Discussion Paper Series

IZA DP No. 18935

September 2026

Perceived versus Realized Income Dynamics: Evidence from Subjective Expectations

Henrique S. Basso
Banco de España and CEMFI

Olympia Bover
CEMFI

Julio Galvez
CUNEF Universidad

Laura Hospido
Banco de España, CEMFI
and IZA@LISER

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18935 September 2026

Perceived versus Realized Income Dynamics: Evidence from Subjective Expectations*

Abstract

Household consumption and saving depend on perceived income risk, yet macroeconomic models typically discipline income processes using realized data. We show that these objects differ systematically. We develop a framework to recover flexible income processes from subjective expectations reported over fixed intervals, accounting for coarse elicitation and focal responses. Households perceive income as more persistent and less dispersed, with weaker state-dependent nonlinear dynamics than realized histories imply. These differences survive adjustments for reporting, measurement error, and horizon differences. In a life-cycle model, they affect precautionary saving and wealth accumulation; when households save according to perceived dynamics but face realized shocks, consumption insurance falls substantially.

JEL classification

D15, D31, D84, E21, C23

Keywords

subjective expectations, income dynamics, nonlinear persistence, consumption, precautionary saving, life-cycle, panel data

Corresponding author

Laura Hospido
laura.hospido@bde.es

* A previous version of the paper circulated under the title "Subjective Income Expectations and Consumption Dynamics". We are grateful to Manuel Arellano, Richard Blundell, Stéphane Bonhomme, Christopher Carroll, Jesús Fernández-Villaverde, Pamela Giustinelli, Luigi Guiso, Nezih Guner, Fatih Guvenen, Gizem Koşar, Thierry Magnac, Benjamin Moll, Wilbert van der Klaauw, and Siqi Wei for valuable suggestions. We thank José María Casado for his early contributions to the project and Jiayi Wen for excellent research assistance. We are also grateful to seminar and conference participants at CEMFI, CUNEF Universidad, Erasmus University Rotterdam, IER-Hitotsubashi University, Encuentro de Economía Pública (Valencia), EEA-ESEM (Cologne and Barcelona), International Panel Data Conference (Vilnius), 6th Conference on Household Finance and Consumption (Dublin), SEHO meeting (Copenhagen), the Conference on Econometric Methods and Empirical Analysis of Micro Data in honor of Manuel Arellano (Madrid), Workshop on consumption and saving over the life-cycle (Laugarvatn, Iceland), IAAE (Oslo and Lisbon), the Workshop on Subjective Expectations (Lisbon), the Econometric Society World Congress (Seoul), the CEPR Workshop on Frontiers in Measurement and Survey Methods (Calabria), T2M (Montreal), the BSE Summer Forum on Expectations in Dynamic Macroeconomic Models and the BdE-CEMFI Workshop on Household Finances for useful feedback. The views expressed here are those of the authors and, therefore, do not necessarily coincide with those of the Banco de España or the Eurosystem.

1 Introduction

Household consumption and saving decisions depend on perceived income risk, not on the income process an econometrician recovers after the fact. Yet structural models typically discipline income risk using statistical income processes estimated from realized income data, implicitly assuming that the conditional distribution inferred from past outcomes coincides with the one households perceive when making decisions. This raises a simple but fundamental question: is the income process households believe they face the same as the income process an econometrician would infer from their realized income histories? In this paper, we show that the two differ systematically. Combining subjective income expectations and realized income data for the same households, we recover perceived and realized income processes whose persistence, dispersion, and nonlinear dynamics differ substantially. These differences are economically large and have quantitatively important implications for precautionary saving, life-cycle wealth accumulation, and consumption insurance.[1]

Distinguishing between subjective expectations and realized income data is especially important in light of recent evidence that income dynamics are highly nonlinear. Realized income processes exhibit state-dependent persistence, asymmetric shocks, and non-Gaussian income changes (Arellano, Blundell, and Bonhomme, 2017; Guvenen, Karahan, Ozkan, and Song, 2021; De Nardi, Fella, and Paz-Pardo, 2020). These features have become central ingredients of modern heterogeneous-agent models because they shape consumption insurance, wealth accumulation, and inequality (Kaplan and Violante, 2010; De Nardi et al., 2020; Gálvez and Paz-Pardo, 2026). Yet nearly all existing estimates of nonlinear income dynamics are inferred from realized income histories. Whether households perceive the same nonlinear persistence, dispersion, and asymmetry remains largely unknown. Our paper brings subjective expectations into the modern nonlinear income dynamics framework and asks whether the process households perceive ex ante resembles the process an econometrician recovers ex post.

[1]Throughout the paper, we use *income risk* to refer to the full conditional distribution of future income given current information, rather than to its dispersion alone. We characterize this distribution along three main dimensions: persistence, which captures the propagation of current income into future income; dispersion, which captures its spread; and skewness, which captures asymmetry in its shape.

There are two main reasons why these two objects could differ. First, econometricians neither observe households' information sets nor the process by which they form expectations (Jappelli and Pistaferri, 2010). Households may hold private information regarding future income (for instance, upcoming promotions, bonuses, contract renewals, or layoffs) that is unobservable to the researcher. Innovations identified from realized income histories may therefore combine genuinely unanticipated shocks with changes households already expected. Second, expectations may depart from rational benchmarks, with households overweighting salient or recent outcomes and underreacting to base rates (Bordalo, Gennaioli, and Shleifer, 2013, 2018). Subjective expectations may therefore themselves reflect systematic biases. Differences between households' perceived and realized income dynamics may consequently reflect differences between households' and the econometrician's information sets, systematic departures from rational benchmarks, or both. We do not attempt to fully decompose these mechanisms. Instead, we ask whether perceived and realized income dynamics differ after conditioning on the state observed by the econometrician, and whether these differences matter.

We study this question using the Spanish Survey of Household Finances (EFF). The EFF combines a rotating panel of realized household income with detailed information on consumption and wealth and, crucially, probabilistic expectations about future household income for the same households. Respondents allocate ten probability points across five fixed common intervals for household income growth over the following twelve months. This provides unusually rich information about households' perceived income distributions, but it also creates an estimation problem that differs fundamentally from the one faced with realized income data. Realized income is observed as a continuous outcome, whereas subjective expectations reveal only probability mass over a small number of intervals. Without additional structure, these reports identify the perceived distribution only at the survey cutoffs and do not reveal its shape within the intervals.

Comparing perceived and realized income dynamics therefore requires more than applying standard income process estimators to two different datasets. In particular, on the subjective side, we do not observe next-period income or the quantiles of its conditional distribution. To overcome this challenge, we develop an estimation framework that recovers continuous income dynamics from the coarse probabilistic allocations observed in the survey. We consider two broad classes of income processes. First, we use a state-dependent AR(1) process as a transparent benchmark, allowing persistence and dispersion to vary across household characteristics and economic states (Chamberlain and Hirano, 1999; Meghir and Pistaferri, 2004; Browning, Ejrnaes, and Alvarez, 2010; Hospido, 2012; Gu and Koenker, 2017; Almuzara, 2020). Second, we consider a flexible, nonlinear quantile process in the spirit of Arellano et al. (2017), which allows nonlinear persistence, conditional dispersion, and skewness to vary with current income and the rank of future shocks.

For the state-dependent AR(1), we derive the probabilities over the survey intervals implied by the latent continuous income process and estimate the resulting model using a multinomial likelihood. The multinomial layer accommodates the discreteness inherent in expressing an underlying continuous probability distribution using ten integer-valued probability points. The nonlinear model poses a more difficult problem because standard quantile regression methods require the future outcome to be observed. We instead interpret the reported probability allocations as revealing the bin membership of latent draws from households' perceived conditional income distributions and develop a stochastic pseudo-EM procedure that alternates between imputing latent income realizations consistent with the reported allocations and re-estimating the conditional quantile process. Because fixed common intervals identify the subjective distribution directly only at the survey cutoffs, recovery of its full within-bin shape is conditional on the maintained parametric or sieve specification; we make this distinction explicit in our identification analysis.

We also separately model systematic focal reporting, motivated by the prevalence of middle-bin bunching in the data: more than 40 percent of respondents assign all probability mass to the central interval. In particular, we introduce a two-regime hurdle model that distinguishes deterministic middle-bin focal responses from allocations generated by the underlying perceived income process, building on de Bresser and van Soest (2013), Kleinjans and van Soest (2014), and Heiss, Hurd, van Rooij, Rossmann, and Winter (2022). We use response behavior in a parallel house price expectations question as an excluded shifter of the focal reporting probability. The validity of this exclusion is a maintained identifying assumption: conditional on household characteristics and wave effects, house price response behavior must affect income expectation responses through reporting behavior rather than through common factors that also affect the underlying perceived income distribution. We discuss this restriction explicitly and examine broader forms of probability distortion as robustness. Together, these methods allow us to recover comparable state-dependent and nonlinear income processes despite the fundamentally different ways in which subjective and realized income dynamics are observed.

The comparison reveals a large and systematic gap between perceived and realized income dynamics. In the state-dependent AR(1) benchmark, average persistence inferred from realized income is about 0.69, whereas perceived persistence is essentially one. Average perceived dispersion is roughly an order of magnitude lower in the baseline subjective specification than in the process estimated from realized income. The nonlinear estimates sharpen this contrast. Realized income displays substantial state dependence: persistence varies considerably across income levels and shock ranks. By contrast, perceived persistence remains high and is remarkably flat across states, while perceived conditional dispersion is substantially lower. Skewness is the dimension along which the two processes are most similar. Accounting for focal reporting changes the subjective skewness profile substantially and makes its state dependence more similar to the realized pattern. The perceived-realized gaps in persistence and dispersion, however, remain substantial.

These gaps are also robust to two natural challenges to the comparison. First, realized household income is itself survey-reported and may contain measurement error. Classical measurement error moves the realized income estimates toward the subjective estimates: at an illustrative reliability ratio of 0.80, the persistence gap falls by about 46 percent and the dispersion gap by about 57 percent. Nevertheless, the two processes remain economically distinct. Second, the EFF realized income histories are observed at three-year intervals, whereas subjective expectations concern income over the following year. Using the Spanish Living Conditions Survey to estimate otherwise comparable income processes at one- and three-year horizons, we find that horizon differences are small relative to the perceived-realized gap.

The differences between perceived and realized dynamics matter for household behavior. We embed the alternative nonlinear income processes in a standard incomplete-markets life-cycle model. The lower dispersion and weaker nonlinearities of the perceived process generate weaker precautionary-saving motives, different life-cycle wealth profiles, and lower cross-sectional wealth dispersion. When preferences are held fixed, the perceived process also implies substantially lower wealth accumulation. Under realized income dynamics, high-income households in particular accumulate larger buffers against downside income risk, producing greater wealth accumulation and cross-sectional wealth dispersion. Average measures of consumption insurance can nevertheless look similar across processes because households achieve that insurance in different ways: under realized dynamics, greater asset accumulation provides more self-insurance, whereas the subjective process produces thinner buffers and more homogeneous consumption responses across households.

We also consider a counterfactual in which households choose saving according to the process they perceive but subsequently experience shocks generated by the process estimated from realizations. This exercise corresponds to interpreting the perceived-realized gap as arising from biased beliefs. Consumption insurance falls substantially, from about 0.78 when households both save and experience shocks according to the realized process to about 0.65 when saving decisions are based on the subjective process but shocks follow realized dynamics. If the perceived-realized gap instead reflects superior private information, however, this counterfactual is not the appropriate interpretation: in that case, processes estimated solely from realizations may overstate the income dispersion relevant for household decisions. Consistent with this possibility, we find that household-specific subjective dispersion predicts subsequent income beyond current income and observable household characteristics and is also associated with consumption and elicited marginal propensities to consume. These exercises do not decisively distinguish biased beliefs from private information, but they show that subjective expectations contain economically relevant information.

The paper builds on a long literature using subjective expectations to study income risk, consumption, and portfolio choice (Guiso, Jappelli, and Terlizzese, 1996; Dominitz and Manski, 1997; Dominitz, 1998, 2001; Pistaferri, 2001; Guiso, Jappelli, and Pistaferri, 2002; Attanasio, Kovacs, and Molnar, 2020; Kovacs, Rondinelli, and Trucchi, 2021; Koşar and Van der Klaauw, 2025), as well as papers that incorporate subjective income expectations into quantitative macroeconomic models (Rozsypal and Schlafmann, 2023; Stoltenberg and Uhlendorff, 2022; Wang, 2023; Koşar and Melcangi, 2025). More recent work has advanced the use of subjective expectations to measure income and earnings risk. Caplin, Gregory, Lee, Leth-Petersen, and Sæverud (2026) introduce a new survey instrument that incorporates job transitions and document that unconditional subjective earnings risk is substantially smaller than estimates based on realized income histories. Arellano, Attanasio, Crossman, and Sancibrián (2026b) develop methods for estimating flexible perceived income processes using household-specific elicitation thresholds and show that subjective income dynamics exhibit heteroskedasticity, skewness, and nonlinear persistence. Finally, Arellano, Attanasio, Borella, De Nardi, and Paz-Pardo (2026a) use earnings and employment expectations from the New York Fed's Survey of Consumer Expectations to identify earnings and employment dynamics while accounting for selection and unobserved heterogeneity.

This question is distinct from, yet is complementary to, the questions explored in this recent work. Rather than asking only how dispersed households perceive future income to be or whether subjective income dynamics are themselves nonlinear, we ask whether the nonlinear features recovered from realized income histories—the features increasingly used to discipline quantitative household models—are the same features households perceive when making decisions. We show that they are not. The perceived-realized gap concerns not only conditional dispersion, but also the dynamic structure of income risk: the strong state dependence and nonlinear propagation present in realized income histories are substantially weaker in households' perceived income processes.

Our contribution is both methodological and substantive. First, we develop a framework for recovering state-dependent and nonlinear income processes from probabilistic expectations reported over fixed common intervals. This requires addressing an estimation problem that does not arise with realized income data and differs from elicitation designs based on household-specific thresholds: the underlying continuous conditional distribution is observed only through coarse probability allocations, potentially contaminated by focal reporting behavior. Our multinomial and stochastic pseudo-EM approaches recover the income process within the maintained model from these reports, while the hurdle specification separately models deterministic focal responses. This makes the framework applicable to the broad class of large-scale surveys that elicit probabilities over common fixed intervals.

Second, using this framework, we provide a same-population comparison of perceived and realized income dynamics. Rather than comparing scalar measures of subjective and realized uncertainty, we characterize the perceived-realized gap along the dimensions that have become central in the modern income dynamics literature: nonlinear persistence, conditional dispersion, and skewness. We also show that the gap survives the principal reporting, measurement, and time-aggregation concerns raised by this comparison. Finally, we quantify the economic implications of the alternative processes for precautionary saving, wealth accumulation, and consumption insurance. Taken together, the results show that income processes recovered from realized histories need not describe the income risks households perceive, and may therefore mischaracterize the income risk relevant for household decisions.

The rest of the paper is organized as follows. Section 2 describes the data and the probabilistic income expectations measure. Section 3 presents the income-process specifications—a state-dependent AR(1) model and a nonlinear quantile model—and defines the main features of income dynamics we study. Section 4 develops our estimation strategy for subjective income processes, introducing the multinomial framework, the stochastic pseudo-EM procedure, and the hurdle model for focal reporting behavior.

Section 5 presents the empirical comparison between income dynamics estimated from subjective expectations and from realized income data. Section 6 studies the implications of our findings for consumption insurance and wealth accumulation, while Section 7 examines the predictive and behavioral content of subjective expectations. Finally, Section 8 concludes. Additional material is provided in the online appendix.

2 Data

We use data from the EFF, a representative survey designed to study household income and wealth in Spain. It contains detailed information on household assets, liabilities, income, consumption, and demographic characteristics. Because the survey was specifically designed to study household wealth, it oversamples wealthy households using information from wealth tax records. This oversampling allows for a richer analysis of fluctuations in the upper tail of the income distribution than would be possible in a conventional survey.

The EFF has a rotating-panel design, with a refreshment sample introduced in each wave to preserve representativeness over time[2]; moreover, panel households are observed for up to four waves. The survey was conducted every three years through the 2020 wave and every two years thereafter. Because the question on subjective income expectations is available starting in 2014, we focus on the 2014, 2017, 2020, and 2022 waves. Following Blundell, Pistaferri, and Saporta-Eksten (2016), we restrict the sample to households whose heads are aged 25–65, with at least one household member active in the labor market, whose main source of income is not pensions, and that are observed in at least two EFF waves. Households in our estimation sample are therefore observed in between two and four waves. The resulting sample contains 8,823 household-wave observations. Table 1 reports the distribution of households by the number of waves in which they are observed. Appendix A provides additional variable definitions and summary statistics.

[2]Roughly 40 to 50 percent of households in each wave are kept in the subsequent waves.

Table 1: Panel structure of the EFF estimation sample

Number of waves observed	Households	Share of households (%)	Household-wave observations
2	1,980	57.10	3,960
3	1,089	31.40	3,267
4	399	11.50	1,596
Total	3,468	100.0	8,823

Notes: The sample includes households whose heads are aged 25–65, with at least one household member active in the labor market, whose main source of income is not pensions, and that are observed in at least two EFF waves. The estimation sample uses the 2014, 2017, 2020, and 2022 waves.

2.1 Eliciting subjective income expectations

Starting in 2014, the EFF included a question designed to elicit households' expectations about future income. Household heads were asked to distribute ten points across five possible scenarios for total household income growth over the following twelve months. Thus, the horizon of the expectations question is always one year, even though consecutive EFF waves in our sample are two or three years apart.

The exact wording of the question is as follows:

We are interested in knowing how you think the total annual income of your household will change in the next 12 months. Share out 10 points among the five options given below, assigning more points to the options you think are more likely (assign 0 points to options you think are impossible):

- *Drop of more than 10%*
- *Drop between 2% and 10%*
- *Approximately steady (falls or rises of no more than 2%)*
- *Increase between 2% and 10%*
- *Increase of more than 10%*
- *Do not know*
- *No answer*

Several features of the income expectations question are worth emphasizing. First, households are asked to allocate probability mass across five pre-specified income-growth intervals. Second, respondents distribute 10 points rather than 100, reducing the cognitive burden associated with probabilistic reporting. Third, the question elicits beliefs using a probability-mass format rather than a cumulative distribution format.[3]

Although probabilistic survey questions may impose additional cognitive demands, a large body of evidence shows that respondents are generally willing and able to answer them meaningfully.[4] In particular, the literature has emphasized that probabilistic expectations tend to be sensible and internally consistent when questions concern well-defined events closely related to respondents' own lives (Van der Klaauw, Bruine de Bruin, Topa, Potter, and Bryan, 2008). Finally, the EFF includes a similar probabilistic expectations question regarding the future value of the household's primary residence, which we later use to help identify reporting behavior and focal responses.

2.2 Response behavior of households

Before discussing the estimation of income processes, we first examine households' response behavior to the subjective income expectations question.[5] Following Bover (2015), we define four indicator variables. *Drop* equals one for households that allocate all 10 points to the first two bins of the income expectations question. *Increase* equals one for households that allocate all 10 points to the last two bins. *Bunching* equals one for households that place all 10 points on the middle bin. Finally, *Other* equals one for households whose responses do not fall into any of the previous three categories.

Table 2 reports summary statistics on households' response behavior. About 41.7% of households bunch all probability mass on the middle bin across the four EFF waves, although this behavior declines over time—from 57.6% in 2014 to 31.8% in 2022.[6]

[3]The survey is administered through computer-assisted face-to-face interviews, with interviewers providing clarification when needed. Respondents are also given a paper version of the question and response options on which they can draft their answers. An automatic prompt appears whenever the responses entered by the interviewer do not sum to 10, in which case respondents are asked to revise their answers. A similarly designed question on house price expectations is administered earlier in the survey.

[4]Approximately 2.5% of households do not respond to the subjective income expectations question across all waves.

[5]We examine in Appendix A cross-question coherence between the probabilistic income expectations question and a qualitative question asking whether households expect their income to increase, decrease, or remain unchanged. The two measures display substantial coherence: households expecting income to increase (decrease) report higher (lower) mean expected income in the probabilistic question, while those expecting income to remain unchanged report substantially lower dispersion.

[6]We also calculate the persistence of bunching in the middle, and find that persistence is 46.7%. Finally, we also examine the proportion of households bunching at any given bin. We find that 58% of

Table 2: Household response behavior across EFF waves

Wave	Drop	Increase	Other	Bunching
2014	.0924	.0899	.2416	.5762
2017	.0794	.1236	.3176	.4793
2020	.1124	.0869	.4562	.3445
2022	.0888	.1438	.4490	.3184
Pooling all waves	.0942	.1107	.3772	.4179

Note: Proportion of households exhibiting the following response patterns in the subjective income expectations question. *Drop* equals one for households that allocate all 10 points to the two lowest income-growth intervals. *Increase* equals one for households that allocate all 10 points to the two highest income-growth intervals. *Other* equals one for households who do not exhibit the specific behaviors highlighted, for example, those who spread points into multiple bins. *Bunching* equals one for households that allocate all 10 points to the middle bin.

We next examine whether these response patterns are correlated with household demographic characteristics using linear probability models for each indicator (Table 3). Households with secondary education are less likely to bunch and more likely to spread probability mass across multiple bins, while female household heads are more likely to bunch. Bunching also becomes more prevalent with age and among households with permanent labor contracts, but is less common among the self-employed. Most notably, response behavior is highly correlated across domains: households that bunch in the house price expectations question are about 25 percentage points more likely to bunch in the income expectations question. This strong cross-domain correlation is consistent with persistent reporting styles or cognitive factors affecting probabilistic survey responses, although it need not exclusively reflect reporting behavior. Because both questions concern nominal outcomes, common household-level beliefs or uncertainty about inflation could also contribute to this association. We return to this issue when discussing the exclusion restriction in Section 4.2.

Finally, we plot the average probability distributions implied by households' responses, both across all waves and separately by survey wave (Figure 1). Overall, respondents place most probability mass on the approximately stable income growth interval, which receives about 60% of the total probability mass on average. However, the concentration of beliefs around income stability declines markedly over time, falling

households bunch at any bin, and the persistence of this behavior is approximately 61%.

Table 3: Demographic correlates of household response behavior

Variables	(1) Bunching	(2) Drop	(3) Increase	(4) Other
Secondary education	-0.031*	-0.030**	0.027**	0.034*
	(0.018)	(0.012)	(0.014)	(0.019)
Higher education	-0.027	-0.045***	0.008	0.064***
	(0.022)	(0.012)	(0.015)	(0.022)
Female household head	0.066***	-0.003	-0.018*	-0.046***
	(0.016)	(0.010)	(0.010)	(0.016)
Reference person aged 35–45	0.125***	0.005	-0.024	-0.105*
	(0.045)	(0.022)	(0.038)	(0.058)
Reference person aged 45–55	0.147***	0.039*	-0.034	-0.152***
	(0.043)	(0.022)	(0.038)	(0.057)
Reference person aged 55–65	0.232***	0.063**	-0.059	-0.237***
	(0.047)	(0.025)	(0.040)	(0.059)
Permanent labor contract	0.060***	-0.029**	-0.053***	0.021
	(0.023)	(0.013)	(0.016)	(0.022)
Self-employed	-0.072***	0.034**	0.009	0.029
	(0.022)	(0.016)	(0.019)	(0.025)
Risk averse	0.124***	-0.000	-0.009	-0.115**
	(0.043)	(0.023)	(0.029)	(0.056)
Medium risk tolerance	0.072	-0.008	-0.020	-0.044
	(0.045)	(0.025)	(0.031)	(0.059)
Bunching in house price expectations	0.249***	0.006	0.020	-0.275***
	(0.019)	(0.011)	(0.012)	(0.016)
Constant	0.132**	0.106***	0.145***	0.617***
	(0.065)	(0.034)	(0.047)	(0.083)
Observations	8,823	8,823	8,823	8,823
R-squared	0.110	0.017	0.011	0.121

Notes: This table reports estimates from linear probability models relating household response patterns to demographic characteristics. *Drop* equals one for households that allocate all 10 points to the two lowest income-growth intervals. *Increase* equals one for households that allocate all 10 points to the two highest income-growth intervals. *Other* equals one for households who do not exhibit the specific behaviors highlighted, for example, those who spread points into multiple bins. *Bunching* equals one for households that allocate all 10 points to the middle bin. All specifications include wave fixed effects. Robust standard errors are reported in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

from 68% in 2014 to 53% in 2022. At the same time, households increasingly assign probability mass to positive income growth scenarios, particularly to moderate income increases. Although the perceived distribution remains centered around stable income growth, the responses display substantial heterogeneity and time variation in reported income beliefs.

Figure 1 also highlights substantial middle-bin bunching, which may reflect genuinely stable income expectations or reporting behavior. Because the correlation between income and house price responses could also reflect common underlying factors, we treat the reporting style interpretation as a maintained identifying assumption; see Section 4.2.

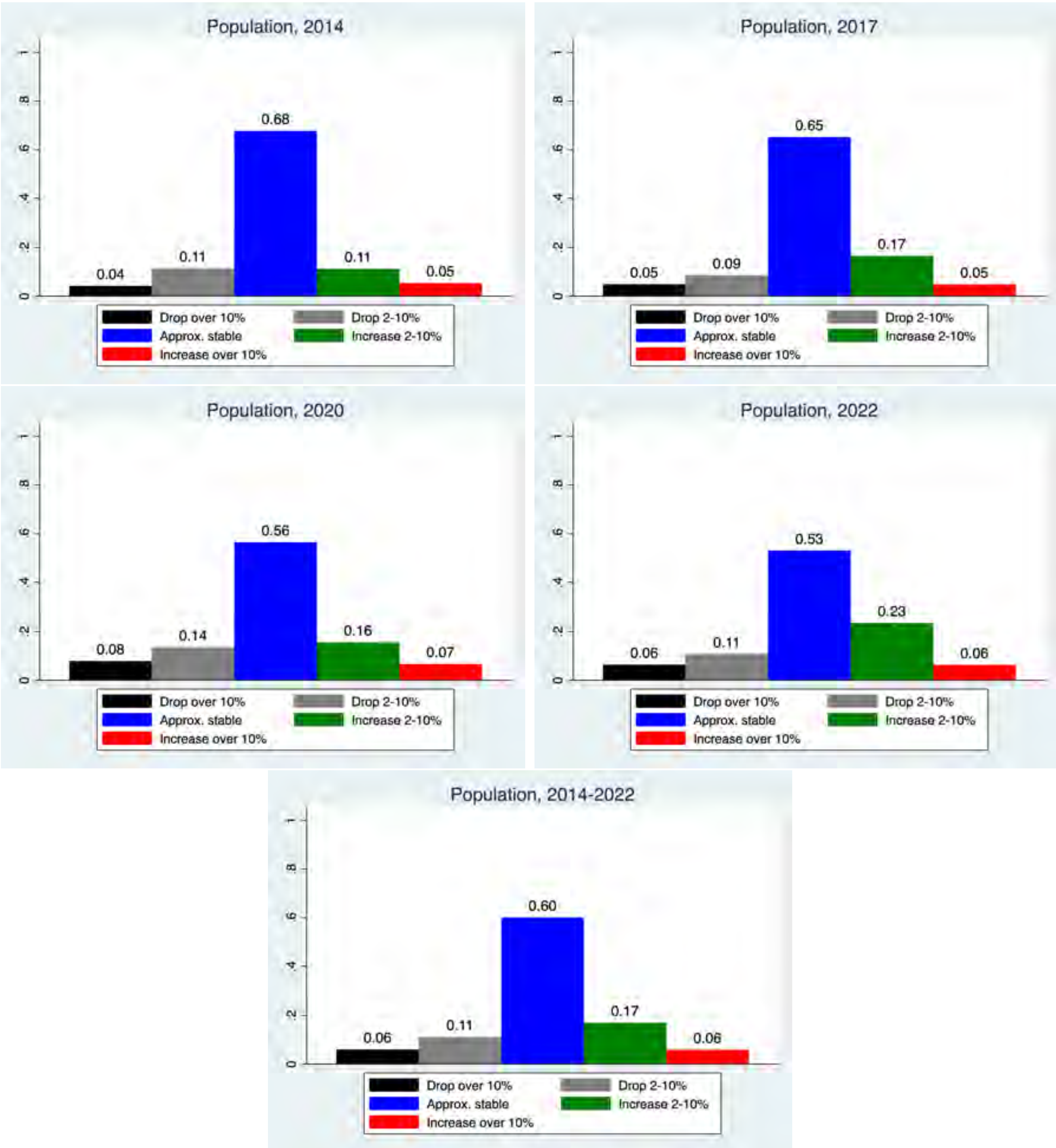
3 Modeling income dynamics

Income processes are central to understanding household consumption and saving decisions and are key inputs in quantitative life-cycle models. These models are widely used to study consumption insurance, precautionary saving, portfolio choice, and household welfare (Kaplan and Violante, 2010; De Nardi et al., 2020; Gálvez and Paz-Pardo, 2026). In this context, income processes summarize the income risk households face when making intertemporal decisions.

