HOUSEHOLD PORTFOLIO CHOICES UNDER (NON-)LINEAR INCOME RISK: AN EMPIRICAL FRAMEWORK

2023

BANCO DE ESPAÑA
Eurosistema

Documentos de Trabajo
N.º 2327

Julio Gálvez

HOUSEHOLD PORTFOLIO CHOICES UNDER (NON-)LINEAR INCOME RISK: AN EMPIRICAL FRAMEWORK

HOUSEHOLD PORTFOLIO CHOICES UNDER (NON-)LINEAR INCOME RISK: AN EMPIRICAL FRAMEWORK

Julio Gálvez [(*)]

BANCO DE ESPAÑA

(*) Contact information: Address: Financial Analysis Division, DG-Economics, Statistics and Research, Banco de España, Alcalá, 48, Madrid 28014, Spain. Phone: +34 91 338 8444. E-mail: julio.galvez@bde.es.

https://doi.org/10.53479/33792

Documentos de Trabajo. N.º 2327

September 2023

ISSN: 1579-8666 (on line)

Abstract

This paper develops a flexible, semi-structural framework to empirically quantify the non-linear transmission of income shocks to household portfolio choice decisions both at the extensive and intensive margins. I model stock market participation and portfolio allocation rules as age-dependent functions of persistent and transitory earnings components, wealth and unobserved taste shifters. I establish non-parametric identification and propose a tractable, simulation-based estimation algorithm, building on recent developments in the sample selection literature. Using recent waves of PSID data, I find heterogeneous income and wealth effects on both extensive and intensive margins, over the wealth and life-cycle dimensions. These results suggest that preferences are heterogeneous across the wealth distribution and over the life cycle. Moreover, in impulse response exercises, I find sizeable extensive margin responses to persistent income shocks. Finally, I find heterogeneity in participation costs across households in the wealth distribution.

Keywords: stock market participation, non-linear income persistence, sample selection, quantile selection models, latent variables.

JEL classification: C23, C24, D31, G50, J24.

Resumen

Este trabajo desarrolla un método flexible y semiestructural para cuantificar empíricamente la transmisión no lineal de los *shocks* de renta a las decisiones de elección de cartera de los hogares, tanto en el margen extensivo como en el intensivo. Modelizo la participación en el mercado de valores y las reglas de asignación de carteras como funciones dependientes de la edad de los componentes persistentes y transitorios de los ingresos, la riqueza y los cambios de gusto no observados. Establezco una identificación no paramétrica y propongo un algoritmo de estimación basado en la simulación, fundado en los recientes desarrollos de la literatura de selección de muestras. Utilizando olas recientes de datos de la PSID, observo efectos heterogéneos sobre la renta y la riqueza en los márgenes extensivo e intensivo, en las dimensiones de riqueza y ciclo vital. Estos resultados sugieren que las preferencias son heterogéneas a lo largo de la distribución de la riqueza y del ciclo vital. Además, en los ejercicios de respuesta al impulso, encuentro respuestas considerables del margen extensivo a las perturbaciones persistentes de la renta. Por último, advierto heterogeneidad en los costes de participación de los hogares en la distribución de la riqueza.

Palabras clave: participación en el mercado de valores, persistencia no lineal de los ingresos, selección de la muestra, modelos de selección cuantílica, variables latentes.

Códigos JEL: C23, C24, D31, G50, J24.

1 Introduction

Households invest in financial assets, such as stocks, to transfer wealth across periods and to pool risks, with the goal of smoothing consumption. When they make their investment decisions, however, households encounter various idiosyncratic and aggregate risks. The primary and most important source of idiosyncratic risk that they face, which they can neither avoid nor fully insure themselves against, is on their labor income. As households experience unique earnings histories, their savings decisions may differ depending on the size and durability of the shocks they receive. In this paper, I empirically assess the impact of earnings shocks on household stock market participation and portfolio choices by developing a novel semi-structural framework, and find sizeable responses to large income shocks.

An extensive literature in macroeconomics and finance has studied how uninsurable labor income shocks affect household consumption, saving, and portfolio allocation decisions over the life cycle, as well as its impact on asset prices (e.g., Gourinchas and Parker (2002) and Storesletten et al. (2004b) in consumption, Constantinides and Duffie (1996) in asset pricing, and Cocco et al. (2005) in portfolio choice). In these models, households accumulate precautionary savings to smooth their consumption against adverse labor market events. Moreover, they may reduce their exposure to avoidable risks by lowering the amount of their wealth invested in equities. The margin of these adjustments, however, depends on the precise nature of earnings dynamics. Previous literature has relied on linear earnings processes as a workhorse model to analyze these decisions. A consensus that has emerged from a wide body of empirical work is that the effect of labor income risk on portfolio allocations, while consistent with theory, is quantitatively small. As a result, earnings risk has seemingly lost its appeal as a candidate explanation for the limited stock market participation puzzle (Guiso and Sodini (2013)).

Yet recent contributions to the earnings dynamics literature document that household labor income substantially departs from the features that characterize these process along two important dimensions (e.g., Arellano et al. (2017), Guvenen et al. (2015), De Nardi et al. (2018)). First, household earnings display varying degrees of persistence that depend on the size of past and current earnings shocks. Second, (log) earnings distributions exhibit significant asymmetries. These features permit asymmetric transmissions of income shocks that have a first-order effect on household portfolios, as shown by recent literature on life cycle models (e.g., Catherine (2022) and Galvez and Paz-Pardo (2021)).

Although there is theoretical work that shows the implications of non-Gaussian earnings dynamics on household portfolio choices, existing empirical work is scarce. In this regard, the first main contribution of this paper is to examine the role of uninsurable income risk on household investment decisions over the life cycle under nonlinear income persistence. The new survey redesign of the PSID allows me to have combined information of income, wealth and consumption for a representative sample of US households. Using the 1999 to 2009 panel survey waves, I estimate the nonlinear nature of income shocks and its implications for the extensive margin of stock market participation and the intensive margin of portfolio allocation.

The second main contribution of this paper is to propose an empirical framework to study *jointly* both the extensive and intensive margins of household portfolio allocation. To do so, I build on the nonlinear panel data framework proposed by Arellano et al. (2017) by introducing non-random sample selection (Heckman (1974)). In this framework, I model empirical portfolio and participation rules as age-specific, non-linear functions of the latent earnings components and wealth. These, together with an equation that governs the dynamics of household wealth, are derived from a general life-cycle model of household portfolio choice with per-period participation costs. As such, the model I propose is compatible with a wide class of structural models, and can be suitably extended to incorporate additional components.

Because households self-select into stock market participation, I require additional assumptions to establish the nonparametric identification of the joint distribution of earnings, assets and portfolio choices that builds on the econometric literature on nonlinear models with latent variables reviewed by Hu (2015). First, the mapping between the latent and observed distributions of risky asset shares must be known. Second, I require an exclusion restriction, that is, a variable that shifts participation costs, but not the subsequent portfolio decision. Provided that both assumptions hold, I identify the empirical participation and portfolio rules from variation in earnings, assets, and participation data. These, in turn, permit the recovery of empirical objects that can be used to study the transmission of income shocks to portfolio choice, and in certain cases, test implications of structural models. In particular, I recover extensive and intensive margins of household portfolio choice, which can then be used to calculate aggregate effects of income and wealth. I also recover impulse response-like functions that assess the extent to which income shocks influence household's participation and portfolio allocation decisions. Finally, I obtain the latent distribution of risky asset shares, which I use to infer participation cost bounds and distributions.

To estimate the empirical policy rules, I rely on a simulation-based algorithm that combines recent developments in sample selection models with sieve estimation approaches, which represents the third main contribution of this paper. To estimate the nonparametric participation and portfolio rules under the presence of sample selection, I consider the quantile selection model of Arellano and Bonhomme (2017a) and the censored quantile regression estimator of Buchinsky and Hahn (1998). Moreover, to deal with the fact that some of the state variables are unobserved to the empirical researcher, I combine the following procedures with a stochastic EM algorithm adapted to time-varying latent variables (Arellano et al. (2017)). The algorithm alternates between two steps: first, simulation draws from the posterior distribution of the latent earnings components, and second, a sequence of Probit and rotated quantile regressions for the policy rules. An added advantage of this approach is its tractability, as the moment conditions for the portfolio rule lead to a convex linear programming problem, one of the appealing features of quantile regressions (Koenker (2005)).

I estimate the semi-structural model using the 1999 to 2009 waves of the US Panel Study of Income Dynamics (PSID), focusing on working-age households. The descriptive statistics indicate that around 40 percent of households re-enter the stock market at least once. These households have higher labor income and wealth than those who never participate in the stock market, but have lower labor income than those who always participate in the stock market.

The estimation of the empirical participation and portfolio rules show that there are heterogeneous effects of increases in income and wealth, respectively, across households over the wealth distribution and the life-cycle, suggesting that there are important interactions between wealth and age. In particular, the extensive margin responses of an increase in wealth is positive and concave along the wealth distribution, while it is positive and monotonically increasing along the life cycle. This result suggests heterogeneities in wealth thresholds that induce households to participate in the stock market. The intensive margin responses, meanwhile of an increase in wealth are positive and increasing along the wealth and age distributions. This result is consistent with the idea that households have heterogeneous preferences that are possibly of the DRRA type (Calvet and Sodini (2014)), which change over the life cycle. I also find that increases in income result in increases in participation and portfolio allocation, although the effects are not as strong. Meanwhile, the impulse response exercises show that introducing nonlinear earnings dynamics into a model of household stock market participation and portfolio choice result in asymmetric participation and portfolio allocation responses across households. To illustrate, the difference in average participation rates for low income households hit by a very positive income shock relative to a median income shock goes up by as much as 12 percentage points; in contrast, the average participation rates for high income households hit by the same shock increases to only two percentage points. Likewise, low income households hit by a very positive income shock increase their risky asset shares by as much as three percentage points, compared to 0.5 percentage points from high income households hit by the same shock. The results also suggest the presence of heterogeneity in participation costs, that range from around almost zero percent of total wealth (for low wealth households) to around 2 percent of total wealth (for high wealth households). Overall, my results highlight the interaction of wealth, income and age in the determination of optimal portfolio choices.

This paper is related to an extensive literature that studies the impact of labor income risk on household portfolio choices (e.g., Guiso et al. (1996), Heaton and Lucas (2000), Vissing-Jørgensen (2002), Angerer and Lam (2009), Palia et al. (2014), and Fagereng et al. (2017b)). Research in this literature has traditionally relied on linear earnings processes and standard econometric methods to investigate the relationship I study here. Relative to these papers, my main contribution is to develop a novel empirical framework that allows for the possibility of studying nonlinear relationships between income risk, stock market participation, and household portfolio choices in a panel data setting. Moreover, I focus on the implications of nonlinear income persistence over the life cycle on household portfolio choices, which to the best of my knowledge, has not been studied.

There is empirical work that has looked at the impact of "unusual" labor market events on stockholding and portfolio choice (Alan (2012), Betermier et al. (2012), Basten et al. (2016), and Knüpfer et al. (2016)). These papers find that households adjust their portfolios in response to events such as unemployment, job switches, and the probability of a zero income realization, which can be considered as "microeconomic disasters". To the extent that a nonlinear earnings process can be thought of as a parsimonious representation of such events, I complement this literature by considering how households' portfolio choice decisions change in response to asymmetric earnings shocks over the life-cycle.[1]

This paper is also related to a small but burgeoning literature that studies the implications of asymmetries and nonlinearities in earnings dynamics on household consumption

[1]Alan (2012) finds that a positive probability of a zero income realization is needed in order to explain household portfolio decisions of younger, poorer households in a structural model. Betermier et al. (2012), using a panel of Swedish households, find that the more volatile the wage is, the lower the exposure of households to risky assets will be, and the less likely they participate in the stock market. Basten et al. (2016), using Norwegian registry data, find some households who can anticipate job loss prepare for unemployment by increasing their saving and shifting toward riskless assets leading up to unemployment, and by depleting their savings after the job loss. Around two years after unemployment, however, they begin to rebalance their portfolio toward risky assets. Finally, Knüpfer et al. (2016) find, within the context of the Finnish Great Depression, that adverse labor market conditions affect both stock market participation and household portfolio choice.

and savings behavior, and on asset prices. These papers argue that nonlinear features of income result in asymmetries in how households insure their consumption against income shocks (e.g., Guvenen et al. (2015) and Arellano et al. (2017)) or in their wealth accumulation patterns (De Nardi et al. (2018)). Higher-order moments of income have also been shown to be a key driver of asset prices (e.g., Schmidt (2015) and Constantinides and Ghosh (2017)). There is some empirical work that studies the covariance between the stock market and income skewness both in the cross-section (Catherine et al. (2022)) and the time series dimensions (Inkmann (2020)). By contrast, this paper focuses on the impact of nonlinear income persistence on household portfolio choice decisions, which is intimately linked to conditional skewness of household earnings (Arellano et al. (2017)). My impulse response analyses also reinforce some of the results in this literature. For instance, I find that a very negative income shock results in high-income households leaving the stock market, which is consistent with Schmidt (2015)'s result that investing in stocks is a poor hedge against adverse labor market events. Moreover, my results also speak to the finding of Catherine et al. (2022) of the importance of per-period participation costs. I go further by quantifying empirically the cost bounds for different households with different wealth levels.

Finally, this paper is related to recent developments in the panel data literature (see the survey of Arellano and Bonhomme (2017b)) that proposes nonlinear reduced forms for a wide class of dynamic economic models. With respect to this literature, I propose an estimation framework that takes into account situations in which sample selection is paramount, which to the best of my knowledge, has not been done. As such, the framework I propose here can also be used to analyze other economic models that exhibit similar features such as models of labor supply (e.g., Heckman (1974)) and occupational choice (e.g., Adda et al. (2017)).

The rest of the paper is structured as follows. Section 2 discusses the semi-structural model that I take to the data. I describe the data and some descriptive evidence in Section 3. In Section 4, I provide details on the nonparametric identification and estimation of the semi-structural model of portfolio choice. Section 5 presents the estimation results. Finally, Section 6 concludes. Additional material is gathered in the online appendices.

2 A semi-structural, life-cycle portfolio choice model

The goal of this section is to develop a flexible empirical framework that is consistent with a wide class of dynamic structural models of life-cycle portfolio choice under several models of earnings dynamics. In particular, for the empirical question at hand, the model is flexible to accommodate both canonical and non-linear models of income risk.

As a motivation for the empirical analysis, I briefly discuss the implications of both earnings processes in a life-cycle portfolio choice model, which draws on Galvez and Paz-Pardo (2021). In the second subsection, I outline the nonparametric model of household portfolio choice. I begin with a description of the standard life cycle model with per-period participation costs, following several papers in the household finance literature (see e.g., Alan (2012), Briggs et al. (2015), and Fagereng et al. (2017b)). I then specify the empirical portfolio and participation rules, and illustrate the objects of interest I can recover from the resulting nonlinear reduced form system.

In what follows, I consider a cohort of households $i = 1, \ldots, N$, and denote by $t = 1, \ldots, T$ the age of the household head.

2.1 Nonlinear earnings dynamics and their implications for portfolio choice

Nonlinear household earnings dynamics. Let y_{it} denote (log) residual household labor income, which I decompose into the following two components:

$$y_{it} = \eta_{it} + \varepsilon_{it}\, , i = 1, \ldots, N, t = 1, \ldots, T. \tag{1}$$

in which η and ε are continuous distributions. The first component η_{it} is persistent, and follows a first-order Markov process. The second component ε_{it} is transitory, and has zero mean, is independent over time and of all realizations of the persistent component.

In most structural models, the usual specification for the persistent component is that of an AR(1) with normal shocks, which I will refer to as the canonical process. Recent empirical literature has uncovered three facts that are inconsistent with this assumption: nonlinearity in persistence, non-normality of shocks, and age-dependence of the persistent component of income and higher-order moments (De Nardi et al. (2018)). In this light, Arellano et al. (2017) propose a quantile-based method to model flexibly richer earnings dynamics. Letting $Q_t(\eta_{it-1}, u_{it})$ denote the τ^{th} conditional quantile of η_{it} given η_{it-1} for all $\tau \in (0, 1)$, the process for η_{it} as:

$$\eta_{it} = Q_t(\eta_{it-1}, u_{it}) \tag{2}$$

where $(u_{it}|\eta_{it-1}, \eta_{it-2}, \ldots) \sim U[0, 1]$ for all t. Clearly, the quantile function $Q_t(\cdot, \cdot)$ maps random draws from the uniform distribution (i.e., the cumulative probabilities) to the corresponding quantile draws for η. Meanwhile, the canonical model can be seen as a special case of model (2) above. In particular, assuming that the cumulative distribution function of η_{it} is Gaussian, the canonical earnings process has the following representation: $Q_t(\eta_{it-1}, u_{it}) = \rho\eta_{it-1} + \Phi^{-1}(u_{it})$.

The quantile-based specification in Arellano et al. (2017) permits the computation of a generalized notion of persistence, which can be summarized as the following quantity:

$$\rho(\eta_{it-1}, \tau) = \frac{\partial Q_t(\eta_{it-1}, \tau)}{\partial \tau}, \tag{3}$$

rank τ. In comparison, the canonical earnings model imposes $\rho(\eta_{it-1}, \tau) = \rho$, regardless of the past realization η_{it-1} and of the shock quantile τ. The nonlinear earnings process, hence, allows for the possibility of current shocks to wipe out the memory of past shocks, or equivalently, for the future persistence of a current shock to depend on future shocks. This notion of persistence, as Arellano et al. (2017) argue, is the persistence of *earnings histories*. In addition, the quantile-based earnings process permits conditional skewness in η_{it}.[2] In comparison, the canonical linear earnings process does not allow for the possibility of non-normality in the shock distributions. As is clear, I can also write a similar unrestricted representation for the transitory component ε_{it} and the initial condition η_1, with the only difference being that they are not persistent.

The nonlinear earnings process of Arellano et al. (2017) is compatible with structural models of the labor market. In particular, one can think of the model as a reduced-form representation of job ladder models of occupational mobility, such as that in Lise (2013). In these models, workers look for good matches with firms while unemployed and while employed. A worker ends employment with a firm when he finds a better opportunity ("climbing up the ladder"), or when he loses a job ("falling off the ladder"). Therefore, on-the-job search and job-to-job transitions generate asymmetries that are captured by the features of the nonlinear earnings process.

I discuss the estimation of both the nonlinear and canonical earnings processes, and compare the differences between the two processes in Appendix A.

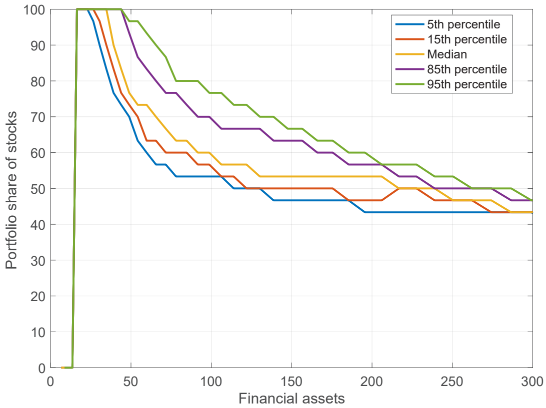
Implications for household portfolio choice. As a guide for the empirical analysis, I discuss the implications of nonlinear and canonical earnings dynamics on household

[2]Specifically, I can define conditional skewness via the following equation, for a given $\tau \in (0.5, 1)$: $sk_t(\eta_{it-1}, \tau) = \frac{Q_t(\eta_{it-1},\tau)+Q_t(\eta_{it-1},1-\tau)-2Q_t(\eta_{it-1},\frac{1}{2})}{Q_t(\eta_{it-1},\tau)-Q_t(\eta_{it-1},1-\tau)}$.

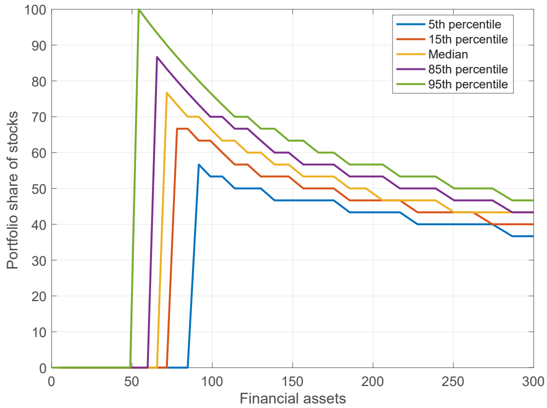
portfolio choices, as in Galvez and Paz-Pardo (2021)[3]. In their model, they consider a discrete-time, life cycle model of households with CRRA utility that decide on optimal consumption and savings. Households in their framework have access to a risk-free and a risky asset, and face participation costs to the stock market. To facilitate the discussion, I reproduce some of the results from their calibrated structural model.[4].

Figure 1: Policy rules from the portfolio choice model of Galvez and Paz-Pardo (2021)

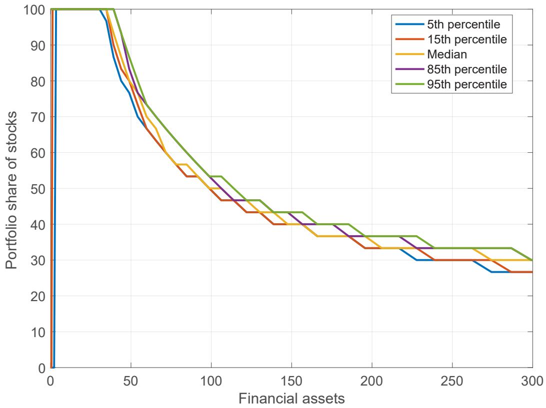
(a) Age 30, canonical model



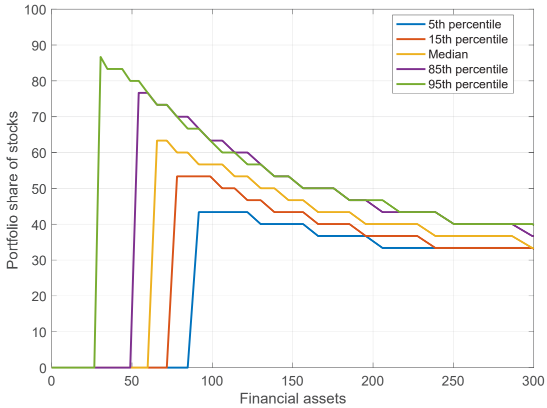
(b) Age 30, nonlinear model



(c) Age 55, canonical model



(d) Age 55, nonlinear model



Note: Panels (a) and (b) show the policy rules from the life-cycle portfolio choice model with the canonical process and the nonlinear process at age 30. Panels (c) and (d) show the policy rules from the life-cycle portfolio choice model with the canonical process and the nonlinear process at age 55. Results are from Galvez and Paz-Pardo (2021). The different lines correspond to the portfolio rules at different percentiles of the income distribution.

[3] In Appendix B, I describe a two-period model based on Campbell and Viceira (2002), and discuss how the features of a more flexible earnings process can influence portfolio choices.

[4] Galvez and Paz-Pardo (2021) find that the structural model under the nonlinear earnings process yields a lower estimate of relative risk aversion and a different structure of participation costs than the model under the canonical earnings process. They also find that the model under nonlinear dynamics fits the profile of stock market participation better.

Figure 1 show the optimal portfolio rules from their estimated structural model. Panels 1a and 1b show the optimal policy rule at age 30. In both economies, low-wealth households do not invest at all in the stock market, as they find it optimal not to pay the participation cost. Once they reach the threshold at which stock market participation is optimal, the resulting policy rules are decreasing functions of wealth. The key driver is the importance of human capital (discounted stream of future labor income) relative to accumulated wealth. However, there are noticeable differences between the model under the two earnings processes. The first difference is related to wealth thresholds, that is, the amount of wealth households believe makes it worthwhile to invest in stocks. In particular, under the canonical process, all households, regardless of their position in the income distribution, start investing in the stock market around the same threshold. Households in the nonlinear process, by contrast, have different thresholds of wealth that depend on their position in the income distribution. The second difference occurs in the portfolio rules at age 55. Under the economy with the canonical earnings process, the uncertainty with respect to their labor income is resolved, leading them to become more aggressive in their stock investments. Under the nonlinear earnings process, the considerable uncertainty with respect to their labor income leads households to remain conservative with respect to their portfolio investments.

In sum, both results underscore the idea that under a more flexible earnings process, income becomes more "stock-like" than before, which implies different wealth thresholds and asset demands.

2.2 Semi-structural model of portfolio choice

I now describe the household optimization problem, and the corresponding empirical policy functions that I will take to the data.

Consider a household with utility function $u(\cdot)$, which is assumed to be concave. It is endowed with wealth w_{it}, and makes an asset allocation decision at time t. The household earns (log) labor income y_{it}, which can be decomposed into deterministic and stochastic parts. The stochastic components of income can be decomposed following equation (1), while the model for the persistent component can be any general first-order Markov process, for example, equation (2). The household cannot borrow against future labor income, thereby making it non tradable. It has access to two assets for investment: a riskless asset that earns a certain return R_f, and a risky asset with a constant expected excess return $\mathbb{E}_t(R_{t+1} - R_f) = \mu$. However, there is a probability that the household might suffer a loss in its investment in risky assets, which can be represented by the unexpected return on the risky asset that has a distribution with mean zero and variance σ_u^2.

