchanged in a model with an endogenous housing choice, which thus explains *total wealth* rather than *financial wealth*.

In the model that we build, households receive utility both from housing and non-housing consumption, with the utility from owner-occupied housing represented by parameter u_h. Households choose to buy a house or to rent, and face transaction costs related to house purchase. Homeowners may finance their house with a mortgage, for which they need to satisfy a downpayment constraint. Because of the new elements in the model, we now estimate five structural parameters: the previous model parameters, plus the utility gain from housing u_h. We target 13 moments to estimate the five parameters: in addition to the moments that we have discussed earlier, we also target the homeownership rate and we augment the stockownership model (9) by including a homeownership dummy. Additional details about the model, together with model fit and other results are summarized in Appendix G.

Table 8 shows the associated parameter estimates of the structural model. We find

Model	Baseline	
Parameter	NL	CA
γ	4.6767 (0.1072)	8.6767 (0.5018)
β	0.7746 (0.0138)	0.87241 (0.0151)
κ^{FC}	0 (0.0021)	0.01 (0.0040)
κ^{PP}	0.06 (0.0041)	0.05 (0.0015)
u^h	1.375 (0.0292)	1.3 (0.0455)

Table 8: Parameter estimates. β is expressed in annual terms. The participation costs are expressed as fractions of average household income (the numeraire in the model)

that the CRRA parameter in the economy under the nonlinear earnings process (4.68) is much lower than that estimated in the economy under the canonical earnings process (8.68). Moreover, the structure of participation costs is similar to that already estimated in the baseline model. That is, the model under the nonlinear earnings process places weight only on the per-period participation costs, while the canonical earnings process provides a role for the entry cost to the stock market. Additionally, as shown in Appendix G, the model with the nonlinear earnings process generates a negative correlation between homeownership and stock market participation once other observables are controlled for, as in our data and unlike in the canonical process, where the correlation is positive. The richer model with housing also implies a Blundell et al. (2008) (BPP) coefficient of 0.39, which is closer to the empirical estimate of 0.36 than our main model (0.44); the model under the canonical earnings process, instead, implies a BPP estimate of 0.10.

7 Conclusion

In this paper, we estimate a richer stochastic process for earnings that features a transitory component and a persistent component that allows for age-dependence in moments, nonlinearity, and non-normality. We use it as an input to an estimated life-cycle portfolio choice model that features a one-time fixed entry cost and a per-period participation cost, and compare the implications of the canonical permanent/transitory linear process, with age-independent, normal shocks and the nonlinear earnings process. Our results indicate that, under a variety of specifications, the model with the nonlinear earnings process exhibits a lower risk aversion coefficient than the canonical earnings process. The model with the nonlinear earnings process also replicates more closely stock market participation by age, consumption insurance, and wealth accumulation patterns.

Our model assumes CRRA preferences, as in the workhorse model of Cocco et al. (2005). However, a wide literature has investigated the portfolio decisions of households under different utility specifications (Guiso and Sodini (2013)), the most prominent of which is Epstein and Zin (1989) utility. Our estimate of 5.64 can be regarded as a lower bound for risk aversion estimates when preferences are Epstein-Zin, and is within the range of the Epstein-Zin risk aversion estimates in Calvet, Campbell, Gomes and Sodini (2021).

This result complements recent literature that showed that countercyclical skewness is important to understand limited stock market participation (Shen (2018) and Catherine (2020)). A promising avenue for future work is to combine both frameworks, using the business-cycle varying earnings process proposed in Paz-Pardo (2021), and study potential complementarities between both approaches.

References

Alan, Sule (2006), 'Entry costs and stock market participation over the life cycle', *Review of Economic Dynamics* **9**(4), 588–611.

Alan, Sule (2012), 'Do disaster expectations explain household portfolios?', *Quantitative Economics* **3**(1), 1–28.

Ameriks, John and Zeldes, Stephen P (2004), How do household portfolio shares vary with age, Technical report, working paper, Columbia University.

Arellano, Manuel (2003), *Panel data econometrics*, Oxford university press.

Arellano, Manuel, Blundell, Richard and Bonhomme, Stéphane (2017), 'Earnings and consumption dynamics: A non-linear panel data framework', *Econometrica* **85**(3), 693–734.

Bell, Felicitie C., Wade, Alice H. and Goss, Stephen C. (1992), 'Life tables for the United States Social Security Area: 1900-2080', (Social Security Administration, Office of the Actuary).

Blundell, Richard, Pistaferri, Luigi and Preston, Ian (2008), 'Consumption inequality and partial insurance', *The American Economic Review* pp. 1887–1921.

Blundell, Richard, Pistaferri, Luigi and Saporta-Eksten, Itay (2016), 'Consumption inequality and family labor supply', *The American Economic Review* **106**(2), 387–435.

Bonaparte, Yosef, Korniotis, George and Kumar, Alok (2020), 'Income risk, ownership dynamics, and portfolio decisions'.

Briggs, Joseph S, Cesarini, David, Lindqvist, Erik and Östling, Robert (2015), Windfall gains and stock market participation, Technical report, National Bureau of Economic Research.

Calvet, Laurent E and Sodini, Paolo (2014), 'Twin picks: Disentangling the determinants of risk-taking in household portfolios', *The Journal of Finance* **69**(2), 867–906.

Calvet, Laurent E, Campbell, John Y and Sodini, Paolo (2007), 'Down or out: Assessing the welfare costs of household investment mistakes', *Journal of Political Economy* **115**(5), 707–747.

Calvet, Laurent E, Campbell, John Y, Gomes, Francisco and Sodini, Paolo (2021), The cross-section of household preferences, Technical report, National Bureau of Economic Research.

Catherine, Sylvain (2020), 'Countercyclical labor income risk and portfolio choices over the life-cycle', *HEC Paris Research Paper No. FIN-2016-1147* .

Catherine, Sylvain, Sodini, Paolo and Zhang, Yapei (2020), 'Countercyclical income risk and portfolio choices: Evidence from sweden', *Swedish House of Finance Research Paper* (20-20).

Chang, Yongsung, Hong, Jay H and Karabarbounis, Marios (2018), 'Labor market uncertainty and portfolio choice puzzles', *American Economic Journal: Macroeconomics* **10**(2), 222–62.

Cocco, Joao F (2005), 'Portfolio choice in the presence of housing', *Review of Financial studies* **18**(2), 535–567.

Cocco, Joao F, Gomes, Francisco J and Maenhout, Pascal J (2005), 'Consumption and portfolio choice over the life cycle', *Review of financial Studies* **18**(2), 491–533.

De Nardi, Mariacristina, Fella, Giulio and Paz-Pardo, Gonzalo (2020), 'Nonlinear household earnings dynamics, self-insurance, and welfare', *Journal of the European Economic Association* **18**(2), 890–926.

De Nardi, Mariacristina, French, Eric and Jones, John B. (2010), 'Why do the elderly save? The role of medical expenses', *Journal of Political Economy* **118**(1), 39–75.

Epstein, Larry G and Zin, Stanley E (1989), 'Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework', *Econometrica: Journal of the Econometric Society* pp. 937–969.

Fagereng, Andreas, Gottlieb, Charles and Guiso, Luigi (2017), 'Asset market participation and portfolio choice over the life-cycle', *The Journal of Finance* **72**(2), 705–750.

Gomes, Francisco and Michaelides, Alexander (2003), 'Portfolio choice with internal habit formation: A life-cycle model with uninsurable labor income risk', *Review of Economic Dynamics* **6**(4), 729–766.

Guiso, Luigi and Sodini, Paolo (2013), Household finance: An emerging field, *in* 'Handbook of the Economics of Finance', Vol. 2, Elsevier, pp. 1397–1532.

Guiso, Luigi, Jappelli, Tullio and Terlizzese, Daniele (1996), 'Income risk, borrowing constraints, and portfolio choice', *The American Economic Review* pp. 158–172.

Guiso, Luigi, Sapienza, Paola and Zingales, Luigi (2008), 'Trusting the stock market', *the Journal of Finance* **63**(6), 2557–2600.

Guvenen, Fatih, Karahan, Fatih, Ozkan, Serdar and Song, Jae (2016), What do data on millions of U.S. workers reveal about life-cycle earnings risk? Working paper, University of Minnesota.

Guvenen, Fatih, Karahan, Fatih, Ozkan, Serdar and Song, Jae (2021), 'What do data on millions of us workers reveal about lifecycle earnings dynamics?', *Econometrica* **89**(5), 2303–2339.

Haliassos, Michael and Bertaut, Carol C (1995), 'Why do so few hold stocks?', *Economic Journal* **105**(432), 1110–1129.

Huggett, Mark and Kaplan, Greg (2016), 'How large is the stock component of human capital?', *Review of Economic Dynamics* **22**, 21–51.

Kaplan, Greg and Violante, Giovanni L. (2010), 'How much consumption insurance beyond self-insurance?', *American Economic Journal: Macroeconomics* **2**(4), 53–87.

Khorunzhina, Natalia (2013), 'Structural estimation of stock market participation costs', *Journal of Economic Dynamics and Control* **37**(12), 2928–2942.

Lise, Jeremy (2013), 'On-the-job search and precautionary savings', *Review of economic studies* **80**(3), 1086–1113.

Nicodano, Giovanna, Bagliano, Fabio-Cesare and Fugazza, Carolina (2021), Life-cycle risk-taking with personal disaster risk, Technical report, University of Torino.

Paiella, Monica (2007), 'The forgone gains of incomplete portfolios', *The Review of Financial Studies* **20**(5), 1623–1646.

Paz-Pardo, Gonzalo (2021), 'Homeownership and portfolio choice over the generations'. ECB Working Paper no. 2522.

Rosen, Harvey S and Wu, Stephen (2004), 'Portfolio choice and health status', *Journal of Financial Economics* **72**(3), 457–484.

Shen, Jialu (2018), Countercyclical risks and portfolio choice over the life cycle: Evidence and theory, *in* '9th Miami Behavioral Finance Conference'.