Our objective is to compare income dynamics inferred from realized income data with those perceived by households themselves. To do so, we consider two representations of income dynamics commonly used in quantitative macroeconomic models: a state-dependent autoregressive process and a fully nonlinear income process. We estimate each specification using both *realized* income data and *subjective* expectations data, and then compare the implied characteristics of the income processes, including persistence, dispersion, and skewness.

Throughout, we use *income risk* to refer to the full conditional distribution of future income given current information. We characterize income dynamics along three main dimensions: *persistence*, which captures the propagation of the current income state into future income; *dispersion*, which captures the spread of the conditional distribution; and

Figure 1: Average probability distributions of future household income



Notes: These are the histograms of household responses to the income expectation questions, for each of the four waves, and for the entire dataset. Each bin corresponds to one of the five intervals of the subjective income expectations question. Drop over 10%: one-year future income growth drops below 10%. Drop 2-10%: one-year future income growth drops to around 2-10%. Approximately stable: income growth falls between -2% to 2%. Increase 2-10%: one-year future income growth increases from 2 to 10%. Increase over 10%: one-year future income growth rises above 10%.

skewness, which captures its asymmetry.

3.1 State-dependent AR(1) income process

In the state-dependent AR(1) specification, log household income for household i at time t, denoted by y_{it}, evolves according to the following law of motion:

$$y_{it+1} = \rho(\mathbf{X}_{it})\, y_{it} + \mathbf{X}_{it}'\beta + \eta_i + \sigma(\mathbf{X}_{it})\, u_{it+1}, \qquad u_{it+1} \sim N(0,1), \tag{1}$$

where $\mathbf{X}_{it}$ denotes observable household characteristics, u_{it+1} is an income innovation, and η_i captures time-invariant unobserved heterogeneity. Persistence, $\rho(\mathbf{X}_{it})$, and dispersion, $\sigma(\mathbf{X}_{it})$, are allowed to vary with household characteristics. Because u_{it+1} has unit variance, conditional on η_i, $\sigma(\mathbf{X}_{it})$ is the conditional standard deviation of the one-period income innovation, and $\sigma^2(\mathbf{X}_{it})$ is its conditional variance. We therefore use $\sigma(\mathbf{X}_{it})$ as our measure of income dispersion in this specification.[7]

In this sense, our specification is closely related to those considered by Chamberlain and Hirano (1999), Meghir and Pistaferri (2004), Browning et al. (2010), Hospido (2012), Gu and Koenker (2017), and Almuzara (2020), which incorporate different dimensions of heterogeneity in income dynamics.[8]

The state-dependent AR(1) specification allows us to characterize the distributions of persistence and dispersion across household-wave observations. In particular, we com-

[7] The canonical autoregressive process is a special case of equation (1), obtained when $\rho(\mathbf{X}_{it}) = \rho$ and $\sigma(\mathbf{X}_{it}) = \sigma$.

[8] A more general specification would allow ρ_{it} and σ_{it} to remain unrestricted:

$$y_{it+1} = \rho_{it} y_{it} + \mathbf{X}_{it}'\beta + \sigma_{it} u_{it+1}, \qquad u_{it+1} \sim N(0,1).$$

Estimating such a model using realized income data requires sufficiently long panel dimensions (Bonhomme and Denis, 2025). Given the relatively short panel dimension of the EFF, we instead parameterize persistence and dispersion as functions of observable household characteristics. This approach also allows us to identify the observable drivers of heterogeneity in income dynamics across both the cross-sectional and time-series dimensions.

Nonetheless, we also consider an alternative specification in which persistence is homogeneous while dispersion is household-specific:

$$y_{it+1} = \rho y_{it} + \mathbf{X}_{it}'\beta + \sigma_i u_{it+1}, \qquad u_{it+1} \sim N(0,1).$$

Details and empirical results for this alternative specification are reported in Appendix D.

pute the mean and standard deviation of income persistence,

$$\mu_\rho = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho(\mathbf{X}_{it}), \qquad \varsigma_\rho = \sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\rho(\mathbf{X}_{it}) - \mu_\rho\right]^2}, \tag{2}$$

and the mean and standard deviation of income dispersion:

$$\mu_\sigma = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sigma(\mathbf{X}_{it}), \qquad \varsigma_\sigma = \sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\sigma(\mathbf{X}_{it}) - \mu_\sigma\right]^2}. \tag{3}$$

Estimating the parameters of this process using *realized* income data can be carried out using standard likelihood-based methods. Specifically, we adopt a random-effects approach and maximize the average log-likelihood, following Alvarez and Arellano (2022). Additional details on the estimation procedure are provided in Appendix B.

3.2 Nonlinear income process

Recent work on income dynamics has challenged the long-standing view that household income processes can be adequately represented by linear models. In particular, Arellano et al. (2017) and Guvenen et al. (2021) provide evidence that, contrary to the implications of the standard AR(1) process, the conditional distribution of pre-tax household income exhibits dispersion and skewness that vary with households' position in the income distribution and over the life-cycle.

In light of this evidence, Arellano et al. (2017) propose a flexible representation of household income dynamics that accommodates nonlinearity, non-normality, and age dependence in the conditional income distribution. We consider a simplified version of their framework, in which next-period income y_{it+1} follows the quantile process

$$y_{it+1} = Q_y(u_{it+1} \mid y_{it}, \mathbf{X}_{it}; \theta), \qquad u_{it+1} \sim U(0,1), \quad t \geqslant 1. \tag{4}$$

Intuitively, the quantile function maps draws from a uniform distribution on $(0,1)$ (i.e., cumulative probabilities) into draws from the conditional distribution of household income (i.e., quantiles).[9]

[9]Both the standard AR(1) and the state-dependent AR(1) specifications can be interpreted as special cases of this more flexible income process.

One way to characterize nonlinearities within this framework is through a generalized notion of persistence. Specifically, persistence conditional on current income, household characteristics, and quantile τ is

$$\rho(y_{it}, \mathbf{X}_{it}, \tau) = \frac{\partial Q_y(\tau \mid y_{it}, \mathbf{X}_{it})}{\partial y_{it}}. \tag{5}$$

This expression measures how strongly the current income state propagates into future income at quantile τ. Unlike in the linear AR(1) model, persistence can vary with both current income and the rank of the future shock. This flexibility allows the model to capture situations in which unusually large positive or negative shocks—such as promotions, job losses, or career transitions—have different degrees of persistence.

The nonlinear model also permits the spread of the conditional income distribution to vary flexibly across states. For some $\tau \in (0.5, 1)$, we define conditional income dispersion as

$$\sigma(y_{it}, \mathbf{X}_{it}, \tau) = Q_y(\tau \mid y_{it}, \mathbf{X}_{it}) - Q_y(1-\tau \mid y_{it}, \mathbf{X}_{it}). \tag{6}$$

This interquantile range measures the spread of the conditional income distribution. Unlike $\sigma^2(\mathbf{X}_{it})$ in the AR(1) specification, it is not a conditional variance, but a distribution-free measure of dispersion.

Finally, we characterize asymmetry in the shape of the conditional distribution using conditional skewness:

$$sk(y_{it}, \mathbf{X}_{it}, \tau) = \frac{Q_y(\tau \mid y_{it}, \mathbf{X}_{it}) + Q_y(1-\tau \mid y_{it}, \mathbf{X}_{it}) - 2Q_y(0.5 \mid y_{it}, \mathbf{X}_{it})}{Q_y(\tau \mid y_{it}, \mathbf{X}_{it}) - Q_y(1-\tau \mid y_{it}, \mathbf{X}_{it})}. \tag{7}$$

We use the estimated process to simulate income paths and construct a discrete transition matrix. Specifically, letting $\Xi = \{y_{1,t+1}, \ldots, y_{n,t+1}\}$ denote a grid of income points, we compute transition probabilities $\Pr(y_{i,t+1} \mid y_{j,t})$ for all $i, j = 1, \ldots, n$, which together define the transition matrix $\mathbb{Q}$.

Estimating this model using realized income data can be implemented using quantile regressions. Details of the estimation procedure are provided in Appendix B.

4 Estimating subjective income processes

In this section, we describe how to estimate household income processes using subjective income expectations together with observed current income and household characteristics. In the subjective expectations estimation, next-period income y_{it+1} is latent rather than observed. Instead, we observe current income y_{it}, covariates in the household information set, and respondents' reported probability mass over coarse bins of *future* income growth.

Estimating income processes from subjective expectations data presents two main challenges. First, the survey provides only coarse and discrete information about the conditional distribution of the continuous outcome y_{it+1} given $(y_{it}, \mathbf{X}_{it})$. Second, reported probabilities may reflect elicitation noise, focal responses, or other reporting behaviors potentially associated with cognitive constraints. Standard methods developed for realized income data are therefore not directly applicable to the estimation of *subjective* income processes.

We propose a framework that addresses these two issues separately. In the first subsection, we show how to estimate the state-dependent and nonlinear income processes using subjective expectations data while abstracting from systematic reporting behavior. We map the underlying income process into observed survey probabilities and model the $M = 10$ reported probability points as multinomial draws. This formulation accommodates the discreteness inherent in expressing an underlying continuous subjective distribution on the survey's 10-point response scale. In the second subsection, we introduce a separate reporting model for the most prominent systematic response pattern in the data: deterministic allocation of all probability mass to the middle bin.[10] The final subsection briefly discusses identification. Throughout, we interpret y_{it} as log household income, so that income growth is defined as

$$\Delta y_{it} \equiv y_{it+1} - y_{it}.$$

When income processes estimated from realized data are used to discipline household

[10] Appendix H considers a broader reporting model that allows for directional and concentration distortions of reported probabilities.

expectations in quantitative models, a common implicit assumption is that the conditional distribution inferred from realizations coincides with the one households perceive. By contrast, our framework estimates income dynamics directly from subjective expectations data and therefore does not require rational expectations. Households observe income realizations and form perceptions about the future distribution of income, and we require only that the estimated process be consistent with these perceived beliefs. Our approach is therefore compatible both with fully rational households that use realized data together with private information to forecast future income and with households that form systematically biased perceptions of future income dynamics.

4.1 Coarse information

Consider household i at time t. The survey asks respondents to allocate M points across $J = 5$ pre-specified intervals of future income growth. Let the interval cutoffs satisfy

$$-\infty = h_0 < h_1 < \cdots < h_{J-1} < h_J = +\infty,$$

so that bin $j \in \{1, \ldots, J\}$ corresponds to the interval $(h_{j-1}, h_j]$. Households observe current income y_{it} and covariates $\mathbf{X}_{it}$ contained in their information set. We denote the observed income state by

$$W_{it} \equiv (y_{it}, \mathbf{X}_{it}).$$

We assume that respondents form beliefs about the future distribution of income according to an underlying subjective income process. In what follows, we describe the estimation strategy for both the state-dependent AR(1) specification and the nonlinear quantile-based process.

4.1.1 State-dependent AR(1) process

Suppose households perceive income dynamics according to the state-dependent AR(1) process:

$$y_{it+1} = \rho(\mathbf{X}_{it})\, y_{it} + \mathbf{X}_{it}'\beta + \eta_i + \sigma(\mathbf{X}_{it})\, u_{it+1}, \qquad u_{it+1} \sim N(0,1).$$

Respondents are assumed to form beliefs about the probability that future income falls into each of the pre-specified intervals implied by model (1).[11]

Given $(y_{it}, \mathbf{X}_{it})$ and a value of η_i, the model implies a conditional Normal distribution for y_{it+1}. Define the corresponding level cutoffs

$$\kappa_{j,it} \equiv y_{it} + h_j,$$

so that the event $\Delta y_{it} \in (h_{j-1}, h_j]$ is equivalent to $y_{it+1} \in (\kappa_{j-1,it}, \kappa_{j,it}]$. Then, the probability that future income falls in bin j is

$$\begin{aligned} p^*_{j,it}(\eta_i) &\equiv \Pr\big(h_{j-1} < \Delta y_{it} \leqslant h_j \,\big|\, W_{it}, \eta_i\big) \\ &= \Pr(y_{it+1} \leqslant \kappa_{j,it} \mid W_{it}, \eta_i) - \Pr(y_{it+1} \leqslant \kappa_{j-1,it} \mid W_{it}, \eta_i) \\ &= \Phi\left(\frac{\kappa_{j,it} - \rho(\mathbf{X}_{it})y_{it} - \mathbf{X}_{it}'\beta - \eta_i}{\sigma(\mathbf{X}_{it})}\right) - \Phi\left(\frac{\kappa_{j-1,it} - \rho(\mathbf{X}_{it})y_{it} - \mathbf{X}_{it}'\beta - \eta_i}{\sigma(\mathbf{X}_{it})}\right), \end{aligned} \quad (8)$$

where $\Phi(\cdot)$ denotes the standard Normal cumulative distribution function (CDF).

Elicitation as multinomial sampling. Let $d_{jit} \in \{0, 1, \ldots, M\}$ denote the number of points assigned to bin j by respondent i at time t, with $\sum_{j=1}^J d_{jit} = M$. We interpret the survey design as requiring respondents to express their beliefs using M discrete probability points, and model

$$\mathbf{d}_{it} \equiv (d_{1it}, \ldots, d_{Jit})' \sim \text{Multinomial}(M, \mathbf{p}^*_{it}(\eta_i)), \; \mathbf{p}^*_{it}(\eta_i) \equiv \big(p^*_{1,it}(\eta_i), \ldots, p^*_{J,it}(\eta_i)\big)'. \quad (9)$$

Importantly, this approach treats observed survey responses as noisy discretized representations of an underlying continuous subjective income distribution. Because respondents allocate exactly $M = 10$ integer-valued points across the five bins, the multinomial layer accommodates the discretization noise inherent in expressing continuous subjective probabilities on this coarse reporting scale. This component is not intended to capture systematic reporting behavior such as focal-point responses, heaping, or other distortions. We model the most prominent such pattern in our data—the allocation of all probability mass to the middle bin—separately in Section 4.2, and consider more general distortions of reported probabilities in Appendix H.[12]

[11] One can perform similar calculations assuming logistic shocks, following Arellano et al. (2026b). We use the Normal distribution because it is the standard assumption in quantitative macroeconomic models of income dynamics.

[12] Alternative approaches to rounding in probabilistic expectations include Manski and Molinari (2010)

Random effects and integrated likelihood. Since η_i is unobserved, we complete the model by assuming

$$\eta_i \sim N(0, \sigma_\eta^2),$$

independently across households. Conditional on η_i, the per-period likelihood contribution is given by the multinomial probability mass function. The integrated likelihood contribution for household i is

$$L_i(\theta) = \int \prod_{t=1}^{T} \left[\frac{M!}{\prod_{j=1}^{J} d_{jit}!} \prod_{j=1}^{J} \left(p^*_{j,it}(\eta_i)\right)^{d_{jit}} \right] \frac{1}{\sigma_\eta} \phi\left(\frac{\eta_i}{\sigma_\eta}\right) d\eta_i, \tag{10}$$

where $\phi(\cdot)$ denotes the standard Normal probability density function (PDF), and θ collects the parameters governing $\rho(\mathbf{X}_{it})$, $\sigma(\mathbf{X}_{it})$, β, and σ_η^2. The sample log-likelihood is therefore

$$\ell_N(\theta) = \sum_{i=1}^{N} \log L_i(\theta), \qquad \widehat{\theta} = \arg\max_{\theta} \ell_N(\theta). \tag{11}$$

Because the integrated likelihood in (10) does not admit a closed-form solution, we approximate the integral numerically using Gauss–Hermite quadrature with 12 nodes. Standard errors are computed using a nonparametric bootstrap with 500 replications.

Implementation. A key modeling choice concerns the specification of $\rho(\mathbf{X}_{it})$ and $\sigma(\mathbf{X}_{it})$. In our empirical implementation, we adopt a logistic specification for persistence and an exponential specification for dispersion:[13]

$$\rho(\mathbf{X}_{it}) = \Lambda(\rho_0 + \mathbf{X}_{it}'\rho_1)\,, \qquad \sigma(\mathbf{X}_{it}) = \exp(\sigma_0 + \mathbf{X}_{it}'\sigma_1)\,,$$

where $\Lambda(z) = \exp(z)/(1 + \exp(z))$. These functional forms ensure that persistence lies in the unit interval and that dispersion remains strictly positive.

Motivated by Meghir and Pistaferri (2004) and Browning et al. (2010), who emphasize age and education as important dimensions of heterogeneity in income dynamics, the

and Giustinelli, Manski, and Molinari (2022), who use interval-valued representations of reported probabilities. Our objective here is different: we retain a parametric representation of the underlying income process and distinguish the discreteness induced by the survey response scale from additional systematic reporting behavior. Respondents may also internally represent probabilities using a finer or coarser grid than the survey value $M = 10$; accommodating such heterogeneity in reporting granularity would require an additional latent reporting layer (for example, a household-specific latent M_i). In our baseline specification, we treat M as fixed and known from the survey design.

[13]We also experimented with linear specifications for $\rho(\mathbf{X}_{it})$ and $\sigma(\mathbf{X}_{it})$. These alternatives provided a substantially poorer fit to the observed bin probabilities when estimated using subjective expectations data.

state variables in $\mathbf{X}_{it}$ include age and education dummies, together with year fixed effects that capture aggregate shocks.[14] We also include log income y_{it} in the dispersion equation to allow for potential nonlinearities, so that $\sigma(\cdot)$ can be rewritten as $\sigma(\mathbf{X}_{it}, y_{it})$.

4.1.2 Nonlinear income process

Motivated by recent work on nonlinear income dynamics, we also consider a nonlinear income process in the spirit of Arellano et al. (2017). In this specification, households' perceived next-period income is modeled through the conditional quantile function

$$y_{it+1} = Q_y(u_{it+1} \mid y_{it}, \mathbf{X}_{it}; \theta), \qquad u_{it+1} \sim U(0,1), \quad t \geqslant 1, \tag{12}$$

which we parameterize following Arellano et al. (2017) as

$$Q_y(\tau \mid y_{it}, \mathbf{X}_{it}) = \sum_{k=0}^{K} a_k(\tau)\, \psi_k(y_{it}) + \mathbf{X}_{it}'\beta(\tau), \tag{13}$$

where $\psi_k(\cdot)$ denotes low-order Hermite polynomials, and the coefficients $\{a_k(\tau)\}_{k=0}^{K}$ and $\beta(\tau)$ are modeled as piecewise-linear splines in τ over a grid $\{\tau_1, \ldots, \tau_L\} \subset (0,1)$. To discipline the behavior of the quantile function near $\tau = 0$ and $\tau = 1$, we impose exponential-tail restrictions at the boundary knots with parameters λ_1 and λ_L.[15]

If next-period income y_{it+1} were directly observed, the parameters of the income process could be estimated using standard quantile regressions. In our setting, however, the observed data consist only of probabilities assigned to intervals of the conditional distribution of y_{it+1} given y_{it} and $\mathbf{X}_{it}$. Moreover, estimating the model by standard maximum likelihood methods as in the state-dependent AR(1) model is infeasible given the quantile representation, the nonlinearities present in the functional form, and the restrictions required to recover a proper distribution. To address these issues, we interpret each of the M reported probability points as corresponding to an underlying latent draw from the household's perceived income distribution. Specifically, we treat the latent outcomes

[14] We also estimate specifications including employment characteristics. Results are largely unchanged.

[15] Because the two extreme survey bins are open-ended, the shape of the perceived distribution within the tails is not directly disciplined by the reported probabilities and depends on the exponential-tail extrapolation. To assess sensitivity to this choice, we re-estimate the nonlinear process imposing finite bounds on income growth and progressively tighten the maximum absolute support from 90% to 20%. The estimated persistence and dispersion profiles, including those obtained under the hurdle specification, remain very similar. Additional results are available upon request.

$\{y_{it+1}^{(m)}\}_{m=1}^{M}$ as missing data and observe only their bin membership through $\mathbf{d}_{it}$. We then estimate the income process using a stochastic pseudo-EM algorithm similar to that proposed by Arellano and Bonhomme (2016).

This approach is feasible because the parameterization of the quantile function, together with the closed-form tail restrictions, allows us to approximate the conditional cumulative distribution function (CDF)

$$F_{it}(y;\theta) \equiv \Pr(y_{it+1} \leqslant y \mid y_{it}, \mathbf{X}_{it}; \theta),$$

which in turn allows us to simulate latent realizations of y_{it+1} consistent with the observed survey responses. We outline the main steps of the algorithm below and relegate additional details to Appendix C.

Stochastic pseudo-EM algorithm. Given a parameter vector $\theta^{(s)}$, the conditional density of a latent draw restricted to bin j is the truncated version of the model-implied density:

$$f\left(y \mid y_{it}, \mathbf{X}_{it},\, y \in I_{j,it}; \theta^{(s)}\right) \propto f\left(y \mid y_{it}, \mathbf{X}_{it}; \theta^{(s)}\right)\, \mathbf{1}\{y \in I_{j,it}\}. \tag{14}$$

Starting from an initial parameter vector $\theta^{(0)}$, we iterate between the following two steps for $s = 1, 2, \ldots$ until convergence:

1. **Stochastic E-step.** For each household-period pair (i, t) and each bin j, we draw d_{jit} latent realizations from the truncated distribution in (14). Pooling these simulated observations across bins yields the augmented sample

$$\{y_{it+1}^{(s,m)}, y_{it}, \mathbf{X}_{it}\}_{i,t,m},$$

   which contains exactly M latent draws for each (i, t) pair.

2. **M-step.** Conditional on the augmented sample, we update the spline parameters by estimating quantile regressions at the spline knots $\{\tau_\ell\}_{\ell=1}^{L}$. Specifically, for each knot τ_ℓ, we estimate

$$(\widehat{a}(\tau_\ell), \widehat{\beta}(\tau_\ell)) = \arg \min_{a(\tau_\ell), \beta(\tau_\ell)} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{m=1}^{M} \rho_{\tau_\ell} \left(y_{it+1}^{(s,m)} - \sum_{k=0}^{K} a_k(\tau_\ell) \psi_k(y_{it}) - \mathbf{X}_{it}' \beta(\tau_\ell) \right),$$

where $\rho_\tau(\cdot)$ denotes the standard quantile-regression check function, $\rho_\tau(u) = u(\tau - \mathbf{1}(u < 0))$.

We monitor convergence by evaluating the observed-data log-likelihood at each iterate $\theta^{(s)}$, and stop the algorithm when

$$\frac{\left|\ell_N(\theta^{(s+1)}) - \ell_N(\theta^{(s)})\right|}{1 + \left|\ell_N(\theta^{(s)})\right|} < \varepsilon,$$

for a pre-specified tolerance level ε. As in the state-dependent AR(1) specification, standard errors are computed using a nonparametric bootstrap with 500 replications.

4.2 A hurdle model for middle-bin focal responses

A large fraction of respondents allocate all M points to the middle bin. As discussed earlier, this response pattern may reflect genuinely low perceived dispersion for some households, but it may also arise from systematic focal-point reporting behavior. This is conceptually distinct from the discretization inherent in the 10-point elicitation scale, which is already accommodated by the multinomial formulation in Section 4.1. Here we address a specific additional reporting pattern: the possibility that some respondents deterministically allocate all probability mass to the middle bin rather than reporting probabilities generated by their underlying perceived income distribution.

To account for this behavior, we augment the model with a hurdle, or two-regime, reporting structure by adapting the finite-mixture approaches of de Bresser and van Soest (2013), Kleinjans and van Soest (2014), and Heiss et al. (2022) to our setting. The baseline hurdle model is specific to middle-bin focal responses; it is not intended as a general model of rounding or arbitrary probability distortions[16]. We illustrate the hurdle model in the context of the state-dependent AR(1) specification and relegate the nonlinear extension to Appendix C.3.

Let $j^\star$ denote the index of the middle bin, and define the observed bunching indicator

$$S_{it} \equiv \mathbf{1}\{d_{j^\star,it} = M\}.$$

[16] One can write a finite mixture model with more than one reporting style, as in Heiss et al. (2022). As other reporting styles (such as allocating all ten points to other bins) are substantially less prevalent than middle-bin bunching, we do not pursue this here. The more general model in Appendix H encompasses these other behaviors.

We assume that respondents follow a sequential reporting process:

1. First, they enter either the deterministic focal-point regime or the non-focal reporting regime.

2. Conditional on entering the non-focal regime, they allocate the M points across bins according to the multinomial probabilities implied by their perceived income process.

Let $R_{it} \in \{0, 1\}$ indicate whether respondent i at time t is in the deterministic focal-point regime. We specify

$$\pi_{it}(b_i; \gamma) \equiv \Pr(R_{it} = 1 \mid \mathbf{Z}_{it}, b_i) = \Lambda(\gamma_0 + \mathbf{Z}_{it}' \gamma_1 + b_i) \,, \tag{15}$$

where

$$\gamma \equiv (\gamma_0, \gamma_1')'$$

collects the parameters of the reporting equation. The term b_i is a time-invariant respondent-specific random effect capturing the propensity to exhibit focal-point reporting. The vector $\mathbf{Z}_{it}$ contains covariates that may overlap with $\mathbf{X}_{it}$ as well as variables excluded from the income process. In particular, we write

$$\mathbf{Z}_{it} = (\mathbf{X}_{it}', \mathbf{V}_{it}')' \,,$$

where $\mathbf{V}_{it}$ affects reporting behavior but does not directly affect the underlying perceived income process.

In the empirical implementation of the hurdle model, we use households' response patterns in the house price expectations question as an exclusion restriction. Because the income and house price expectations questions share a nearly identical probabilistic format, it is plausible that cognitive limitations or reporting styles affect responses similarly across domains, consistent with the evidence reported in Table 3. The identifying assumption is that, conditional on the household characteristics included in the model and wave fixed effects, house price response behavior affects the probability of focal reporting in the income expectations question but does not directly affect the household's underlying perceived income distribution.

This exclusion restriction is necessarily a maintained assumption. In particular, because both questions concern nominal outcomes, heterogeneous inflation expectations or inflation uncertainty could affect responses to both income growth and house price expectations. Wave fixed effects absorb common aggregate inflation developments, but they do not remove such household-specific variation. Hence, part of the observed cross-domain association in focal responses could reflect common subjective inflation uncertainty rather than reporting style. We therefore interpret house price bunching as an excluded reporting shifter under the assumption that, after conditioning on observed household characteristics and wave effects, its remaining association with income expectation bunching operates through reporting behavior.

In the baseline specification, we assume that the household-specific random effect in the income process, η_i, and the random effect in the reporting equation, b_i, are independent:[17]

$$\eta_i \sim N(0, \sigma_\eta^2), \qquad b_i \sim N(0, \sigma_b^2), \qquad \eta_i \perp b_i.$$

We also maintain that (η_i, b_i) is independent of the observed covariate histories.

Let θ denote the vector of income-process parameters introduced in Section 4.1, including σ_η^2. We collect all parameters of the hurdle model in

$$\vartheta \equiv \left(\theta', \gamma', \sigma_b^2\right)'.$$

Likelihood function. Conditional on η_i, define the probability that the non-focal multinomial reporting mechanism places all M points in the middle bin as

$$A_{it}(\eta_i; \theta) \equiv \Pr(d_{j^\star,it} = M \mid \mathbf{W}_{it}, \eta_i; \theta) = \left[p^*_{j^\star,it}(\eta_i; \theta)\right]^M,$$

where

$$\mathbf{W}_{it} \equiv (y_{it}, \mathbf{X}_{it})$$

and $p^*_{j^\star,it}(\eta_i; \theta)$ denotes the model-implied probability of the middle bin under the state-dependent AR(1) process in equation (1).

[17] Allowing for correlated random effects is straightforward by replacing the diagonal covariance matrix with a full bivariate Normal covariance matrix.

Conditional on (η_i, b_i), the per-period likelihood contribution is

$$L_{it}^{H}(\eta_i, b_i; \vartheta) = \begin{cases} \pi_{it}(b_i;\gamma) + [1-\pi_{it}(b_i;\gamma)]\, A_{it}(\eta_i;\theta), & \text{if } S_{it}=1, \\ [1-\pi_{it}(b_i;\gamma)]\, \dfrac{M!}{\prod_{j=1}^{J} d_{jit}!} \displaystyle\prod_{j=1}^{J} \left[p_{j,it}^{*}(\eta_i;\theta)\right]^{d_{jit}}, & \text{if } S_{it}=0. \end{cases} \tag{16}$$

The two cases admit a natural interpretation:

- If $S_{it} = 1$, the observed bunching response may arise either because the respondent is in the deterministic focal-point regime, with probability $\pi_{it}(b_i;\gamma)$, or because the respondent is in the non-focal regime and all M points happen to be allocated to the middle bin, with probability

$$[1-\pi_{it}(b_i;\gamma)]\, A_{it}(\eta_i;\theta).$$

- If $S_{it} = 0$, the deterministic focal-point regime assigns zero probability to the observed allocation. The likelihood contribution is therefore the non-focal multinomial likelihood multiplied by

$$1-\pi_{it}(b_i;\gamma).$$

The integrated likelihood contribution for household i is

$$L_i^{H}(\vartheta) = \iint \prod_{t=1}^{T_i} L_{it}^{H}(\eta_i, b_i;\vartheta) f_\eta(\eta_i;\sigma_\eta^2) f_b(b_i;\sigma_b^2)\, d\eta_i\, db_i, \tag{17}$$

where

$$f_\eta(\eta_i;\sigma_\eta^2) = \phi(\eta_i;0,\sigma_\eta^2)$$

and

$$f_b(b_i;\sigma_b^2) = \phi(b_i;0,\sigma_b^2).$$

The sample log-likelihood is

$$\ell_N^{H}(\vartheta) = \sum_{i=1}^{N} \log L_i^{H}(\vartheta),$$

and the maximum-likelihood estimator is

$$\widehat{\vartheta} = \arg\max_{\vartheta\in\mathcal{V}} \ell_N^{H}(\vartheta),$$

where $\mathcal{V}$ denotes the parameter space of the hurdle model.