Finally, the household faces some frictions in financial markets. First, if the household wants to invest in stocks, it must pay a per-period participation cost F. One can rationalize this cost as a way of capturing several explanations proposed for limited participation in financial markets. These include the presence of trading costs (e.g., Vissing-Jørgensen (2002)), financial sophistication and financial literacy, or the lack of it (e.g., Calvet et al. (2007), Van Rooij et al. (2011)), and trust in financial markets (e.g., Guiso et al. (2008)). Second, the household can neither borrow nor short-sell, constraining its risky share to be between zero and one, as is typical in household portfolio choice problems (Cocco et al. (2005)).

To make its decision, the household solves two maximization subproblems. In the first, called the participation subproblem, it maximizes the following objective function by choosing optimal consumption c_{it} and risky share α_{it} given the state variables w_{it}, η_{it} and ε_{it}:

$$V_{p,t}(w_{it}, \eta_{it}, \varepsilon_{it}) = \max_{c_{it}, 0<\alpha_{it}\leq 1} u(c_{it}) + \beta E_t(V_{t+1}(w_{it+1}, \eta_{it+1}, \varepsilon_{it+1}))$$

subject to the following intertemporal budget constraint:

$$w_{it+1} = [\alpha_{it} R_{t+1} + (1-\alpha_{it})R_f][w_{it} + y_{it} - c_{it} - F] + y_{it+1}.$$

In the second subproblem, called the non-participation subproblem, the household computes optimal consumption by solving the following Bellman equation:

$$V_{np,t}(w_{it}, \eta_{it}, \varepsilon_{it}) = \max_{c_{it}} u(c_{it}) + \beta E_t(V_{t+1}(w_{it+1}, \eta_{it+1}, \varepsilon_{it+1}))$$

subject to the following intertemporal budget constraint:

$$w_{it+1} = R_f[w_{it} + y_{it} - c_{it}].$$

Finally, to make its optimal choice, the household compares the utility gained from each scenario and chooses optimal consumption, and the share of wealth in risky assets that maximizes utility:

$$V_t = \max\{V_{p,t}(w_{it}, \eta_{it}, \varepsilon_{it}), V_{np,t}(w_{it}, \eta_{it}, \varepsilon_{it})\} \tag{4}$$

Deriving empirical portfolio and participation rules. The life-cycle model I described earlier results in the following portfolio rule that corresponds to the solution of the participation subproblem:

$$\alpha_{it}^{*} = g_t(w_{it}, \eta_{it}, \varepsilon_{it}, \xi_{it}) \tag{5}$$

where g_t is an age-specific function of wealth and the stochastic components of income. To take the portfolio rule to the data, I introduce an unobserved argument ξ_{it}, which can be thought of as a taste shifter (e.g., risk aversion) that affects household portfolio choice decisions. In the baseline model, ξ_{it} is an unobserved preference shifter that increases households' marginal utility; this implies that g_t is monotone in ξ_{it}.[5]

However, the portfolio rule above is latent. That is, the empirical researcher only observes the solution of the complete portfolio choice problem, which is given by the following equation:

$$\alpha_{it} = \alpha_{it}^{*} \cdot d_{it}, \tag{6}$$

wherein the observed portfolio rule is the interaction of the latent portfolio rule solved in the participation subproblem, and the participation rule of the households, which is represented by the indicator d_{it}. A reduced form version of the indirect utility comparison in (4) can be written as:

$$d_{it} = \begin{cases} 1, & \text{if } m_t(w_{it}, \eta_{it}, \varepsilon_{it}) \leq \chi_{it} \\ 0, & \text{otherwise} \end{cases} . \tag{7}$$

Similar to the portfolio rule (5), to take the participation rule to the data, I introduce an unobserved argument χ_{it} that corresponds to participation cost shifters. It is crucial to note that while the unobserved arguments ξ_{it} and χ_{it} are different, they are correlated to each other, given the nature of the household problem. In this sense, I can stack them into a vector $\mathbf{v}_{it} = (\xi_{it}, \chi_{it})'$, that represents the unobserved errors of the portfolio choice problem.[6] Allowing for correlation between the error terms suggests the presence of sample selection bias, an issue that I will resolve in section 4. Moreover, the presence of equation (7) permits the mapping of the semi-structural model to the Bellman equation of the entire economic problem.

[5] This will be true in the participation subproblem if $\frac{\partial u(C,u')}{\partial u'} > \frac{\partial u(C,u)}{\partial u}$ where $u' > u$. This implies, hence, that the Bellman equation of the participation subproblem is monotonic.

[6] One can also plausibly think that $\xi_{it} = \chi_{it}$, as it is highly likely that the same unobserved shifters drive both participation costs and portfolio choice decisions.

As is clear from the household problem, wealth w_{it} is endogenous. To ensure identification of the portfolio and participation rules, I assume that w_{it} is predetermined.[7] Specifically, I write the following equation:

$$w_{it} = h_t(w_{it-1}, \eta_{it-1}, \varepsilon_{it-1}, \alpha_{it-1}, \upsilon_{it}) \tag{9}$$

in which household wealth is an age-dependent function of lagged assets, the lagged persistent and transitory components of income, and the lagged portfolio choice decision of the household. The unobserved argument υ_{it} can be thought of as a catch-all that captures two important aspects of the problem that I do not model explicitly. The first is that of the return process for risky assets.[8] The second, meanwhile, is that of the consumption rule. Finally, to close the model, I specify the initial wealth distribution w_{i0} as unrestricted.

Together with the income process (2), equations (5)-(9) constitute a system that describes the life-cycle model of portfolio choice for a wide class of structural models, as it does not impose a specific functional form[9]. One can use this model to estimate directly households' stock market participation and portfolio rules under several models of earnings dynamics. Moreover, the model's flexibility permits interactions between the different state variables of the economic problem at hand. This stands in contrast to reduced-form models, which come from first-order approximations of the economic model. A drawback of the semi-structural model that I outlined here, compared to dynamic structural models, is that it cannot be used to analyze counterfactual scenarios. However, the model permits the recovery of objects of interest that can be useful targets for structural estimation, and provide some suggestive evidence on the nature of stock market participation, or the nature of preferences, and in particular, whether they are of the CRRA or DRRA type.

---

[7] An alternative to introducing the wealth accumulation rule is the introduction of the empirical consumption rule as in Arellano et al. (2017):

$$c_{it} = h_t(\eta_{it}, \varepsilon_{it}, w_{it}, u_{it}) \tag{8}$$

wherein consumption is a function of the stochastic components of income, wealth and an unobserved taste shifter. If one introduces this in the estimation, though, one must include the imposition of the budget constraint.

[8] It could be the case that there is some correlation between the persistent earnings component, and the returns to the risky asset. The flexibility of the wealth accumulation, as shown by the interaction of the income components, wealth, and the share of wealth in risky assets, allows for this correlation to be captured.

[9] A clear exception, however, are models that feature ambiguity aversion (Peijnenburg (2018)). One would need a suitable modification of the dynamic programming problem to incorporate these models.

Objects of interest. The semi-structural model of portfolio choice allows me to recover the following objects of interest. To fix ideas, I calculate all objects with respect to the persistent component of income η_{it}, though I can also calculate these with respect to wealth w_{it} and the transitory component ε_{it}.

First, the average stock market participation rate, for a specific value of wealth and the stochastic components of income is:

$$\mathbb{E}[(\mathbf{1}(m_t(w_{it}, \eta_{it}, \varepsilon_{it}) \leq \chi_{it})|w_{it}, \eta_{it}, \varepsilon_{it}] = \Pr(d_{it} = 1|w_{it}, \eta_{it}, \varepsilon_{it}).$$

From this, I can recover the extensive margin of stock market participation, which is simply:

$$\phi_E(w, \eta, \varepsilon) = \frac{\partial \Pr(d_{it} = 1|w, \eta, \varepsilon)}{\partial \eta_{it}} \tag{10}$$

Meanwhile, the average share of wealth in risky assets for stock market participants can be written as:

$$\mathbb{E}(\alpha^*_{it}|w, \eta, \varepsilon, d_{it} = 1) = \mathbb{E}[g_t(w_{it}, \eta_{it}, \varepsilon_{it}, \xi_{it})|w, \eta, \varepsilon, d_{it} = 1]$$

Given this, I can then recover the intensive margin of household portfolio choice, which can then be written as:

$$\phi_I(w, \eta, \varepsilon) = \mathbb{E}\left(\left.\frac{\partial g_t(w_{it}, \eta_{it}, \varepsilon_{it}, \xi_{it})}{\partial \eta}\right| d_{it} = 1\right) \tag{11}$$

The extensive and intensive margins of portfolio choice can then be combined to compute the aggregate effect of income on portfolio choice:

$$\phi_A(w, \eta, \varepsilon) = \frac{\partial \Pr(d_{it} = 1|w, \eta, \varepsilon)}{\partial \eta_{it}} \times \mathbb{E}\left(\left.\frac{\partial g_t(w_{it}, \eta_{it}, \varepsilon_{it}, \xi_{it})}{\partial \eta}\right| d_{it} = 1\right), \tag{12}$$

which after taking logs on both sides, can then be interpreted as the elasticity of participation and portfolio choice with respect to labor income, respectively.

Apart from the derivative effects, I can also recover other objects of interest. In particular, to study the dynamic effects of a persistent income shock, I can compute "impulse response"-like functions:

and

$$\Delta_I(\eta + \eta', \eta) = \mathbb{E}(\alpha^*_{it}|\eta + \eta', \varepsilon, w, \mathbf{x}, d_{it} = 1) - \mathbb{E}(\alpha^*_{it}|\eta, \varepsilon, w, \mathbf{x}, d_{it} = 1)$$

which can be thought of as finite difference counterparts of the average derivative effects (10) and (11) with respect to a shock u_{it} to persistent labor income. Finally, I can recover the latent distribution of risky asset shares $F(\alpha^*_{it}|w_{it}, \eta_{it}, \varepsilon_{it})$ and utilize this object to compute participation cost distributions.

3 Data and descriptive evidence

3.1 Dataset description and sample selection

The main dataset for my empirical analysis is a balanced panel of households from the 1999 to 2009 waves of the Panel Study of Income Dynamics (PSID). The primary aim of the survey was to study the dynamics of income and poverty of US households. Hence, the original 1968 study was drawn from two independent subsamples: 2,000 poor families that were under the Survey of Economic Opportunity (SEO), and a nationally representative sample of approximately 3,000 families. The survey waves were annual from 1968 until 1997, when the data was collected biennially. A distinct advantage of the PSID is that since the 1999 wave, it has collected detailed data on consumption expenditures and asset holdings, in addition to information on household earnings. This makes it one of the few longitudinal surveys in the US with comprehensive information on assets, consumption, and earnings for a representative sample of households. As I need continuous information on labor earnings and portfolio choices over the life cycle, I focus on the 1999 to 2009 waves, which correspond to calendar years 1998 to 2008. I deflate all of the variables with 2000 as the base year.

Sample selection criteria. I focus on non-SEO households with participating and married household heads between 25 to 60 years old. I exclude households that have missing information on key demographic variables (age, race, education, and state of residence) and on the main variables in the study, in logs or in levels. To reduce the influence of measurement error, I remove households that have more than \$20 million in total household assets, following Blundell et al. (2016). I also drop households that have "extreme jumps" in their earnings and implied hourly wages[10], and those who have transfer incomes that are more than twice household labor income. The sample selection

[10] A jump is defined as an extremely positive (negative) change from year $t-2$ to t, followed by an extremely negative (positive) change from year t to $t+2$. Formally, for each variable, I construct the biennial log difference $\Delta^2 \log(x_t)$ and drop the relevant variables for observation in the bottom 0.25 percentile of the product $\Delta^2 \log(x_t)\Delta^2 \log(x_{t-2})$, following Blundell et al. (2016).

criteria results in a balanced panel of 661 households. A detailed description of the data cleaning process is in Appendix C.1.[11]

3.2 Main variables

Earnings Y_{it} is total pre-tax household labor earnings. I construct y_{it} as the residuals from regressing log household earnings on a set of demographics, which include cohort dummies interacted with education categories for both husband and wife, race, state, and large city dummies, a family size indicator, number of kids, a dummy for income recipient other than husband and wife, and a dummy for kids out of the household.

I consider risky assets as the sum of two components: (i.) the value of stockholdings held in publicly traded corporations, mutual funds or investment in trusts, and (ii.) the part of Individual Retirement Accounts (IRAs) that are held in stocks.[12] To identify the part of the IRA allocated in stocks, I follow the treatment of Vissing-Jørgensen (2002), Malmendier and Nagel (2011) and Palia et al. (2014). Specifically, the PSID asks a household about the allocation of its pension account, if it has any. I assume that all investments in IRAs are in stocks if the household reports that most of the money is allocated in them. If the household reports that the money in the IRA is split between stocks and interest-earning assets, I assume that half the value is in stocks and half the value is in bonds.

Wealth W_{it} is constructed as the sum of financial assets, real estate value, pension funds, and car value, net of mortgage and other debt. I define financial assets as the sum of: stocks; cash, defined as checking or savings accounts, money market accounts, or Treasury bills, including those held in IRAs; and bonds, which includes bonds, the cash value in life insurance policies, valuable collections, rights in trusts or estates. All of

[11] I also calculate summary statistics to compare my baseline sample with a sample of all married household heads (independently of work status) and with a sample of all household heads headed by a male recorded at least once in the 1998 to 2008 period (again, independently of work status), which are also shown in Appendix C. The results indicate that there does not seem to be substantial differences across samples.

[12] Albeit some papers in the empirical literature (such as Brunnermeier and Nagel (2008) and Chiappori and Paiella (2011)) include home equity in the definition of risky wealth, as it can be interpreted as such (see Flavin and Yamashita (2002)), doing so requires a more involved estimation framework in which I would also need to model homeownership, and the evolution of house prices, an aggregate state variable (see Hahn et al. (2015) for a discussion on the challenges of estimating models with aggregate shocks). Moreover, empirically disentangling the effects of house price risk from labor income risk is an arduous task, as underscored by Chetty et al. (2017). This definition of wealth recognizes, however, that households indeed have most of their wealth in housing. In this sense, I can also interpret the error term in the evolution of wealth equation as one that also captures the return process of housing value. Moreover, my sample selection criteria is such that almost all households in my sample are homeowners at any given point in time.

the estimations that I present use the log of total household wealth, w_{it}, as the relevant independent variable. Finally, the risky share α_{it} is computed as the proportion of risky assets to total household wealth.

3.3 Descriptive evidence

Table 1 presents pooled cross-section/time-series summary statistics for all relevant variables, grouped by income quartiles. The table is divided into two panels. The first panel corresponds to all households that satisfy the sample selection criteria. The second panel, meanwhile, corresponds to the subset of risky asset market participants.

Table 1: Sample means of main variables, by income quartiles

		Income quartile			
	TOTAL	First	Second	Third	Fourth
			All participants		
Household income	111,612.60	42,206.75	71,354.16	99,518.06	222,091.50
Total assets	371,189.60	184,597.80	212,898.20	332,963.40	719,382.40
Liquid wealth	95,529.77	43,204.98	48,433.76	73,681.34	206,119.10
Risky wealth	81,353.17	35,366.25	36,714.01	60,835.09	182,757.20
Stocks	51,547.79	23,514.87	19,774.00	34,475.20	121,831.80
Share of stocks in total wealth	0.077	0.048	0.057	0.073	0.124
Share of risky assets in total wealth	0.148	0.087	0.120	0.154	0.224
Ownership of risky assets	0.591	0.372	0.526	0.649	0.793
			Risky market participants		
Household income	130,668.70	44,692.65	71,214.30	99,955.57	226,578.30
Total assets	512,872.20	317,459.30	289,769.80	399,048.60	818,180.50
Liquid wealth	145,598.80	101,981.00	73,786.93	97,593.12	244,521.50
Risky wealth	137,676.70	95,196.34	69,818.05	93,735.98	230,589.40
Stocks	87,236.05	63,295.65	37,603.70	53,120.10	153,718.30
Ownership of stocks	0.695	0.621	0.653	0.684	0.759
Share of risky assets in total wealth	0.250	0.233	0.229	0.237	0.283

Note: Data from 1999 to 2009 PSID waves. This table presents sample means of the main economic variables (in 2000 US dollars) related to this empirical study, calculated across income quartiles. The first column calculates the mean across all households in the sample, while the second to fifth columns calculate the mean for different income quartiles. The first panel presents results across all households. The second panel presents results for risky market participants, defined as households who have direct and indirect stockholdings in stocks, and who have part of their pension funds invested in stocks.

The table indicates a wide dispersion across households in their earnings and assets. The average participation rate in risky wealth is around 59.1 percent for the households in my sample, slightly higher from those in other observational studies, such as the US Survey of Consumer Finances (SCF)[13]. Furthermore, once I condition on the subset of

[13]The proportions in the SCF are 48.9 (1998), 52.2 (2001), 50.2 (2003), and 51.1 (2007). (Bucks et al. (2009))

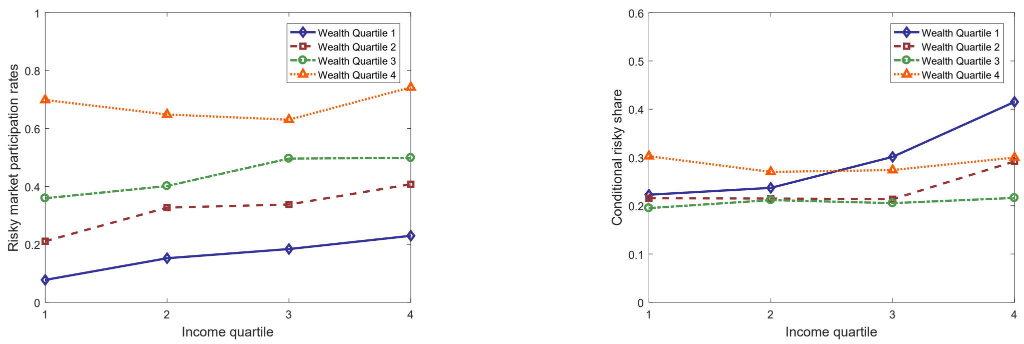
risky asset market participants, I find that household assets are not monotonic in income. In fact, households at the lowest and highest income quartiles have higher liquid wealth and stockholdings than households at the middle income quintiles. While there may be a host of other reasons why this could be the case, one can surmise that differences in the income risks that these households face could possibly drive this phenomenon.

Figure 2 presents the average stock market participation rate and conditional risky share for households of different income and wealth quartiles. As the graphs indicate, with the exception of the highest wealth quartile, participation rates increase as income increases. For households in the highest wealth quartile, however, the graphs suggest that those at the extreme income quartiles have slightly higher participation rates than those in the middle income quartiles. The conditional risky shares, meanwhile, reveal non-monotonic patterns across income and wealth quartiles. The graphs also indicate that among households in the lowest income quartile, those in the highest wealth quartile tend to be the most aggressive in their risky asset investment. At the same time, among households in the highest income quartile, those in the lowest wealth quartile tend to invest the most in risky assets, on average. Overall, these plots suggest an interaction between wealth and income in terms of risky asset market investment.

Figure 2: Participation and the conditional risky share, by income and wealth quartiles

(a) Risky market participation rates

(b) Conditional risky shares



Note: Data from 1999 to 2009 PSID waves. The following figures show the average stock market participation rates and the average conditional risky shares for households of different income and wealth quartiles. The x-axis corresponds to the income quartiles, while the y-axis corresponds to the average participation rate or risky share. The blue line corresponds to households in the poorest wealth quartile; the red line corresponds to households in the second wealth quartile; the green line corresponds to households in the third wealth quartile; and the orange line corresponds to households in the richest wealth quartile.

Table 2 presents the frequency distribution of household risky market participation sequences disaggregated by age quartiles. I distinguish between the following groups of households: (i.) those who have never participated; (ii.) those who have always participated; (iii.) "pure entry" and "pure exit" households, that is, those who, after observing a stock market entry after nonparticipation, choose to remain in the stock market and likewise, those who, after observing a stock market exit after an entry, remain to be out[14]; and (iv.) households who move from one participation state to another more than once. Looking at the panel of risky market participation sequences, households with at least one re-entry comprise the majority in the sample, at 39.03 percent. This is followed by households who have always participated, who are approximately 30.10 percent of the sample. Households who have never participated in the stock markets comprise 16.64 percent, and the rest are the "pure entry" and "pure exit" households.

Table 2: Frequency distribution of risky market participation sequences

		Age quartile			
	TOTAL	First	Second	Third	Fourth
Never participated	110	15	32	52	11
Always participated	199	11	40	108	40
Pure entry households	51	9	9	29	4
Pure exit households	43	1	15	25	2
Households with re-entries	258	17	70	141	30
TOTAL		53	166	365	87

Note: Data from 1999 to 2009 PSID waves. This table presents the frequency distribution of household stock market participation sequences disaggregated by age quartiles. In this table, households who have never bought stocks are those who have participation sequences (0,0,0,0,0,0), while households who have always participated are those who have participation sequences (1,1,1,1,1,1). Pure entry and pure exit households are those who have the sequences that are described in the text. Households with re-entries are those who have had more than one re-entry in the stock market.

Finally, Table 3 presents summary statistics according to participation status. Households who have never participated tend to be less educated, have lower household incomes, and have lower household wealth. The rest of the households in my sample tend to have studied at least a year of university education and are roughly of the same age category. Interestingly, those who have re-entered the stock market have labor incomes that are less than those of who have always participated in the stock markets. Arguably, this result suggests that these households are at the margin between participation and non-participation in the stock markets. Furthermore, the labor incomes these households have suggest that they face substantial earnings risk.

[14] A pure entry household is one with a participation sequence of (0,1,1,1,1,1), (0,0,1,1,1,1), etc. Meanwhile, a pure exit household is one with a participation sequence of (1,0,0,0,0,0), (1,1,0,0,0,0), etc.

Table 3: Summary statistics, by risky market participation sequence

	Participation sequence status				
	Never	Always	Pure entry	Pure exit	Re-entries
Age of household head	43.32	46.55	43.88	44.16	44.91
Education of household head	12.24	15.23	14.94	13.81	14.24
Household income	61,808.22	145,525.20	96,973.63	152,104.60	95,064.78
Net household wealth	110,145.50	640,761.60	191,249.00	435,548.50	247,425.80
Liquid wealth	8,921.91	202,606.90	30,637.71	95,406.39	44,432.09
Risky wealth	-	196,810.00	25,291.58	59,151.12	25,136.02
Ownership of risky assets	-	1.00	0.58	0.60	0.48
Share of wealth in risky assets	-	0.29	0.11	0.13	0.10

Note: Data from 1999 to 2009 PSID waves. This table presents summary statistics of the main variables in the PSID subsample that I consider, disaggregated by stock market participation status. The first column corresponds to households who have never participated in the stock market. The second column corresponds to households who have always participated in the stock market. The third and foruth columns correspond to households who have either "purely entered" or "purely exited" the stock market. Finally, the last column corresponds to households who have at least one stock market re-entry.

4 Identification and estimation strategy

There are two main challenges to overcome in identifying and estimating the semi-structural model of portfolio choice. First, households select themselves into stock market participation, which implies dealing with the well-known sample selection problem (Heckman (1974)). Second, the stochastic components of income are also unobserved, and the way it translates to stock market participation and portfolio choice can be highly nonlinear. I address these two concerns in this section.

4.1 Nonparametric identification

In the current setting, the goal is to recover the empirical portfolio and participation rules, and the latent distribution of risky asset shares. Because the semi-structural model of portfolio choice takes the form of a nonlinear state-space model, I leverage techniques used in the literature that studies nonparametric identification of the joint dynamic distributions of the observed and latent variables in these nonlinear models (surveyed in Hu (2017)) to outline a formal argument in Appendix D.

As I show, the empirical participation and portfolio rules, the average derivative effects, and the impulse response functions are non-parametrically identified given at least two periods of earnings, assets, and the observed participation and portfolio choices, provided that two assumptions are satisfied. First, the mapping between the latent and observed distributions of risky asset shares must be known. Second, there should be a variable that affects participation, but not the subsequent portfolio allocation. Within the context of the model, the exclusion restriction can be thought of as a cost shifter.

It is crucial that both assumptions must be satisfied. Knowledge of the mapping between the latent and observed distributions, which is represented by the conditional copula, permits the calculation of the latent quantiles of risky asset shares from the observed data. However, this function will not be informative of the extent of participation in the stock market without the presence of an exclusion restriction. Otherwise, the quantiles, and subsequently, the distribution of risky asset shares, are only set identified (Arellano and Bonhomme (2017a)).

The intuition behind the identification argument comes from the connection to nonparametric instrumental variable problems (see, e.g., Newey and Powell (2003) and Blundell et al. (2007)). In my set-up, the endogenous variable is the persistent component of income, which is unobserved. As I argue in appendix D, given the assumptions of the model, the "excluded instruments" are the lagged portfolio choices, participation indicators, assets, and the leads and lags of earnings. The availability of these instruments allows me to identify average derivative effects of the persistent component of income, and participation and portfolio choice responses with respect to an income shock.

