Heterogeneity of Consumption Responses to Income Shocks in the Presence of Nonlinear Persistence*

Manuel Arellano† Richard Blundell‡ Stéphane Bonhomme § Jack Light §

January 7, 2023

Abstract

In this paper we use the enhanced consumption data in the Panel Survey of Income Dynamics (PSID) from 2005-2017 to explore the transmission of income shocks to consumption. We build on the nonlinear quantile framework introduced in Arellano, Blundell and Bonhomme (2017). Our focus is on the estimation of consumption responses to persistent nonlinear income shocks in the presence of unobserved heterogeneity. To reliably estimate heterogeneous responses in our unbalanced panel, we develop Sequential Monte Carlo computational methods. We find substantial heterogeneity in consumption responses, and uncover latent types of households with different life-cycle consumption behavior. Ordering types according to their average log-consumption, we find that low-consumption types respond more strongly to income shocks at the beginning of the life cycle and when their assets are low, as standard life-cycle theory would predict. In contrast, high-consumption types respond less on average, and in a way that changes little with age or assets. We examine various mechanisms that might explain this heterogeneity.

JEL code: C23, D31, D91.

Keywords: nonlinear income persistence, consumption dynamics, partial insurance, heterogeneity, panel data.

---

*We thank the editor Xiaohong Chen, an associate editor, two anonymous referees, and participants at the Whitney Newey Conference at MIT, and other venues for helpful comments. Arellano acknowledges research funding from the Ministerio de Economia y Competitividad, Grant ECO2016-79848-P. Blundell would like to thank the ESRC Centre for the Microeconomic Analysis of Public Policy (CPP) at the Institute for Fiscal Studies (ES/T014334/1) for support. Bonhomme acknowledges support from the NSF, Grant SES-1658920.

†CEMFI, Madrid.

‡University College London and Institute for Fiscal Studies.

§University of Chicago.

1 Introduction

The empirical analysis of consumption and income dynamics has an important place in a number of key areas of economic research and policy design. A large literature aims at understanding income persistence, income inequality and income volatility, see Moffitt and Gottschalk (1995), Baker and Solon (2003) and references in Meghir and Pistaferri (2011). A parallel literature studies how income shocks impact consumption and savings decisions, see Hall and Mishkin (1982) and Blundell, Pistaferri and Preston (2008) among many other references. In this paper our goal is to empirically document the nature of consumption responses, with a particular focus on household heterogeneity and nonlinear persistence.

Economic models inform the empirical analysis of consumption and income. In a standard incomplete markets model of the life cycle, how much a household consumes in a given period is determined by the level of assets, the stage of the life cycle, as well as the income stream, see Deaton (1992) for a comprehensive review. Changes to income components with different degrees of persistence lead to different consumption responses. In addition, the shape of the consumption function may differ among households for a variety of reasons, such as heterogeneity in preferences or discounting, household-specific returns to assets, or heterogeneous access to other sources of insurance.

Our starting point is the nonlinear panel data framework proposed by Arellano, Blundell and Bonhomme (2017, ABB hereafter) which involves a Markovian permanent-transitory model of income, and a flexible age-dependent nonlinear consumption rule that is a function of assets, permanent income and transitory income. ABB found that individual income dynamics feature nonlinearities that matter for economic decisions. Specifically, they found evidence that the persistence of past earnings varies substantially with the sign and magnitude of shocks across the past earnings distribution. Thus, *ex ante* identical individuals may have experienced a very different propagation of a past shock into their income depending on their history of subsequent shocks. Using a balanced panel from the PSID, from 1999 to 2009, ABB showed how nonlinear income dynamics lead to nonlinear responses of consumption to income shocks.[1]

Given this background we make three main contributions. First, we exploit the important extension to the set of consumption goods in the recent waves of the PSID to produce new estimates of the degree of nonlinear persistence and consumption insurance. The improved panel survey redesign in the 1999 PSID was further enhanced in 2005 and, in addition to food at home and food away from home, includes health expenditures, utilities, gasoline, car maintenance, transportation, education, clothing, and leisure activities, see Andreski, Geng, Samancioglu and Schoeni (2014). We bring this together with the detailed data on earnings, family income, and financial and real estate assets. Using the 2005-2017 PSID panel survey waves, we estimate the nonlinear nature of income shocks and the consumption

[1] See De Nardi, Fella and Paz-Pardo (2020) and Anghel, Basso, Bover, Casado, Hospido, Izquierdo and Vozmediano (2018) for recent applications of the nonlinear dynamic approach introduced in ABB.

implications of the insurance to income shocks. In addition, unlike ABB we do not restrict the sample to be balanced. This leads us to consider a larger and more comprehensive sample, more than 2000 households compared to approximately 800 in ABB.

Our second main contribution is to empirically document household heterogeneity in consumption responses. To do so, we move away from the partial insurance consumption growth framework of Blundell, Pistaferri and Preston (2008) and estimate a dynamic model where we specify the entire conditional distribution of consumption given assets, age, and the income components. This modeling approach contrasts with that adopted in ABB, who specified the link between consumption and its determinants using a nonlinear mean model with separable heterogeneity. Allowing for non-separabilities, we show how to estimate the joint distribution of latent and observed variables, and to consistently estimate log-derivatives of the consumption function as a result.[2]

The average log-derivatives of the consumption function that we focus on are nonlinear coefficients quantifying how well insured households are, at different points of the life cycle and depending on their level of assets. Importantly, we model the consumption function as heterogeneous across households, by indexing consumption on a latent time-invariant continuous type. This unobserved consumer type may reflect heterogeneity in economic primitives, and leads to different consumption derivative responses for two households that are at the same point of the life cycle, face the same income stream, and own the same level of assets. We show this heterogeneity to be a salient feature of the PSID.

To study a larger sample using a more complex model, we modify the computational techniques that ABB relied on. The use of new computational tools represents our third main contribution. Specifically, we examine improved sequential computational methods for the estimation of the nonlinear latent/hidden quantile Markov model. The Markovian structure for latent earnings components allows us to make use of Sequential Monte-Carlo (SMC) methods to improve the Markov Chain Monte Carlo algorithm, see Creal (2012) for a review. SMC methods can be used to generate efficient proposals within a Particle Markov Chain Monte Carlo (PMCMC) algorithm, as proposed by Andrieu, Doucet and Hollenstein (2010). We develop an implementation in the latent Markov setting of this paper. The PMCMC approach allows us to produce numerically robust estimates of derivatives of log-consumption with respect to the latent income components, in a nonlinear quantile model that allows for unobserved types.

Empirically, we confirm the nonlinear income dynamics found in ABB while documenting new

[2] As we will explain below, our approach exploits the weak exogeneity of the observed state variables (i.e., assets and income components), conditional on a latent time-invariant type, to identify average response functions, see Matzkin (2013) for a review of identification results in models with non-separable heterogeneity. Relaxing exogeneity would require valid instruments and appropriate structure on the first stage (Imbens and Newey, 2009). Also, while the distribution of consumption responses is generally not identified beyond its mean, partial information about this distribution can be obtained by using a result from Hoderlein and Mammen (2007). We will apply this strategy to compute a lower bound on the variance of responses.

patterns in consumption responses. The estimated quantile Markovian permanent-transitory model of income reveals asymmetric persistence of earnings and income shocks. We show the use of enhanced computational techniques leads to essentially the same results as ABB in their balanced sample. However, estimates based on SMC techniques are more stable numerically. The use of sequential Monte Carlo methods allows us to draw robust conclusions in our larger unbalanced sample, and to document nonlinear patterns in the dynamics of income.

Our main results concern the nature of consumption responses to income shocks. We find that older and wealthier households adjust their consumption less as a response to an income shock than younger and less wealthy households. For our main sample of dual earners the average derivative of log-consumption to the persistent income component is 0.33 on overall average, yet it can be much higher for younger households with low levels of assets and, conversely, as low as 0.10 for older and wealthy households. These findings are qualitatively consistent with the implications of standard life-cycle models of consumption and saving behavior. We show that accounting for latent income components with varying degrees of persistence, and for unobserved heterogeneity in consumption, are both important to accurately document these patterns quantitatively. Heterogeneity in consumer responses to income shocks matters for understanding the impact not only of fiscal policies but also of monetary policies which, as Auclert (2019) notes, can create large redistribution in favor of high MPC agents and be expansionary over and beyond the effect on real interest rates.

Our key finding is that consumption responses vary substantially with unobserved types. Our results clearly separate lower consumption types, who appear to follow the life-cycle patterns in consumption responses implied by standard models, from higher types, whose consumption responses to income shocks vary little with either assets levels or the stage of the life cycle. High-type households consistently have higher consumption levels, and relative to low-type households they have slightly higher incomes and levels of assets. For the younger low types, consumption responses to persistent income shocks exceed 0.50 while for older low types this falls below 0.20. Moreover, based on bootstrapped confidence intervals we conclude the difference between the two coefficients is significant at conventional levels. For the higher types, consumption responses are flatter across age and assets, and differences across age and assets are insignificant. These findings shed new light on the presence of heterogeneity in consumption behavior across households, on which there has been extensive micro- and macroeconomic research, see Alan, Browning and Ejrnæs (2018), Crawley and Kuchler (2020), and references therein.

We examine several mechanisms that could lead to such heterogeneous consumption responses. First, the fact that high types consume more and hold more assets is difficult to reconcile with an explanation based on heterogeneity in preferences or discounting. Second, we estimate a specification that allows for latent heterogeneity in asset accumulation and find that the heterogeneity in consumption responses is virtually unaffected. Lastly, we link a subset of household heads in our sample (33%)

to their parents, using the inter-generational linkages that the PSID provides. We find that high-type household heads have on average parents with higher consumption and income levels, suggesting that the heterogeneous responses that we find might in part reflect heterogeneity in access to other sources of insurance such as parental insurance.

We show the main results are robust to a number of specification changes. In particular, while we use disposable income in most of the analysis, we find similar patterns when using pre-tax labor income, with some quantitative differences. In addition, we find that including households where one member may not be working does not lead to major changes in our results. Lastly, we probe the robustness of our scalar individual effect modeling approach by allowing for a separate effect of education on consumption responses, in addition to the latent type. While the heterogeneity results remain qualitatively similar, the findings based on this specification allow us to discuss some limitations of our scalar individual effect modeling approach and to motivate future work.

The outline of the paper is as follows. In Section 2 we describe the sample and present motivating evidence on the nature of consumption responses. In Section 3 we provide a general description of the model, and in Section 4 we discuss implementation and present the computational methods we use. We then show our main empirical results in Section 5. In Section 6 we study possible mechanisms for those results. In Section 7 we show results based on extensions of our main model. We conclude in Section 8. An appendix describes implementation and provides additional empirical results. Lastly, replication codes are available in an online repository.

2 Data

In this section we describe the PSID sample, and we provide preliminary motivating evidence about how consumption responds to income changes.

2.1 The PSID sample

We rely on the newly redesigned PSID, from 2005 to 2017. Since 1999, the PSID presents a unique combination of longitudinal data on income, consumption, and assets holdings for the US. Unlike the annual information available every year before 1997, after 1999 a new wave is only available every other year. Since 2005, the consumption information has been enhanced, with additional categories, see Li, Schoeni, Danziger and Charles (2010). The recent waves include food at home and away from home, gasoline, health, transportation, utilities, clothing, and leisure activities. Andreski, Geng, Samancioglu and Schoeni (2014) provide a detailed analysis of the post-2005 data and assess the new methodology developed by the PSID for collecting household expenditure data. The new survey methodology allows unfolding brackets as well as choice of time-frame for different consumption categories. They show that since 2005 the PSID has captured almost all expenditures measured in the cross-sectional

Consumer Expenditure Survey (CE) and suggest the new measurement design is likely to improve on the accuracy of the expenditure data. For this reason, we expect the post-2005 PSID to provide more accurate information about household consumption patterns than the earlier period used in ABB.

Another difference with ABB is that we do not restrict the panel to be balanced. Following Blundell, Pistaferri and Saporta-Eksten (2016), we focus on a sample of household heads that participate in the labor market and are between 25 and 60 years old. Since we do not model labor supply, either at the extensive or intensive margin, in our baseline sample we focus on households where both adult members are working and present in at least two waves, and we keep their first spell of non-zero income observations. We refer to this baseline as the "dual earners" sample. However, in Section 7 we will also present results based on a broader sample that includes households where only one member is employed.

Our final dual earners sample contains 2,113 households and seven biennial waves from 2005-2017. In Table 1 we report some descriptive statistics about this sample. Food consumption, which was the only consumption item available in the PSID prior to the redesign of the data set, accounts for approximately one fourth of total non-durables consumption. Net disposable income is approximately 30% lower than pre-tax labor income. Since it is disposable income and not pre-tax income that should affect consumption decisions, we will focus on disposable income in most of the analysis. In Section 7 we will also present results using pre-tax labor income.

Table 1 also shows that total wealth tends to decrease around the 2008 recession, whereas income and especially consumption seem more stable over the period. See Krueger, Mitman and Perri (2015) for an analysis of consumption, income and wealth using the PSID with a focus on the great recession. In our analysis we will not focus on business cycle fluctuations, and we will attempt to remove calendar time effects in a prior partialling-out estimation step.

In Appendix Table A1 we show additional statistics in order to describe the unbalanced structure of the panel sample. In the first column of that table we report statistics for households who are only observed for one wave, although we do not include these households in our main sample due to our focus on unobserved heterogeneity. More than half of households in our main sample are observed for at most three waves. For this reason, it will be important to account for the unbalancedness of the PSID in the modeling of income and consumption dynamics.

Following a common practice in the previous literature on income dynamics, we will work with residuals of log-disposable income on a set of demographics and time indicators. This partialling-out is meant to make household demographics as comparable to each other as possible, and to control for aggregate time effects. Specifically, we net out household size, year of birth, state indicators, number of kids, race of both adults, a higher education indicator for both adults interacted with age indicators, and a full set of age indicators interacted with year indicators. We similarly construct residuals of log-consumption and log-assets net of the same set of controls. Working with logarithms

Table 1: Descriptive Statistics

	(1) 2005	(2) 2007	(3) 2009	(4) 2011	(5) 2013	(6) 2015	(7) 2017
Food	10,681.46	10,652.44	10,356.33	10,516.91	10,778.89	11,287.65	11,916.79
	(5,280.66)	(5,497.57)	(5,035.15)	(5,107.21)	(5,744.91)	(5,385.16)	(5,673.31)
Non-durables (excl. food)	28,476.06	29,563.67	28,264.68	28,694.76	30,310.30	29,906.71	28,432.69
	(19,445.13)	(19,881.54)	(19,295.93)	(18,331.37)	(18,247.37)	(17,265.61)	(14,547.69)
Total Non-durables	39,179.31	40,233.90	38,669.21	39,265.89	41,129.95	41,246.63	40,383.30
	(22,220.87)	(22,516.17)	(21,678.39)	(21,154.18)	(20,962.80)	(19,845.41)	(17,547.23)
Home equity	161560.91	169580.40	137089.26	121021.37	111956.54	113269.94	130350.80
	(216942.00)	(229763.44)	(197997.93)	(166538.89)	(154874.43)	(143419.48)	(144146.96)
Negative Equity Dummy	0.01	0.01	0.03	0.03	0.02	0.01	0.01
	(0.08)	(0.10)	(0.16)	(0.16)	(0.15)	(0.09)	(0.10)
Wealth (excl. home)	206679.75	278971.16	269420.39	247951.44	231130.23	256813.63	333757.83
	(709285.07)	(1.00e+06)	(933414.69)	(536086.47)	(516957.59)	(566105.75)	(1.06e+06)
Total wealth	446917.54	512678.86	448989.83	388763.07	349033.92	370083.56	448654.75
	(970857.51)	(1.25e+06)	(1.14e+06)	(656915.67)	(621844.77)	(636801.00)	(1.07e+06)
Labor income	126181.76	127847.66	133105.34	129458.55	128366.66	124779.30	131051.39
	(143916.08)	(148500.93)	(194142.24)	(129247.51)	(128479.97)	(72,585.03)	(69,355.95)
Net income	95,598.70	97,089.32	100204.10	99,234.77	98,238.57	95,004.23	99,192.91
	(86,212.45)	(89,857.83)	(116281.39)	(78,750.29)	(77,931.32)	(46,552.59)	(45,252.48)
Observations	1288	1544	1400	1149	1023	948	755

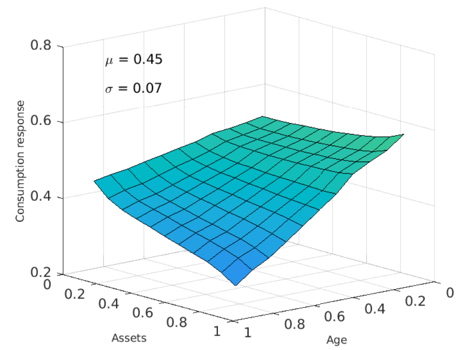
Notes: PSID, 2005-2017. Means of variables, standard deviations in parentheses. Our baseline measure of consumption includes the following categories: food at home, food delivery, eating out, food stamps, clothing, gasoline, utilities, telephone bills, automobile insurance, parking, transport, education, childcare, institutional medical services, doctor services, prescriptions, health insurance, and trips and other recreation.

requires removing observations with zero or negative assets, which reduces the number of observations by approximately 200 households per year. In Appendix Table A2 we report additional statistics for a sample which includes households with negative asset balances.

2.2 A first look at consumption responses

We will analyze the PSID sample using a dynamic model of income, consumption, and assets holdings. The model is flexibly parameterized and it features various latent variables. Before describing how we

Figure 1: Average derivative of log-consumption with respect to log-income



Notes: The graph shows averages of the derivative of log-consumption with respect to log-income, conditional on log-income, age and log-assets. Estimates are based on a linear regression of log-consumption on a second-order polynomial in log-income, age, and log-assets. The two horizontal axes show age and assets percentiles. μ and σ denote the mean and standard deviation of the average derivatives, respectively.

specify the model and estimate it, here we provide preliminary motivating evidence about consumption and income, only using observed covariates and simple econometric methods. We highlight two features of the data in turn.

In Figure 1, we show average derivatives of log-consumption with respect to log-income, controlling for age and log-assets.[3] The derivative effect is 0.45 on average, with a standard deviation of 0.07. In particular, wealthier and older households have a lower derivative (i.e., lower than 0.30), suggesting that they are relatively well insured against income shocks. In contrast, younger and less wealthy households have a higher derivative (i.e., higher than 0.50), suggesting less ability to insure.

In Figure 2, we show quantile derivatives of log-consumption with respect to log-income. In the left graph, we average quantile derivatives over the bottom tercile, while in the right graph we report an average over the top tercile. We see that these quantile derivative coefficients tend to be somewhat higher at the bottom of the consumption distribution (0.48 on average) than at the top (0.43 on average). The main difference between the two graphs concerns the younger and less wealthy households, for whom the derivative drops from 0.60 to 0.40 when moving from the bottom tercile to the top tercile.[4]

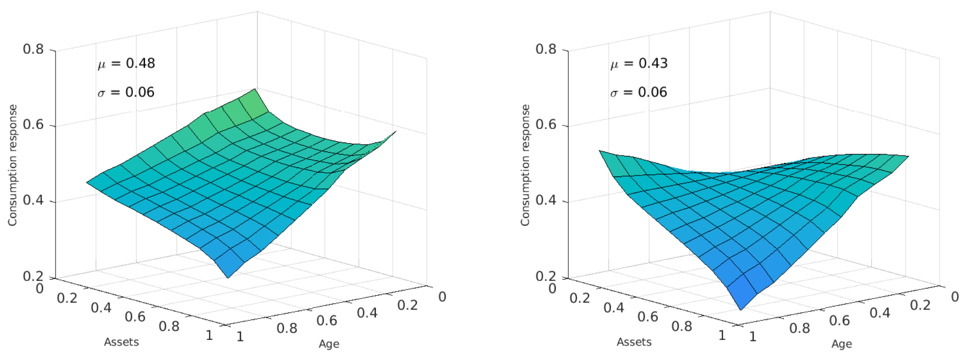
This evidence is suggestive of the presence of heterogeneity in consumption responses and insurance. However, there are several reasons why it may be incomplete and quantitatively inaccurate. Standard consumption models imply that income components with varying degrees of persistence

[3] Here and in the following we simply refer to log-income residuals in a regression on demographics and time indicators as "log-income", and we similarly refer to log-consumption residuals and log-assets residuals as "log-consumption" and "log-assets", respectively.

[4] In Appendix Figure A3 we show bootstrapped confidence bands corresponding to both Figures 1 and 2.

Figure 2: Quantile derivatives of log-consumption with respect to log-income

(a) Bottom Tercile (b) Top Tercile



Notes: The graphs show averages of the derivatives of quantile functions of log-consumption with respect to log-income, conditional on income, age and assets. In the left graph we report results for the bottom tercile (averaged over a fine grid of percentiles), in the right graph we report results for the top tercile. Estimates are based on quantile regressions of log-consumption on a second-order polynomial in log-income, age, and log-assets. The two horizontal axes show age and assets percentiles.

have a different impact on consumption. Hence, while in Figure 1 we report derivatives with respect to observed income, in a model where log-income is the sum of a persistent and a transitory component, economically-relevant consumption derivatives should be computed with respect to the latent components of income. To do so, a dynamic model with latent variables is needed. The heterogeneity suggested by Figure 2 is similarly ambiguous. Indeed, consumption quantiles are likely to reflect a combination of time-invariant household heterogeneity and time-varying shocks. Distinguishing the two requires estimating a dynamic panel data model that features latent heterogeneity explicitly. In the next two sections we describe such a model, and we explain how we estimate it using the PSID.

3 Overview of the model

3.1 Consumption behaviour

Our primary interest is to understand how shocks to income translate into consumption for different types of consumers. Consumers are allowed to differ along a number of dimensions, specifically according to their assets, the stage in their life cycle, observable characteristics, and unobserved heterogeneity. Our underlying framework is one where households act as single agents with access to a single risk-free asset. They receive income shocks each period and make consumption decisions subject to a period-to-period budget constraint. We assume all distributions are known to households, and there is no aggregate uncertainty.

In modeling the dynamic responses of consumption to earnings shocks, one strategy is to specify the functional form of the utility function and the distributions of the shocks, and to calibrate or estimate the model's parameters by comparing the model's predictions with the data, see Kaplan and Violante (2010) and references therein. Another strategy is to follow the partial insurance approach of Blundell, Pistaferri and Preston (2008) and linearize the Euler equation, with the help of the budget constraint. The approach we follow in this paper builds on the framework introduced in ABB. It differs from the earlier strategies as we directly estimate the consumption rule that comes from the optimization problem. In this approach the level of consumption is modeled as a function of beginning of period assets, income components, consumer characteristics and individual heterogeneity. The framework we develop here is a generalization of the main specification in ABB to allow for individual unobserved heterogeneity and a more flexible policy rule. The shape of the consumption function and its derivatives will depend on the distributions of beliefs about future incomes and characteristics. We are therefore able to document a rich set of derivative effects but, as our model does not separate the role of preferences from expectations, we cannot recover counterfactuals that involve a change in the income process.

In our approach, the income process is modeled using the framework of ABB which allows for nonlinear persistence. In this framework, log-income is decomposed into a predetermined life-cycle component and two latent stochastic factors that represent the level of persistent income and the level of transitory income. We consider an unbalanced panel of households, $i = 1, ..., N$, in which household i is observed T_i consecutive time periods. For any household i at time t we denote the persistent income component as η_{it} and assume it follows a nonlinear first-order Markov process. The transitory income component ε_{it} is assumed to be distributed independently across time and independent of the $\eta' s$. Log-income residuals are then $y_{it} = \eta_{it} + \varepsilon_{it}$. The details of the income specification are developed in the next subsection.

Given beginning-of-period-t assets a_{it}, and the realizations of the persistent and transitory income components η_{it} and ε_{it}, consumers make their consumption choices according to the policy rule

$$c_{it} = g_t\left(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, \nu_{it}\right), \quad i = 1, ..., N, \quad t = t_i, ..., t_i + T_i - 1, \tag{1}$$

where t_i denotes the period when i enters the panel, c_{it} is log-consumption for household i in period t, a_{it} is log-assets, age_{it} is the age of the household head in period t, and unobserved heterogeneity is given by the "fixed effect" ξ_i.[5] As mentioned above, both c_{it} and a_{it} are net of common effects of age and other demographics, and of time indicators. We also allow consumption choices to depend on transitory preference shocks ν_{it}, with arbitrary dimension.