Vestman, Roine (2019), 'Limited stock market participation among renters and homeowners', *The Review of Financial Studies* **32**(4), 1494–1535.

Vissing-Jorgensen, Annette (2002), Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures, Technical report.

Yao, Rui and Zhang, Harold H (2005), 'Optimal consumption and portfolio choices with risky housing and borrowing constraints', *Review of Financial studies* **18**(1), 197–239.

A Data and summary statistics

As we mentioned in section 2.2 of the main text, we use a combination of the PSID and the SCF for the estimation of the earnings process and the calculation of the auxiliary statistics for the structural estimation.

A.1 PSID

The PSID follows a large number of US households and their potential spin-offs since 1968. While the survey was originally designed to track income and poverty, the PSID has since evolved into tracking household consumption and wealth in more recent waves. When it originally started, the PSID was composed of two main samples: the Survey Research Center (SRC) sample, which was designed to be representative of the US population, and the Survey of Economic Opportunity (SEO), which oversamples the poor.

For the purposes of this study, we focus on the biennial waves that started in 1999. This is because starting from this wave, the PSID has continuous information on household earnings, assets, and consumption.

To construct the statistics that we use for estimation, we follow the sample selection criterion in Blundell, Pistaferri and Saporta-Eksten (2016). In particular, we consider households with heads aged 25 to 60 years old, who are continuously married, and who have continuously participated in the labor force. This leaves us with 10,655 household-year observations. We exclude individuals who are part of the SEO to obtain a representative sample.

A.1.1 Variable definitions

The main variables that we use for the calculation of auxiliary statistics and the earnings process are income, wealth, and the risky share.

The definition of income that we use follows De Nardi et al. (2020). In particular, we use *disposable household earnings*, which are defined as the sum of household labor income and transfers, such as welfare payments, net of taxes and Social Security contributions paid. The reason for this is due to our focus on understanding how households choose between different assets to insure their consumption against income risk. Wealth is defined as total financial wealth, which is the sum of households' holdings in stocks, bonds, and cash, plus any amount invested in retirement accounts. The risky share, then, is defined as the share of stocks in total financial wealth.

A.2 SCF

The SCF is a repeated cross-sectional survey that studies the wealth of US households. It is triennial in nature. The main advantage of the SCF as opposed to the PSID is that it is more detailed with respect to information on wealth. A disadvantage of using the SCF is that as it is a cross-sectional survey, we wouldn't be able to follow households over time; moreover, the SCF does not have information on consumption.

In order to calculate the statistics that we use for comparison with the PSID, we use similar criteria as in Blundell et al. (2016). We also remove households with incomplete information on education, age, and other demographic information. We also remove households that have zero labor income, and who have less than $100 in financial assets, following Fagereng et al. (2017). This criteria gives us a sample of 54,321 households. Given that the SCF oversamples the wealthy, we use weights in the calculation of the auxiliary statistics. To have a comparable sample period as with the PSID, we work with the 1998-2016 waves.

A.3 Some summary statistics

We now compare some summary statistics that we obtain with the PSID and the SCF. In particular, we show statistics with respect to income, wealth, the risky share, and financial wealth to income ratios, which we show in Table 9. As the table illustrates, the resulting distributions and summary statistics are similar in both datasets. Moreover, in the main text, we show the age profiles of stock market participation, and find that both datasets exhibit similar stock market participation patterns.

	Mean	SD	Min	P25	Median	P75	Max
				SCF			
Household income	122378.9	195459.4	380.8253	62653.93	95062.34	137097.1	7.19E+07
Financial wealth	246234.7	1158090	106.0424	8983.125	49769.22	187084.3	6.36E+08
Wealth-to-income	1.635655	47.98129	0.0009804	0.1266945	0.5252401	1.575386	26984.11
Risky share	0.0395985	0.1257449	0	0	0	0	1
Stock ownership	0.2160427	0.4115481	0	0	0	0	1
				PSID			
Household income	95281.69	107221.8	447.9695	52716.41	76858.77	110110.1	4239712
Financial wealth	144018.1	352617.4	105.6337	7650.093	36734.98	136267.4	8214712
Wealth-to-income	1.707796	6.516661	0.0013736	0.1160714	0.4550562	1.527094	321.4286
Risky share	0.0682506	0.1474159	0	0	0	0.0546448	1
Stock ownership	0.3564551	0.4789766	0	0	0	1	1

Table 9: Comparison, PSID vs. SCF

B Estimating the canonical earnings process

We outline the identification and estimation of the canonical earnings process in this appendix.

B.1 Identification

A more formal statement of the assumptions behind the canonical earnings process are the following:

1. $|\rho| < 1$.

2. $\eta_{it} \perp u_{it} \perp \varepsilon_{it}$.

3. $\eta_{it} \sim iidN(0, \sigma_z^2)$, $u_{it} \sim iidN(0, \sigma_u^2)$, $\varepsilon_{it} \sim iidN(0, \sigma_\varepsilon^2)$

Given these assumptions, we can formally identify the parameters of interest in this model from the auto-covariance function alone, following standard arguments. Identification of the parameters requires four periods of data. The arguments for identification are reproduced below. To ease exposition, we assume that $\nu_{it} = \eta_{it} + \varepsilon_{it}$.

First, we can identify ρ from the slope:

$$\begin{aligned}\frac{\text{Cov}(\nu_{i0},\nu_{i3})-\text{Cov}(\nu_{i0},\nu_{i2})}{\text{Cov}(\nu_{i0},\nu_{i2})-\text{Cov}(\nu_{i0},\nu_{i1})} &= \frac{\rho^3\sigma_z^2-\rho^2\sigma_z^2}{\rho^2\sigma_z^2-\rho\sigma_z^2}\\ &= \frac{(\rho^3-\rho^2)(\sigma_z^2)}{(\rho^2-\rho)(\sigma_z^2)}=\rho.\end{aligned}$$

The difference between the covariances allows us to obtain σ_z:

$$\begin{aligned}\text{Cov}(\nu_{i0},\nu_{i2})-\text{Cov}(\nu_{i0},\nu_{i1}) &= \rho^2\sigma_z^2-\rho\sigma_z^2\\ &= (\rho^2-\rho)(\sigma_z^2).\end{aligned}$$

The difference between the variances allows us to obtain σ_u:

$$\begin{aligned}\text{Var}(\nu_{i1})-\text{Var}(\nu_{i0}) &= (\rho\sigma_z^2+\sigma_u^2+\sigma_\varepsilon^2)-(\sigma_z^2+\sigma_\varepsilon^2)\\ &= (\rho-1)\sigma_z^2+\sigma_u^2.\end{aligned}$$

Finally, the variance allows us to identify σ_ε:

$$\text{Var}(\nu_{i0})=\sigma_z^2+\sigma_\varepsilon^2.$$

B.2 Estimation

The standard estimation strategy is to use minimum distance estimation, where the goal is to choose the parameters that minimize the distance between the empirical and theoretical moments. An alternative, which we implement here, is to estimate the parameters via pseudo-maximum likelihood estimation, following Arellano (2003). That is, if $u_i \sim \mathcal{N}(0,\Omega(\theta))$, then the pseudo maximum likelihood estimator of θ solves:

$$\hat{\theta}_{PML}=\arg\min_c\left\{\log\det(\Omega(c))+\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i\Omega(c)^{-1}\hat{u}_i\right\}.$$

This is equivalent to:

$$\hat{\theta}_{PML}=\arg\min_c\left\{\log\det(\Omega(c))+\text{tr}(\Omega(c)^{-1}\hat{\Omega})\right\},$$

where **tr** is the trace of the resulting matrix, and $\hat{\Omega}=\sum\hat{u}_i'\hat{u}_i$. We can then use the asymptotic covariance matrix to compute the standard errors.

The assumptions on the earnings process imply the following moments:

$$\nu_{it} = \rho^{t-1}\eta_{i0} + \sum_{j=2}^{t} \rho^{t-j} u_{ij} + \varepsilon_{it} \tag{18}$$

from which

$$var(\nu_{it}) = \rho^{2(t-1)}\sigma_z^2 + \sum_{j=2}^{t} \rho^{2(t-j)}\sigma_u^2 + \sigma_\varepsilon^2 \tag{19}$$

$$cov(\nu_{it}, \nu_{it-1}) = \rho^{2t-1}\sigma_z^2 + \sum_{j=2}^{t} \rho^{1+2(t-j)}\sigma_u^2 \tag{20}$$

follow, allowing us to identify the moments.

The estimation results are in Table B.2.

Parameter	Estimate (Std. Err.)
Autoregressive component	0.904
	(0.136)
Std. dev. of the initial distribution	0.401
	(0.091)
Std. dev. of the persistent component	0.210
	(0.064)
Std. dev. of the transitory component	0.206
	(0.091)

Table 10: Parameters of the linear AR(1) process. Note: We report the parameter estimates of the linear AR(1) process for earnings. Standard errors are computed via the asymptotic variance calculation. Data from the PSID, 1999 to 2017. All measures are biennial.

C Additional results – earnings process

This section presents additional results in relation to the nonlinear earnings process of Arellano et al. (2017).