Using leads and lags of earnings is a common strategy in identifying consumption responses (see, e.g., Blundell et al. (2008)) with respect to an income shock. To the best of my knowledge, this strategy has not been used to identify the impact of income shocks on portfolio allocation. The usual approach is to estimate measures of income risk (commonly the variance of labor income) (as in Angerer and Lam (2009) and Fagereng et al. (2017b)) or to use information on subjective income expectations (as in Guiso et al. (1996) and Hochguertel (2003)), and use these as an independent variable in a linear regression. The approach taken here provides the possibility of directly estimating these responses from the available data on earnings, assets and participation.

4.2 Model specification

As the goal is to estimate a flexible life-cycle model that is compatible with a wide class of structural models, I rely on sieve estimation approaches (Chen (2007)) that allow me to specify the empirical portfolio rules as semiparametric, age-dependent, nonlinear functions of the latent earnings components and of wealth. In the rest of this subsection, I discuss the specifications I use. In what follows, let $\varphi_k(\cdot)$, $k = 1, \ldots, K$ denote a dictionary of functions, with $\varphi_0 = 1$.

Portfolio rule. I first discuss the specification for the portfolio rule (5). Letting age_{it} denote the age of the household head i at period t, I specify the empirical portfolio rule as:

$$\Lambda^{-1}(\alpha^*_{it}) = \sum_{k=0}^{K} b^a_k(\tau)\varphi_k(\eta_{it}, \varepsilon_{it}, w_{it}, age_{it}) + \gamma^a(\tau)'\mathbf{X}_{it} \tag{13}$$

in which I introduce a logit transformation $\Lambda^{-1}(\cdot)$ to constrain the risky asset shares to be in between zero and one; this reflects the fact that in the economic problem, households can neither borrow nor can they short sell. This does not affect the recovery of the latent distribution of risky asset shares, as quantiles are invariant to monotonic transformations (Koenker (2005)).[15] In practice, $\varphi_k(\cdot)$ is a product of Hermite polynomials. The function depends on the quantiles of the distribution of risky shares, which implies that I consider a series quantile model.

The empirical specification is composed of two parts: a nonparametric function that corresponds to the state variables of the economic model, and a linear function that corresponds to variables that proxy for preference shifters/life-cycle controls. It is conceptually straightforward to allow all variables to interact with each other, but this would result to a less parsimonious specification. Moreover, as I am interested in the average derivative effects of the state variables, I reduce the dimensions of the nonlinear function I aim to approximate by introducing the life-cycle controls linearly, which are contained in the vector $\mathbf{X}_{it}$. I allow for flexibility by specifying $\gamma^a(\tau)$ as a function of τ.

Participation rule. I specify the participation rule given current earnings components, assets, age and life-cycle controls as follows:

$$\Pr(d_{it} = 1|\eta_{it}, \varepsilon_{it}, w_{it}, age_{it}, \mathbf{Z}_{it}) = \Phi\left(\sum_{k=0}^{K} b^p_k\varphi_k(\eta_{it}, \varepsilon_{it}, w_{it}, age_{it}) + \gamma^p\mathbf{Z}_{it}\right) \tag{14}$$

in which $\Phi(\cdot)$ is the Normal CDF[16]. As opposed to the portfolio rule, the vector of life-cycle controls/preference shifters is in $\mathbf{Z}_{it} = (\mathbf{X}_{it}, b_{it})'$, in which b_{it} is an exclusion restriction, i.e., a variable that affects participation, but not the subsequent portfolio decision. In the next subsection, I discuss the variable I use as an exclusion restriction.

[15] Chamberlain (1994) and Buchinsky (1995) apply Box-Cox transformations in a censored quantile model where they study female wage distributions, while Bottai et al. (2010) use the logistic transformation in the context of studying adolescent depression.

[16] In estimations that are not shown here, I also consider a situation wherein I use a logistic cdf. The results are quite similar.

Evolution of wealth. I specify the initial distribution of wealth w_{i1} conditional on the persistent component η_{i1}, age at the start of the period age_{i1} and life-cycle controls during the initial period when I observe them as follows:

$$Q_w(\eta_{i1}, age_{i1}) = \sum_{k=0}^{K} b_k^w(\tau)\widetilde{\varphi}_k(\eta_{i1}, age_{i1}) + \gamma^w(\tau)'\mathbf{X}_{i1} \quad (15)$$

for different choices of K and $\widetilde{\varphi}_k$. As in the empirical specification of the portfolio rule, I consider a series quantile model.

Meanwhile, I specify household wealth dynamics[17] via the following equation:

$$\begin{aligned} w_{it} &= h_t(\eta_{it-1}, \varepsilon_{it-1}, w_{it-1}, \alpha_{it-1}, \mathbf{X}_{it}, \tau) \\ &= h(\eta_{it-1}, \varepsilon_{it-1}, w_{it-1}, \alpha_{it-1}, \mathbf{X}_{it}, age_{it}, \tau) \\ &= \sum_{k=1}^{K} b_k^m \widetilde{\varphi}_k(\eta_{it-1}, \varepsilon_{it-1}, w_{it-1}, \alpha_{it-1}, age_{it}) + \gamma^m \mathbf{X}_{it} + b_0^m(\tau) \end{aligned} \quad (16)$$

for some K and $\widetilde{\varphi}_k$. In contrast with (15), I specify equation (16) as a nonlinear regression model. Notice as well that the model is additive in τ. In principle, it can also be specified as a series quantile model; in light of sample size, I resort to this model specification.

Implementation. The functions b_k^a, γ^a, b_k^w, γ^w and b_0^m are indexed by a finite dimensional parameter vector θ, which also contains the coefficients b_k^p's, b_k^m's, γ^p's, and γ^m's. I model the functions b_k^a as piecewise-polynomial interpolating splines on a grid $[\tau_1, \tau_2]$, $[\tau_2, \tau_3]$, …, $[\tau_{L-1}, \tau_L]$, contained in the unit interval. I extend the specification of the intercept coefficient b_0^a on $(0, \tau_1]$ and $[\tau_L, 1)$ with a parametric model indexed by λ^a. All other b_k^a for $k \geq 1$ are constant on these two intervals. Thus, denoting $b_{kl}^a = b_k^a(\tau_l)$, the functions b_k^a depend on $\{b_{11}^a, \dots, b_{KL}^a, \lambda^a\}$. I implement the same modelling for the other functions.

I estimate the portfolio rule by defining a grid from $\tau_1 = 0.20$ to $\tau_L = 0.80$, with a step size equal to 0.10. The functions b_k^a are piecewise-linear, which allows the likelihood to

[17] Implicit in this model of wealth dynamics are two components of the portfolio choice model: first, the empirical consumption rule, and second, the returns on the stock market. In the case of the consumption function, modelling it explicitly would require me to explicitly impose the budget constraint. Meanwhile, as I do not observe individual returns, I do not model them explicitly here, but control for them by introducing time fixed effects. However, the average derivative effect of wealth with respect to income can provide information on the correlation between income and the stock return.

be specified in closed form. In addition, b_0^Q is specified as the quantile of an exponential distribution on $[0, \tau_1)$ (with parameter λ_-^a) and $[\tau_L, 1)$ (with parameter λ_+^a).[18]

Meanwhile, I define $\tau_l = l/L$ and $L = 10$ for the functions that correspond to the initial wealth distribution. I set the wealth accumulation functions b_0^m equal to $\mu + \sigma\Phi^{-1}(\tau)$, where (μ, σ) are parameters to be estimated. I use tensor products of Hermite polynomials for φ_k and $\widetilde{\varphi}_k$, although in practice, other specifications could be used, such as B-splines or wavelets. Each component of the product takes a standardized variable as an argument.[19] Lastly, the life-cycle controls in all of the empirical specifications are cohort and geographical dummies, family size, number of children, and education. I also control for time dummies to take into account aggregate effects.

4.3 Estimation strategy

From the point of estimation, I need to overcome two key challenges. These are: (i.) the well-known sample selection problem and (ii.) the fact that both the persistent and transitory components of labor income, η_{it} and ε_{it}, are unobserved. I discuss how I deal with each of these problems separately. A more technical treatment of the estimation strategy is outlined in Appendix E.

Sample selection. To deal with sample selection, I consider a quantile generalization of the sample selection model (Arellano and Bonhomme (2017a)) and a randomly censored quantile model (Buchinsky and Hahn (1998)) as an alternative estimation strategy. The main idea of the quantile selection model is to suitably shift the percentile ranks from the latent to the observed quantiles of risky asset shares. Through this, I will be able to recover the latent distribution of risky asset shares, and consequently, the empirical portfolio and participation rules. The quantile selection model is appealing in that it does not require a distributional assumption on the error terms, which yields consistency with a wide class of structural models. In order to describe the conditional distribution of the error terms, Arellano and Bonhomme (2017a) resort to a conditional copula model.[20]

[18] More specifically,

$$b_0^Q = \frac{1}{\lambda_-^Q}\log\left(\frac{\tau}{\tau_1}\right)\mathbf{1}\{0 < \tau < \tau_l\} + \sum_{l=1}^{L-1}\left(b_{kl}^Q + \frac{b_{kl+1}^Q - b_{kl}^Q}{\tau_{l+1} - \tau_l}(\tau - \tau_l)\right)\mathbf{1}\{\tau_l \leq \tau < \tau_{l+1}\}$$
$$-\frac{1}{\lambda_+^Q}\log\left(\frac{1-\tau}{1-\tau_L}\right)\mathbf{1}\{\tau_L \leq \tau < 1\}$$

[19] For example, the portfolio rule arguments are $(\eta_{it} - mean(\eta))/sd(\eta)$,$(\varepsilon_{it} - mean(\varepsilon))/sd(\varepsilon)$, $(w_{it} - mean(w))/sd(w)$and $(age_{it} - mean(age))/sd(age)$.

[20] Formally, a copula is a joint distribution function that permits the characterization of dependence between two or more variables. Specifically, in the bivariate case, the joint c.d.f. between two variables x and y is: $F_{X,Y}(x, y) = C(F_X(x), F_Y(y))$.

The estimation procedure of the quantile selection model consists of three steps. To illustrate, I will first assume that the persistent and transitory components are observable. The first step consists of estimating the empirical participation rule (7) via sieve maximum likelihood:

$$\max_{(b_0^P,\ldots,b_K^P,\gamma^P)} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} d_{it}\log\left[\Phi\left(\sum_{k=0}^{K} b_k^p\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})+\gamma^p\mathbf{Z}_{it}\right)\right]$$
$$+(1-d_{it})\log\left[1-\Phi\left(\sum_{k=0}^{K} b_k^p\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})+\gamma^p\mathbf{Z}_{it}\right)\right]. \quad (17)$$

This allows me to recover the conditional choice probabilities $p(\mathbf{Z}_{it})$ that will be used to shift the percentile ranks.

In the second step, I estimate the correlation parameter of the conditional copula that describes the joint distribution of the unobserved arguments of the empirical participation and portfolio rules, ρ_c, by considering the following objective function:

$$\rho_c = \arg\min_{c\in\mathcal{C}}\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} d_{it}\Upsilon(\tau_l,\mathbf{x}_{it})\left[\mathbf{1}\left\{\Lambda^{-1}(\alpha_{it}^*)\leq\sum_{k=0}^{K} b_{kl}^a(c)\varphi_k(\cdot)\right\}-G(\tau_l,p(\mathbf{z}_{it});c)\right]\right\| \quad (18)$$

where $\tau_1,\ldots,\tau_L$ is a finite grid on (0,1), $\|\cdot\|$ is the Euclidean norm,

$$G(\tau_l,p(\mathbf{z}_{it});c)=\frac{C(\tau_l,p(\mathbf{z}_{it});\rho_c)}{p(\mathbf{z}_{it})}$$

is the conditional copula of the error terms of the portfolio and participation rules, $\Upsilon(\cdot)$ are instrument functions, and

$$b_{kl}^a(c)=\arg\min_{b\in\mathcal{B}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} d_{it}\left[G(\tau_l,p(\mathbf{z}_{it});c)\left(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K} b_k^a(\tau)\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})\right)^+\right.$$
$$\left.+(1-G(\tau_l,p(\mathbf{z}_{it});c))\left(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K} b_k^a(\tau)\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})\right)^-\right] \quad (19)$$

The objective function (18) is the result of the conditional moment restrictions that permit the mapping of the distribution of latent outcomes to the distribution of observed outcomes conditional on participation. In the context of the economic model, the latent outcomes are the portfolio shares of the participation subproblem; meanwhile, the observed outcomes are the observed portfolio shares, the solution of the entire problem. More formally, the mapping is:

$$\Pr(\alpha_{it}^*\leq g_t(\eta_{it},\varepsilon_{it},w_{it},age_{it}))=G(\tau_l,p(\mathbf{z}_{it});\rho_c). \quad (20)$$

As can be observed from (20), $G(\tau_l,p(\mathbf{z}_{it});\rho_c)$ is the function that maps the two distributions.

Third, given the estimates of the conditional choice probabilities and the estimates of the correlation parameter, I can then rotate the "check" (or asymmetric absolute loss) function, and estimate the parameters of the empirical portfolio rule (5) via nonparametric quantile regression. More formally, for a given quantile τ_l, I solve the following optimization problem:

$$\min_{(b_0^a,\ldots,b_K^a)} \sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{m=1}^{M} d_{it}\Bigg[G(\tau_l,\widehat{p}(\mathbf{z}_{it});\widehat{\rho}_c)\left(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K} b_{kl}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})+\gamma^a(\tau_l)'\mathbf{X}_{it}\right)^+$$

$$+(1-G(\tau_l,\widehat{p}(\mathbf{z}_{it});\widehat{\rho}_c)))\left(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K} b_{kl}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})+\gamma^a(\tau_l)'\mathbf{X}_{it}\right)^-\Bigg] \quad (21)$$

The optimization problem (21) is equivalent to minimizing a rotated check function, with an individual-specific perturbed τ (Arellano and Bonhomme (2017a)). The rotation preserves the linear programming formulation, and thus, the computational simplicity, of quantile regression. To correct for selection, I replace τ with the individual specific-perturbed τ, $G(\tau_l,\widehat{p}(\mathbf{z}_{it});\widehat{\rho}_c)$.

Notice that the model relies on an exclusion restriction to achieve identification of the empirical participation and portfolio rules. In this regard, I consider the lagged value of lifetime wealth, following Vissing-Jørgensen (2002), Bonaparte et al. (2014), and Fagereng et al. (2017a). The motivation behind this can be easily seen in the portfolio rule of the two-period model with CRRA risk preferences (Campbell and Viceira (2002)), which can be written as:

$$\alpha_t = \left(1+\frac{\widetilde{H}}{W_t}\right)\left[\frac{\mathbb{E}_t(r_{t+1}-r_f)+\frac{1}{2}\sigma_r^2}{\gamma\sigma_r^2}\right],$$

in which $\tilde{H}$ is human wealth, or the discounted present value of future labor income, γ is the risk aversion parameter, and σ_r^2 is the variance of the portfolio return. As can be observed, the optimal portfolio rule is a function of the ratio between human and total household wealth, and not on the level of lifetime wealth. I explain how to calculate this variable in Appendix C. [21]

Alternative identification scheme. A potential concern is the validity of the exclusion restriction. In other words, it could be the case that the variables that determine

[21]In results that I do not present here, I examine the validity of the restriction by running Heckman (1979)-type regressions on simulated data from the structural model in Galvez and Paz-Pardo (2021). The results suggest that the exclusion restriction is valid, as the coefficient of the lagged value of lifetime wealth is significant in the participation regression at a level of 8 percent, while insignificant in the asset allocation regression.

participation are the same ones that determine the portfolio rule. To address this issue, I consider the censored quantile regression estimator of Buchinsky and Hahn (1998).

The choice of this estimator over similar estimation methods (e.g., Powell (1986), Chernozhukov and Hong (2002), Honore et al. (2002), Khan and Tamer (2009)) is primarily motivated by three reasons. First, the nonlinear semi-reduced form described by equations (5)–(9) when $\mathbf{X}_{it} = \mathbf{Z}_{it}$ whittles down to a model with a random censoring point, which can be considered as the household's wealth threshold. This rules out Powell (1986) and Chernozhukov and Hong (2002), who both consider models with a fixed censoring point. Second, Buchinsky and Hahn (1998) propose an estimation method that is computationally tractable, as it also results in a convex optimization problem. Though both Honore et al. (2002) and Khan and Tamer (2009) consider more general models with random censoring, their proposed estimation methods are computationally more demanding.[22] Third, and most importantly, the estimator can be interpreted as the limiting case of the more general quantile selection model of Arellano and Bonhomme (2017a). A distinction between the two estimators is that the quantile selection model allows the recovery of the latent distribution of risky shares, while the censored regression model allows the recovery of the observed distribution. I outline the model specification and the estimation algorithm in Appendix E.2.

Simulation-based algorithm. If the persistent and transitory components were observable, I can recover the empirical participation and portfolio rules from the quantile selection model I described earlier. Given that this is not the case, I rely on simulation-based approaches to obtain consistent parameters of the empirical participation and portfolio rules, plus the parameters of the wealth dynamics equations. In particular, I utilize a stochastic-EM like algorithm for time-varying latent variables, which is a simulated version of the classical EM algorithm (Dempster et al. (1977)).

Starting from an initial guess $\widehat{\theta}$, the algorithm iterates on the following two steps until convergence of the sequence of parameter estimates:

1. *Stochastic E-step:* Draw pseudo-data of the persistent component $\eta_i^{(m)}$, for $m = 1, \ldots, M$, from the posterior distribution of the latent persistent component of earnings given the observed data, $f_i(\cdot; \widehat{\theta}^{(s)})$.

2. *M-step:* Estimate the parameters $\widehat{\theta}^{(s+1)}$ of the empirical portfolio and participation rules via the quantile selection model with moment conditions (17)–(21), and the

[22]Honore et al. (2002) propose a procedure that does not yield a convex optimization problem, while Khan and Tamer (2009) propose a moment inequality-based approach.

parameters of the wealth dynamics equations via nonlinear least squares. To see how this is operationalized, the corresponding estimation for the conditional choice probability (17) is:

$$\max_{(b_0^P,\ldots,b_K^P,\gamma^P)} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} d_{it}\log\left[\Phi\left(\sum_{k=0}^{K} b_k^p\varphi_k(\eta_{it}^{(m)},\varepsilon_{it},w_{it},age_{it})+\gamma^p\mathbf{Z}_{it}\right)\right]$$
$$+(1-d_{it})\log\left[1-\Phi\left(\sum_{k=0}^{K} b_k^p\varphi_k(\eta_{it}^{(m)},\varepsilon_{it},w_{it},age_{it})+\gamma^p\mathbf{Z}_{it}\right)\right].$$

Notice that in this step, I use the pseudo-data previously generated as if they were true observations of persistent income.

Because the likelihood function has a closed form, the stochastic E-step is straightforward to implement. In practice, I take $M = 1$, stop the chain after a large number of iterations, and report an average across the last $\bar{S}$ values $\hat{\theta} = \frac{1}{\bar{S}}\sum_{s=S-\bar{S}+1}^{S}\hat{\theta}^{(s)}$, where I take $\bar{S} = S/2$. Each estimation is based on $S = 200$ iterations, with 200 random walk Metropolis-Hastings draws per iteration. I target the variance of the proposal distributions to obtain an acceptance rate of approximately 30 percent. Inference is based on parametric and nonparametric bootstrap.

5 Results

In this section, I present empirical results. I first begin by showing average derivative effects of the empirical participation and portfolio rules.[23] I then report simulation exercises based on the estimated model. Finally, I compute implied stock market participation costs based on a framework by Vissing-Jørgensen (2002).

5.1 Average derivative effects

Figure 3 shows the average derivative effect, with respect to η_{it}, of the propensity score $p(\mathbf{z}_{it})$ and the conditional mean of the latent risky asset share α_{it}, respectively, evaluated at different percentiles of wealth τ_{wealth} and age τ_{age}, and averaged over η_{it}, based on the estimated nonlinear model. In general, with the exception of young households with little wealth, the average derivative effect of an increase in persistent income is positive, and ranges from 0.97 to 5 percent. Meanwhile, with respect to the risky asset share, the

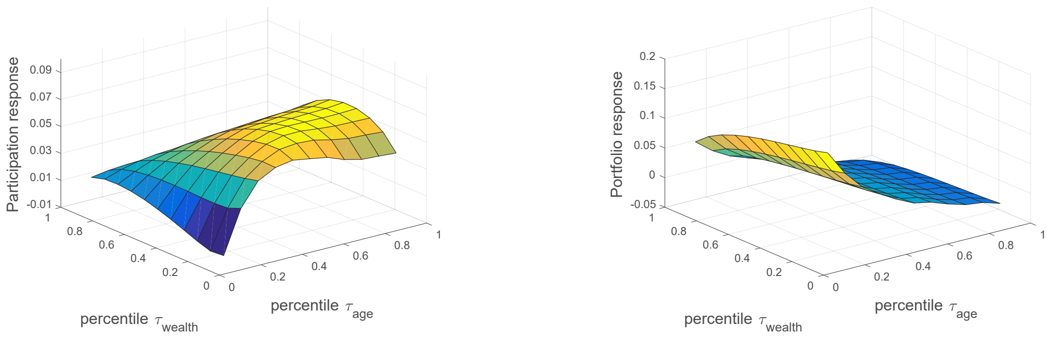
[23] In the estimation of the participation and portfolio rules, I use tensor products of Hermite polynomials with degrees (2,1,1,1). The choice of the following specification is motivated by previous literature (e.g., Guiso et al. (1996) and Vissing-Jørgensen (2002)), among others, who model income with quadratic terms.

average derivative effect is positive for young households (regardless of wealth levels), and negative for older households. This result suggests that even if there are increases in income, older households might prefer to rebalance their portfolios, as they are approaching retirement. However, this interpretation would not be exact without specific assumptions on preferences. Figure F1 in Appendix F shows the results with the censored quantile regression estimator of Buchinsky and Hahn (1998). As the graphs indicate, the results are quite similar. Meanwhile, Figure F11 in Appendix F presents the 95% confidence bands calculated via parametric bootstrap. As the results indicate, however, the effects do not seem to be precisely estimated.

Figure 3: Average derivative effect of the persistent component of income η_{it}

(a) Propensity score

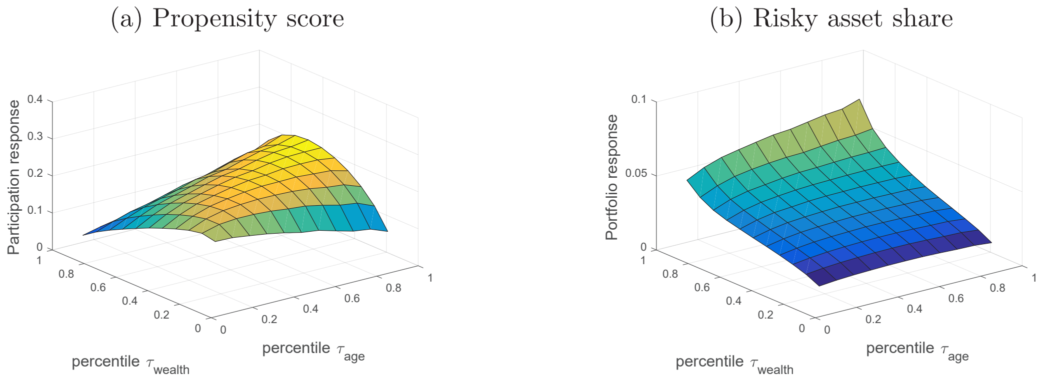
(b) Risky asset share



Note: The graphs show average derivatives of the propensity score and the risky asset share, respectively, with respect to the persistent income η_{it}, given w_{it}, persistent component η_{it}, income y_{it}, and age age_{it}, and evaluated at different values of w_{it} and age_{it} that correspond to their τ_{wealth} and τ_{age} percentiles. All results are based on estimates from the semi-structural model with the Arellano and Bonhomme (2017a) quantile selection model estimator.

Figure 4 shows the average derivative effect, with respect to wealth w_{it}, of the propensity score $p(\mathbf{z}_{it})$ and the conditional mean of the latent risky asset share α_{it}, respectively, evaluated at percentiles of wealth τ_{wealth} and age τ_{age}, and averaged over η_{it}, based on the estimated nonlinear model. The average derivative effect of wealth on the extensive margin is positive and nonlinear across the wealth and age dimensions. In particular, the effect is positive and concave along the wealth distribution, while it is positive and mostly increasing along the age distribution. One can rationalize this result as being consistent with the idea of heterogeneous participation costs, an object I will recover in section 5.3. This result also reinforces the idea that households have different wealth thresholds when they decide on their portfolio allocations. Meanwhile, the average derivative effect of wealth along the intensive margin are positive, and increasing across different quantiles of wealth and age. While these results might suggest that households exhibit DRRA preferences, which is consistent with the findings of Calvet and Sodini (2014), note that these average derivative effects might be a mixture of unobserved preferences and risk. The increasing result, meanwhile, suggests that perhaps, the DRRA preferences are heterogeneous over the age and wealth distributions. Figure F2 in Appendix F shows the results with the censored quantile regression estimator of Buchinsky and Hahn (1998). The results are quite similar. Figure F12 in Appendix F shows the confidence bands calculated using parametric bootstrap. Again, as in the results with respect to persistent income, the estimates seem to be precisely estimated.

Figure 4: Average derivative effect of wealth



Note: The graphs show average derivatives of the propensity score and the risky asset share, respectively, with respect to wealth w_{it} given w_{it}, persistent component η_{it}, income y_{it}, and age age_{it}, evaluated at different values of w_{it} and age_{it} that correspond to their τ_{wealth} and τ_{age} percentiles. All results are based on estimates from the semi-structural model with the Arellano and Bonhomme (2017a) quantile selection model estimator.