Our main goal is to estimate the empirical consumption response parameters

$$\phi(age_{it}, a_{it}, \eta_{it}, \varepsilon_{it}, \xi_i) = \mathbb{E}_{\nu_{it}}\left[\frac{\partial g_t\left(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, \nu_{it}\right)}{\partial \eta}\right]. \tag{2}$$

[5] Below we will postulate that ξ_i follows a certain distribution (albeit a rather flexible one) conditional on cohort, education and income. An alternative description of ξ_i would thus be as a "correlated random effect".

Average derivative effects such as (2) can be identified without restricting the dimensionality of ν_{it}, see Matzkin (2013) and references therein.[6] Reporting features of estimates of the individual transmission parameters

$$\phi_{it} = \phi(age_{it}, a_{it}, \eta_{it}, \varepsilon_{it}, \xi_i)$$

in the PSID will shed light on how much variation there is in consumption responses and insurance, over the life cycle and as a function of assets and income. Importantly, the dependence of the consumption function on the latent type ξ_i will allow us to document individual heterogeneity in consumption responses. Exploring the relationship between ϕ_{it} and ξ_i is a main objective of this paper.

In order to estimate the consumption function g_t in (1), one needs to recover the persistent and transitory income components η_{it} and ε_{it}, and the time-invariant consumption type ξ_i, all of which are unobserved to the econometrician. For this purpose, we will estimate a dynamic model of income and consumption with latent variables, following ABB.

Asset accumulation. Estimation of the consumption function g_t requires taking a stand on the accumulation of assets. A simple case is when current assets only depend on lagged assets, income, and consumption, but not on the latent income components and heterogeneity separately. This would hold in a textbook asset accumulation rule with a constant risk-free interest rate, for example. Under the assumption that asset accumulation does not depend on the latent variables, one can estimate the consumption function consistently without having to model the assets process, in the spirit of partial likelihood estimation. We will use this approach in our main results. More generally, our approach can allow the latent income components and type heterogeneity to affect current assets, and we will report results based on such a specification as well, see Subsection 6.3.

Dispersion of consumption derivatives. Lastly, while we focus on recovering the average response parameters ϕ_{it}, the distribution of the consumption derivatives

$$\frac{\partial c_{it}}{\partial \eta} = \frac{\partial g_t\left(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, \nu_{it}\right)}{\partial \eta},$$

conditional on $(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i)$, is generally not identified unless ν_{it} is scalar and has a monotone effect on g_t. Yet, using an insight from Hoderlein and Mammen (2007), one can compute a lower bound on the variance of the consumption derivatives $\frac{\partial c_{it}}{\partial \eta}$, even though the variance itself is not identified. We make this point formally in Appendix C, and we will report empirical estimates of bounds on variances as a complement to our main average coefficients.

[6] However, in our setting, some of the arguments of the structural function g_t (i.e., η_{it}, ε_{it}, and ξ_i) are latent. Identification of average derivatives thus requires showing that the distribution of $(c_{it}, a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i)$ is identified.

3.2 Income and consumption

Our modeling of the income process closely follows ABB, with the main difference that we extend the model to an unbalanced panel. Specifically, let y_{it} be the log-disposable income of household i in year t, net of common effects of age and other demographics, and time indicators. We specify the following persistent-transitory model

$$y_{it} = \eta_{it} + \varepsilon_{it}, \qquad i = 1, ..., N, \ t = t_i, ..., t_i + T_i - 1, \tag{3}$$

where the persistent and transitory components η_{it} and ε_{it}, respectively, are zero-mean continuous latent variables given age.

We model the processes η_{it} and ε_{it} using their quantile representations. Let $Q_A(B, v)$ be a generic notation for the conditional quantile of A given B, evaluated at the percentile v in the unit interval. The quantile representation of A given B implies that $A = Q_A(B, V)$, where V is standard uniform independent of B.[7]

The persistent income component η_{it} follows a nonlinear first-order Markov process with age-specific transitions; that is,[8]

$$\eta_{it} = Q_\eta(\eta_{i,t-1}, age_{it}, u^\eta_{it}), \qquad \left(u^\eta_{it} \mid \eta_{i,t-1}, age_{it}\right) \sim iid \, \text{Uniform}\,(0,1)\,, \qquad t > t_i. \tag{4}$$

In order to model entry in the panel, we let the initial persistent latent component η_{i,t_i} depend on years of education and birth cohort of the household head, and on age at entry in the sample:

$$\eta_{i,t_i} = Q_{\eta_1}(cohort_i, educ_i, age_{i,t_i}, u^{\eta_1}_i), \quad \left(u^{\eta_1}_i \mid cohort_i, educ_i, age_{i,t_i}\right) \sim iid \, \text{Uniform}\,(0,1)\,. \tag{5}$$

In turn, the transitory component ε_{it} is assumed to be independent over time and independent of η_{is} for all s with an age-specific distribution,

$$\varepsilon_{it} = Q_\varepsilon(age_{it}, u^\varepsilon_{it}), \quad (u^\varepsilon_{it} \mid age_{it}) \sim iid \, \text{Uniform}\,(0,1)\,. \tag{6}$$

Note that the income process is common across households. In this paper we do not attempt to model latent time-invariant heterogeneity in the income process beyond heterogeneity in initial conditions. However, we allow for an unobserved type that affects consumption and may be correlated with income.

Turning to consumption, we let the unobserved heterogeneity variable ξ_i be correlated with birth cohort, education, and income; that is, we specify

$$\xi_i = Q_\xi(cohort_i, educ_i, income_i, u^\xi_i), \quad \left(u^\xi_i \mid cohort_i, educ_i, income_i\right) \sim iid \, \text{Uniform}\,(0,1)\,. \tag{7}$$

[7] For example, $Q_A(B, 0.50)$ is the conditional median of A given B, and $Q_A(B, 0.90)$ is the conditional 90th percentile of A given B. The fact that $A = Q_A(B, V)$, where V is standard uniform independent of B, is referred to as the Skorohod representation in the literature, see, e.g., Chernozhukov and Hansen (2005).

[8] In our sequential model, we assume that $u^\eta_{it} \mid \eta^{t-1}_i, age^t_i$ is standard uniform, where η^{t-1}_i and age^t_i denote sequences of lags of η and age. For conciseness we leave the full conditioning implicit in the notation.

Here $income_i$ is a measure of the household's "normal" income. In our baseline specification we will take $income_i$ to be the average log-income over the period of observation. In addition, note that the age at entry in the panel does not affect ξ_i given cohort, education, and income. Hence, ξ_i is a time-invariant household characteristic that does not depend on when the household starts being recorded in the PSID, whereas the value of the initial persistent latent component in (5) depends on the stage of the life cycle the household was at when she entered the panel.

We then specify the log-consumption function as

$$c_{it} = Q_c(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, u_{it}^c), \qquad (u_{it}^c \mid a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i) \sim iid\,\text{Uniform}\,(0,1)\,. \tag{8}$$

For the purpose of documenting consumption responses, it is important to know under which conditions estimating (8) allows one to learn about features of the household's consumption function g_t in (1). Suppose that the transitory preference shocks ν_{it} in (1) are i.i.d., independent of past assets and income components, age, and latent type ξ_i. If in addition ν_{it} are scalar and have a monotone impact on the consumption function g_t, then g_t will be identified based on (8), up to a nonlinear transformation of its last argument. Moreover, when the economic primitives are such that ν_{it} are multidimensional or have a non-monotone impact on consumption, the conditional mean function of log-consumption implied by (1) will still be identified based on (8), even though the individual consumption function g_t will not be identified in general. Indeed, under our assumptions we have

$$\phi_{it} = \mathbb{E}_{\nu_{it}}\left[\frac{\partial g_t\left(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, \nu_{it}\right)}{\partial \eta}\right] = \mathbb{E}_{u_{it}^c}\left[\frac{\partial Q_c(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, u_{it}^c)}{\partial \eta}\right].$$

In other words, using quantile methods to flexibly estimate the function Q_c in (8), we will be able to consistently estimate our main target parameters, which are the average derivative quantities ϕ_{it}.

Note that, under mild assumptions, the consumption response parameters in (2) are equal to the derivatives of the conditional mean of consumption given the state variables,

$$\phi_{it} = \frac{\partial}{\partial \eta}\mathbb{E}\big[c_{it} \,|\, a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i\big].$$

However, η_{it}, ε_{it} and ξ_i are unobserved in the data, so it is not enough to model the conditional mean $\mathbb{E}\left[c_{it} \,|\, a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i\right]$ to recover our key parameters ϕ_{it}. ABB specified a nonlinear mean model with separable heterogeneity. A concern with their specification is that it might be too restrictive as a model of the conditional distribution of c_{it} given $(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i)$. In contrast, in this paper we employ a quantile specification to achieve a more flexible modeling of that conditional distribution.

In our baseline model where assets do not depend on the latent variables η_{it}, ε_{it}, and ξ_i directly, a specification of the assets process is not needed. However, assuming that asset accumulation does not depend on the latent variables might be restrictive if, for example, assets returns are heterogeneous and the assets process is not independent of ξ_i. For this reason, we will also estimate a model where we specify a reduced-form assets process as

$$a_{i,t+1} = Q_a(a_{it}, \eta_{it}, \varepsilon_{it}, c_{it}, age_{it}, \xi_i, u^a_{i,t+1}), \qquad \left(u^a_{i,t+1} \mid a_{it}, \eta_{it}, \varepsilon_{it}, c_{it}, age_{it}, \xi_i\right) \sim iid\, \text{Uniform}\,(0,1)\,, \tag{9}$$

where in addition $u^a_{i,t+1}$ and $u^c_{i,t+1}$ are independent. In this model, we will specify initial assets holdings as

$$\begin{aligned} a_{i,t_i} &= Q_{a_1}(\eta_{i,t_i}, age_{i,t_i}, cohort_i, educ_i, \xi_i, u^{a_1}_{i,t_i}), \\ \left(u^{a_1}_{i,t_i} \mid \eta_{i,t_i}, age_{i,t_i}, cohort_i, educ_i, \xi_i\right) &\sim iid\, \text{Uniform}\,(0,1)\,. \end{aligned} \tag{10}$$

To summarize the framework laid out in this section, we have described a model with three latent components. The time-invariant type ξ_i is intended to capture household pre-sample-period observed and unobserved heterogeneity. The other two latent components enter the income process. The persistent component η_{it} captures household heterogeneity that results from the accumulation of persistent shocks over time. Finally, independent transitory shocks ε_{it} with an age-specific distribution combine with the persistent component and its profile to produce observed labor income.

The presence of the latent type ξ_i as an argument of the consumption function may potentially reflect several mechanisms. For example, ξ_i may indicate preference or discounting heterogeneity. Alternatively, it may capture heterogeneity in returns to assets. Yet another possible interpretation of ξ_i is as additional resources that are available to the household but not observed in the data, such as consumption insurance provided by parents. We will examine the plausibility of these various mechanisms empirically in Section 6. We let the latent type ξ_i correlate with income through the conditioning on $income_i$ in (7). In addition, although here we will use our most parsimonious specification as a baseline when reporting results, in an extension we will let ξ_i enter asset accumulation directly, see equations (9)-(10).

The model thus features two levels of heterogeneity: (a) demographics and time effects, which we partial out linearly in an initial step, and (b) the latent type ξ_i, which we include as part of our nonlinear model. We will study the possibility of an additional nonlinear impact of demographic heterogeneity in Subsection 7.4.

4 Estimation methodology and implementation

To specify and estimate the model, we closely follow ABB, with some differences. While in this section we focus on estimation and practical implementation, we note that given the similarity of the model's structure to that of ABB, nonparametric identification can be shown using the arguments they provide. Those arguments rely on insights from the literature on nonparametric instrumental variable models and nonlinear models with latent variables (see, among others, Newey and Powell, 2003, Hu and Schennach, 2008, and Wilhelm, 2015).

4.1 Specification

Following ABB, we model all conditional quantile functions using linear quantile specifications at a grid of percentiles. As an example, we model the conditional quantile function of the persistent latent component of income in (4) as

$$Q_\eta(\eta_{i,t-1}, age_{it}, \tau) = \sum_{k=0}^{K} a_k^\eta(\tau)\varphi_k(\eta_{i,t-1}, age_{it}), \tag{11}$$

where φ_k are low-order products of Hermite polynomials in age and the lagged persistent latent component of income, and $a_k^\eta(\tau)$ are piecewise-linear polynomial functions of τ. In practice we use a grid of 11 equidistant percentiles. In addition, following ABB we augment the model by specifying $a_k^\eta(\tau)$ using an exponential modeling of the tails of the intercept coefficients. We use similar specifications for all the other equations (6)-(10). We provide details in Appendix A.

A difference with ABB is that, while they modeled the nonlinear mean of log-consumption and assumed separable errors, here we flexibly estimate the entire conditional quantile function of log-consumption in (8) without imposing separability between u_{it}^c and the other determinants of consumption. This is important for estimating the average consumption derivative parameters ϕ_{it} in the presence of latent variables, in a way which is robust to the presence of non-separabilities implied by the economic model.

Fully nonlinear estimation of consumption quantiles has implications for the econometric specification of the model, given that the type ξ_i is a latent variable. Indeed, note that ξ_i and the conditional quantile function Q_c are not separately nonparametrically identified, since it is always possible to take a transformation of ξ_i, and to undo it in Q_c.[9] In a general quantile model such as (8), we impose the following restriction:

$$\mathbb{E}[c_{it} \,|\, a_{it} = \overline{a}, \eta_{it} = \overline{\eta}, \varepsilon_{it} = \overline{\varepsilon}, age_{it} = \overline{age}, \xi_i = \xi] = \int_0^1 Q_c(\overline{a}, \overline{\eta}, \overline{\varepsilon}, \overline{age}, \xi, \tau)d\tau = \xi, \quad \text{for all } \xi, \tag{12}$$

where $\overline{a}, \overline{\eta}, \overline{\varepsilon}, \overline{age}$ are some fixed reference values for log-assets, persistent and transitory income components, and age. Imposing this restriction resolves the indeterminacy.[10] In this way, ξ_i is measured in consumption units, which is meaningful when studying its distribution. In the implementation we set $\overline{a}, \overline{\eta}, \overline{age}$ to be the unconditional sample averages of log-assets, log-income and age, respectively, and we set $\overline{\varepsilon}$ to zero.

4.2 Estimation

To estimate the model we adapt the multi-step approach proposed by ABB to our setting. *In a first step*, we compute regression residuals of log-income, log-consumption, and log-assets on a set of

[9] For example, for any invertible function ψ we can write $Q_c(\xi) = (Q_c \circ \psi^{-1})(\psi(\xi))$.

[10] If Q_c is linear, (12) selects a form of the fixed effect that is inclusive of all the intercept components. See Hu and Schennach (2008) and the subsequent literature for related assumptions.

controls, which includes demographics and time indicators, see Section 2 for the full list of controls. This allows us to construct the residualized variables y_{it}, c_{it}, and a_{it}.

In a second step, we estimate the income process. To this end, we use a stochastic EM algorithm (Nielsen, 2000), which alternates between draws of the latent income components η_{it} and ε_{it}, and parameter updates based on the latent draws. The updates are performed using quantile regressions, similarly to ABB. For example, to estimate the parameters $a_k^{\eta}(\tau)$ at a grid of τ values in (11), we run multiple quantile regressions.[11]

To generate the latent draws, we depart from ABB who relied on Metropolis Hastings, and use a Sequential Monte Carlo sampling method. We describe this method in the next subsection. The reason for using a different sampler compared to ABB is numerical stability. Indeed, the performance of Metropolis Hastings tends to deteriorate as the length of the panel and the number of households increase. In the longer and larger panel sample we use in this paper, Sequential Monte Carlo methods tend to be more robust to numerical issues such as initialization and seeding than Metropolis Hastings in our experience. A feature of Sequential Monte Carlo methods is that they take advantage of the Markovian structure of the model to improve performance relative to naive importance sampling.

In a third step, we estimate the consumption function, for given values of the parameters governing the income process. We perform this step using a similar strategy to the one we use for income. In this case also, we depart from ABB in the sampling step of the stochastic EM algorithm. However, the presence of the latent type ξ_i further complicates implementation, since one needs to repeatedly draw ξ_i together with the sequences of persistent and transitory components. To generate valid draws, we rely on the pseudo-marginal Markov Chain Monte Carlo algorithm proposed by Andrieu, Doucet and Hollenstein (2010), which itself makes use of Sequential Monte Carlo sampling. We describe our implementation in the next subsection.

Quantile monotonicity. Given our quantile modeling, the parameters satisfy monotonicity restrictions (e.g., Chernozhukov, Fernández-Val and Galichon, 2010). For example, in (11) the mapping $\tau \mapsto a_k^{\eta}(\tau)\varphi_k(\eta_{i,t-1}, age_{it})$ is non-decreasing. In practice we do not enforce monotonicity in estimation. However, in each expectation step of the stochastic EM algorithm we draw from the likelihood implied by the estimated parameters. This ensures that we obtain posterior draws from a valid distribution of η's and ξ's, irrespective of the lack of monotonicity of the quantile parameter estimates. To provide intuition in a simple setup, note that to draw η_{it} according to model (11) one can compute, as in Machado and Mata (2005),

$$\widetilde{\eta}_{it} = \sum_{k=0}^{K} \widehat{a}_k^{\eta}(u_{it}^{\eta})\varphi_k(\widetilde{\eta}_{i,t-1}, age_{it}) \text{ for } t > t_i, \quad \widetilde{\eta}_{i,t_i} = \eta_{i,t_i},$$

[11] Before every update step, we compute an empirical counterpart of the left-hand side in (12) by regressing log-consumption on the draws of η, ε, ξ, log-assets, and age, and we set ξ_i to be the corresponding predicted value. See Appendix A for details.

where u_{it}^{η} are i.i.d. standard uniform. Although the estimates $\widehat{a}_k^{\eta}(\tau)$ may not satisfy monotonicity restrictions, this approach produces $\widetilde{\eta}_{it}$ draws from a valid distribution function. In our setting we use this strategy to generate *posterior* draws of η's and ξ's, see Appendix A for details.

Asymptotic distribution and inference. Under the assumption that the parametric model is correctly specified,[12] averages of parameter draws are consistent and asymptotically normal with an asymptotic variance that can be estimated by bootstrap or analytical approximations, see Arellano and Bonhomme (2016) and ABB for details. We will report confidence bands computed using two versions of the bootstrap: a parametric bootstrap that relies on the model's structure for simulations, and a nonparametric bootstrap clustered at the household level.

4.3 Computational sampling techniques

Here we describe how we draw latent variables in every step of the stochastic EM algorithm. We present, in turn, the methods we use for the latent income components $\eta_{it}, \varepsilon_{it}$, and for the latent consumption type ξ_i. In practice we run these simulation steps in parallel across households, which makes it easy to estimate the model on an unbalanced panel.

Income components: Sequential Monte Carlo. Estimating the income process requires solving a nonlinear filtering problem, where $\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}$ are latent variables. To draw from their posterior distribution given the income data we use a Sequential Monte Carlo (SMC) approach, see Creal (2012) and Kantas, Doucet, Singh and Maciejowski (2009) for surveys.

To describe the SMC approach, we focus on the problem of sampling $\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}$ for a single household i from the posterior distribution $f(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}|y_{i,t_i}, ..., y_{i,t_i+T_i-1})$. In practice we sample in parallel across households. With importance sampling, one might first sample directly from some proposal distribution $\pi(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1})$, and then re-sample using importance sampling weights

$$w_i \propto \frac{f(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}|y_{i,t_i}, ..., y_{i,t_i+T_i-1})}{\pi(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1})},$$

where $\propto$ is a proportionality symbol. However, finding a suitable proposal distribution in our flexible nonlinear model is challenging. Instead, we try and generate draws (also called "particles") sequentially.

At $t = t_i$, we initialize S particles $\eta_{i,t_i}^{(s)}$ from a suitable proposal distribution $\pi(\eta_{i,t_i})$. Re-sampling with weights

$$w_{i,t_i}^{(s)} \propto \frac{f\left(\eta_{i,t_i}^{(s)}|y_{i,t_i}\right)}{\pi\left(\eta_{i,t_i}^{(s)}\right)}$$

[12] One may also view the parametric model as a sieve approximation to a nonparametric distribution, where the size of the grid of τ values, and hence the number of parameters, would grow with the sample size at an appropriate rate. The theoretical justification we mention here is for a well-specified parametric model.

gives S particles approximately distributed according to $f(\eta_{i,t_i}|y_{i,t_i})$.

At $t = t_i + 1$, we now aim to approximate

$$f(\eta_{i,t_i}, \eta_{i,t_i+1}|y_{i,t_i}, y_{i,t_i+1}) = \frac{f(y_{i,t_i+1}|\eta_{i,t_i+1}) f(\eta_{i,t_i+1}|\eta_{i,t_i})}{f(y_{i,t_i+1}|y_{i,t_i})} f(\eta_{i,t_i}|y_{i,t_i}).$$

Since we already have S particles approximately distributed according to $f(\eta_{i,t_i}|y_{i,t_i})$, we can simply use a second proposal distribution $\pi(\eta_{i,t_i+1}|\eta_{i,t_i})$ to extend these existing particles. Re-sampling with weights

$$w^{(s)}_{i,t_i+1} \propto \frac{f\left(\eta^{(s)}_{i,t_i+1}|y_{i,t_i+1}, \eta^{(s)}_{i,t_i}\right)}{\pi\left(\eta^{(s)}_{i,t_i+1}|\eta^{(s)}_{i,t_i}\right)}$$

gives S particles approximately distributed according to $f(\eta_{i,t_i}, \eta_{i,t_i+1}|y_{i,t_i}, y_{i,t_i+1})$. The process continues until we obtain S particles approximately distributed as $f(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}|y_{i,t_i}, ..., y_{i,t_i+T_i-1})$.[13]

The choice of proposal distributions π is important for numerical performance. We found that a simple generalization of a linear permanent-transitory earnings model with Gaussian errors performed well. Specifically, we postulate the following model:

$$y_{it} = \eta_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim iid\ \mathcal{N}(0, \sigma^2_{\varepsilon}), \tag{13}$$

$$\eta_{it} = m(\eta_{i,t-1}, age_{it}) + v^{\eta}_{it}, \quad v^{\eta}_{it} \sim iid\ \mathcal{N}(0, \sigma^2_{v}), \tag{14}$$

where ε_{it} and v^{η}_{it} are independent at all lags, and m is a Hermite polynomial. We re-estimate this model at each iteration of the stochastic EM algorithm, and then set $\pi(\eta_{it}|\eta_{i,t-1})$ to be the posterior distribution based on it. We provide details about the implementation of the SMC sampler in Appendix A. In addition, we provide a comparison of the SMC and Metropolis Hastings sampling methods in the ABB sample in Appendix B. We find that, while our SMC algorithm recovers similar estimates of nonlinear persistence to those reported in ABB, the SMC method is less sensitive to numerical instability than Metropolis Hastings.

Unobserved type in consumption: Particle Markov Chain Monte Carlo. In order to incorporate unobserved heterogeneity ξ_i, we embed the SMC sampler into a Particle Markov Chain Monte Carlo (PMCMC) algorithm, following Andrieu, Doucet and Hollenstein (2010). We use this method to estimate the parameters of the consumption process, after having estimated the parameters of the income process.

To outline the PMCMC approach, suppose we wish to sample $\xi_i, \eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1}$ from the posterior distribution $f(\xi_i, \eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1} \,|\, w_{i,t_i}, ..., w_{i,t_i+T_i-1})$, where $w_{it} = (y_{it}, c_{it}, a_{it})$ is a vector of household i's observed income, consumption and assets at time t. In the PMCMC approach, SMC algorithms are used to generate efficient proposals to be used within a Metropolis Hastings algorithms.

[13] In practice, re-sampling at every time increment can result in degeneracy among the available particles. For this reason, we instead use an adaptive rule which avoids degeneracy (see Creal, 2012).

