Heterogeneity in Transitory Income Risk*

Martín Almuzara†

December 2024

[Link to the latest version]

Abstract

I propose a new framework to distinguish between permanent and transitory shocks in short panels that permits general forms of cross-sectional heterogeneity. This includes a model in which log income is the sum of a nonlinear permanent component and a transitory innovation with individual-specific variance. I establish nonparametric identification results, and analyze flexible estimators that use series approximations to latent variable distributions and integrated moment conditions. I also develop a novel local-projections-based approach to recovering dynamic responses to these shocks. Using data from the Panel Study of Income Dynamics, I find that (i) heterogeneity in transitory risk is sizable, (ii) it implies pronounced inequality in total income risk, and (iii) it has a significant but short-lived passthrough into consumption.

Keywords: nonparametric identification, income process, latent variables, series estimation, impulse responses, consumption passthrough.

*I have accumulated an immense debt of gratitude towards Manuel Arellano and Enrique Sentana for all their support and guidance during my PhD. I thank Dante Amengual, Dmitry Arkhangelsky, Robert Moffitt, and seminar participants at CEMFI, UPF, Sciences Po, Cambridge, UCL, NYU and Columbia for useful comments and discussions. All errors are my sole responsibility. The views expressed in this paper are those of the author and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

†Federal Reserve Bank of New York: martin.almuzara@ny.frb.org

1 Introduction

The distinction between permanent and transitory shocks occupies a central place in many areas of economic analysis. It matters for earnings mobility (Lillard and Willis (1978), Meghir and Pistaferri (2004), and Gottschalk and Moffitt (2009)) since the relative size of permanent and transitory shocks determines the persistence of income inequality. It does too for studying labor supply and consumption choices (Hall and Mishkin (1982), Abowd and Card (1989), Deaton and Paxson (1994), Blundell and Preston (1998), Blundell, Pistaferri, and Preston (2008), Altonji, Smith, and Vindangos (2013), Arellano, Blundell, and Bonhomme (2017)), and for quantitative macroeconomic models (e.g., Kaplan and Violante (2010)). Understanding the nature of changes in inequality and household responses to income risk is relevant for both academic and policy discussions.

In practice, permanent and transitory income shocks are not directly observed. For the framework to be useful, one has to identify them from income data. One approach uses observed events—long illness, involuntary job loss, unemployment spells, tax rebates, or lotteries—as proxies for permanent and transitory shocks (e.g., Cochrane (1991), Souleles (1999), Parker, Souleles, Johnson, and McClelland (2013), Misra and Surico (2014)). Another approach relies on deconvolution techniques that exploit restrictions on the dependence among unobservables (e.g., MaCurdy (1982), Horowitz and Markatou (1996), and Bonhomme and Robin (2010)). Some common restrictions include statistical independence between permanent and transitory components and serial independence of transitory shocks. While convenient, those restrictions rule out potentially interesting aspects of income risk, such as unobserved heterogeneity in the variance or skewness of income shocks.

Motivated by that, in this paper I propose a framework to separate permanent and transitory shocks that allows for general forms of time-invariant cross-sectional heterogeneity. A model that belongs in this framework is one where permanent income is a nonlinear first-order Markov process (as in Arellano et al. (2017)), and the variance of transitory shocks is cross-sectionally heterogeneous (as in Chamberlain and Hirano (1999)) and possibly correlated with permanent income. This setup therefore allows for flexible state-dependence and heterogeneity, and substantially weakens the restrictive independence and homogeneity assumptions of previous analyses. In addition, I propose a novel approach to quantifying the dynamic responses to these shocks in panel data, which I apply to measure the consumption passthrough of transitory income risk.

The *first* contribution of my paper is to establish nonparametric identification results in short panels for a class of heterogeneous permanent-transitory models. I show that, under weak conditions, the distribution of time-invariant cross-sectional heterogeneity and those of permanent and transitory components can be recovered from the distribution of observables if enough periods are available.