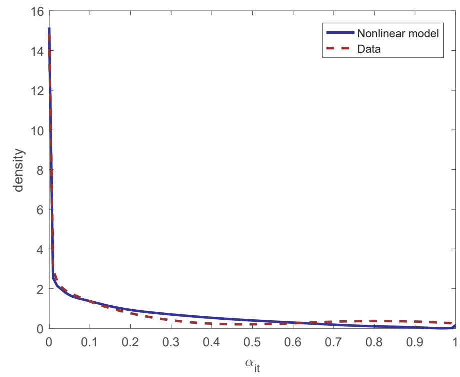
Finally, I evaluate the fit implied by the nonlinear model with that of the data. As the results indicate, 57.28 percent of households participate according to the nonlinear model, which is close to the observed 57.03 percent in the sample that I use for estimation.[24] Moreover, as the graph of the density of risky asset shares indicates, the nonlinear model provides a close fit with the data. Looking at the figure implied by the estimation with the Buchinsky and Hahn (1998) estimator (Figure F3), I find that the fit is quite similar.

[24]The average participation rate in the sample I use for estimation is different from that in Section 3, as my exclusion restriction is a lagged variable.

5.2 The impact of persistent earnings shocks

In this subsection, I simulate the life-cycle portfolio choice model with participation according to the nonlinear model, and show the extensive and intensive margin response with respect to a persistent income shock. In the simulation exercise, I report the difference between two types of households: households who are hit at age 37 by a large negative shock to the persistent component ($\tau_{shock} = 0.1$), or by a large positive shock ($\tau_{shock} = 0.9$), and households who are hit by a median shock ($\tau_{shock} = 0.5$) to the persistent component.[25] I report age-specific means across 250,000 simulations. I compare three different kinds of households: those who have low income ($\tau_{init} = 0.1$), middle income ($\tau_{init} = 0.5$), or high income ($\tau_{init} = 0.9$).

Figure 5: Observed and implied densities of the risky asset share



Note: The graph shows the observed and predicted unconditional densities of the share of household wealth in risky asset share based on the nonlinear model. The blue line corresponds to the density implied by the nonlinear model, while the red line corresponds to the density implied by the data. All results are based on estimates from the semi-structural model with the Arellano and Bonhomme (2017a) quantile selection model estimator.

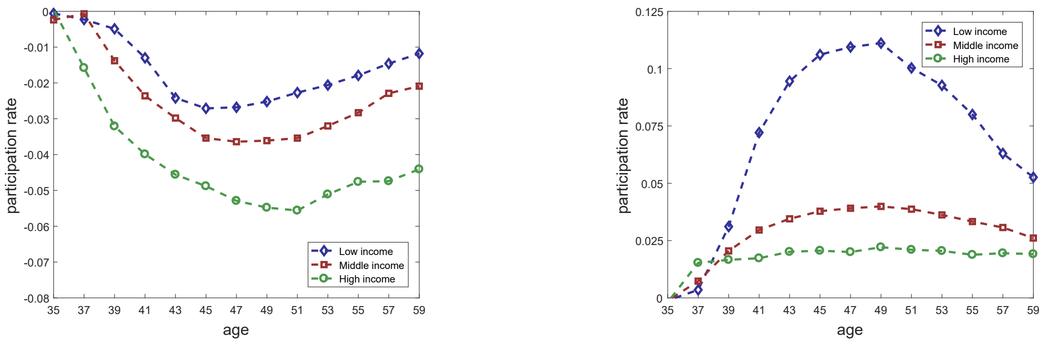
Extensive margin responses. In Figure 6, I report the results with respect to the conditional probability of participation, i.e., the extensive margin. The results show asymmetric extensive margin responses to large income shocks. The results also highlight the interaction between the rank of the household in the distribution of the initial earnings (τ_{init}) and the size of the shock received (τ_{shock}). In particular, a large negative shock results in a decrease in participation of as much as 6 percentage points for high income

[25]Note that the notion of shocks here are taken with respect to the rank of the household in the income distribution.

households, and a 2.9 percentage point decrease for low income households. Meanwhile, a large positive shock yields an increase in participation of as much as 12 percentage points for low-income households, compared to 2.2 percentage points for low-income households. I observe similar patterns for the Buchinsky and Hahn (1998) censored quantile regression estimator, as can be observed in Figure F4 of Appendix F. In addition, I report results from estimating a linear portfolio choice rule with the standard linear earnings process.

Figure 6: Impulse response, participation rule

(a) $\tau_{shock} = 0.1$ (b) $\tau_{shock} = 0.9$



Note: The graphs show the difference in average participation rates between a household hit by a shock τ_{shock} at age 37, and a household hit by 0.5 shock at the same age. The blue line corresponds to low-income households (i.e., rank of $\tau_{init} = 0.1$ in the income distribution). The red line corresponds to middle-income households (i.e., rank of $\tau_{init} = 0.5$ in the income distribution). The green line corresponds to high-income households (i.e., rank of $\tau_{init} = 0.9$ in the income distribution). All results are based on estimates from the semi-structural model with the Arellano et al. (2017) quantile selection estimator.

As Figure F9 indicates, the fact that the standard model does not allow for interactions between the household's rank in the income distribution and the type of shock received appears to be at odds with the data.[26]

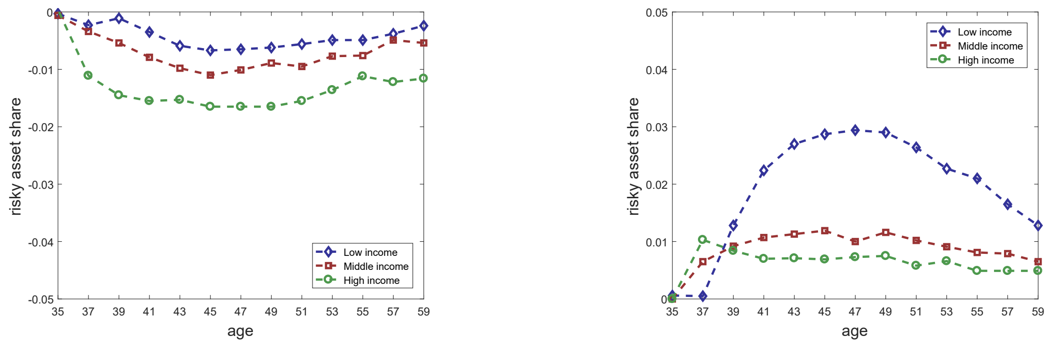
Intensive margin responses. In Figure 7, I look at the differences with respect to the average risky asset share for stock market participants at age 35. As the results indicate, a large negative income shock yields a 2.1 percentage point decrease in average risky asset shares for high income households. In comparison, the same shock yields a 0.45 percentage point decrease for low income households. A large positive income shock, meanwhile, results in a 3 percentage point increase in the share of wealth invested

[26]The estimated model is a Tobit regression where α_{it} is modeled as a linear function of η_{it}, ε_{it}, w_{it}, and age, and was estimated via maximum likelihood methods with the stochastic EM algorithm.

in risky assets for low income households. In comparison, the same shock results in a 1 percentage point increase for high income households. Looking at the corresponding figure in the appendix (Figure F6) for the censored quantile regression estimator, I observe similar results. These results suggest that the persistence of earnings histories is crucial in understanding changes in risky asset shares with respect to asymmetries in income risk. Figure F10 presents results from estimating a linear portfolio choice rule with the standard linear earnings process. Again, the results assume away the presence of interaction effects between the household's rank in the income distribution and the type of shock received.

Figure 7: Impulse response, portfolio rule

(a) $\tau_{shock} = 0.1$ (b) $\tau_{shock} = 0.9$

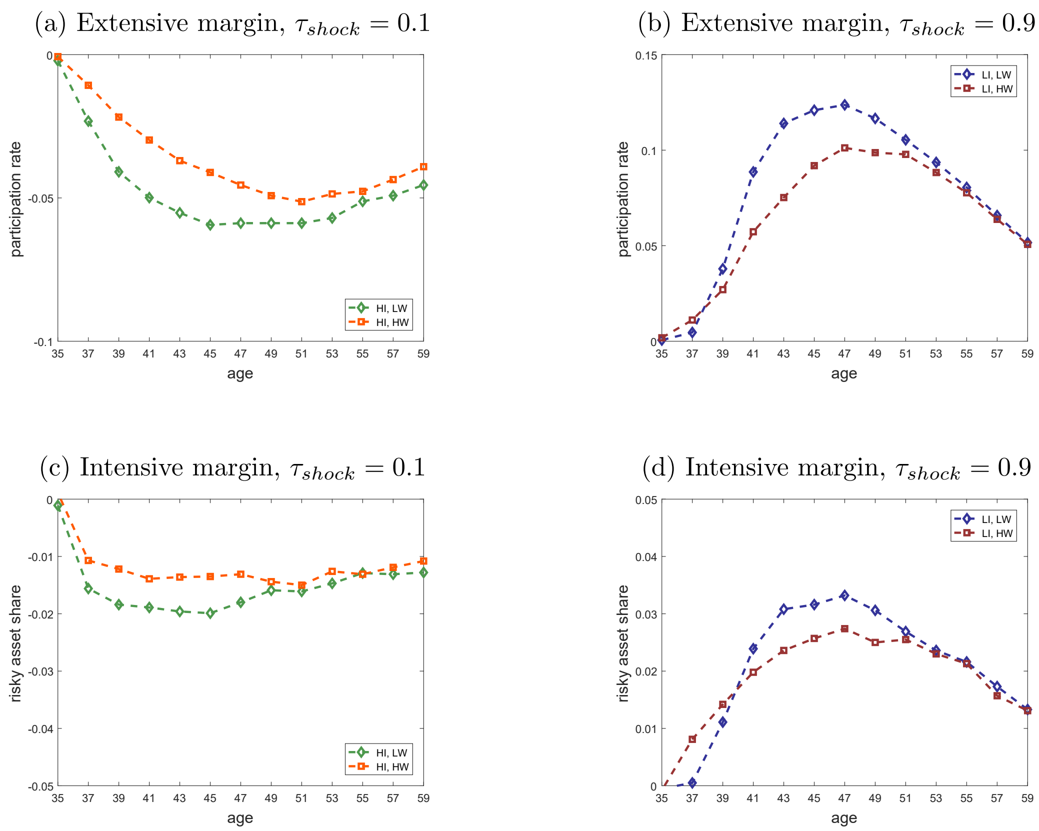


Note: The graphs show the difference in average portfolio shares conditional on participation between a household hit by a shock τ_{shock} at age 37, and a household hit by 0.5 shock at the same age. The blue line corresponds to low-income households (i.e., rank of $\tau_{init} = 0.1$ in the income distribution). The red line corresponds to middle-income households (i.e., rank of $\tau_{init} = 0.5$ in the income distribution). The green line corresponds to high-income households (i.e., rank of $\tau_{init} = 0.9$ in the income distribution). All results are based on estimates from the semi-structural model with the Arellano et al. (2017) quantile selection estimator.

Interactions with wealth and age. Finally, Figures 8 and 9 present similar simulation exercises, but varying the timing of the shocks and the amount of wealth that the households possess. Comparing the magnitudes to households who are hit by the same shock when they are old, as shown in Figure 9, I find that the extensive margin responses are stronger for younger households than for older households. Meanwhile, intensive margin responses seem to be similar for both households. The results suggest the importance of human capital as a factor in portfolio choice decisions. Figures F7 and F8 in the appendix reveal similar patterns when the relevant estimator is the censored regression estimator of Buchinsky and Hahn (1998), with the exception of a very positive income shock for low income households at old age.

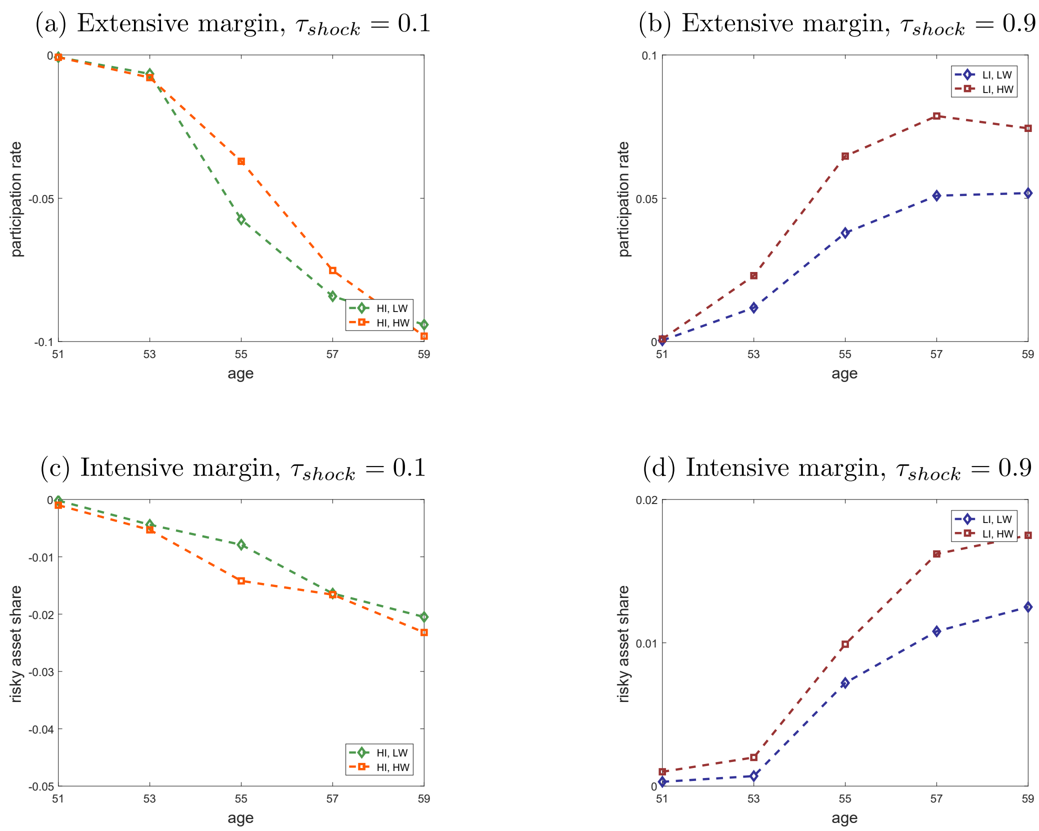
Discussion. The results of the simulation exercises highlight the importance of the persistence of earnings histories and large income shocks for household portfolio choice decisions. In particular, the analyses suggest that for low-income households, labor mar-sifying their investments into potentially risky, but high return assets such as stocks. In contrast, for high-income households who are more likely to be stock market participants, an "unusually" negative income shock (such as the risk of job loss or an adverse health shock, as in Rosen and Wu (2004) and Edwards (2008), for example) results in an exit from the stock markets, as the risk that they face from their future labor income becomes sufficiently high enough that they would like to reduce their exposures to other sources of risk, and in particular, to financial risk. Hence, the impulse response exercises suggest that the nonlinear earnings process allows for the possibility of "unusual" income shocks to wipe out the memory of past income shocks.

Figure 8: Impulse responses by income and wealth quantiles, at age 35



Note: The graphs show the difference between a household hit by a shock τ_{shock} at age 37, and a household hit by a 0.5 shock at the same age, by income and wealth categories. The blue line corresponds to low income, low wealth households. The red line corresponds to low income, high wealth households. The green line corresponds to high income, low wealth households. The orange line corresponds to high income, high wealth households. All results are based on estimates from the semi-structural model with the Arellano et al. (2017) quantile selection estimator.

Figure 9: Impulse responses by income and wealth quantiles, at age 51



(a) Extensive margin, $\tau_{shock} = 0.1$

(b) Extensive margin, $\tau_{shock} = 0.9$

(c) Intensive margin, $\tau_{shock} = 0.1$

(d) Intensive margin, $\tau_{shock} = 0.9$

Note: The graphs show the difference between a household hit by a shock τ_{shock} at age 53, and a household hit by a 0.5 shock at the same age, by income and wealth categories. The blue line corresponds to low income, low wealth households. The red line corresponds to low income, high wealth households. The green line corresponds to high income, low wealth households. The orange line corresponds to high income, high wealth households. All results are based on estimates from the semi-structural model with the Arellano et al. (2017) quantile selection estimator.

5.3 What are the implied costs of stock market participation?

In this final subsection, I calculate the implied costs of stock market participation. The approach that I follow builds on the framework outlined in Vissing-Jørgensen (2002). Assuming that preferences are time-separable and homothetic, the per-period benefit to stock market participation is approximately:

$$Benefit_{it} \approx (r_{it}^{ce} - r_f) \times \alpha_{it}^{*} \times W_{it} \tag{22}$$

where r_{it}^{ce} is the certain return that would make a household indifferent between investing in a risky asset and investing in an asset with a certain return of r_{it}^{ce}. α_{it}^{*} is the desired risky asset share, and W_{it} is household i's wealth at time t. An estimate of the per-period participation cost c_{it} can be obtained, hence, as the dollar amount required to offset this benefit.

As the nonlinear model allows me to recover the latent distribution of risky asset shares, I can calculate the following two objects. The first corresponds to c^l_{it}, which can be interpreted as the lower bound for participation costs for household i in period t[27], that is,

$$c^l_{it} = (r^{ce}_{it} - r_f) \times \alpha^*_{it} \times W^{np}_{it}, \tag{23}$$

in which W^{np}_{it} is the wealth of a non-participating household i. The second, meanwhile, corresponds to c^u_{it}, which can be interpreted as an upper bound for participation costs for household i in period t:

$$c^u_{it} = (r^{ce}_{it} - r_f) \times \alpha^*_{it} \times W^{p}_{it}, \tag{24}$$

in which W^p_{it} is the wealth of a participating household i. An interpretation of c^l_{it} is that it is the minimum per-period participation cost that deters nonparticipants from entering the stock market, while c^u_{it} is the minimum per-period benefit that stock market participants gained from buying stocks. Hence, conditioning on wealth W_{it}, I can calculate the following cost distributions:

$$F(c^l_{it}|d_{it} = 0, W_{it}) \text{ and } F(c^u_{it}|d_{it} = 1, W_{it}).$$

To estimate participation cost bounds, I perform the following procedure. As in Vissing-Jørgensen (2002), I assume that $r^{ce}_{it} - r_f = 0.04$. I then calculate the quantiles of latent risky asset shares as implied by the nonlinear model. Finally, I calculate c^l_{it} and c^u_{it} following formulas (23) and (24), for a given W_{it}.

Table 4 provides estimates of the participation cost distributions faced by nonparticipants and participants, respectively. I calculate this for three types of households: those who have low wealth, those who have median wealth, and those who have high wealth. As the results indicate, there is a wide dispersion on the participation costs faced by both non-participants and participants, respectively. Notice as well, that the costs of stock market participation lie in between 0 to 2 percent of a household's wealth, regardless of household type. The fact that these are costs are small suggests that there are other fac-

[27] Luttmer (1999) (using aggregate data) and Paiella (2007) (using the US Consumer Expenditure Survey) both estimate the foregone gains to stock market participation via a moment inequality approach.

tors that are play which deter households from buying stocks, for example, risk aversion or investment in housing.

Table 4: Implied distribution of per-period stock market participation costs

	Quantiles of latent risky asset shares				
	10	25	50	75	90
α_{it}^*	0.001	0.019	0.076	0.293	0.773
			Non-participants		
Low wealth	0.63	5.58	47.53	194.89	475.93
Median wealth	3.41	30.05	256.06	1,409.98	2,563.84
High wealth	13.06	114.97	979.84	4,017.56	9.810.93
			Participants		
Low wealth	2.86	25.15	214.37	878.96	2,146.43
Median wealth	10.75	94.61	806.22	3,305.69	8,072.52
High wealth	39.77	350.07	2,983.25	12,232.06	29,870.72

Note: Data from 1999 to 2009 PSID waves. The table shows the implied per-period stock market participation cost, for non-participants and participants, respectively. The columns correspond to the quantiles of latent risky asset shares recovered from the estimates of the semi-structural model of portfolio choice with the Arellano and Bonhomme (2017a) quantile selection estimator. The rows correspond to households with different wealth levels. Low-wealth households are those at the 10^{th} quantile of the wealth distribution ($15,870 for non-participants (NP) and $71,682 for participants (P)); median-wealth households are those at the 50^{th} quantile of the wealth distribution ($85,819.37 for NP and $268,337.29 for P); and high-wealth households are those at the 90^{th} quantile of the wealth distribution ($327,747.90 for NP $996,097 for P).

6 Conclusion

In this paper, I develop a semi-structural framework to understand the nonlinear transmission of income shocks on household investment behavior. I model stock market participation and portfolio rules as age-dependent functions of the persistent and transitory components of income, and of assets. The model reveals asymmetric participation and portfolio adjustment responses with respect to “unusual” income shocks, which to the best of my knowledge, has not been uncovered in previous literature. These results suggest that the persistence of households earnings histories drive their stock market participation decision; however, “unusual” income shocks, which can correspond to moves up and down the job ladder, or adverse health shocks, can wipe out the past memory of income shocks and induce changes in participation and portfolio choice behavior. The empirical evidence also shown in this paper supports the evidence that households present DRRA preferences, as shown by Calvet and Sodini (2014).

In this paper, I have abstracted from an explicit modelling of aggregate shocks, such as those related to the volatility of stock market returns and its correlation with labor income. On the one hand, as Benzoni et al. (2007), Betermier et al. (2012), and Bonaparte et al. (2014) show, the correlation between the stock market return and human capital might induce households to hedge against their income risks through the use of stocks. Catherine et al. (2022) also emphasize the importance of skewness risk and its covariance with the stock return as a driver of portfolio allocations. Catherine (2022) also shows, in a structural model of portfolio choice with countercyclical income risk, that portfolio rules might have monotonicities in wealth and income. On the other hand, Schmidt (2015) emphasizes that investing in stocks is a poor hedge against potentially disastrous income shocks. A rigorous analysis of the income hedging motives requires extending the empirical framework I present here to incorporate aggregate shocks.

Finally, it might be important to simultaneously consider the demand for other assets; in particular, housing. As underscored by a large literature, for most households, their abode constitutes the biggest share of their wealth (see Davis et al. (2015)). To this end, extending the framework I present here to consider features inherent in studying models of housing demand, such as movements from renting to homeownership (e.g., Han (2010) and Bajari et al. (2013)), might be relevant to understanding household investment decisions.

Acknowledgements. I am extremely grateful to Manuel Arellano, Javier Mencía and Enrique Sentana for their support. I also thank Dante Amengual, Steffen Andersen (discussant), Stéphane Bonhomme, Agar Brugiavini (discussant), Lukas Freund (discussant), Luigi Guiso, Nezih Guner, Anson T.Y. Ho (discussant), Laura Hospido, Borja Petit, Josep Pijoan, Kjetil Storesletten and several seminar and conference participants for helpful comments and suggestions. I acknowledge funding from the Spanish Ministry of Economics and Competitiveness, grant no. BES-2014-070515-P. A previous version of this paper circulated under the name "Household portfolio choices and nonlinear income risk", and received the 2017 CEPR Household Finance Best Ph.D. Paper prize. All remaining errors are mine.