An important feature of these methods is that they only rely upon the availability of unbiased estimates of the marginal likelihood $f(w_{i,t_i}, ..., w_{i,t_i+T_i-1}|\xi_i)$, which are readily available as a by-product of the SMC algorithm. The use of unbiased estimates of a target distribution within a Metropolis Hastings algorithm can be viewed more generally as an example of a pseudo-marginal approach in which the resulting algorithms can be presented as *bona fide* Metropolis Hastings samplers whose marginal distribution is the target distribution of interest. We provide details about the implementation of the PMCMC sampler in Appendix A.

5 Main results

In this section we present the main empirical results on income and consumption, obtained using our baseline nonlinear model with unobserved heterogeneity.

5.1 Income persistence

We start by reporting the results on nonlinear income persistence. In the left graph of Figure 3 we show the derivative of the conditional quantile function of log-income given lagged log-income and age, with respect to lagged log-income. Formally, we compute an estimate of

$$\rho_y(y, age, \tau) = \frac{\partial Q_y(y, age, \tau)}{\partial y}, \qquad \text{for } \tau \in (0, 1),$$

where Q_y is the conditional quantile function of log-income given lagged log-income and age, and average it with respect to age. The nonlinear persistence parameters $\rho_y(y, age, \tau)$ can be interpreted as heterogeneous autoregressive coefficients, which may depend on both the income level y and the income shock τ.[14] We plot the derivative as a function of lagged log-income (which we refer to as "initial income") and of the innovation in the quantile model (which we refer to as "income shock").

The results show that most households, for most shocks, have current disposable incomes that are quite persistent, with a derivative coefficient that is above 0.80. However, households with low initial income and high income shocks have incomes that are substantially less persistent, with a coefficient as low as 0.40. Likewise, persistence is also low for households with high initial income and low income shocks, with a coefficient of a similar magnitude. These nonlinear persistence estimates are closely related to those found by ABB on a smaller balanced sample drawn from the earlier pre-recession years of the PSID.

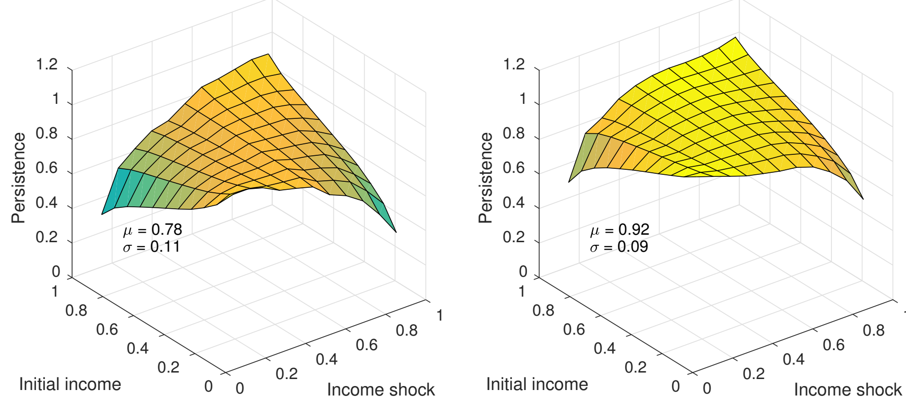
In the right graph of Figure 3 we show nonlinear income persistence, but now for the persistent latent component η_{it}. That is, we show

$$\rho_\eta(\eta, age, \tau) = \frac{\partial Q_\eta(\eta, age, \tau)}{\partial \eta}, \qquad \text{for } \tau \in (0, 1),$$

[14] In Appendix Figure A4 we show a different projection of the same three-dimensional surfaces, to ease visualization.

Figure 3: Nonlinear income persistence

(a) Log-income (b) Persistent component



Notes: PSID, 2005-2017 sample, disposable income, dual earners. The left graph shows quantile derivatives of log-income with respect to lagged log-income, $\rho_y(y, age, \tau)$ averaged over y. The right graph shows quantile derivatives of the persistent latent component η_{it} with respect to η_{it-1}, $\rho_\eta(\eta, age, \tau)$ averaged over η, in a model estimated using sequential Monte Carlo with a stochastic EM algorithm. In this case, the two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

where Q_η is the conditional quantile function of η_{it} given $\eta_{i,t-1}$ and age, see (4). We plot the derivative as a function of $\eta_{i,t-1}$ ("initial income") and the innovation in the quantile model ("income shock").[15] We see that average persistence is higher than for the case of log disposable income — it is 0.92 in the right graph, versus 0.78 in the left graph — due to the removal of the transitory income component. For households with high values of initial persistent income and high shocks, persistence is close to unity, and similarly for households with low initial persistent income and low shocks.[16] The nonlinear pattern for the persistent latent component η_{it} is qualitatively similar to the one for log-income, although it is quantitatively less pronounced.

These nonlinear persistence patterns are rather precisely estimated, see the parametric bootstrap 95% confidence bands in Appendix Figure A5 and the nonparametric bootstrap 95% confidence bands in Appendix Figure A6. In addition, comparing Figure 3 to Appendix Figure A1, we see that, while nonlinearities are somewhat more salient in our larger and more recent sample compared to the balanced sample used in ABB, the persistence patterns in both cases are comparable.[17]

[15]To produce the plot, we use posterior draws computed from the model. We proceed similarly when plotting all subsequent results involving latent variables.

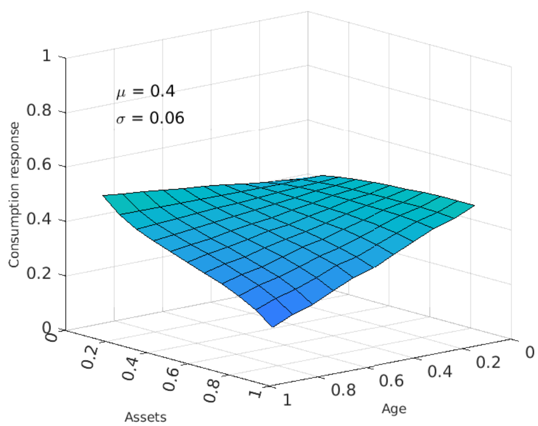
[16]Note that it is possible for the nonlinear income persistence measure to exceed one.

[17]In Figure 3 we average the persistence measure across age values. In contrast, the main nonlinear persistence figures in ABB are evaluated at a reference age value. The analog of Figure 3(a) in ABB is Figure S3 in their supplemental appendix.

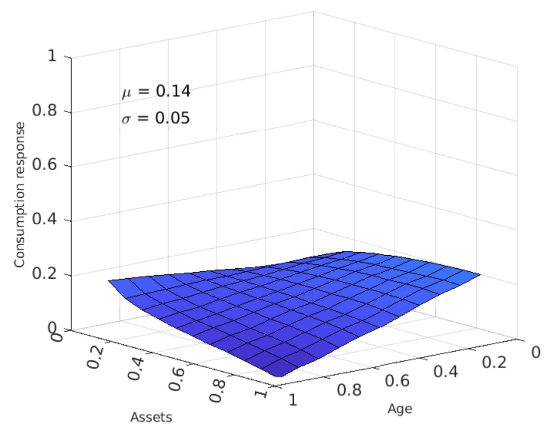
Figure 4: Average consumption responses

A. Models without filtering

(a) No heterogeneity

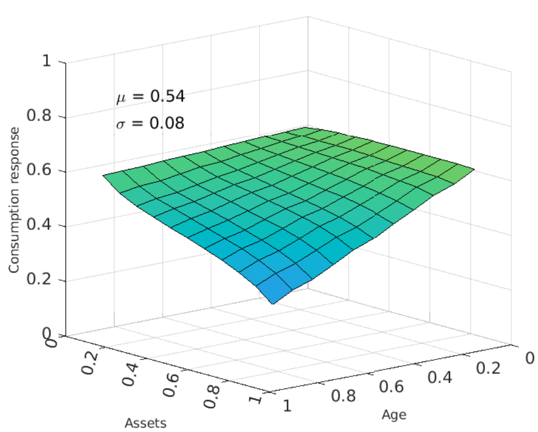


(b) Heterogeneity

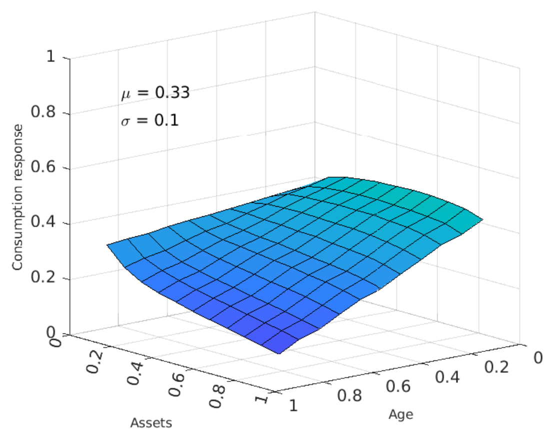


B. Models with filtering

(c) No heterogeneity



(d) Heterogeneity



Notes: PSID, 2005-2017 sample, dual earners. The graphs show the average derivative of log-consumption with respect to log-income (in the top panel) and the persistent latent component η_{it} (in the bottom panel). The left graphs correspond to a model without unobserved heterogeneity ξ_i in consumption, whereas the right graphs correspond to a model with unobserved heterogeneity ξ_i. The two horizontal axes show age and assets percentiles, respectively.

5.2 Average consumption responses to income shocks

The main goal of the paper is to study heterogeneity in consumption responses to unexpected changes in income. That is, the way income shocks are transmitted into consumption which underpins the degree of "partial insurance" achieved by the household. In this subsection, and the next, we document several key features of household partial insurance, which we measure using the household-and-time-varying transmission coefficients

$$\phi_{it} = \phi(age_{it}, a_{it}, \eta_{it}, \varepsilon_{it}, \xi_i)$$

given by the average derivative effects (2) introduced in Subsection 3.1. The transmission coefficient ϕ_{it} quantifies the change in consumption induced by an exogenous marginal change in the persistent latent component of income.

In Figure 4 we start by showing how the mean of the estimated transmission parameters ϕ_{it} varies with assets levels and over the life cycle. We compare four specifications. The "models without filtering" in the upper panel correspond to specifications without transitory component ε_{it}, so the derivative on the right-hand side of (2) is taken with respect to log current disposable income y_{it} instead of the persistent latent component η_{it}. The "models with filtering" in the lower panel allow for a separate role of η_{it} and ε_{it}. For both models with and without filtering, we distinguish two specifications with and without unobserved heterogeneity ξ_i, in the left and right columns, respectively.

Figure 4 shows that all specifications agree quite well qualitatively. In particular, the association between consumption and income or its persistent latent component is weaker for older and wealthier households. At the same time, there are important quantitative differences between the four specifications. We find that allowing for unobserved heterogeneity ξ_i tends to dampen the consumption impacts of income shocks, the difference being particularly noticeable for the models without filtering where average responses decrease from 0.40 to 0.14. The impact of heterogeneity can be explained by the fact that, according to our estimates, ξ_i is positively correlated with income, see Section 6. In contrast, allowing for a transitory income component tends to increase consumption responses to income shocks, as is typically the case in estimates that correct for measurement error bias. As a result, in our main model with unobserved heterogeneity and a transitory component, the lower right hand graph shows an estimated average response parameter of 0.33. There are strong differences by assets and age too, with the estimated average transmission coefficient dropping toward 0.10 for older and wealthier households, while for younger households the estimated mean transmission rises to around 0.40.[18]

[18]The relative magnitudes of the nonlinear estimates in Figure 4 are reminiscent of the situation in a linear model with a mismeasured persistent regressor and fixed effects, where the (positive) fixed effects bias and the (negative) measurement error bias tend to offset each other, while only accounting for fixed effects exacerbates the measurement error bias (Griliches and Hausman, 1986).

Comparison with ABB. It is informative to compare the average responses in Figure 4 to the results obtained by ABB. In a model without heterogeneity but with a transitory component, ABB found an average transmission coefficient of 0.38. This is lower than the responses in Figure 4 (c), which are 0.54 on average.[19] As we previously noted, the period of observation, the sample of households, and the income measure used in ABB all differ from the ones we focus on in the current paper. In particular, ABB focus on labor income as opposed to disposable income. Our estimates of consumption responses based on labor income are substantially lower than the responses based on disposable income shown in Figure 4, see Section 7.

Test of homogeneity. By comparing average response coefficients in models with and without household-specific heterogeneity, one can assess whether the data supports an homogeneous model without latent types. To do this, in Appendix Figure A7 we report confidence bands based on the nonparametric bootstrap clustered at the household level for the average responses depicted in the lower panel of Figure 4. We find a 95% confidence interval for the mean across these responses of $[0.50, 0.59]$ in the model without heterogeneity, and of $[0.21, 0.44]$ in the model with heterogeneity. The fact that the two confidence intervals do not overlap represents a formal rejection, at the 5% level, of the null hypothesis of homogeneity. The same conclusion holds when we use the parametric bootstrap to produce confidence intervals, see Appendix Figure A8.

5.3 Heterogeneity in consumption responses to income shocks

We have already seen that the introduction of unobserved heterogeneity has a systematic effect on the estimated average response of consumption to changes in income. We hypothesize that there are also systematic differences in responses across consumers that differ according to unobserved heterogeneity. To examine this, we study how consumption responses differ among households that are at the same point in the life cycle and have the same level of assets. For this purpose, we show how the transmission coefficients ϕ_{it} vary by quantiles of the unobserved type ξ_i, in addition to showing how they vary with assets levels and over the life cycle.

In Figure 5 we show transmission parameters as a function of assets and age, for five different percentiles of ξ_i, and we also show the average across ξ_i values. The results show clear evidence of household heterogeneity in consumption responses to income shocks. Consider the 10th percentile of ξ_i, in the top left graph. For these "low consumption type" households, average transmission is 0.36, yet the magnitude of the transmission coefficient varies substantially with age and assets. Indeed, while younger and less wealthy households have transmission coefficients of close to 0.60, the coefficient is

[19] The consumption responses in a model without heterogeneity in ABB can be found in their Figure 5(c). In addition, ABB also reported average responses based on a model with unobserved heterogeneity, albeit using a different specification for the consumption rule. They found lower responses in this case, amounting to 0.32 on average, see Figure S24(b) in the supplementary appendix of ABB.

as low as 0.10 for older and wealthier households. This pattern is qualitatively consistent with the implications of a standard life-cycle model of consumption and saving behavior in which persistent shocks are harder to self-insure for young consumers and for those consumers with low levels of net assets.

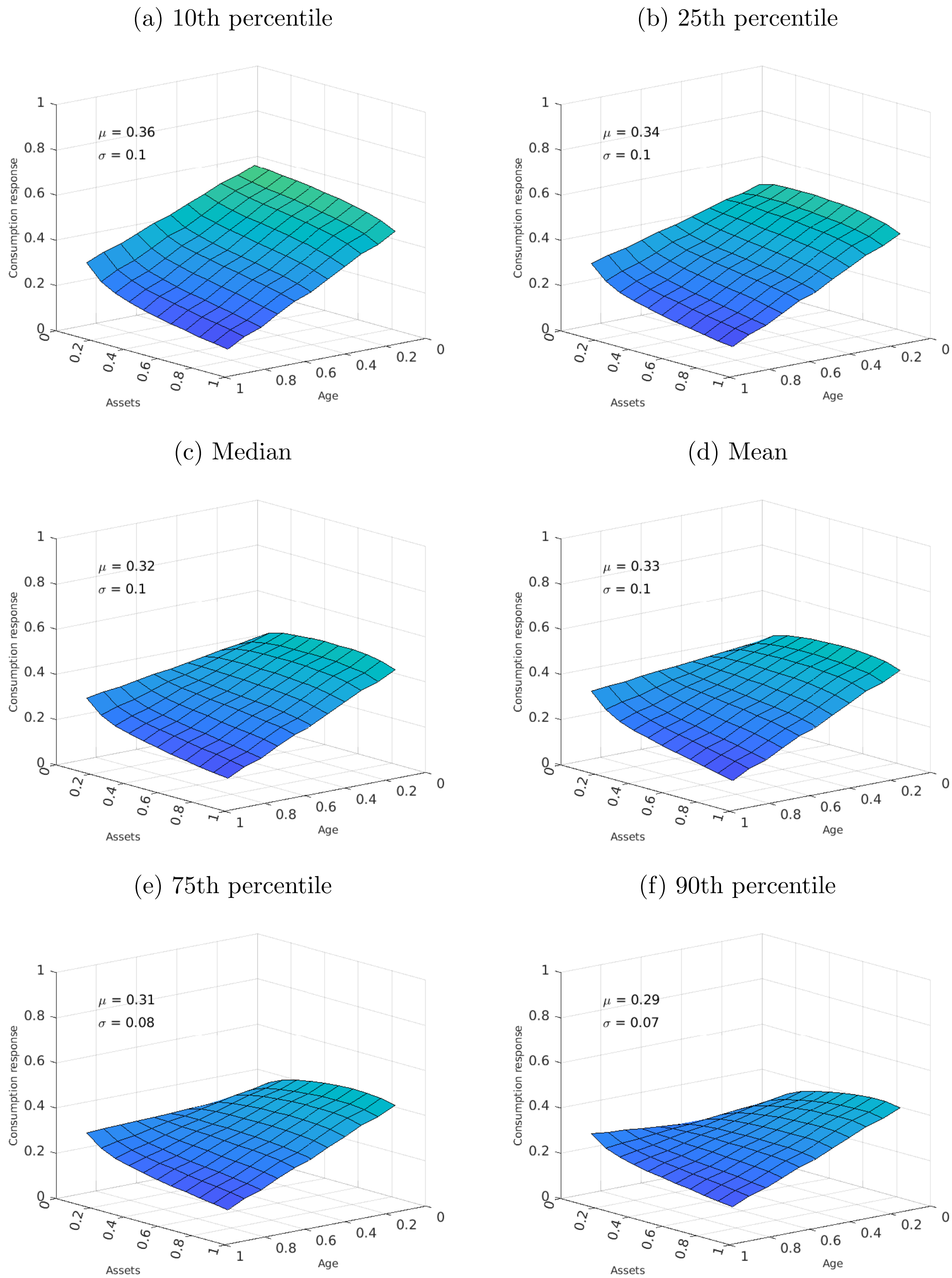
This "life-cycle consistent" pattern of responses is maintained through to the median type, albeit less pronounced. As we move to the higher consumer types, a pattern that is much less sensitive to assets and age appears. Consider the 90th percentile of ξ_i, in the bottom right graph of Figure 5. For these high-type households, the transmission coefficients are 0.29 on average, hence lower than the coefficients of the low-type households. In addition, the variation of the transmission coefficients with assets and age is less pronounced than for the low types. Indeed, while coefficients are approximately 0.15 for the older and wealthier households, the young and less wealthy households have coefficients that do not exceed 0.40. These patterns for the high-types are less in accordance with the forces at play in conventional life-cycle models of the individual household.

In order to provide measures of uncertainty associated with our main results, we rely on the bootstrap. We report results based on a parametric bootstrap approach, where we use the model to simulate bootstrapped data sets given parameter estimates. In Appendix Figure A9 we report pointwise 95% bands for the transmission parameters of Figure 5. We see that our estimates are rather precise. As a complement to the parametric bootstrap, in Appendix Figure A10 we report pointwise 95% bands based on the nonparametric bootstrap clustered at the household level. Precision is lower in this case, which is is not surprising, since, relative to the clustered nonparametric bootstrap, the parametric bootstrap exploit our modeling of the time-series dependence.

As a summary measure of the salient dimensions of heterogeneity that we find, in the top panel of Table 2 we report estimates of average transmission parameters for various categories of households: high and low types, corresponding to ξ_i being at the 90th percentile or the 10th percentile, young/low assets for whom age and assets are below the median, and old/high assets for whom age and assets are above the median. In the bottom panel we repeat the exercise for high types corresponding to ξ_i being at the 75th percentile and low types corresponding to ξ_i being at the 25th percentile. Alongside point estimates, we report 95% confidence intervals based on the parametric bootstrap.

We find that, while for high consumption types at the 90th percentile the transmission of income shocks is only 0.09 higher for young/low assets households and insignificant at conventional levels, for low types at the 10th percentile the average response coefficient is 0.27 higher for the young and low assets and significant at the 5% level. This supports our main conclusion regarding the fact that the behavior of low types appears to be consistent with a standard life-cycle model of consumption and saving, yet the behavior of high types appears less consistent with the mechanisms of the model. In addition, the cross-type difference $0.09 - 0.27 = -0.18$ between these two estimates, which is akin to

Figure 5: Heterogeneity in consumption responses



Notes: See the notes to Figure 4. Here we report the results by percentiles of heterogeneity ξ_i in consumption.

Table 2: Summarizing heterogeneity across types, parametric bootstrap

A. 90th vs 10th percentile of ξ

	Young, low assets	Old, high assets	Δ
High ξ	0.31	0.22	0.09
	[0.20,0.39]	[0.12,0.34]	[-0.03,0.19]
Low ξ	0.48	0.21	0.27
	[0.40,0.62]	[0.13,0.33]	[0.16,0.38]
Δ	-0.17	0.01	-0.18
	[-0.36,-0.06]	[-0.15, 0.13]	[-0.34,-0.06]

B. 75th vs 25th percentile of ξ

	Young, low assets	Old, high assets	Δ
High ξ	0.36	0.21	0.15
	[0.28,0.42]	[0.14,0.31]	[0.04,0.20]
Low ξ	0.45	0.21	0.24
	[0.38,0.55]	[0.15,0.31]	[0.14,0.31]
Δ	-0.09	0.00	-0.09
	[-0.17,-0.03]	[-0.08, 0.06]	[-0.17,-0.03]

Notes: See the notes to Figure 4. Here we report average consumption responses for young and low assets households compared to old and high assets households, for different percentiles of heterogeneity ξ_i in consumption. Values are calculated by evaluating the average consumption response for households at a fixed percentile of ξ_i when assets and age are fixed at the τth percentile. Reported values for young and low assets households are then shown by averaging over $\tau \in (0, 0.5)$. Reported values for old and high assets households are then shown by averaging over $\tau \in (0.5, 1)$. Parametric bootstrap 95% confidence intervals based on 200 replications are shown in brackets.

a difference-in-differences estimate, is significant at the 5% level.[20]

The results in this section, based on a dynamic model with latent income components and unobserved heterogeneity, provide evidence for the presence of heterogeneous types of consumers, confirming what Figure 2 suggested. In the next section, we develop the implications of these results for life-cycle patterns of consumption and savings, and we examine various possible mechanisms for the patterns in transmission parameters displayed in Figure 5.

Dispersion of consumption responses around their means ϕ_{it}. While our main focus is on the average consumption response parameters ϕ_{it}, there may be dispersion around those averages. In Appendix C we show how to compute an upper bound on the share of variance in responses $\frac{\partial c_{it}}{\partial \eta}$ explained by the means ϕ_{it}, obtained by calculating a lower bound on the variance of $\frac{\partial c_{it}}{\partial \eta}$ conditional on $(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i)$. The reason why only bounds are available is because transitory preference shocks ν_{it}, which may generate additional heterogeneity in responses beyond the mean transmission parameters ϕ_{it}, may be multi-dimensional. We report estimates of the upper bounds on the variance shares in Appendix Figure A11. We find high variance shares, in many cases higher than 80%, suggesting that the ϕ_{it} parameters capture a large part of the heterogeneity in responses (although we note that, since those are upper bounds, this evidence does not strictly speaking rule out the presence of substantial additional heterogeneity).

6 Candidate mechanisms to explain the heterogeneity

In this section we study various mechanisms that might potentially explain the type heterogeneity that we find.

6.1 Three candidate mechanisms

Informed by standard models of consumption and saving decisions, which guide our empirical analysis, we can outline three candidate mechanisms to explain the heterogeneous types that we document.

A first possible explanation is heterogeneity in preferences and discounting. There is a long history of incorporating discount rate heterogeneity to help explain lifetime wealth accumulation, for example Krusell and Smith (1998) and Hendricks (2007). Everything else equal, individuals with higher marginal utility of consumption will consume more, and hold fewer assets. Individuals with higher discount factors will delay consumption relative to those with lower discount factors, and hold more assets. This type of heterogeneity should lead to high-type households consuming more and holding

[20] In Appendix Table A3 we report confidence intervals based on the nonparametric bootstrap clustered at the household level. In this case, for low types below the 10th percentile the average response coefficient remains significantly higher for the young and low assets. However, the cross-type difference is insignificant at the 5% level.

fewer assets. We examine this hypothesis by showing how consumption and assets profiles depend on the latent type.