We estimate the model by maximum likelihood and approximate the two-dimensional integrals in equation (17) using Gauss–Hermite quadrature. Standard errors are computed using a nonparametric bootstrap with 500 replications.

The baseline hurdle model is intentionally narrow: it allows systematic deviations from the multinomial reporting mechanism only through deterministic focal responses in which all probability mass is allocated to the middle bin. It is therefore not a general model of rounding or reporting error. As a robustness exercise, Appendix H considers a broader reporting model in which households may systematically distort reported probabilities by tilting their perceived distribution toward optimism or pessimism, or toward greater concentration or dispersion. The resulting estimates are qualitatively very similar to those obtained under the baseline hurdle specification. Thus, our main conclusions do not depend on restricting systematic reporting distortions to middle-bin focal responses.

4.3 Identification

The object that we would like to recover is the perceived distribution of future log income y_{it+1} conditional on the state observed by the econometrician,

$$F_y(\cdot \mid W_{it}), \tag{18}$$

where $W_{it} \equiv (y_{it}, \mathbf{X}_{it})$ collects current household income and household characteristics. Households report their perceived distribution by allocating M probability points across pre-specified intervals of income growth. In this subsection, we briefly discuss the identification of the conditional distribution. A more formal analysis of identification is in Appendix E.

Given minimal assumptions on the data, the conditional distribution is only partially identified.[18] This is because the data pin down the perceived CDF at the fixed cutoff points, while leaving its shape within each interval unrestricted. Multiple conditional

[18] One potential route to nonparametric identification would be to allow cutoff values to vary across households, as in the min–max elicitation approach of Arellano et al. (2026b). If the support of the thresholds were sufficiently rich and the corresponding probabilities observed, the full conditional distribution could in principle be recovered.

distributions can therefore generate the same set of observed bin probabilities; any values within the interval defined by the bins can be generated by some conditional distribution consistent with the observed data, as in Manski and Tamer (2002). Hence, in the absence of additional assumptions, subjective expectations data alone do not nonparametrically identify the full conditional distribution of future income or its associated moments.[19]

State-dependent AR(1) process. In the state-dependent AR(1) specification, the household-specific perceived income dynamics satisfy

$$y_{i,t+1} \mid W_{it}, \eta_i \sim N\left(\rho(X_{it})y_{it} + X_{it}'\beta + \eta_i, \sigma^2(X_{it})\right),$$

where

$$\eta_i \sim N(0, \sigma_\eta^2)$$

is a household-specific random effect that is exogenous. Under this assumption, integrating over η_i gives

$$y_{i,t+1} \mid W_{it} \sim N\left(\mu(W_{it}), \sigma_e^2(X_{it})\right),$$

where

$$\mu(W_{it}) = \rho(X_{it})y_{it} + X_{it}'\beta$$

and

$$\sigma_e^2(X_{it}) = \sigma^2(X_{it}) + \sigma_\eta^2.$$

The Normal location–scale restriction determines the full marginal CDF from $\mu(W_{it})$ and $\sigma_e(X_{it})$. At a given state, two distinct finite cutoffs whose cumulative probabilities are strictly between zero and one identify these two objects. Variation in current income conditional on X_{it}, together with the relevant full-rank conditions, then identifies the parameters governing $\rho(X_{it})$ and the conditional mean.

[19] Appendix E focuses on the identification of households' perceived distribution $F(\cdot \mid W_{it})$. If instead household i's perceived distribution is $F^I(\cdot \mid W_{it}, \xi_{it+1})$, where ξ_{it+1} denotes private information about the evolution of future income, then provided reported bin probabilities are unbiased conditional on the household's information, the conditional bin probabilities correspond, by the law of iterated expectations, to those of

$$F(\cdot \mid W_{it}) = \int F^I(\cdot \mid W_{it}, \xi)\, dG(\xi \mid W_{it}).$$

The current estimator $\widehat{F}(\cdot \mid W_{it})$ can therefore be interpreted, within the maintained model, as recovering this integrated perceived distribution. The subjective answers do not separately identify ξ_{it+1} or $F^I(\cdot \mid W_{it}, \xi_{it+1})$.

The cross-sectional distribution identifies only the conditional variance of the composite disturbance, $\sigma_e^2(X_{it})$. Under random-effects exogeneity and conditional independence of reporting draws across waves given the household effect, cross-wave dependence in the reported allocations identifies σ_η^2. The within-household innovation variance can then be recovered as

$$\sigma^2(X_{it}) = \sigma_e^2(X_{it}) - \sigma_\eta^2.$$

Thus, the panel separates persistent cross-household heterogeneity from the within-household innovation variance[20]. The dispersion measure reported in the empirical results is the corresponding innovation standard deviation, $\sigma(X_{it})$, rather than the total marginal standard deviation $\sigma_e(X_{it})$.

Nonlinear income process. Under the nonlinear income process, we restrict the conditional quantile function $Q_y(\tau \mid W_{it})$ to a finite-dimensional sieve class. At a fixed sieve dimension, the observed allocations identify the model-implied CDF at the bin cutoffs. The parameters of the nonlinear process are point identified within the maintained sieve model provided that the mapping from those parameters to the collection of cutoff CDF values is injective. Monotonicity and the tail restrictions ensure that each admissible parameter vector defines a proper conditional distribution, while the sieve specification determines its shape within the bins.

Under these conditions, the conditional quantile function and the associated measures of persistence, dispersion—measured by the interquantile range in equation (6)—and skewness are identified within the maintained model. This is a parametric or sieve-based identification result: the within-bin shape is supplied by the model restrictions rather than identified nonparametrically from the reported bin probabilities.

Hurdle model. The hurdle model introduces a second reporting regime. With probability $\pi_{it}(b_i)$, the household enters the deterministic focal-point regime and allocates all

[20] This decomposition additionally relies on the maintained random effects exogeneity and cross-wave conditional-independence assumptions; unlike the period-specific identification of the marginal perceived distribution, these restrictions may be violated by time-varying omitted household information correlated with subsequent observed income.

M points to the middle bin. With probability $1 - \pi_{it}(b_i)$, the allocation is generated by the underlying perceived income distribution.

Middle-bin bunching can therefore arise either from genuinely concentrated beliefs or from deterministic focal-point reporting. To separate these two sources of bunching, we rely on an excluded reporting variable that shifts the probability of entering the focal regime while, conditional on the observed state and year effects, being excluded from the perceived income distribution. As discussed in Section 4.2, our empirical implementation uses house price response behavior for this purpose. The validity of this restriction requires that any remaining cross-domain association after conditioning on observables operates through reporting behavior rather than through common latent factors such as household-specific inflation uncertainty.

Point identification additionally requires that the non-focal allocation distribution identify the parameters of the underlying income process and that the mapping from the reporting equation parameters to focal-reporting probabilities be injective. Under these conditions, the income-process parameters can first be recovered from the distribution of non-focal allocation patterns. Given the identified income process, the observed frequency of non-focal responses then identifies the focal-reporting probabilities and the parameters of the reporting equation.

5 Results

This section compares the income processes recovered from subjective expectations data with those estimated from realized income data, focusing on both the state-dependent AR(1) specification and the nonlinear quantile model. Across specifications, a consistent pattern emerges: households' subjective expectations imply substantially more persistent and less dispersed income dynamics than those inferred from realized income histories. These differences are economically meaningful and have important implications for precautionary saving, consumption smoothing, and the extent of self-insurance against income risk.

5.1 Coarse information

We begin by presenting results from the baseline specifications, which account for the coarse nature of the survey responses but abstract from systematic reporting biases or focal-point behavior.

5.1.1 State-dependent AR(1) process

Table 4 reports the estimated mean and cross-household standard deviation of persistence and dispersion for the realized and subjective income processes. Persistence estimated from realized income data is 0.685, whereas subjective expectations imply a much higher value, close to unity (0.999). In contrast, average income dispersion is substantially lower in the subjective process: the realized-income estimates imply average dispersion of 0.472, compared with 0.041 in the subjective process.

Table 4: Persistence and dispersion in the state-dependent AR(1) process

	Mean (μ_ρ, μ_σ)	Standard deviation $(\varsigma_\rho, \varsigma_\sigma)$
Panel A: Realized income process		
Persistence	0.6853	0.0407
	[0.6584,0.7155]	[0.0323,0.077]
Dispersion	0.4722	0.0558
	[0.4543,0.4877]	[0.0428,0.0752]
Panel B: Subjective income process		
Persistence	0.9998	0.0004
	[0.9539, 0.9999]	[1.0×10^{-5}, 0.0013]
Dispersion	0.0411	0.0105
	[0.0320, 0.1426]	[0.0084, 0.5629]

Notes: This table reports estimates of the mean and standard deviation of persistence, μ_ρ and ς_ρ in (2), and of income dispersion, measured by the conditional standard deviation of the income innovation, $\sigma(X_{it})$, with mean μ_σ and cross-household standard deviation ς_σ in (3), for the realized and subjective income processes. The specifications include wave dummies and controls for age and education. Brackets report 95% block-bootstrap confidence intervals computed using 500 bootstrap replications.

Figure 2 further illustrates these differences. The distributions of persistence and dispersion differ markedly depending on whether the process is estimated from realized income data or from subjective expectations. Overall, the results point to two main empirical patterns: households perceive future income as substantially more persistent than estimates based on realized income histories imply,[21] and they perceive considerably less dispersion in future income than is observed in realized income dynamics.

The model fit further illustrates the limitations of the baseline specification. As the bottom panel of Figure 2 shows, the subjective AR(1) model fails to reproduce the sharp concentration of probability mass in the middle bin and underpredicts the probability assigned to more extreme income-growth outcomes, yielding a root-mean-squared error (RMSE) of 0.25. This relatively poor fit highlights the difficulty of reconciling coarse and highly bunched probability allocations with a smooth Gaussian income process and motivates the extension below that explicitly accounts for focal-point reporting behavior.

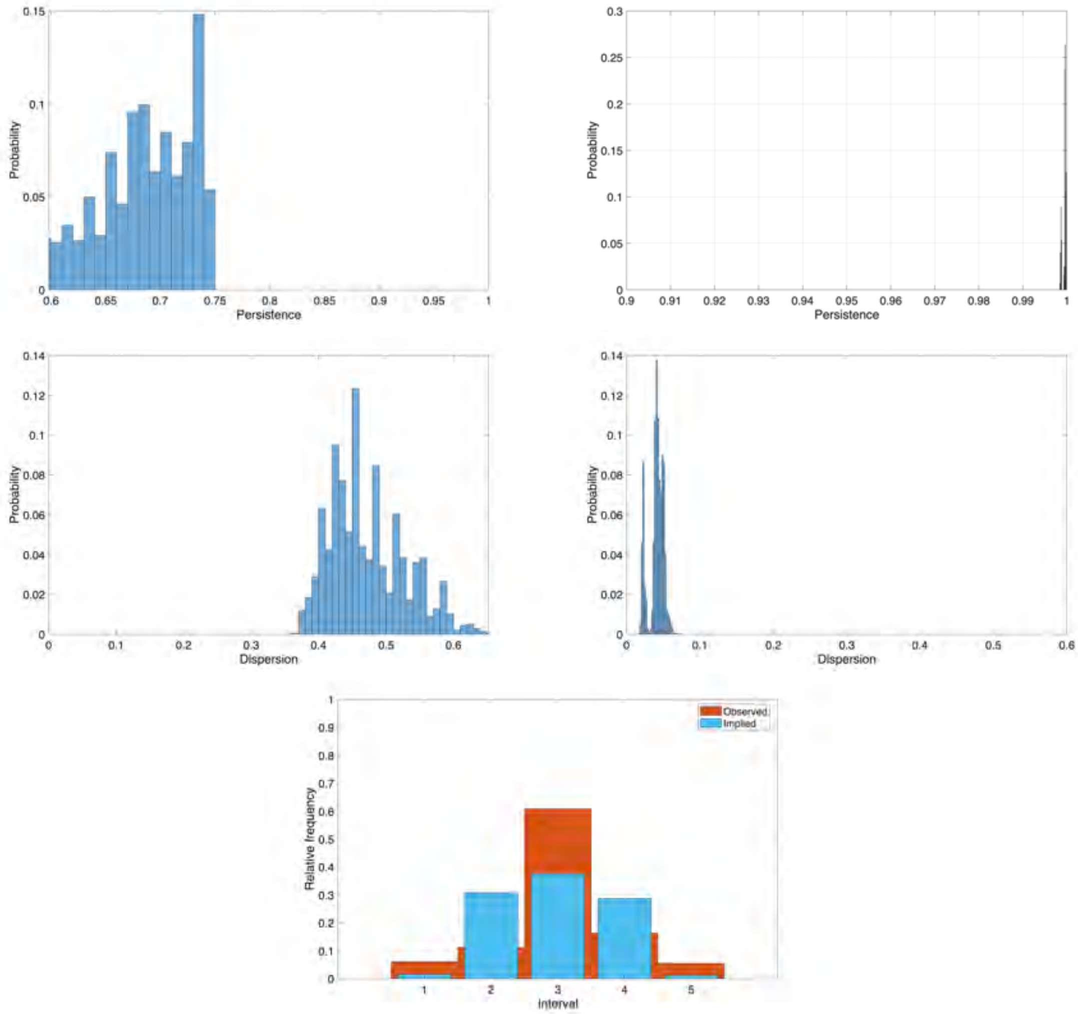
Measurement error in reported income. An important concern is that realized household income is survey-reported and may therefore contain measurement error. Classical serially uncorrelated measurement error attenuates estimated persistence and inflates innovation variance, and hence innovation dispersion, moving the realized income estimates in the same direction as the differences we document. To assess the quantitative importance of this channel, Appendix F re-estimates both the realized and subjective state-dependent AR(1) processes over a range of assumed reliability ratios for reported income.

The correction is quantitatively important but does not eliminate the gap. At $\lambda = 0.80$, an illustrative benchmark broadly consistent with validation evidence (see Bound, Brown, and Mathiowetz (2001) for a survey) mean realized persistence rises from 0.685 to 0.832, while mean realized dispersion falls from 0.472 to 0.224. By comparison, subjective persistence remains essentially one and subjective dispersion is 0.038. Thus, relative to the no measurement-error benchmark, the persistence gap falls by about 46 percent and the dispersion gap by about 57 percent, but both remain economically large. Lower reliability ratios narrow the differences further without reconciling the two processes.[22]

[21]This pattern is consistent with the evidence on over-persistence in subjective income expectations documented by Rozsypal and Schlafmann (2023).

[22]This exercise is intended as a sensitivity analysis rather than as an estimate of the amount of reporting error in the EFF. Survey income-reporting errors may be non-classical or serially correlated, in which case the direction of the resulting bias is not generally signed.

Figure 2: Persistence, dispersion, and model fit in the state-dependent AR(1) process



Notes: The top row shows the distribution of persistence, and the middle row shows the distribution of income dispersion, measured by the innovation standard deviation $\sigma(X_{it})$, for the state-dependent AR(1) process estimated without modeling focal-point behavior. The left column corresponds to the realized income process, while the right column corresponds to the subjective income process. The bottom panel compares model-implied probabilities from the subjective income process with the observed probabilities in the survey data.

Household-specific dispersion. Appendix D reports an alternative AR(1) specification with household-specific dispersion. This improves the fit of the subjective expectations data substantially, reducing the RMSE from 0.25 to 0.05, while leaving the qualitative conclusions unchanged: subjective expectations imply much more persistent and less dispersed income dynamics than realized income histories.

5.1.2 Nonlinear income process

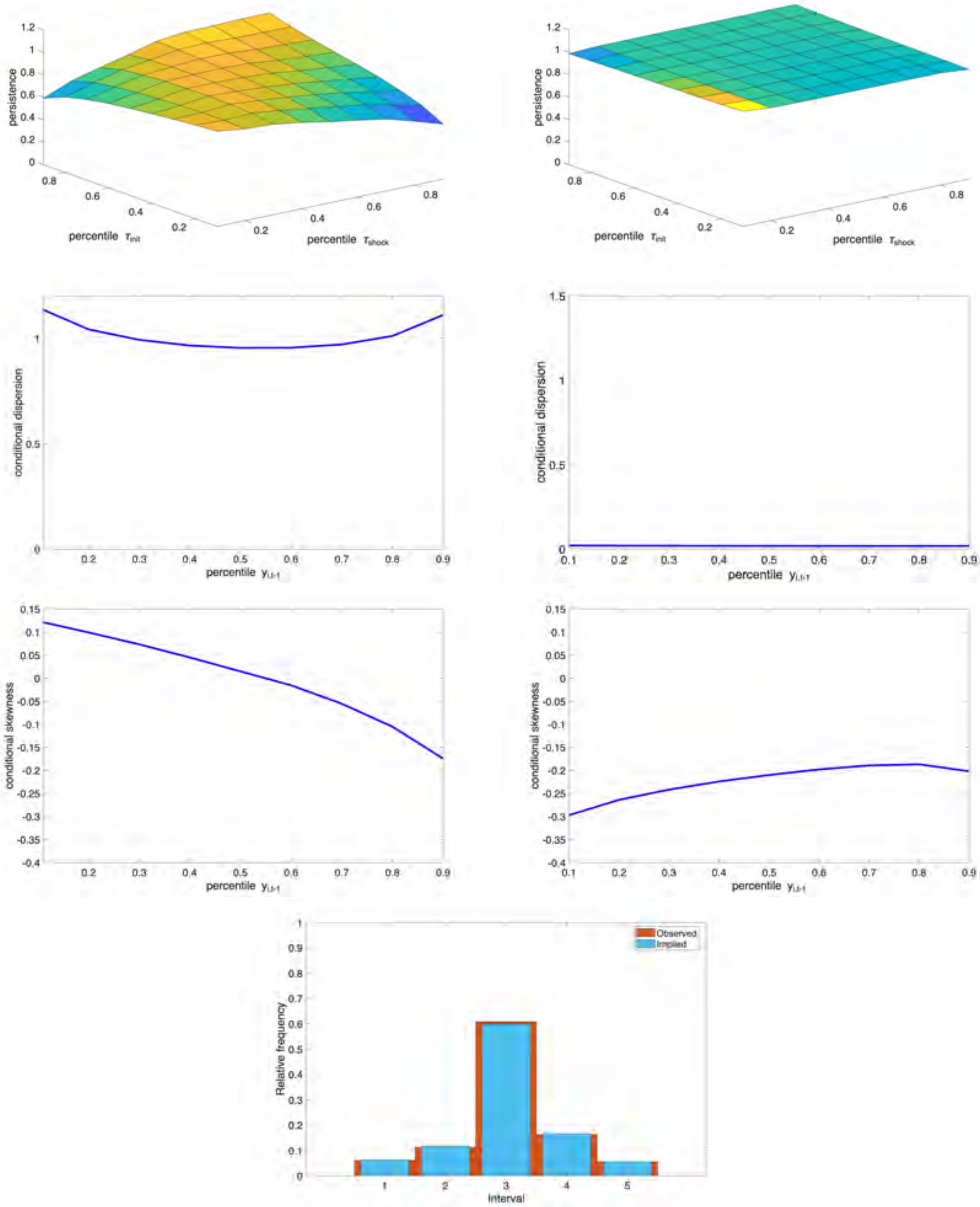
Figure 3 reports results from the nonlinear quantile-based income model estimated without accounting for focal-point behavior. Three patterns stand out. First, the generalized persistence surface estimated from realized income data displays rich nonlinearities: persistence varies with both current income and the rank of the shock. By contrast, the subjective model generates an almost flat persistence surface, suggesting that households perceive shocks as similarly persistent across the income distribution and across different types of shocks. Second, subjective expectations imply substantially lower conditional dispersion. The dispersion measure is uniformly lower than that inferred from realized income data. Third, conditional skewness varies with household rank in the realized income process, while the subjective process implies a much flatter skewness profile. In particular, the subjective conditional distribution is negatively skewed across much of the income distribution.

The nonlinear subjective model fits the observed probability allocations well, with an RMSE of around 0.0045. Appendix J reports 95% nonparametric bootstrap confidence intervals for the main nonlinear estimates. Figures J.1 and J.2 report the confidence intervals for the processes estimated from realized income data and subjective expectations, respectively, showing that the estimated nonlinear objects are precisely estimated.

The differences between income processes inferred from realized outcomes and from subjective expectations are substantial, regardless of whether we use the state-dependent AR(1) specification or the nonlinear quantile model. One potential concern is that the two sources of information refer to different horizons. The realized income estimates use income observations from the 2014, 2017, 2020 EFF waves,[23] so household income is

[23] We obtain similar results when using waves from 2002–2022.

Figure 3: Nonlinear persistence, dispersion, skewness, and model fit



Notes: The first row shows nonlinear persistence, the second row conditional dispersion, measured by the interquantile range in equation (6), and the third row conditional skewness, measured by the quantile-based skewness measure (7) for the nonlinear income process estimated without modeling focal-point behavior. The left column corresponds to the process estimated from realized income data, while the right column corresponds to the process estimated from subjective expectations data. The bottom panel compares model-implied probabilities from the subjective income process with the observed survey probabilities.

observed at three-year intervals. By contrast, the subjective expectations question refers to expected income growth over the next twelve months.

Could this horizon difference explain the large gap between realized and subjective income processes? To assess this possibility, we use data from the Spanish Living Conditions Survey (ECV), which allows us to estimate both the state-dependent AR(1) and nonlinear income processes using one-year and three-year income changes. This exercise isolates how much estimated income dynamics change solely because of the length of the observation interval. The results and the corresponding bootstrap confidence intervals are reported in Appendix G.

The evidence suggests that horizon differences are not driving our main findings. First, income processes estimated using three-year intervals in the ECV are very similar to those obtained from the EFF. Second, although persistence is slightly lower and dispersion slightly higher when moving from one-year to three-year income changes, these differences are small relative to the uncovered gap between realized and subjective income processes.

While the absence of annual realized income data for the same EFF households prevents a definitive comparison, the ECV evidence indicates that differences in observation horizons are unlikely to account for the large discrepancies we document. In Section 6, when embedding the estimated processes into a life-cycle model, we explicitly account for the different horizons by adjusting the underlying transition matrices.

5.2 Hurdle model for middle-bin focal responses

We now present the empirical results obtained when we explicitly account for middle-bin focal-point reporting behavior through the hurdle model. These specifications allow the observed concentration of probability mass in the middle bin to reflect a separate focal-reporting regime rather than the underlying perceived income distribution. We can therefore assess how much of the differences between the realized and subjective income processes can be accounted for by this specific concentration distortion.

5.2.1 State-dependent AR(1) process

We report results for the following state-dependent AR(1) specification:

$$y_{it+1} = \rho y_{it} + \eta_i + \mathbf{X}_{it}'\beta + \sigma(\mathbf{X}_{it})u_{it+1}, \qquad u_{it+1} \sim N(0,1).$$

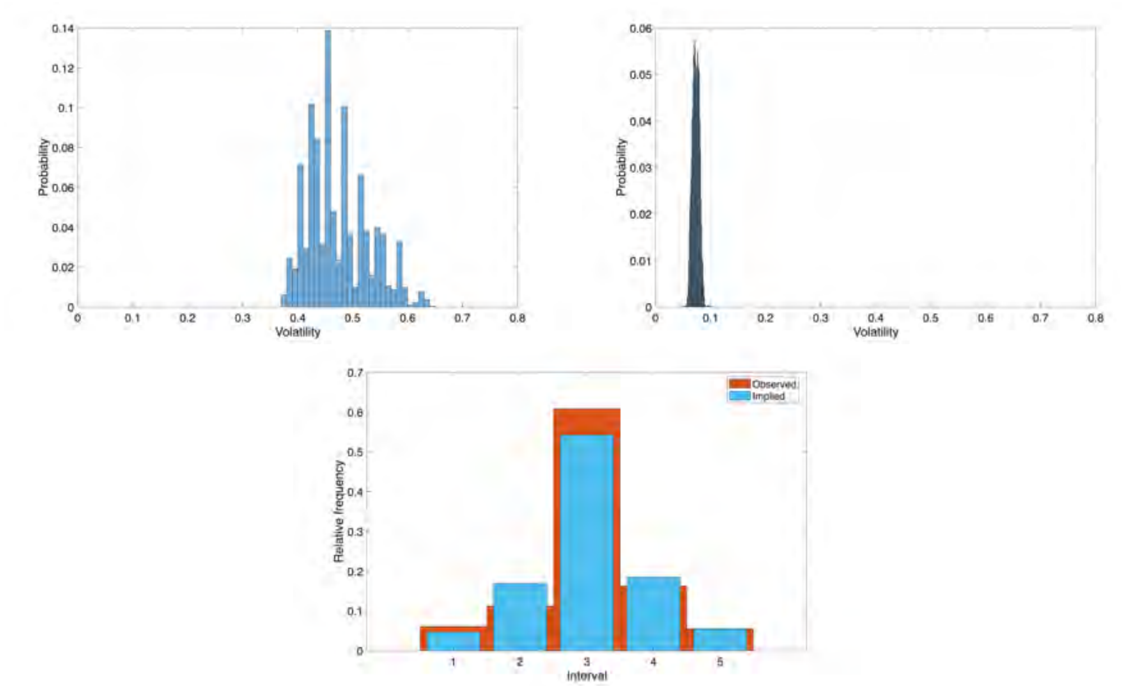
Relative to the baseline state-dependent AR(1) model, we restrict persistence to be homogeneous across households. This simplification is motivated by the fact that, in the baseline subjective income process, the estimated distribution of $\rho(\mathbf{X}_{it})$ is tightly concentrated around 0.999. We therefore focus on the single estimated persistence parameter and on the distribution of state-dependent dispersion.

Allowing for focal-point reporting behavior improves model fit substantially (Figure 4). The hurdle extension increases the estimated dispersion of the subjective income process while leaving persistence essentially unchanged. Relative to the baseline specification without the hurdle, the perceived-realized dispersion gap narrows by about 27 percent, but remains economically large. The model also captures the shape of the observed bin probabilities more closely, especially in the tails, reducing the RMSE to 0.12. Nevertheless, the middle bin remains difficult to match exactly, suggesting that the degree of focal-point behavior is sufficiently large that even a two-regime mixture cannot fully reproduce the concentration of responses around income stability.