References

Adda, Jérôme, Christian Dustmann and Katrien Stevens. (2017). "The career costs of children". *Journal of Political Economy*, 125(2), pp. 293-337. https://doi.org/10.1086/690952

Alan, Sule. (2012). "Do disaster expectations explain household portfolios?". *Quantitative Economics*, 3(1), pp. 1-28. https://doi.org/10.3982/QE128

Angerer, Xiaohong, and Pok-Sang Lam. (2009). "Income risk and portfolio choice: an empirical study". *The Journal of Finance*, 64(2), pp. 1037-1055. https://doi.org/10.1111/j.1540-6261.2009.01456.x

Arellano, Manuel, Richard W. Blundell and Stéphane Bonhomme. (2017). "Earnings and consumption dynamics: a nonlinear panel data framework". *Econometrica*, 85(3), pp. 693-734. https://doi.org/10.3982/ECTA13795

Arellano, Manuel, and Stéphane Bonhomme. (2017a). "Quantile selection models with an application to understanding changes in wage inequality". *Econometrica*, 85(1), pp. 1-28. https://doi.org/10.3982/ECTA14030

Arellano, Manuel, and Stéphane Bonhomme. (2017b). "Nonlinear panel data methods for dynamic heterogeneous agent models". *Annual Review of Economics*, 9, pp. 471-496. https://doi.org/10.1146/annurev-economics-063016-104346

Bajari, Patrick, Phoebe Chan, Dirk Krueger and Daniel Miller. (2013). "A dynamic model of housing demand: Estimation and policy implications". *International Economic Review*, 54(2), pp. 409-442. https://doi.org/10.1111/iere.12001

Basten, Christoph, Andreas Fagereng and Kjetil Telle. (2016). "Saving and portfolio allocation before and after job loss". *Journal of Money, Credit and Banking*, 48(2-3), pp. 293-324. https://doi.org/10.1111/jmcb.12301

Benzoni, Luca, Pierre Collin-Dufresne and Robert S. Goldstein. (2007). "Portfolio choice over the life-cycle when the stock and labor markets are cointegrated". *The Journal of Finance*, 62(5), pp. 2123-2167. https://doi.org/10.1111/j.1540-6261.2007.01271.x

Betermier, Sebastien, Thomas Jansson, Christine Parlour and Johan Walden. (2012). "Hedging labor income risk". *Journal of Financial Economics*, 105(3), pp. 622-639. https://doi.org/10.1016/j.jfineco.2012.05.001

Blundell, Richard, Xiaohong Chen and Dennis Kristensen. (2007). "Semi-nonparametric IV estimation of shape-invariant Engel curves". *Econometrica*, 75(6), pp. 1613-1669. https://doi.org/10.1111/j.1468-0262.2007.00808.x

Blundell, Richard, Luigi Pistaferri and Ian Preston. (2008). "Consumption inequality and partial insurance". *The American Economic Review*, 98(5), pp. 1887-1921. https://doi.org/10.1257/aer.98.5.1887

Blundell, Richard, Luigi Pistaferri and Itay Saporta-Eksten. (2016). "Consumption inequality and family labor supply". *The American Economic Review*, 106(2), pp. 387-435. https://doi.org/10.1257/aer.20121549

Bonaparte, Yosef, George M. Korniotis and Alok Kumar. (2014). "Income hedging and portfolio decisions". *Journal of Financial Economics*, 113, pp. 300-324. https://doi.org/10.1016/j.jfineco.2014.05.001

Bottai, Matteo, Bo Cai and Robert E. McKeown. (2010). "Logistic quantile regression for bounded outcomes". *Statistics in Medicine*, 29(2), pp. 309-317. https://doi.org/10.1002/sim.3781

Briggs, Joseph S., David Cesarini, Erik Lindqvist and Robert Östling. (2015). "Windfall gains and stock market participation". Technical Report, National Bureau of Economic Research. https://doi.org/10.3386/w21673

Brunnermeier, Markus K., and Stefan Nagel. (2008). "Do wealth fluctuations generate time-varying risk aversion? Micro-evidence on individuals". *American Economic Review*, 98(3), pp. 713-736. https://doi.org/10.1257/aer.98.3.713

Buchinsky, Moshe. (1995). "Quantile regression, box-cox transformation model, and the U.S. wage structure, 1963-1987". *Journal of Econometrics*, 65(1), pp. 109-154. https://doi.org/10.1016/0304-4076(94)01599-U

Buchinsky, Moshe, and Jinyong Hahn. (1998). "An alternative estimator for the censored quantile regression model". *Econometrica*, pp. 653-671. https://doi.org/10.2307/2998578

Bucks, Brian K., Arthur B. Kennickell, Traci L. Mach and Kevin B. Moore. (2009). "Changes in U.S. family finances from 2004 to 2007: evidence from the survey of consumer finances". *Federal Reserve Bulletin*. https://doi.org/10.17016/bulletin.2009.95-2

Calvet, Laurent E., John Y. Campbell and Paolo Sodini. (2007). "Down or out: Assessing the welfare costs of household investment mistakes". *Journal of Political Economy*, 115(5), pp. 707-747. https://doi.org/10.1086/524204

Calvet, Laurent E., and Paolo Sodini. (2014). "Twin Picks: Disentangling the determinants of risk-taking in household portfolios". *The Journal of Finance*, 69(2), pp. 867-906. https://doi.org/10.1111/jofi.12125

Campbell, John Y., and Luis M. Viceira. (2002). *Strategic asset allocation: portfolio choice for long-term investors*. Oxford University Press. https://doi.org/10.1093/0198296940.001.0001

Catherine, Sylvain. (2022). "Countercyclical labor income risk and portfolio choices over the life cycle". *The Review of Financial Studies*, 35(9), pp. 4016-4054. https://doi.org/10.1093/rfs/hhab136

Catherine, Sylvain, Paolo Sodini and Yapei Zhang. (2022). "Countercyclical income risk and portfolio choices: Evidence from Sweden". *Swedish House of Finance Research Paper*, 20-20.

Chamberlain, Gary. (1994). "Quantile regression, censoring, and the structure of wages". In *Advances in Econometrics: Sixth World Congress*, 2, pp. 171-209. https://doi.org/10.1017/CCOL0521444594.005

Chen, Xiaohong. (2007). "Large sample sieve estimation of semi-nonparametric models". *Handbook of Econometrics*, 6, pp. 5549-5632. https://doi.org/10.1016/S1573-4412(07)06076-X

Chernozhukov, Victor, and Han Hong. (2002). "Three-step censored quantile regression and extra-marital affairs". *Journal of the American Statistical Association*, 97(459). https://doi.org/10.1198/016214502388618663

Chetty, Raj, László Sándor and Adam Szeidl. (2017). "The effect of housing on portfolio choice". *The Journal of Finance*, 72(3), pp. 1171-1212. https://doi.org/10.1111/jofi.12500

Chiappori, Pierre-André, and Monica Paiella. (2011). "Relative risk aversion is constant: Evidence from panel data". *Journal of the European Economic Association*, 9(6), pp. 1021-1052. https://doi.org/10.1111/j.1542-4774.2011.01046.x

Cocco, Joao F., Francisco J. Gomes and Pascal J. Maenhout. (2005). "Consumption and portfolio choice over the life cycle". *Review of Financial Studies*, 18(2), pp. 491-533. https://doi.org/10.1093/rfs/hhi017

Constantinides, George M., and Darrell Duffie. (1996). "Asset pricing with heterogeneous consumers". *Journal of Political Economy*, 104(2), pp. 219-240. https://doi.org/10.1086/262023

Constantinides, George M., and Anisha Ghosh. (2017). "Asset pricing with countercyclical household consumption risk". *The Journal of Finance*, 72(1), pp. 415-460. https://doi.org/10.1111/jofi.12471

Davis, Morris A., Stijn Van Nieuwerburgh. (2015). "Housing, finance, and the macroeconomy". *Handbook of Regional and Urban Economics*, 5, pp. 753-811. https://doi.org/10.1016/B978-0-444-59531-7.00012-0

De Nardi, Mariacristina, Giulio Fella and Gonzalo Paz Pardo. (2018). "Nonlinear household earnings dynamics, self-insurance and welfare". *Journal of the European Economic Association*, 18(2), pp. 890-926. https://doi.org/10.1093/jeea/jvz010

Dempster, Arthur P., Nan M. Laird and Donald B. Rubin. (1977). "Maximum likelihood from incomplete data via the EM algorithm". *Journal of the Royal Statistical Society. Series B (methodological)*, pp. 1-38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Edwards, Ryan D. (2008). "Health risk and portfolio choice". *Journal of Business & Economic Statistics*, 26(4), pp. 472-485. https://doi.org/10.1198/073500107000000287

Fagereng, Andreas, Charles Gottlieb and Luigi Guiso. (2017). "Asset market participation and portfolio choice over the life-cycle". *The Journal of Finance*, 72(2), pp. 705-750. https://doi.org/10.1111/jofi.12484

Fagereng, Andreas, Luigi Guiso and Luigi Pistaferri. (2017). "Portfolio choices, firm shocks and uninsurable wage risk". *Review of Economic Studies*, 85(1), pp. 437-474. https://doi.org/10.1093/restud/rdx023

Flavin, Marjorie, and Takashi Yamashita. (2002). "Owner-occupied housing and the composition of the household portfolio". *American Economic Review*, 92(1), pp. 345-362. https://doi.org/10.1257/000282802760015775

Galvez, Julio, and Gonzalo Paz-Pardo. (2022). "Richer earnings dynamics, consumption and portfolio choice over the life cycle". Documentos de Trabajo, 2241, Banco de España. https://doi.org/10.53479/23706

Gourinchas, Pierre-Olivier, and Jonathan A. Parker. (2002). "Consumption over the life cycle". *Econometrica*, 70(1), pp. 47-89. https://doi.org/10.1111/1468-0262.00269

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

Guiso, Luigi, Paola Sapienza and Luigi Zingales. (2008). "Trusting the stock market". *The Journal of Finance*, 63(6), pp. 2557-2600. https://doi.org/10.1111/j.1540-6261.2008.01408.x

Guiso, Luigi, and Paolo Sodini. (2013). "Household finance: An emerging field". In G. Constandinides, M. Harris and R. Stulz, *Handbook of the Economics of Finance.* Elsevier. https://doi.org/10.1016/B978-0-44-459406-8.00021-4

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan and Jae Song. (2015). "What do data on millions of US workers reveal about life-cycle earnings risk?". Technical Report, National Bureau of Economic Research. https://doi.org/10.3386/w20913

Hahn, Jinyong, Guido Kuersteiner and Maurizio Mazzocco. (2020). "Estimation with aggregate shocks". *The Review of Economic Studies*, 87(3), pp. 1365-1398. https://doi.org/10.1093/restud/rdz016

Han, Lu. (2010). "The effects of price risk on housing demand: empirical evidence from US markets". *The Review of Financial Studies*, 23(11), pp. 3889-3928. https://doi.org/10.1093/rfs/hhq088

Heaton, John, and Deborah Lucas. (2000). "Portfolio choice and asset prices: The importance of entrepreneurial risk". *The Journal of Finance*, 55(3), pp. 1163-1198. https://doi.org/10.1111/0022-1082.00244

Heckman, James. (1974). "Shadow prices, market wages, and labor supply". *Econometrica*, 42(4), pp. 679-694. https://doi.org/10.2307/1913937

Heckman, James J. (1979). "Sample selection bias as a specification error". *Econometrica*, pp. 153-161. https://doi.org/10.2307/1912352

Hochguertel, Stefan. (2003). "Precautionary motives and portfolio decisions". *Journal of Applied Econometrics*, 18(1), pp. 61-77. https://doi.org/10.1002/jae.658

Honore, Bo, Shakeeb Khan and James L. Powell. (2002). "Quantile regression under random censoring". *Journal of Econometrics*, 109(1), pp. 67-105. https://doi.org/10.1016/S0304-4076(01)00142-7

Hu, Yingyao. (2015). "Microeconomic models with latent variables: applications of measurement error models in empirical industrial organization and labor economics". Technical Report, CEMMAP. https://doi.org/10.2139/ssrn.2555111

Hu, Yingyao. (2017). "The econometrics of unobservables: Applications of measurement error models in empirical industrial organization and labor economics". *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2017.06.002

Inkmann, Joachim. (2020). "Industry affiliation and the value of portfolio choice". Technical Report, University of Mellbourne. https://doi.org/10.2139/ssrn.3665862

Kaplan, Greg, and Giovanni L. Violante. (2010). "How much consumption insurance beyond self- insurance?". *American Economic Journal: Macroeconomics*, 2(4), pp. 53-87. https://doi.org/10.1257/mac.2.4.53

Khan, Shakeeb, and Elie Tamer. (2009). "Inference on endogenously censored regression models using conditional moment inequalities". *Journal of Econometrics*, 152(2), pp. 104-119. https://doi.org/10.1016/j.jeconom.2009.01.006

Knüpfer, Samuli, Elias Rantapuska and Matti Sarvimäki. (2016). "Formative experiences and portfolio choice: Evidence from the finnish great depression". *The Journal of Finance*. https://doi.org/10.1111/jofi.12469

Koenker, Roger. (2005). *Quantile regression*. Cambridge University Press. https://doi.org/10.1017/CBO9780511754098

Lise, Jeremy. (2013). "On-the-job search and precautionary savings". *Review of Economic Studies*, 80(3), pp. 1086-1113. https://doi.org/10.1093/restud/rds042

Luttmer, Erzo G. J. (1999). "What level of fixed costs can reconcile consumption and stock returns?". *Journal of Political Economy*, 107(5), pp. 969-997. https://doi.org/10.1086/250087

Malmendier, Ulrike, and Stefan Nagel. (2011). "Depression babies: Do macroeconomic experiences affect risk taking?". *The Quarterly Journal of Economics*, 126(1), pp. 373-416. https://doi.org/10.1093/qje/qjq004

Martin, Ian W. R. (2012). "Consumption-based asset pricing with higher cumulants". *Review of Economic Studies*, 80(2), pp. 745-773. https://doi.org/10.1093/restud/rds029

Newey, Whitney K., and James L. Powell. (2003). "Instrumental variable estimation of nonparametric models". *Econometrica*, 71(5), pp. 1565-1578. https://doi.org/10.1111/1468-0262.00459

Paiella, Monica. (2007). "The Forgone Gains of Incomplete Portfolios". *Review of Financial Studies*, 20(5), pp. 1623-1646. https://doi.org/10.1093/rfs/hhm022

Palia, Darius, Yaxuan Qi and Yangru Wu. (2014). "Heterogeneous background risks and portfolio choice: Evidence from micro-level data". *Journal of Money, Credit and Banking*, 46(8), pp. 1687-1720. https://doi.org/10.1111/jmcb.12163

Peijnenburg, Kim. (2018). "Life-cycle asset allocation with ambiguity aversion and learning". *Journal of Financial and Quantitative Analysis*, 53(5), pp. 1963-1994. https://doi.org/10.1017/S0022109017001144

Powell, James L. (1986). "Censored regression quantiles". *Journal of Econometrics*, 32(1), pp. 143-155. https://doi.org/10.1016/0304-4076(86)90016-3

Rooij, Maarten Van, Annamaria Lusardi and Rob Alessie. (2011). "Financial literacy and stock market participation". *Journal of Financial Economics*, 101(2), pp. 449-472. https://doi.org/10.1016/j.jfineco.2011.03.006

Rosen, Harvey S., and Stephen Wu. (2004). "Portfolio choice and health status". *Journal of Financial Economics*, 72(3), pp. 457-484. https://doi.org/10.1016/S0304-405X(03)00178-8

Schmidt, Lawrence D. W. (2015). "Climbing and falling off the ladder: Asset pricing implications of labor market event risk". https://dx.doi.org/10.2139/ssrn.2471342

Storesletten, Kjetil, Chris I. Telmer and Amir Yaron. (2004a). "Cyclical dynamics in idiosyncratic labor market risk". *Journal of Political Economy*, 112(3), pp. 695-717. https://doi.org/10.1086/383105

Storesletten, Kjetil, Christopher I. Telmer and Amir Yaron. (2004b). "Consumption and risk sharing over the life cycle". *Journal of Monetary Economics*, 51(3), pp. 609-633. https://doi.org/10.1016/j.jmoneco.2003.06.005

Vissing-Jørgensen, Annette. (2002). "Towards an explanation of household portfolio choice heterogeneity: Nonfinancial income and participation cost structures". Technical Report, National Bureau of Economic Research. https://doi.org/10.3386/w8884

Online Appendix for the paper "Household portfolio choices under (non)linear income risk: an empirical framework"

A Estimation of the canonical and nonlinear earnings processes

A.1 Linear earnings process estimation

The linear earnings process I specify is similar to those considered by Storesletten et al. (2004a) and Kaplan and Violante (2010), among others. In this earnings process, the transitory component ε_{it} is assumed to be independently and identically distributed (i.i.d.) Gaussian, with $N(0, \sigma_\varepsilon^2)$. The persistent component, meanwhile, is modelled as

$$\eta_{it} = \rho\eta_{it-1} + \nu_{it}$$

where ρ is a persistence parameter. The idiosyncratic part of the persistent component of income, ν_{it}, is also i.i.d. Gaussian, with $\nu_{it} \sim N(0, \sigma_\nu^2)$. I assume that households have different initial conditions; that is, $\eta_{i0} \sim i.i.d.N(0, \sigma_z^2)$. Moreover, I assume that the initial condition, the idiosyncratic component of the persistent shock, and the transitory shock are independent of each other.

The standard estimation strategy is minimum distance estimation. An alternative, which I implement here, is to estimate the parameters via pseudo maximum likelihood estimation. 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:

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

where **tr** is the trace of the resulting matrix, and $\hat{\Omega} = \sum \hat{u}_i'\hat{u}_i$. The parameter estimates are reported in the table below.

A.2 Specification and estimation of Arellano et al. (2017) earnings process

Persistent component. Denote the persistent component of the household head i at period t by η_{it} and age_{it} the age of the household head. Then, the conditional quantile of the persistent component as a function of the past persistent component and age is:

$$Q_t(\eta_{it-1}, \tau) = \sum_{k=0}^{K} a_K^P(\tau)\varphi_k(\eta_{it-1}, age_{it})$$

	Std. Error
Autoregressive parameter	0.8784
	(0.0987)
SD of transitory component	0.2333
	(0.0550)
SD of initial persistent component	0.3148
	(0.0666)
SD of idiosyncratic persistent component	0.1968
	(0.0365)

Note: These are estimates of the parameters of the linear earnings process. Standard errors are in parentheses, and are calculated using the asymptotic covariance matrix.

Initial condition of the persistent component. The conditional quantile function of the initial persistent component as a function of age is:

$$Q_{\eta_{i1}}(age_{i1}, \tau) = \sum_{k=0}^{K} a_K^I(\tau)\tilde{\varphi}_k(age_{i1})$$

Transitory component. The conditional quantile function of the transitory component as a function of age is:

$$Q_{\varepsilon}(age_{it}, \tau) = \sum_{k=0}^{K} a_K^T(\tau)\bar{\varphi}_k(age_{it})$$

The estimation of this earnings process follows the stochastic EM algorithm described in Arellano et al. (2017). I refer the reader to Arellano et al. (2017) for a full description of the estimation procedure, and the likelihood function of the earnings process.

A.3 Comparing the nonlinear and canonical earnings processes

Figure A1 presents the results of the estimation of the earnings process, and the estimation of a quantile autoregression of earnings. Specifically, the first three graphs are plots of the average derivative of the conditional quantile function y_{it} on y_{it-1}, with respect to the simulated earnings data. Meanwhile, the last graph is the average derivative of the conditional quantile function η_{it} on η_{it-1} using the nonlinear earnings process.

To check whether the results that I have obtained fit the model well, I compare the estimated persistence from the true data, which is depicted in Figure panel (a) of Figure A1, with the estimated persistence that comes from simulated data according to the nonlinear and the linear earnings models, which are in panels (b) and (c), respectively. What I find is that the nonlinear earnings process is able to fit the data well in terms of persistence. Panel (d) describes the results of the persistence in the persistent component η_{it} in the nonlinear earnings model. As the results indicate, there seems to be heterogeneity in persistence of income shocks.

Figure A2 shows estimates of conditional skewness, which were calculated using quantile-based skewness measures. The results in panel (b) indicate evidence of conditional asymmetry, which goes in the same direction as in Arellano et al. (2017). There is less evidence of it though in the simulated data and in the earnings data of the PSID.

B Two-period model

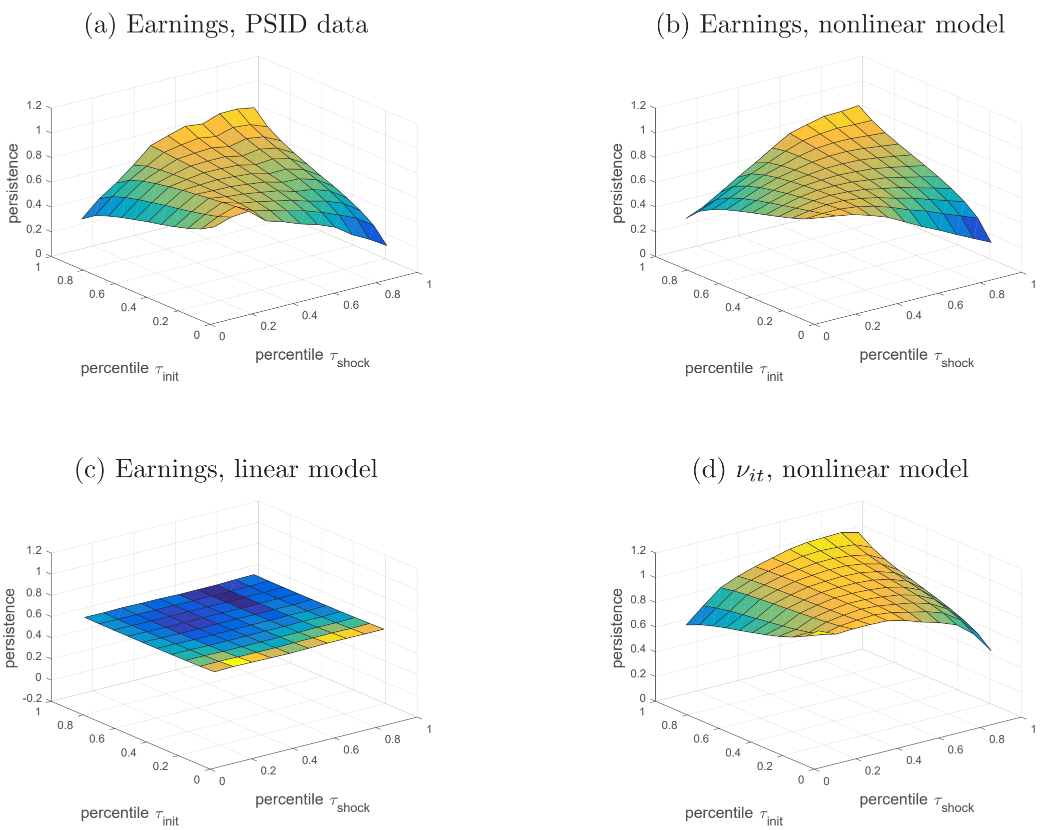
In this appendix, I show how optimal portfolio choice is affected by higher-order moments of income in a simple two-period set-up. The model I present is heavily based on the one presented in Campbell and Viceira (2002).

Consider a model where a household with CRRA utility and wealth W_t that makes a portfolio decision at time t. It consumes the liquidation value of its portfolio at time $t+1$ plus household labor income Y_{t+1} one period later. Labor income is stochastic and follows a general distribution $H(Y)$. The household cannot borrow against future labor income, thereby making it non-tradeable. It has access to two assets for investment: a riskless asset that has a certain return r_f and a risky asset with a constant expected log excess return $\mathbb{E}_t(r_{t+1} - r_f) \equiv \mu$. The unexpected log return on the risky asset, denoted by u_{t+1}, is conditionally Normal, with mean zero and variance σ_r^2. To invest in the stock market, the household must pay a participation cost q.

To make its optimal decision, the household considers two subproblems. The first corresponds to the situation in which it does not participate in the stock market. In this case, it solves for optimal consumption, $C_{np,t+1}$, via the following maximization problem:

$$V_{np} = \max_{C_{np,t+1}} \mathbb{E}[U(C)]$$
$$\text{s.t. } C_{np,t+1} = W_t(1 + R_f) + Y_{t+1}$$

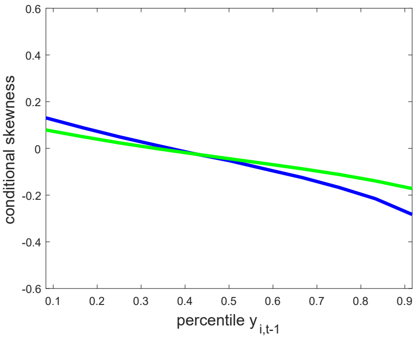
Figure A1: Nonlinear persistence



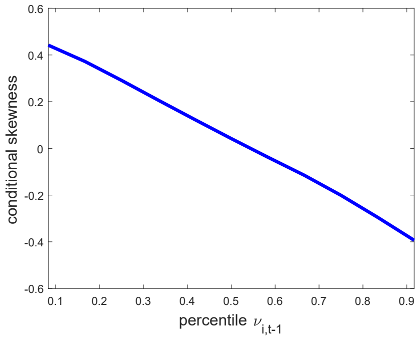
Note: Panel (a), (b), and (c) show estimates of the conditional quantile function of y_{it} given y_{it-1} with respect to y_{it-1}, evaluated at τ_{shock} and at a value of y_{it-1} that corresponds to the τ_{init} percentile of the distribution of y_{it-1}. Panel (a) is based on the PSID data in my subsample, panel (b) comes from the simulated data based on the nonlinear earnings process, and panel (c) comes from the simulated data based on the linear earnings process. Finally, panel (d) is the average derivative of the the conditonal quantile function of η_{it} on η_{it-1} with respect to η_{it-1}, based on estimates from the nonlinear earnings model.

Figure A2: Conditional skewness

(a) Log-earnings residuals y_{it}



(b) Persistent component ν_{it}



Note: These are estimates of the quantile-based skewness of residualized log household income (left panel) from the data (blue) and from simulated data coming from the nonlinear earnings model of Arellano et al. (2017) (green). The right panel is an estimate of the conditional skewness of the persistent income component, η_{it}.

In the second subproblem, the household invests part or all of its wealth in stocks. It then solves for the optimal portfolio share, α_t, via the following maximization problem:

$$V_p = \max_{\alpha_t} \mathbb{E}[U(C)]$$
$$\text{s.t. } C_{p,t+1} = W_{c,t}[\alpha_t R_{t+1} + (1-\alpha_t)R_f] + Y_{t+1}$$

in which I have defined $W_{c,t} = W_t - q$. As the household can neither borrow nor short-sell, the optimal portfolio share is constrained to be in between zero and one. Moreover, it neither knows the realization of future stock market returns nor future labor income when it makes the portfolio choice decision.