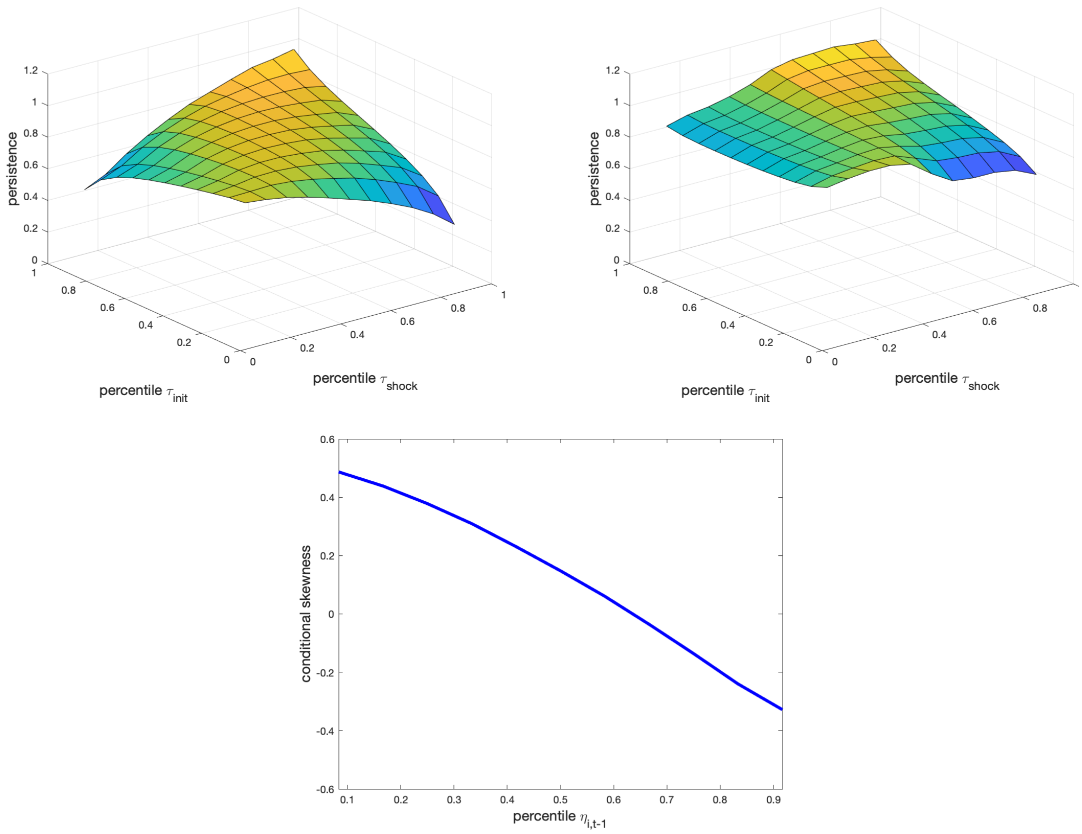

Figure 8: Persistence and conditional skewness in the PSID, nonlinear model. The upper left panel presents the graph of the average derivative of y_{it} given y_{it-1}, with respect to y_{it-1}, which was estimated from a quantile autoregression of y_{it} on a third-order Hermite polynomial on y_{it-1}. The upper right panel presents the persistence from the persistent component of income, η_{it}. The bottom graph presents the conditional skewness of the persistent component of income, η_{it}, from the nonlinear earnings model.

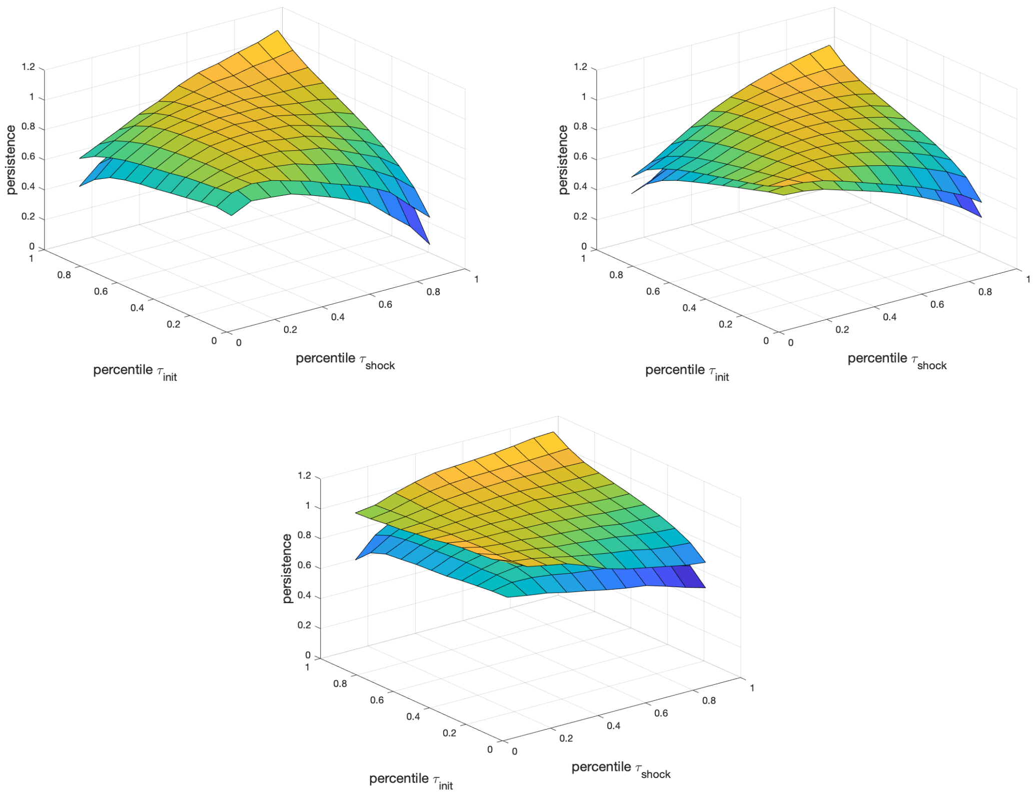

Figure 9: Persistence in the PSID, nonparametric bootstrap. The graphs presented here show the uniform 95% confidence bands calculated from nonparametric bootstraps. The top left panel presents the graph of the average derivative of y_{it} given y_{it-1}, with respect to y_{it-1}, which was estimated from a quantile autoregression of y_{it} on a third-order Hermite polynomial on y_{it-1}. The top right panel presents the average derivative based on simulated data from the nonlinear earnings model. The bottom right graph presents the average derivative of η_{it} given η_{it-1}, with respect to η_{it-1}, based on estimates from the nonlinear earnings model.

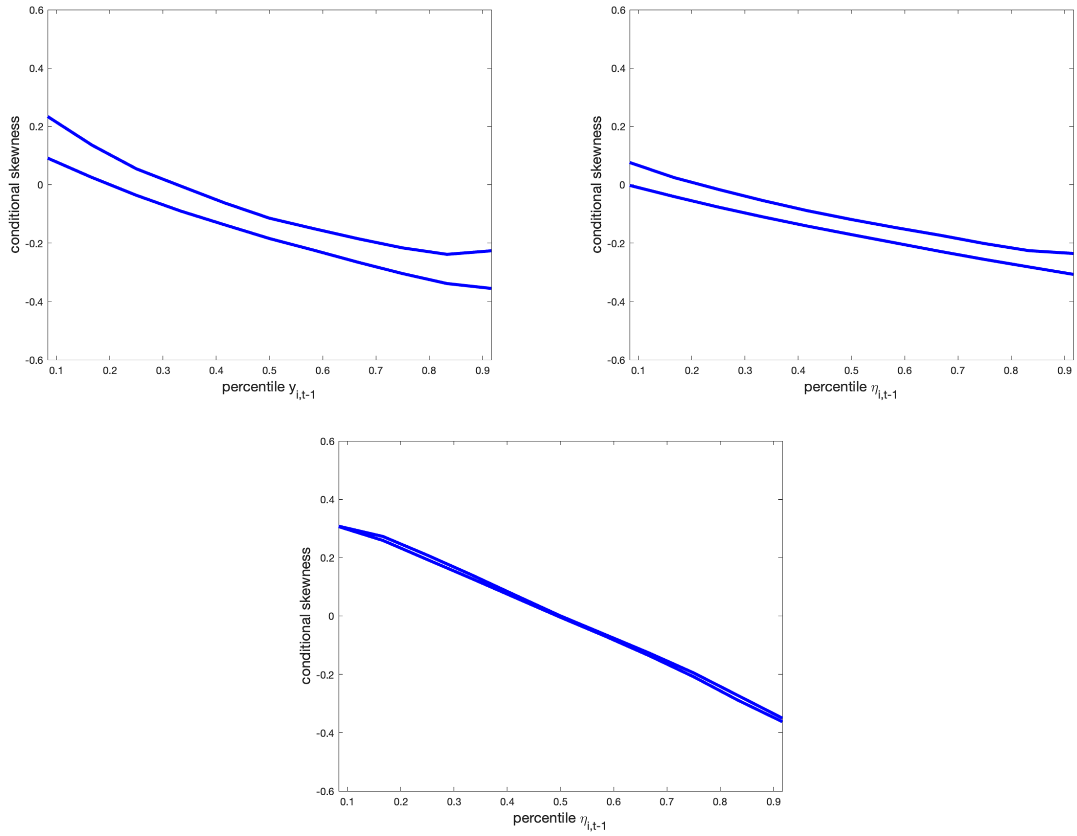

Figure 10: Conditional skewness in the PSID, bootstrap confidence intervals, nonparametric bootstrap. The graphs presented here show the uniform 95% confidence bands. The top left panel presents the graph of the conditional skewness of earnings data y_{it}. The top right panel presents the conditional skewness of earnings simulated from the nonlinear model. The bottom panel presents the conditional skewness of the persistent component η. The graphs were computed via a non-parametric bootstrap with 500 replications.

D Model

D.1 SMM estimation

In our SMM estimation, we pick 4 parameters (risk aversion γ, discount rate β, and participation costs κ^{FC}, κ^{PP}) to match 11 targets in the data (percentage of people that own stocks directly, median financial wealth to income, conditional risky share, and 8 parameters from the OLS regression in Table 2.4).

In order to measure the variability of these moments in the data, we use a bootstrap procedure in which we draw 1,000 samples with replacement from our PSID data. With this, we construct a data variance-covariance matrix S.

We are particularly interested that our model under both earnings processes closely replicates stock market participation, conditional risky shares, and median financial wealth to income ratios. Thus, we choose to increase the weight of these moments in our estimation procedure. Thus, our weighting matrix W is diagonal and is formed of the inverse of the standard deviations of the data moments (for the OLS parameters), 10 (for the wealth to income ratio), and 1000 (participation ratio and risky share). We have experimented with alternative values for these weights and results are very similar.