5.2.2 Nonlinear process

Figure 5 reports results from the nonlinear income process estimated with the hurdle model for middle-bin focal responses. The main qualitative findings from Figure 3 remain: subjective expectations imply income dynamics that are more persistent and less dispersed than those inferred from realized income data. Conditional dispersion remains substantially lower in the subjective income process than in the realized income process. Accounting for middle-bin focal reporting increases average subjective dispersion from around 0.02 in the baseline specification to approximately 0.06. Despite this increase, the realized-subjective dispersion gap remains large. Thus, allowing for this specific concentration distortion moves subjective dispersion toward the realized estimates, but is far from reconciling the two processes. The most notable change concerns conditional

Figure 4: Dispersion and model fit in the AR(1) process with hurdle behavior



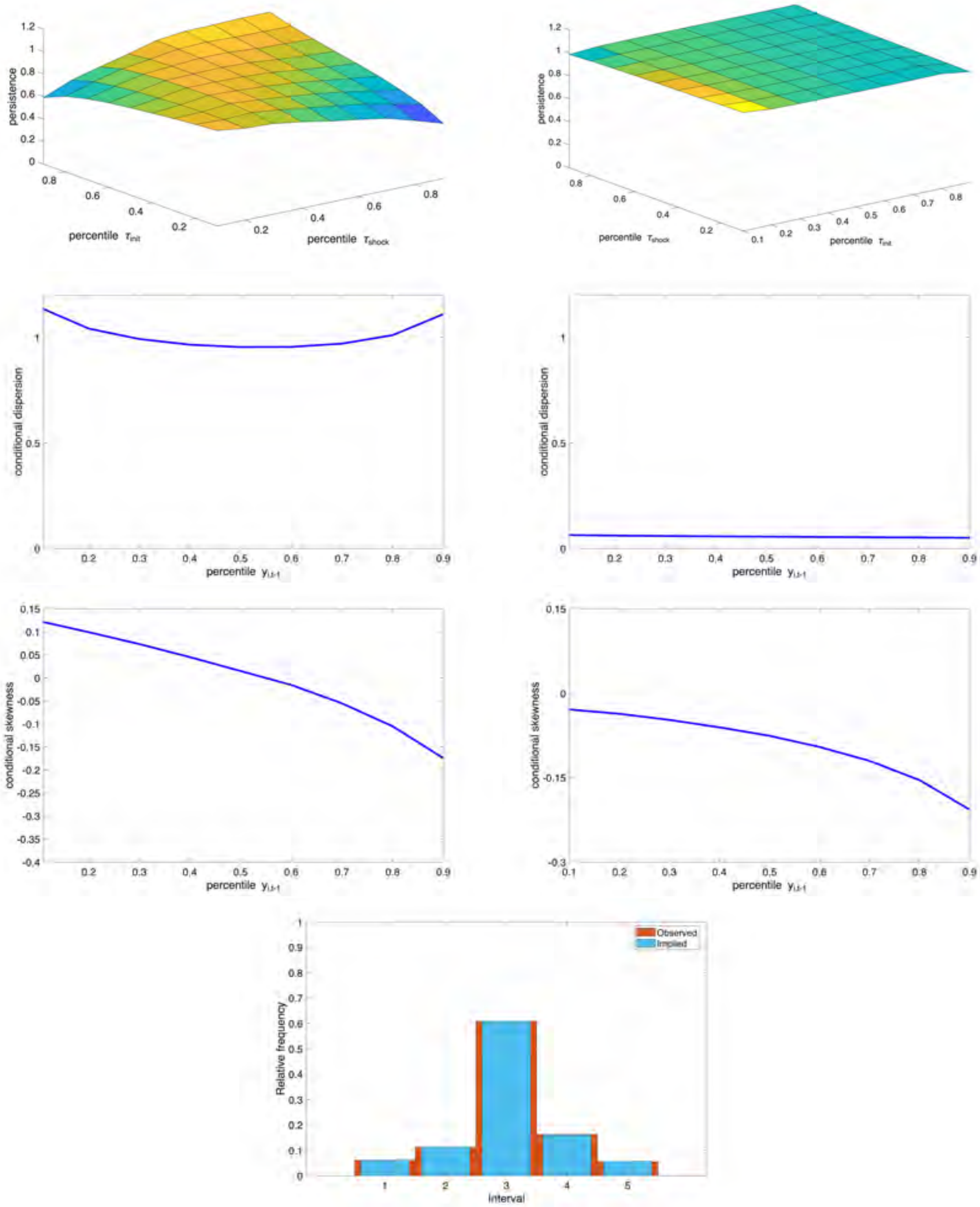
Notes: The top row shows the distribution of dispersion for the AR(1) process estimated with the hurdle model for middle-bin focal responses. The left panel corresponds to the process estimated from realized income data, while the right panel corresponds to the process estimated from subjective expectations data. The bottom panel compares model-implied probabilities from the subjective income process with the observed survey probabilities.

skewness. With the hurdle correction, the subjective income process generates a skewness profile that is more consistent with the patterns documented by Arellano et al. (2017): high-income households perceive negative conditional skewness, whereas low-income households perceive little skewness. These results highlight the importance of distinguishing genuine beliefs about income risk from focal-point reporting behavior.

Figure J.3 in Appendix J reports the corresponding 95% nonparametric bootstrap confidence intervals for the subjective income process estimated with the hurdle model. As in the specification without the hurdle, the estimated nonlinear objects are precisely estimated.

Taken together, the nonlinear estimates confirm the main message from the AR(1) analysis: households perceive income as highly persistent and substantially less dispersed than realized income histories imply. At the same time, accounting for focal-point behavior is important for recovering richer features of the subjective conditional distribution, especially skewness.

Figure 5: Nonlinear persistence, dispersion, skewness, and model fit with hurdle behavior



Notes: The first row shows nonlinear persistence, the second row conditional dispersion, measured by the interquantile range in equation (6), and the third row conditional skewness, measured by the quantile-based skewness measure (7) for the nonlinear income process estimated with the hurdle model for middle-bin focal responses. The left column corresponds to the process estimated from realized income data, while the right column corresponds to the process estimated from subjective expectations data. The bottom panel compares model-implied probabilities from the subjective income process with the observed survey probabilities.

6 Implications for consumption insurance and wealth accumulation

We uncover substantial differences between the income processes inferred from realized income data and those recovered from subjective expectations. In this section, we employ a life-cycle model with heterogeneous agents and incomplete asset markets to study the key implications of these differences for wealth accumulation over the life-cycle and for consumption insurance.

Our aim is to assess whether each data-generating process, taken in isolation, leads to different economic outcomes, rather than to compare their relative performance within a structural framework in matching a specific set of moments. Given the distinct patterns of wealth accumulation and consumption insurance we obtain, such a comparison would only be informative if the underlying drivers of these differences were explicitly modeled and accounted for—a task we leave for future research.[24]

We first present the main features of our model economy and describe its calibration, discussing how we embed the estimated income processes into the framework. The main insights are presented next.

6.1 Model

Each period a continuum of agents is born. Agents live a maximum of T periods, facing a probability s_j of surviving up to age j conditional upon being alive at age $j-1$ (survival probability is zero at period $T+1$). Population grows at a constant rate g_n.[25] Furthermore, we assume agents are born with no assets and face a mandatory retirement age of $j = J_R$. The asset holdings of agents who die are discarded.

Age-j agents select how much to consume c_j and next-period asset holdings a_{j+1} (subject to a borrowing limit A) to maximize their utility $u(c_j)$. We assume retirees

[24] As discussed in the next section, we conjecture that the different data-generating processes we obtain may reflect that households have superior information about their income risk relative to the econometrician, or that they face cognitive limitations or exhibit biases when assessing the probability of future states of the world.

[25] We assume that age-j agents always constitute a fraction μ_j of the population at any point in time, so that the demographic structure is stationary. The weights μ_j are normalized to sum to 1, and are given by the recursion $\mu_{j+1} = \frac{s_{j+1}}{(1+g_n)}\mu_j$.

do not work and receive a pension PB_j that depends on their income before retirement (y_{J_R}). Working-age agents ($j < J_R$) supply one unit of labor inelastically at zero disutility, receiving stochastic log income y_j, which follows either the nonlinear process estimated from realized data (NL) or the nonlinear process estimated from subjective expectations, with or without accounting for focal-point behavior (NL-SE Focal and NL-SE, respectively). Agents pay income taxes, denoted by T_j, depending on their total income $I_j = exp(y_j) + ra_j$, where r is the net return on asset holdings a_j. Income taxes are given by $\tau(I_j)I_j$, where the tax rate $\tau(I_j)$ is a function of total income. Finally, agents may receive transfers $G(I_j)$.[26]

Workers of age $j < J_R$ solve the recursive problem

$$\begin{aligned}
V_j(y,a) &= \max_{a',c} \left\{ \frac{c^{1-\sigma}}{1-\sigma} + \beta s_{j+1} \mathbb{E}\left[V_{j+1}(y',a') \mid y\right] \right\}, \\
\text{s.t.} \quad a' &= exp(y) - T_j(I_j) + G(I_j) - c + (1+r)a, \qquad a' \geqslant A, \\
y' &= Q_j^P(y,u'), \qquad u' \sim U[0,1], \\
P &\in \{NL,\ NL\text{-}SE,\ NL\text{-}SE \text{ Focal}\}.
\end{aligned}$$

Retirees of age $J_R \leqslant j \leqslant T$ solve the recursive problem

$$\begin{aligned}
W_j(y_{J_R},a) &= \max_{a',c} \left\{ \frac{c^{1-\sigma}}{1-\sigma} + \beta s_{j+1} W_{j+1}(y_{J_R},a') \right\}, \\
\text{s.t.} \quad a' &= PB_j(y_{J_R}) - c + (1+r)a, \qquad a' \geqslant A.
\end{aligned}$$

The model is solved recursively by backward induction over the life-cycle holding interest rates fixed and then simulated for 100,000 households given a set of structural parameters, discussed next.

[26] We assume

$$\tau(I_j) = \begin{cases} 0 \text{ if } I_j \leqslant \tilde{I} \\ max\{1 - \lambda(I_j/\bar{I})^{-\tau_I}, 0\} \text{ if } I_j > \tilde{I}, \end{cases} \tag{19}$$

where $\bar{I}$ is the mean income and $\tilde{I}$ is a tax income threshold, and government transfers are given by

$$G(I_j) = \begin{cases} g_0 \bar{I} \text{ if } I_j = 0 \\ max\{g_1 - g_2(I_j/\bar{I}), 0\}\bar{I} \text{ if } I_j > 0. \end{cases} \tag{20}$$

6.2 Calibration

Individuals start life at age 25, retire at age 65 and live up to a maximum possible age of 85. This implies that $J_R = 41$ and $T = 61$. The population growth rate is 0.6% per year, corresponding to the actual growth rate for the period 1990-2024 in Spain (Source: World Development Indicators). Survival probabilities (s_j, for $j = 25$ to 85) are set according to the data from the Spanish Statistical Office (INE) for the year 2023.

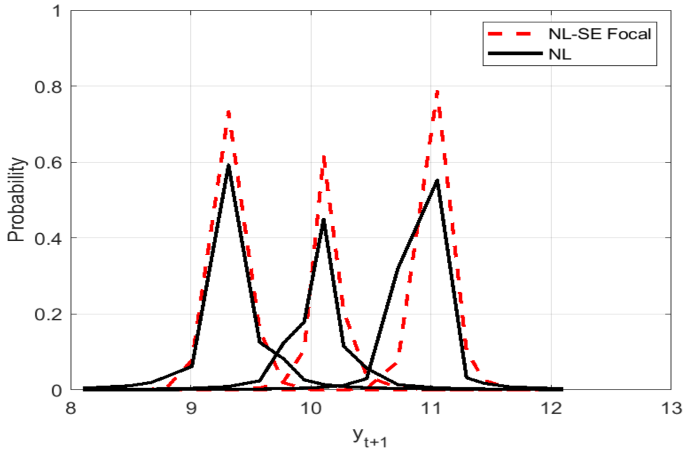
The coefficient of relative risk aversion is set to 2, a standard value. The risk-free rate is 2.5% and the discount factor β is calibrated to match a wealth-to-income ratio of 4, calculated from the EFF data. Finally, we set the exogenous borrowing limit A to 16% of average disposable household earnings, matching the liquid debt-to-income ratio of households who have liquid debt in the EFF.

Parameters λ, τ_I and $\tilde{I}$ determine the nonlinear income tax function in the model. We set $\lambda = 0.91255$ and $\tau_I = 0.1209$ and $\tilde{I} = 0.47\bar{I}$, following García-Miralles, Guner, and Ramos (2019). We set pension benefits such that the average benefit ratio is equal to 64.1% (the benefit ratio in 2022; see the European Commission's 2024 Ageing Report). Furthermore, the minimum pension is set to match income at the 15th percentile of the pre-retirement income distribution, and the maximum pension is set to match income at the 75th percentile, roughly matching the redistribution features of the pension system in Spain. The values of g_0, g_1 and g_2 of the transfer function $G(I)$ are estimated by Guner, Kaya, and Sánchez-Marcos (2024), using the data from the EU-Statistics on Income and Living Conditions. We follow their results and set $g_0 = 0.138$, $g_1 = 0.03$ and $g_2 = 0.0129$.

We use the estimation results of the income process presented above to simulate data for 100,000 households. We then discretize $z \in [-\infty, \infty]$ into a vector $z = [z_1, ..., z_n]$ and utilize the simulated data to obtain the transition matrix $\mathbb{Q}(z_i \mid z_h)$ for $i, h \in [1, n]$. As an illustration, Figure 6 depicts a subset of rows of the transition matrices for $n = 18$.[27]

[27]The EFF is conducted at three-year intervals; thus, the time period in the simulated data is three years. In the model, each period corresponds to one year, and hence we must adjust the transition matrix based on realized data. For the subjective expectations case the question explicitly refers to income variation in 12 months, so no adjustment is needed. Let $\mathbb{Q}_3$ denote the matrix from the empirical estimates and $\mathbb{Q}_1$ the modified matrix to be used for the yearly transitions. To ensure consistency, we must have $\mathbb{Q}_1^3 = \mathbb{Q}_3$. We first calculate $\tilde{\mathbb{Q}}_1 = \mathbb{Q}_3^{(1/3)}$. However, given the nonlinearities present in the estimated income process and reflected in matrix $\mathbb{Q}_3$, $\tilde{\mathbb{Q}}_1$ may contain negative elements. There is no unique way to adjust the transition matrices. Thus, if that occurs, we modify $\tilde{\mathbb{Q}}_1$ such that, for each

Figure 6: Transition Matrices



These matrices further illustrate the differences between the estimated processes using realized data (NL) and subjective expectations data accounting for focal behavior (NL-SE Focal). First, the variance of next-period income implied by the transition matrix is smaller under NL-SE Focal. Second, the process estimated from realized data assigns a substantially higher probability to large negative shocks for high-income households and to large positive shocks for low-income households, relative to the corresponding probabilities for a median-earning household; these transition probabilities vary nonlinearly with current income. We find little evidence of these nonlinearities in the NL-SE Focal process.

Our aim is to compare wealth accumulation and consumption insurance across the different estimated income processes. Given that the variance of next-period income implied by the transition matrices under NL-SE Focal and NL-SE is lower, keeping the structural parameters unchanged while varying only the income process would result in asset-to-income ratios that are significantly lower than those obtained under NL, the process estimated using realized data. To avoid comparing economies with substantially different average saving rates, we keep all structural parameters other than the discount factor constant and recalibrate β separately for each income process so that the asset-

row (initial z), the expected income in the next period, $E(z' \mid z)$, and its variance, $Var(z' \mid z)$, remain unchanged, and the sum of the three elements around the diagonal remains the same (for row j, the elements in columns $j-1$, j, and $j+1$ are summed, while for the first and last rows, the two elements to the right and to the left of the diagonal, respectively, are used in the sum). For robustness we also consider the model in which the period is set to three years, where such adjustments are not necessary.

to-income ratio matches that observed in the data. We present the results with a fixed β in Appendix I.

Finally, we set the model period to one year, as is standard in life-cycle models, and adjust the transition matrix of NL to account for the fact that the estimation is based on three-year intervals. In Appendix I, we also present an alternative specification in which the model period is set to three years and the transition matrices of NL-SE Focal and NL-SE are adjusted accordingly ($\mathbb{Q}_3 = \mathbb{Q}_1^3$, where $\mathbb{Q}_1$ is the transition matrix obtained from the simulation exercises for these two income processes). Our main qualitative conclusions remain unchanged across all the alternative specifications.

6.3 Results

We focus on two key simulated outcomes of our life-cycle model: wealth accumulation and consumption insurance.

Figure 7 plots average wealth over the life-cycle. Under the process estimated from realized income (NL), its greater dispersion and, crucially, its higher probability of large negative shocks for high-income households induce high-income households to save substantially more early in the life-cycle. Under the estimated processes using subjective expectations data (NL-SE Focal and NL-SE), income-shock dispersion is substantially lower, and wealth accumulation is slower earlier in the life-cycle and follows a different age profile: households that can save engage in less precautionary saving. Because high-income households save more in response to these nonlinear income dynamics, the economy under the NL also displays a significantly higher cross-sectional variance of wealth in the early part of the life-cycle, as illustrated in Figure 8.

Second, we focus on consumption insurance. We follow De Nardi et al. (2020) and calculate the insurance coefficient of household i's consumption (c_{it}) with respect to income shocks (μ_{it}), as proposed by Blundell, Pistaferri, and Preston (2008), which is given by[28]

[28]This measure can be obtained if the shock μ_{it} is known, which is the case in our model economy. An alternative measure under AR(1) shocks can be obtained by $\tilde{\psi} = 1 - \frac{\text{cov}(\Delta c_{it}, y_{it+1} - y_{it-2})}{\text{cov}(\Delta y_{it}, y_{it+1} - y_{it-2})}$. Results are largely unchanged using this alternative measure.

$$\psi = 1 - \frac{\operatorname{cov}(\Delta c_{it}, \mu_{it})}{\operatorname{var}(\mu_{it})} \tag{21}$$

Table 5 shows the results. We consider four versions of the model: *NL*; *NL-SE*; *NL-SE* Focal, which accounts for focal-point behavior; and a fourth case in which households expect the income process to follow *NL-SE* Focal, consistent with their subjective expectations, but realized income shocks follow the *NL* process. This exercise effectively assumes that the difference between the two estimated processes is entirely due to cognitive biases in household expectation formation. In the next section, we discuss the other main alternative explanation, which relies on households possessing superior private information.

Figure 7: Wealth Accumulation

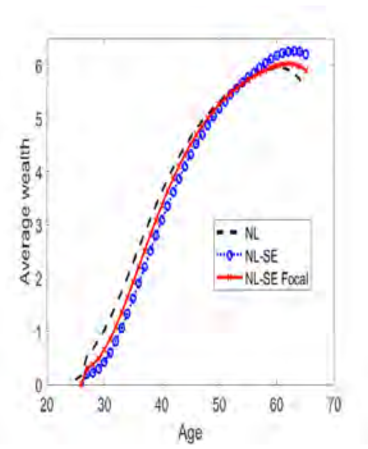


Figure 8: Cross-sectional Variance of Wealth

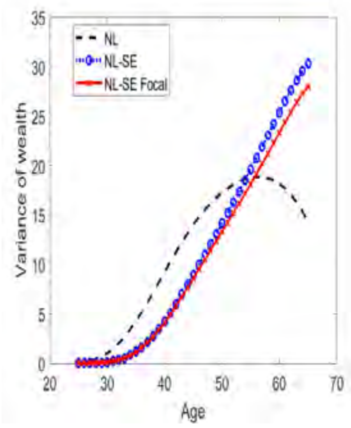


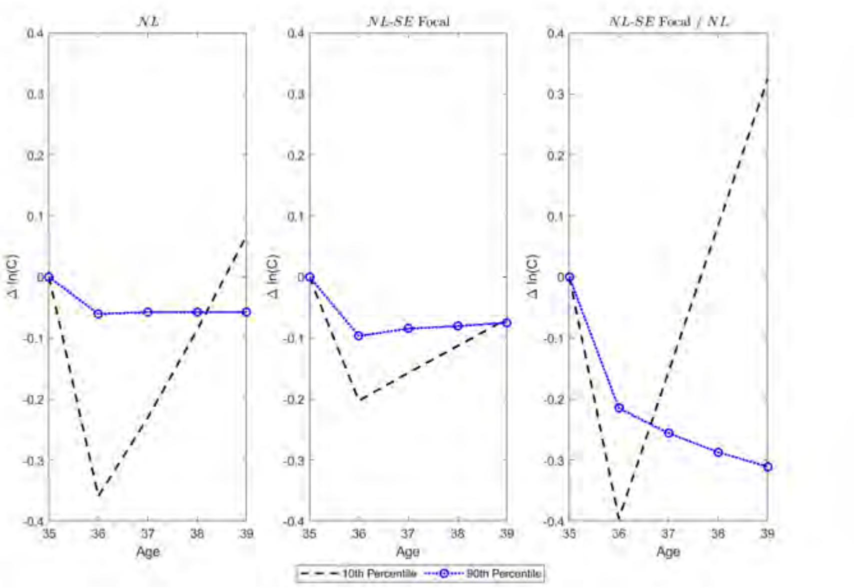
Table 5: Consumption Insurance

Income process	ψ
NL	0.781
NL-SE	0.730
NL-SE Focal	0.779
NL-SE Focal / NL	0.653

We find that the degree of consumption insurance in the first three cases is very similar (around 0.78 in two cases, a bit lower for the *NL-SE* where the variance of simulated income shocks is the lowest and thus households save very little in the beginning of the life-cycle). First, note that in our model income shocks can be either transitory or persistent, while subjective expectations concern total income without distinguishing between the two. Accordingly, the implied degree of consumption insurance lies between the benchmark values typically associated with transitory shocks (around 0.4) and persistent shocks (around 0.9) (see De Nardi et al. (2020)). Second, although the overall level of consumption insurance is similar across the first three cases, the underlying mechanisms differ. In the first case, households rely on asset accumulation to smooth relatively large income shocks. In the second, income variance is much lower, so consumption fluctuations are correspondingly limited. Finally, if households save in a manner consistent with their subjective expectations—anticipating the lower dispersion implied by the perceived process—but actual shocks follow the higher-variance, nonlinear process estimated from realized data, consumption insurance becomes more limited and income and consumption move more closely together. Thus, if cognitive biases distort households' assessment of the probabilities of extreme income changes and influence their saving decisions, consumption self-insurance may be impaired.

To illustrate the different degrees of insurance that each specification affords across the distribution, we plot (Figure 9) the consumption response for households in the 10th and 90th percentiles of the wealth distribution after they face a negative shock equivalent to the 5th percentile of their future income distribution. We impose these shocks on 35-year-old individuals, assuming that throughout their life-cycle up to this point, and thereafter, they receive a series of shocks consistent with their income process.

Figure 9: Consumption Impulse Responses



Under the NL income process, rich households, anticipating the likelihood of nonlinear shocks, accumulate more assets and, despite receiving large negative shocks, are able to smooth consumption to a significant extent. Such shocks effectively wipe out the memory of previous favorable shocks, permanently reducing consumption. Poor households hold significantly fewer assets and, when faced with large shocks, experience substantial consumption losses. However, given the relatively high likelihood of receiving favorable shocks that offset past negative shocks (the "good" aspect of the nonlinearity in NL), consumption eventually returns to its pre-shock level.

By contrast, under the NL-SE Focal process, both poor and rich households accumulate insufficient assets. As a result, even though they face relatively smaller shocks, they are similarly unable to smooth consumption, and consumption responses across the wealth distribution are more homogeneous than in the NL case. Therefore, while average consumption smoothing appears similar across the two income processes in the cross section, this masks substantial heterogeneity across households.

Finally, we present impulse responses for the case in which households choose their savings assuming a low-dispersion, high-persistence income process (NL-SE Focal), but instead experience shocks consistent with a higher-dispersion, lower-persistence process (NL). In this scenario, both rich and poor households incur significant consumption losses, implying that insurance is severely compromised. Due to the unexpected nonlinearity, rich households are unable to smooth consumption and continue to experience losses as they struggle to return to their previous income level. Poor households, in contrast, eventually recover their consumption, as they are subsequently hit by unanticipated positive shocks.

7 Predictive and behavioral content of subjective expectations

The preceding analysis shows that income dynamics recovered from subjective expectations differ substantially from those inferred from realized income data and that these differences have important implications for household behavior. Interpreting this gap, however, depends in part on whether subjective expectations reflect biased beliefs or private information unavailable to the econometrician. In the former case, households may underweight low-probability or unusually large income changes when forming expectations, consistent with mechanisms emphasized by Bordalo et al. (2013). In the latter case, income dynamics inferred from realizations alone may overstate the income dispersion relevant for some households. We therefore examine both the predictive content of subjective income expectations for subsequent income and their relationship with household consumption choices.

For each household-wave observation, we construct household-specific measures of expected future income and expected dispersion following Arellano, Bonhomme, De Vera, Hospido, and Wei (2021). Specifically, we summarize the household's one-year-ahead subjective distribution using an auxiliary random-walk representation with household-specific drift and shock dispersion. Expected future log income is given by current log income plus the household-specific drift, while expected dispersion is measured by the household-specific standard deviation of income growth.[29]

Persistence does not enter these regressions because, in this auxiliary random-walk representation, it is fixed at one rather than estimated as a household-specific forecast moment. The purpose of the exercise is therefore to assess whether variation across households in the location and dispersion of their one-year-ahead subjective distributions contains predictive or behavioral information.

We first examine whether incorporating subjective expectations into a standard autoregressive model with household controls and wave fixed effects improves the prediction of future household income (Table 6). Expected income adds little predictive power once current income is included. By contrast, expected dispersion is economically and statistically significant, indicating that the subjective distribution contains information about subsequent income beyond current income and observable household characteristics.

Second, we examine whether subjective expectations are related to household consumption and elicited marginal propensities to consume (Table 7). Household consumption is measured as expenditure on food and other non-durable goods. The elicited MPC is based on a hypothetical windfall equal to one month of household income and asks what fraction the household would spend over the following twelve months rather than

[29]This auxiliary representation is used only to summarize the household-specific forecast distribution in the exercises of this section; it is distinct from the income-process specifications estimated in Sections 3–5.

Table 6: Predictability of household income

Variable	Future household income (in logs)
Expected household income (in logs)	0.047
	(0.067)
Household income	0.641***
	(0.068)
Expected dispersion of household income	-0.581***
	(0.210)
Constant	3.283***
	(0.216)
Dependent-variable mean	10.443
Observations	5,303
R-squared	0.608

Note: This table reports an OLS regression of future log household income on household-specific expected log income and expected dispersion, controlling for current log income, demographic characteristics, and wave fixed effects. Expected income and dispersion are constructed from the one-year-ahead subjective income distribution following Arellano et al. (2021), as described in the text. Standard errors are clustered by household. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

save.[30] Conditional on current income and wealth, expected income again adds little explanatory power, whereas greater expected dispersion is associated with lower consumption and a lower elicited MPC. These results provide complementary evidence that household-specific subjective expectations contain economically relevant information and are related to household decision-making.

[30]The exact question is: *Imagine that in a raffle, you won an amount of money equal to your household's monthly income. What percentage would you spend over the following twelve months, rather than save?* Responses range from 0 to 100 and are rescaled to the unit interval for the empirical analysis.

Table 7: Subjective income expectations and household consumption

Variables	(1) Log consumption	(2) Elicited MPC
Expected household income (in logs)	-0.035	0.067
	(0.038)	(0.062)
Household income (in logs)	0.451***	-0.037
	(0.042)	(0.064)
Expected dispersion of household income	-0.293**	-0.342*
	(0.130)	(0.192)
Household wealth (in logs)	0.034***	-0.013**
	(0.006)	(0.007)
Constant	4.675***	0.015
	(0.144)	(0.180)
Dependent-variable mean	9.653	0.381
Observations	8,823	7,221
R-squared	0.460	

Note: Column (1) reports an OLS regression of log household consumption on expected income and dispersion, current income, wealth, demographic characteristics, and wave fixed effects. Column (2) reports a Tobit regression of the elicited MPC, rescaled to the unit interval, on the same covariates. Expected income and dispersion are constructed from the one-year-ahead subjective income distribution following Arellano et al. (2021), as described in the text. Standard errors are clustered by household. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

8 Conclusions

Household consumption and saving decisions depend on the income process that households perceive, while most quantitative models discipline income risk using processes recovered from realized income data. This paper asks whether these two income processes coincide. To do so, we develop and estimate flexible income processes using probabilistic subjective expectations data and contrast them with processes inferred from realized income alone. We document systematic and economically meaningful differences between perceived and realized income dynamics. Households perceive income to be substantially more persistent and less dispersed than implied by ex-post realizations; the nonlinearities present in realized income data are substantially attenuated in subjective expectations data. Embedding these alternative income processes into a life-cycle model reveals that such differences have quantitatively important implications for precautionary saving, wealth accumulation, and the degree of consumption insurance.

The broader implication is not that we should replace realized income histories with subjective income expectations data in quantitative models. Rather, the two sources of information measure different objects that need not coincide. The more recent income dynamics literature has substantially improved our understanding of income dynamics ex post. Our results show that those dynamics need not coincide with the income dynamics households perceive ex ante and use when making intertemporal decisions. Distinguishing between these objects is therefore important both for measuring household income risk and for disciplining models of consumption, saving, and self-insurance.

An important open question is whether these discrepancies reflect biases in expectation formation or instead arise because households possess private, household-specific information that is not observed by the econometrician estimating income processes from realized data alone. Distinguishing between these two interpretations is crucial for both measurement and policy analysis: if the gap is driven by cognitive distortions, models calibrated on subjective expectations may misrepresent actual income risk; if it reflects superior private information, models based solely on realized income may overstate the income dispersion relevant for household decision-making.

References

Almuzara, Martín. (2020). "Heterogeneity in transitory income risk". Tech. rep., Federal Reserve Bank of New York.

Alvarez, Javier, and Manuel Arellano. (2022). "Robust likelihood estimation of dynamic panel data models". *Journal of Econometrics*, 226(1), pp. 21–61.

Arellano, Manuel, Orazio Attanasio, Margherita Borella, Mariacristina De Nardi and Gonzalo Paz-Pardo. (2026a). "Subjective earnings and employment dynamics". Tech. rep., National Bureau of Economic Research.

Arellano, Manuel, Orazio Attanasio, Samuel Crossman and Víctor Sancibrián. (2026b). "Estimating flexible income processes from subjective expectations data: evidence from india and colombia". Tech. rep., National Bureau of Economic Research.

Arellano, Manuel, Richard Blundell and Stéphane Bonhomme. (2017). "Earnings and consumption dynamics: a nonlinear panel data framework". *Econometrica*, 85(3), pp. 693–734.

Arellano, Manuel, and Stéphane Bonhomme. (2016). "Nonlinear panel data estimation via quantile regressions". *The Econometrics Journal*, 19(3), pp. C61–C94.