Relaxing independence and homogeneity of unobserved shocks is challenging as it invalidates the linear and nonlinear deconvolution techniques typically used in the literature to establish identification (e.g., Kotlarski (1967), Székely and Rao (2000) and Wilhelm (2015)). Instead, I build on the spectral decomposition approach developed by Hu and Schennach (2008) in the context of models with nonclassical measurement error.[1] Consider a model with a heterogeneous parameter vector θ_i and observables $y_{i1}, \ldots, y_{iT}$. The idea is to divide the observables in three groups $\{y_{i1}, ..., y_{i,s-1}\}$, y_{is}, $\{y_{i,s+1}, ..., y_{iT}\}$ so that $\{y_{i1}, ..., y_{i,s-1}\}$ and $\{y_{i,s+1}, ..., y_{iT}\}$ can act as noisy measurements of θ_i in a regression of y_{is} on θ_i.[2] In that setup, the key identifying assumption—Assumption 1 below—is that θ_i can be pinned down by a known functional of the distribution of y_{is} given θ_i (for example, the conditional variance of y_{is}). Naturally, for this approach to work one needs that there be at least as many measurements as heterogeneous parameters, and this creates a trade-off between fixed-T identification and heterogeneity. For instance, the model outlined above with one-dimensional heterogeneity in transitory variances needs $T \geq 5$; a model augmented with heterogeneity in permanent variances needs $T \geq 7$, and so on.[3]

The *second* contribution of my paper is to analyze the statistical properties of a class of panel data estimators that relies on series approximation of latent variable distributions and filtered moment conditions. That class offers a computationally and statistically attractive estimation approach for the heterogeneous permanent-transitory models studied in this paper. It also covers the quantile-based estimators proposed by Arellano and Bonhomme (2016) and Arellano et al. (2017) for nonlinear panel data models. In fact, the Stochastic EM algorithm recommended below to implement the estimation approach (Algorithm 1) simply extends those developed in Arellano et al. (2017) and Arellano,

[1]For other applications of this and related approaches see also Hu (2008), Cunha, Heckman, and Schennach (2010), Hu and Shum (2012) and Fryberger (2018). Complete reviews of the literature can be found in Schennach (2020), Schennach (2022) and Hu (2024).

[2]Except that the regression takes place in a space of linear operators. See, e.g., Carrasco, Florens, and Renault (2007) for an overview of the use of linear operator theory in econometrics.

[3]This trade-off is reminiscent of that in random coefficient models (Arellano and Bonhomme, 2012). In the problem I study, the precise condition is that $T \geq 2\dim(\theta_i) + 3$; see Proposition 1.

Blundell, Bonhomme, and Light (2023) by accommodating general forms of heterogeneity. The novelty here is that I provide large-sample approximations to estimators of certain functionals of latent variable distributions that allow the flexibility of the series approximation to grow with the cross-sectional sample size N (the time series dimension T is held constant). To the best of my knowledge, this is the first treatment of the nonparametric asymptotics for this type of problem and may be of independent interest more generally.[4]

The *third* contribution of my paper is to develop methodology to recover impulse responses with respect to the permanent and transitory shocks of the model. One approach, pursued in the early literature, derives identifying restrictions for the responses of outcomes (e.g., consumption) from linearized optimality conditions (e.g., the Euler equation; see Blundell et al. (2008)). One limitation is that usually this only identifies contemporaneous effects but it is silent about the dynamics.[5] Instead, I propose a strategy closer in spirit to the quasi-experimental literature: if the shock of interest were directly observed, one could recover dynamic responses by running local projections (Jordà, 2005). When the shock is not observed but its distribution is nonparametrically identified from the data, one can transform those local projections into integrated moment conditions. In practice, this can be done by running local projections on shocks simulated from their posterior distribution, averaging the projection coefficients over a large number of iterations.

Empirics of transitory income risk. The remaining contributions of my paper are empirical. Using household data from the Panel Study of Income Dynamics (PSID), I estimate a model with nonlinear permanent income and heterogeneity in the variance of transitory shocks. I then use this model to document several novel features of income risk. First, differences across households in transitory income risk are sizable: according to my estimates, almost 70% of households have half the average variance or less, whereas around 20% have twice the average variance or more. Second, even after allowing for richer heterogeneity, there is evidence of nonlinear persistence and decreasing conditional

[4]For example, Hu and Schennach (2008) develop asymptotic approximations to the sieve maximum likelihood estimator of their nonparametric model of nonclassical measurement error, building on the techniques of Shen (1997), Chen and Shen (1998) and Ai and Chen (2003). In contrast, the estimators I analyze allow for pseudo-likelihood moment-based estimation of latent variable distributions.