Finally, we compute a matrix D that measures the responsiveness of our parameter estimates to changes in the moments in the data. We estimate this gradient matrix numerically.

In order to compute the standard errors for our parameter estimates, we compute a variance covariance matrix V determined by (following the notation in De Nardi, French and Jones (2010)):

$$V = (1+\tau)(D'WD)^{-1}D'WSWD(D'WD)^{-1} \tag{21}$$

where τ is the ratio between the number of simulated households in the model and the number of households in the data. Our results do not change if we consider that S has zeros outside of the main diagonal.

E Consumption pass-through

In this section, we discuss the implications of the nonlinear and canonical earnings processes on consumption insurance in the model with portfolio choice. To do so, we estimate semi-structural empirical consumption rules of the form:

$$c_{it} = f_t(\eta_{it}, \varepsilon_{it}, a_{it}, u_{it}), \tag{22}$$

in which c_{it} is log consumption, η_{it} and ε_{it} are the persistent and transitory components of income, a_{it} is log assets, and u_{it} is an unobserved taste shifter. The model allows us to compute consumption insurance coefficients that are a function of age and position in the asset distribution. To see this, we can write average consumption for a given observation of the earnings components and assets as:

$$E(c_{it}|\eta_{it} = \eta, \varepsilon_{it} = \varepsilon, a_{it} = a) = E(f_t(\eta, \varepsilon, a, u_{it})), \tag{23}$$

We can then report the average derivative effect $\phi_t(\eta, \epsilon, a) = E\left(\frac{\partial f_t(\eta,\varepsilon,a,u_{it})}{\partial \eta}\right)$, and, averaging over the earnings components, $\overline{\phi}_t(a) = E(\phi_t(\eta_{it}, \varepsilon_{it}, a))$. The quantity

$$\psi^{\eta} = 1 - \overline{\phi}_t(a)$$

can then be understood as the degree of partial insurance to shocks to the persistent component, as a function of age and assets. Similarly, we can define the same quantity for the transitory component.

Following Arellano et al. (2017), we approximate the consumption function with the following specification:

$$c_{it} = \sum_{k=1}^{K} a_k f_k(\eta_{it}, \varepsilon_{it}, a_{it}, age_{it}) + a_0(\tau), \tag{24}$$

where a_k are piecewise polynomial interpolating splines, and f_k's are dictionaries of functions, which are assumed to be Hermite polynomials. We estimate this model on a simulated panel of households from 25 to 60 years old coming from the economy with the nonlinear earnings process, and the economy with the canonical earnings process. To be consistent with Arellano et al. (2017), we use the same approximating polynomials as their paper.[10] As this is a nonlinear regression model, we estimate the parameter estimates via OLS. Given that we can observe the otherwise latent earnings components, we do not have to resort to a simulation-based estimation algorithm.

We report estimates of the average derivative effect $\overline{\phi}_t(a)$, as a function of age and assets, for both economies. The results show that, on average, the estimated parameter $\overline{\phi}_t(a)$ lies between 0.25 to 0.75, suggesting that on average, around half of post-tax household earnings fluctuations is effectively insured in the economy with the nonlinear earnings process. The equivalent parameter estimates for the economy with the canonical earnings process is around 0.45 to 0.95, which suggests that less than half of earnings is effectively insured. Remarkably, the results for the nonlinear economy mirror the results in Arellano et al. (2017). We also calculate the implied insurance parameters using the estimators proposed by Blundell et al. (2008), and find that the insurance parameters for the persistent component are 0.56 for the nonlinear process, and 0.18 for the canonical process.

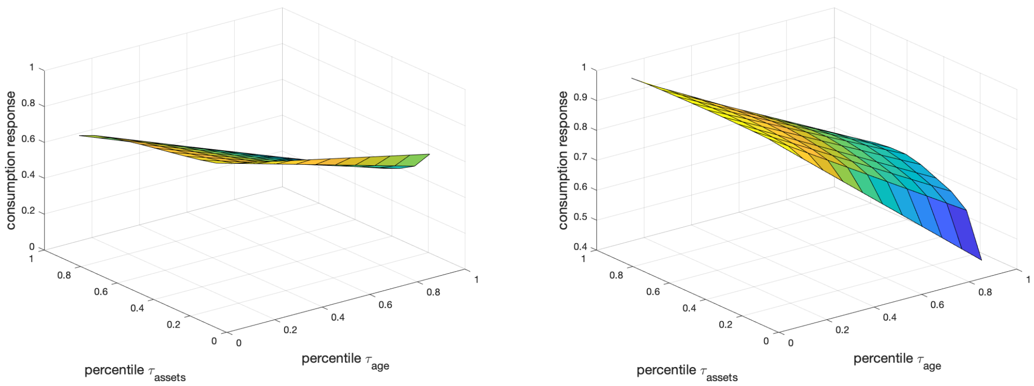

Figure 11: Consumption response to earnings shocks, nonlinear vs. linear model. Note: The graphs presented here show the average derivative effect of η_{it} on c_{it}, computed at percentiles of a_{it} and age_{it}. Data simulated from structural model of life cycle portfolio choice with the nonlinear earnings process (left) and the canonical earnings process (right).

We also compute the corresponding parameter estimates the average derivative effect with respect to assets, which we show in Figure 12. The estimated average derivative effects for the economy with the nonlinear earnings process are from 0.2 to 0.7, and

[10] This is (2,1,2,1), where the order is (persistent,transitory,wealth,age).

are shown to be increasing in age and assets, consistent with the results in Arellano et al. (2017). The implied derivative effects for the economy with the canonical earnings process, meanwhile, are from -0.02 to 0.15, and are decreasing in age.

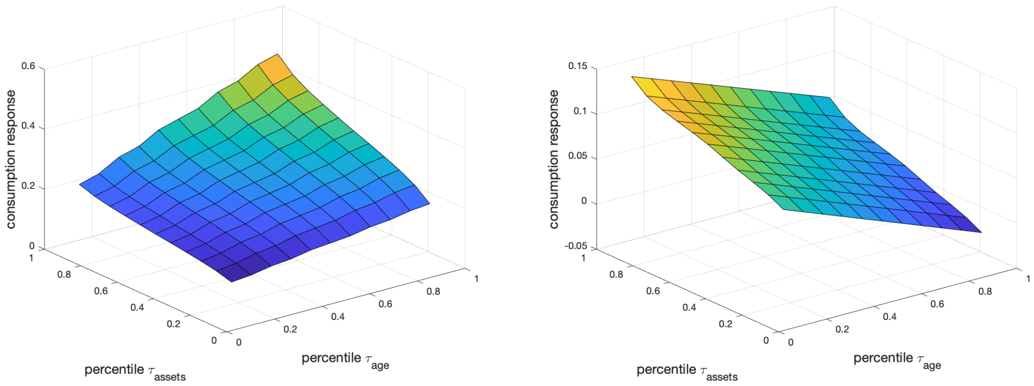

Figure 12: Consumption response to assets, nonlinear vs. linear model. Note: The graphs presented here show the average derivative effect of a_{it} on c_{it}, computed at percentiles of a_{it} and age_{it}. Data simulated from structural model of life cycle portfolio choice with the nonlinear earnings process (left) and the canonical earnings process (right).

We finally compute the average derivative effects for non-stockholders and stockholders, in the case of the economy with the nonlinear earnings process. The results that we obtain are in Figure 13. We find that the derivative effects with respect to consumption for non-stockholders is higher than that of stockholders. These results imply that stockholders are able to effectively insure their consumption with respect to shocks to persistent income. The MPC's with respect to wealth are also higher for households who own stocks than for non-stockholders.

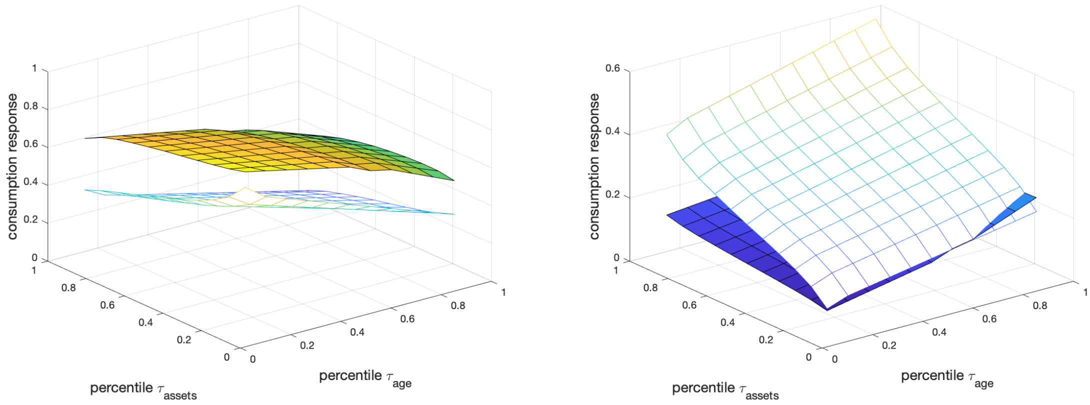

Figure 13: Consumption response to earnings shocks, stockholders vs. non-stockholders. Note: The left panel graphs show the average derivative effect of η_{it} on c_{it} (left), computed at percentiles of a_{it} and age_{it} for non-stockholders (surface graph) and stockholders (mesh graph). The right panel graphs show the average derivative effect of a_{it} on c_{it} (left), computed at percentiles of a_{it} and age_{it} for non-stockholders (surface graph) and stockholders (mesh graph). Data simulated from structural model of life cycle portfolio choice with the nonlinear earnings process.