A second candidate explanation is heterogeneity in returns to assets. The rate of return is a key determinant of consumption choice in standard models, so the types we find might in fact reflect heterogeneity in those returns across households. Fagereng, Guiso, Malacrino and Pistafferi (2020) find evidence of individual heterogeneity in returns to wealth using administrative records from Norway. We examine this heterogeneity in the PSID by estimating an extension of the model with heterogeneity in the asset accumulation rule (see equations (9)-(10)), and by empirically documenting the form of this rule.

A third candidate explanation is heterogeneity in access to external resources, such as parental insurance. Individuals with access to other forms of insurance would be expected to consume more, for a comparable level of income and assets. Altonji, Hayashi, and Kotlikoff (1992), Hayashi, Altonji, and Kotlikoff (1996) and, more recently, Charles, Danziger, Li and Schoeni (2014) and Attanasio, Meghir and Mommaerts (2019), use the generational links in the PSID to document a significant role for parents and family networks in providing additional insurance. To probe this hypothesis, we link the household heads in the PSID to their parents, and study how the latent types relate to parental income, wealth, and consumption.

6.2 Life-cycle profiles

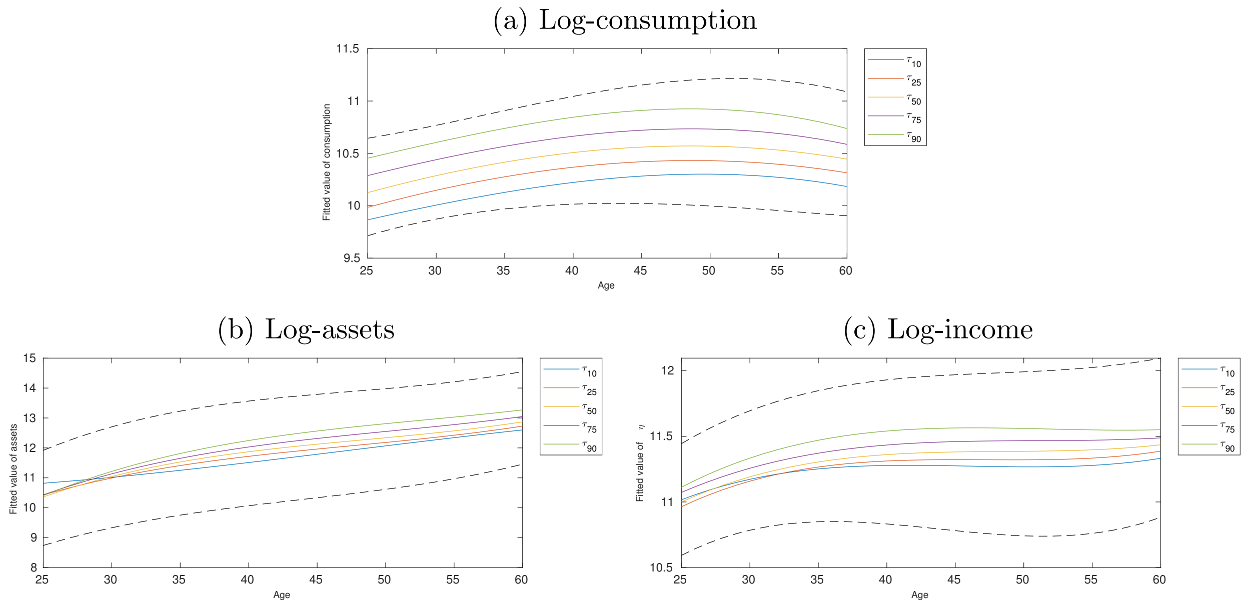
As a step towards examining the plausibility of a preference and discounting channel, we show the life-cycle profiles implied by our dynamic model, for various percentiles of the unobserved heterogeneity ξ_i. In the top panel of Figure 6 we show consumption profiles, in logs.[21] We see that consumption levels are monotone in the types. This is partly a result of our restriction in (12), which implies monotonicity at the reference age. In addition, comparing the dispersion of the solid lines (which correspond to the ξ_i percentiles) with the dashed lines (which correspond to 10th and 90th unconditional percentiles of log-consumption), we see that type heterogeneity explains a large part of the overall variation in log-consumption. Our results imply that ξ_i accounts for 25% of the variance of log-consumption.[22]

In the bottom panel of Figure 6 we show the profiles of assets and income, in logs. In the left graph we see that, similarly to consumption, assets are monotone in types. This suggests that, while high-type households consume more than low types, they also hold more assets. However, the variation

[21]To draw these profiles we proceed by simulation, using a similar strategy to ABB. In addition, in the graphs we show non-residualized variables; that is, we add back the predictions of the first-stage regressions to the residuals of log-consumption, log-assets, and log-income. Note these predictions include the effects of calendar time in addition to those of demographics.

[22]In Appendix Figure A12 we plot the median and 10th and 90th percentile of log-consumption, over the life cycle, for three percentiles of ξ_i (10th, median, and 90th). This confirms that the between-ξ_i dispersion of consumption is substantial, even though there is large within-ξ_i variation as well.

Figure 6: Life-cycle profile



Notes: Average non-residualized log-consumption in graph (a), log-assets in graph (b), and persistent latent component of log-income in graph (c), for different ages and percentiles of ξ_i (10%, 25%, Median, 75%, 90%). The dashed lines show the age-specific unconditional 10th and 90th percentiles for each outcome measure.

in types explains a relatively small share of the overall variation in log-assets. Note that, while the restriction in (12) imposes that log-consumption increases with the type ξ_i at particular covariates values, nothing in our approach restricts log-assets to be monotone in the type. Quantitatively, we find that ξ_i accounts for 3% of the variance of log-assets. In the right graph we show the results for the persistent latent component of income. We see the same monotone behavior in the type as for consumption and assets. Our results imply that ξ_i accounts for 4% of the variance of the persistent latent component of log-income.[23] We have already seen that the correlation between the latent type and income is sufficient to generate sizable differences between specifications with and without latent heterogeneity, see Figure 4.

Overall, our results show that high-type households consume more, hold more assets, and have higher income. Quantitatively, individual types mainly differ in their consumption profiles. While these findings do not rule out that differences in preferences and discounting may be present in the data, they are difficult to reconcile with this channel being the main driver of the heterogeneity in consumption responses that we find.

[23] In Appendix Figure A13 we plot the median and 10th and 90th percentile of log-assets and the persistent latent component of log-income, over the life cycle, for three percentiles of ξ_i. The results confirm that most of the dispersion in assets and income is within-ξ_i.

6.3 Heterogeneity in consumption and assets

We next assess the role of heterogeneity in assets returns as an explanation for type heterogeneity. For this purpose, we estimate a specification where asset accumulation depends on the latent type ξ_i, see equations (9)-(10). The results based on this specification are similar to the baseline ones for both income and consumption. In Appendix Figure A14 we show the type heterogeneity in consumption responses to variation in the persistent latent component of income, and find overall very similar responses to the ones based on a specification without assets heterogeneity. In Appendix Figures A15 and A16 we report estimates of assets responses, by type, in this generalized specification that allows the latent type to enter the asset accumulation rule. We find that the association between lagged assets and current assets conditional on lagged income and consumption increases with the latent type, and that assets responses are higher for the young, decrease with the level of lagged assets, and increase with the type ξ_i, especially for older households.

Overall, the results based on the extended specification with latent heterogeneity in assets and consumption suggest that returns to assets are indeed heterogeneous across households in the data. However, allowing the heterogeneity to enter asset accumulation does not materially affect the conclusions regarding the heterogeneity in consumption responses.

6.4 Heterogeneity in parental insurance

A third candidate mechanism is heterogeneity in access to other forms of insurance, such as parental insurance. In order to examine the plausibility of this mechanism, we take advantage of the intergenerational linkages available in the PSID to match households to their parents. This aspect makes the PSID uniquely suited to study income and consumption dynamics in the presence of links across generations. Specifically, we start by matching the heads of each household to those households headed by a parent of the head. If matches to the household head are not available, we alternatively try and match the spouse of each household to those households headed by a parent of the spouse. In our baseline sample we are able to successfully match approximately 33% of households.

Given this matched panel dataset, we then regress posterior means of the types ξ_i on various parental outcomes, such as consumption, income, and assets. In Table 3 we report the results of various specifications with different sets of controls. For robustness, in addition to the results for all households (in the top panel) we also report results for household heads who are less than 45 years old (in the bottom panel). We find that parental income and consumption correlate positively with the mean type, although the correlation with assets is insignificant from zero at conventional levels. When including all parental variables together, parental consumption remains significantly positively correlated with the type. This suggests that, indeed, the latent type ξ_i may partly reflect heterogeneous access to parental insurance. This interpretation is further supported by the monotonicity of assets in the type documented in Figure 6.

Table 3: Heterogeneity and parental outcomes

A. All households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parent consumption	0.05	0.07					0.04	0.05
	(0.02)	(0.02)					(0.02)	(0.02)
Parent income			0.03	0.04			0.02	0.02
			(0.01)	(0.01)			(0.02)	(0.02)
Parent assets					0.01	0.01	0.00	0.00
					(0.01)	(0.01)	(0.01)	(0.01)
Controls	No	Yes	No	Yes	No	Yes	No	Yes

B. Young adults only

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Parent consumption	0.05	0.06					0.03	0.04
	(0.02)	(0.02)					(0.02)	(0.02)
Parent income			0.03	0.04			0.02	0.02
			(0.01)	(0.01)			(0.02)	(0.02)
Parent assets					0.01	0.01	0.00	0.00
					(0.01)	(0.01)	(0.01)	(0.01)
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: PSID, 2005-2017 sample, household heads aged 25-60 (top panel) and 25-45 (bottom panel). Regressions of posterior ξ_i draws on parental outcomes. Parental links are obtained for approximately 33% of panel. Parental outcomes are obtained as average residuals net of cohort and year effects. Results are based on 10 posterior draws per household. Controls include an education dummy for the household head and a quadratic specification for first period age. Standard errors clustered at the household level do not account for the uncertainty in the posterior parameter estimates.

However, these results are purely indicative and we leave it to future work to assess whether this channel is quantitatively important.

7 Other results and extensions

In this section we report results based on extensions of the model and other robustness checks.

7.1 Impulse responses

We start by reporting impulse responses implied by the model's estimates. In Figure 7 we estimate the impact of a shock to the persistent latent component of income, η_{it}, at age 34. The figure is divided into three parts. In the upper part, we report the difference between the average persistent latent component of income for households hit by the shock and the average persistent latent component of income for households hit by a "median" shock, i.e., corresponding to the 50th percentile of η_{it} conditional on $\eta_{i,t-1}$. To highlight the heterogeneity in impulse responses, we show results for various percentiles of the latent type distribution. In the middle and bottom parts of the figure we proceed similarly for log-consumption and log-assets, respectively, instead of the income component.

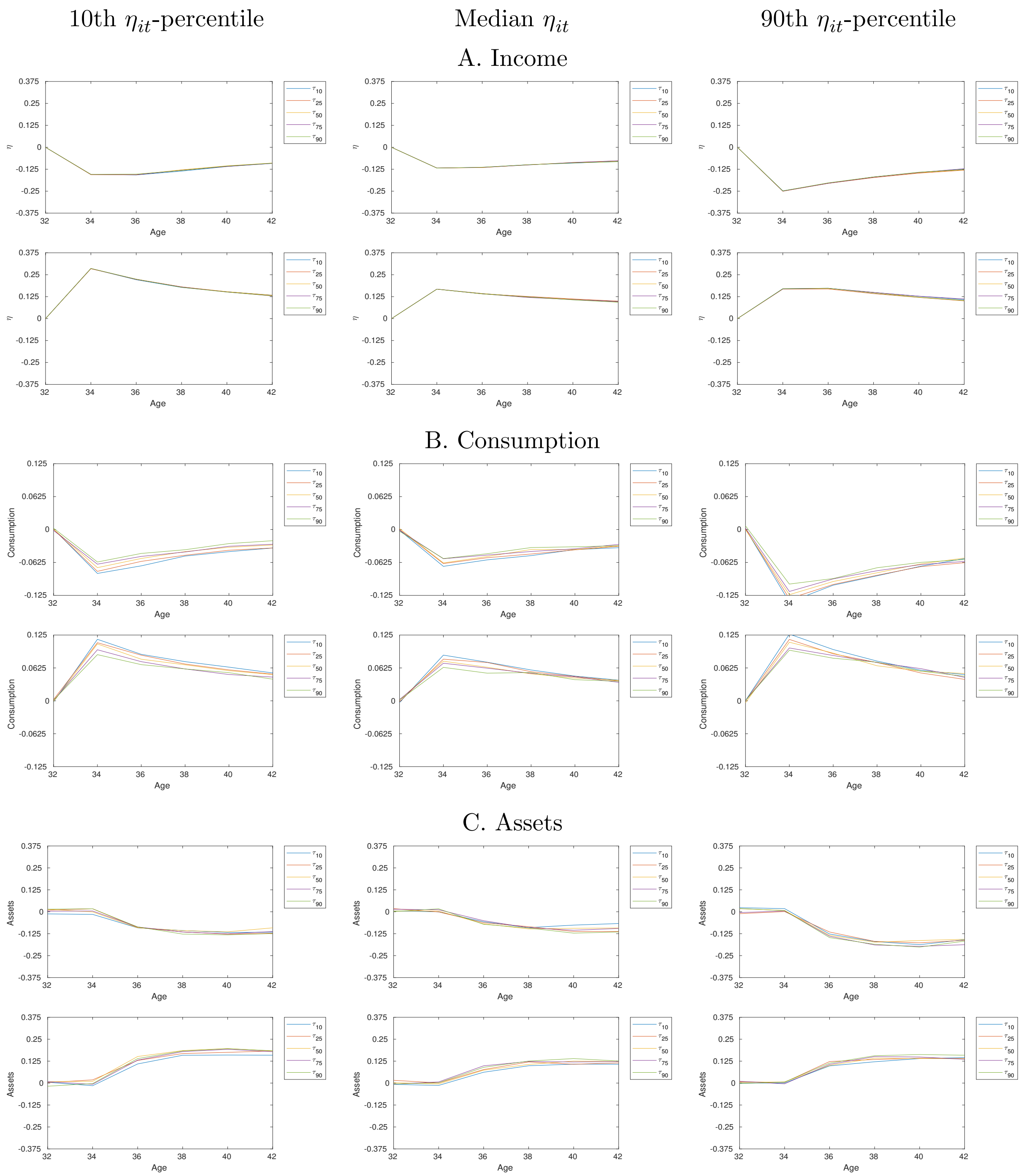
Within each part of the figure, we show impulse responses for various values of initial income and the shock. In the left, middle and right columns we consider households who are at the 10th, 50th and 90th percentile of the distribution of the persistent income component at age 32, respectively. In the top (respectively, bottom) subpanels, we show the results for a shock at the 10th (respectively, 90th) percentile of the distribution of shocks. Hence, top subpanels correspond to negative income shocks, whereas bottom subpanels correspond to positive income shocks.

Focusing first on the upper part of Figure 7, and moving across columns, we observe that negative shocks tend to have a stronger impact for those on higher income, and that positive shocks have a stronger impact for those on lower income. This illustrates the nonlinear persistence in the income process documented in ABB. In addition, the fact that all lines corresponding to different values of the latent type ξ_i reflects our assumption that the income process does not depend on ξ_i.

Moving then to the middle part of Figure 7, we also observe nonlinearities in consumption responses, although those are stronger for the negative income shocks than for the positive ones. In addition, the differences between lines reflect the heterogeneity between types. In particular, low types with higher income tend to respond more strongly to negative shocks than other types. To further illustrate this heterogeneity, in Appendix Figure A17 we show how consumption levels evolve, on impact, after an income shock.

Lastly, focusing on the bottom part of Figure 7 we see only moderate differences in assets evolution after a shock depending on the initial income level. In Appendix Figure A18 we show impulse responses based on the model that allows for heterogeneity in both assets and consumption, see equations (9)-

Figure 7: Heterogeneity in impulse responses



Notes: Impulse responses shown for shocks at the 10th (top subpanels) and 90th (bottom subpanels) percentiles, relative to median. See the text for a description. The different lines correspond to different percentiles of ξ_i.

(10) and the results discussed in Subsection 6.3. The responses to a shock to the persistent latent component of income are overall similar to the ones based on the model without heterogeneity in the asset accumulation rule.

7.2 Robustness to the complexity of the quantile model used in estimation

The complexity of our empirical specification is controlled in part by the number of knots at which we evaluate the quantiles of the variables in the model (i.e., the income components, consumption, and the latent type). Our estimates of the functions, such as $a_k^{\eta}(\tau)$ in (11), interpolate between those τ values. Hence, a large number of knots can approximate any continuous quantile function well, while a small number of knots may provide a worse approximation. However, in estimation one faces the usual bias/variance trade-off, and the impact of the number of knots on the estimates is *a priori* unclear. To probe the sensitivity of our main results to the number of knots, we report average consumption derivatives based on 19 knots in Appendix Figure A19. By comparison, our baseline results were obtained using 11 knots (see Figure 5). Overall the two sets of estimates agree very well.

7.3 Robustness to income definition and sample restriction

Next, we probe the robustness of our results to changes in income definition and sample restriction. While our main results rely on using disposable, post-tax income, in Appendix Figure A20 we report results on nonlinear income persistence based on pre-tax labor income. In Appendix Figure A21 we report the corresponding results for heterogeneity in consumption responses. The findings suggest a higher degree of nonlinearity in income persistence, and a higher degree of consumption insurance, compared to the results based on disposable income. This is not surprising, as the non-proportionality in the tax system can be interpreted as a source of insurance to households. Moreover, since the results in ABB were based on labor income, these findings help explain the quantitative differences between the results in ABB and the ones we report in this paper when relying on disposable income.

Another important feature of our sample is the restriction to dual earner households. While this restriction is motivated by the goal to abstract from extensive labor supply decisions, it also results in a smaller and potentially more insured sample. We have estimated our model on a larger sample that also includes single earners, where the second member of the household is not working.[24] In Appendix Figure A22 we report the results for income persistence, and in Appendix Figure A23 we reproduce our main results on heterogeneity in consumption responses to income shocks. Our findings are qualitatively unchanged relative to our baseline sample of dual earners.

[24] In Appendix Table A4 we show descriptive statistics for this broader sample.

7.4 Additional dimensions of heterogeneity

Our specification of the consumption function flexibly allows for heterogeneity in income, assets, age effects, and the effect of the latent type, see equation (1). However, it is possible that the effects of additional observed and unobserved factors might matter for consumption insurance. For example, differences in education and birth cohorts might be associated with different consumption responses to income shocks. In Appendix Figure A25 we show that neither education nor cohort are strongly associated with the latent type ξ_i. Yet, it is theoretically possible that they enter the consumption function, and interact with income components in meaningful ways, even though our modeling approach rules out this possibility.

To tentatively explore this question, in Appendix Figure A26 we report consumption responses to income shocks, by type ξ_i, in a specification that also controls for a fully interacted education indicator. Since we do not re-estimate the model with latent variables, we view this exercise as indicative. We see that the consumption responses across types are qualitatively similar to the baseline ones, yet those responses appear somewhat muted. This motivates future work extending our framework to allow for multiple observed and unobserved sources of heterogeneity in consumption insurance and income processes across households.

8 Conclusion

The motivation for this research has been to better understand nonlinear income dynamics and heterogeneous consumption responses to changes in income. In this paper we have developed methods that build on and extend Arellano, Blundell and Bonhomme (2017), and we have applied them to a larger and more comprehensive sample from the PSID which includes a richer set of consumption categories. We have developed computational tools to better handle larger and more complex models, including in settings with unbalanced panels, within a nonlinear quantile-based latent variables framework. These new data and tools allow us to go beyond confirming the presence of nonlinear income and consumption dynamics, and to document rich heterogeneity in consumption responses across households.

Our results point to consumption responses to income shocks that vary substantially with unobserved types. We distinguish lower types, who appear to follow the life-cycle patterns in consumption responses implied by standard models, from higher types, whose consumption responses to income shocks vary little with either assets levels or the stage of the life cycle. High-type households consistently have higher consumption levels and, relative to low-type households, have slightly higher incomes and levels of assets. For the younger low types, consumption responses to persistent income shocks are close to 0.60 while for older low types this falls to 0.10. For the higher types, consumption responses are flatter across age and assets.

We examined alternative mechanisms that could lead to such heterogeneous consumption responses. The fact that high types both consume more and hold more assets is difficult to reconcile with an explanation based on heterogeneity in preferences or discounting. We also argue that it is difficult to align with a specification that allows for latent heterogeneity in asset accumulation, finding that the heterogeneity in consumption responses is virtually unaffected by this extension. To explore a third mechanism, parental insurance, we used the inter-generational linkages in the PSID to link a subset of household heads in our sample to their parents. We found that high-type household heads have on average parents with higher consumption and income levels, suggesting that the heterogeneous responses might in part reflect heterogeneity in access to other sources of insurance such as parental insurance.

Our findings motivate further work on two fronts. First, whilst we have examined several mechanisms and found a correlation between the latent types and parental consumption, we lack a quantitative understanding of how these and other factors shape the household differences in consumption responses and insurance. Second, although we have leveraged a single-latent-factor model to maintain tractability in the presence of heterogeneous responses, generalizing the model to account for other sources of heterogeneity is an important next step. In particular, it would be valuable to extend the model to allow for time-invariant heterogeneity in income, in addition to the latent consumption type.

References

[1] Alan, S., Browning, M., and Ejrnæs, M. (2018). Income and consumption: A micro semistructural analysis with pervasive heterogeneity. Journal of Political Economy, 126(5), 1827-1864.

[2] Altonji, J., Hayashi, F., and Kotlikoff, L. (1992). Is the Extended Family Altruistically Linked? Direct Tests Using Micro Data.American Economic Review, vol. 82(5). 1177–98.

[3] Andreski, P., Geng, L., Samancioglu, M.Z., and Schoeni, R. (2014). Estimates of Annual Consumption Expenditures and Its Major Components in the PSID in Comparison to the CE. American Economic Review: Papers & Proceedings 2014, 104(5): 132–135.

[4] Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle markov chain monte carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(3), 269-342.

[5] Anghel, B., Basso, H., Bover, O., Casado, J. M., Hospido, L., Izquierdo, M., and Vozmediano, E. (2018). Income, consumption and wealth inequality in Spain. SERIEs, 9(4), 351-387.

[6] Arellano, M., Blundell, Richard and S. Bonhomme. 2017. "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework", *Econometrica*, 85, 693- 734.

[7] Arellano, M., and Bonhomme, S. (2016). Nonlinear panel data estimation via quantile regressions. The Econometrics Journal, 19(3), C61-C94.

[8] Attanasio, O., Meghir, C. and Mommaerts, C. (2019). Insurance in extended family networks. NBER Working Paper 21059.

[9] Auclert, A. (2019). Monetary Policy and the Redistribution Channel. American Economic Review 109(6), 2333–2367, June.

[10] Baker, Michael, and Gary Solon. 2003. "Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records." Journal of Labor Economics, 21(2): 267-288.

[11] Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98(5): 1887–1921.

[12] Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016): "Consumption Smoothing and Family Labor Supply," *American Economic Review*, 106(2), 387–435.

[13] Charles, K. K., S. Danziger, G. Li, and R. Schoeni (2014): "The Intergenerational Correlation of Consumption Expenditures," *American Economic Review*, 104(5), 136–40.

[14] Chernozhukov, V., I. Fernández-Val, I., and A. Galichon (2010): "Quantile and Probability Curves without Crossing," *Econometrica*, 78(3), 1093–1125.

[15] Chernozhukov, V., and Hansen, C. (2005). An IV model of quantile treatment effects. Econometrica, 73(1), 245-261.

[16] Crawley, E., and Kuchler, A. (2020). Consumption Heterogeneity: Micro Drivers and Macro Implications.

[17] Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. Econometric reviews, 31(3), 245-296.

[18] Deaton, A. (1992). Understanding consumption. Oxford University Press.