[5]Another limitation is that imposing linearity may obscure the interpretation. One alternative is to postulate a nonparametric optimal policy function as in Arellano et al. (2017) but this requires correctly specifying (and having data on) the state vector in the agent's problem. For example, if the state vector of the consumer includes liquid and illiquid assets, an optimal policy function that assumes the state to be total wealth will generally not recover the true impulse responses, whereas the approach I propose will still recover an interpretable (linear) estimand.

skewness in permanent income as in Arellano et al. (2017). Moreover, permanent and transitory income risk go hand in hand as household units with higher transitory risk also experience higher dispersion and more negative skewness in persistent income shocks.

Third, transitory risk heterogeneity explains a large extent of the cross-sectional differences in total income risk; the importance of nonlinear dependence by comparison is less evident. Specifically, the variance and skewness of the predictive distribution of future total income have large gradients with respect to unit-specific transitory variances; they are practically flat with respect to permanent income even at long horizons. Lastly, using consumption data, I estimate the dynamic passthrough of transitory income shocks. I find a significantly positive but short-lived response of consumption which, combined with the previous point, indicates potentially large welfare costs of transitory risk.

To be clear, the econometric tools developed here can be easily extended to handle various forms of heterogeneity in the persistent component. Also, the applicability of the permanent-transitory framework goes well beyond income risk. It has been used, e.g., to model firm-level productivity, worker-level wages and household-level portfolio choices. Yet the focus on heterogeneity in transitory income risk has both substantive and empirical justification.

In fact, everyday experience is rich in situations where people differ in transitory income risks. A large literature attributes those differences to risk allocation, to the need to protect investments from hold-up problems, and to asymmetric information (see Malcomson (1999) for a review). Moreover, individuals tend to remain in the same firm and within the same activity for very long periods (for example, because of the accumulation of firm- or activity-specific human capital). The result is a cross-sectional distribution of transitory risks that changes only slowly over the working life of individuals. In short panels, if the researcher does not observe all the determinants of transitory risks, it is then reasonable to treat the distribution as permanent unobserved heterogeneity—the approach I adopt in this paper.

There are three additional reasons why measuring differences in transitory risk empirically is important. First, transitory shocks explains a large fraction of the variance of yearly changes in household incomes (possibly as large as 90% for the US economy). They compose, thus, the bulk of the variability affecting households decisions and welfare. Second, transitory shocks are policy relevant because many economic policies (e.g., tax rebates and stimulus checks) are often perceived as temporary. Third, in models with incomplete markets, heterogeneity in transitory risks implies heterogeneity in self-insurance. This suggests that estimates of insurance coefficients that neglect differences in transitory risks may be misleading.

Related literature. There is a vast literature on income dynamics. Three branches of it are most closely related to my paper. One branch deals with models with nonparametric heteroskedasticity (Botosaru and Sasaki (2018) and Botosaru (2023)). These typically allow the heterogeneity in variances to be time-varying but at the cost of restrictive assumptions needed for identification compared to the ones I use. Moreover, dependence between the persistent component and the heterogeneous variances is ruled out in those papers but it is allowed in mine. Second, Browning, Ejrnæs, and Álvarez (2010) and Alan, Browning, and Ejrnæs (2018) study models with lots of heterogeneity (i.e., with high-dimensional potentially correlated heterogeneous parameters), while Hospido (2012) explicitly includes a worker-specific factor in the variance of wages in her model. My paper differs from them in that I consider a model that distinguishes between permanent and transitory shocks (a two-error model as opposed to their one-error models) and my approach is nonparametric. The latter turns out to be important to capture the nonlinear relation between latent components in the empirical analysis. The two-error model I study is closer to the parametric model in Chamberlain and Hirano (1999) except that I allow for potential nonlinearities in the persistent component and dependence between permanent and transitory risk.

The third strand that matters for my paper is the nonlinear transmission of income shocks, and particularly Arellano et al. (2017). Flexible estimates of predictive distributions reveal that the persistence of past income histories varies with the size and sign of shocks and with the current level of income (Arellano et al., 2017, Figure 1). They also reveal conditional skewness that decreases with current income (Guvenen, Ozkan, and Song (2014) and Arellano et al. (2017)). As I show below, those patterns are consistent with a model with no nonlinear persistence but heterogeneity in transitory risk. This implies a reassessment of the evidence of nonlinearities in total income risk and calls for building models that encompass both features. The message from my paper is that those encompassing models can be nonparametrically identified and estimated in short panels.