Although a closed form solution generally does not exist, an approximate formula for optimal portfolio shares in the participation subproblem can be obtained as a function of the certainty equivalent of future labor earnings $\widetilde{H}$ and wealth net of participation costs, $W_{c,t}$, under idiosyncratic labor income risk[28]:

$$\alpha_t = \left(1 + \frac{\widetilde{H}}{W_{c,t}}\right)\left[\frac{\mathbb{E}_t(r_{t+1} - r_f) + \frac{1}{2}\sigma_r^2}{\gamma \sigma_r^2}\right] \tag{B1}$$

To solve the optimization problem, the household compares the indirect utilities calculated from the two subproblems. The household will clearly participate if the expected utility from equity investment is at least as high as that of non-investment, which is

[28]A more formal treatment of a static portfolio choice problem when income risk is not lognormal requires working with higher-order cumulants as in Martin (2012), but is left for further research.

effectively summarized by this inequality:

$$\mathbb{E}_t\left(\delta\frac{C_{p,t+1}^{1-\gamma}}{1-\gamma}\right) \geq \mathbb{E}_t\left(\delta\frac{C_{np,t+1}^{1-\gamma}}{1-\gamma}\right)$$

where $C_{i,t}$, $i = p, np$ denotes the consumption if the household bought stocks or not, respectively. Taking logs:

$$\mathbb{E}_t\left(\delta\frac{C_{i,t+1}^{1-\gamma}}{1-\gamma}\right) = \mathbb{E}_t\left[\exp\left\{\log\left(\delta\frac{C_{i,t+1}^{1-\gamma}}{1-\gamma}\right)\right\}\right] \approx \mathbb{E}_t(\exp\{(1-\gamma)c_{i,t+1}\})$$

Because the risk aversion parameter is a constant, I focus on $\mathbb{E}_t(\exp\{c_{i,t+1}\})$. Taking a first order Taylor expansion around zero, I obtain:

$$\mathbb{E}_t(\exp\{c_{i,t+1}\}) \approx 1 + \mathbb{E}_t(c_{i,t+1}),$$

which, when substituted to the inequality yields the condition

$$\mathbb{E}_t(c_{p,t+1}) \geq \mathbb{E}_t(c_{np,t+1}).$$

that effectively states that to invest in the stock market, households must have at least the same amount of consumption in both

Thus, if it decides to participate in the stock market, the optimal portfolio rule is characterized by equation (B1). Otherwise, the optimal portfolio rule is characterized by $\alpha_t = 0$.

Comparative statics. Equation (B1) allows me to study the effect of increases in wealth and labor income under idiosyncratic labor income risk, respectively.[29] First, keeping labor income constant, an exogenous increase in wealth reduces the portfolio share, as total household wealth becomes a more important source to draw consumption from than labor income. Hence, the household will not invest in stocks, and might prefer to save in riskless bonds, or to spend part of the wealth gain on goods. This result holds regardless of whether labor income is lognormal or not.

Second, an increase in labor income has an ambiguous effect on household portfolio shares when I relax lognormality, keeping wealth fixed. This is because now, labor income

[29] The comparative statics results I discuss here apply to a household who is currently a stock investor. For the marginal investor who is indifferent between entering and exiting the stock market, he will continue to participate in the stock market if the consumption gained from investment is at least as high as that of non-investment, regardless of whether the increases are with respect to labor income or wealth. The results, again, depend on the higher-order moments of income.

will be affected by higher-order moments. For example, an increase in labor income might lead the household to invest less in stocks if the distribution of its earnings is negatively skewed. In this case, even if the household experiences an increase in labor income, the possibility of highly negative income realizations in the future might lead the household to become more conservative in its investments. In contrast, under lognormality, an increase in labor income will lead the household to invest in stocks if the expected labor income is greater than the risk the household faces, which is captured by its variance.

C Data and descriptive statistics

C.1 Sample selection criteria in detail

Table C1 shows the detailed sample selection criteria that was implemented for the main sample in the empirical study that largely follows the criteria of Blundell et al. (2016). In the sample selection, I first construct a subsample of all households who I can follow for the entire six waves of the PSID. I then apply each of the criteria in Blundell et al. (2016).

Table C1: Sample selection criteria, PSID

Total households eligible from the PSID interviews		4304
Less:		
Households who were not continuously married	2064	
Households with missing information on state	20	
Households with missing information on age	0	
Households with missing information on education	292	
Non-SEO households	273	
Households with missing information on race	7	
Households who are not within the age range	561	
Households with higher than $20,000,000 assets	1	
Households with missing information on assets	192	
Households with wages less than half the state minimum wage	187	
Households with missing labor income	26	
Households with jumps on their labor income	20	
		661

Note: Data from 1999 to 2009 PSID waves. This table presents the sample selection criteria I operationalized for the paper. The sample selection criteria mostly follows that of Blundell et al. (2016).

Table C2 compares the households in my baseline sample with a sample of all married male household heads (regardless of participation status) and with a sample of all male household heads who were recorded as married at least once in the 1998 to 2008 period (again, independently of participation status). The table shows very small differences in the observables across households. Household earnings are only slightly smaller for the more comprehensive households than for the baseline sample, and total household wealth is smaller for the more comprehensive samples than for the baseline sample. The proportion of stock market participants is also roughly similar across these households, which suggests that concerns on whether I am removing households whose heads are facing extremely large income shocks that result in long unemployment spells seem to be disspelled.

Table C2: Baseline sample comparisons

	Baseline sample		With nonworking males		All ever married	
	Mean	Median	Mean	Median	Mean	Median
			All households			
Household income	82,699.47	54,000.00	81,744.46	54,000.00	81,071.35	54,000.00
Total assets	371,189.60	164,043.40	367,263.70	163,715.20	364,863.50	160,581.20
Liquid wealth	95,529.77	17,083.33	97,067.34	17,083.33	95,278.02	17,083.33
Risky wealth	81,353.17	4,741.95	81,998.92	4,365.16	80,484.74	4,000.00
Stocks	51,547.79	-	52,983.55	-	51,758.05	-
Share of stocks in total wealth	0.077	-	0.077	-	0.076	-
Share of risky assets in total wealth	0.148	0.030	0.148	0.029	0.147	0.029
Ownership of stocks	0.410		0.408		0.408	
Ownership of risky assets	0.591		0.587		0.586	
			Risky market participants			
Household income	98,524.98	65,000.00	97,221.93	65,000.00	96,468.87	64,250.00
Total assets	512,872.20	269,187.60	510,379.10	269,682.60	507,263.20	264,412.70
Liquid wealth	145,598.80	41,201.23	148,873.50	42,548.95	146,207.80	41,314.15
Risky wealth	137,676.70	43,994.58	139,767.20	44,223.90	137,445.20	43,073.93
Stocks	87,236.06	9,597.64	90,310.51	10,000.00	88,388.12	9,571.98
Share of stocks in total wealth	0.130	0.046	0.131	0.047	0.130	0.047
Share of risky assets in total wealth	0.251	0.192	0.252	0.192	0.251	0.190

Note: Data from 1999 to 2009 PSID waves. This table presents summary statistics for different subsamples. The first two columns correspond to the baseline sample that I use in my empirical study. The next two columns correspond to all male married households heads (independently of work status). The last two columns correspond to all male households heads who have been married at least once, again independently of work status. The first panel corresponds to all households in a given sample; the second panel corresponds to the risky market participants in a given sample.

C.2 Calculating human wealth

I follow the definition of Calvet and Sodini (2014) and Fagereng et al. (2017a) in the calculation of human wealth, which is crucial in the calculation of lifetime wealth, the exclusion restriction I follow in the paper. To be specific, the formula is the following:

$$HW_{i,t} = Y_{i,t} + \sum_{\tau=1}^{T-t} \pi_{\tau+1|\tau} \frac{\mathbb{E}_t(Y_{i,t+\tau})}{(1+r)^{\tau}} \tag{C2}$$

in which $HW_{i,t}$ denotes human wealth, $Y_{i,t+\tau}$ is the labor income of the household at age $t+\tau$ and π_τ is the survival probability of the household head being alove at age $t+\tau$

given that he was alive at age t. I set the discount rate to the one I use in the calibrations, approximately 2 percent. Lifetime wealth, in this case, is the sum of human wealth and accumulated assets during that period.

D Nonparametric identification

This section covers the nonparametric identification of the semi-structural model. I state the formal assumptions, and then outline the proof, which relies on an induction argument.

Consider a given household i observed for T periods. Let us denote by x_i^t as the following vector $x_i^t = (x_{i1}, x_{i2}, \ldots, x_{iT})'$, and suppose that we observe the sequence of household wealth w_i^t, the resulting portfolio choice α_i^t, the observed participation choice d_i^t, and household earnings y_i^t. Let us call by η_i^t the latent earnings component, and by $\alpha_i^{*,t}$ the latent risky asset share. Our goal is to identify the joint distribution of η's, assets, earnings, portfolio choices, and participation decisions. To do so, we will need the following assumptions on the semi-structural model:

Assumption 1 *For all $t \geq 1$:*

a. The unobserved errors $(\mathbf{v}_{i1}, \ldots, \mathbf{v}_{iT}, \eta_{i1}, \ldots, \eta_{iT}, \varepsilon_{i1}, \ldots, \varepsilon_{iT}, \zeta_{i2}, \ldots, \zeta_{iT})$ are mutually independently distributed.

b. The bivariate distribution of $\mathbf{v}_{it} = (\chi_{it}, \xi_{it})'$ given $w_i^t, \alpha_i^t, d_i^t, y_i^t$ and η_i^t are absolutely continuous with respect to the Lebesgue measure, with standard uniform marginals and rectangular support. I denote the c.d.f. as $C_x(\chi, \xi)$.

c. The conditional c.d.f. $F_{\alpha^*|Z}(\alpha|Z)$ is strictly increasing. Moreover, $C_x(\chi, \xi)$ is strictly increasing in χ.

d. $\Pr(d_{it} = 1|z_{it}) > 0$ *for all t with probability one.*

The first assumption implies the following: first, that current and future earnings shocks are independent of current and past wealth; second, a Markovian assumption on wealth dynamics; and third, that the unobserved errors of the portfolio and participation rules are independent over time, independent of earnings components, and independent of current and past assets. The second assumption states that the dependence between (χ_{it}, ξ_{it}) is the source of sample selection bias. In the context of the economic model, the dependence is the link between the participation subproblem, the result of which is the latent risky asset shares α_{it}^*, and the economic problem at hand, which results in the participation rule denoted by d_{it} and the observed outcome α_{it}. The third assumption restricts the analysis to continuous outcomes, and the fourth is a standard assumption in sample selection models. The proof proceeds sequentially, starting with the first period.

First period wealth. Using f as a generic notation for a density function:

$$f(w_1|y) = \int f(w_1|\eta_1) f(\eta_1|y) d\eta_1 \tag{D3}$$

where, by Assumption 1a, $f(w_1|\eta_1, y)$ and $f(w_1|\eta_1)$ coincide. Provided that the distribution of $(\eta_{i1}|y_i)$ is complete (given that this is identified from the earnings process, see Arellano et al. (2017)), the density $f(w_1|y)$ is identified.

First period participation. Next, I turn into the participation rule. Notice that via the definition of the participation rule, and by assumption 1a, I have:

$$f(d_1|w_1, q_1, y) = \int f(d_1|w_1, \eta_1, q_1, y_1) f(\eta_1|w_1, y) d\eta_1 \tag{D4}$$

where under completeness of $(\eta_1|w_1, y_i)$ in y_i, the density $f(d_1|w_1, q_1, y_1)$ is identified.

First period portfolio rule. Identification of the joint density of risky shares and the participation indicator is not straightforward, however. To illustrate, write the density $f(\alpha_1^*, d_1|w_1, y)$ as:

$$f(\alpha_1^*, d_1|w_1, q_1, y) = \int [\Pr(\alpha_1^*, d_1 = 1|q_1, w_1, \eta_1, y)]^{d_1} [\Pr(d_1 = 0|q_1, w_1, \eta_1, y)]^{(1-d_1)} f(\eta_1|w_1, y) d\eta_1.$$

$$\begin{aligned} f(\alpha_1^*, d_1|w_1, q_1, y) = \int & [f(\alpha_1^*|d_1 = 1, w_1, \eta_1, y)]^{d_1} \times \\ & [\Pr(d_1 = 1|w_1, q_1, \eta_1, y)]^{d_1} [\Pr(d_1 = 0|w_1, q_1, \eta_1, y)]^{(1-d_1)} f(\eta_1|w_1, y) d\eta_1, \end{aligned}$$

which, after, using the definition of the participation rule, yields:

$$f(\alpha_1^*, d_1|w_1, y) = \int [f(\alpha_1^*|d_1 = 1, w_1, \eta_1, y)]^{d_1} f(d_1|w_1, \eta_1, y) f(\eta_1|w_1, y) d\eta_1 \tag{D5}$$

If α_1^* were not latent, I can proceed with the same identification arguments as those for wealth and the participation rule.[30] To prove that this density is identified under the presence of sample selection, note that, conditional on participation, for all $\tau \in (0,1)$ and $t \geq 1$:

$$\Pr(\alpha_{it}^* \leq q(\tau, x))|d_{it} = 1, \mathbf{Z}_{it} = \mathbf{z}_{it}) = G_x(\tau, p(\mathbf{z}_{it})) \tag{D6}$$

where I have used Assumption 1b and Bayes' rule. The mapping $G_x(\tau, p(\mathbf{z}_{it}))$ provides the link between the latent and the observed distribution of risky asset shares in my model. Notably, what equation (D6) implies is that if $G_x(\cdot,\cdot)$ is known, then I can recover $q(\tau, x)$, as a quantile of the observed portfolio shares, by suitably shifting percentile ranks. Subsequently, I would be able to recover the portfolio rule. More formally, under Assumption 1, and given that the mapping G_x is known, Proposition 1 of Arellano and Bonhomme (2017a) will hold. Given that this is satisifed, $F(\alpha_1^*|d_1 = 1, w_1, \eta_1, y)$ and subsequently, $f(\alpha_1^*|d_1 = 1, w_1, \eta_1, y)$ are identified nonparametrically.

Second period wealth. Notice that:

$$f(w_2|\alpha_1, d_1, w_1, y, q) = \int f(w_2|\alpha_1, d_1, w_1, \eta_1, y) f(\eta_1|\alpha_1, d_1, w_1, y) d\eta_1 \tag{D7}$$

Provided that the distribution $(\eta_1|\alpha_1, d_1, w_1, y)$ is complete in $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})$, I can identify the density $f(w_2|\alpha_1, d_1, w_1, \eta_1, y)$.

Via Bayes' rule and Assumption 1a,

$$f(\eta_2|w_2, w_1, \alpha_1, d_1, y, b) = \int \frac{f(\eta_1, \eta_2|w_1, w_2, \alpha_1, d_1, y_1) f(y|y_1, \eta_1, \eta_2)}{f(y|w_2, w_1, \alpha_1, y_1)} d\eta_2 \tag{D8}$$

As $f(\eta_1, \eta_2|w_1, w_2, \alpha_1, d_1, y_1) = f(\eta_1|w_1, w_2, \alpha_1, d_1, y_1) f(\eta_2|\eta_1)$ is identified (given that $f(\eta_2|\eta_1)$ is identified from the earnings process, and $f(\eta_1|w_1, w_2, \alpha_1, d_1, y_1)$ is identified from above), it follows that $f(\eta_2|w_2, w_1, \alpha_1, d_1, y)$ is identified.

Subsequent periods. To prove the participation rule, notice that:

$$f(d_2|w_2, w_1, \alpha_1, y) = \int f(d_2|w_2, \eta_2, y, q) f(\eta_2|w_2, w_1, \alpha_1, d_1, y) d\eta_2. \tag{D9}$$

[30] Alternatively, if the economic problem is such that there is no selection into stock market participation, the same arguments will still apply. This is because, in the absence of the participation rule, the density of risky asset shares is simply:

$$f(\alpha_1|w_1, y) = \int [f(\alpha_1|w_1, \eta_1, y)]^{\mathbf{1}(\alpha_1 > 0)} [F(0|w_1, \eta_1, y)]^{\mathbf{1}(\alpha_1 = 0)} f(\eta_1|w_1, y) d\eta_1$$

It can then be shown that the density is nonparametrically identified under completeness of $(\eta_{i1}|w_{i1}, y_i)$ on $(y_{i1}, y_{i2}, \ldots, y_{iT})$, following similar arguments.

Given that $(\eta_{i2}|w_{i2}, w_{i1}, \alpha_{i1}, d_{i1}, y)$ (which is identified from the previous paragraph) is complete in $(\alpha_{i1}, d_{i1}, w_{i1}, y_{i1}, y_{i3}, \ldots, y_{iT})$, the density $f(d_2|w_2, \eta_2, y, q)$ is nonparametrically identified.

In the case of the portfolio rule,

$$f(\alpha_2^*, d_2|w_2, w_1, y_1, \alpha_1) = \int [f(\alpha_2^*|d_2 = 1, w_2, \eta_2, y)]^{d_2} f(\eta_2|d_2, b_2, w_2, w_1, \alpha_1, d_1, y)d\nu_2. \tag{D10}$$

As the distribution of $(\eta_{i2}|w_{i2}, w_{i1}, \alpha_{i1}, d_{i1}, y)$ is complete in $(d_{i2}, \alpha_{i1}, d_{i1}, w_{i1}, y_{i1}, y_{i3}, \ldots, y_{iT})$ (which is identified from the participation rule) and under Proposition 1 of Arellano and Bonhomme (2017a), the distribution $F(\alpha_2^*|d_2 = 1, w_2, \eta_2, y)$, and subsequently, the density $f(\alpha_2^*|d_2 = 1, w_2, \eta_2, y)$ is nonparametrically identified.

Finally, by induction, and using assumptions 1 and Proposition 1 of Arellano and Bonhomme (2017a) from the third period onward, the joint density of η's, assets, earnings, portfolio choices, and participation decisions are nonparametrically identified. This is provided that, for all $t \geq 1$, the distributions of $(\eta_{it}|\alpha_i^t, w_i^t, d_i^t, y_i)$ are complete in $(\alpha_i^{t-1}, w_i^{t-1}, d_i^{t-1}, y_i^{t-1}, y_{it+1}, y_{iT})$.

Intuitively, the identification argument comes from the link to nonparametric instrumental variables problems. To see this, note that I can rewrite equation (D3) as

$$f(w_1|y) = \mathbb{E}\left[f(w_1|\eta_{i1})|y_i = y\right] \tag{D11}$$

where I am taking the expectation of the distribution of η_{i1} conditional on y_i for a given fixed value of w_{i1}. This is analogous to a nonparametric IV problem where the endogenous regressor is η_{i1} and y_i are the excluded instruments. Likewise, I can rewrite equations (D4) and (D5) as the following functional equations:

$$f(d_1|w_1, y) = \mathbb{E}\left[f(d_1|\eta_{i1}, w_{i1}, y_{i1})|w_{i1} = w, y_i = y\right] \tag{D12}$$

$$f(\alpha_1^*, d_1|w_1, y) = \mathbb{E}\left[\{f(\alpha_{i1}^*|\eta_{i1}, w_{i1}, d_{i1} = 1, y_{i1})\}^{\mathbf{1}\{d_{i1}=1\}}|d_{i1} = d, q_{i1} = q, w_{i1} = w, y_i = y\right] \tag{D13}$$

In these particular cases, conditional on (w_{i1}, y_{i1}), $(y_{i2}, \ldots, y_{iT})$ are the "excluded instruments" for η_{i1} with respect to the functional equation that corresponds to the participation rule. With respect to the portfolio rule, however, I require not only these "excluded instruments", but also a participation cost shifter that does not affect the subsequent portfolio choice.

The identification arguments also rely on completeness conditions, which relate to the relevance of the excluded instruments. The notion of completeness that I refer to here relates to the concept of operator injectivity. More formally, a linear operator $\mathcal{L}$ is injective if the only solution $h \in \mathcal{H}_1$ to the equation $\mathcal{L}h = 0$ is $h = 0$. In this paper, $\mathcal{L}$ is the conditional expectation operator. For example, in equation (D11), completeness implies that the only solution to the equation $[\mathcal{L}h](y_i) = \mathbb{E}\left[f(w_1|\eta_{i1})|y_i = y\right]$ is $h = 0$.

E Estimation strategy

E.1 Model estimation: details

E.1.1 Likelihood function

The likelihood function is:

$$f(\alpha_i^T, \eta_i^T, \varepsilon_i^T, w_i^T, \mathbf{Z_i}, d_i^T; \bar{\mu}), = \prod_{t=1}^{T} \left[f(\alpha_{it}^*|\eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{x}_{it}) p(d_{it} = 1|\nu_{it}, \varepsilon_{it}, w_{it}, \mathbf{z}_{it}) \nabla C(u, v; c)\right]^{d_{it}}$$
$$\times \prod_{t=1}^{T} \left[p(d_{it} = 0|\eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{z}_{it})\right]^{1-d_{it}} \prod_{t=2}^{T} f(w_{it}|w_{it-1}, \eta_{it-1}, y_{it-1}, \alpha_{it-1}, \mathbf{x}_{it})$$
$$\times f(w_{i1}|\eta_{i1}, \mathbf{x}_{i1}) \prod_{t=1}^{T} f(y_{it}|\eta_{it}) \prod_{t=2}^{T} f(\eta_{it}|\eta_{it-1}) f(\eta_{i1}) \quad \text{(E14)}$$

where $u = F(\alpha_{it}^*|\eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{x}_{it})$, $v = p(d_{it} = 1|\eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{z}_{it})$ and $\nabla C(\cdot, \cdot; \cdot)$ is the first derivative of the conditional copula with respect to the first argument.

I can simplify the likelihood function further by noting that I can rewrite the conditional copula as follows:

$$C(u, v; c) = \frac{G(F(\alpha_{it}^*|\eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{x}_{it}), p(d_{it} = 1|\eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{z}_{it}); \rho_c)}{p(d_{it} = 1|\eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{z}_{it})} \quad \text{(E15)}$$

where $G(\cdot, \cdot; \rho_c)$ is the Gaussian copula. It follows that the first derivative of this function with respect to the first argument is:

$$\nabla G(u, v; c) = \Phi\left(\frac{\Phi^{-1}(v) - \rho_c \Phi^{-1}(u)}{\sqrt{1 - \rho_c^2}}\right) \quad \text{(E16)}$$

Substituting the resulting expression for $\nabla C(u,v;c)$ to the expression above, the likelihood function simplifies to:

$$f(\alpha_i^T,\eta_i^T,\varepsilon_i^T,w_i^T,\mathbf{Z}_i^T,d_i^T;\bar{\mu}),=\prod_{t=1}^{T}\left[f(\alpha_{it}^*|\eta_{it},\varepsilon_{it},w_{it},\mathbf{x}_{it})\Phi\left(\frac{\Phi^{-1}(v)-\rho_c\Phi^{-1}(u)}{\sqrt{1-\rho_c^2}}\right)\right]^{d_{it}}$$
$$\times\prod_{t=1}^{T}[p(d_{it}=0|\eta_{it},\varepsilon_{it},w_{it},\mathbf{z}_{it})]^{1-d_{it}}\prod_{t=2}^{T}f(w_{it}|w_{it-1},\eta_{it-1},y_{it-1},\alpha_{it-1},\mathbf{x}_{it})f(w_{i1}|\eta_{i1},\mathbf{x}_{i1})$$
$$\times\prod_{t=1}^{T}f(y_{it}|\eta_{it})\prod_{t=2}^{T}f(\eta_{it}|\eta_{it-1})f(\eta_{i1}) \quad \text{(E17)}$$

As the model is fully specified, I can write the likelihood function in closed form. To focus my discussion, I illustrate the specification for the approximating density function $f(\alpha_{it}^*|\eta_{it},\varepsilon_{it},w_{it},\mathbf{x}_{it})$ and cumulative distribution function $F(\alpha_{it}^*|\eta_{it},\varepsilon_{it},w_{it},\mathbf{x}_{it})$. Notice that I can write the approximating outcome density as:

$$f(\alpha_{it}^*|\eta_{it},\varepsilon_{it},w_{it},\mathbf{x}_{it})=\left[\sum_{l=1}^{L-1}\frac{\tau_{l+1}-\tau_l}{\sum_{k=0}^{K}\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})(b_{kl+1}^a-b_{kl}^a)}\right.$$
$$\left.\times\mathbf{1}\left\{\sum_{k=0}^{K}b_{kl}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})<\Lambda^{-1}(\alpha_{it}^*)\leq\sum_{k=0}^{K}b_{kl+1}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})\right\}\right]$$
$$+\lambda_-^a\tau_1 e^{\lambda_-^a(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K}b_{k1}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it}))}\mathbf{1}\left\{\Lambda^{-1}(\alpha_{it}^*)<\sum_{k=0}^{K}b_{k1}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})\right\}$$
$$+\lambda_+^a(1-\tau_L)e^{-\lambda_+^a(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K}b_{kL}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it}))}\mathbf{1}\left\{\Lambda^{-1}(\alpha_{it}^*)\geq\sum_{k=0}^{K}b_{kL}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})\right\}$$

where I proceed with an exponential modelling of the tails.