F Robustness

F.1 Disaster risk

Model	Case with disaster risk	
	NL	CA
γ	4.981	9.2759
β	0.83193	0.78473
κ^{FC}	0	0.0125
κ^{PP}	0.05	0.025

Table 11: Parameter estimates. β is expressed in annual terms. The participation costs are expressed as fractions of average household income (the numeraire in the model)

Model		Case with disaster risk	
Moment	Data	NL	CA
Med W/I	0.35	0.56328	0.099254
Participation	0.32	0.31615	0.38757
OLS constant	-0.492 (2.474)	-3.8759	16.242
OLS partic	0.432 (0.011)	0.75898	0.45269
OLS age	0.0288 (0.234)	0.28271	-0.97776
OLS age2	-0.00105 (0.008)	-0.0071807	0.015103
OLS age3	9.96e-6 (1.28e-4)	7.2168e-05	3.3624e-05
OLS age4	-2.06e-8 (7.25e-7)	-2.0588e-07	-1.4634e-06
OLS log income	0.0286	0.20979	0.65952
OLS log wealth	0.0458	0.0020205	0.0034425
Risky share	0.60	0.59705	0.63119

Table 12: Targeted parameters

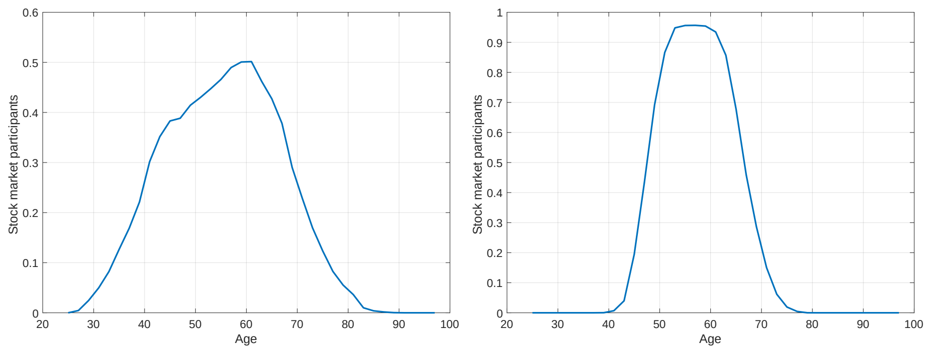

Figure 14: Stock market participation over the life cycle. Case with disaster risk (left: NL; right: canonical)

F.2 Only per-period or only entry costs

Model	Risky share, only PP		Risky share, only FC	
Parameter	NL	CA	NL	CA
γ	5.64	10.87	5.64	8.99
β	0.81117	0.74162	0.80623	0.74673
κ^{FC}	0	0	0.17	0.0814
κ^{PP}	0.0563	0.02	0	0

Table 13: Parameter estimates. β is expressed in annual terms. The participation costs are expressed as fractions of average household income (the numeraire in the model)

Model		Risky share, only PP		Risky share, only FC	
Moment	Data	NL	CA	NL	CA
Med W/I	0.35	0.6158	0.092308	0.61759	0.084513
Participation	0.32	0.35171	0.41528	0.50356	0.42001
OLS constant	-0.492 (2.474)	-5.1892	26.475	-5.0502	36.82
OLS partic	0.432 (0.011)	0.75924	0.39213	0.81841	0.36699
OLS age	0.0288 (0.234)	0.38465	-1.8718	0.34894	-2.7742
OLS age2	-0.00105 (0.008)	-0.0099974	0.043817	-0.0079468	0.072776
OLS age3	9.96e-6 (1.28e-4)	0.00010501	-0.00036682	6.4793e-05	-0.0007706
OLS age4	-2.06e-8 (7.25e-7)	-3.4043e-07	5.9224e-07	-8.5688e-08	2.6596e-06
OLS log income	0.0286	0.21063	0.58235	0.11535	0.51634
OLS log wealth	0.0458	0.0024687	0.0051995	0.010087	0.0076908
Risky share	0.60	0.60722	0.6854	0.6323	0.74061

Table 14: Targeted parameters

F.3 Case without targeting risky share

Model	Baseline		Without risky share	
Parameter	NL	CA	NL	CA
γ	5.64	8.82	0.55	1.1
β	0.81117	0.81117	0.90105	0.91652
κ^{FC}	0	0.0142	0	0.0091
κ^{PP}	0.0563	0.045	0	0

Table 15: Parameter estimates. β is expressed in annual terms. The participation costs are expressed as fractions of average household income (the numeraire in the model)

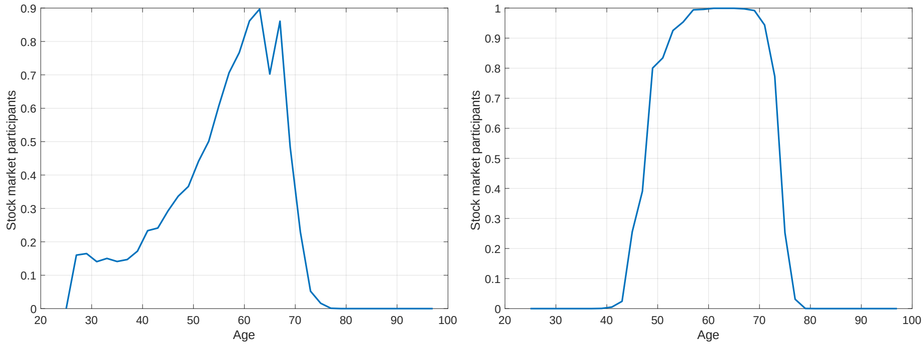

Figure 15: Stock market participation over the life cycle. Left: NL process, right: canonical process.

Model		Baseline		Without risky share	
Moment	Data	NL	CA	NL	CA
Med W/I	0.35	0.62083	0.11638	0.26801	0.079028
Participation	0.32	0.35575	0.34414	0.34135	0.39585
OLS constant	-0.492 (2.474)	-7.714	5.6635	-4.5765	29.473
OLS partic	0.432 (0.011)	0.75693	0.5436	0.49759	0.43368
OLS age	0.0288 (0.234)	0.60866	-0.054792	0.38865	-2.1838
OLS age2	-0.00105 (0.008)	-0.017322	-0.014545	-0.011953	0.055796
OLS age3	9.96e-6 (1.28e-4)	0.00020987	0.0004491	0.00015878	-0.00056395
OLS age4	-2.06e-8 (7.25e-7)	-8.9583e-07	-3.623e-06	-7.3336e-07	1.764e-06
OLS log income	0.0286	0.21117	0.70433	0.044904	0.46848
OLS log wealth	0.0458	0.002837	0.00099915	0.029351	0.010393
Risky share	0.60	0.60862	0.66384	0.86939	0.84445

Table 16: Targeted parameters

F.4 Main case, possible alternative parametrizations

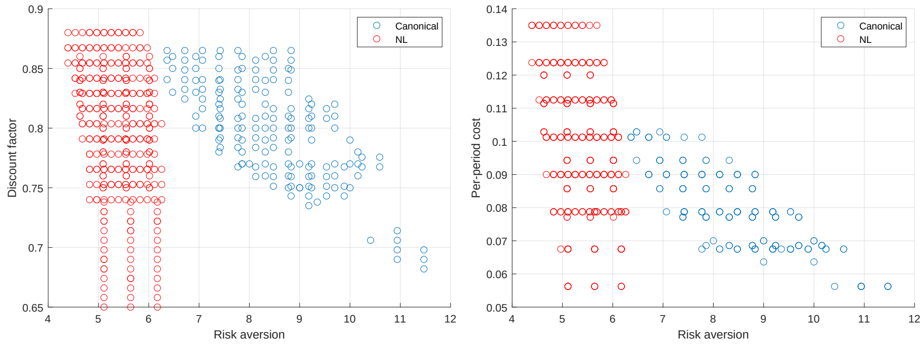

Figure 16: Parametrizations that yield participation shares between 30 and 40%, conditional risky shares between 50 and 70%, and median wealth above 0.25, given zero fixed entry costs to the stock market. NL vs canonical process

F.5 Empirical initial wealth

In this section, we assume that households, instead of starting out life with zero wealth, they do so with a draw from the empirical wealth distribution of the 2016 SCF (20-30 year olds). For simplicity, we assume that this initial wealth draw is uncorrelated with initial income.

Model	Baseline	
Parameter	NL	CA
γ	5.5411	8.1628
β	0.79668	0.79668
κ^{FC}	0.05	0.0625
κ^{PP}	0.0375	0.025

Table 17: Parameter estimates. β is expressed in annual terms. The participation costs are expressed as fractions of average household income (the numeraire in the model)

Model		Baseline	
Moment	Data	NL	CA
Med W/I	0.35	0.60429	0.10065
Participation	0.32	0.34035	0.3709
OLS constant	-0.492 (2.474)	-3.5008	2.9142
OLS partic	0.432 (0.011)	0.84375	0.52762
OLS age	0.0288 (0.234)	0.26182	0.19368
OLS age2	-0.00105 (0.008)	-0.0068823	-0.022781
OLS age3	9.96e-6 (1.28e-4)	7.3702e-05	0.00057005
OLS age4	-2.06e-8 (7.25e-7)	-2.4971e-07	-4.2834e-06
OLS log income	0.0286	0.14421	0.59737
OLS log wealth	0.0458	0.0033967	0.0035551
Risky share	0.60	0.60094	0.59874

Table 18: Targeted parameters

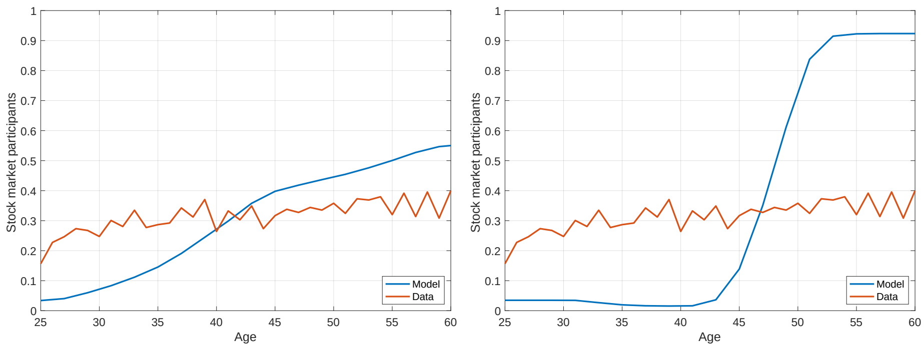

Figure 17: Stock market participation over the life cycle. Left: NL process, right: canonical process.