Arellano, Manuel, Stéphane Bonhomme, Micole De Vera, Laura Hospido and Siqi Wei. (2021). "Income risk inequality: evidence from spanish administrative records". Tech. rep., Banco de España.

Attanasio, Orazio, Agnes Kovacs and Krisztina Molnar. (2020). "Euler equations, subjective expectations and income shocks". *Economica*, 87(346), pp. 406–441.

Blundell, Richard, Luigi Pistaferri and Ian Preston. (2008). "Consumption inequality and partial insurance". *American Economic Review*, 98(5), pp. 1887–1921.

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

Bonhomme, Stéphane, and Angela Denis. (2025). "Fixed effects and beyond. bias reduction, groups, shrinkage and factors in panel data". Tech. rep., Banco de España.

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. (2013). "Salience and consumer choice". *Journal of Political Economy*, 121(5), pp. 803–843.

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. (2018). "Diagnostic expectations and credit cycles". *The Journal of Finance*, 73(1), pp. 199–227.

Bound, John, Charles Brown and Nancy Mathiowetz. (2001). "Measurement error in survey data". In *Handbook of econometrics*, vol. 5. Elsevier, pp. 3705–3843.

Bound, John, and Alan B Krueger. (1991). "The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?" *Journal of labor economics*, 9(1), pp. 1–24.

Bover, Olympia. (2015). "Measuring expectations from household surveys: new results on subjective probabilities of future house prices". *SERIEs*, 6(4), pp. 361–405.

Browning, Martin, Mette Ejrnaes and Javier Alvarez. (2010). "Modelling income processes with lots of heterogeneity". *The Review of Economic Studies*, 77(4), pp. 1353–1381.

Caplin, Andrew, Victoria Gregory, Eungik Lee, Søren Leth-Petersen and Johan Sæverud. (2026). "Subjective earnings risk". *The Review of Economic Studies*, p. rdag070. https://doi.org/10.1093/restud/rdag070

Chamberlain, Gary, and Keisuke Hirano. (1999). "Predictive distributions based on longitudinal earnings data". *Annales d'Economie et de Statistique*, pp. 211–242.

de Bresser, Jochem, and Arthur van Soest. (2013). "Survey response in probabilistic questions and its impact on inference". *Journal of Economic Behavior & Organization*, 96, pp. 65–84.

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

Dominitz, Jeff. (1998). "Earnings expectations, revisions, and realizations". *Review of Economics and statistics*, 80(3), pp. 374–388.

Dominitz, Jeff. (2001). "Estimation of income expectations models using expectations and realization data". *Journal of Econometrics*, 102(2), pp. 165–195.

Dominitz, Jeff, and Charles F Manski. (1997). "Using expectations data to study subjective income expectations". *Journal of the American statistical Association*, 92(439), pp. 855–867.

Gálvez, Julio, and Gonzalo Paz-Pardo. (2026). "Richer earnings dynamics, consumption and portfolio choice over the life cycle". *Journal of Financial Economics*, 176, p. 104206.

García-Miralles, Esteban, Nezih Guner and Roberto Ramos. (2019). "The spanish personal income tax: facts and parametric estimates". *SERIEs: Journal of the Spanish Economic Association*, 10(3), pp. 439–477. https://doi.org/10.1007/s13209-019-0197-5

Giustinelli, Pamela, Charles F Manski and Francesca Molinari. (2022). "Tail and center rounding of probabilistic expectations in the health and retirement study". *Journal of Econometrics*, 231(1), pp. 265–281.

Gu, Jiaying, and Roger Koenker. (2017). "Unobserved heterogeneity in income dynamics: An empirical bayes perspective". *Journal of Business & Economic Statistics*, 35(1), pp. 1–16.

Guiso, Luigi, Tullio Jappelli and Luigi Pistaferri. (2002). "An empirical analysis of earnings and employment risk". *Journal of Business & Economic Statistics*, 20(2), pp. 241–253.

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

Guner, Nezih, Ezgi Kaya and Virginia Sánchez-Marcos. (2024). "Labor market institutions and fertility". *International Economic Review*, 65(3), pp. 1551–1587. https://doi.org/10.1111/iere.12708

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

Heiss, Florian, Michael Hurd, Maarten van Rooij, Tobias Rossmann and Joachim Winter. (2022). "Dynamics and heterogeneity of subjective stock market expectations". *Journal of Econometrics*, 231(1), pp. 213–231.

Hospido, Laura. (2012). "Modelling heterogeneity and dynamics in the volatility of individual wages". *Journal of Applied Econometrics*, 27(3), pp. 386–414.

Jappelli, Tullio, and Luigi Pistaferri. (2010). "The consumption response to income changes". *Annu. Rev. Econ.*, 2(1), pp. 479–506.

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

Kleinjans, Kristin J, and Arthur van Soest. (2014). "Rounding, focal point answers and nonresponse to subjective probability questions". *Journal of Applied Econometrics*, 29(4), pp. 567–585.

Koşar, Gizem, and Davide Melcangi. (2025). "Subjective uncertainty and the marginal propensity to consume". Tech. rep., CESifo Working Paper.

Koşar, Gizem, and Wilbert Van der Klaauw. (2025). "Workers' perceptions of earnings growth and employment risk". *Journal of Labor Economics*, 43(S1), pp. S83–S121.

Kovacs, Agnes, Concetta Rondinelli and Serena Trucchi. (2021). "Permanent versus transitory income shocks over the business cycle". *European Economic Review*, 139, p. 103873.

Lee, Wooyong. (2025). "Identification and estimation of dynamic random coefficient models". *arXiv preprint arXiv:2505.01600.*

Manski, Charles F, and Francesca Molinari. (2010). "Rounding probabilistic expectations in surveys". *Journal of Business & Economic Statistics*, 28(2), pp. 219–231.

Manski, Charles F, and Elie Tamer. (2002). "Inference on regressions with interval data on a regressor or outcome". *Econometrica*, 70(2), pp. 519–546.

Meghir, Costas, and Luigi Pistaferri. (2004). "Income variance dynamics and heterogeneity". *Econometrica*, 72(1), pp. 1–32.

Pischke, Jörn-Steffen. (1995). "Measurement error and earnings dynamics: Some estimates from the psid validation study". *Journal of Business & Economic Statistics*, 13(3), pp. 305–314.

Pistaferri, Luigi. (2001). "Superior information, income shocks, and the permanent income hypothesis". *Review of Economics and Statistics*, 83(3), pp. 465–476.

Rozsypal, Filip, and Kathrin Schlafmann. (2023). "Overpersistence bias in individual income expectations and its aggregate implications". *American Economic Journal: Macroeconomics*, 15(4), pp. 331–371.

Stoltenberg, Christian, and Arne Uhlendorff. (2022). "Consumption choices and earnings expectations: Empirical evidence and structural estimation". Tech. rep., Tinbergen Institute Discussion Paper.

Van der Klaauw, Wilbert, Wändi Bruine de Bruin, Giorgio Topa, Simon Potter and Michael F Bryan. (2008). "Rethinking the measurement of household inflation expectations: preliminary findings". *FRB of New York Staff Report*, (359).

Wang, Tao. (2023). "Perceived versus calibrated income risks in heterogeneous-agent consumption models". Tech. rep., Bank of Canada.

Online Appendix for "Perceived versus Realized Income Dynamics: Evidence from Subjective Expectations"

A Additional data details and summary statistics

Section 2 describes the EFF panel structure and the construction of our estimation sample. This appendix provides additional information on the variables used in the empirical analysis and reports summary statistics for that sample.

Variable definitions. Realized income is reported total pre-tax household income, including transfers. Household net wealth is constructed as the sum of total real and financial assets (including real estate value, car value, cash, pension funds, stocks, and bonds) net of total debt (mortgages and other debt). Household consumption is the sum of expenditure on food and other non-durable goods. We also use the subjective house price expectations question described in Bover (2015). Like the income expectations question, it elicits households' subjective probability distribution over changes in the value of their primary residence.

Summary statistics. Table A.1 reports summary statistics for the estimation sample.

Table A.1: Summary statistics, EFF 2014–2022 panel

Variable	Mean	Standard Deviation	Minimum	Maximum
Household income	45355.34	45645.72	104.46	7885194
Household wealth	264038	1161995	1.0111	2.54e+08
Non-durable consumption	15089.32	9472.724	1007.056	662566
Age of head	48.2669	9.298195	25	65
Head is woman	.4090239	.4916816	0	1
Number of household members	3.108887	1.300956	1	9
Second adult income earner	.5275598	.4992682	0	1
Third (or more) adult income earner	.1197938	.3247387	0	1

Notes: All monetary amounts are in 2017 euros. Statistics are calculated for the estimation sample described in Section 2.

Coherence with alternative income expectations measures. To check the coherence of the probabilistic income expectations question, we compare summary measures of expected income and dispersion following Arellano et al. (2021) to a qualitative question in the EFF that asks households if income will increase, decrease or stay the same.

The results, which are in Table A.2, indicate that there is cross-question coherence between the income expectations measures. In particular, households who respond that income will increase (decrease) in the future tend to have a higher (lower) mean expected income. Both groups also display higher dispersion than households reporting that income will stay the same.

Table A.2: Mean and Dispersion by Income Change Category

	Mean	Std. Error	95% Confidence Interval	Observations
Panel A: Mean, μ				
Income increases	0.0713	0.0033	[0.0649, 0.0777]	2,819
Income decreases	-0.1389	0.0061	[-0.1510, -0.1269]	1,197
Income stays the same	-0.0059	0.0011	[-0.0081, -0.0037]	4,807
Panel B: Dispersion, σ				
Income increases	0.0683	0.0010	[0.0664, 0.0703]	2,819
Income decreases	0.0861	0.0020	[0.0823, 0.0900]	1,197
Income stays the same	0.0362	0.0004	[0.0354, 0.0371]	4,807

Notes: Expected mean and dispersion are computed via the method in Arellano et al. (2021).

B Estimation of the realized income processes

In this section, we outline the estimation of the income processes we described in section 3 using realized income data.

B.1 State-dependent AR(1) process

Recall the specification of the state-dependent AR(1) process:

$$y_{it+1} = \rho(\mathbf{X}_{it})\, y_{it} + \mathbf{X}'_{it}\beta + \eta_i + \sigma(\mathbf{X}_{it})\, u_{it+1}, \qquad u_{it+1} \sim N(0,1), \tag{B1}$$

To estimate this model, we follow the approach in Alvarez and Arellano (2022) and consider a random effects specification for η_i. In particular, we assume that

$$\eta_i \sim N(0, \sigma_\eta^2), \quad \eta_i \perp u_{it+1} \text{ for all } t, \quad \eta_i \perp \mathbf{X}_{it} \text{ for all } t. \tag{B2}$$

Let us define $\mu_{it}(\theta, \eta_i) \equiv \rho(\mathbf{X}_{it}; \pi) y_{it} + \mathbf{X}'_{it}\beta + \eta_i$, and $\sigma_{it}(\theta) \equiv \sigma(\mathbf{X}_{it}; \gamma)$. Then, the conditional distribution of y_{it+1} given y_{it}, $\mathbf{X}_{it}$ and η_i is

$$y_{it+1} | y_{it}, \mathbf{X}_{it}, \eta_i \sim N(\mu_{it}(\theta, \eta_i), \sigma_{it}^2(\theta))$$

with density

$$f(y_{it+1} | y_{it}, \mathbf{X}_{it}, \eta_i; \theta) = \frac{1}{\sigma_{it}(\theta)} \phi\left(\frac{y_{it+1} - \mu_{it}(\theta, \eta_i)}{\sigma_{it}(\theta)}\right), \tag{B3}$$

where $\phi(\cdot)$ is the standard Normal pdf. Assuming conditional independence over t given η_i,

$$L_i(\theta|\eta_i) = \prod_{t=1}^{T} f(y_{it+1}|y_{it}, \mathbf{X}_{it}, \eta_i; \theta). \tag{B4}$$

Integrating over the distribution of η_i, we obtain the unconditional likelihood contribution for household i:

$$L_i(\theta) = \int \prod_{t=1}^{T} f(y_{it+1} \mid y_{it}, \mathbf{X}_{it}, \eta_i; \theta) \frac{1}{\sigma_\eta} \phi\left(\frac{\eta_i}{\sigma_\eta}\right) d\eta_i. \tag{B5}$$

The maximum likelihood estimator of the parameters of the income process can be obtained by maximizing over the average log-likelihood

$$\hat{\theta} = \arg\max_{\theta} \ell(\theta), \text{ where } \ell(\theta) = \sum_{i=1}^{N} \log L_i(\theta). \tag{B6}$$

Because the optimization problem requires integrating over the distribution of η, we evaluate the integral numerically by Gauss-Hermite quadrature methods. In our empirical implementation, we assume that $\rho(\mathbf{X}_{it})$ is a logistic function and that $\sigma(\mathbf{X}_{it})$ is an exponential function.

B.2 Nonlinear income process

Following Arellano et al. (2017), we approximate the conditional quantile function of y_{it+1} given y_{it} and covariates $\mathbf{X}_{it}$ via the following

$$Q_y(\tau \mid y_{it}, \mathbf{X}_{it}) = \sum_{k=0}^{K} a_k(\tau)\, \psi_k(y_{it}) + \mathbf{X}_{it}'\beta(\tau), \tag{B7}$$

where $a_k(\tau)$ and $\beta(\tau)$ are piecewise-linear splines, and ψ_k is a low-order Hermite polynomial. Given this specification, we solve, for a given set of quantiles the following optimization problem:

$$(\widehat{\alpha}(\tau_l), \widehat{\beta}(\tau_l)) = \arg \min_{a_k(\tau),\beta(\tau)} \sum_{i=1}^{N}\sum_{t=1}^{T} \rho_{\tau_l}\left(y_{it+1} - \sum_{k=0}^{K} a_k(\tau)\, \psi_k(y_{it}) - \mathbf{X}_{it}'\beta(\tau)\right).$$

C Estimating the nonlinear income process using subjective expectations data

C.1 Approximation of the conditional distributions

In this subsection, we write the analytical form of the approximating cdf and pdf. Following Arellano and Bonhomme (2016), the conditional distribution of next-period income can be represented through the conditional quantile function. To simplify the discussion, we first ignore the conditioning on state variables $\mathbf{X}_{it}$. Recall that the conditional quantile function of y_{it+1} given y_{it} is

$$Q_y(\tau \mid y_{it}) = \sum_{k=0}^{K} a_k(\tau)\psi_k(y_{it}), \tag{C8}$$

where $a_k(\tau)$ are piecewise-linear interpolating splines in τ.

Given this specification, we can write the approximating CDF and PDF implied by the quantile function. Let

$$A_{it}(\ell) \equiv Q_y(\tau_\ell \mid y_{it})$$

denote the value of the conditional quantile function at knot τ_ℓ. Then the approximating CDF $F_{it}(y)$ and PDF $f_{it}(y)$ are defined by three regions.

For the left tail, if $y < A_{it}(1)$,

$$F_{it}(y) = \tau_1 \exp\left\{\lambda_1\big(y - A_{it}(1)\big)\right\},$$

and therefore

$$f_{it}(y) = \lambda_1\tau_1 \exp\left\{\lambda_1\big(y - A_{it}(1)\big)\right\}.$$

For the interior segments, if $A_{it}(\ell) \leqslant y < A_{it}(\ell+1)$ for some $\ell = 1, \ldots, L-1$,

$$F_{it}(y) = \tau_\ell + \frac{\tau_{\ell+1} - \tau_\ell}{A_{it}(\ell+1) - A_{it}(\ell)}\big(y - A_{it}(\ell)\big).$$

This implies that the pdf is constant within each segment:

$$f_{it}(y) = \frac{\tau_{\ell+1} - \tau_\ell}{A_{it}(\ell+1) - A_{it}(\ell)}.$$

For the right tail, if $y > A_{it}(L)$,

$$F_{it}(y) = 1 - (1 - \tau_L) \exp\left\{-\lambda_L\big(y - A_{it}(L)\big)\right\},$$

and therefore

$$f_{it}(y) = \lambda_L(1 - \tau_L) \exp\left\{-\lambda_L\big(y - A_{it}(L)\big)\right\}.$$

When the state variables $\mathbf{X}_{it}$ are included, the conditional quantile function becomes

$$Q_y(\tau \mid y_{it}, \mathbf{X}_{it}) = \sum_{k=0}^{K} a_k(\tau)\psi_k(y_{it}) + \mathbf{X}_{it}'\beta(\tau), \tag{C9}$$

and we write

$$A_{it}(\ell;\theta) \equiv Q_y(\tau_\ell \mid y_{it}, \mathbf{X}_{it}; \theta).$$

The corresponding CDF and PDF are denoted by $F_{it}(y;\theta)$ and $f_{it}(y;\theta)$.

Observed data likelihood. Let $I_{j,it} \equiv (\kappa_{j-1,it}, \kappa_{j,it}]$ denote bin j for respondent i at time t. Given the nonlinear income process, the model-implied bin probabilities are

$$p_{j,it}(\theta) = \Pr(y_{it+1} \in I_{j,it} \mid y_{it}, \mathbf{X}_{it}; \theta) = F_{it}(\kappa_{j,it}; \theta) - F_{it}(\kappa_{j-1,it}; \theta). \tag{C10}$$

Interpreting the M reported points as M exchangeable draws from the subjective distribution, the observed-data log-likelihood is

$$\ell_N(\theta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\log \frac{M!}{\prod_{j=1}^{J} d_{jit}!} + \sum_{j=1}^{J} d_{jit} \log \left(p_{j,it}(\theta) \right) \right]. \tag{C11}$$

C.2 Coarse information

Maximizing the observed-data log-likelihood (C11) directly is computationally demanding because of the constraints needed to guarantee monotonicity and because the resulting objective function is highly nonlinear. Hence, to estimate the parameters of the nonlinear income process, we appeal to a stochastic pseudo-EM algorithm in the spirit of Arellano and Bonhomme (2016). In particular, we interpret the M different points as possible realizations from the conditional distribution $f_{it}(y_{it+1}|y_{it}; \theta)$, and estimate its parameters via quantile regressions. The estimated parameters of the conditional distribution are those that are used to evaluate the model-implied bin probabilities (C10), and the observed data log-likelihood (C11).

In practice, the algorithm works in the following manner. For a given initial parameter estimate $\theta^{(0)}$, we iterate on $s = 0, 1, 2, \ldots$ until convergence using the following two steps.

E-step. For each (i, t) and for each bin j, draw d_{jit} latent values from the posterior density

$$f_{it}(y \mid y \in I_{j,it}; \theta^{(s)}).$$

In practice, for each point in bin j, the steps to draw the latent value are:

1. Compute
$$a_{j,it}^{(s)} = F_{it}(\kappa_{j-1,it}; \theta^{(s)}), \qquad b_{j,it}^{(s)} = F_{it}(\kappa_{j,it}; \theta^{(s)}),$$
where $F_{it}(\cdot; \theta^{(s)})$ is the approximating CDF defined in Subsection C.1.

2. Draw
$$u_{itjm}^{(s)} \sim U\left[a_{j,it}^{(s)}, b_{j,it}^{(s)}\right].$$

3. Set
$$y_{itjm}^{(s)} = Q_y\left(u_{itjm}^{(s)} \mid y_{it}, \mathbf{X}_{it}; \theta^{(s)}\right).$$

We finally collect the pooled augmented sample

$$\{y_{itjm}^{(s)}, y_{it}, \mathbf{X}_{it}\}_{i,t,j,m},$$

with exactly M simulated draws for each (i, t).

M-step. In the M-step, we update both the parameters of the quantile function and the tail parameters. To update the parameters of the quantile function, we estimate quantile regressions at the spline knots $\{\tau_\ell\}_{\ell=1}^L$ on the augmented sample. For each knot τ_ℓ, estimate

$$(\widehat{a}(\tau_\ell), \widehat{\beta}(\tau_\ell)) = \arg\min_{a(\tau_\ell),\beta(\tau_\ell)} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{J}\sum_{m=1}^{d_{jit}} \rho_{\tau_\ell}\left(y_{itjm}^{(s)} - \sum_{k=0}^{K} a_k(\tau_\ell)\psi_k(y_{it}) - \mathbf{X}_{it}'\beta(\tau_\ell) \right), \tag{C12}$$

where $\rho_\tau(\cdot)$ is the standard check function. After estimating the quantile regressions, we apply the ABB rearrangement step to enforce monotonicity of $\tau \mapsto Q_y(\tau \mid y_{it}, \mathbf{X}_{it})$.

We update the tail parameters using the exponential tail restrictions. Let

$$A_{it}^{(s+1)}(1) = Q_y(\tau_1 \mid y_{it}, \mathbf{X}_{it}; \theta^{(s+1)}), \qquad A_{it}^{(s+1)}(L) = Q_y(\tau_L \mid y_{it}, \mathbf{X}_{it}; \theta^{(s+1)})$$

denote the updated boundary quantiles. Index the pooled simulated draws by $q \equiv (i, t, j, m)$ and write

$$y_q \equiv y_{itjm}^{(s)}, \qquad A_q(1) \equiv A_{it}^{(s+1)}(1), \qquad A_q(L) \equiv A_{it}^{(s+1)}(L).$$

Define

$$N_{\text{left}} = \#\{q : y_q < A_q(1)\}, \qquad N_{\text{right}} = \#\{q : y_q > A_q(L)\}.$$

Then the MLE updates for the exponential tail rates are

$$\widehat{\lambda}_1^{(s+1)} = \left[\frac{1}{N_{\text{left}}} \sum_{q:\, y_q < A_q(1)} \big(A_q(1) - y_q\big) \right]^{-1}, \qquad \widehat{\lambda}_L^{(s+1)} = \left[\frac{1}{N_{\text{right}}} \sum_{q:\, y_q > A_q(L)} \big(y_q - A_q(L)\big) \right]^{-1}. \tag{C13}$$

C.3 Hurdle model in the nonlinear process

This subsection extends the hurdle reporting model to the nonlinear income process defined in subsections C.1 and C.2. The underlying income process is unchanged. For a given parameter vector θ, which includes the parameters of the conditional quantile function and the exponential-tail parameters, the model-implied probability of bin j is

$$p_{j,it}(\theta) = F_{it}(\kappa_{j,it}; \theta) - F_{it}(\kappa_{j-1,it}; \theta),$$

where $F_{it}(\cdot;\theta)$ is the approximating conditional CDF implied by the nonlinear conditional quantile function.

The additional component is a reporting equation that allows some respondents to place all M probability points in the middle bin deterministically. Let j^* denote the middle bin and define

$$S_{it} = \mathbf{1}\{d_{j*,it} = M\}.$$

Let $R_{it} \in \{0,1\}$ indicate whether household i at time t is in the deterministic focal-point regime.

Reporting equation. We use the non-centered random-effects representation

$$b_i = \sigma_b v_i, \qquad v_i \sim N(0,1),$$

and specify

$$\pi_{it}(v_i;\gamma,\sigma_b) \equiv \Pr(R_{it} = 1 \mid Z_{it}, v_i) = \Lambda\left(\gamma_0 + Z_{it}'\gamma_1 + \sigma_b v_i\right). \tag{C14}$$

The non-centered representation is equivalent to assuming $b_i \sim N(0,\sigma_b^2)$, but is more convenient for estimation because the Gauss–Hermite nodes associated with v_i remain fixed when σ_b is updated.

Under the non-deterministic component, the probability that all M points are allocated to the middle bin is

$$B_{it}(\theta) \equiv \left[p_{j*,it}(\theta)\right]^M.$$

Conditional on v_i, the per-period observed-data likelihood contribution is

$$L_{it}^f(v_i;\Theta) = \begin{cases} \pi_{it}(v_i;\gamma,\sigma_b) + \left[1 - \pi_{it}(v_i;\gamma,\sigma_b)\right] B_{it}(\theta), & S_{it} = 1, \\ \left[1 - \pi_{it}(v_i;\gamma,\sigma_b)\right] \dfrac{M!}{\prod_{j=1}^J d_{j,it}!} & \\ \qquad \times \displaystyle\prod_{j=1}^J \left[p_{j,it}(\theta)\right]^{d_{j,it}}, & S_{it} = 0. \end{cases} \tag{C15}$$

where

$$\Theta = (\theta,\gamma,\sigma_b).$$

The individual likelihood contribution is

$$L_i^f(\Theta) = \int \prod_{t=1}^{T_i} L_{it}^f(v;\Theta)\phi(v)\,dv, \tag{C16}$$

where $\phi(\cdot)$ denotes the standard Normal density.

Stochastic pseudo-EM algorithm. As in Subsection C.2, direct maximization of the observed-data likelihood is computationally demanding. We therefore use a stochastic pseudo-EM algorithm that augments the observed allocations with latent income draws and the latent reporting regimes.

Let $\{v_r, \varpi_r\}_{r=1}^{R}$ denote fixed Gauss–Hermite nodes and weights, normalized to integrate with respect to the standard Normal distribution. Starting from an initial parameter vector

$$\Theta^{(0)} = \left(\theta^{(0)}, \gamma^{(0)}, \sigma_b^{(0)}\right),$$

iterate over $s = 0, 1, 2, \ldots$.

E-step

Posterior quadrature-node probabilities. For each household i and node r, calculate

$$\omega_{ir}^{(s)} = \frac{\varpi_r \prod_{t=1}^{T_i} L_{it}^{f}\left(v_r; \Theta^{(s)}\right)}{\sum_{q=1}^{R} \varpi_q \prod_{t=1}^{T_i} L_{it}^{f}\left(v_q; \Theta^{(s)}\right)}, \qquad \sum_{r=1}^{R} \omega_{ir}^{(s)} = 1. \tag{C17}$$

The reporting probability inside $L_{it}^{f}(v_r; \Theta^{(s)})$ is evaluated as

$$\pi_{it}^{(s)}(v_r) = \Lambda\left(\gamma_0^{(s)} + Z_{it}'\gamma_1^{(s)} + \sigma_b^{(s)} v_r\right).$$

Posterior focal-regime probabilities. For an observation with $S_{it} = 1$, the response may have arisen either from the deterministic focal regime or from the non-focal multinomial component. Conditional on quadrature node v_r, define

$$\chi_{itr}^{(s)} \equiv \Pr\left(R_{it} = 1 \mid S_{it} = 1, v_r, \Theta^{(s)}\right) = \frac{\pi_{it}^{(s)}(v_r)}{\pi_{it}^{(s)}(v_r) + \left[1 - \pi_{it}^{(s)}(v_r)\right] B_{it}\left(\theta^{(s)}\right)}. \tag{C18}$$

For observations with $S_{it} = 0$, the deterministic regime is impossible, and we set

$$\chi_{itr}^{(s)} = 0.$$

Integrating over the posterior distribution of the reporting random effect, the posterior probability that observation (i, t) belongs to the deterministic focal regime is

$$\bar{r}_{it}^{(s)} = \sum_{r=1}^{R} \omega_{ir}^{(s)} \chi_{itr}^{(s)}. \tag{C19}$$

The posterior probability that the response was generated by the non-focal income distribution is therefore

$$w_{it}^{(s)} = 1 - \bar{r}_{it}^{(s)} = \sum_{r=1}^{R} \omega_{ir}^{(s)} \left[1 - \chi_{itr}^{(s)}\right]. \tag{C20}$$

Notice that $w_{it}^{(s)} = 1$ whenever $S_{it} = 0$.

Simulation of latent incomes. Conditional on the current income-process parameter vector $\theta^{(s)}$, simulate the latent income draws as in Subsection C.2. For each household-period pair (i, t) and each bin j, calculate

$$a_{j,it}^{(s)} = F_{it}\left(\kappa_{j-1,it}; \theta^{(s)}\right), \qquad b_{j,it}^{(s)} = F_{it}\left(\kappa_{j,it}; \theta^{(s)}\right).$$

For each of the $d_{j,it}$ points allocated to bin j, draw

$$u_{itjm}^{(s)} \sim U\left[a_{j,it}^{(s)}, b_{j,it}^{(s)}\right], \qquad m = 1, \dots, d_{j,it},$$

and set

$$y_{itjm}^{(s)} = Q_y\left(u_{itjm}^{(s)} \mid y_{it}, X_{it}; \theta^{(s)}\right).$$

For observations with $S_{it} = 1$, all M simulated values lie in the middle bin, but their contribution to the income-process update is weighted by $w_{it}^{(s)}$. The portion attributed to the deterministic focal regime does not contribute to estimation of the income process.

M-step

Update of the reporting equation. Holding the posterior quantities $\omega_{ir}^{(s)}$ and $\chi_{itr}^{(s)}$ fixed, update the parameters of the reporting equation by solving

$$\left(\widehat{\gamma}^{(s+1)}, \widehat{\sigma}_b^{(s+1)}\right) = \arg\max_{\gamma,\,\sigma_b>0} \sum_{i=1}^{N}\sum_{r=1}^{R} \omega_{ir}^{(s)} \sum_{t=1}^{T_i} \Bigg\{ \chi_{itr}^{(s)} \log \Lambda\left(\gamma_0 + Z_{it}'\gamma_1 + \sigma_b v_r\right) + \left[1 - \chi_{itr}^{(s)}\right] \log\left[1 - \Lambda\left(\gamma_0 + Z_{it}'\gamma_1 + \sigma_b v_r\right)\right] \Bigg\}. \tag{C21}$$

Unlike the previous formulation of (C21), this objective depends explicitly on σ_b. In the numerical implementation, it is convenient to optimize over

$$\alpha_b = \log \sigma_b, \qquad \sigma_b = \exp(\alpha_b),$$

so that the positivity restriction is automatically satisfied.