[19] De Nardi, M., Fella, G., and Paz-Pardo, G. (2020). Nonlinear household earnings dynamics, self-insurance, and welfare. Journal of the European Economic Association, 18(2), 890-926.

[20] Fagereng, A., Guiso, L., Malacrino, D., and Pistaferri, L. (2016). Heterogeneity in returns to wealth and the measurement of wealth inequality. American Economic Review, 106(5), 651-55.

[21] Griliches, Z., and J. A. Hausman (1986): "Errors in Variables in Panel Data," Journal of Econometrics, 31, 93–118.

[22] Hall, R., and F. Mishkin (1982): "The sensitivity of Consumption to Transitory Income: Estimates from Panel Data of Households," *Econometrica*, 50(2): 261–81.

[23] Hayashi, F., Altonji, J., and Kotlikoff, L. (1996). Risk-Sharing between and within Families. Econometrica, vol. 64(2), 261–294.

[24] Hendricks, L. (2007). How Important is Discount Rate heterogeneity for Wealth Inequality. Journal of Economic Dynamics and Control, 31(9), 3042-3068.

[25] Hoderlein, S., and E. Mammen (2007). "Identification of Marginal Effects in Nonseparable Models Without Monotonicity," Econometrica, 75(5), 1513–1518.

[26] Hu, Y., and Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. Econometrica, 76(1), 195-216.

[27] Imbens, G. W., and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Econometrica, 77(5), 1481–1512.

[28] Kantas, N., Doucet, A., Singh, S. S., Maciejowski, J., and Chopin, N. (2015). On particle methods for parameter estimation in state-space models. Statistical science, 30(3), 328-351.

[29] Kaplan, G. and Violante, G.L., (2010). How much Consumption Insurance Beyond Self-Insurance? American Economic Journal: Macroeconomics 2 (October): 53–87

[30] Koenker, R., (2005). Quantile Regression. Econometric Society Monograph Series.

[31] Krueger, D., K. Mitman, and F. Perri (2015): "Macroeconomics and Heterogeneity, Including Inequality," unpublished manuscript.

[32] Krusell, P. and A. Smith, (1998). Income and Wealth Heterogeneity in the Macroeconomy. Journal of Political Economy, 2:1, 245-272.

[33] Li, G., R. F. Schoeni, S. Danziger, and K. K. Charles (2010). New Expenditure Data in the PSID: Comparisons with the CE. *Monthly Lab. Rev.*, 133, 29.

[34] Machado, J. A., and J. Mata (2005): "Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression," *Journal of Applied Econometrics*, 20(4), 445–465.

[35] Matzkin, R. L. (2013). Nonparametric identification in structural economic models. Annu. Rev. Econ., 5(1), 457-486.

[36] Meghir, C., and L. Pistaferri (2011): "Earnings, Consumption and Life Cycle Choices," *Handbook of Labor Economics*, Elsevier.

[37] Moffitt, R., and Gottschalk, P. (1993). Trends in the covariance structure of earnings in the United States: 1969-1987 (No. 1001-93). University of Wisconsin Institute for Research on Poverty.

[38] Newey, W. K., and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," Econometrica, 71(5), 1565–1578.

[39] Nielsen, S. F. (2000). The stochastic EM algorithm: estimation and asymptotic results. Bernoulli, 457-489.

[40] Wilhelm, D. (2015): "Identification and Estimation of Nonparametric Panel Data Regressions with Measurement Error," Cemmap working paper No. CWP34/15.

APPENDIX

A Modeling and estimation details

A.1 Empirical specification

Earnings components. Let φ_k, for $k = 0, 1, ...,$ denote a dictionary of functions, with $\varphi_0 = 1$. In practice we use low-order products of Hermite polynomials for φ_k. We specify, for $t \in \{t_i + 1, ..., t_i + T_i - 1\}$, the conditional quantile function of η_{it} given $\eta_{i,t-1}$ and age_{it} as in (11). We specify the quantile function of ε_{it} (for $t = 1, ..., T$) given age_{it}, and that of η_{i1} given age at the start of the period age_{i1}, in a similar way. Specifically, we set

$$Q_\varepsilon(age_{it}, \tau) = \sum_{k=0}^{K} a_k^\varepsilon(\tau)\varphi_k(age_{it}),$$
$$Q_{\eta_1}(cohort_i, educ_i, age_{i,t_i}, \tau) = \sum_{k=0}^{K} a_k^{\eta_1}(\tau)\varphi_k(cohort_i, educ_i, age_{i,t_i}),$$

with outcome-specific choices for K and φ_k.

Consumption type. To specify the latent type we set

$$Q_\xi(cohort_i, educ_i, income_i, \tau) = \sum_{k=0}^{K} a_k^\xi(\tau)\varphi_k(cohort_i, educ_i, income_i).$$

Consumption rule. To specify the consumption process we set

$$Q_c(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i, \tau) = \sum_{k=1}^{K} a_k^c(\tau)\varphi_k(a_{it}, \eta_{it}, \varepsilon_{it}, age_{it}, \xi_i). \quad \text{(A1)}$$

To fix the scale of the function we impose that

$$\int_0^1 Q_c(\overline{a}, \overline{\eta}, \overline{\varepsilon}, \overline{age}, \xi, \tau)d\tau = \xi,$$

which translates into linear restrictions on the parameters $\int_0^1 a_k^c(\tau)d\tau$.

Assets evolution. For initial assets we set

$$Q_{a_1}(\eta_{i,t_i}, age_{i,t_i}, cohort_i, educ_i, \xi_i, \tau) = \sum_{k=0}^{K} a_k^{a_1}(\tau)\varphi_k(\eta_{i,t_i}, age_{i,t_i}, cohort_i, educ_i, \xi_i). \quad \text{(A2)}$$

For assets evolution we set

$$Q_a(a_{it}, \eta_{it}, \varepsilon_{it}, c_{it}, age_{it}, \xi_i, \tau) = \sum_{k=0}^{K} a_k^a(\tau)\varphi_k(a_{it}, \eta_{it}, \varepsilon_{it}, c_{it}, age_{it}, \xi_i, \tau). \quad \text{(A3)}$$

Implementation. We base our implementation on ABB, and model the functions $a_k(\tau)$ as piecewise-linear interpolating splines on a grid $[\tau_1, \tau_2]$, $[\tau_2, \tau_3]$, ... , $[\tau_{L-1}, \tau_L]$, contained in the unit interval. We extend the specification of the intercept coefficient a_0 on $(0, \tau_1]$ and $[\tau_L, 1)$ using a Laplace model indexed by λ_- (for the left tail) and λ_+ (for the right tail). All a_k for $k \geq 1$ are constant on $[0, \tau_1]$ and $[\tau_L, 1]$, respectively. We denote $a_{k\ell} = a_k(\tau_\ell)$. In practice, we take $L = 11$ and $\tau_\ell = \ell/(L+1)$. We use tensor products of Hermite polynomials for φ_k, each component of the product taking as argument a standardized variable.

A.2 Estimation

Overview of the estimation strategy. We start by estimating the earnings parameters. Next, we recover estimates of the consumption, assets, and type parameters, given the previous earnings estimates.

Parameters. We collect all parameters governing the income process into a vector θ, given by

$$\theta = \left(a^{\eta}, \lambda^{\eta}, a^{\varepsilon}, \lambda^{\varepsilon}, a^{\eta_1}, \lambda^{\eta_1} \right).$$

Likewise, we collect all parameters governing the consumption process into a vector μ, given by

$$\mu = \left(a^{\xi}, \lambda^{\xi}, a^{c}, \lambda^{c}, a^{a_1}, \lambda^{a_1}, a^{a}, \lambda^{a} \right).$$

We estimate θ and μ sequentially.

Model's restrictions Let $\rho_\tau(u) = u(\tau - \mathbf{1}\{u \leq 0\})$ denote the "check" function of quantile regression. Consider the parameters of Q_η; that is, the $a^{\eta}_{k\ell}$ and the corresponding Laplace parameters λ^{η}. The true values of $a^{\eta}_{k\ell}$ maximize

$$E\left[\sum_{t=t_i+1}^{t_i+T_i-1} \int \rho_{\tau_\ell} \left(\eta_t - \sum_{k=0}^{K} a^{\eta}_{k\ell} \varphi_k(\eta_{t-1}, age_{it}) \right) f_i(\eta) d\eta \right] = 0,$$

where f_i is the posterior distribution of the $(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1})$ given the data and the true parameter values. In turn, the true values of λ^{η} satisfy

$$\overline{\lambda}^{\eta}_{-} = -\frac{E\left[\sum_{t=t_i+1}^{t_i+T_i-1} \int \mathbf{1}\left\{ \eta_t \leq \sum_{k=0}^{K} \overline{a}^{\eta}_{k1} \varphi_k(\eta_{t-1}, age_{it}) \right\} f_i(\eta) d\eta \right]}{E\left[\sum_{t=t_i+1}^{t_i+T_i-1} \int \left(\eta_t - \sum_{k=0}^{K} \overline{a}^{\eta}_{k1} \varphi_k(\eta_{t-1}, age_{it}) \right) \mathbf{1}\left\{ \eta_t \leq \sum_{k=0}^{K} \overline{a}^{\eta}_{k1} \varphi_k(\eta_{t-1}, age_{it}) \right\} f_i(\eta) d\eta \right]},$$

with an analogous formula for the upper tail parameter λ^{η}_{+}. The model implies related restrictions on all the other quantile and tail parameters in θ and μ.

Likelihood function. The likelihood function is, letting $z_i = (cohort_i, educ_i)$ and $\mathcal{T}_i = \{t_i, ..., t_i + T_i - 1\}$,

$$
\begin{aligned}
& f(y_i^{\mathcal{T}_i}, c_i^{\mathcal{T}_i}, a_i^{\mathcal{T}_i}, \eta_i^{\mathcal{T}_i}, \xi_i \,|\, age_i^{\mathcal{T}_i}, z_i; \theta, \mu) \\
& = \prod_{t \in \mathcal{T}_i} f(c_{it}|a_{it}, \eta_{it}, y_{it}, \xi_i, age_{it}; \mu) \\
& \quad \times \prod_{t \in \mathcal{T}_i, t > t_i} f(a_{it}|a_{i,t-1}, y_{i,t-1}, c_{i,t-1}, \eta_{i,t-1}, \xi_i, age_{it}; \mu) \\
& \quad \times \prod_{t \in \mathcal{T}_i} f(y_{it}|\eta_{it}, age_{it}; \theta) \prod_{t \in \mathcal{T}_i, t > t_i} f(\eta_{it}|\eta_{i,t-1}, age_{it}; \theta) \\
& \quad \times f(a_{i,t_i}|\eta_{i,t_i}, age_{i,t_i} z_i, \xi_i; \mu) f(\eta_{i,t_i} \,|\, z_i, age_{i,t_i}; \theta) f(\xi_i \,|\, z_i, income_i; \mu),
\end{aligned}
$$

where notice we have imposed the assumption that ξ_i is independent of $(y_i^{\mathcal{T}_i}, \eta_i^{\mathcal{T}_i})$ given $(z_i, income_i)$.

Similarly to ABB, the likelihood function is available in closed form. For example, we have

$$
\begin{aligned}
f(y_{it}|\eta_{it}, age_{it}; \theta) \;=\; & \mathbf{1}\left\{y_{it} - \eta_{it} < A_{it}^{\varepsilon}(1)\right\} \tau_1 \lambda_-^{\varepsilon} \exp\left[\lambda_-^{\varepsilon}\left(y_{it} - \eta_{it} - A_{it}^{\varepsilon}(1)\right)\right] \\
& + \sum_{\ell=1}^{L-1} \mathbf{1}\left\{A_{it}^{\varepsilon}(\ell) \le y_{it} - \eta_{it} < A_{it}^{\varepsilon}(\ell+1)\right\} \frac{\tau_{\ell+1} - \tau_\ell}{A_{it}^{\varepsilon}(\ell+1) - A_{it}^{\varepsilon}(\ell)} \\
& + \mathbf{1}\left\{A_{it}^{\varepsilon}(L) \le y_{it} - \eta_{it}\right\} (1 - \tau_L) \lambda_+^{\varepsilon} \exp\left[-\lambda_+^{\varepsilon}\left(y_{it} - \eta_{it} - A_{it}^{\varepsilon}(L)\right)\right],
\end{aligned}
$$

where

$$
A_{it}^{\varepsilon}(\ell) \equiv \sum_{k=0}^{K} a_{k\ell}^{\varepsilon} \varphi_k(age_{it}) \text{ for all } (i, t, \ell).
$$

Note that the likelihood function is non-negative by construction. In particular, drawing from the posterior density of η automatically produces rearrangement of the various quantile curves (Chernozhukov, Fernández-Val and Galichon, 2010).

Estimation algorithm. Like in ABB, starting from initial parameter values, we iterate between two steps.

In the stochastic E-step, we draw M values $\eta_i^{(m)} = (\eta_{i,t_i}^{(m)}, ..., \eta_{i,t_i+T_i-1}^{(m)})$ and $\xi_i^{(m)}$ from their posterior distribution. In practice we take $M = 1$.

In the M-step, we estimate parameters by solving empirical counterparts of the population restrictions. This involves running multiple quantile regressions in order to estimate the $a_{k\ell}$ parameters, and estimating the λ parameters which are available in closed form.

Solving the indeterminacy in consumption. To impose the restriction (12), which solves the indeterminacy in the relationship between consumption and the latent type, we proceed as follows. At the start of every M-step, given draws $\eta_i^{(m)}$ and $\xi_i^{(m)}$, we regress c_{it} on polynomials in a_{it}, $\eta_{it}^{(m)}$, $\varepsilon_{it}^{(m)} = y_{it} - \eta_{it}^{(m)}$, age_{it}, and $\xi_i^{(m)}$, using the same polynomial specification as in the quantile model for log-consumption. Letting $\widehat{c}_{it}$ denote the predicted value at $(\overline{a}, \overline{\eta}, \overline{\varepsilon}, \overline{age}, \xi_i^{(m)})$, we then reset $\widehat{c}_{it} \mapsto \xi_i^{(m)}$.

Stochastic E-step (income estimation). The target for a given household i is the posterior distribution

$$f(\eta_{i,t_i},...,\eta_{i,t_i+T_i-1}|y_{i,t_i},...,y_{i,t_i+T_i-1}).$$

At $t = t_i$, we initialize S particles $\eta_{i,t_i}^{(s)}$ from the following proposal distribution π:

$$\begin{aligned}
\eta_{i,t_i} &\sim \mathcal{N}(\mu_i, \sigma^2),\\
\mu_i &= \left(1 - \frac{\sigma_{\eta_1}^2}{\sigma_{\eta_1}^2 + \sigma_\varepsilon^2}\right)\sum_{k=0}^{K}\beta_k^\varepsilon \varphi_k(cohort_i, educ_i, age_{i,t_i}) + \frac{\sigma_{\eta_1}^2}{\sigma_{\eta_1}^2 + \sigma_\varepsilon^2} y_{i,t_i},\\
\sigma^2 &= \frac{c}{\frac{1}{\sigma_{\eta_1}^2} + \frac{1}{\sigma_\varepsilon^2}},
\end{aligned}$$

where the β_k^ε, $\sigma_{\eta_1}^2$ and σ_ε^2 are parameters estimated by running OLS counterparts to the M-step quantile regressions (in the previous stochastic EM iteration), and $c \geq 1$ is a constant (we take $c = 2$). Time $t = t_i$ re-sampling weights are then given by

$$w_{i,t_i}^{(s)} \propto \frac{f(\eta_{i,t_i}^{(s)}|y_{i,t_i})}{\pi(\eta_{i,t_i}^{(s)})},$$

where π is the normal density with mean μ_i and variance σ^2. These weights, which are available in closed form, are used to re-sample particles with replacement from the set of particles $\eta_{i,t_i}^{(s)}$, if the effective sample size $\frac{1}{\sum_{s=1}^{S}(w_{i,t_i}^{(s)})^2}$ exceeds some threshold (see below). This simple adaptive rule avoids degeneracy of the particles. After re-sampling we reset $w_{i,t_i}^{(s)} = \frac{1}{S}$. Otherwise we keep all the existing particles and weights.

At $t = t_i + r > t_i$, we use the following proposal distribution, again denoted as π, to generate new draws to append to the existing set of particles:

$$\begin{aligned}
\eta_{i,t_i+r}\,|\,\eta_{i,t_i+r-1} &\sim \mathcal{N}(\widetilde{\mu}_{i,r}, \widetilde{\sigma}^2),\\
\widetilde{\mu}_{i,r} &= \left(1 - \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2}\right)\sum_{k=0}^{K}\beta_k^\varepsilon \varphi_k(\eta_{i,t+r-1}, age_{i,t_i+r}) + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\varepsilon^2} y_{i,t_i+r},\\
\widetilde{\sigma}^2 &= \frac{c}{\frac{1}{\sigma_\eta^2} + \frac{1}{\sigma_\varepsilon^2}},
\end{aligned}$$

where again the β_k^ε, σ_η^2 and σ_ε^2 are parameters estimated by running OLS counterparts to the M-step quantile regressions. The re-sampling weights are given by

$$w_{i,t_i+r}^{(s)} \propto w_{i,t_i+r-1}^{(s)} \frac{f(\eta_{i,t_i+r}^{(s)}|y_{i,t_i+r}, \eta_{i,t_i+r-1}^{(s)})}{\pi(\eta_{i,t_i+r}^{(s)}|\eta_{i,t_i+r-1}^{(s)})},$$

which are used to re-sample particles if the effective sample size $\frac{1}{\sum_{s=1}^{S}\left(w_{i,t_i+r}^{(s)}\right)^2}$ exceeds the threshold.

Stochastic E-step (consumption estimation). The target for a given household is the posterior distribution

$$f(\xi_i, \eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1} \,|\, x_{i,t_i}, ..., x_{i,t_i+T_i-1}),$$

where $x_{it} = (y_{it}, c_{it}, a_{it}, age_{it})$ is a vector of household i's observed income, consumption, assets and age at time t. Algorithm A1 below provides a pseudo-code for the implementation. The SMC sampling steps (used to generate efficient proposals within a Metropolis Hastings algorithm) are identical to those outlined above with the exception that re-sampling weights at times $t = t_i$ and $t > t_i$ are now given by

$$w_{i,t_i}^{(s)} \propto \frac{f\left(\eta_{i,t_i}^{(s)} | \xi_i^*, x_{i,t_i}\right)}{\pi\left(\eta_{i,t_i}^{(s)}\right)},$$

and

$$w_{i,t_i+r}^{(s)} \propto \frac{f\left(\eta_{i,t_i+r}^{(s)} | \xi_i^*, x_{i,t_i+r}, \eta_{i,t_i+r-1}^{(s)}\right)}{\pi\left(\eta_{i,t_i+1}^{(s)} | \eta_{i,t_i}^{(s)}\right)},$$

respectively, where ξ_i^* is a draw from a random walk proposal. We make use of the same proposal distributions π as in the income estimation.

In the very first iteration of the stochastic EM algorithm we initialize the Metropolis Hastings chains using random draws from the following proposal:

$$\xi_i^* \sim \mathcal{N}(\nu_i, \omega^2),$$

where $\nu_i = \sum_{k=0}^{K} \beta_k^{\xi} \varphi_k(cohort_i, educ_i, income_i)$. The parameters β_k^{ξ} and ω^2 are estimated by running OLS counterparts to the corresponding M-step quantile regressions. At subsequent iterations of the stochastic EM we initialize the Metropolis Hastings chains using draws from the previous iteration. After initialization we use a Gaussian random walk proposal with variance $3.5\omega^2$.

Whilst running the SMC samplers we obtain unbiased estimates of the marginal likelihood which can be calculated recursively as $\hat{p}(x_{i,t_i+r}, ...|\xi_i^*) = \sum_{s=1}^{S} \hat{p}(x_{i,t_i+r-1}, ...|\xi_i^*) w_{i,t_i+r}^{(s)}$. The unbiasedness of these marginal likelihood estimates implies that the resulting algorithm can be represented as a *bona fide* Metropolis Hastings algorithm yielding the desired target as its marginal.

Pseudo-code of the stochastic EM algorithm. A short pseudo-code for the algorithm we use is presented in Algorithm A1.

Algorithm A1 *(Stochastic EM)*

1: ***for*** ℓ*=1:L* ***do***
2:   ***Stochastic E-Step:***
3:   *Set* ξ_i^0 *and* $(\eta_{i,t_i}^0, ..., \eta_{i,t_i+T_i-1}^0)$ *to some starting values.*[1]
4:   ***for*** *k=1:K* ***do***
5:   *Sample* $\xi_i^* \sim q(.|\xi_i^{k-1})$*, where q is a proposal distribution.*[2]

[1]When $\ell > 1$ we simply take ξ_i^0 to be the ξ_i draw from the previous $(\ell - 1)$ step. When $\ell = 1$ we always accept the first proposal. In both cases, we run an SMC algorithm (see line 6 in the pseudo-code) based on ξ_i^0 to generate a draw $(\eta_{i,t_i}^0, ..., \eta_{i,t_i+T_i-1}^0)$.

[2]In practice, we use a random walk proposal. We tune the variance of the proposal so that the acceptance rate is approximately 30%.

6: *Run an SMC algorithm targeting* $p(\eta_{i,t_i}, ..., \eta_{i,t_i+T_i-1} | \xi_i^*, w_{i,t_i}, ..., w_{i,t_i+T_i-1})$.

7: *Store the marginal likelihood estimate,* $\hat{p}(\xi_i^*) = p(w_{i,t_i}, ..., w_{i,t_i+T_i-1} | \xi_i^*)$, *and the resulting particles* $\eta_{i,t_i}^*, ..., \eta_{i,t_i+T_i-1}^*$, *both of which are available as output of the SMC algorithm in line 6.*

8: *Let* f *denote the density of* ξ_i, *whose expression is given in Appendix A. With probability* $\min\left(1, \frac{\hat{p}(\xi_i^*) f(\xi_i^*) q(\xi_i^{k-1} | \xi_i^*)}{\hat{p}(\xi_i^{k-1}) f(\xi_i^{k-1}) q(\xi_i^* | \xi_i^{k-1})}\right)$ *set* $\xi_i^k = \xi_i^*$ *and* $(\eta_{i,t_i}^k, ..., \eta_{i,t_i+T_i-1}^k) = (\eta_{i,t_i}^*, ..., \eta_{i,t_i+T_i-1}^*)$; *otherwise set* $\xi_i^k = \xi_i^{k-1}$ *and* $(\eta_{i,t_i}^k, ..., \eta_{i,t_i+T_i-1}^k) = (\eta_{i,t_i}^{k-1}, ..., \eta_{i,t_i+T_i-1}^{k-1})$.

9: ***end for***

10: *Keep the last values* ξ_i^K *and* $(\eta_{i,t_i}^K, ..., \eta_{i,t_i+T_i-1}^K)$.

11: ***M-Step:***

12: *Estimate the quantile parameters by quantile regressions given the draws* ξ_i^K *and* $(\eta_{i,t_i}^K, ..., \eta_{i,t_i+T_i-1}^K)$, *as explained in Appendix A. Estimate the Laplace tail parameters.*

13: *Update the parameters of the proposal distribution, as explained in Appendix A.*

14: ***end for***

Practical issues: number of particles and threshold for effective sample size. In practice, we set an i-specific number of particles equal to $S_i = 50T_i$, where T_i is the number of observations of household i. We set the threshold for effective sample size to $S_i/2$.