F.6 Empirical initial wealth and stock market participation

In this section, we assume that households, instead of starting out life with zero wealth, they do so with a draw from the empirical wealth distribution of the 2016 SCF (20-30 year olds). For simplicity, we assume that this initial wealth draw is uncorrelated with initial income.

Here, we also assume that some households start their working lifes already being stock market participants. Given that stock market participation at age 25 is 15.5% in our data, we assume that the richest 15.5% initial households start out by having some of their wealth in stocks and thus must not pay the stock entry cost if they wish to continue investing in the risky asset.

Model	Baseline	
Parameter	NL	CA
γ	4.981	7.2741
β	0.80784	0.79668
κ^{FC}	0.07	0
κ^{PP}	0.025	0.0125

Table 19: Parameter estimates. β is expressed in annual terms. The participation costs are expressed as fractions of average household income (the numeraire in the model)

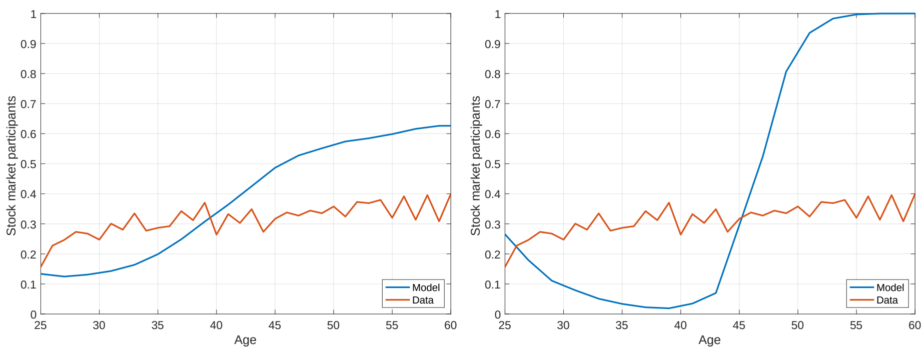

Figure 18: Stock market participation over the life cycle. Left: NL process, right: canonical process.

Model		Baseline	
Moment	Data	NL	CA
Med W/I	0.35	0.55001	0.092308
Participation	0.32	0.42151	0.44076
OLS constant	-0.492 (2.474)	2.8739	41.63
OLS partic	0.432 (0.011)	0.83736	0.43612
OLS age	0.0288 (0.234)	-0.31948	-3.2824
OLS age2	-0.00105 (0.008)	0.012643	0.092336
OLS age3	9.96e-6 (1.28e-4)	-0.00021173	-0.0010944
OLS age4	-2.06e-8 (7.25e-7)	1.284e-06	4.6018e-06
OLS log income	0.0286	0.13151	0.4928
OLS log wealth	0.0458	0.0061351	0.0075691
Risky share	0.60	0.69436	0.68547

Table 20: Targeted parameters

F.7 Alternative pension system

In this section, we assume that, instead of a flat pension, households receive a replacement rate of the last realization of their earnings that is consistent with U.S. data. Namely, we follow Kaplan and Violante (2010) and assume that there is a 90% replacement rate for any earnings below 18% of the average, 32% for the earnings between 18% and 110% of the average, and 15% for the remainder. For simplicity, we only apply these replacement rates to the persistent component of earnings.

This different pension system affects the incentives to save, and thus implies a lower calibrated discount factor for both the NL and canonical process. However, the implications in terms of a lower coefficient of risk aversion and a more reasonable profile of stock market participation for the NL process still hold true.

Model	Baseline	
Parameter	NL	CA
γ	4.9	10.362
β	0.74162	0.74733
κ^{FC}	0	0.05
κ^{PP}	0.05	0.025

Table 21: Parameter estimates. β is expressed in annual terms. The participation costs are expressed as fractions of average household income (the numeraire in the model)

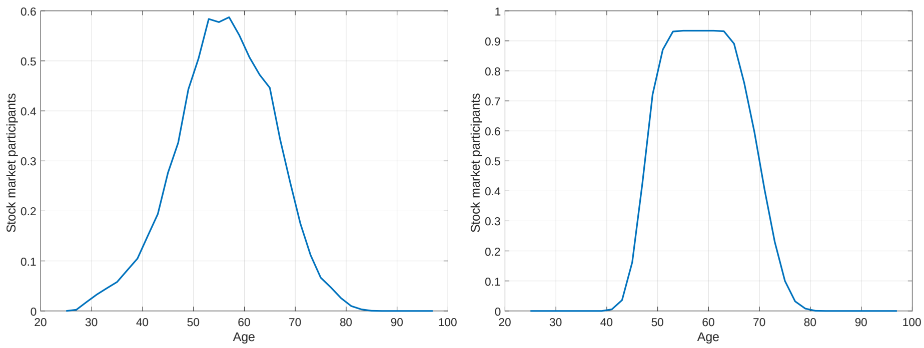

Figure 19: Stock market participation over the life cycle. Left: NL process, right: canonical process.

Model		Baseline	
Moment	Data	NL	CA
Med W/I	0.35	0.42985	0.096833
Participation	0.32	0.29373	0.38149
OLS constant	-0.492 (2.474)	6.1606	21.646
OLS partic	0.432 (0.011)	0.72255	0.46778
OLS age	0.0288 (0.234)	-0.46166	-1.4625
OLS age2	-0.00105 (0.008)	0.011743	0.031197
OLS age3	9.96e-6 (1.28e-4)	-0.00011348	-0.00020064
OLS age4	-2.06e-8 (7.25e-7)	2.9798e-07	-2.0358e-07
OLS log income	0.0286	0.21241	0.66139
OLS log wealth	0.0458	0.0026219	0.0030189
Risky share	0.60	0.70001	0.66091

Table 22: Targeted parameters

F.8 Correlation (0.20) between labor income and stock market returns

Model	Positive correlation (0.20)		Negative correlation (-0.20)	
Parameter	NL	CA	NL	CA
γ	5.4666	9.3507	5.7093	9.5934
β	0.83193	0.78307	0.79969	0.7746
κ^{FC}	0	0.04	0	0.04
κ^{PP}	0.0625	0.025	0.0625	0.025

Table 23: Parameter estimates

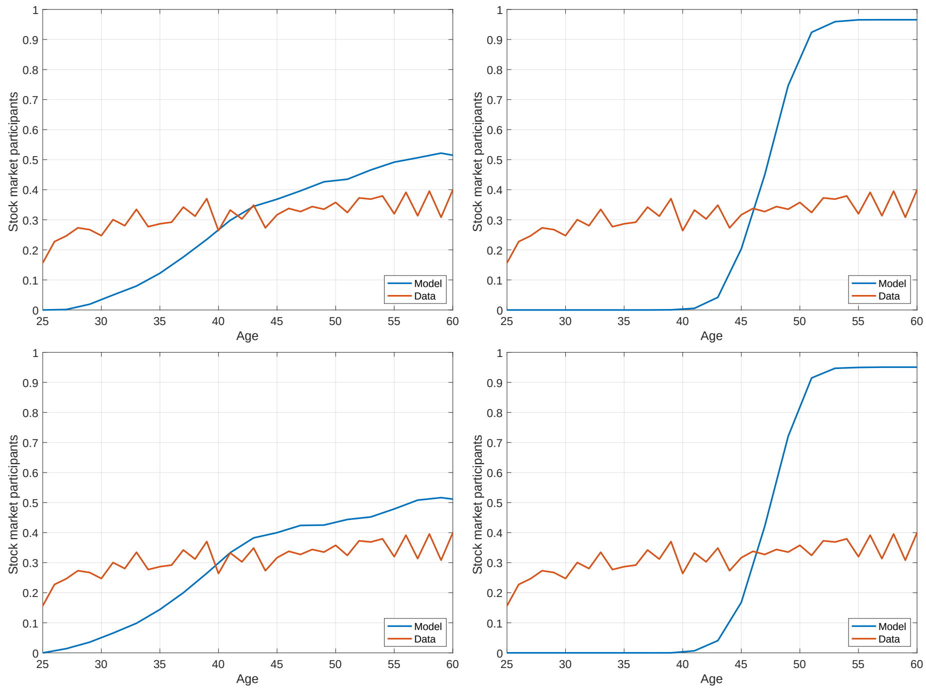

Figure 20: Stock market participation over the life cycle. Left: NL process, right: canonical process. Top: positive correlation; bottom: negative correlation

F.9 Case in which both processes match median wealth to income more precisely

As argued in the main text, our baseline parametrization for the NL process slightly overestimates wealth accumulation in the economy, whilst the canonical process underestimates the total amount of wealth. In this section, we show that our results in terms of life-cycle profiles (e.g., the earlier accumulation of wealth and stock market participation in the NL process and the strong age profile in the canonical process, where almost everyone starts participating very late in life) are not driven by this feature of the calibration. To do so, we pick the parametrization that minimizes our objective criteria whilst requiring it to generate a median wealth to income ratio between 0.3 and 0.5 (vs. 0.35 in the data). The tables below show the associated parameters, targets, stock market participation profile over the life cycle, and wealth accumulation patterns.