Update of the conditional quantile function. Let

$$z_{it} = \left(\psi_0(y_{it}), \dots, \psi_K(y_{it}), X_{it}'\right)'.$$

For each spline knot τ_ℓ, update the corresponding coefficient vector $\delta(\tau_\ell)$ using weighted quantile regression:

$$\widehat{\delta}^{(s+1)}(\tau_\ell) = \arg\min_{\delta} \sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{j=1}^{J}\sum_{m=1}^{d_{j,it}} w_{it}^{(s)} \rho_{\tau_\ell}\left(y_{itjm}^{(s)} - z_{it}'\delta\right), \qquad \ell = 1, \dots, L. \tag{C22}$$

Here,

$$\delta(\tau_\ell) = (a_0(\tau_\ell), \ldots, a_K(\tau_\ell), \beta(\tau_\ell)')' .$$

After estimating the quantile regressions, apply the ABB rearrangement step to enforce monotonicity of

$$\tau \mapsto Q_y(\tau \mid y_{it}, X_{it}).$$

Update of the exponential-tail parameters. Let

$$A_{it}^{(s+1)}(1) = Q_y\left(\tau_1 \mid y_{it}, X_{it}; \theta^{(s+1)}\right)$$

and

$$A_{it}^{(s+1)}(L) = Q_y\left(\tau_L \mid y_{it}, X_{it}; \theta^{(s+1)}\right)$$

denote the updated boundary quantiles.

Index the simulated draws by

$$q = (i, t, j, m),$$

and write

$$y_q = y_{itjm}^{(s)}, \qquad w_q^{(s)} = w_{it}^{(s)},$$
$$A_q^{(s+1)}(1) = A_{it}^{(s+1)}(1), \qquad A_q^{(s+1)}(L) = A_{it}^{(s+1)}(L).$$

The weighted maximum-likelihood update for the left-tail rate is

$$\widehat{\lambda}_1^{(s+1)} = \frac{\displaystyle\sum_q w_q^{(s)} \mathbf{1}\left\{y_q < A_q^{(s+1)}(1)\right\}}{\displaystyle\sum_q w_q^{(s)} \left[A_q^{(s+1)}(1) - y_q\right] \mathbf{1}\left\{y_q < A_q^{(s+1)}(1)\right\}}. \tag{C23}$$

Similarly, the weighted maximum-likelihood update for the right-tail rate is

$$\widehat{\lambda}_L^{(s+1)} = \frac{\displaystyle\sum_q w_q^{(s)} \mathbf{1}\left\{y_q > A_q^{(s+1)}(L)\right\}}{\displaystyle\sum_q w_q^{(s)} \left[y_q - A_q^{(s+1)}(L)\right] \mathbf{1}\left\{y_q > A_q^{(s+1)}(L)\right\}}. \tag{C24}$$

Thus, both the number of effective tail observations and their total exceedances are weighted by the posterior non-focal probabilities.

If there are no simulated observations with positive weight in one of the tails, the corresponding tail parameter is left at its previous value for that iteration.

Convergence. At the end of each iteration, evaluate the quadrature approximation to the observed-data log-likelihood:

$$\ell^f(\Theta) = \sum_{i=1}^{N} \log \left[\sum_{r=1}^{R} \varpi_r \prod_{t=1}^{T_i} L_{it}^f(v_r;\Theta) \right]. \tag{C25}$$

The algorithm is stopped when

$$\frac{\left|\ell^f\left(\Theta^{(s+1)}\right) - \ell^f\left(\Theta^{(s)}\right)\right|}{1 + |\ell^f\left(\Theta^{(s)}\right)|} < \varepsilon$$

and the change in the parameter vector is also below a pre-specified tolerance.

Because the latent incomes are simulated and the conditional quantile function is updated through weighted quantile regressions followed by rearrangement, this procedure is a stochastic pseudo-EM rather than an exact EM algorithm. The observed-data log-likelihood therefore need not increase at every single iteration. In practice, the simulation noise can be reduced by using common random numbers across nearby iterations or by averaging the parameter estimates and likelihood over several final iterations.

Interpretation. The hurdle component does not alter the household's underlying nonlinear perceived income process. It changes the estimation by separating observations that are likely to reflect deterministic focal-point reporting from observations generated by the subjective income distribution. Non-focal responses receive full weight, whereas middle-bin bunching observations contribute to the income-process estimation in proportion to their posterior probability of having arisen from the non-focal component.

D Heterogeneous AR(1) process

In this appendix, we consider the following income process:

$$y_{it+1} = \rho y_{it} + \mathbf{X}_{it}'\beta + \sigma_i u_{it+1}, \qquad u_{it+1} \sim N(0,1). \tag{D14}$$

where σ_i is household-specific. Estimating individual-specific parameters is not feasible, as we have a short panel (see Bonhomme and Denis (2025) for a discussion of issues related to the estimation of σ_i via fixed effects methods, and Lee (2025) for a discussion of identification and estimation of random coefficient models of this type via fixed effects.) However, we can estimate features of the distribution of σ_i, such as its mean and standard deviation. To do so, we take a random effects approach, and assume parametric distributions for σ_i. We discuss the estimation for both the realized and subjective income processes.

D.1 Realized income process

Define $e_{it}(\rho, \beta) = y_{it+1} - \rho y_{it} - \mathbf{X}'_{it}\beta$ for a given household. Note that for a given individual i at a given time t,

$$e_{it}|\sigma_i \sim N(0, \sigma_i^2), \text{i.i.d. over } t.$$

The conditional density for an individual i at period t is:

$$f(e_{it}|\sigma_i; \beta, \rho) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{e_{it}^2}{2\sigma_i^2}\right).$$

The marginal likelihood for an individual i is

$$L_i(\theta) = \int_0^\infty \prod_{t=1}^{T} f(e_{it}|\sigma_i; \beta, \rho) g(\sigma_i; \delta) d\sigma_i. \tag{D15}$$

The sample log-likelihood is

$$\ell(\theta) = \sum_{i=1}^{N} \log L_i(\theta). \tag{D16}$$

To estimate this model, we appeal to maximum likelihood methods. We estimate this model assuming that the distribution of σ_i is lognormal[31] and integrate the log-likelihood via Gauss-Hermite quadrature methods.

D.2 Subjective income process

We discuss the estimation for the baseline model, as the estimation of the model with the hurdle behavior is an extension of the model in the main text. Under the assumption that households answer the income expectations question with income process (D14) in mind, the true probabilities $p^*_{j,it}(\sigma_i)$ are calculated by the households as:

$$p^*_{j,it}(\sigma_i) = \Phi\left(\frac{\kappa_{j,it} - \rho y_{it} - \mathbf{X}'_{it}\beta}{\sigma_i}\right) - \Phi\left(\frac{\kappa_{j-1,it} - \rho y_{it} - \mathbf{X}'_{it}\beta}{\sigma_i}\right). \tag{D17}$$

The integrated likelihood contribution for household i is:

$$L_i(\theta) = \int \prod_{t=1}^{T} \left[\frac{M!}{\prod_{j=1}^{J} d_{jit}!} \prod_{j=1}^{J} \left(p^*_{j,it}(\sigma_i)\right)^{d_{jit}}\right] f_\sigma(\sigma_i; \delta)\, d\sigma_i. \tag{D18}$$

In practice, we assume that σ_i is distributed lognormally, and numerically integrate the likelihood via Gauss-Hermite quadrature methods.

[31] We also estimate the model under the assumption that the distribution is inverse-Gamma, which is convenient because we can derive the analytical formulas for the maximum likelihood estimators of the parameters.

D.3 Results

We first discuss the results of the realized and the subjective income processes without taking bunching behavior into account. Table D.1 reports persistence and the distribution of household-specific income dispersion, where dispersion is measured by the innovation standard deviation σ_i. As in the state-dependent AR(1) model, we find that subjective expectations imply substantially higher persistence and lower dispersion than the income process estimated from realized data. In particular, the persistence parameter is 0.7558 for the realized income process, and 1.0000 for the subjective income process. Meanwhile, on average, income dispersion is 0.4189 in the realized income process, while it is 0.0426 in the subjective income process. The top panels of Figure D.1 also demonstrate the differences in dispersion between the implied distribution of σ_i according to realized income data and subjective income data. As the results indicate, the distribution of σ_i under the realized income process is more dispersed than under the subjective income process.

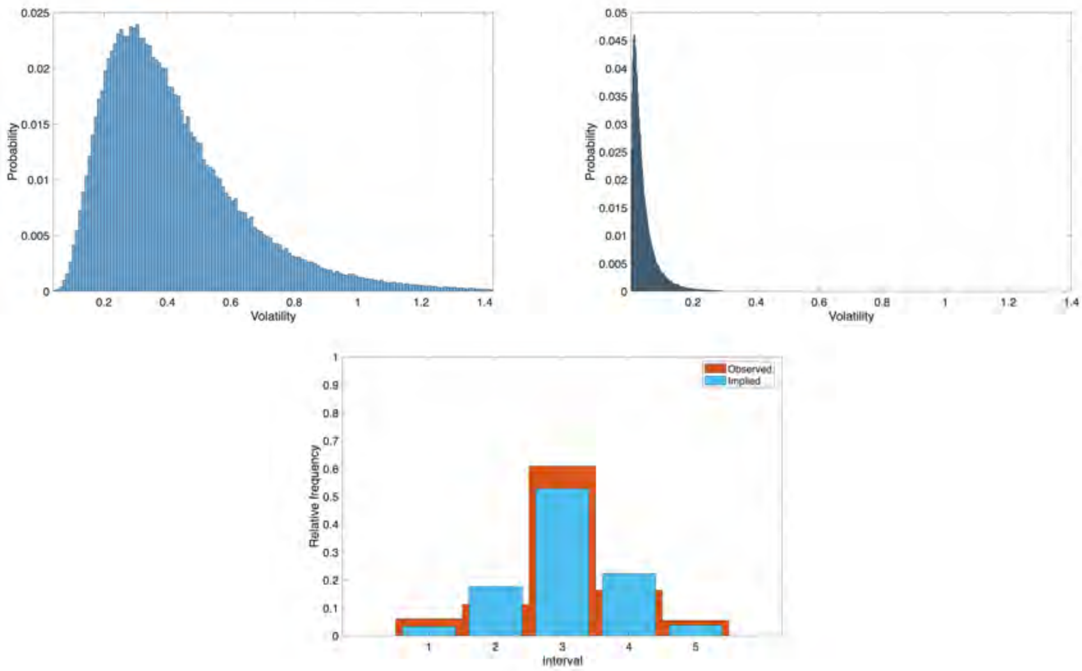
The bottom panel of Figure D.1 shows the model fit of the heterogeneous AR(1) process without a hurdle. We find that the model fit is good, with an RMSE of 0.0522, which is a remarkable improvement over the model with the state-dependent process. We do observe that the model underestimates the bunching in the middle, while it spreads out more points in the left and right bins.

Table D.1: Persistence and dispersion in the heterogeneous AR(1) process without hurdle

	Mean (ρ, μ_σ)	Standard deviation $(--, \varsigma_\sigma)$
Panel A: Realized income process		
Persistence	0.7558	–
	[0.7283,0.7797]	–
Dispersion	0.4189	0.2294
	[0.4133,0.4393]	[0.2200,0.2599]
Panel B: Subjective income process		
Persistence	1.0000	–
	(*)	–
Dispersion	0.0426	0.0471
	[0.0378,0.0527]	[0.0355,0.0475]

Notes: This table reports persistence and the mean and cross-household standard deviation of income dispersion, where household-specific dispersion is the innovation standard deviation σ_i. The specifications include wave dummies and controls for education and employment characteristics. Brackets report 95% block-bootstrap confidence intervals. (*) The subjective persistence estimate equals one in all bootstrap replications, so no bootstrap confidence interval is reported.

Figure D.1: Distributions of dispersion, and model fit.



Notes: These plots show the distribution of household-specific innovation standard deviations σ_i (top row) for the heterogeneous AR(1) process without modeling bunching behavior. The left column corresponds to the realized income process, while the right column corresponds to the subjective income process. The bottom row compares model fit by plotting the implied probabilities of the subjective income process to the observed probabilities.

We then look at how the results change when the subjective income process is estimated with the hurdle model. Table D.2 presents the results associated with the subjective income process with the hurdle model. The results that we obtain indicate that, while persistence decreases from 1.00 to 0.98, and the estimates of the distributional characteristics of dispersion increase, the estimates are still far from those implied by realized income data. Figure D.2 plots the implied and observed bin probabilities of the income process with heterogeneous σ_i. We get a better model fit, as the RMSE is 0.0432. We also observe, though, that while we are able to match the middle bin well, it comes at the expense of the left and right bins.

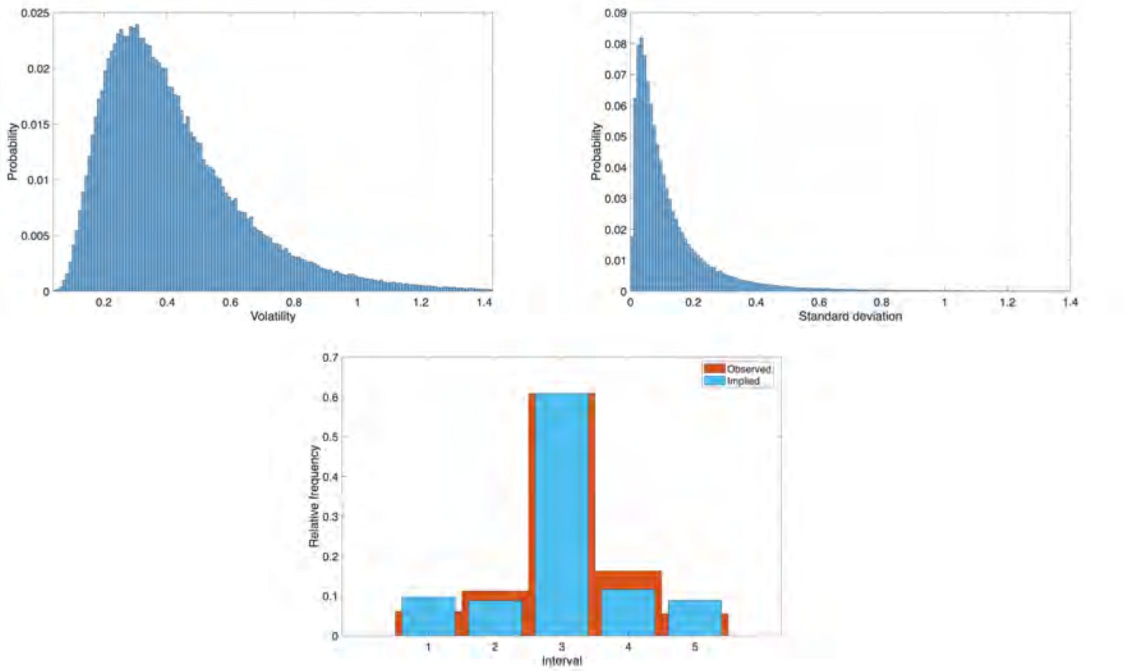
In sum, these results confirm the insight of the state-dependent AR(1) model: subjective expectations imply higher persistence and lower dispersion than the income process estimated from realized data.

Table D.2: Persistence and dispersion in the heterogeneous AR(1) process with hurdle

	Mean (ρ, μ_σ)	Standard deviation $(--, \varsigma_\sigma)$
Panel A: Realized income process		
Persistence	0.7558	–
	[0.7283,0.7797]	–
Dispersion	0.4189	0.2294
	[0.4133,0.4393]	[0.2200,0.2599]
Panel B: Subjective income process		
Persistence	0.9862	–
	[0.9499,0.9914]	–
Dispersion	0.1176	0.1073
	[0.0840,1.0426]	[0.0824,3.5679]

Notes: This table reports persistence and the mean and cross-household standard deviation of income dispersion, where household-specific dispersion is the innovation standard deviation σ_i. The specifications include wave dummies and controls for education and employment characteristics. Brackets report 95% block-bootstrap confidence intervals.

Figure D.2: Distributions of dispersion, and model fit.



Notes: These plots show the distribution of household-specific innovation standard deviations σ_i (top row) for the heterogeneous AR(1) process accounting for bunching behavior. The left column corresponds to the realized income process, while the right column corresponds to the subjective income process. The bottom row compares model fit by plotting the implied probabilities of the subjective income process to the observed probabilities.

E Identification

In this appendix, we discuss identification of the conditional distribution of future log income perceived by household i at time t, given the state

$$W_{it} = (y_{it}, X_{it}),$$

which we denote by

$$F_y(c \mid W_{it}) = \Pr(y_{i,t+1} \leqslant c \mid W_{it}).$$

The observed data are

$$\{d_{it}, W_{it}\}_{t=1}^{T_i}, \qquad d_{it} = (d_{1,it}, \ldots, d_{J,it})', \qquad \sum_{j=1}^{J} d_{j,it} = M.$$

Let the income-growth cutoffs satisfy

$$-\infty = h_0 < h_1 < \cdots < h_{J-1} < h_J = +\infty.$$

The corresponding income-level cutoffs and bins are

$$\kappa_{j,it} = y_{it} + h_j, \qquad I_{j,it} = (\kappa_{j-1,it}, \kappa_{j,it}] = (y_{it} + h_{j-1}, y_{it} + h_j].$$

The allocation vector d_{it} provides interval-censored information about the perceived distribution of $y_{i,t+1}$. It reveals the probability assigned to each interval, but does not reveal how probability mass is distributed within an interval.

Assumption E.1 (Correct reporting of subjective bin probabilities)**.** For every $j = 1, \ldots, J$,

$$E\left[\frac{d_{j,it}}{M} \,\middle|\, W_{it} = (y, X)\right] = F_y(y + h_j \mid y, X) - F_y(y + h_{j-1} \mid y, X).$$

Assumption E.1 can be interpreted either as exact reporting of subjective probabilities or as unbiased reporting at the population level. The hurdle model considered below relaxes this assumption by allowing respondents to enter a deterministic focal-point regime.

E.1 Nonparametric content of the data

Define the cumulative reported probability

$$\Pi_j(y, X) \equiv E\left[\sum_{k=1}^{j} \frac{d_{k,it}}{M} \,\middle|\, W_{it} = (y, X)\right], \qquad j = 1, \ldots, J-1,$$

and set $\Pi_0(y, X) = 0$ and $\Pi_J(y, X) = 1$.

Proposition E.1 (Sharp partial identification)**.** *Under Assumption E.1, the conditional CDF is identified at each finite bin cutoff:*

$$F_y(y + h_j \mid y, X) = \Pi_j(y, X), \qquad j = 1, \ldots, J - 1.$$

Without additional restrictions on the within-bin shape of the distribution, the conditional quantile is set identified. In particular, if

$$\Pi_{j-1}(y, X) < \tau \leqslant \Pi_j(y, X),$$

then

$$Q_y(\tau \mid y, X) \in (y + h_{j-1}, y + h_j].$$

This identified set is sharp.

Proof. By Assumption E.1,

$$E\left[\frac{d_{k,it}}{M} \,\middle|\, W_{it} = (y, X)\right] = F_y(y + h_k \mid y, X) - F_y(y + h_{k-1} \mid y, X).$$

Summing from $k = 1$ to $k = j$ gives

$$\begin{aligned}
\Pi_j(y, X) &= \sum_{k=1}^{j} E\left[\frac{d_{k,it}}{M} \,\middle|\, W_{it} = (y, X)\right] \\
&= \sum_{k=1}^{j} \{F_y(y + h_k \mid y, X) - F_y(y + h_{k-1} \mid y, X)\} \\
&= F_y(y + h_j \mid y, X) - F_y(y + h_0 \mid y, X).
\end{aligned}$$

Because $h_0 = -\infty$,

$$F_y(y + h_0 \mid y, X) = 0,$$

and hence

$$F_y(y + h_j \mid y, X) = \Pi_j(y, X).$$

Now fix $\tau \in (0, 1)$ and suppose

$$\Pi_{j-1}(y, X) < \tau \leqslant \Pi_j(y, X).$$

Since the CDF is nondecreasing,

$$F_y(y + h_{j-1} \mid y, X) < \tau \leqslant F_y(y + h_j \mid y, X).$$

Therefore, under the generalized-inverse definition

$$Q_y(\tau \mid y, X) = \inf\{c : F_y(c \mid y, X) \geqslant \tau\},$$

we have

$$Q_y(\tau \mid y, X) \in (y + h_{j-1}, y + h_j].$$

To establish sharpness, fix any

$$q \in (y + h_{j-1}, y + h_j].$$

Construct an alternative CDF $\widetilde{F}_y(\cdot \mid y, X)$ that agrees with $F_y(\cdot \mid y, X)$ at every cutoff $y + h_\ell$, but reallocates the probability mass inside bin j so that

$$\widetilde{F}_y(q- \mid y, X) < \tau \leqslant \widetilde{F}_y(q \mid y, X).$$

This reallocation leaves all bin probabilities unchanged and is therefore observationally equivalent to the original distribution. Its τ-quantile is q. Since every point in the interval can be attained in this way, the bound is sharp. □

Remark E.1 (Finite lower cutoffs). If the first cutoff h_0 is finite rather than $-\infty$, the correct identity is

$$F_y(y + h_j \mid y, X) = F_y(y + h_0 \mid y, X) + \Pi_j(y, X).$$

Thus, the CDF at the cutoffs is identified only if the probability below the first cutoff is known or separately modeled.

Remark E.2 (Additional panel waves). Additional waves identify the CDF at the cutoffs associated with the states observed in those waves. In the absence of restrictions linking the within-bin shape across states, however, additional waves do not tighten the within-bin quantile bounds at a fixed state (y, X).

E.2 State-dependent AR(1) model

We next impose the state-dependent Gaussian income process

$$y_{i,t+1} = \rho(X_{it}) y_{it} + X_{it}' \beta + \eta_i + \sigma(X_{it}) u_{i,t+1}, \qquad u_{i,t+1} \sim N(0, 1),$$

where

$$\rho(X) = \Lambda(\rho_0 + X' \rho_1), \qquad \sigma(X) = \exp(\sigma_0 + X' \sigma_1).$$

Assumption E.2 (Normal location–scale structure and random-effects exogeneity)**.** The household effect satisfies

$$\eta_i \sim N(0, \sigma_\eta^2),$$

and is independent of the full observed state history:

$$\eta_i \perp (W_{i1}, \ldots, W_{iT_i}).$$

Conditional on η_i and the observed state history, the income innovations are independent over time.

Assumption E.3 (Support and rank)**.** For every relevant value of the discrete components of X, the conditional support of y contains a nondegenerate interval. Moreover, the relevant covariate design matrices have full column rank.

After integrating over η_i,

$$y_{i,t+1} \mid W_{it} \sim N\left(\mu(W_{it}), \sigma_e^2(X_{it})\right),$$

where

$$\mu(y, X) = \rho(X)y + X'\beta$$

and

$$\sigma_e^2(X) = \sigma^2(X) + \sigma_\eta^2.$$

E.2.1 Identification of the marginal distribution

Corollary E.2 (Identification of the marginal perceived distribution)**.** *Suppose Assumptions E.1, E.2, and E.3 hold.*

Fix a state (y, X). If there are two distinct finite cutoffs $h_j \neq h_k$ such that

$$0 < \Pi_j(y, X) < 1, \qquad 0 < \Pi_k(y, X) < 1,$$

then $\mu(y, X)$ and $\sigma_e(X)$ are identified. Consequently, the full marginal CDF

$$F_y(\cdot \mid y, X)$$

is identified within the Normal location–scale model.

If this condition holds over a sufficiently rich set of states, then (ρ_0, ρ_1, β) are identified.

Proof. By Proposition E.1,

$$\Pi_j(y, X) = F_y(y + h_j \mid y, X).$$

Under the marginal Normal model,

$$\Pi_j(y, X) = \Phi\left(\frac{y + h_j - \mu(y, X)}{\sigma_e(X)}\right).$$

For an interior cutoff probability, define

$$z_j(y, X) = \Phi^{-1}\left(\Pi_j(y, X)\right).$$

Then

$$z_j(y, X) = \frac{y + h_j - \mu(y, X)}{\sigma_e(X)}.$$

For two distinct cutoffs,

$$z_k(y, X) - z_j(y, X) = \frac{h_k - h_j}{\sigma_e(X)}.$$

Because $h_k \neq h_j$ and $\sigma_e(X) > 0$,

$$\sigma_e(X) = \frac{h_k - h_j}{z_k(y, X) - z_j(y, X)}$$

is uniquely identified. The conditional mean is then

$$\mu(y, X) = y + h_j - \sigma_e(X) z_j(y, X).$$

Hence

$$F_y(c \mid y, X) = \Phi\left(\frac{c - \mu(y, X)}{\sigma_e(X)}\right)$$

is identified for every $c \in \mathbb{R}$.

It remains to identify the coefficients of the conditional mean. For a fixed X and two distinct income values $y_1 \neq y_2$ in the conditional support,

$$\rho(X) = \frac{\mu(y_2, X) - \mu(y_1, X)}{y_2 - y_1}.$$

Thus, $\rho(X)$ is identified over the support of X. Since Λ is known and strictly increasing,

$$\Lambda^{-1}(\rho(X)) = \rho_0 + X'\rho_1.$$

Full-rank variation in $(1, X')'$ identifies (ρ_0, ρ_1). Finally,

$$\mu(y, X) - \rho(X)y = X'\beta,$$

so full-rank variation in X identifies β. □

Remark E.3 (Interiority). The inverse Normal CDF is finite only on $(0, 1)$. Thus, two strictly interior cumulative probabilities are sufficient for pointwise identification of both location and scale. If both objects are unrestricted at a given state, a single interior cutoff is not sufficient.

Remark E.4 (Interpretation under misspecification). If the true perceived distribution is not Gaussian, the Normal specification selects the location and scale parameters that best reconcile the model with the observed cutoff probabilities. The implied within-bin distribution is then supplied by Normality rather than identified nonparametrically from the survey responses.

E.2.2 Separating within-household innovation variance and fixed heterogeneity

The cross-sectional distribution identifies

$$\sigma_e^2(X) = \sigma^2(X) + \sigma_\eta^2,$$

but does not separately identify $\sigma^2(X)$ and σ_η^2. Panel dependence provides the additional information required for this decomposition.

Define the cumulative allocation

$$C_{j,it} = \sum_{\ell=1}^{j} \frac{d_{\ell,it}}{M}.$$

Assumption E.4 (Conditional independence across waves)**.** For $t \neq s$, the probability-point draws underlying d_{it} and d_{is} are independent conditional on

$$\eta_i \quad \text{and} \quad (W_{i1}, \ldots, W_{iT_i}).$$

Assumption E.5 (Panel interiority)**.** For at least one pair $t \neq s$ and finite cutoffs h_j, h_k,

$$0 < \Pi_j(W_{it}) < 1, \qquad 0 < \Pi_k(W_{is}) < 1.$$

The within-type variances satisfy $\sigma^2(X) > 0$ over the relevant support.

Corollary E.3 (Identification of the variance components)**.** *Suppose Assumptions E.1–E.5 hold and $T_i \geqslant 2$ for a positive fraction of households. Then the joint panel distribution identifies*

$$\sigma_\eta^2, \quad \sigma^2(X), \quad (\sigma_0, \sigma_1).$$

Together with Corollary E.2, this identifies

$$(\rho_0, \rho_1, \beta, \sigma_0, \sigma_1, \sigma_\eta^2).$$

Proof. Fix $t \neq s$ and cumulative cutoffs j and k. Conditional on η_i,

$$E[C_{j,it} \mid W_{it}, \eta_i] = \Phi\left(\frac{y_{it} + h_j - \mu(W_{it}) - \eta_i}{\sigma(X_{it})}\right).$$

By Assumption E.4,

$$\begin{aligned} &\operatorname{Cov}\left(C_{j,it}, C_{k,is} \mid W_{it}, W_{is}\right) \\ &\qquad = \operatorname{Cov}_\eta\left(E[C_{j,it} \mid W_{it}, \eta_i], E[C_{k,is} \mid W_{is}, \eta_i] \mid W_{it}, W_{is}\right). \end{aligned}$$

Equivalently, consider two latent draws

$$Y_t^* = \mu(W_{it}) + \eta_i + \sigma(X_{it}) u_t,$$

and

$$Y_s^* = \mu(W_{is}) + \eta_i + \sigma(X_{is})u_s,$$

where u_t and u_s are independent standard Normal variables. Conditional on (W_{it}, W_{is}), the pair (Y_t^*, Y_s^*) is bivariate Normal, with marginal variances

$$\sigma_e^2(X_{it}) \quad \text{and} \quad \sigma_e^2(X_{is}),$$

and correlation

$$r_{ts} = \frac{\sigma_\eta^2}{\sigma_e(X_{it})\sigma_e(X_{is})}.$$

Let

$$z_{j,it} = \Phi^{-1}(\Pi_j(W_{it})), \qquad z_{k,is} = \Phi^{-1}(\Pi_k(W_{is})).$$

It follows that

$$\begin{aligned} &\operatorname{Cov}\left(C_{j,it}, C_{k,is} \mid W_{it}, W_{is}\right) \\ &\quad = \Phi_2(z_{j,it}, z_{k,is}; r_{ts}) - \Phi(z_{j,it})\Phi(z_{k,is}), \end{aligned}$$

where $\Phi_2(a, b; r)$ is the standard bivariate Normal CDF with correlation r.