Outline. The rest of the paper is organized as follows. Section 2 introduces the heterogeneous permanent-transitory framework to be studied. Section 3 establishes the main nonparametric identification result while Section 4 discusses estimation. The empirical analysis is developed in Section 5. Proofs and additional empirical results are relegated to the appendix and supplemental material.

2 Model

Let y_{it} be the log income of household i at time t observed for units $i = 1, \dots, N$ and over periods $t = 1, \dots, T$. I consider models that decompose y_{it} as the sum of persistent (η_{it}) and transitory (ε_{it}) unobserved components:

$$y_{it} = \eta_{it} + \varepsilon_{it}, \quad t = 1, \dots, T. \tag{1}$$

I use $f_{z|w}$ to denote the conditional density (with respect to Lebesgue measure on the Borel sets of $\mathbb{R}^d$) of the d-dimensional random vector z given the random vector w. I use f_z for the marginal density of z and $f_{z,w}$ for the joint density of z and w.

Given a unit-specific quantity θ_i, the persistent component is first-order Markov,

$$\eta_{i1}|\theta_i \sim f_{\eta_1|\theta}, \tag{2}$$

$$\eta_{it}|\eta_{i,t-1}, \dots, \eta_{i1}, \theta_i \sim f_{\eta_t|\eta_{t-1},\theta}, \quad t = 2, \dots, T, \tag{3}$$

while the transitory component is serially independent,

$$\varepsilon_{it}|\varepsilon_{i,t-1}, \dots, \varepsilon_{i1}, \theta_i \sim f_{\varepsilon_t|\theta}, \quad t = 1, \dots, T. \tag{4}$$

Conditional on θ_i, persistent and transitory components are independent.

Thus, θ_i introduces a new element of cross-sectional unobserved heterogeneity. In the context of income dynamics, I think of it as capturing permanent differences across households in transitory risk (embodied in $f_{\varepsilon_t|\theta}$), although θ_i may also affect the distribution of $\eta_{i1}, \dots, \eta_{iT}$. For example, in my empirical analysis, θ_i governs the variance of transitory income shocks and, therefore, heterogeneity in θ_i reflects differences in households' exposure to transitory risk. To be more specific, in the analysis of Section 5 equation (4) simplifies to

$$\begin{aligned} \varepsilon_{it} &= \theta_i e_{it}, \\ e_{it}|e_{i,t-1}, \dots, e_{i1}, \theta_i &\sim f_{e_t}, \quad t = 1, \dots, T. \end{aligned} \tag{4'}$$

The variance of the transitory income component is $\operatorname{Var}(\varepsilon_{it}|\theta_i) = \theta_i^2 \sigma_{e_t}^2$ where $\sigma_{e_t}^2$ is the variance corresponding to density f_{e_t}. Households with large θ_i are therefore subject to higher transitory income risk than those with small θ_i.

To close the model, I assume θ_i is an absolutely continuous random vector with

$$\theta_i \sim f_\theta. \tag{5}$$

In summary, the parameters of the model are the densities

$$f_\theta, f_{\eta_1|\theta}, f_{\eta_2|\eta_1,\theta}, \ldots, f_{\eta_T|\eta_{T-1},\theta}, f_{\varepsilon_1|\theta}, \ldots, f_{\varepsilon_T|\theta}.$$

The task for the researcher is to recover (a subset of) the parameters from the joint density of observables $f_{y_1,\ldots,y_T}$ — This is the nonparametric identification question I address in Section 3.