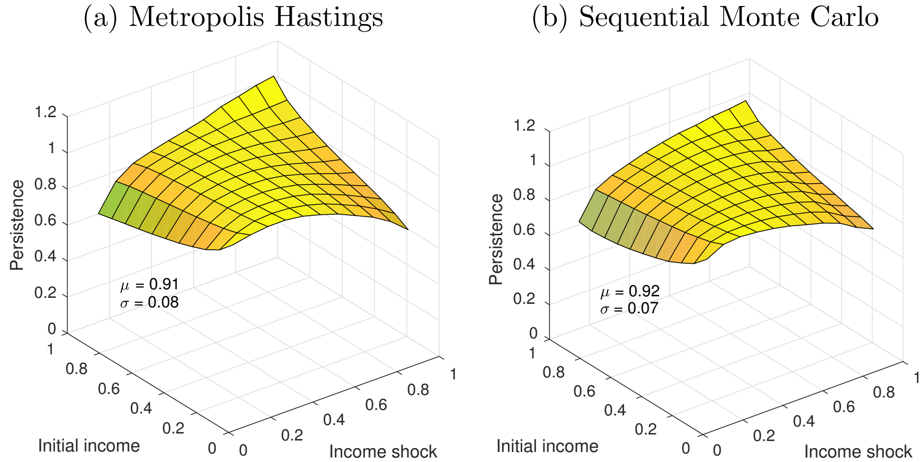
Practical issues: specification. In practice we set the following polynomial degrees K for our baseline specification, chosen after some experimentation:

- Q_η: $K^\eta = 3$, $K^{age} = 2$.
- Q_{η_1}: $K^{educ} = 1$, $K^{cohort} = 1$, $K^{age} = 2$.
- Q_ε: $K^{age} = 2$.
- Q_c: $K^{age} = 1$, $K^a = 2$, $K^\eta = 2$, $K^\varepsilon = 1$, $K^\xi = 1$.
- Q_a: $K^{age} = 1$, $K^a = 2$, $K^y = 1$, $K^c = 1$.
- Q_{a_1}: $K^{age} = 1$, $K^\eta = 1$, $K^\xi = 1$, $K^{education} = 1$, $K^{cohort} = 1$.
- Q_ξ: $K^{income} = 1$, $K^{educ} = 1$, $K^{cohort} = 1$.

Practical issues: starting values. In practice, we start the algorithm from different parameter values. For example, for the initial values of the quantile parameters in η_{it}, we run quantile regressions of log-earnings on lagged log-earnings and age. We proceed similarly to set other starting parameter values, including those for the proposal distributions. In addition, we use latent draws from the income model as initial draws when estimating the consumption model. We experimented with a number of other choices.

Practical issues: numerical performance. Our aim is to ensure that the stochastic EM parameter Markov chains mix well. Among the factors that influence mixing (as measured by the decay rate of auto-correlations along the parameter Markov chains), we found three key ones to be the number of particles, the length of the Metropolis chains, and the number of iterations in the overall EM algorithm. Given our experiments, we found that setting moderate numbers for the first two (we set $S_i = 50T_i$

Figure A1: Comparing Metropolis Hastings and Sequential Monte Carlo in the balanced panel used in Arellano, Blundell and Bonhomme (2017, ABB)



Notes: 6-wave balanced sample from the PSID used in ABB, 1999-2009. The graphs show the quantile derivatives of the persistent income component η_{it} with respect to η_{it-1}, averaged over ages in the sample. In the left graph we show the result obtained using a Metropolis Hastings, using the codes from ABB. In the right graph we show the results obtained using the Sequential Monte Carlo algorithm. The two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

particles, as indicated above, and we run each Metropolis chain for 50 iterations), and relatively large numbers for the third (we run the stochastic EM for 2000 iterations), gave best performance given computation constraints in our short panel data setting.

B Numerical comparison with ABB

The SMC approach differs from the Metropolis Hastings method that was used in ABB. Here we compare the income persistence implied by SMC and Metropolis Hastings, when using the original 6-wave balanced panel from ABB.

In Figure A1 we show the nonlinear income persistence predicted by the algorithm using SMC, and compare it to the estimates based on the Metropolis Hastings algorithm from ABB. We see that the results are little affected by the change in method. In particular, we see that, for households with a low persistent income component, high shocks are associated with less income persistence, and for households with a high persistent income component, low shocks are associated with more income persistence. These patterns differ from the implications of a linear process such as a random walk, where income persistence would be flat, independent of both the income level and the income shock. Formally, the income persistence measure proposed by ABB is, in the case of the persistent income component η_{it},

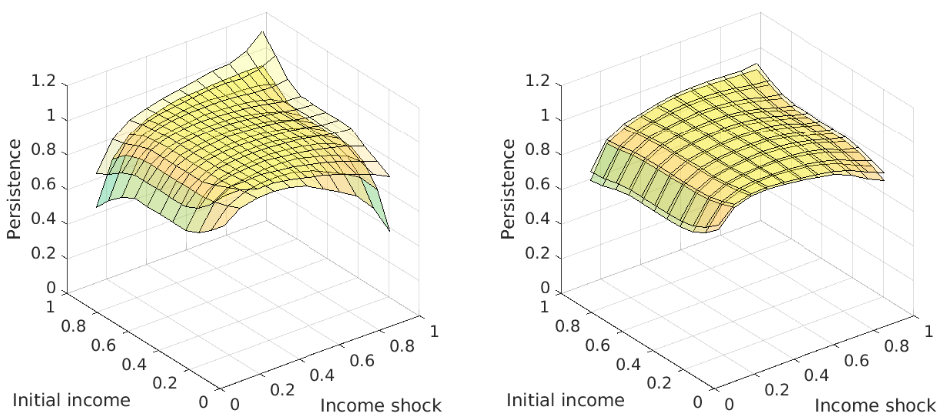
$$\rho(\eta, age, \tau) = \frac{\partial Q_\eta(\eta, age, \tau)}{\partial \eta}, \quad \tau \in (0,1), \tag{A4}$$

where Q_η is the quantile function appearing in (4).[3]

[3] Note that $\rho(\eta, age, \tau)$ also depends on age, which we average out in Figure A1.

Figure A2: Pointwise numerical stability bands of nonlinear persistence estimates

(a) Metropolis Hastings (b) Sequential Markov Chain



Notes: 6-wave balanced sample from the PSID used in ABB, 1999-2009. The graphs show the quantile derivatives of the persistent income component η_{it} with respect to η_{it-1}, averaged over ages in the sample, and evaluated pointwise at the 2.5th and 97.5th percentiles over 200 runs of the stochastic EM algorithm, using different seeds every time. In the left graph we show the result obtained using a Metropolis Hastings sampler, using the codes from ABB. In the right graph we show the results obtained using the Sequential Monte Carlo algorithm. The two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

The income persistence results reported in ABB are based on comparing various estimation runs, and selecting the one that provides the highest value of the likelihood. However, compared to Metropolis Hastings used in ABB, we found the SMC approach to be more effective at reducing the numerical instability across estimation runs. To illustrate this, in Figure A2 we report numerical stability bands that indicate the variability of income persistence estimates obtained from 200 runs of our estimation algorithm using different seeds, based on the two different sampling methods. In the left graph of the figure we report results based on Metropolis Hastings. In the right graph we report results based on the SMC algorithm we rely on in this paper. The SMC results show substantially less numerical variability.

Lastly, although reported estimates in ABB appear reliable in the shorter balanced sample, in our experience increasing the number of households and the length of the panel makes it more challenging to rely on Metropolis Hastings for sampling. In contrast, we found our SMC implementation to remain numerically stable in such cases.

C Which features of the consumption policy rule can be identified?

Consider a structural policy rule of the form

$$C = g(X, \nu),$$

where ν, of unrestricted dimension, is independent of X. To simplify the presentation we assume that X is scalar. In this paper, C denotes consumption, and X contains all state variables, including the income components. Denote the conditional quantile function of C given X as $Q(X, \tau)$. Hence, for U uniform independent of X, we can write

$$C = Q(X, U).$$

We are interested in moments of the marginal effects

$$\Delta_x C = \frac{\partial g(x, \nu)}{\partial x}.$$

The key challenge is that, while Q is identified from data on (C, X), g is generally not.

Average responses. We have, under standard conditions,

$$\mathbb{E}\left[\Delta_x C\right] = \frac{\partial}{\partial x} \mathbb{E}\left[g(x, \nu)\right],$$

hence

$$\mathbb{E}\left[\Delta_x C\right] = \frac{\partial}{\partial x} \mathbb{E}\left[C \mid X = x\right],$$

or, equivalently,

$$\mathbb{E}\left[\Delta_x C\right] = \frac{\partial}{\partial x} \mathbb{E}\left[Q(x, U)\right],$$

that is,

$$\mathbb{E}\left[\Delta_x C\right] = \mathbb{E}\left[\frac{\partial}{\partial x} Q(x, U)\right].$$

Hence, average marginal effects are identified, irrespective of the dimensionality of ν and the monotonicity properties of g.

Variance of responses. By Theorem 2.1 in Hoderlein and Mammen (2007) we have

$$\mathbb{E}\left[\Delta_x C \mid X = x, C = Q(x, \tau)\right] = \frac{\partial}{\partial x} Q(x, \tau),$$

for all τ and x. We thus can write

$$\frac{\partial}{\partial x} Q(x, U) = \mathbb{E}\left[\Delta_x C\right] + V,$$

where

$$V = \frac{\partial}{\partial x} Q(x, U) - \mathbb{E}\left[\Delta_x C\right].$$

Now, V has mean zero, and variance

$$\begin{aligned} \operatorname{Var}\left(\frac{\partial}{\partial x} Q(x, U)\right) &= \operatorname{Var}\left(\mathbb{E}\left[\Delta_x C \mid X = x, C = Q(x, U)\right]\right) \\ &= R^2 \operatorname{Var}\left(\Delta_x C\right), \end{aligned}$$

where R^2 corresponds to the nonparametric regression of $\Delta_x C$ on C and X. Hence, the variance of $\frac{\partial}{\partial x} Q(x, U)$ underestimates the variance of $\Delta_x C$, by an amount that depends on how well C and X explain $\Delta_x C$.

For example, if ν is scalar and has a monotone effect on g, then $R^2 = 1$ and the variances are equal. In that case, $Q = g$, and g is identified. More generally, even though g is may not be identified, the mean of $\frac{\partial g(x, \nu)}{\partial x}$ is identified and one can compute a lower bound on the variance of $\frac{\partial g(x, \nu)}{\partial x}$.

D Additional tables and figures

D.1 Tables and figures for Section 2

Table A1: Additional descriptive statistics about the unbalanced panel

	(1) Waves 1	(2) Waves 2	(3) Waves 3	(4) Waves 4	(5) Waves 5	(6) Waves 6	(7) Waves 7
Age	38.29	39.66	40.58	40.95	40.90	40.08	38.25
	(10.51)	(10.70)	(10.13)	(9.55)	(9.23)	(8.52)	(6.71)
Education	4.88	4.95	5.00	5.06	5.05	5.16	5.14
	(1.09)	(1.10)	(1.10)	(0.99)	(1.05)	(0.95)	(0.98)
Kids	1.22	1.05	1.08	0.90	1.02	1.17	1.35
	(1.16)	(1.15)	(1.17)	(1.00)	(1.22)	(1.08)	(1.01)
Food	10,224.82	10,297.24	10,231.33	10,417.36	10,618.86	10,873.81	10,339.80
	(5,618.54)	(4,871.19)	(4,884.57)	(5,322.02)	(5,205.25)	(5,295.29)	(5,566.94)
Non-durables (excl. food)	24,446.69	25,271.07	27,640.10	26,705.96	27,597.78	29,553.66	27,365.30
	(23,423.94)	(14,975.07)	(20,170.04)	(19,519.81)	(18,453.14)	(18,044.51)	(18,732.42)
Total Non-durables	34,818.81	35,657.00	37,929.68	37,137.98	38,269.81	40,427.48	37,731.48
	(26,171.72)	(17,197.60)	(22,674.04)	(22,345.36)	(21,752.17)	(20,778.80)	(21,432.31)
Home equity	94,353.18	93,634.64	134445.57	142168.95	146854.98	144322.37	145431.99
	(221908.96)	(157549.09)	(218194.70)	(196533.44)	(231684.48)	(171917.47)	(182450.48)
Negative Equity Dummy	0.03	0.01	0.02	0.02	0.03	0.02	0.01
	(0.16)	(0.12)	(0.13)	(0.12)	(0.16)	(0.15)	(0.10)
Wealth (excl. home)	236379.23	151718.99	192237.00	207947.79	245846.12	149537.06	144836.15
	(1.85e+06)	(452508.81)	(480574.21)	(1.03e+06)	(713417.52)	(437249.71)	(607971.51)
Total wealth	369397.05	283854.13	387068.75	414928.35	464594.00	349645.50	352285.71
	(2.26e+06)	(648961.96)	(714758.54)	(1.30e+06)	(1.00e+06)	(604613.97)	(791042.37)
Labor income	105504.37	106842.60	121094.14	134196.63	136728.34	118852.50	117218.18
	(131690.40)	(90,625.35)	(131051.85)	(226750.06)	(132471.38)	(65,851.87)	(53,500.79)
Net income	83,800.29	84,063.03	92,061.60	100974.16	101662.12	90,869.67	90,045.96
	(80,287.10)	(57,270.53)	(80,307.77)	(132837.00)	(79,228.69)	(42,442.62)	(34,698.92)
Observations	1002	668	484	263	223	177	299

Notes: PSID, 2005-2017. Means of variables, standard deviations in parentheses.

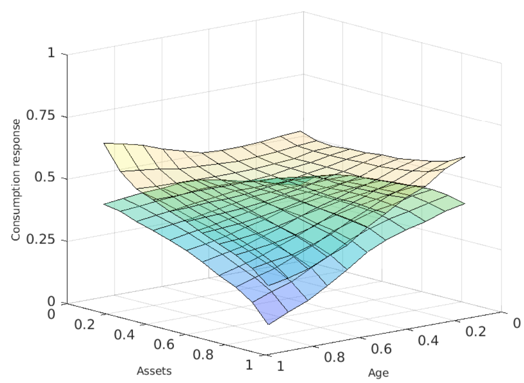
Table A2: Additional descriptive statistics about the main sample, including negative assets

	(1) 2005	(2) 2007	(3) 2009	(4) 2011	(5) 2013	(6) 2015	(7) 2017
Food	10,632.46	10,584.68	10,231.23	10,517.05	10,701.87	11,154.70	11,761.09
	(5,299.94)	(5,480.66)	(4,985.32)	(5,039.51)	(5,575.75)	(5,262.43)	(5,514.86)
Non-durables (excl. food)	28,005.27	29,138.06	27,784.64	28,336.56	30,089.00	29,597.75	28,312.56
	(18,936.74)	(19,416.39)	(18,768.90)	(17,696.35)	(17,860.16)	(17,018.40)	(14,572.47)
Total Non-durables	38,669.31	39,750.28	38,081.61	38,921.75	40,869.60	40,824.46	40,119.38
	(21,699.98)	(22,033.74)	(21,113.38)	(20,391.50)	(20,440.32)	(19,538.10)	(17,414.63)
Home equity	150404.41	156582.18	117029.77	97,240.09	91,229.90	94,851.48	108298.61
	(212201.08)	(224409.31)	(192280.79)	(161856.05)	(146908.40)	(135356.38)	(135913.11)
Negative Equity Dummy	0.01	0.01	0.07	0.08	0.06	0.02	0.02
	(0.12)	(0.12)	(0.26)	(0.28)	(0.24)	(0.14)	(0.12)
Wealth (excl. home)	188962.86	255179.00	230841.97	201148.10	183919.95	203580.64	272524.94
	(683870.01)	(964936.86)	(874673.16)	(497471.78)	(476877.85)	(519691.15)	(1.01e+06)
Total wealth	411875.17	470628.74	384224.26	314392.55	279919.36	298432.12	368142.89
	(940132.05)	(1.20e+06)	(1.07e+06)	(617956.79)	(578419.27)	(590056.80)	(1.01e+06)
Labor income	122972.70	124391.48	126510.00	123237.46	121745.86	120544.04	125475.14
	(139187.13)	(143195.31)	(182296.90)	(119741.17)	(118132.57)	(72,546.62)	(66,226.60)
Net income	93,504.28	94,804.55	95,893.52	95,289.90	94,087.25	92,224.02	95,572.56
	(83,501.16)	(86,771.32)	(109386.52)	(73,204.74)	(71,919.39)	(46,205.59)	(43,329.49)
Observations	1397	1684	1616	1399	1269	1192	968

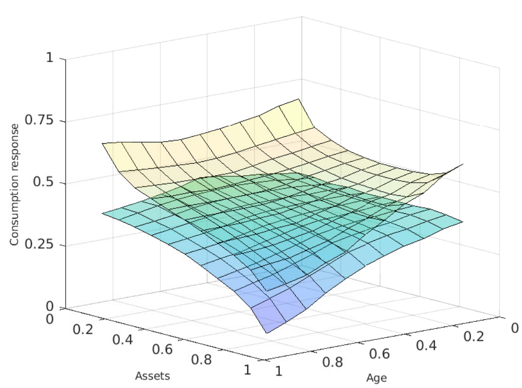
Notes: PSID, 2005-2017. Means of variables, standard deviations in parentheses.

Figure A3: Consumption responses at various quantiles, confidence bands

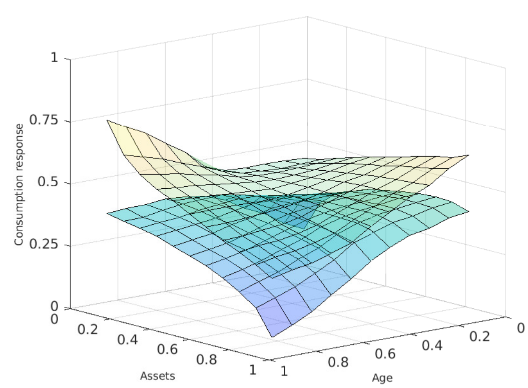
(a) Average



(b) Bottom tercile



(c) Top tercile



Notes: See the notes to Figures 1 and 2. Bootstrapped pointwise 95% confidence bands clustered at the household level.

D.2 Tables and figures for Section 5

Table A3: Summarizing heterogeneity across types, nonparametric bootstrap

A. 90th vs 10th percentile of ξ

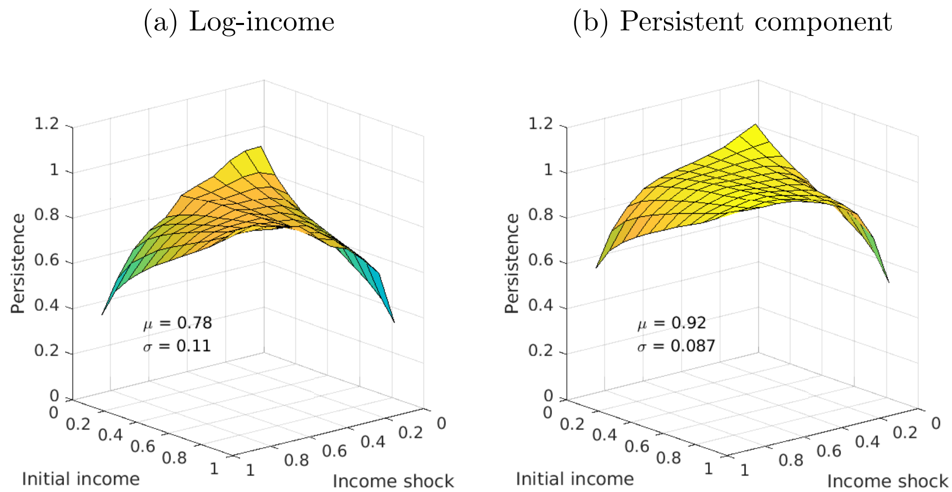
	Young, low assets	Old, high assets	Δ
High ξ	0.31	0.22	0.09
	[0.08,0.48]	[0.04,0.44]	[-0.15,0.22]
Low ξ	0.48	0.21	0.27
	[0.27,0.68]	[0.04,0.42]	[0.01,0.43]
Δ	-0.17	0.01	-0.18
	[-0.56,0.08]	[-0.27,0.28]	[-0.54,0.08]

B. 75th vs 25th percentile of ξ

	Young, low assets	Old, high assets	Δ
High ξ	0.36	0.21	0.15
	[0.19,0.46]	[0.08,0.40]	[-0.02,0.24]
Low ξ	0.45	0.21	0.24
	[0.27,0.58]	[0.09,0.39]	[0.02,0.34]
Δ	-0.09	0.00	-0.09
	[-0.27,0.04]	[-0.13,0.14]	[-0.25,0.04]

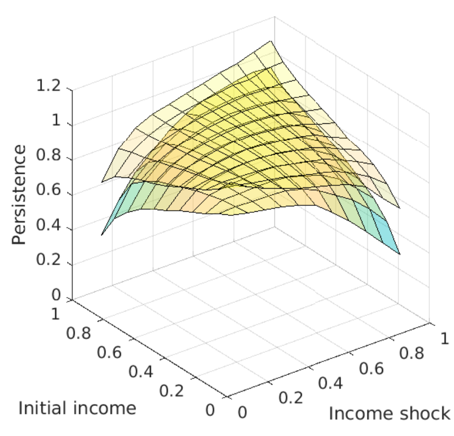
Notes: See the notes to Figure 4. Here we report average consumption responses for young and low assets households compared to old and high assets households, for different percentiles of heterogeneity ξ_i in consumption. Values are calculated by evaluating the average consumption response for households at a fixed percentile of ξ_i when assets and age are fixed at the τth percentile. Reported values for young and low assets households are then shown by averaging over $\tau \in (0, 0.5)$. Reported values for old and high assets households are then shown by averaging over $\tau \in (0.5, 1)$. Nonparametric bootstrap 95% confidence intervals clustered at the household level based on 200 replications are shown in brackets.

Figure A4: Nonlinear income persistence



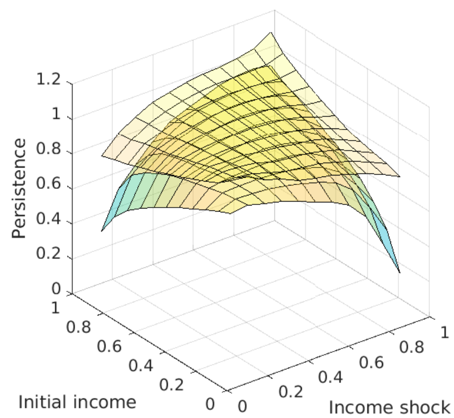
Notes: PSID, 2005-2017 sample, disposable income, dual earners from an alternative perspective. The left graph shows quantile derivatives of log-income with respect to lagged log-income. The right graph shows quantile derivatives of the persistent latent component η_{it} with respect to η_{it-1}, model estimated using sequential Monte Carlo with a stochastic EM algorithm. The two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

Figure A5: Nonlinear persistence in η_{it}, 95% pointwise confidence bands (parametric bootstrap)



Notes: Pointwise 95% confidence bands based on the parametric bootstrap. 200 replications.

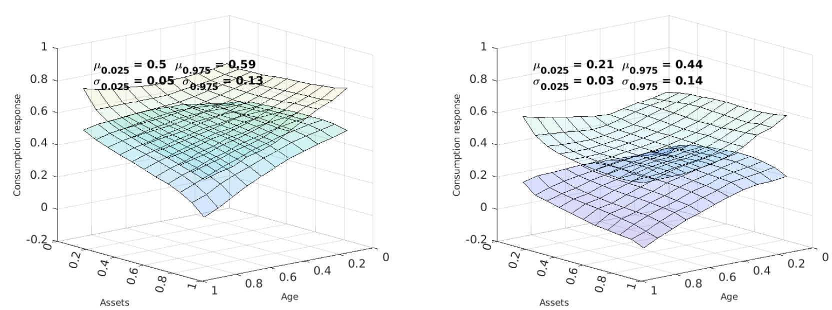
Figure A6: Nonlinear persistence in η_{it}, 95% pointwise confidence bands (nonparametric bootstrap)



Notes: Pointwise 95% confidence bands based on nonparametric bootstrap. 200 replications. Bootstrap is clustered at the household level.

Figure A7: Average insurance in model with and without heterogeneity 95% pointwise confidence bands (nonparametric bootstrap)

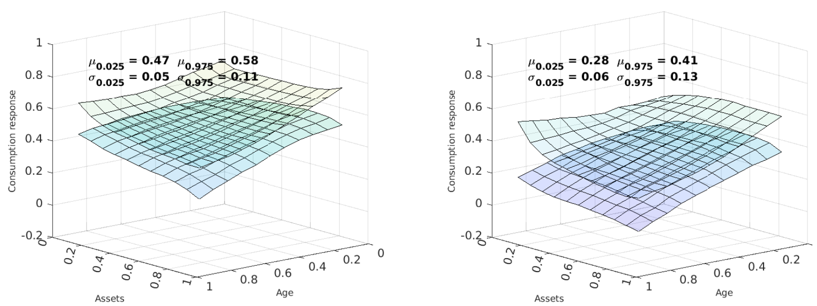
(a) Without heterogeneity (b) With heterogeneity



Notes: Pointwise 95% confidence bands based on nonparametric bootstrap. 200 replications. Bootstrap is clustered at the household level.