The left-hand side is identified from the joint panel distribution. The marginal quantities $z_{j,it}$, $z_{k,is}$, $\sigma_e(X_{it})$, and $\sigma_e(X_{is})$ are identified from Corollary E.2.

For finite a and b, the bivariate Normal CDF is strictly increasing in its correlation parameter:

$$\frac{\partial}{\partial r}\Phi_2(a, b; r) = \phi_2(a, b; r) > 0.$$

Therefore, the observed cross-wave covariance uniquely identifies r_{ts}. It then identifies

$$\sigma_\eta^2 = r_{ts}\, \sigma_e(X_{it})\sigma_e(X_{is}).$$

Having identified σ_η^2, the within-household innovation variance is

$$\sigma^2(X) = \sigma_e^2(X) - \sigma_\eta^2.$$

Finally,

$$\log \sigma^2(X) = 2\sigma_0 + 2X'\sigma_1.$$

Full-rank variation in $(1, X')'$ identifies (σ_0, σ_1). □

Remark E.5 (Role of the panel). The panel identifies σ_η^2 through persistent cross-wave dependence in allocations made by the same household. It does not identify the shape of the distribution within the survey bins; that shape continues to be supplied by the Normal location–scale assumption.

Remark E.6 (Overidentification). One pair of waves and one pair of interior cumulative cutoffs is sufficient in principle. Additional waves and cutoffs provide overidentifying covariance restrictions that can be used to assess the random-effects and conditional-independence assumptions.

E.3 Nonlinear income process

Let the conditional quantile function belong to a finite-dimensional sieve class

$$\mathcal{Q}_{L,K} = \{Q_y(\tau \mid y, X; \theta) : \theta \in \Theta_{L,K}\}.$$

Assumption E.6 (Support)**.** The support $\mathcal{W}$ of (y, X) contains sufficient variation to identify the basis coefficients used in the quantile specification.

Assumption E.7 (Correctly specified finite sieve)**.** For fixed sieve dimensions (L, K), there exists $\theta_0 \in \Theta_{L,K}$ such that

$$Q_y^0(\tau \mid y, X) = Q_y(\tau \mid y, X; \theta_0).$$

For every (y, X), the function

$$\tau \mapsto Q_y(\tau \mid y, X; \theta)$$

is continuous and strictly increasing.

For $j = 1, \ldots, J - 1$, define

$$m_j(y, X; \theta) = F_y(y + h_j \mid y, X; \theta).$$

Assumption E.8 (Cutoff injectivity)**.** For fixed (L, K), the map

$$\theta \mapsto \{m_j(y, X; \theta) : j = 1, \ldots, J - 1, \ (y, X) \in \mathcal{W}\}$$

is injective over $\Theta_{L,K}$. That is, if

$$m_j(y, X; \theta) = m_j(y, X; \theta')$$

for every j and almost every $(y, X) \in \mathcal{W}$, then

$$\theta = \theta'.$$

Corollary E.4 (Identification of the finite-dimensional nonlinear model)**.** *Suppose Assumptions E.1, E.6, E.7, and E.8 hold. For fixed sieve dimensions (L, K), θ_0 is point identified from the cross-sectional distribution of (d_{it}, W_{it}). Consequently, the conditional quantile function and the associated conditional CDF are identified within the maintained sieve model.*

Proof. By Proposition E.1, the data identify

$$\Pi_j(y, X) = F_y(y + h_j \mid y, X)$$

for every $j = 1, \ldots, J - 1$ and almost every $(y, X) \in \mathcal{W}$.

Under the quantile specification and strict monotonicity,

$$F_y(c \mid y, X; \theta) = \sup \{\tau \in (0, 1) : Q_y(\tau \mid y, X; \theta) \leqslant c\} .$$

Therefore,

$$m_j(y, X; \theta_0) = \Pi_j(y, X)$$

for every cutoff and state.

Suppose that another parameter $\theta' \in \Theta_{L,K}$ generates the same observed cutoff probabilities. Then

$$m_j(y, X; \theta') = m_j(y, X; \theta_0)$$

for all j and almost every $(y, X) \in \mathcal{W}$. By Assumption E.8,

$$\theta' = \theta_0.$$

Thus, θ_0 is point identified. Since the parameter uniquely determines the quantile function, the full conditional quantile function and its associated CDF are identified within the maintained sieve class. □

Remark E.7 (Approximation versus identification)*.* The preceding result concerns identification at a fixed sieve dimension under correct specification. If the true quantile function does not belong exactly to $\mathcal{Q}_{L,K}$, the fixed-dimensional model identifies a pseudo-true sieve approximation rather than the exact true distribution. Approximation results as $(L, K) \to \infty$ are therefore conceptually distinct from fixed-dimensional point identification.

Remark E.8 (Local identification)*.* A sufficient condition for local identification is that the Jacobian of the collection of cutoff probabilities with respect to θ has full column rank at θ_0. Global point identification requires the stronger injectivity condition in Assumption E.8.

E.4 Hurdle model

Let

$$e^* = (0, \ldots, 0, M, 0, \ldots, 0)'$$

denote the allocation that places all M points in the middle bin. The observed bunching indicator is

$$S_{it} = 1\{d_{it} = e^*\}.$$

With probability

$$\pi(Z_{it}, b_i; \gamma) = \Lambda(\gamma_0 + Z_{it}'\gamma_1 + b_i),$$

the household enters the deterministic focal regime and reports e^*. Otherwise, it reports according to the underlying income-process model.

Let

$$f_\theta(d \mid W)$$

denote the probability of allocation d under the non-focal income-process component, after integrating over any income-process random effect. Define

$$A_\theta(W) = f_\theta(e^* \mid W).$$

Also define the marginal focal probability

$$q_{\gamma,\sigma_b}(Z) = \int \pi(Z, b; \gamma)\, dG_{\sigma_b}(b), \qquad b \sim N(0, \sigma_b^2).$$

Assumption E.9 (Exclusion and random-effects exogeneity)**.** The reporting shifter V_{it} is contained in Z_{it} but not in W_{it}. Conditional on W_{it} and year effects, V_{it} does not affect the perceived income distribution.

Moreover,

$$(b_i, \eta_i) \perp \{Z_{it}, W_{it}\}_{t=1}^{T_i}, \qquad b_i \perp \eta_i.$$

Assumption E.10 (Positivity and nondegeneracy)**.** Over the relevant support,

$$0 < q_{\gamma,\sigma_b}(Z) < 1$$

and

$$A_\theta(W) < 1.$$

Assumption E.11 (Identification from off-focal allocations)**.** The map from θ to the conditional distribution of non-focal allocation patterns,

$$\left\{ \frac{f_\theta(d \mid W)}{1 - A_\theta(W)} : d \neq e^*,\ W \in \mathcal{W} \right\},$$

is injective. For models with an income-process random effect, this condition applies to the corresponding joint cross-wave distribution of off-focal allocations.

Assumption E.12 (Identification of the reporting equation)**.** The map

$$(\gamma, \sigma_b^2) \mapsto \{q_{\gamma,\sigma_b}(Z) : Z \in \mathcal{Z}\}$$

is injective over the maintained parameter space. The excluded variable V has nondegenerate conditional variation and shifts the focal-point probability.

Proposition E.5 (Identification of the hurdle model)**.** *Suppose Assumptions E.9–E.12 hold and the underlying income-process component satisfies the relevant identification conditions in Corollary E.3 or Corollary E.4. Then*

$$\theta, \quad \gamma, \quad \sigma_b^2$$

are jointly point identified.

Proof. Under Assumption E.9, the observed conditional probability of an allocation d is

$$g(d \mid Z, W) = \begin{cases} q_{\gamma,\sigma_b}(Z) + [1 - q_{\gamma,\sigma_b}(Z)]\, A_\theta(W), & d = e^*, \\ [1 - q_{\gamma,\sigma_b}(Z)]\, f_\theta(d \mid W), & d \neq e^*. \end{cases}$$

Consider the distribution conditional on observing a non-focal allocation, $S = 0$. For $d \neq e^*$,

$$\begin{aligned} \Pr(d \mid S = 0, Z, W) &= \frac{[1 - q_{\gamma,\sigma_b}(Z)] f_\theta(d \mid W)}{[1 - q_{\gamma,\sigma_b}(Z)][1 - A_\theta(W)]} \\ &= \frac{f_\theta(d \mid W)}{1 - A_\theta(W)}. \end{aligned}$$

Thus, conditioning on $S = 0$ eliminates the reporting probability. The resulting distribution is independent of the excluded reporting shifter V. By Assumption E.11, this distribution identifies θ.

Once θ is identified, $A_\theta(W)$ is known. Moreover,

$$\Pr(S = 0 \mid Z, W) = [1 - q_{\gamma,\sigma_b}(Z)]\,[1 - A_\theta(W)]\,.$$

Since $A_\theta(W) < 1$,

$$q_{\gamma,\sigma_b}(Z) = 1 - \frac{\Pr(S = 0 \mid Z, W)}{1 - A_\theta(W)}$$

is identified over the support of Z. Assumption E.12 then identifies (γ, σ_b^2). Hence all income-process and reporting-process parameters are jointly point identified. ∎

Remark E.9 (Role of the exclusion restriction). The exclusion restriction provides variation in the mixture probability while holding the income-process component fixed. For a continuous excluded variable,

$$\frac{\partial}{\partial V} \Pr(S = 1 \mid Z, W) = \frac{\partial q_{\gamma,\sigma_b}(Z)}{\partial V}\,[1 - A_\theta(W)]\,.$$

For a discrete excluded variable, the analogous expression is a finite contrast. Because $A_\theta(W) < 1$, a genuine shift in the focal probability changes observed middle-bin bunching. This establishes relevance of the excluded variable. Point identification additionally requires the component and reporting-equation injectivity conditions stated above.

In our empirical application, the validity of the exclusion restriction is a maintained identifying assumption rather than a directly testable condition. House price response behavior is a valid exclusion only if, conditional on the included covariates and wave effects, its remaining association with income-expectation responses operates through reporting behavior rather than through common household-level factors that also affect the perceived income distribution. In particular, heterogeneous inflation expectations or inflation uncertainty could affect responses to both nominal income-growth and house price expectations, thereby violating the exclusion restriction.

F Measurement error in reported income

A potential concern is that measurement error in reported household income may contribute to the gap between realized and subjective income dynamics. To assess the quantitative importance of this channel, we examine how the estimated realized and subjective income processes change as we impose progressively larger amounts of classical measurement error. This benchmark is informative because classical measurement error moves the realized-income estimates in the direction most favorable to this explanation: ignoring it attenuates estimated persistence and inflates innovation variance, and hence innovation dispersion. If economically important differences remain after correcting for substantial classical measurement error, standard errors-in-variables attenuation cannot by itself reconcile the two processes.

Let reported log household income y_{it} satisfy

$$y_{it} = y_{it}^* + e_{it}, \qquad e_{it} \sim N(0, \tau^2), \tag{F19}$$

where y_{it}^* denotes latent error-free income. The benchmark assumes that e_{it} is independent across time and independent of latent income and income innovations. This provides the standard classical errors-in-variables benchmark (Bound et al., 2001).

We parameterize the amount of measurement error using the conditional reliability ratio

$$\lambda = \frac{V(y_{it}^* \mid H_{it})}{V(y_{it} \mid H_{it})}, \qquad 0 < \lambda \leqslant 1, \tag{F20}$$

where H_{it} denotes the controls in an auxiliary conditional-income regression used to calibrate the distribution of latent current income. Let

$$m_{it} = \mathbb{E}[y_{it} \mid H_{it}], \qquad V_c = V(y_{it} - m_{it}), \tag{F21}$$

where V_c is the residual variance from this auxiliary regression. Under (F19) and Gaussianity,

$$\tau^2 = (1-\lambda)V_c, \qquad y^*_{it} \mid y_{it}, Z_{it}; \lambda \sim N(m_{it} + \lambda(y_{it} - m_{it}), \Omega)\,, \tag{F22}$$

where

$$\Omega = \lambda(1-\lambda)V_c. \tag{F23}$$

Thus, $\lambda = 1$ corresponds to no measurement error, whereas lower values place progressively more weight on the conditional mean m_{it} and increase posterior uncertainty about latent current income. We consider $\lambda \in \{1, .95, .90, .85, .80, .75, .70, .65\}$. A reliability ratio around 0.8 is broadly consistent with validation evidence for annual earnings levels, although such evidence does not directly identify the reliability of household income in the EFF.[32]

F.1 State-dependent AR(1) process

We implement the exercise for the state-dependent AR(1) specification underlying the headline comparison in Table 4. For realized income, the latent state-dependent process is

$$y^*_{i,t+1} = \rho_{\mathrm{R}}(X_{it})y^*_{it} + X'_{it}\beta_{\mathrm{R}} + \eta_i + \sigma_{\mathrm{R}}(X_{it}, y^*_{it})u_{i,t+1}, \qquad u_{i,t+1} \sim N(0,1). \tag{F24}$$

Because future reported income is also measured with error, conditional on y^*_{it} and η_i we have

$$y_{i,t+1} \mid y^*_{it}, X_{it}, \eta_i \sim N\big(\rho_{\mathrm{R}}(X_{it})y^*_{it} + X'_{it}\beta_{\mathrm{R}} + \eta_i, \sigma_{\mathrm{R}}(X_{it}, y^*_{it})^2 + \tau^2\big)\,. \tag{F25}$$

The sensitivity likelihood integrates this density over the conditional distribution of y^*_{it} in (F22) and over the household random effect. Because current income enters the dispersion function in the state-dependent specification, measurement error affects both the autoregressive term and $\sigma_{\mathrm{R}}(X_{it}, y^*_{it})$.

For subjective expectations, the object observed at date t is a probability allocation over fixed bins of future income growth. Conditional on latent current income and the household effect, perceived income growth has mean

$$\mu^g_{it}(y^*, \eta_i) = \big[\rho_{\mathrm{S}}(X_{it}) - 1\big]y^* + X'_{it}\beta_{\mathrm{S}} + \eta_i \tag{F26}$$

[32] See, e.g., Bound and Krueger (1991) and Pischke (1995).

and standard deviation $\sigma_S(X_{it}, y^*)$. If $h_0 = -\infty < h_1 < \cdots < h_J = +\infty$ denote the income-growth bin cutoffs, the probability assigned by the model to bin j is

$$p^S_{jit}(y^*, \eta_i) = \Phi\left(\frac{h_j - \mu^g_{it}(y^*, \eta_i)}{\sigma_S(X_{it}, y^*)}\right) - \Phi\left(\frac{h_{j-1} - \mu^g_{it}(y^*, \eta_i)}{\sigma_S(X_{it}, y^*)}\right). \tag{F27}$$

The multinomial likelihood for the reported probability points is then integrated over $y^*_{it} \mid y_{it}, X_{it}$ and the household random effect. Importantly, there is no additional $+\tau^2$ term in the conditional variance entering the subjective likelihood: the survey reports beliefs about future income rather than a noisy future income realization. Measurement error nevertheless affects the subjective estimates because reported current income is used as a state variable and current income enters the state-dependent dispersion function.

F.2 Results

Table F.1 summarizes the results of the exercise. The $\lambda = 1$ row reports the baseline estimates, while the remaining rows report the measurement-error sensitivity estimates. For dispersion, we report the posterior expected innovation standard deviation

$$\overline{\sigma}(\lambda) = \frac{1}{N_{\text{obs}}} \sum_{it} \mathbb{E}[\sigma(X_{it}, y^*_{it}) \mid y_{it}, X_{it}; \lambda], \tag{F28}$$

which averages the model's dispersion measure over the posterior distribution of the latent current-income state.

Table F.1: Measurement error sensitivity: state-dependent AR(1)

	Persistence			Dispersion		
λ	Realized	Subjective	Gap	Realized	Subjective	Gap
1.00	0.6853	0.9998	0.3146	0.4722	0.0411	0.4311
0.95	0.7139	1.0000	0.2861	0.4333	0.0410	0.3923
0.90	0.7438	1.0000	0.2562	0.3887	0.0412	0.3475
0.85	0.7766	1.0000	0.2234	0.3317	0.0404	0.2913
0.80	0.8316	1.0000	0.1684	0.2242	0.0376	0.1866
0.75	0.8616	1.0000	0.1384	0.1387	0.0367	0.1020
0.70	0.8700	1.0000	0.1300	0.1048	0.0333	0.0715
0.65	0.8750	1.0000	0.1250	0.0864	0.0341	0.0523

Notes: The $\lambda = 1$ row reports the no-measurement-error baseline estimates. For $\lambda < 1$, realized and subjective estimates are obtained from the classical measurement-error sensitivity exercise. Dispersion is measured by the posterior expected innovation standard deviation in (F28). The gap is subjective minus realized for persistence and realized minus subjective for dispersion. Gaps are computed using the unrounded estimates.

The correction moves the realized process substantially toward the subjective process, but does not eliminate the difference. At our benchmark reliability ratio of $\lambda = 0.80$, average realized persistence increases from 0.685 in the no-error specification to 0.832, whereas subjective persistence remains essentially one. The persistence gap therefore falls from about 0.315 to 0.168, a reduction of about 46 percent. Posterior expected realized dispersion falls from 0.472 to 0.224, whereas subjective dispersion is 0.038. The dispersion gap therefore falls from 0.431 to 0.187, a reduction of about 57 percent.

The same qualitative conclusion holds under more severe corrections. At $\lambda = 0.75$, realized persistence is about 0.862 and realized dispersion about 0.139, compared with subjective persistence of approximately one and subjective dispersion of about 0.037. Even at $\lambda = 0.65$, realized income remains less persistent (0.875 versus approximately one) and more dispersed (0.086 versus 0.034). The lower-reliability cases should nevertheless be viewed as stress tests: as λ falls, the cross-sectional distribution of realized dispersion becomes increasingly heterogeneous, with its lower tail approaching zero for some states.

The subjective process is much less sensitive to the correction, particularly in persistence. This is intuitive because perceived persistence is already close to one: the coefficient on latent current income in perceived income growth, $\rho_S(X) - 1$, is therefore close to zero. Subjective dispersion changes somewhat because current income enters the state-dependent dispersion function, but these movements are small relative to those in the realized process.

Overall, classical measurement error can account for a quantitatively important fraction of the perceived-realized gap, especially in dispersion, but it does not reconcile the two processes over the reliability range considered. This is informative because the classical benchmark moves realized persistence and dispersion in exactly the direction required to explain our headline result. At the same time, survey income-reporting errors need not be classical. Validation studies document mean reversion and serial correlation in reporting errors (Bound and Krueger, 1991; Pischke, 1995). Under such non-classical error, the direction of the bias in estimated persistence and dispersion is not generally signed. We therefore interpret this exercise as a sensitivity analysis rather than as a structural model of the reporting-error process.

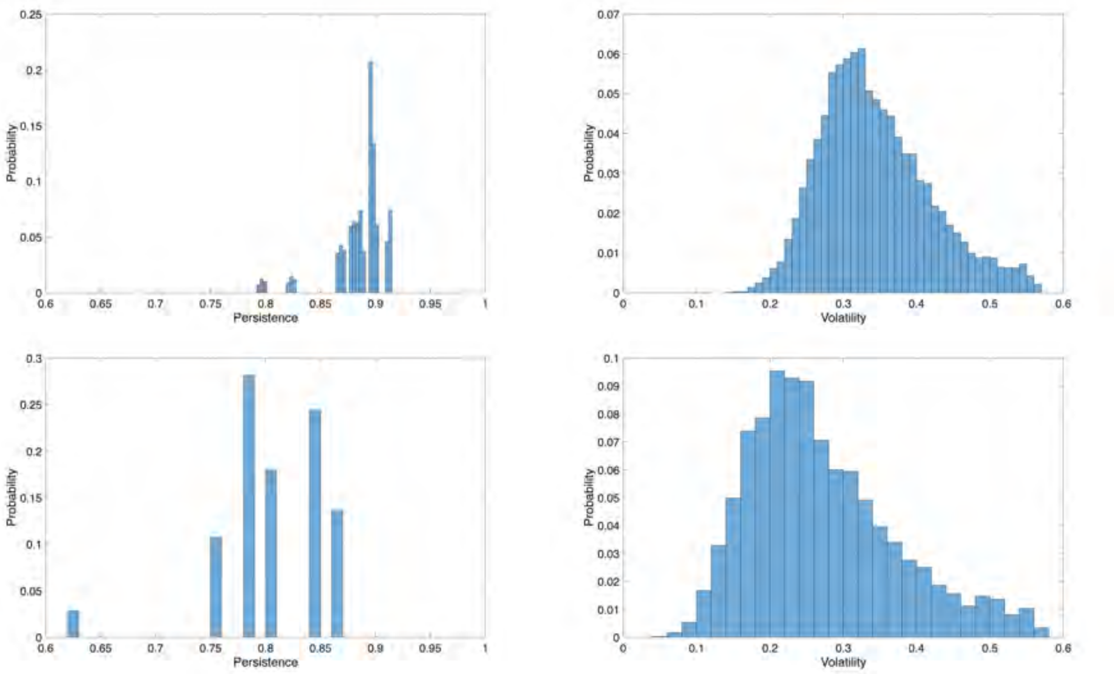
G Results from the ECV

As discussed in Section 5, one potential explanation for the differences between the realized and subjective income processes is the difference in observation horizons. The estimation using realized EFF income data is based on household income observed at three-year intervals, whereas the subjective income expectations question refers to annual changes in household income. To assess whether this difference in horizons can account for the gap between the two sets of estimates, we use data from the *Encuesta de Condiciones de Vida* (ECV), or Living Conditions Survey, conducted by the Spanish Statistical Institute (INE), and estimate the income processes using both one-year and three-year intervals.

To obtain a sample comparable to our EFF estimation sample, we restrict the ECV data to households whose heads are aged 25–65, with at least one household member active in the labor market, and whose main source of income is not pensions. This leaves us with a balanced panel of 3,863 households observed for four consecutive periods from 2021 to 2024.

Figure G.1 reports the estimates of the state-dependent AR(1) process using income observed at one-year and three-year intervals. Average persistence decreases from approximately 0.85 for the one-year income process to 0.72 for the three-year income process, while average dispersion increases slightly. These changes are small relative to the differences between the income processes estimated from realized and subjective expectations data.

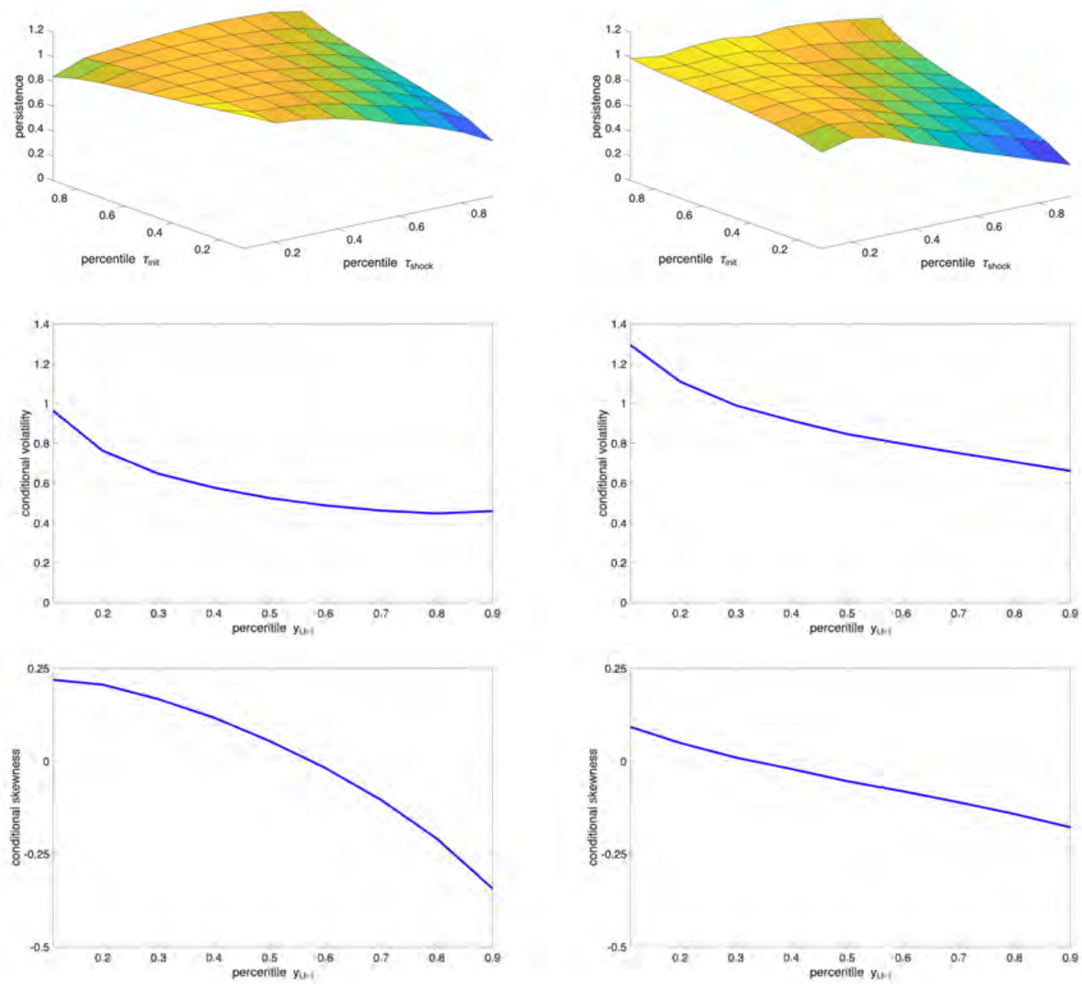
Figure G.1: Distributions of persistence and dispersion, state-dependent AR(1) process, ECV



Notes: These plots show the distribution of persistence (left panel) and dispersion, measured by the innovation standard deviation $\sigma(X_{it})$ (right panel), for the state-dependent AR(1) process using one-year (top row) and three-year (bottom row) income data.

Figure G.2 reports the corresponding estimates for the nonlinear income process. We find similar patterns for persistence and dispersion. The estimated skewness profiles also differ somewhat across observation horizons, with more pronounced estimates in magnitude in the three-year specification. Overall, however, the differences between the one-year and three-year estimates remain small relative to the gap between the realized and subjective income processes documented in the main text.

Figure G.2: Nonlinear persistence, dispersion, and skewness, ECV

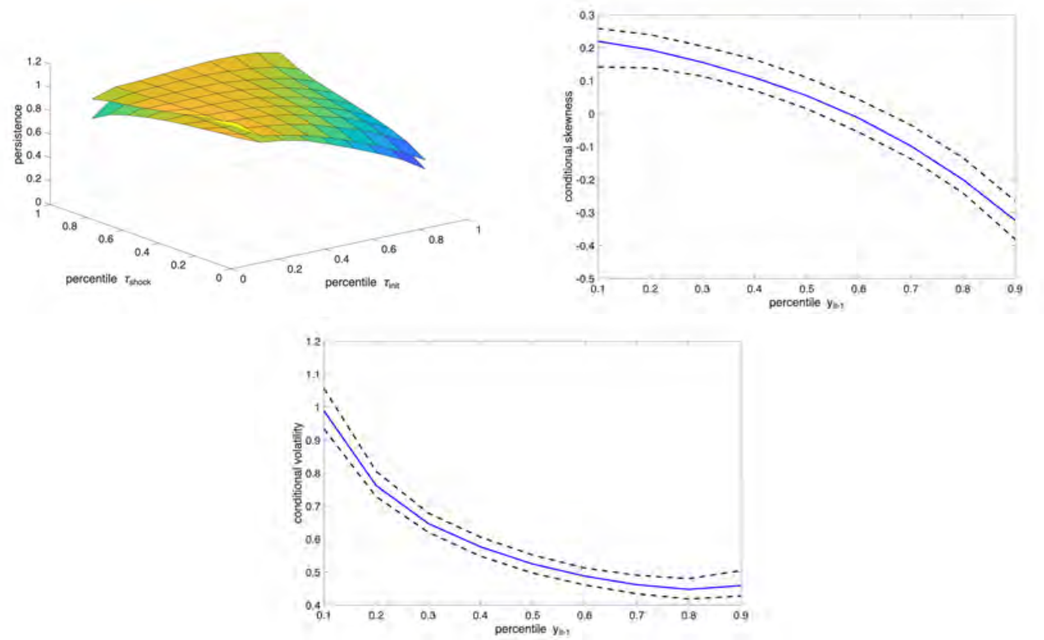


Notes: These plots show estimates of persistence (top), conditional dispersion, measured by the interquantile range in equation (6) (middle), and conditional skewness (bottom) for the nonlinear income process using one-year (left column) and three-year (right column) income data.