Model (1), (2), (3), (4) and (5) encompasses some of the most prominent models in the income dynamics literature. The presence of θ_i builds extra heterogeneity into the nonlinear income framework of Arellano et al. (2017). The nonparametric nature of latent variables and heterogeneity, on the other hand, makes my model a complement to approaches that emphasize second-order features of persistent and transitory components, such as the heterogeneous-income-profile (HIP) model (Guvenen (2007)), and parametric approaches, such as the heterogeneous-transitory-risk (HTR) model (Chamberlain and Hirano (1999)) or the lots-of-heterogeneity model (Browning et al. (2010)).[6]

Often, state-dependence and heterogeneity have different economic implications but are difficult to distinguish empirically when T is small. The model above is no exception as I illustrate in Section 2.2. Thus, it is important to consider whether identification is possible and under what conditions. Since unobserved heterogeneity is permanent, I am interested in identification for a fixed and small T as it is rare for individual-specific

[6] Let U_{it}, V_{it} be normally distributed and mutually, cross-sectionally and serially independent. In Chamberlain and Hirano (1999) the persistent component is $\eta_{it} = \theta_{i,1} + \rho\eta_{it-1} + V_{it}$ and the transitory $\varepsilon_{it} = \theta_{i,2}U_{it}$, while in Guvenen (2007) the persistent component is $\eta_{it} = \theta_{i,1} + \theta_{i,2}t + \rho\eta_{i,t-1} + V_{it}$ and the transitory $\varepsilon_{it} = U_{it}$. Browning et al. (2010) model $y_{it} = \tilde{y}_{it} + e_{it}$ where $\tilde{y}_{it}$ is an ARMA(1, 1) process with a special trend function and ARCH(1) errors,

$$\tilde{y}_{it} = \mu(\theta_{i,1}, \theta_{i,2}, \theta_{i,3}, \theta_{i,4}, t) + \theta_{i,1}\tilde{y}_{i,t-1} + \sigma_{it}V_{it} + \theta_{i,5}\sigma_{i,t-1}V_{i,t-1},$$
$$\sigma^2_{it} = \nu(\theta_{i,6}, t) + \theta_{i,7}\sigma^2_{i,t-1}V^2_{i,t-1}$$

and a white noise measurement error $e_{it} = \theta_{i,8}U_{it}$. The reduced-form for y_{it} will also be ARMA(1, 1) and, under certain conditions, will be well approximated by the sum of an AR(1) around a heterogeneous deterministic trend function η_{it} and a white noise process ε_{it}.

features to remain constant over long periods.[7] Section 3 will reveal that there is a tension between allowing for additional dimensions of heterogeneity and the number of periods T needed for identification.

In practice, researchers may have access to covariates x_i (age, education, state of residence, etc.) that allow her to control for observable forms of heterogeneity. It is straightforward to include them in the framework of this paper. One simply adds an additional conditioning variable to the densities above, i.e.,

$$f_{\theta|x}, f_{\eta_1|\theta,x}, f_{\eta_2|\eta_1,\theta,x}, \ldots, f_{\eta_T|\eta_{T-1},\theta,x}, f_{\varepsilon_1|\theta,x}, \ldots, f_{\varepsilon_T|\theta,x},$$

and carries the analysis conditioning on x_i.[8] The identification argument of Section 3 applies with minor modifications. I highlight the most salient changes and how to flexibly accommodate covariates during estimation later on.

The rest of this section is devoted to three additional tasks. First, I particularize model (1)–(5) to one with simpler state-dependence and heterogeneity in transitory variances (Section 2.1). Second, I show that, despite its simplicity, that model can generate nonlinear patterns in income data that resemble those from models with richer state-dependence but simpler heterogeneity (Section 2.2). Third, Section 2.3 discusses some of the interesting empirical objects that can be recovered from model (1)–(5) and how it can serve as a building block for an analysis of consumption.

2.1 Heterogeneous transitory risk

The canonical model of income risk assumes that η_{it} evolves as a pure random walk with innovations independent of the transitory shocks ε_{it}. A natural extension is the heterogeneous-transitory-risk (HTR) model that allows variances of transitory shocks to differ across households.

The persistent component in the HTR model simplifies to

$$\eta_{it} = \eta_{i,t-1} + v_{it},$$

[7] A good example is the life-cycle pattern found in the persistence and variance of income shocks (e.g., Karahan and Ozkan (2013)).