The approximating conditional distribution functions are:

$$F(\alpha_{it}^*|\eta_{it},\varepsilon_{it},w_{it},\mathbf{x}_{it})=\left[\sum_{l=1}^{L-1}\tau_l+(\tau_{l+1}-\tau_l)\frac{\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K}b_{kl}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})}{\sum_{k=0}^{K}\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})(b_{kl+1}^a-b_{kl}^a)}\right.$$
$$\left.\times\mathbf{1}\left\{\sum_{k=0}^{K}b_{kl}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})<\Lambda^{-1}(\alpha_{it}^*)\leq\sum_{k=0}^{K}b_{kl+1}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})\right\}\right]$$
$$+\tau_1\left(e^{\lambda_-^a(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K}b_{k1}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it}))}\right)\mathbf{1}\left\{\Lambda^{-1}(\alpha_{it}^*)<\sum_{k=0}^{K}b_{k1}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})\right\}$$
$$+\left[\tau_L+(1-\tau_L)\left(1-e^{-\lambda_+^a(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K}b_{kL}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it}))}\right)\right]$$
$$\mathbf{1}\left\{\Lambda^{-1}(\alpha_{it}^*)\times\geq\sum_{k=0}^{K}b_{kL}^a\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it}).\right\}$$

E.1.2 Estimation algorithm: details

Start with $\theta^{(0)}$. Then, iterate on $s = 0, 1, 2, \ldots$ the following two steps:

Stochastic E-step: Draw M values $\eta_i^{(m)} = \left(\eta_{i1}^{(m)}, \ldots, \eta_{iT}^{(m)}\right)$ from

$$f(\alpha_i^T, \eta_i^T, \varepsilon_i^T, w_i^T, \mathbf{Z}_i^T, d_i^T; \widehat{\theta}^{(s)}), = \prod_{t=1}^{T} \left[f(\alpha_{it}^* | \eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{x}_{it}; \widehat{\theta}^{(s)}) \Phi \left(\frac{\Phi^{-1}(v) - \rho_c \Phi^{-1}(u)}{\sqrt{1-\rho_c^2}}; \widehat{\theta}^{(s)} \right) \right]^{d_{it}}$$

$$\times \prod_{t=1}^{T} \left[p(d_{it} = 0 | \eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{z}_{it}; \widehat{\theta}^{(s)}) \right]^{1-d_{it}} \prod_{t=2}^{T} f(w_{it} | w_{it-1}, \eta_{it-1}, y_{it-1}, \alpha_{it-1}, \mathbf{x}_{it}; \widehat{\theta}^{(s)}) f(w_{i1} | \eta_{i1}, \mathbf{x}_{i1}; \widehat{\theta}^{(s)})$$

$$\times \prod_{t=1}^{T} f(y_{it} | \eta_{it}; \hat{\theta}) \prod_{t=2}^{T} f(\eta_{it} | \eta_{it-1}; \hat{\theta}) f(\eta_{i1}; \hat{\theta}) \quad \text{(E18)}$$

M-step: Compute, for all $l = 1, \ldots, L$:

$$(\widehat{b}_0^{P,(s+1)}, \ldots, \widehat{b}_K^{P,(s+1)}) = \underset{(b_0^{P,(s+1)}, \ldots, b_K^{P,(s+1)})}{\arg\max} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{m=1}^{M} d_{it} \log \Lambda \left(\sum_{k=0}^{K} b_k^p \varphi_k(\eta_{it}^{(m)}, \varepsilon_{it}, w_{it}, age_{it}) \right)$$
$$+ (1 - d_{it}) \log \left(1 - \Lambda \left(\sum_{k=0}^{K} b_k^p \varphi_k(\eta_{it}^{(m)}, \varepsilon_{it}, w_{it}, age_{it}) \right) \right) \quad \text{(E19)}$$

$$\widehat{\rho}_c^{(s+1)} = \underset{\rho_c}{\arg\min} \left\| \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} d_{it} \Upsilon(\tau_l, \mathbf{x}_{it}) \left[\mathbf{1} \left\{ \Lambda^{-1}(\alpha_{it}^*) \leq \sum_{k=0}^{K} b_{kl}^a(c) \varphi_k(\cdot) \right\} - G(\tau_l, p(\mathbf{x}_{it}); \rho_c) \right] \right\| \quad \text{(E20)}$$

$$(\widehat{b}_{0l}^{a,(s+1)}, \ldots, \widehat{b}_{Kl}^{a,(s+1)}) = \underset{(b_{0l}^{a,(s+1)}, \ldots, b_{Kl}^{a,(s+1)})}{\arg\min} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{m=1}^{M} d_{it} \left[G(\tau_l, p(\mathbf{x}_{it}); \rho_c) \left(\Lambda^{-1}(\alpha_{it}^*) - \sum_{k=0}^{K} b_{kl}^a \varphi_k(\eta_{it}^{(m)}, \varepsilon_{it}, w_{it}, age_{it}) \right)^{-} \right.$$
$$\left. + (1 - G(\tau_l, p(\mathbf{x}_{it}); \rho_c)) \left(\Lambda^{-1}(\alpha_{it}^*) - \sum_{k=0}^{K} b_{kl}^a \varphi_k(\eta_{it}^{(m)}, \varepsilon_{it}, w_{it}, age_{it}) \right)^{-} \right] \quad \text{(E21)}$$

$$(\widehat{b}_0^{m,(s+1)}, \ldots, \widehat{b}_K^{m,(s+1)}) = \underset{(b_0^{m,(s+1)}, \ldots, b_K^{m,(s+1)})}{\arg\min} \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{m=1}^{M} \left(w_{it} - \sum_{k=1}^{K} b_k^m \widetilde{\varphi}_k(\eta_{it-1}^{(m)}, \varepsilon_{it-1}, w_{it-1}, \alpha_{it-1}, age_{it}) \right)^2 \quad \text{(E22)}$$

$$(\widehat{b}_{0l}^{w,(s+1)}, \ldots, \widehat{b}_{Kl}^{w,(s+1)}) = \underset{(b_{0l}^{w,(s+1)}, \ldots, b_{Kl}^{w,(s+1)})}{\arg\min} \sum_{i=1}^{N} \sum_{m=1}^{M} \tau_l \left(w_{i1} - \sum_{k=0}^{K} b_{kl}^w \widetilde{\varphi}_k(\eta_{i1}^{(m)}, age_{i1}) \right)^{+}$$
$$+ (1 - \tau_l) \left(w_{i1} - \sum_{k=0}^{K} b_{kl}^w \widetilde{\varphi}_k(\eta_{i1}^{(m)}, age_{i1}) \right)^{-} \quad \text{(E23)}$$

For the tail parameters, I calculate the following:

$$\lambda_-^{a,(s+1)} = -\frac{\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{m=1}^{M}\mathbf{1}\{\Lambda^{-1}(\alpha_{it}^*) \leq \sum_{k=0}^{K}\widehat{b}_{k1}^{a}\varphi_k(\cdot)\}}{\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{m=1}^{M}\left(\Lambda^{-1}(\alpha_{it}^*) - \sum_{k=0}^{K}\widehat{b}_{k1}^{a}\varphi_k(\cdot)\right)\mathbf{1}\{\Lambda^{-1}(\alpha_{it}^*) \leq \sum_{k=0}^{K}\widehat{b}_{k1}^{a}\varphi_k(\cdot)\}}, \tag{E24}$$

where $\varphi_k(\cdot) = \varphi_k(\eta_{it}^{(m)}, \varepsilon_{it}, w_{it}, age_{it})$, with similar updating rules for the other tail parameters.

E.2 Buchinsky and Hahn (1998) censored quantile regression estimation

E.2.1 Nonlinear reduced form and model specification

The equivalent nonlinear reduced form that corresponds to the Buchinsky and Hahn (1998) model is the following:

$$\alpha_{it}^* = g_t(\eta_{it}, \varepsilon_{it}, w_{it}, \mathbf{X}_{it}, u_{it}) \tag{E25}$$

$$\alpha_{it} = \alpha_{it}^* \cdot d_{it} \tag{E26}$$

$$d_{it} = \begin{cases} 1, & \text{if } m_t(\eta_{it}, \varepsilon_{it}, w_{it}, q(\mathbf{X}_{it})) \leq u_{it} \\ 0, & \text{otherwise} \end{cases} \tag{E27}$$

$$w_{it} = h_t(\eta_{it-1}, \varepsilon_{it-1}, w_{it-1}, \alpha_{it-1}, \mathbf{X}_{it}, \zeta_{it}) \tag{E28}$$

$$w_{i0} \text{ unrestricted} \tag{E29}$$

The specification outlined here is similar to the one outlined in the main text. There are two main differences: the first is that the variables that determine participation are the same as the ones that determine the outcome, and the second is that the error terms of equations (E25) and (E27) are the same.

E.2.2 Model specification and estimation algorithm

Participation rule. Most of the model specifications outlined in the main text remain to be the same when I move to the model of Buchinsky and Hahn (1998); the main difference is in the participation rule, equation (E27). The specification now becomes:

$$\Pr(d_{it} = 1 | \eta_{it}, \varepsilon_{it}, w_{it}, age_{it}, \mathbf{X}_{it}) = \Lambda\left(\sum_{k=0}^{K} b_k^p \varphi_k(\eta_{it}, \varepsilon_{it}, w_{it}, age_{it}) + \gamma^p \mathbf{X}_{it}\right) \tag{E30}$$

where $\Lambda(\cdot)$ is the logistic function and ϕ_k is a dictionary of functions.[31]

[31] Buchinsky and Hahn (1998) propose to estimate the propensity score with a nonparametric kernel density estimator, as the propensity score depends on the latent distribution of outcomes. However, this leads to a less computationally tractable estimation procedure in the context of the nonlinear reduced form model. Hence, I specify the propensity score with this model. An added advantage is the possibility of calculating extensive margins of income components and wealth.

Overview of the estimation algorithm. The M-step that corresponds with Buchinsky and Hahn (1998) is characterized by the following steps. First, I estimate the participation rule:

$$\max_{(b_0^P,\ldots,b_K^P,\gamma^P)} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{m=1}^{M} d_{it}\log\left[\Lambda\left(\sum_{k=0}^{K} b_k^p \varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})+\gamma^p\mathbf{X}_{it}\right)\right]$$
$$+(1-d_{it})\log\left[1-\Lambda\left(\sum_{k=0}^{K} b_k^p \varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})+\gamma^p\mathbf{X}_{it}\right)\right]. \quad \text{(E31)}$$

From here, I can compute the propensity score $p(\mathbf{x}_{it})$; that is, the probability that a household participates in the stock market. In the second step, I estimate the following censored quantile regression, which updates the parameters of the portfolio rule:

$$\min_{(b_0^a,\ldots,b_K^a)} \sum_{i=1}^{N}\sum_{t=2}^{T}\sum_{m=1}^{M} d_{it}\mathbf{1}\{h_\tau(\mathbf{x}_{it})>0\}\left[h_\tau(\mathbf{x}_{it})\left(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K} b_k^a(\tau)\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})+\gamma^a(\tau)'\mathbf{X}_{it}\right)^+\right.$$
$$\left.+(1-h_\tau(\mathbf{x}_{it}))\left(\Lambda^{-1}(\alpha_{it}^*)-\sum_{k=0}^{K} b_k^a(\tau)\varphi_k(\eta_{it},\varepsilon_{it},w_{it},age_{it})+\gamma^a(\tau)'\mathbf{X}_{it}\right)^-\right] \quad \text{(E32)}$$

where $h_\tau(\mathbf{x}_{it}) = \frac{\tau+p(\mathbf{x}_{it})-1}{p(\mathbf{x}_{it})}$. The role of this function is to "shift" the mass from the unobserved to the observed part of the distribution of portolio shares. In fact, $h_\tau(\mathbf{x}_{it})$ provides the link between Buchinsky and Hahn (1998) and Arellano and Bonhomme (2017a). This is because the conditional copula of the error terms of the participation and portfolio rules when there is no exclusion restriction and where the error terms are the same is the lower Fréchet bound, i.e., $G^-(\tau,p)=\max\left\{\frac{\tau+p(\mathbf{x}_{it})-1}{p(\mathbf{x}_{it})},0\right\}$.

As the model restrictions and implementation are similar as in the main text, I do not outline them here. I show, however, the likelihood function implied by the model.

Likelihood function. The corresponding likelihood function has the following form:

$$f(\alpha_i^T,\eta_i^T,\varepsilon_i^T,m_i^T,\mathbf{X_i},d_i^T;\bar{\theta}) = \prod_{t=1}^{T}\left[f(\alpha_{it}^*|\eta_{it},\varepsilon_{it},m_{it},\mathbf{x}_{it})p(d_{it}=1|\eta_{it},\varepsilon_{it},m_{it},\mathbf{x}_{it})\nabla C(u,v;c)\right]^{d_{it}}$$
$$\times\prod_{t=1}^{T}\left[p(d_{it}=0|\eta_{it},\varepsilon_{it},m_{it},\mathbf{x}_{it})\right]^{1-d_{it}}\prod_{t=2}^{T} f(m_{it}|m_{it-1},\eta_{it-1},y_{it-1},\alpha_{it-1},\mathbf{x}_{it})$$
$$\times f(m_{i1}|\eta_{i1},\mathbf{x}_{i1})\prod_{t=1}^{T} f(y_{it}|\eta_{it})\prod_{t=2}^{T} f(\eta_{it}|\eta_{it-1})f(\eta_{i1}) \quad \text{(E33)}$$

I can simplify the likelihood function further by noting that I can rewrite the lower Fréchet bound as follows:

$$C(u,v;c)=\begin{cases}\frac{\tau+p(\mathbf{x}_{it})-1}{p(\mathbf{x}_{it})}, \text{ if } p(\mathbf{x}_{it})>1-\tau\\ 0, \text{ otherwise}\end{cases}$$

It follows that the first derivative of this function with respect to the first argument is:

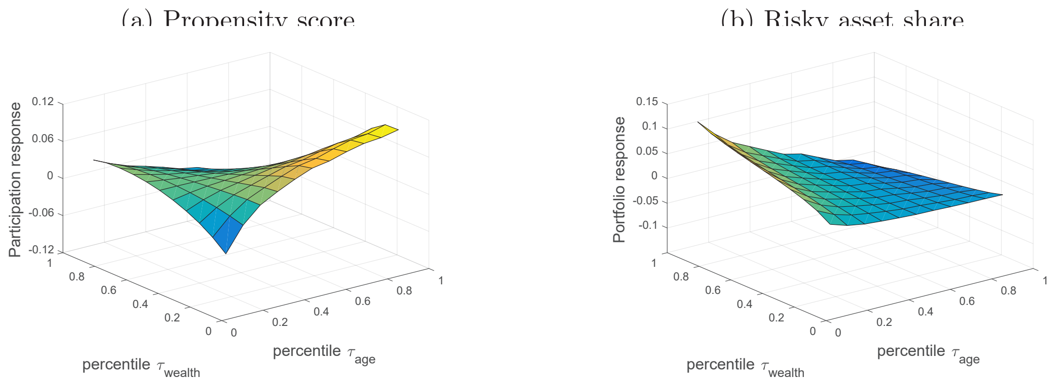
$$\nabla C(u,v;c)=\begin{cases}\frac{1}{p(\mathbf{x}_{it})}, \text{ if } p(\mathbf{x}_{it})>1-\tau\\ 0, \text{ otherwise}\end{cases}$$

Substituting this, the likelihood function above simplifies to:

$$\begin{aligned}f(\alpha_i^T,\eta_i^T,\varepsilon_i^T,m_i^T,\mathbf{X_i},d_i^T;\bar{\theta}) &= \prod_{t=1}^{T}[f(\alpha_{it}^*|\eta_{it},\varepsilon_{it},m_{it},\mathbf{x}_{it})]^{d_{it}}\prod_{t=1}^{T}[p(d_{it}=0|\eta_{it},\varepsilon_{it},m_{it},\mathbf{x}_{it})]^{1-d_{it}}\\ &\times\prod_{t=2}^{T}f(m_{it}|m_{it-1},\eta_{it-1},y_{it-1},\alpha_{it-1},\mathbf{x}_{it})f(m_{i1}|\eta_{i1},\mathbf{x}_{i1})\\ &\times\prod_{t=1}^{T}f(y_{it}|\eta_{it})\prod_{t=2}^{T}f(\eta_{it}|\eta_{it-1})f(\eta_{i1}) \qquad \text{(E34)}\end{aligned}$$

F Additional empirical evidence

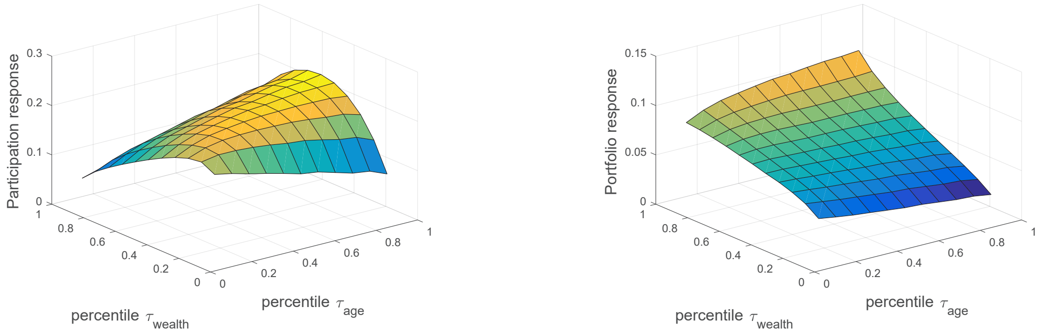
Figure F1: Average derivative effect of the persistent component of income η_{it}, Buchinsky and Hahn (1998)



Note: The graphs show average derivatives of the propensity score and the risky asset share of stock market participants, respectively, with respect to wealth w_{it} given w_{it}, persistent component η_{it}, income y_{it}, and age age_{it}, evaluated at different values of w_{it} and age_{it} that correspond to their τ_{wealth} and τ_{age} percentiles. All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator.

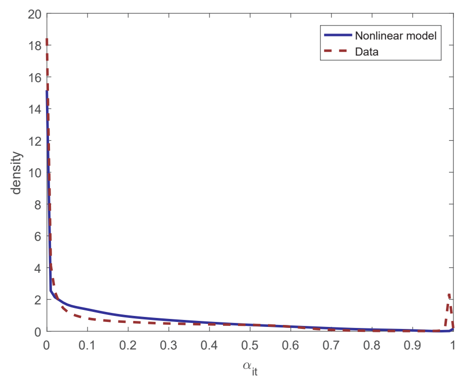
Figure F2: Average derivative effect of wealth, Buchinsky and Hahn (1998)

(a) Propensity score (b) Risky asset share



Note: The graphs show average derivatives of the portfolio rule with respect to wealth w_{it} and persistent income η_{it}, respectively, given w_{it}, persistent component η_{it}, income y_{it}, and age age_{it}, evaluated at different values of w_{it} and age_{it} that correspond to their τ_{wealth} and τ_{age} percentiles. All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator.

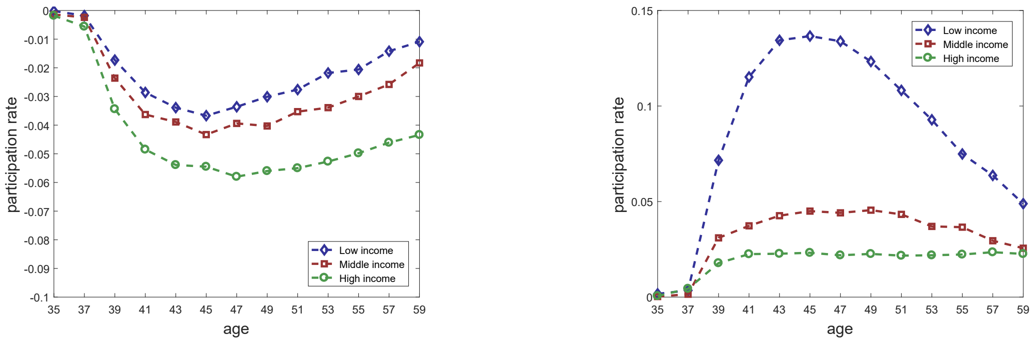
Figure F3: Observed and implied densities of the risky asset share, Buchinsky and Hahn (1998)



Note: The graph shows the observed and predicted unconditional densities of the share of household wealth in risky assset share based on the nonlinear model. The blue line corresponds to the density implied by the nonlinear model, while the red line corresponds to the density implied by the data. All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator.

Figure F4: Impulse response, participation rule, Buchinsky and Hahn (1998)

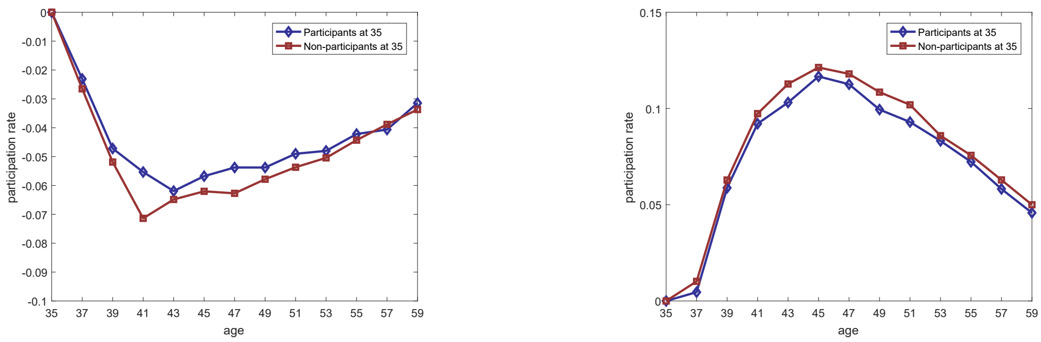
(a) $\tau_{shock} = 0.1$ (b) $\tau_{shock} = 0.9$



Note: The graphs show the difference in average participation rates between a household hit by a shock τ_{shock} at age 37, and a household hit by 0.5 shock at the same age. The blue line corresponds to low-income households (i.e., rank of $\tau_{init} = 0.1$ in the income distribution). The red line corresponds to middle-income households (i.e., rank of $\tau_{init} = 0.5$ in the income distribution). The green line corresponds to high-income households (i.e., rank of $\tau_{init} = 0.9$ in the income distribution). All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator.

Figure F5: Impulse response, participation rule, by participation status, Buchinsky and Hahn (1998)

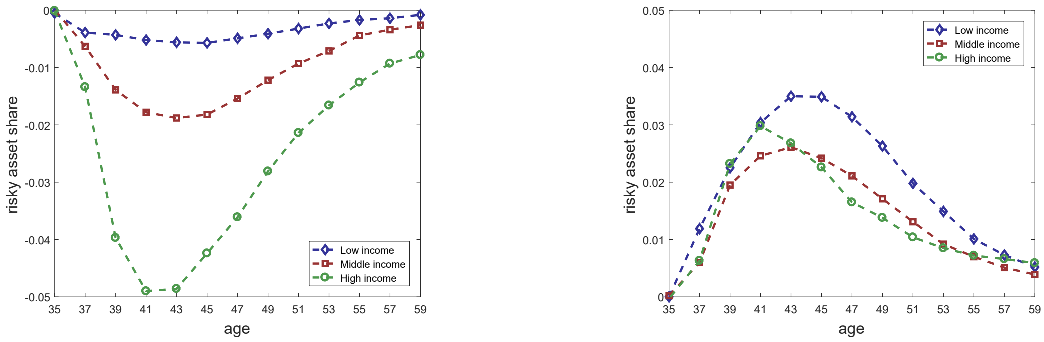
(a) $\tau_{init} = 0.9$, $\tau_{shock} = 0.1$ (b) $\tau_{init} = 0.1$, $\tau_{shock} = 0.9$



Note: The graphs show the difference in average participation rates between a household with rank τ_{init} hit by a shock τ_{shock} at age 37, and a household hit by 0.5 shock at the same age, conditional on participation status at age 35. The blue line corresponds to stock market participants. The red line corresponds to stock market non-participants. All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator.

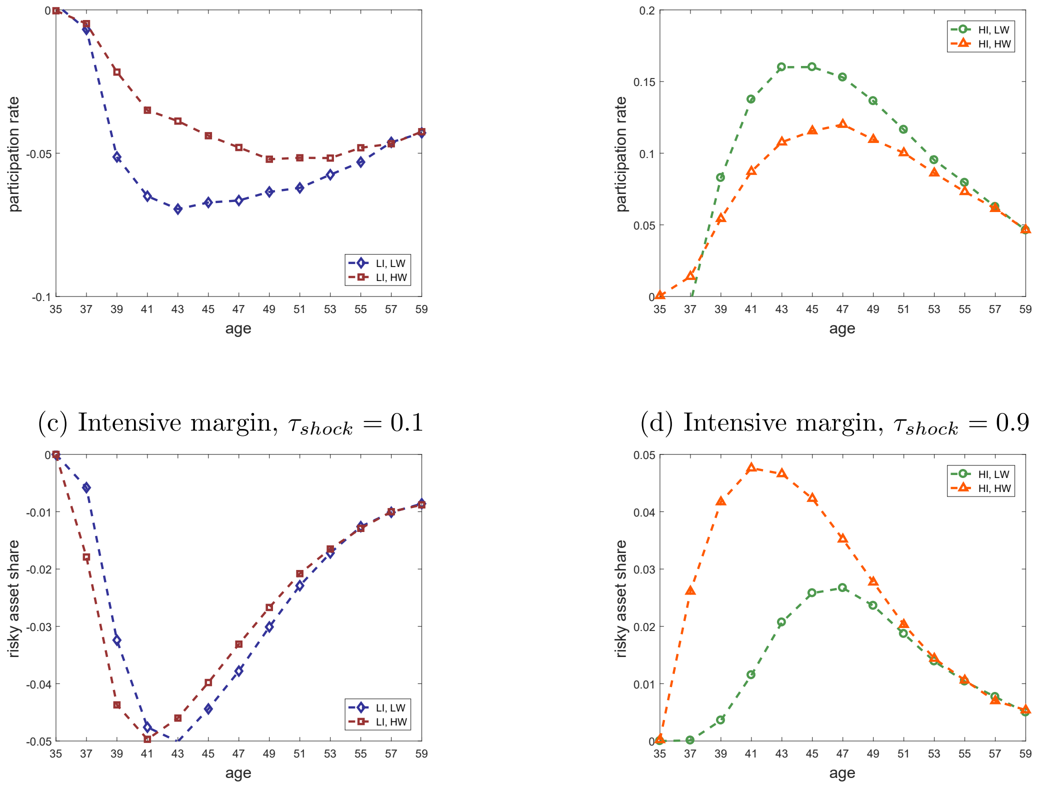
Figure F6: Impulse response, portfolio rule, Buchinsky and Hahn (1998)

(a) $\tau_{shock} = 0.1$ (b) $\tau_{shock} = 0.9$



Note: The graphs show the difference in average portfolio shares conditional on participation between a household hit by a shock τ_{shock} at age 37, and a household hit by 0.5 shock at the same age. The blue line corresponds to low-income households (i.e., rank of $\tau_{init} = 0.1$ in the income distribution). The red line corresponds to middle-income households (i.e., rank of $\tau_{init} = 0.5$ in the income distribution). The green line corresponds to high-income households (i.e., rank of $\tau_{init} = 0.9$ in the income distribution). All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator.