Figures G.3 and G.4 report the corresponding 95% nonparametric bootstrap confidence intervals for the nonlinear income-process estimates using one-year and three-year ECV data, respectively.

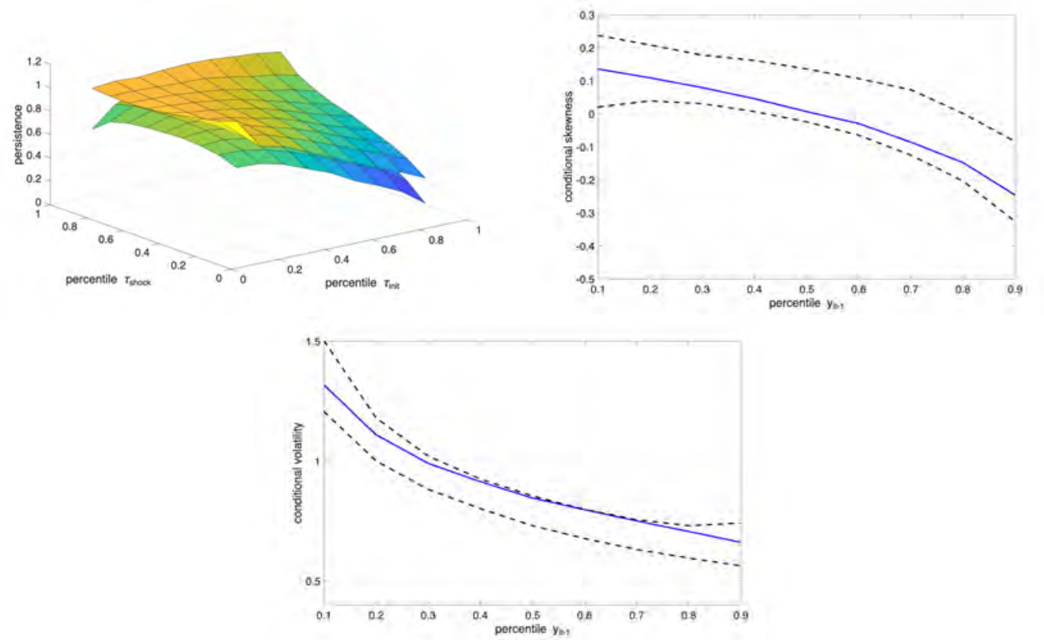
Taken together, the ECV estimates suggest that horizon differences have only modest effects and are unlikely to explain the much larger gap between realized and subjective income processes.

Figure G.3: Nonparametric bootstraps, nonlinear income process, ECV one-year data



Notes: These plots show the 95% confidence bands associated with the nonlinear persistence measure (top left), conditional skewness (top right), and conditional dispersion (bottom) for the income process estimated using one-year ECV data. These were constructed from 500 block bootstrap replications.

Figure G.4: Nonparametric bootstraps, nonlinear income process, ECV three-year data



Notes: These plots show the 95% confidence bands associated with the nonlinear persistence measure (top left), conditional skewness (top right), and conditional dispersion (bottom) for the income process estimated using three-year ECV data. These were constructed from 500 block bootstrap replications.

H Model with tilted probabilities

H.1 Model specification and estimation

The baseline hurdle model in Section 4.2 focuses on a specific reporting pattern: deterministic allocation of all probability mass to the middle bin. To assess whether our results are sensitive to this restriction, we consider here a more general reporting model in which respondents may systematically distort the probabilities implied by their underlying perceived income distribution. In particular, reported probabilities may be tilted toward optimism or pessimism and toward greater concentration or dispersion. We discuss the model set-up and estimation using the state-dependent AR(1) model as an illustration, and report results for all three income-process specifications considered in the paper.

Suppose that households use the state-dependent AR(1) model to answer the subjective income expectations question:

$$y_{it+1} = \rho(\mathbf{X}_{it})y_{it} + \mathbf{X}_{it}'\beta + \eta_i + \sigma(\mathbf{X}_{it})u_{it+1}, \qquad u_{it+1} \sim N(0,1). \tag{H29}$$

Respondents may report the probabilities implied by the underlying income process without additional systematic distortion, or they may report a tilted version of those probabilities. In particular, a fraction ζ_{it} of households draw their responses from a Multinomial distribution with probabilities defined by the underlying income process $p^*_{j,it}(\eta_i)$, while the remaining households draw their points from a tilted probability distribution $\omega_{j,it}(\eta_i, \lambda)$, defined as:

$$\omega_{j,it}(\eta_i, \lambda) = \frac{p^*_{j,it}(\eta_i)w_j(\lambda)}{\sum_{k=1}^{J} p^*_{k,it}(\eta_i)w_k(\lambda)} \tag{H30}$$

where

$$w_j(\lambda) = \exp\{\lambda_1(j - j^\star) + \lambda_2(j - j^\star)^2\}.$$

The tilted probabilities are governed by two parameters. The parameter λ_1 captures directional distortions toward relatively higher or lower income-growth bins, while λ_2 changes the concentration of reported probability mass around the middle of the distribution. This specification therefore allows reporting behavior to reflect optimism or pessimism as well as excessive concentration or spreading, rather than restricting systematic deviations to all-or-nothing middle-bin responses.

Let $T_{it} = 1$ denote reporting without this additional tilt. We model the probability of undistorted reporting as

$$\zeta_{it} \equiv \Pr(T_{it} = 1 \mid \mathbf{Z}_{it}) = \Lambda(\mathbf{Z}_{it}'\gamma), \tag{H31}$$

where T_{it} is a latent indicator equal to one when reported probability points are generated from the probabilities implied directly by the underlying perceived income process.

Conditional on η_i, the likelihood for individual i at time t is:

$$L_{it}(\eta_i) = \zeta_{it} L_{it}^U(\eta_i) + (1 - \zeta_{it}) L_{it}^B(\eta_i),$$

where

$$L_{it}^U(\eta_i) = \prod_{j=1}^{J} \left[p_{j,it}^*(\eta_i)\right]^{d_{jit}}$$

is the likelihood under undistorted reporting, and

$$L_{it}^B(\eta_i) = \prod_{j=1}^{J} \left[\omega_{j,it}(\eta_i, \lambda)\right]^{d_{jit}}$$

is the likelihood under tilted reporting.

The conditional panel likelihood for an individual i is:

$$L_i(\eta_i) = \prod_{t=1}^{T} L_{it}(\eta_i), \tag{H32}$$

To complete the model, we need a specification for the unobserved heterogeneity. The integrated likelihood is:

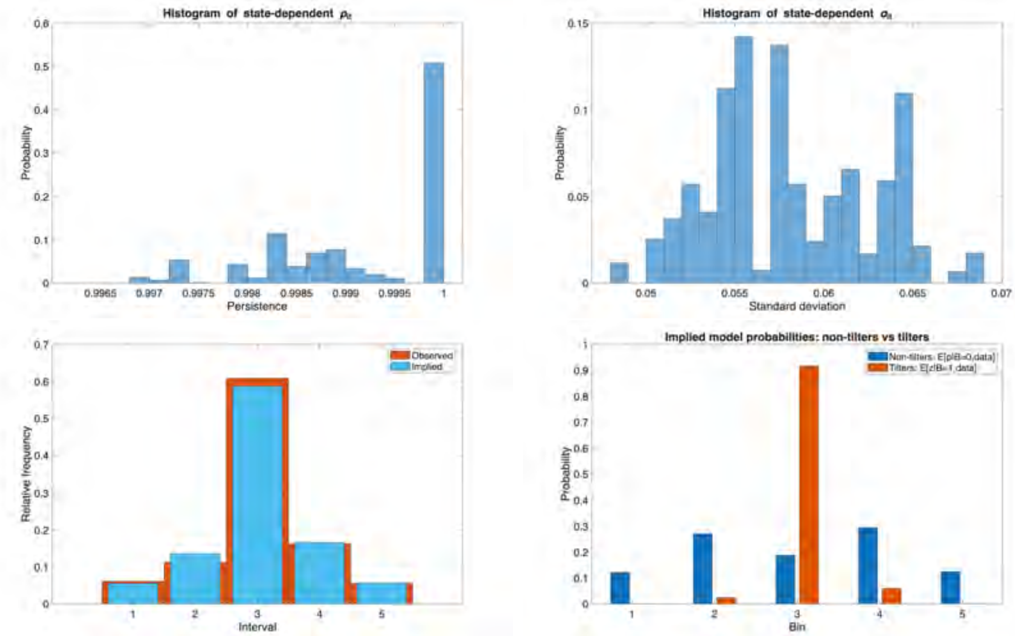
$$L_i = \int \prod_{t=1}^{T} L_{it}(\eta_i) f(\eta_i) d\eta_i. \tag{H33}$$

We estimate this model via the EM algorithm and conduct a nonparametric bootstrap to compute standard errors.

H.2 Results

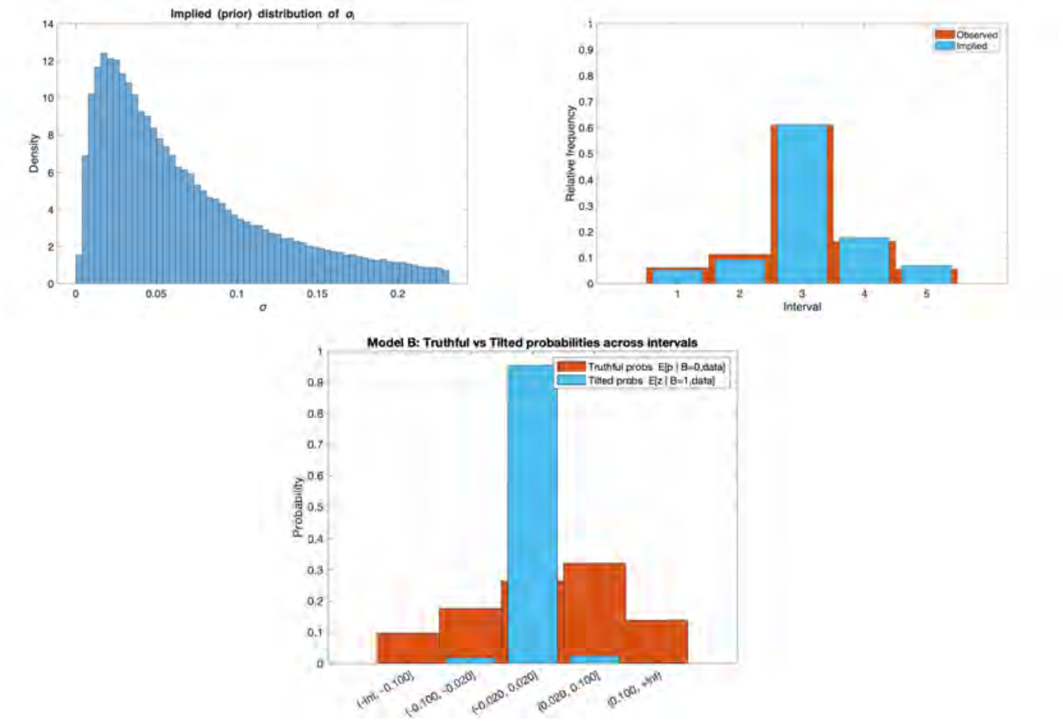
Figures H.1–H.3 present the results from the models with tilted probabilities. The estimates are qualitatively very similar to those obtained under the baseline hurdle model, and the main empirical conclusions remain unchanged: subjective expectations imply substantially more persistent and less dispersed income dynamics than realized income data. Allowing for directional distortions and more general concentration or spreading of reported probabilities therefore does not materially alter our findings. This provides reassurance that the baseline results are not driven by restricting systematic reporting behavior to deterministic middle-bin focal responses.

Figure H.1: Results of the state-dependent AR(1) process, model with tilted probabilities



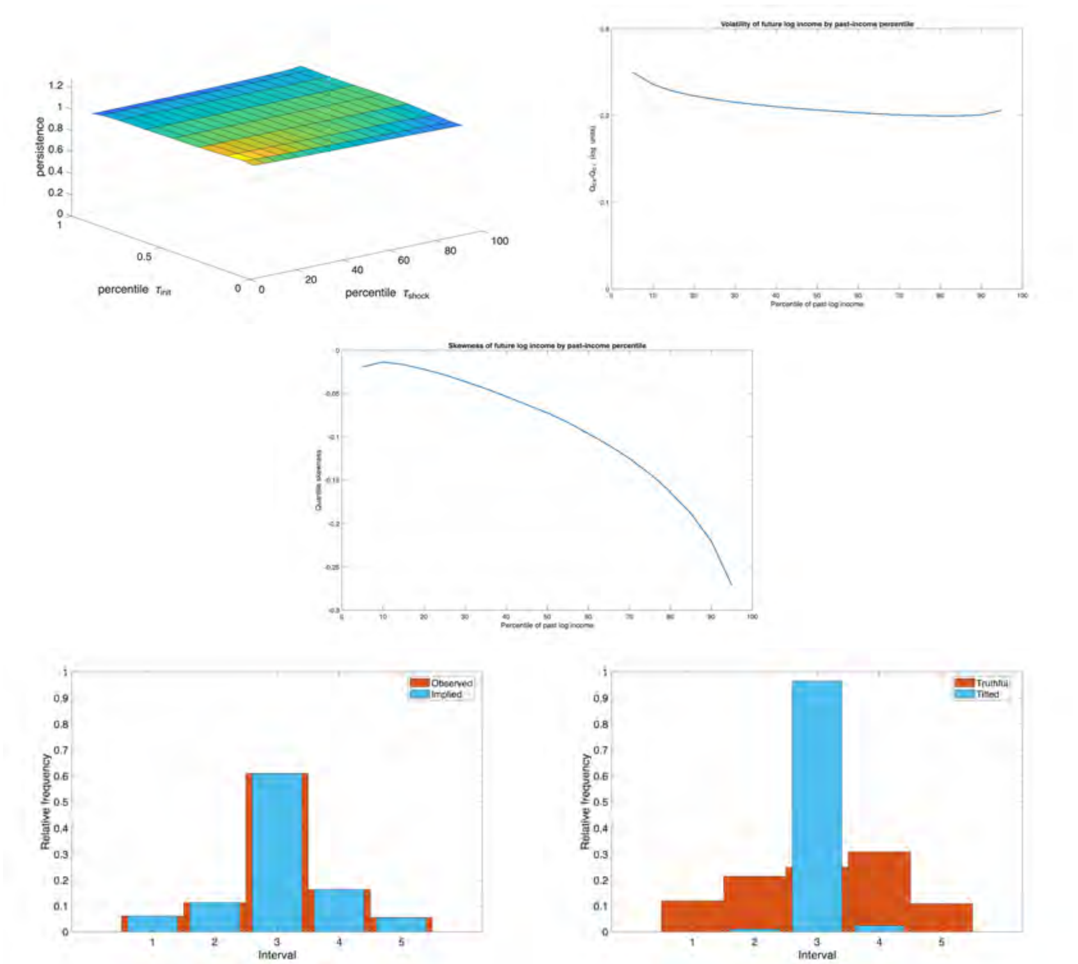
Notes: These plots show estimates of persistence (top left), dispersion (top right), the comparison between the observed and model-implied bin probabilities (bottom left), and the decomposition of bin probabilities between undistorted and tilted reporting (bottom right).

Figure H.2: Results of the heterogeneous AR(1) process, model with tilted probabilities



Notes: These plots show estimates of dispersion (top left), the comparison between the observed and model-implied bin probabilities (top right), and the decomposition of bin probabilities between undistorted and tilted reporting (bottom).

Figure H.3: Results of the nonlinear process, model with tilted probabilities



Notes: These plots show estimates of persistence (top left), conditional dispersion (top right), skewness (middle), the comparison between the observed and model-implied bin probabilities (bottom left), and the decomposition of bin probabilities between undistorted and tilted reporting (bottom right).

I Additional results: life-cycle model

We report additional results from the life-cycle model that complement the baseline exercises described in the main text. As discussed in the last paragraphs of Section 6, we consider two alternative specifications to assess the robustness of our findings to modeling choices regarding time aggregation and preference calibration.

First, we present results for the case in which the discount factor β is kept constant across income processes, rather than being adjusted to match the asset-to-income ratio observed in the data. This exercise allows us to isolate the implications of the estimated income processes for wealth accumulation and consumption insurance without imposing identical average saving rates across specifications. Second, we consider an alternative timing structure in which the model period is set to three years, consistent with the frequency of income observations in the EFF, and adjust the transition matrices of the income processes accordingly. In particular, for the processes estimated using subjective expectations data, we construct the three-year transition matrix as $\mathbb{Q}_3 = \mathbb{Q}_1^3$, where $\mathbb{Q}_1$ is the one-year transition matrix obtained from the estimation.

The results (displayed in Figures I.1, I.2 and Table I.1) confirm the main qualitative findings of the baseline analysis. When β is held fixed, differences in wealth accumulation across income processes become even more pronounced. In particular, income processes estimated using realized data continue to generate substantially higher precautionary saving because they imply greater income dispersion and more pronounced nonlinearities. In contrast, the processes estimated using subjective expectations imply much flatter wealth profiles, reflecting the lower dispersion implied by the processes estimated from subjective expectations. These differences mirror the patterns documented in the main text and highlight the central role of perceived income risk in shaping life-cycle saving behavior.

Similarly, the alternative specification with a three-year model period yields qualitatively similar results (see Figures I.3, I.4 and Table I.2). Adjusting the transition matrices to account for the longer horizon slightly reduces the persistence of income dynamics, but the key conclusions remain. The lower dispersion implied by subjective expectations leads households to save less and accumulate wealth more gradually. The nonlinear process estimated from realized income generates particularly strong precautionary saving motives among high-income households and produces substantially greater wealth accumulation early in life and higher cross-sectional wealth dispersion than the processes recovered from subjective expectations.

Overall, these exercises confirm that differences in income dynamics—particularly in

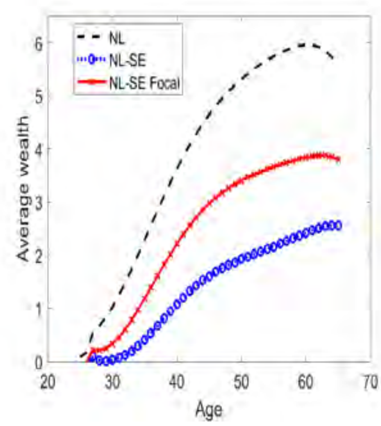


Figure I.1: Wealth Accumulation - Constant β

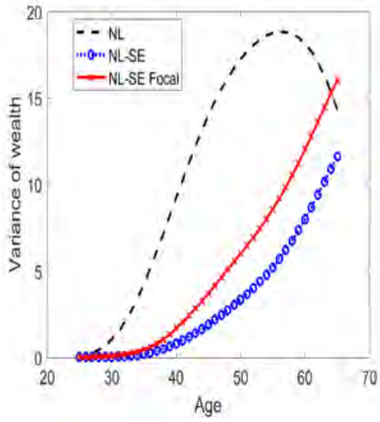


Figure I.2: Cross-sectional Variance of Wealth - Constant β

Income Process	ψ
NL	0.781
NL-SE	0.676
NL-SE Focal	0.756
NL-SE Focal / *NL*	0.609

Table I.1: Consumption Insurance - Constant β

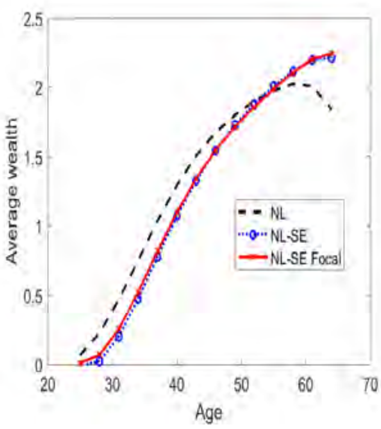


Figure I.3: Wealth Accumulation - 3 Year

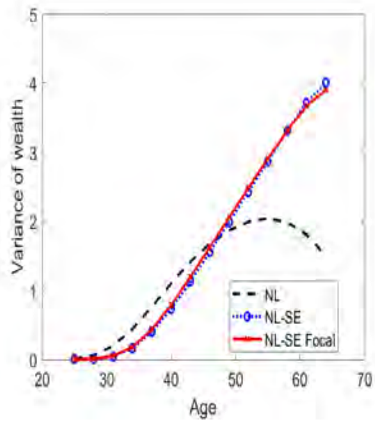


Figure I.4: Cross-sectional Variance of Wealth - 3 Year

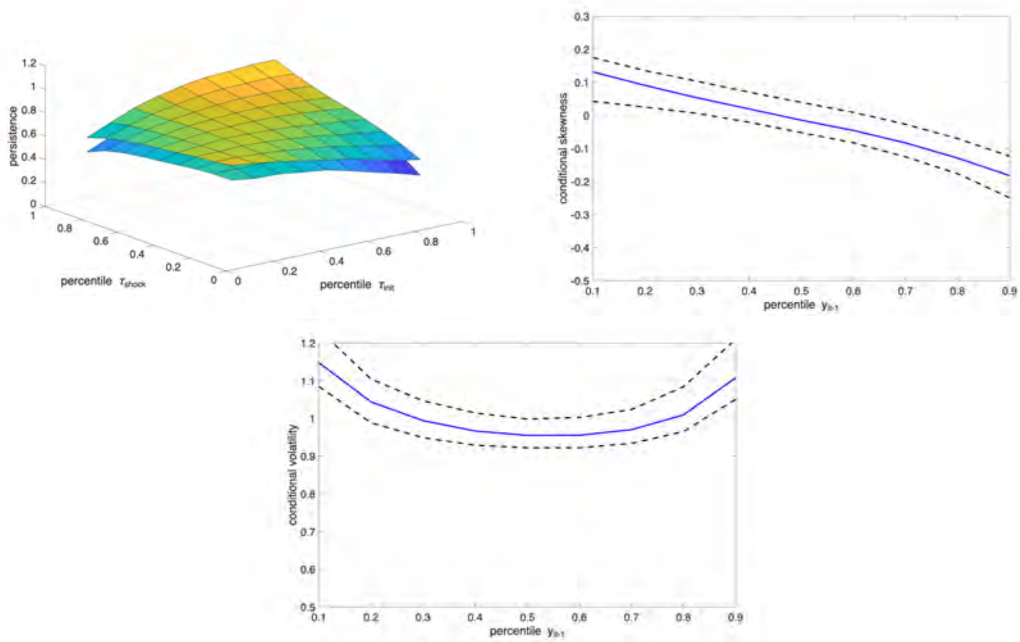
Income Process	ψ
NL	0.70
NL-SE	0.56
NL-SE Focal	0.60
NL-SE Focal / *NL*	0.57

Table I.2: Consumption Insurance - 3 Year dispersion and nonlinearities—have meaningful implications for consumption insurance and life-cycle wealth accumulation, and are robust to alternative preference calibrations and time aggregation.

J Additional confidence intervals

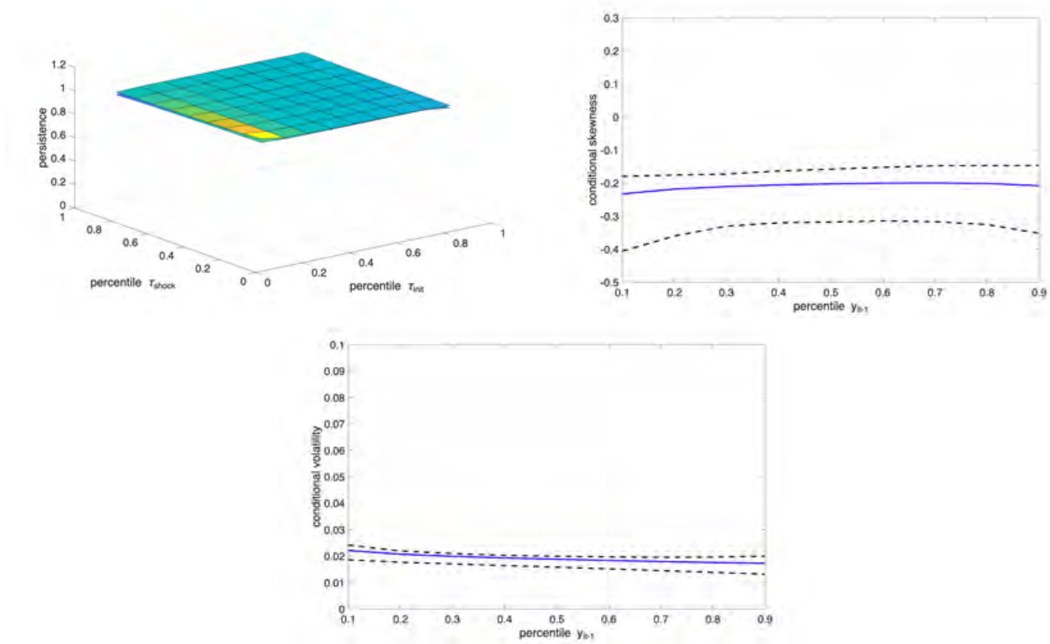
This appendix reports 95% nonparametric bootstrap confidence intervals for the main nonlinear income-process estimates presented in Section 5. Figure J.1 reports the confidence intervals for the nonlinear process estimated from realized EFF income data. Figures J.2 and J.3 report the corresponding confidence intervals for the nonlinear process estimated from subjective expectations without and with the hurdle model for middle-bin focal responses, respectively.

Figure J.1: Nonparametric bootstraps, nonlinear income process with realized EFF data



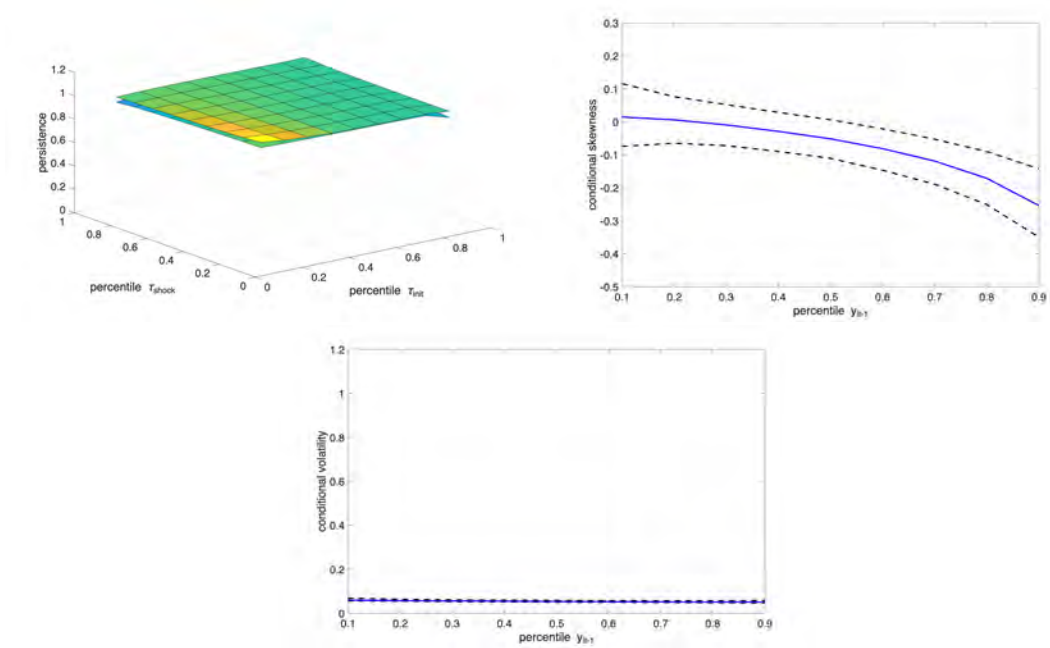
Notes: These plots show the 95% confidence bands associated with the nonlinear persistence measure (top left), conditional skewness (top right), and conditional dispersion (bottom) for the income process estimated from realized EFF income data. These were constructed from 500 block bootstrap replications.

Figure J.2: Nonparametric bootstraps, nonlinear income process with subjective expectations



Notes: These plots show the 95% confidence bands associated with the nonlinear persistence measure (top left), conditional skewness (top right), and conditional dispersion (bottom) for the subjective income process estimated without the hurdle model. These were constructed from 500 block bootstrap replications.

Figure J.3: Nonparametric bootstraps, nonlinear income process with subjective expectations and hurdle model



Notes: These plots show the 95% confidence bands associated with the nonlinear persistence measure (top left), conditional skewness (top right), and conditional dispersion (bottom) for the subjective income process estimated with the hurdle model for middle-bin focal responses. These were constructed from 500 block bootstrap replications.