[8] The measurement equation (1) sometimes includes a deterministic component: $y_{it} = \mu_{it} + \eta_{it} + \varepsilon_{it}$ where $\mu_{it} = m_t(x_i)$ for some function m_t known up to a finite-dimensional parameter. The deterministic component is identified by requiring that η_{it} and ε_{it} are zero-mean independent of x_i. Alternatively, one can absorb μ_{it} into the conditional mean of η_{it} given x_i.

$$\eta_{i1}|\theta_i \sim f_{\eta_1|\theta},$$
$$v_{it}|\eta_{i,t-1},\ldots,\eta_{i1},\theta_i \sim f_{v_t}, \quad t=2,\ldots,T,$$

and the transitory component becomes

$$\varepsilon_{it} = \theta_i e_{it},$$
$$e_{it}|e_{i,t-1},\ldots,e_{i1},\theta_i \sim f_{e_t}, \quad t=1,\ldots,T.$$

Here, $v_{i2},\ldots,v_{iT},e_{i1},\ldots,e_{iT}$ are mutually independent and independent of (η_{i1},θ_i), the two sources of cross-sectional heterogeneity. Therefore, the parameters of the HTR model reduce to $f_\theta, f_{\eta_1|\theta}, f_{v_2},\ldots,f_{v_T}, f_{e_1},\ldots,f_{e_T}$.

Also, the model assumes zero-mean shocks, $E\left[v_{it}\right] = E\left[e_{it}\right] = 0$, implying

$$E\left[y_{it}\middle|\eta_{i1},\theta_i\right] = \eta_{i1},$$
$$\text{Var}\left(y_{it}\middle|\eta_{i1},\theta_i\right) = \sum_{s=2}^{t}\sigma_{v_s}^2 + \theta_i^2\sigma_{e_t}^2,$$

where $\sigma_{v_t}^2 = \text{Var}\left(v_t\right)$ and $\sigma_{e_t}^2 = \text{Var}\left(e_t\right)$, so that η_{i1} captures permanent differences in the *level* of income and θ_i captures permanent differences in its *variability*. In the presence of covariates x_i, η_{i1} and θ_i would capture permanent differences in level and variability beyond what could be explained by x_i. In empirical calculations, even when very good controls are available, the explanatory power of x_i is typically low which suggests that accounting for η_{i1} and θ_i is key.

Dependence between θ_i and η_{i1} is allowed and it is important from a substantive point of view. A model with homogeneous transitory income variances ignores potential correlation between income level and income risk, which can have implications for consumption insurance and welfare calculations. From the statistical point of view, dependence between permanent and transitory income components raises challenges for identification as it invalidates Kotlarski (1967) type techniques.

2.2 Heterogeneity vs state dependence

The approach to quantifying nonlinearities proposed by Arellano et al. (2017) for their income process (the ABB model, henceforth) can be extended to (1)–(5). Let $\tau \mapsto Q_t(\eta_{i,t-1},\theta_i,\tau)$ be the conditional quantile function associated to density $f_{\eta_t|\eta_{t-1},\theta}$. The measures of nonlinear persistence and skewness of η_{it} are

$$\rho_t(\eta_{i,t-1}, \theta_i, \tau) = \left.\frac{\partial Q_t(\eta, \theta_i, \tau)}{\partial \eta}\right|_{\eta=\eta_{i,t-1}},$$

$$\text{sk}_t(\eta_{i,t-1}, \theta_i) = \frac{Q_t(\eta_{i,t-1}, \theta_i, \tau_0) + Q_t(\eta_{i,t-1}, \theta_i, 1-\tau_0) - 2Q_t(\eta_{i,t-1}, \theta_i, 1/2)}{Q_t(\eta_{i,t-1}, \theta_i, \tau_0) - Q_t(\eta_{i,t-1}, \theta_i, 1-\tau_0)},$$

for some $\tau_0 > 1/2$.

The HTR model has $\rho_t(\eta_{i,t-1}, \theta_i, \tau) = 1$ and $\text{sk}_t(\eta_{i,t-1}, \theta_i)$ constant in $\eta_{i,t-1}$ while the ABB model has nonlinear persistence and conditional skewness decreasing in $\eta_{i,t-1}$ (albeit with ρ_t and sk_t independent of θ_i). If we use $\tilde{\rho}_t$ and $\tilde{\text{sk}}_t$ to denote analogous measures obtained from the conditional density of observables $f_{y_t|y_{t-1}}$, a noteworthy fact is that both models produce similar nonlinearities in $\tilde{\rho}_t$ and a decreasing $\tilde{\text{sk}}_t$ as a function of $y_{i,t-1}$. To illustrate this, I conduct the following simulation experiment. I generate data from the HTR model with

$$\theta_i \sim \Gamma\left(\nu/2, 2/\sqrt{\nu(\nu+2)}\right),\ \eta_{i1} \sim N\left(0, \sigma^2_{\eta_1}\right),\ v_{it} \sim N\left(0, \sigma^2_{v_t}\right),\ e_{it} \sim N\left(0, \sigma^2_{e_t}\right).$$