Figure A8: Average insurance in model with and without heterogeneity 95% pointwise confidence bands (parametric bootstrap)

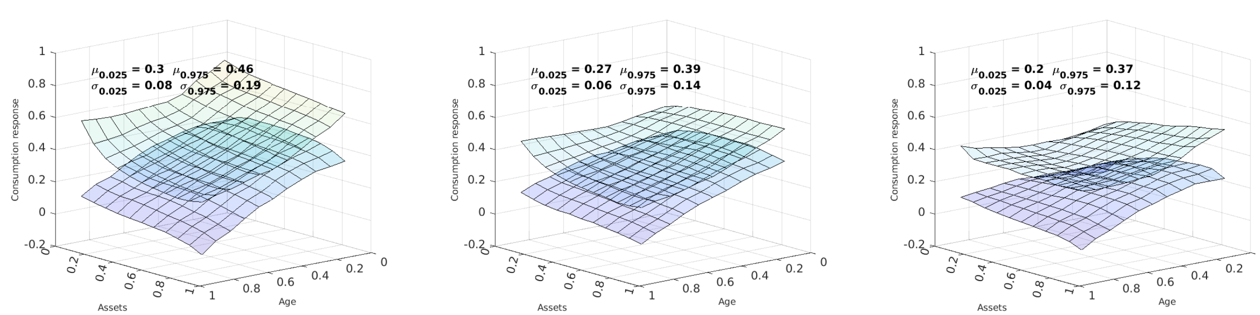
(a) Without heterogeneity (b) With heterogeneity



Notes: Pointwise 95% confidence bands based on parametric bootstrap.

Figure A9: Heterogeneity in consumption responses, 95% pointwise confidence bands (parametric bootstrap)

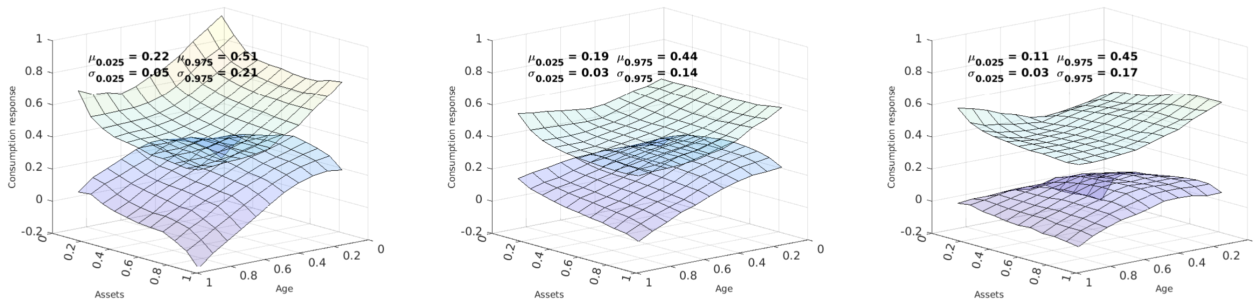
(a) 10th percentile (b) 50th percentile (b) 90th percentile



Notes: Pointwise 95% confidence bands based on parametric bootstrap. 200 replications.

Figure A10: Heterogeneity in consumption responses, 95% pointwise confidence bands (nonparametric bootstrap)

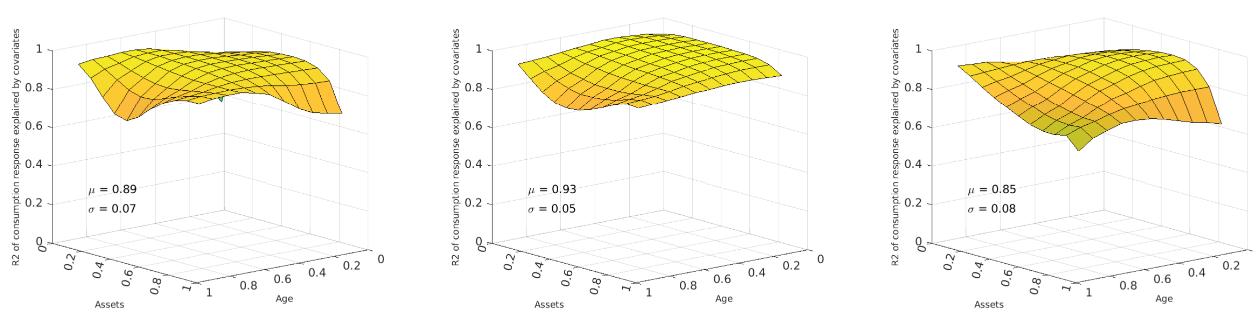
(a) 10th percentile (b) 50th percentile (b) 90th percentile



Notes: Pointwise 95% confidence bands based on nonparametric bootstrap. 200 replications. Bootstrap is clustered at the household level.

Figure A11: Heterogeneity in residual variation of consumption responses

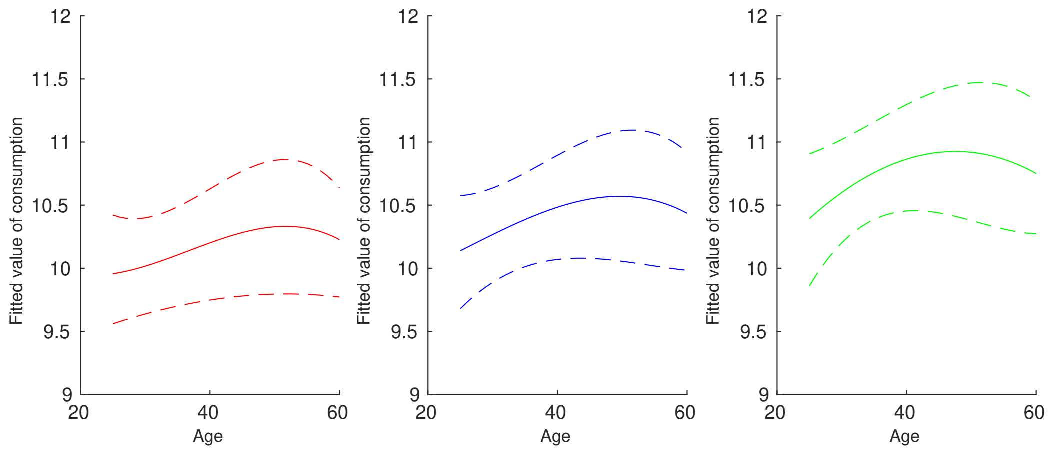
(a) 10th percentile (b) 50th percentile (c) 90th percentile



Notes: See the notes to Figure 5. The figure shows an upper bound on the proportion of the variation in consumption responses to η_{it} explained by the average consumption response, conditional on age and assets, see Section C of the appendix. The various graphs corresponds to different percentiles of ξ_i.

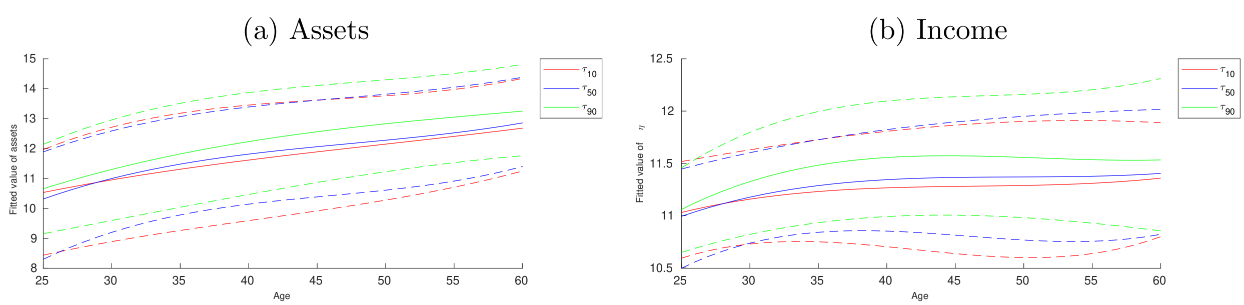
D.3 Tables and figures for Section 6

Figure A12: Life-cycle profile of log-consumption, for different percentiles of unobserved types



Notes: Average non-residualized log-consumption, for different ages and percentiles of ξ_i (10%, Median, 90%). The dashed lines show the age-specific and ξ_i-specific 10th and 90th percentiles of log-consumption.

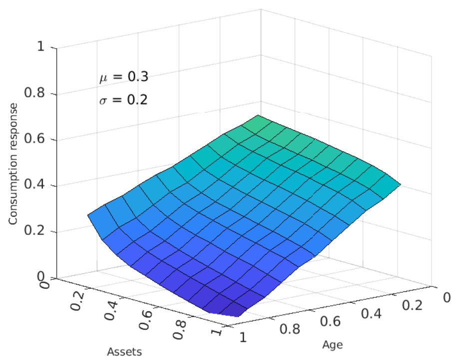
Figure A13: Life-cycle profiles of log-assets and log-income, for different percentiles of unobserved types



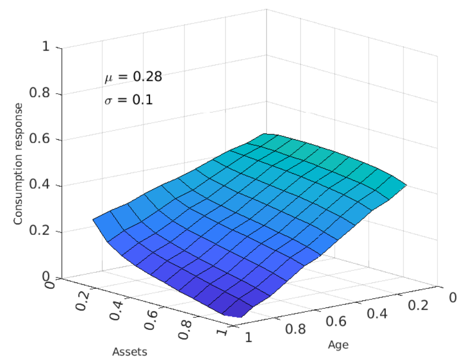
Notes: Average non-residualized log-assets and persistent latent component of log-income, for different ages and percentiles of ξ_i (10%, Median, 90%). The dashed lines show the age-specific and ξ_i-specific 10th and 90th percentiles for each outcome measure.

Figure A14: Heterogeneity in consumption responses, model with heterogeneity in assets

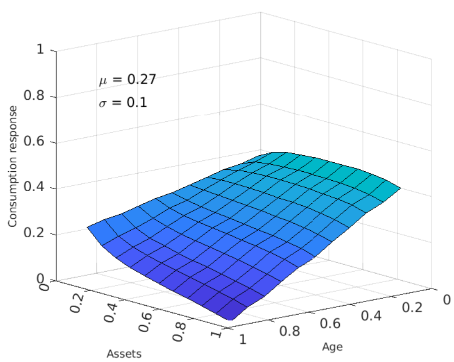
(a) 10th Percentile



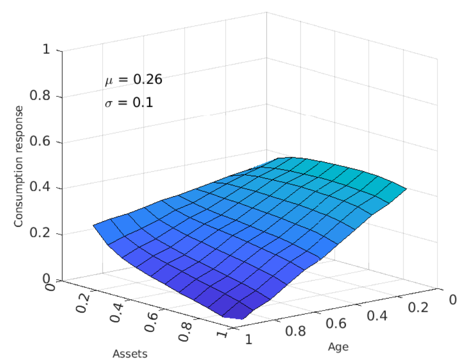
(b) 25th Percentile



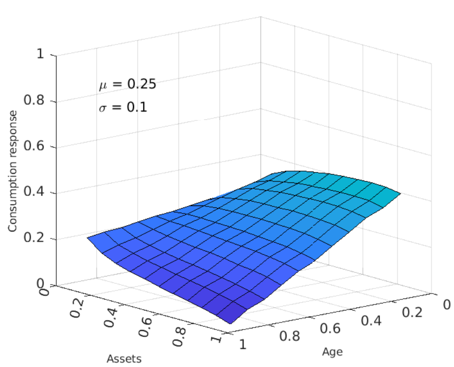
(c) 50th Percentile



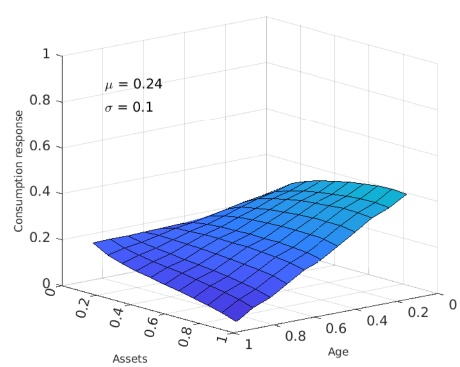
(d) Average



(e) 75th Percentile



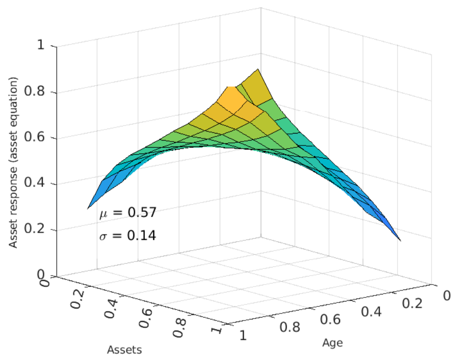
(f) 90th Percentile



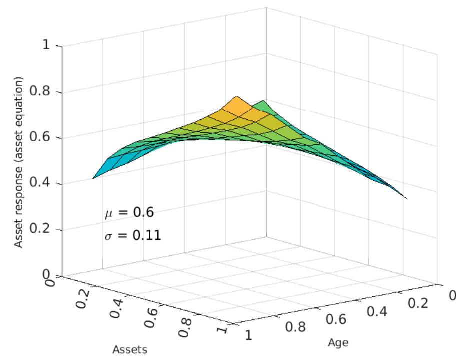
Notes: See the notes to Figure 5. The results are based on a model with latent heterogeneity ξ_i in consumption and assets. Here we report the results by percentiles of heterogeneity ξ_i.

Figure A15: Heterogeneity in assets dynamics, model with heterogeneity in assets

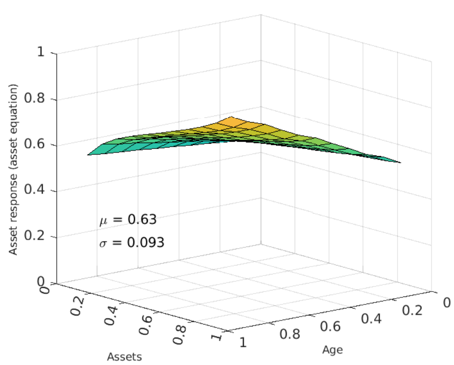
(a) 10th Percentile



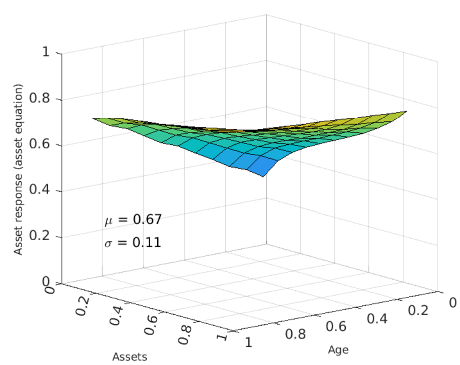
(b) 25th Percentile



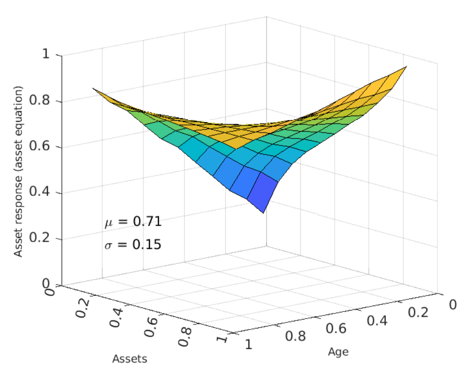
(c) 50th Percentile



(d) 75th Percentile



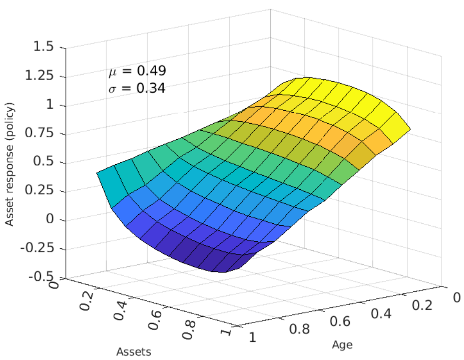
(e) 90th Percentile



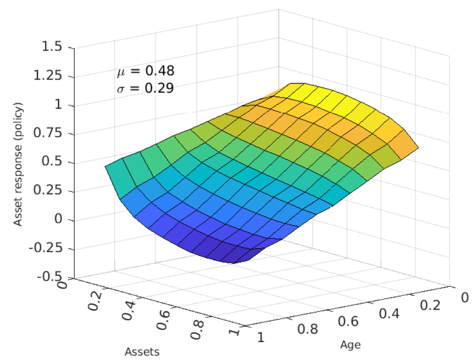
Notes: The figure shows the average total derivative of log-assets with respect to lagged log-assets, conditional on lags of log-assets, income components, log-consumption, age, and the latent type. Here we report the results by percentiles of heterogeneity ξ_i.

Figure A16: Heterogeneity in assets responses, model with heterogeneity in assets

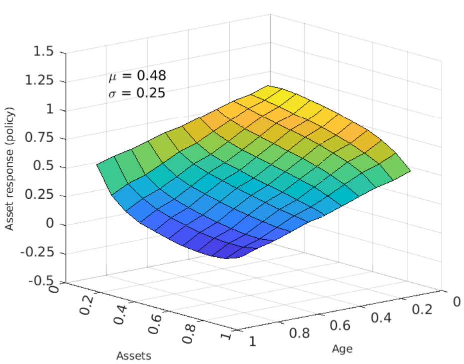
(a) 10th Percentile



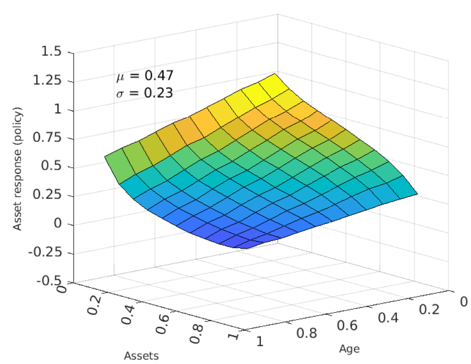
(b) 25th Percentile



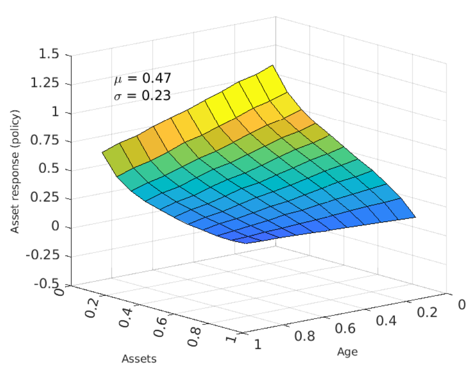
(c) 50th Percentile



(d) 75th Percentile



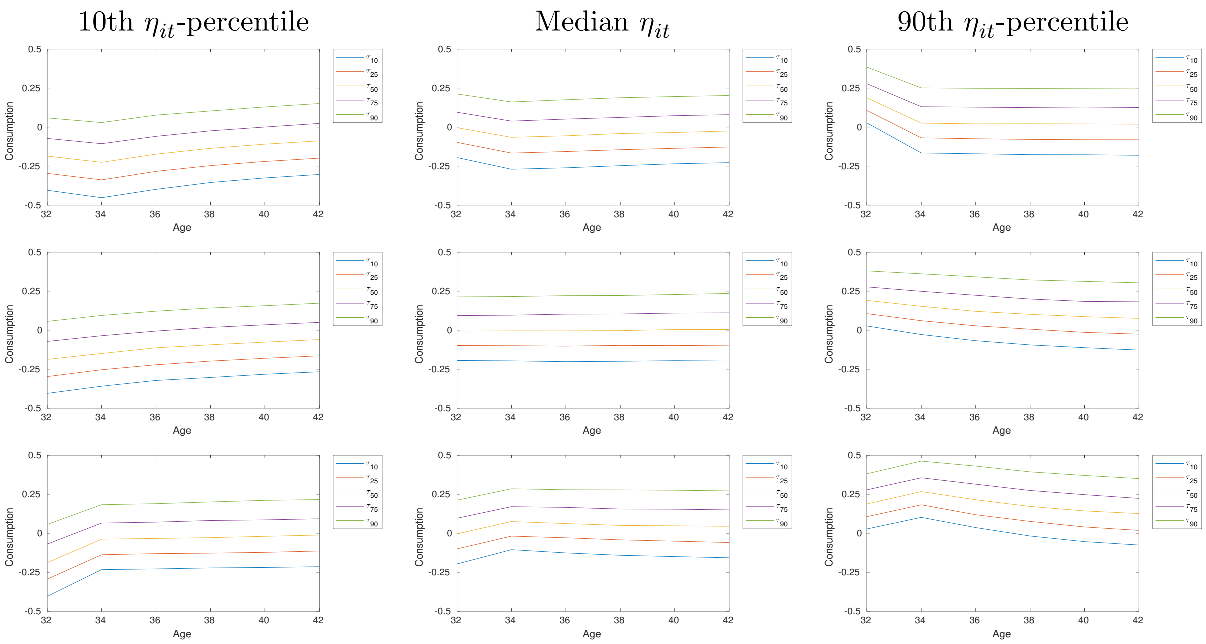
(e) 90th Percentile



Notes: The figure shows the average total derivative of log-assets with respect to lagged η*, conditional on lags of log-assets, income components, age and the latent type. Here we report the results by percentiles of heterogeneity* ξ_i*.*

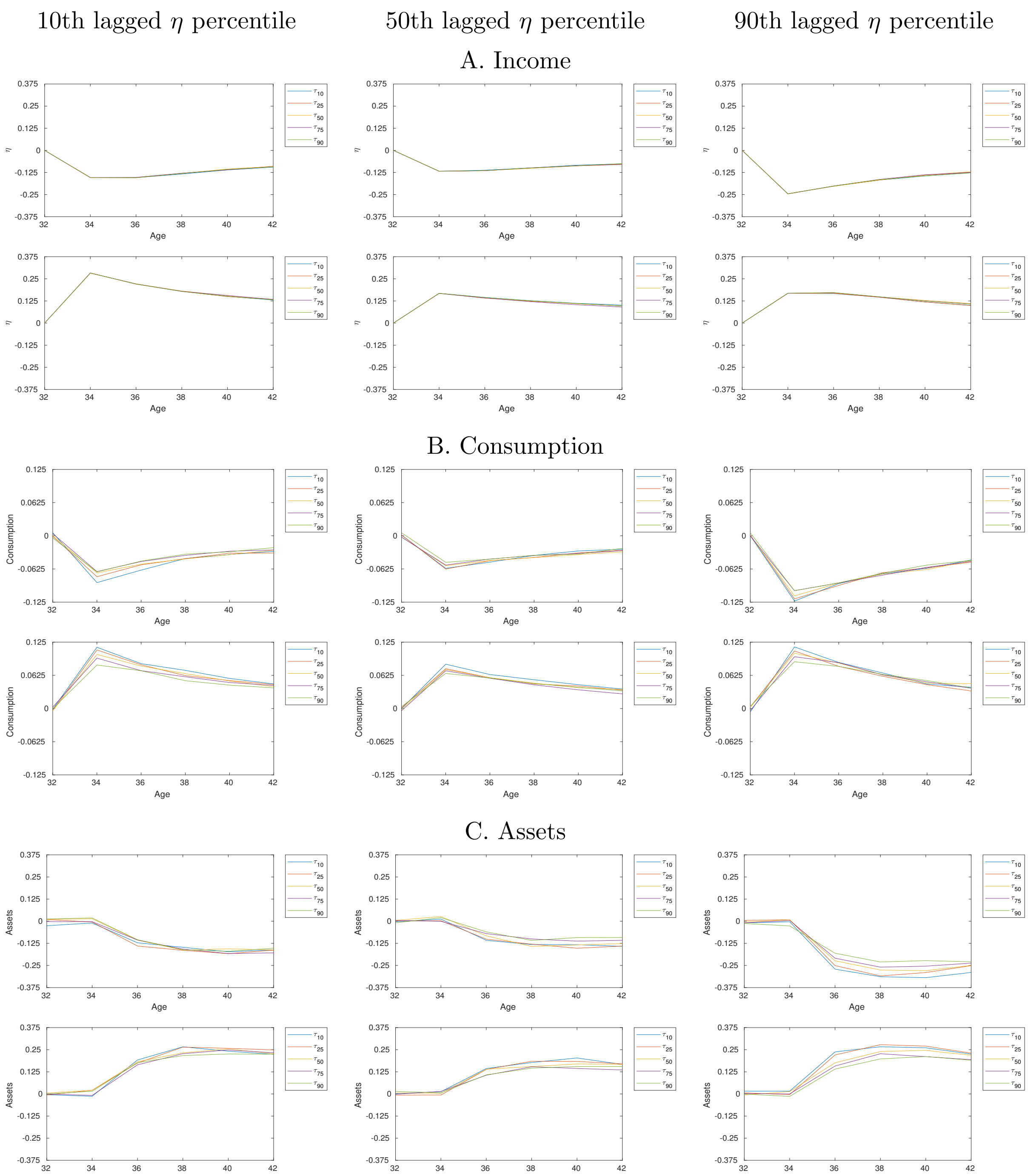
D.4 Tables and figures for Section 7

Figure A17: Heterogeneity in impulse responses: consumption trajectories



Notes: Trajectories shown for shocks at the 10th (top subpanel), 50th (middle subpanel) and 90th (bottom subpanel) percentiles.