Model	Baseline	
Parameter	NL	CA
γ	4.6642	5.5526
β	0.74675	0.91652
κ^{FC}	0.01162	0.01071
κ^{PP}	0.03	0.10286

Table 24: Parameter estimates. β is expressed in annual terms. The participation costs are expressed as fractions of average household income (the numeraire in the model)

Model		Baseline	
Moment	Data	NL	CA
Med W/I	0.35	0.39193	0.30637
Participation	0.32	0.30912	0.28384
OLS constant	-0.492 (2.474)	10.03	17.252
OLS partic	0.432 (0.011)	0.79034	0.66434
OLS age	0.0288 (0.234)	-0.86356	-1.1981
OLS age2	-0.00105 (0.008)	0.027095	0.026768
OLS age3	9.96e-6 (1.28e-4)	-0.00036727	-0.00019902
OLS age4	-2.06e-8 (7.25e-7)	1.8246e-06	8.5682e-08
OLS log income	0.0286	0.17586	0.76839
OLS log wealth	0.0458	0.0029703	-0.0029158
Risky share	0.60	0.77098	0.7936

Table 25: Targeted parameters

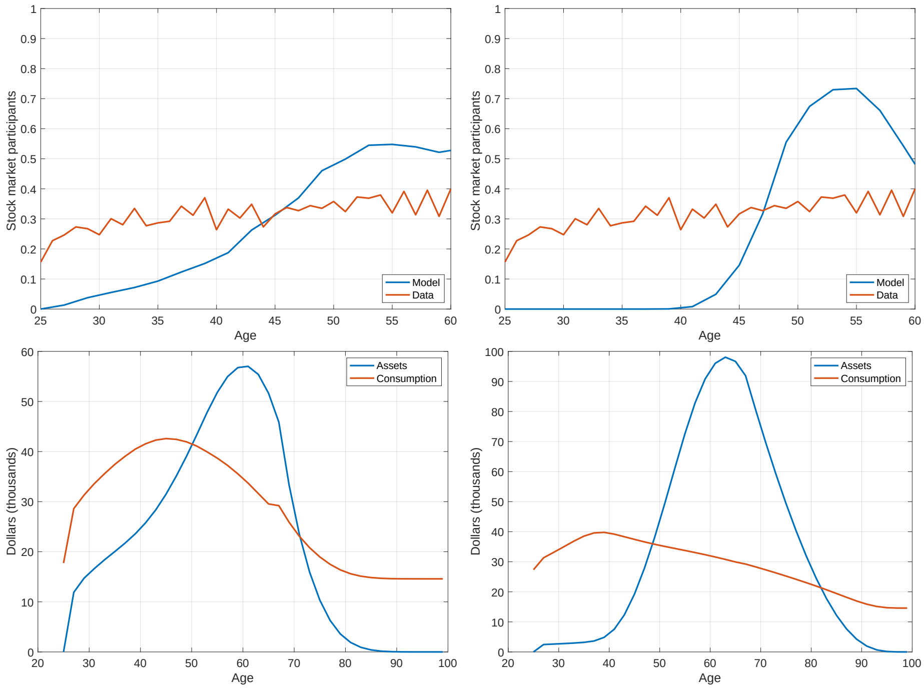

Figure 21: Stock market participation over the life cycle. Left: NL process, right: canonical process.

G Portfolio choice under the presence of housing

This section provides a description of the life-cycle structural model of portfolio choice with an endogenous housing decision. We also discuss the moments that we use for the structural estimation.

G.1 Model

In this subsection, we describe the structural model that we take to the data.

Demographics. Agents start working life at 25, face age-dependent positive death probabilities, and die with certainty at age 100. The model period is two years and our unit of observation is the household.

Preferences. Households maximize:

$$\max E_t \left[\sum_{t=0}^{t=T} \beta^t \mathcal{S}_t \frac{[c_t \phi(h_t)]^{1-\gamma}}{1-\gamma} \right] \tag{25}$$

where c is nondurable consumption, γ is the coefficient of relative risk aversion, β is the discount factor, $\phi(h_t)$ is a function of the current housing status, normalized such that $\phi(0) = 1$ and $\phi(1) = u_h$, and $\mathcal{S}_t$ is the probability of arriving alive at time t.

Earnings process. As in the main text, the stochastic component of earnings can be decomposed to a persistent and transitory component (Equation 1), and we use alternatively a canonical linear specification and a non-linear, non-normal specification for both components of the earnings process. There is no earnings risk after retirement (age 65), from which households get a public pension.

Budget constraint. Households can save in two types of financial assets and in housing:

$$c_{t+1} + s_{t+1} + a_{t+1} + \kappa^f(I_{t+1}, I_t) + p_h h_{t+1} + \kappa^h(h_{t+1}, h_t) + r^h I(h_t = 0) + m_{t+1} = \tag{26}$$

$$(1 + r^s_{t+1})s_t + (1 + r)a_t + y_{t+1} + p_h h_t + (1 + r^m)m_t \tag{27}$$

where s_t is the amount of wealth invested in the risky asset and a_t the amount of wealth invested in the risk-free asset at time t. r^s_{t+1} represents the risky return of stocks (which is i.i.d.), while r is the risk-free rate. κ^f represents potential costs of participation in the stock market, which depend on the households' stock market participation status I_t. We define

$$I_t = (s_t > 0). \tag{28}$$

These may be per-period participation costs (just dependent on I_{t+1}), fixed participation costs (only paid if $I_t = 0$ and $I_{t+1} = 1$, and zero if $I_t = 1$) or a combination of both.

p_h represents the (fixed) price of housing, $\kappa^h(h_{t+1}, h_t)$ represents the transaction costs associated with the purchase of the house, which we model as a fixed fraction of the house price for both buyer and seller, and r^h represents the rental rate for those who do not own a house, which is a fixed fraction of house prices. m_t represents the outstanding mortgage balance ($m_t \leq 0$) and r^m the associated mortgage balance.

Households can borrow in order to buy a house. Borrowing constraints are thus:

$$a_{t+1} \geq 0, s_{t+1} \geq 0 \tag{29}$$

for renters and

$$a_{t+1} \geq 0, s_{t+1} \geq 0, m_{t+1} \geq -\lambda_h p_h h_{t+1} \tag{30}$$

for homeowners.

For simplicity, we assume that the borrowing constraint on mortgages is always binding $m_{t+1} = -\lambda_h p_h h_{t+1}$ (as in Vestman (2019)).

Finally, we define

$$x_t = (1 + r^s_{t+1})s_t + (1 + r)a_t + (1 + r^m)m_t \tag{31}$$

as the amount of cash-on-hand that an individual has at the beginning of period t.

Households' problem Households thus solve the following problem:

$$V_t(x_t, y_t, I_t, h_t) = \max_{c_t, a_{t+1}, s_{t+1}, h_{t+1}} \left\{ \frac{(c_t \phi(h_t))^{1-\gamma}}{1-\gamma} + \beta \frac{\mathcal{S}_t}{\mathcal{S}_{t-1}} E_t V_{t+1}(x_{t+1}, y_{t+1}, I_{t+1}, h_{t+1}) \right\} \tag{32}$$

subject to the budget constraint (27) and the borrowing and short-sale constraints (29) and (30), and where the expectation E_t is taken with respect to future realizations of persistent income, transitory income, and stock market returns.

G.2 Targeted moments and model fit

All externally calibrated parameters are identical to those in the main version of the model. Additionally, we assume that the house price is fixed at $p^h = 4$ years of average disposable household income, transaction costs are $k^h(h_{t+1} = 1, h_t = 0) = k^h(h_{t+1} = 0, h_t = 1) = 0.05p^h$, annual rental costs are 2.5% of the house price, and the interest rate on mortgages is $r^m = r^a = 2\%$ annually.

We estimate the life-cycle model of portfolio choice under the presence of housing via SMM. The difference between the model we present here and the baseline model, is that we target additional moments that are informative of the homeownership status of households. Specifically, we target the following additional moments: homeownership rate (0.76), the median total wealth-to-income ratio (1.27), and the parameters of an OLS regression which now includes a dummy for homeownership, following Yao and Zhang (2005). We report the results of these regressions below in Table 26. The results that we obtain from the estimations confirm our previous results that past participation, income and wealth are important determinants of participation.

Table 26: Determinants of portfolio choice with housing

	(1)	(2)	(3)
VARIABLES	OLS	Probit	CRE Probit
Past participation	0.491***	1.382***	1.042***
	(0.00814)	(0.0266)	(0.0558)
Assets (in logs)	0.0401***	0.327***	0.418***
	(0.00174)	(0.0124)	(0.0193)
Income (in logs)	0.0309***	0.120***	0.156***
	(0.00353)	(0.0190)	(0.0258)
Homeownership	-0.0503***	-0.185***	-0.209***
	(0.00723)	(0.0451)	(0.0567)
Constant	-1.277	-9.838	-8.114
	(1.425)	(7.310)	(7.746)
Observations	17,987	17,982	17,982
R-squared	0.398		
Number of person			4,211

Model fit. The fit of the model with respect to the targeted moments is below in Table 27. As can be observed, both processes fit the data well, with the non-linear earnings process fitting the homeownership rate better than the canonical process.