I calibrate the parameters as $\sigma^2_{\eta} = 0.15$, $\sigma^2_{v_t} = 0.01$, $\sigma^2_{e_t} = 0.05$, and $\nu = 11.45$. These numbers are in line with the nonparametric estimates I obtain for the HTR model in Section 5.[9] I simulate $1,000$ samples with $N = 1,000$ and $T = 6$ and, in each of them, I estimate quantile autoregressions of y_{it} on a third-order Hermite polynomial of $y_{i,t-1}$ to obtain $\tilde{\rho}_t$ and $\tilde{\text{sk}}_t$. Figure 1 reports the average across simulations, which is to be compared with figures 2(b) and 4(a) in Arellano et al. (2017).

How can the HTR model generate nonlinear persistence in $\tilde{\rho}_t$? The most relevant pattern is that high-income (low-income) units subject to a bad (good) shock display lower persistence. The intuition is as follows. Households whose income (minus the mean) is high in absolute value in $t-1$ experiencing a big change of the opposite sign in t are likely to have a high θ_i. Therefore, for them, transitory income shocks explain a larger fraction of total income variation and the link between past and current income is weaker.

An implication of this experiment is that the evidence of nonlinearities in $f_{y_t|y_{t-1}}$, which appears to be pervasive in both survey and administrative datasets across countries and periods, is just as favorable to permanent unobserved heterogeneity in exposure to transitory shocks as it is to nonlinear transmission stories. I will show in Section 3 that model

[9]The experiment assumes η_{i1} and θ_i independent of each other, but Section 5 permits dependence.

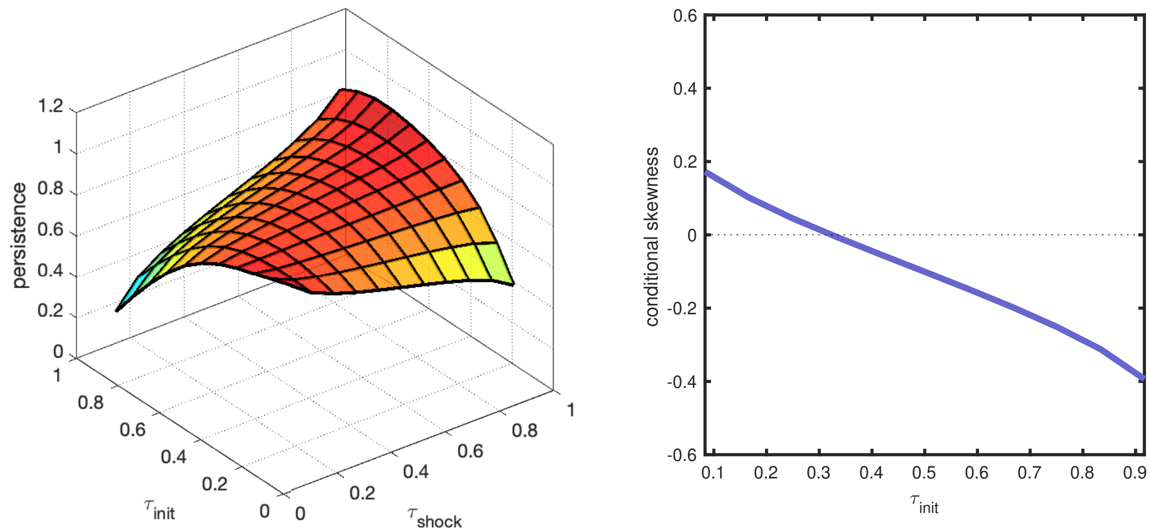

FIGURE 1. Nonlinear persistence and skewness in data simulated from the HTR model.
Note: Average of $\tilde{\rho}_t$ and $\tilde{\text{sk}}_t$ over 1,000 samples simulated from the HTR model with $N = 1,000$ and $T = 6$ estimated by quantile autoregressions of y_{it} on a third-order Hermite polynomial in y_{it-1}.