Figure F7: Impulse responses to an income shock, by income and wealth at age 35, Buchinsky and Hahn (1998)



(a) Extensive margin, $\tau_{shock} = 0.1$

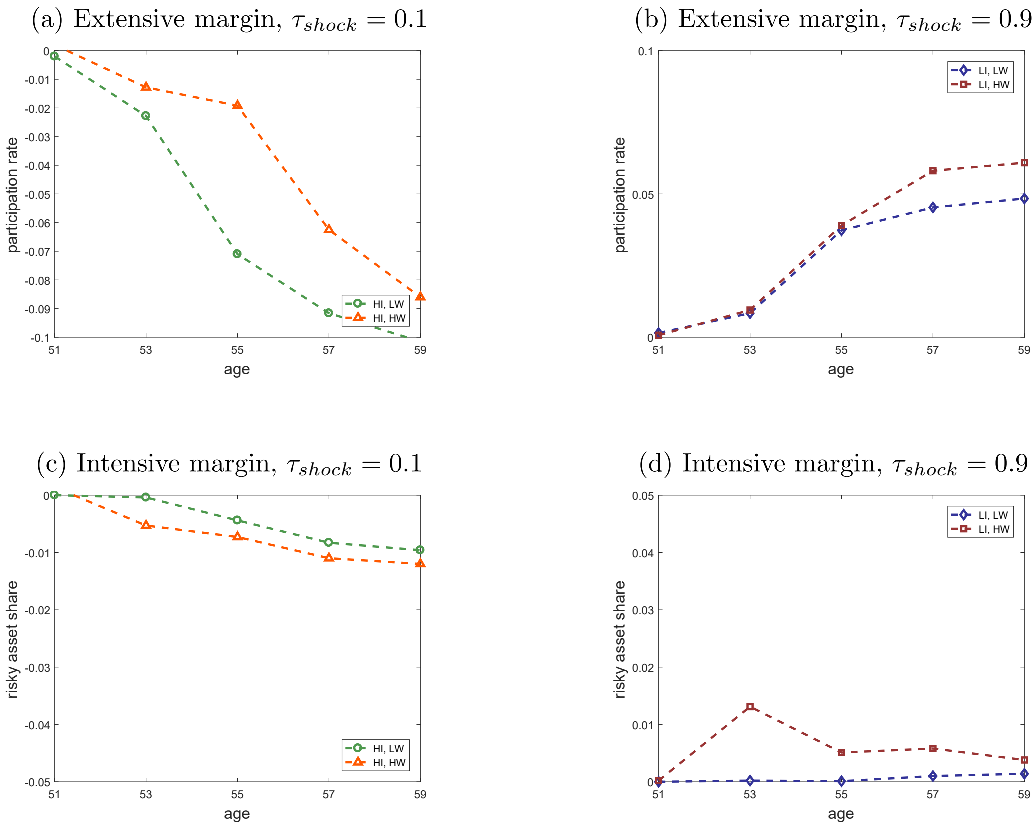
(b) Extensive margin, $\tau_{shock} = 0.9$

(c) Intensive margin, $\tau_{shock} = 0.1$

(d) Intensive margin, $\tau_{shock} = 0.9$

Note: The graphs show the difference between a household hit by a shock τ_{shock} at age 37, and a household hit by a 0.5 shock at the same age, by income and wealth categories. The blue line corresponds to low income, low wealth households. The red line corresponds to low income, high wealth households. The green line corresponds to high income, low wealth households. The orange line corresponds to high income, high wealth households. All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator.

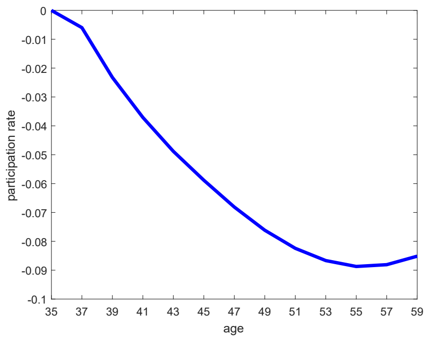
Figure F8: Impulse responses to an income shock, by income and wealth at age 51, Buchinsky and Hahn (1998)



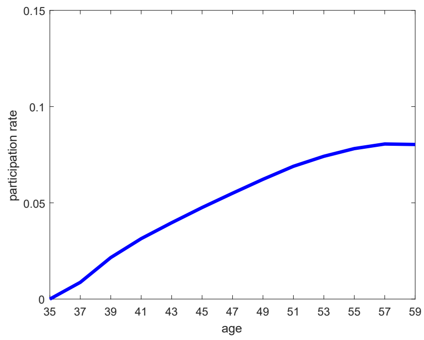
Note: The graphs show the difference between a household hit by a shock τ_{shock} at age 53, and a household hit by a 0.5 shock at the same age, by income and wealth categories. The blue line corresponds to low income, low wealth households. The red line corresponds to low income, high wealth households. The green line corresponds to high income, low wealth households. The orange line corresponds to high income, high wealth households. All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator.

Figure F9: Impulse response, participation rule, linear earnings process

(a) $\tau_{shock} = 0.1$



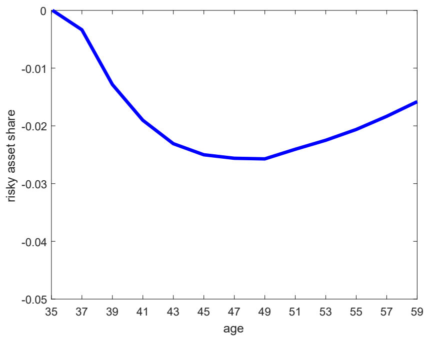
(b) $\tau_{shock} = 0.9$



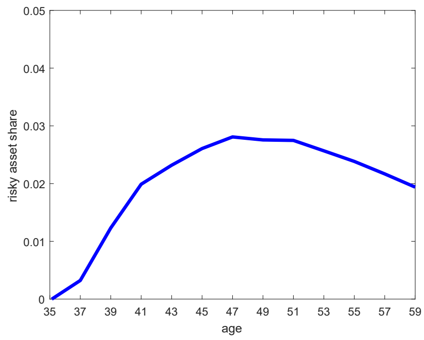
Note: The graphs show the difference in average participation rates between a household hit by a shock τ_{shock} at age 37, and a household hit by a median shock at the same age. All results are based on estimates from a linear portfolio choice rule with a linear earnings process via Tobit regressions.

Figure F10: Impulse response, linear approximation to the portfolio rule, linear earnings process

(a) $\tau_{shock} = 0.1$



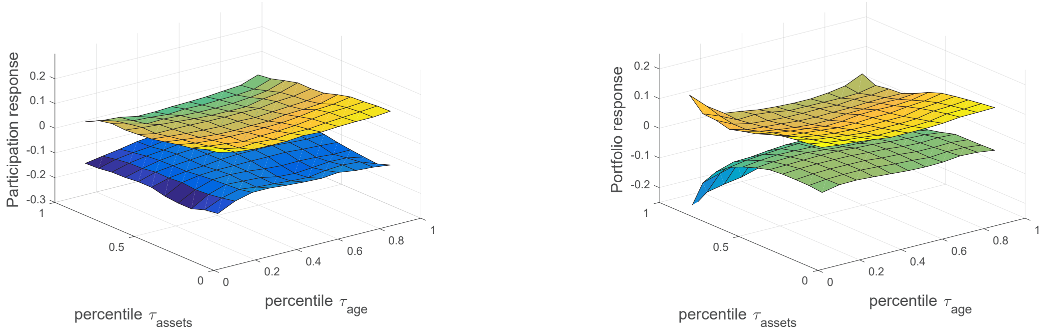
(b) $\tau_{shock} = 0.9$



Note: The graphs show the difference in average risky asset shares between a household hit by a shock τ_{shock} at age 37, and a household hit by a median shock at the same age. All results are based on estimates from a linear portfolio choice rule with a linear earnings process via Tobit regressions.

Figure F11: Parametric bootstraps, average derivative effect of the persistent component η_{it}, Arellano et al. (2017) estimation

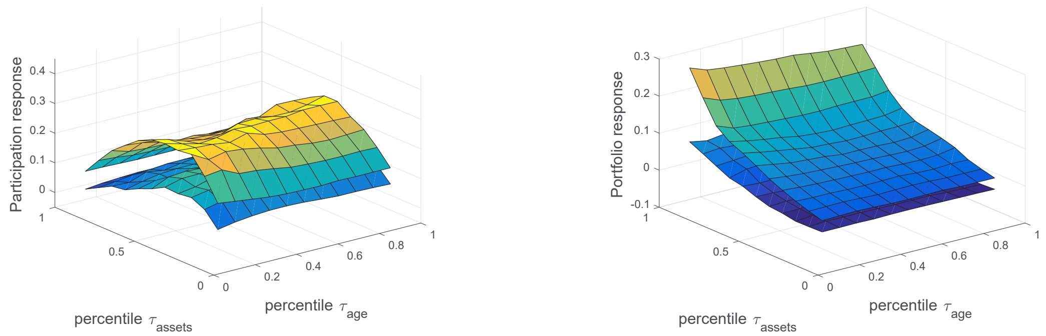
(a) Propensity score (b) Risky asset share



Note: The graphs show the 95% pointwise confidence bands of the average derivatives of the propensity score and the risky asset share of stock market participants, respectively, with respect to wealth w_{it} given w_{it}, persistent component η_{it}, income y_{it}, and age age_{it}, evaluated at different values of w_{it} and age_{it} that correspond to their τ_{wealth} and τ_{age} percentiles. All results are based on estimates from the semi-structural model with the Arellano and Bonhomme (2017a) quantile selection model estimator. Parametric bootstrap with 100 replications.

Figure F12: Parametric bootstraps, average derivative effect of wealth, Arellano et al. (2017) estimation

(a) Propensity score (b) Risky asset share

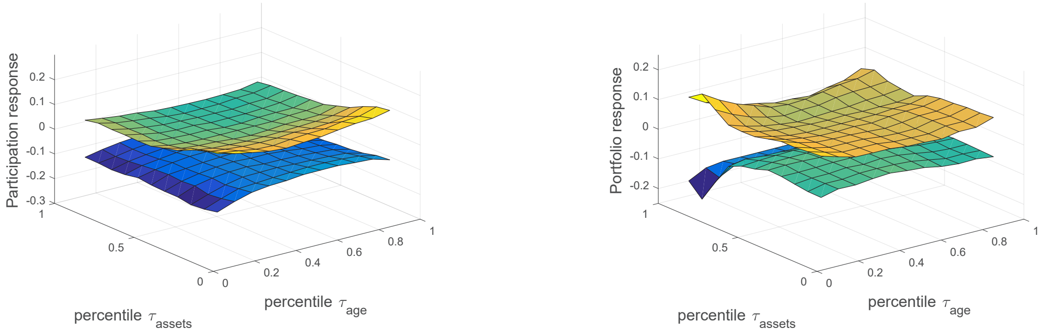


Note: The graphs show the 95% pointwise confidence bands of the average derivatives of the portfolio rule with respect to wealth w_{it} and persistent income η_{it}, respectively, given w_{it}, persistent component η_{it}, income y_{it}, and age age_{it}, evaluated at different values of w_{it} and age_{it} that correspond to their τ_{wealth} and τ_{age} percentiles. All results are based on estimates from the semi-structural model with the Arellano and Bonhomme (2017a) quantile selection model estimator. Parametric bootstrap with 100 replications.

Figure F13: Parametric bootstraps, average derivative effect of the persistent component η_{it}, Buchinsky and Hahn (1998) estimation

(a) Propensity score

(b) Risky asset share

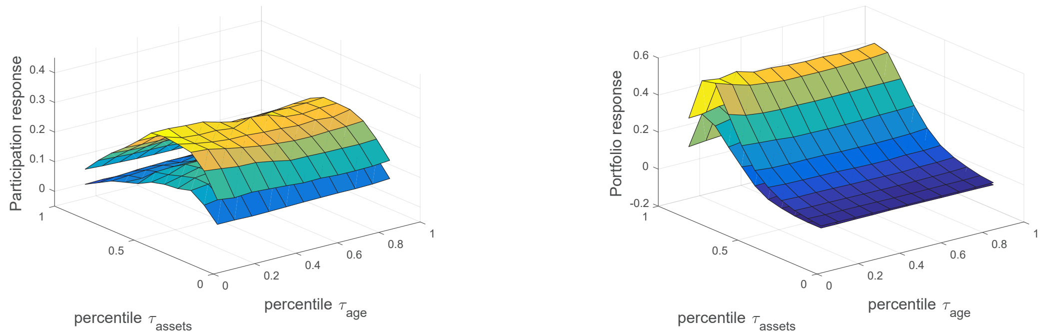


Note: The graphs show the 95% pointwise confidence bands of the average derivatives of the propensity score and the risky asset share of stock market participants, respectively, with respect to wealth w_{it} given w_{it}, persistent component η_{it}, income y_{it}, and age age_{it}, evaluated at different values of w_{it} and age_{it} that correspond to their τ_{wealth} and τ_{age} percentiles. All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator. Parametric bootstrap with 100 replications.

Figure F14: Parametric bootstraps, average derivative effect of wealth, Buchinsky and Hahn (1998) estimation

(a) Propensity score

(b) Risky asset share



Note: The graphs show the 95% pointwise confidence bands of the average derivatives of the portfolio rule with respect to wealth w_{it} and persistent income η_{it}, respectively, given w_{it}, persistent component η_{it}, income y_{it}, and age age_{it}, evaluated at different values of w_{it} and age_{it} that correspond to their τ_{wealth} and τ_{age} percentiles. All results are based on estimates from the semi-structural model with the Buchinsky and Hahn (1998) censored quantile regression estimator. Parametric bootstrap with 100 replications.

BANCO DE ESPAÑA PUBLICATIONS

WORKING PAPERS

2215 JOSÉ MANUEL CARBÓ and SERGIO GORJÓN: Application of machine learning models and interpretability techniques to identify the determinants of the price of bitcoin.

2216 LUIS GUIROLA and MARÍA SÁNCHEZ-DOMÍNGUEZ: Childcare constraints on immigrant integration.

2217 ADRIÁN CARRO, MARC HINTERSCHWEIGER, ARZU ULUC and J. DOYNE FARMER: Heterogeneous effects and spillovers of macroprudential policy in an agent-based model of the UK housing market.

2218 STÉPHANE DUPRAZ, HERVÉ LE BIHAN and JULIEN MATHERON: Make-up strategies with finite planning horizons but forward-looking asset prices.

2219 LAURA ÁLVAREZ, MIGUEL GARCÍA-POSADA and SERGIO MAYORDOMO: Distressed firms, zombie firms and zombie lending: a taxonomy.

2220 BLANCA JIMÉNEZ-GARCÍA and JULIO RODRÍGUEZ: A quantification of the evolution of bilateral trade flows once bilateral RTAs are implemented.

2221 SALOMÓN GARCÍA: Mortgage securitization and information frictions in general equilibrium.

2222 ANDRÉS ALONSO and JOSÉ MANUEL CARBÓ: Accuracy of explanations of machine learning models for credit decisions.

2223 JAMES COSTAIN, GALO NUÑO and CARLOS THOMAS: The term structure of interest rates in a heterogeneous monetary union.

2224 ANTOINE BERTHEAU, EDOARDO MARIA ACABBI, CRISTINA BARCELÓ, ANDREAS GULYAS, STEFANO LOMBARDI and RAFFAELE SAGGIO: The Unequal Consequences of Job Loss across Countries.

2225 ERWAN GAUTIER, CRISTINA CONFLITTI, RIEMER P. FABER, BRIAN FABO, LUDMILA FADEJEVA, VALENTIN JOUVANCEAU, JAN-OLIVER MENZ, TERESA MESSNER, PAVLOS PETROULAS, PAU ROLDAN-BLANCO, FABIO RUMLER, SERGIO SANTORO, ELISABETH WIELAND and HÉLÈNE ZIMMER. New facts on consumer price rigidity in the euro area.

2226 MARIO BAJO and EMILIO RODRÍGUEZ: Integrating the carbon footprint into the construction of corporate bond portfolios.

2227 FEDERICO CARRIL-CACCIA, JORDI PANIAGUA and MARTA SUÁREZ-VARELA: Forced migration and food crises.

2228 CARLOS MORENO PÉREZ and MARCO MINOZZO: Natural Language Processing and Financial Markets: Semi-supervised Modelling of Coronavirus and Economic News.

2229 CARLOS MORENO PÉREZ and MARCO MINOZZO: Monetary Policy Uncertainty in Mexico: An Unsupervised Approach.

2230 ADRIAN CARRO: Could Spain be less different? Exploring the effects of macroprudential policy on the house price cycle.

2231 DANIEL SANTABÁRBARA and MARTA SUÁREZ-VARELA: Carbon pricing and inflation volatility.

2232 MARINA DIAKONOVA, LUIS MOLINA, HANNES MUELLER, JAVIER J. PÉREZ and CRISTOPHER RAUH: The information content of conflict, social unrest and policy uncertainty measures for macroeconomic forecasting.

2233 JULIAN DI GIOVANNI, MANUEL GARCÍA-SANTANA, PRIIT JEENAS, ENRIQUE MORAL-BENITO and JOSEP PIJOAN-MAS: Government Procurement and Access to Credit: Firm Dynamics and Aggregate Implications.

2234 PETER PAZ: Bank capitalization heterogeneity and monetary policy.

2235 ERIK ANDRES-ESCAYOLA, CORINNA GHIRELLI, LUIS MOLINA, JAVIER J. PÉREZ and ELENA VIDAL: Using newspapers for textual indicators: which and how many?

2236 MARÍA ALEJANDRA AMADO: Macroprudential FX regulations: sacrificing small firms for stability?

2237 LUIS GUIROLA and GONZALO RIVERO: Polarization contaminates the link with partisan and independent institutions: evidence from 138 cabinet shifts.

2238 MIGUEL DURO, GERMÁN LÓPEZ-ESPINOSA, SERGIO MAYORDOMO, GAIZKA ORMAZABAL and MARÍA RODRÍGUEZ-MORENO: Enforcing mandatory reporting on private firms: the role of banks.

2239 LUIS J. ÁLVAREZ and FLORENS ODENDAHL: Data outliers and Bayesian VARs in the Euro Area.

2240 CARLOS MORENO PÉREZ and MARCO MINOZZO: “Making text talk”: The minutes of the Central Bank of Brazil and the real economy.

2241 JULIO GÁLVEZ and GONZALO PAZ-PARDO: Richer earnings dynamics, consumption and portfolio choice over the life cycle.

2242 MARINA DIAKONOVA, CORINNA GHIRELLI, LUIS MOLINA and JAVIER J. PÉREZ: The economic impact of conflict-related and policy uncertainty shocks: the case of Russia.

2243 CARMEN BROTO, LUIS FERNÁNDEZ LAFUERZA and MARIYA MELNYCHUK: Do buffer requirements for European systemically important banks make them less systemic?

2244 GERGELY GANICS and MARÍA RODRÍGUEZ-MORENO: A house price-at-risk model to monitor the downside risk for the Spanish housing market.

2245 JOSÉ E. GUTIÉRREZ and LUIS FERNÁNDEZ LAFUERZA: Credit line runs and bank risk management: evidence from the disclosure of stress test results.

2301 MARÍA BRU MUÑOZ: The forgotten lender: the role of multilateral lenders in sovereign debt and default.

2302 SILVIA ALBRIZIO, BEATRIZ GONZÁLEZ and DMITRY KHAMETSHIN: A tale of two margins: monetary policy and capital misallocation.

2303 JUAN EQUIZA, RICARDO GIMENO, ANTONIO MORENO and CARLOS THOMAS: Evaluating central bank asset purchases in a term structure model with a forward-looking supply factor.

2304 PABLO BURRIEL, IVÁN KATARYNIUK, CARLOS MORENO PÉREZ and FRANCESCA VIANI: New supply bottlenecks index based on newspaper data.

2305 ALEJANDRO FERNÁNDEZ-CEREZO, ENRIQUE MORAL-BENITO and JAVIER QUINTANA: A production network model for the Spanish economy with an application to the impact of NGEU funds.

2306 MONICA MARTINEZ-BRAVO and CARLOS SANZ: Trust and accountability in times of pandemic.

2307 NATALIA FABRA, EDUARDO GUTIÉRREZ, AITOR LACUESTA and ROBERTO RAMOS: Do Renewables Create Local Jobs?

2308 ISABEL ARGIMÓN and IRENE ROIBÁS: Debt overhang, credit demand and financial conditions.

2309 JOSÉ-ELÍAS GALLEGOS: Inflation persistence, noisy information and the Phillips curve.

2310 ANDRÉS ALONSO-ROBISCO, JOSÉ MANUEL CARBÓ and JOSÉ MANUEL MARQUÉS: Machine Learning methods in climate finance: a systematic review.

2311 ALESSANDRO PERI, OMAR RACHEDI and IACOPO VAROTTO: The public investment multiplier in a production network.

2312 JUAN S. MORA-SANGUINETTI, JAVIER QUINTANA, ISABEL SOLER and ROK SPRUK: Sector-level economic effects of regulatory complexity: evidence from Spain.

2313 CORINNA GHIRELLI, ENKELEJDA HAVARI, ELENA MERONI and STEFANO VERZILLO: The long-term causal effects of winning an ERC grant.

2314 ALFREDO GARCÍA-HIERNAUX, MARÍA T. GONZÁLEZ-PÉREZ and DAVID E. GUERRERO: How to measure inflation volatility. A note.

2315 NICOLÁS ABBATE, INÉS BERNIELL, JOAQUÍN COLEFF, LUIS LAGUINGE, MARGARITA MACHELETT, MARIANA MARCHIONNI, JULIÁN PEDRAZZI and MARÍA FLORENCIA PINTO: Discrimination against gay and transgender people in Latin America: a correspondence study in the rental housing market.

2316 SALOMÓN GARCÍA: The amplification effects of adverse selection in mortgage credit suply.

2317 METTE EJRNÆS, ESTEBAN GARCÍA-MIRALLES, METTE GØRTZ and PETTER LUNDBORG: When death was postponed: the effect of HIV medication on work, savings and marriage.

2318 GABRIEL JIMÉNEZ, LUC LAEVEN, DAVID MARTÍNEZ-MIERA and JOSÉ-LUIS PEYDRÓ: Public guarantees and private banks' incentives: evidence from the COVID-19 crisis.

2319 HERVÉ LE BIHAN, DANILO LEIVA-LEÓN and MATÍAS PACCE: Underlying inflation and asymmetric risks.

2320 JUAN S. MORA-SANGUINETTI, LAURA HOSPIDO and ANDRÉS ATIENZA-MAESO: Los números de la regulación sobre igualdad. Cuantificación de la actividad normativa sobre no discriminación en España y su relación con las brechas de género en el mercado de trabajo.

2321 ANDRES ALONSO-ROBISCO and JOSÉ MANUEL CARBÓ: Analysis of CBDC Narrative of Central Banks using Large Language Models.

2322 STEFANIA ALBANESI, ANTÓNIO DIAS DA SILVA, JUAN F. JIMENO, ANA LAMO and ALENA WABITSCH: New technologies and jobs in Europe.

2323 JOSÉ E. GUTIÉRREZ: Optimal regulation of credit lines.

2324 MERCEDES DE LUIS, EMILIO RODRÍGUEZ and DIEGO TORRES: Machine learning applied to active fixed-income portfolio management: a Lasso logit approach.

2325 SELVA BAHAR BAZIKI, MARÍA J. NIETO and RIMA TURK-ARISS: Sovereign portfolio composition and bank risk: the case of European banks.

2326 ÁNGEL IVÁN MORENO and TERESA CAMINERO: Assessing the data challenges of climate-related disclosures in european banks. A text mining study.

2327 JULIO GÁLVEZ: Household portfolio choices under (non-)linear income risk: an empirical framework.