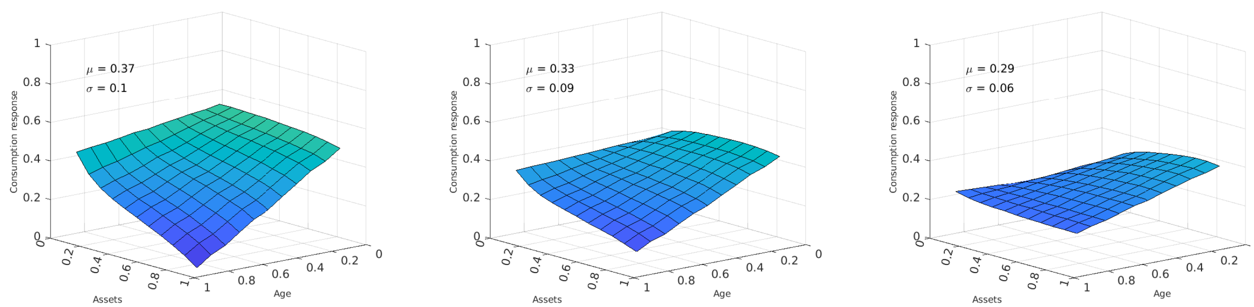
Figure A18: Heterogeneity in impulse responses, model with heterogeneity in assets



Notes: Impulse responses shown for shocks at the 10th (top subpanels) and 90th (bottom subpanels) percentiles, relative to median.

Figure A19: Heterogeneity in consumption responses based on 19 knots

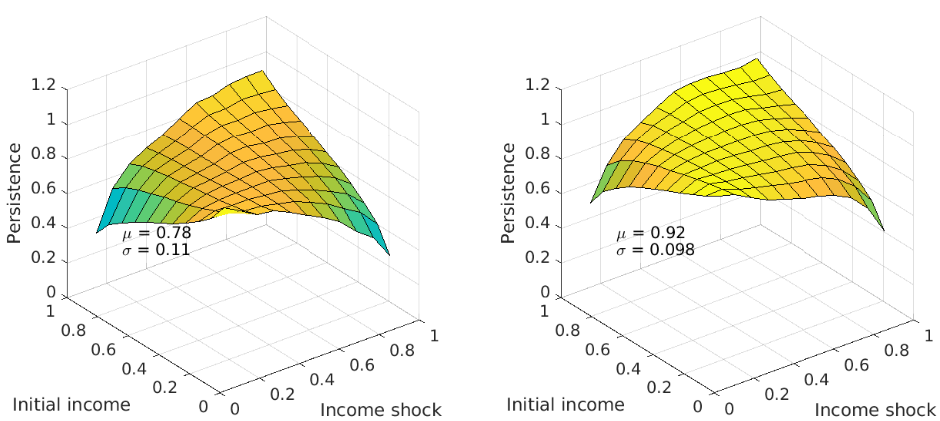
(a) 10th percentile (b) 50th percentile (b) 90th percentile



Notes: See the notes to Figure 5. In this figure we use 19 knots in estimation. For our baseline results in Figure 5 we used 11 knots.

Figure A20: Nonlinear income persistence, labor income

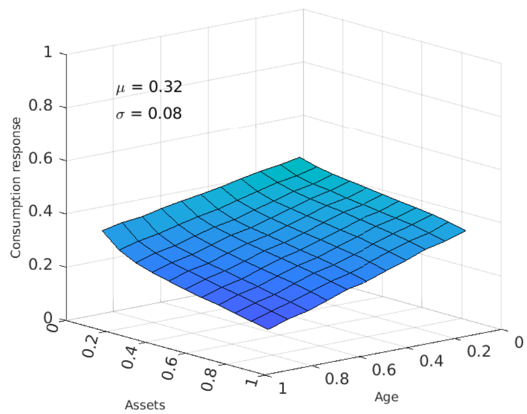
(a) Log-income (b) Persistent component



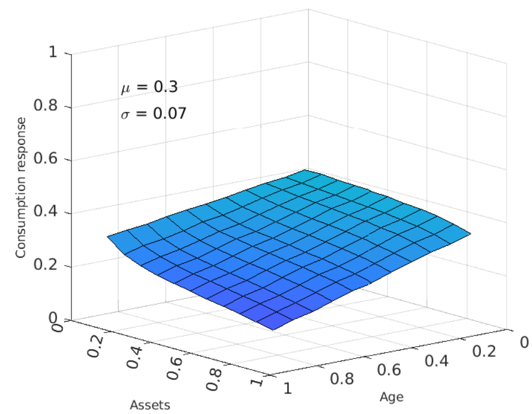
Notes: PSID, 2005-2017 sample, household labor income. The left graph shows quantile derivatives of log-income with respect to lagged log-income. The right graph shows quantile derivatives of the persistent latent component η_{it} with respect to η_{it-1}, model estimated using sequential Monte Carlo with a stochastic EM algorithm. The two horizontal axes show percentiles of η_{it-1} ("initial income") and conditional percentiles of η_{it} given η_{it-1} ("income shock"), respectively.

Figure A21: Heterogeneity in consumption responses, labor income

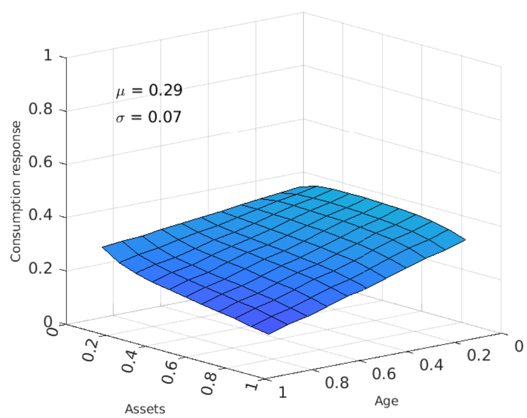
(a) 10th percentile



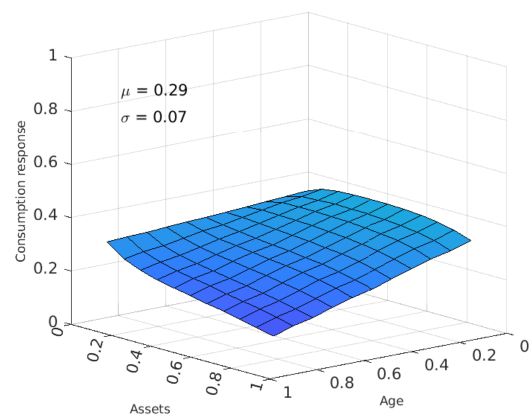
(b) 25th percentile



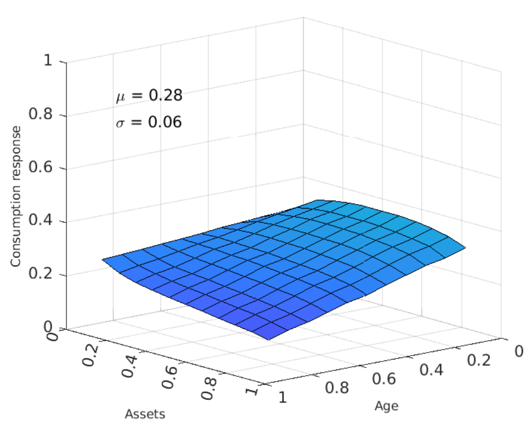
(c) Median



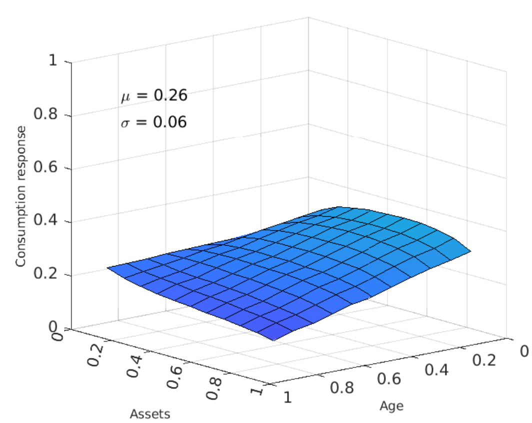
(d) Mean



(e) 75th percentile



(f) 90th percentile



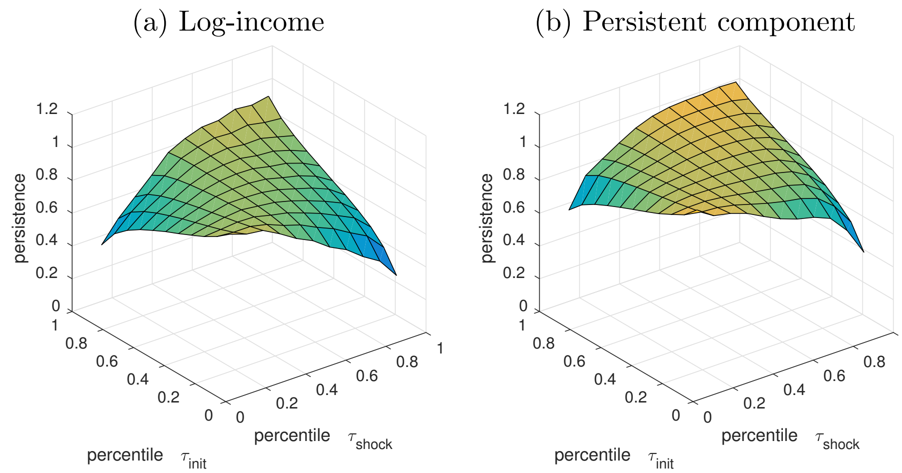
Notes: See the notes to Figure 5. Household labor income. Here we report the results by percentiles of heterogeneity ξ_i *in consumption.*

Table A4: Descriptive statistics about the main sample without dual earners restriction

	(1) 2005	(2) 2007	(3) 2009	(4) 2011	(5) 2013	(6) 2015	(7) 2017
Food	10,739.58	10,629.51	10,294.05	10,523.17	10,728.10	11,195.42	12,049.05
	(5,602.43)	(5,617.21)	(5,131.22)	(5,066.30)	(5,701.24)	(5,290.79)	(5,872.98)
Non-durables (excl. food)	27,847.42	28,588.68	27,339.21	27,883.75	29,368.71	29,549.63	27,907.17
	(23,625.00)	(20,214.59)	(19,243.79)	(19,340.37)	(19,382.75)	(19,794.72)	(16,322.99)
Total Non-durables	38,625.09	39,265.87	37,731.22	38,532.60	40,205.65	40,843.42	40,002.57
	(26,482.07)	(23,195.91)	(21,701.55)	(21,932.92)	(22,194.81)	(22,563.15)	(19,525.89)
Home equity	168358.82	176300.55	136154.76	121783.91	116463.18	118089.14	133596.76
	(262246.00)	(283429.69)	(207398.72)	(175957.85)	(165962.97)	(152785.21)	(150451.14)
Negative Equity Dummy	0.01	0.01	0.03	0.03	0.03	0.01	0.01
	(0.08)	(0.10)	(0.17)	(0.17)	(0.16)	(0.10)	(0.10)
Wealth (excl. home)	211547.79	279544.52	278268.96	268297.79	260584.33	291511.77	346692.78
	(1.09e+06)	(1.16e+06)	(1.02e+06)	(704058.90)	(656770.57)	(765195.01)	(1.07e+06)
Total wealth	461075.98	521015.10	457730.90	411004.15	383583.34	409600.91	464296.87
	(1.51e+06)	(1.51e+06)	(1.23e+06)	(841641.87)	(762537.13)	(834428.29)	(1.09e+06)
Labor income	121962.17	120618.90	124276.87	121469.61	127809.57	124560.33	129948.36
	(155403.02)	(143009.67)	(181097.03)	(129296.65)	(241344.84)	(172615.50)	(115383.29)
Net income	93,333.70	93,262.22	95,306.98	95,476.95	98,924.10	95,790.47	99,431.91
	(92,700.09)	(86,962.24)	(108935.54)	(82,869.14)	(145844.22)	(98,144.09)	(69,882.14)
Observations	1730	2004	1843	1578	1436	1321	1090

Notes: PSID, 2005-2017. Means of variables, standard deviations in parentheses.

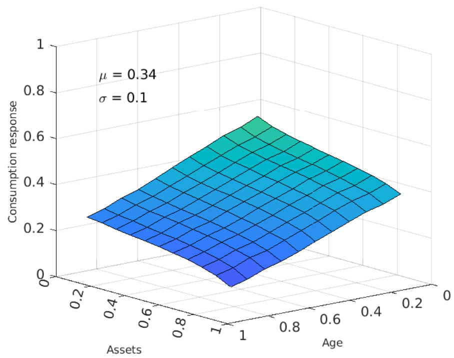
Figure A22: Nonlinear income persistence, no dual earners restriction



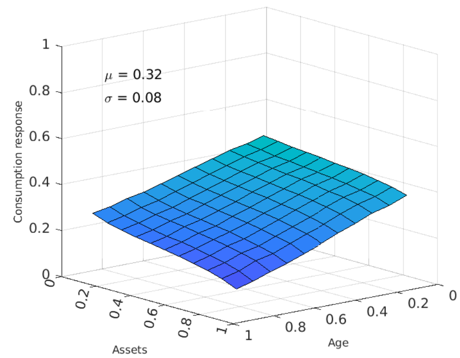
Notes: PSID sample, no dual earners restriction. See the notes to Figure 3.

Figure A23: Heterogeneity in consumption responses, no dual earners restriction

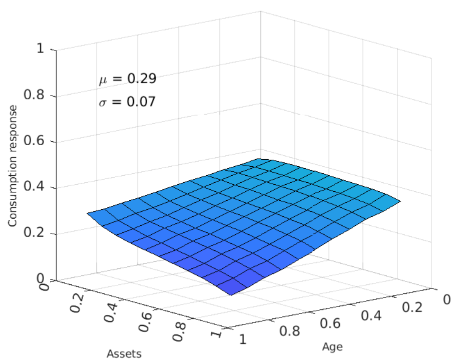
(a) 10th Percentile



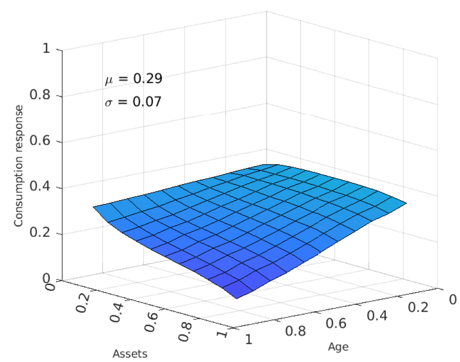
(b) 25th Percentile



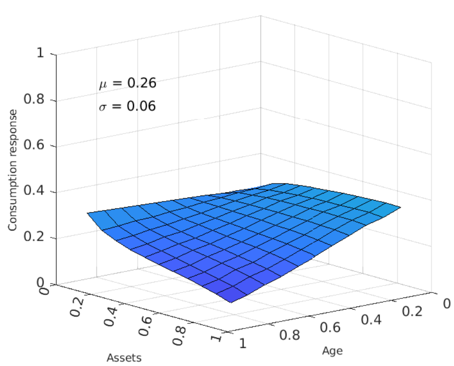
(c) 50th Percentile



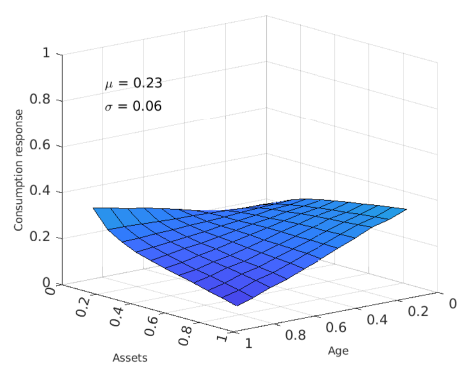
(d) Average



(e) 75th Percentile

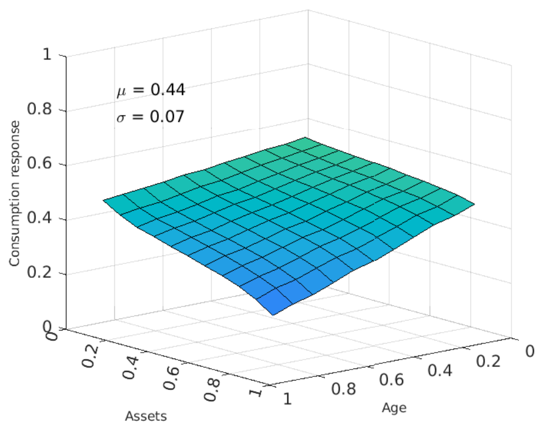


(f) 90th Percentile



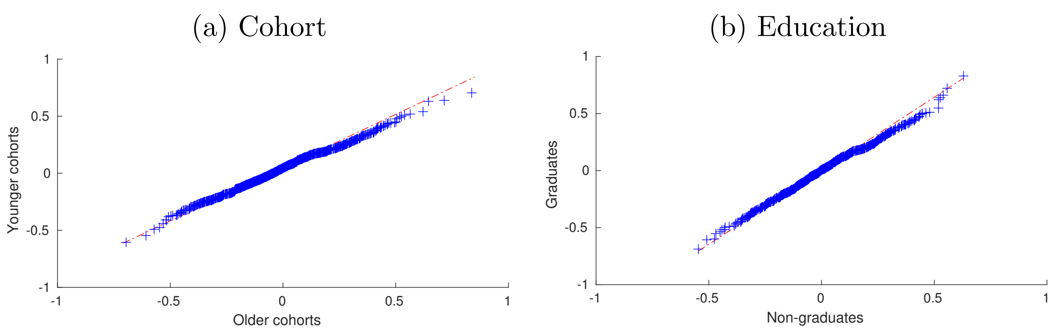
Notes: See the notes to Figure 4. No dual earners restriction. Here we report the results by percentiles of heterogeneity ξ_i in consumption.

Figure A24: Average consumption responses to labor income



Notes: PSID, 2005-2017 sample, dual earners, labor income. The graph shows the average derivative of log-consumption with respect to the persistent latent component η_{it} in a model without unobserved heterogeneity ξ_i in consumption. The two horizontal axes show age and assets percentiles, respectively.

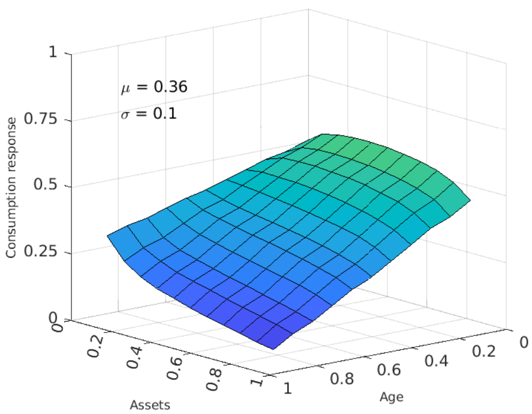
Figure A25: Quantile-quantile plots for ξ_i by observables



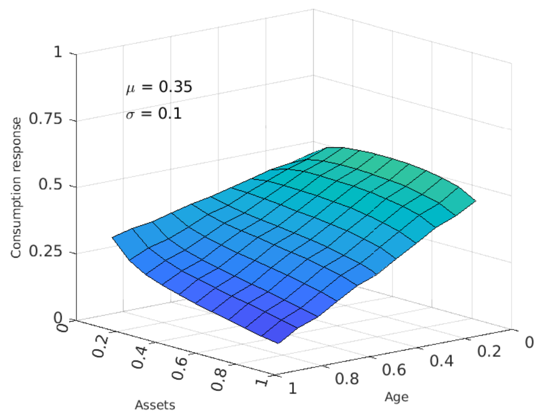
Notes: Quantile-quantile plots shown for (a) graduates and non-graduates (b) born before 1969 and born after 1969. The graphs show the quantiles of ξ_i indicated on the x-axis against the quantiles of ξ_i indicated on the y-axis.

Figure A26: Heterogeneity in consumption responses controlling for education

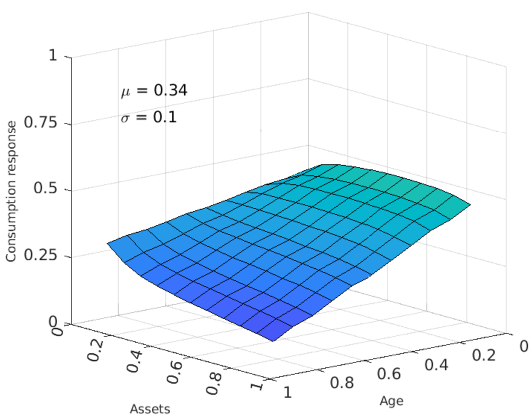
(a) 10th percentile



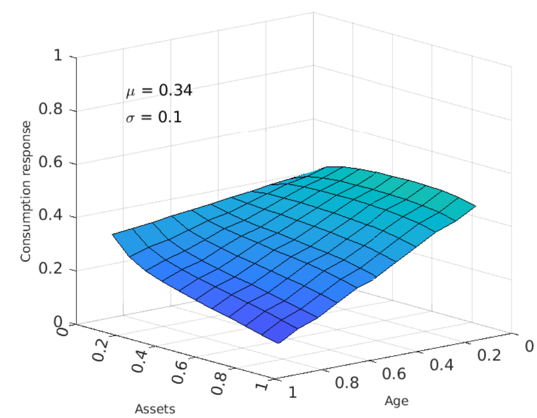
(b) 25th percentile



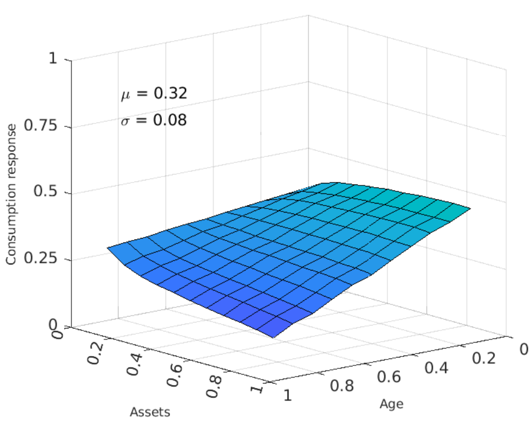
(c) Median



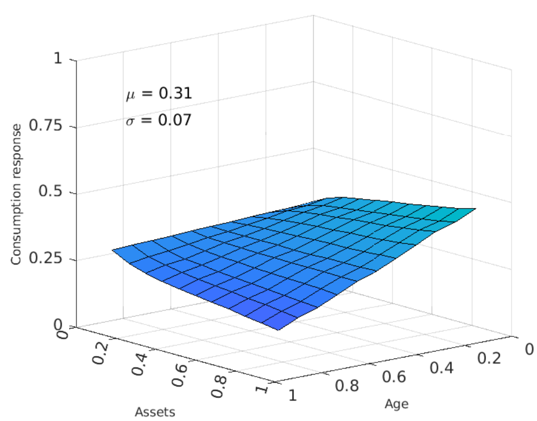
(d) Mean



(e) 75th percentile



(f) 90th percentile



Notes: See the notes to Figure 4. We report average derivatives in a regression that includes a full set of interactions with a binary higher education indicator. Here we report the results by percentiles of heterogeneity ξ_i in consumption.