Model		Baseline	
Moment	Data	NL	Ca
Participation	0.32	0.2934	0.37642
Risky share	0.60	0.6467	0.67011
Med total W/I	1.27	1.6713	1.039
OLS constant	-0.492 (2.474)	-4.3281	0.73367
OLS partic	0.491 (0.008)	0.68899	0.56212
OLS age	0.0699 (0.139)	0.41358	0.070539
OLS age2	-0.00235 (0.00497)	-0.014718	-0.0074202
OLS age3	3.07e-5 (7.75e-5)	0.0002323	0.00018509
OLS age4	-1.42e-7 (4.45e-7)	-1.3572e-06	-1.4219e-06
OLS log income	0.0309 (0.00353)	0.20522	0.28195
OLS log wealth	0.0401 (0.00174)	0.025285	0.0048599
OLS homeownership	-0.0503 (0.007)	-0.039296	0.34376
Homeownership rate	0.76	0.77137	0.49091

Table 27: Targeted parameters for the portfolio choice model under housing

G.3 Homeownership and life-cycle profiles of portfolio choice

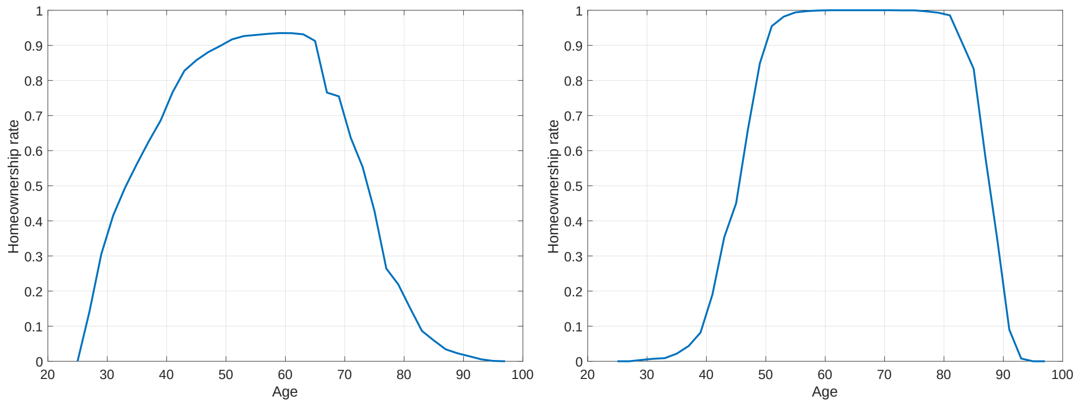

Figure 22: Homeownership rates by age (left: NL; right: CA)

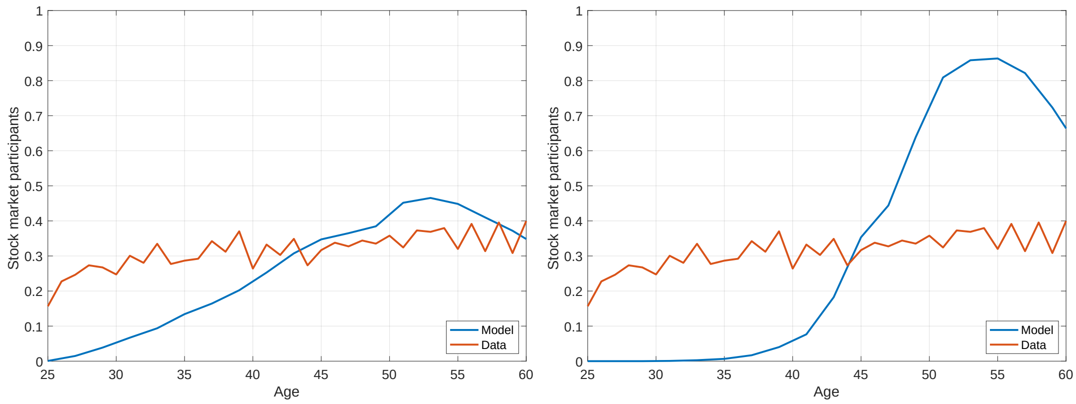

Figure 23: Stock market participation by age (left: NL; right: CA).

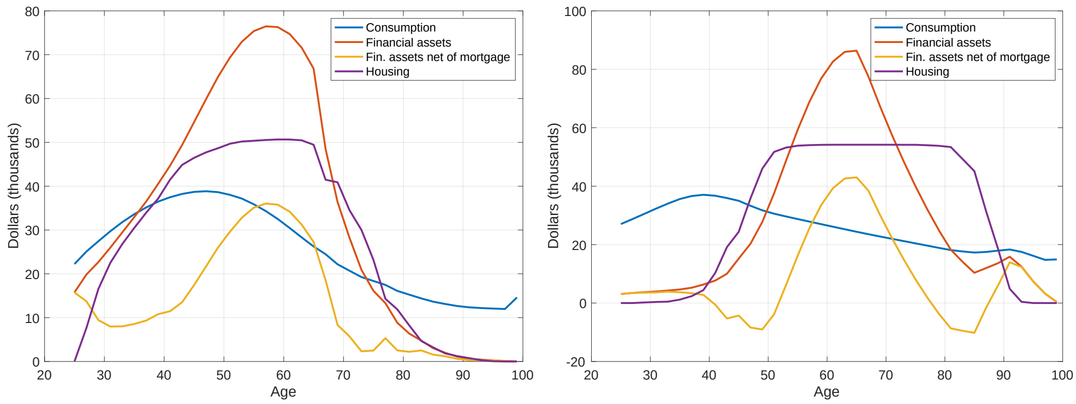

Figure 24: Asset accumulation and consumption over the life cycle (left: NL; right: CA).

G.4 Consumption pass-throughs

We also estimate the consumption pass-throughs implied by the model with housing. As in Appendix E, we estimate the semi-structural consumption policy function (22) specified in Arellano et al. (2017). The main difference between the estimation in this exercise and that in the baseline model is that a_{it} is now defined as the sum of financial and housing wealth, which is closer to the definition in ABB.

As we show in Figure 25, the results are quite similar to the baseline model. We estimate a marginal propensity to consume out of persistent income η_{it} which ranges from 0.3-0.8; the resulting consumption surface also is similar to the results in Arellano et al. (2017). The canonical model, on the other hand, results in a widely different estimate of the MPC. We also compute the associated Blundell et al. (2008) coefficients and find that the associated BPP estimate is 0.39, much closer to the BPP estimate of 0.36. The canonical model, meanwhile, implies a BPP estimate of 0.10. We also estimate the marginal propensity to consume out of assets, and as Figure 26 illustrates, the economy with the nonlinear earnings process gets closer estimates to Arellano et al. (2017) than the canonical one.

Finally, we find in Figure 27 that stockholders are able to insure their consumption against income shocks, similar to what we found in the baseline model.

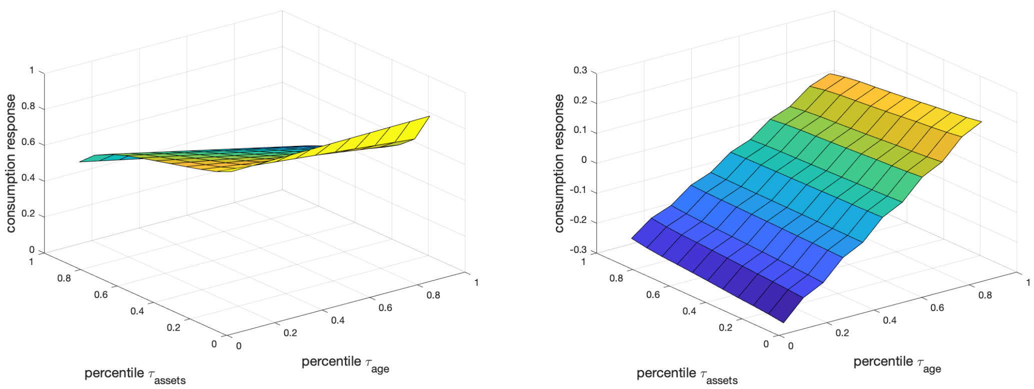

Figure 25: MPCs with respect to persistent component of income η_{it}, model with housing, evaluated at percentiles of assets τ_{assets} and age τ_{age}. Left: nonlinear earnings process. Right: canonical earnings process.

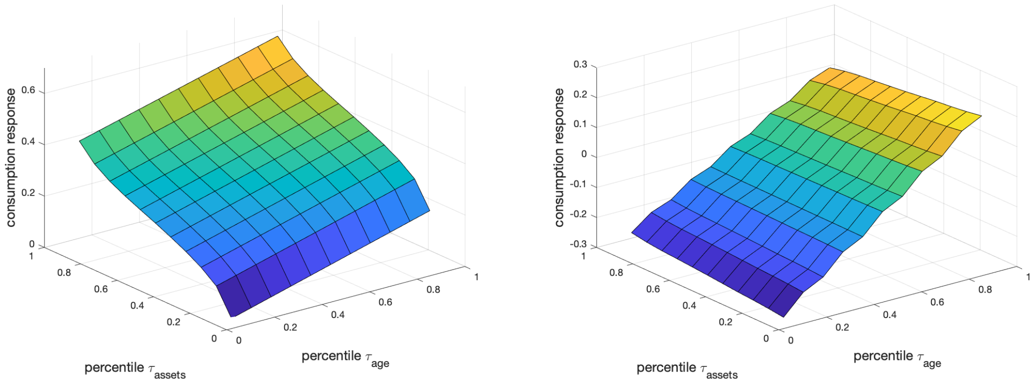

Figure 26: MPCs with respect to assets, model with housing, evaluated at percentiles of assets τ_{assets} and age τ_{age}. Left: nonlinear economy. Right: canonical earnings process.

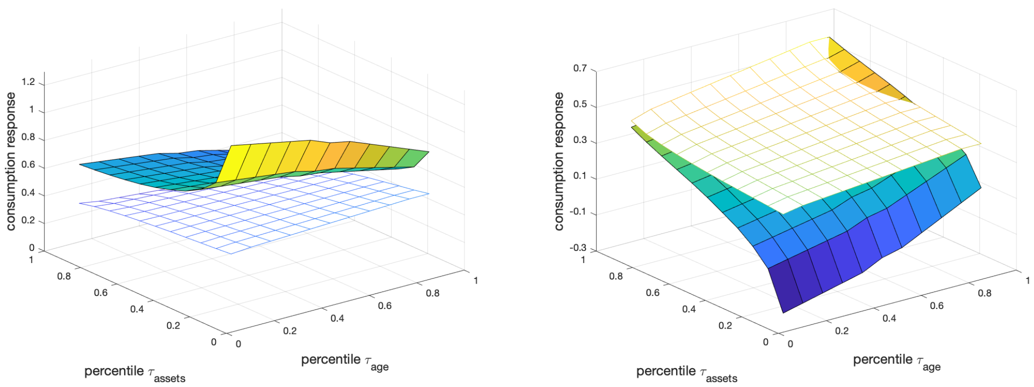

Figure 27: MPCs with respect to persistent component of income η_{it}, stockholders (mesh) vs. non-stockholders (surface plot), model with housing, evaluated at percentiles of assets τ_{assets} and age τ_{age}. Left: nonlinear economy. Right: canonical earnings process.