(1)–(5), which allows for rich forms of both state-dependence and heterogeneity, is nonparametrically identified for $T \geq 5$. Estimating the model using short panels of household income can then help elucidate the source of the nonlinearities in $f_{y_t|y_{t-1}}$, a discussion I develop in Section 5.

2.3 Decomposing income risk

In the income dynamics context, recovering f_θ is of interest because θ_i summarizes a key dimension of inequality. To develop an interpretable measure, consider the predictive distribution of $y_{i,t+h}$ for a unit who knows $\eta_{i1}, \ldots, \eta_{it}, \varepsilon_{i1}, \ldots, \varepsilon_{it}$ and θ_i. Its variance can be decomposed into persistent and transitory contributions:

$$\text{Var}\left(y_{i,t+h} \middle| \eta_{i1}, \ldots, \eta_{it}, \varepsilon_{i1}, \ldots, \varepsilon_{it}, \theta_i\right) = \text{Var}\left(\eta_{i,t+h} \middle| \eta_{it}, \theta_i\right) + \text{Var}\left(\varepsilon_{i,t+h} \middle| \theta_i\right).$$

Let $\sigma^2_{\eta_{t+h}|\eta_t,\theta}$ be the variance corresponding to the conditional density $f_{\eta_{t+h}|\eta_t,\theta}$. In the version of model (1)–(5) that I use in the empirical analysis (specializing (4) to (4′)), the decomposition becomes

$$\text{Var}\left(y_{i,t+h}\middle|\eta_{i1},\ldots,\eta_{it},\varepsilon_{i1},\ldots,\varepsilon_{it},\theta_i\right)=\sigma^2_{\eta_{t+h}|\eta_t,\theta}(\eta_{it},\theta_i)+\theta_i^2\sigma^2_{e_{t+h}}. \tag{6}$$

Arellano, Bonhomme, De Vera, Hospido, and Wei (2022) show that the predictive variance is proportional to a first-order approximation of the welfare gains from eliminating income risk for a household with power utility if log-income shocks are normally distributed. Consequently, whether η_{it} or θ_i explain differences in $\text{Var}\left(y_{i,t+h}\middle|\eta_{i1},\ldots,\eta_{it},\varepsilon_{i1},\ldots,\varepsilon_{it},\theta_i\right)$ across households is a substantive question.

If income shocks were non-normal, higher-order moments should also be taken into account. The additive decomposition of the variance easily extends to other cumulants.[10] For example, let $\kappa^3_{\eta_{t+h}|\eta_t,\theta}$ and $\kappa^3_{e_{t+h}}$ be the skewness (i.e., the third central moment) corresponding to densities $f_{\eta_{t+h}|\eta_t,\theta}$ and $f_{e_{t+h}}$. The skewness of the predictive distribution decomposes into

$$\text{Skew}\left(y_{i,t+h}\middle|\eta_{i1},\ldots,\eta_{it},\varepsilon_{i1},\ldots,\varepsilon_{it},\theta_i\right)=\kappa^3_{\eta_{t+h}|\eta_t,\theta}(\eta_{it},\theta_i)+\theta_i^3\kappa^3_{e_{t+h}}. \tag{7}$$

In the ABB model both the variance and skewness vary with η_{it} but there is no role for θ_i and differences in income risk stem mostly from state dependence. In the HTR model (Section 2.1), the reverse holds: η_{it} plays no role since $\sigma^2_{\eta_{t+h}|\eta_t,\theta}$ and $\kappa^3_{\eta_{t+h}|\eta_t,\theta}$ are the variance and skewness of permanent-income shocks v_{it}, and differences in income risk reflect differences in transitory risk. I report decompositions (6) and (7) from model (1)–(5) and from the ABB and HTR models in Section 5.

Of course, a key goal of analyses that decompose income risk into persistent and transitory sources is to produce empirical estimates of the transmission of income shocks to consumption. This is informative about households' self-insurance and about the welfare implications of income risk, and it is the topic of Section 5.4.

[10] Quantile-based measures of dispersion and asymmetry, such as the Kelley measure of skewness used in Section 2.2, do not admit an additive decomposition. This is why in this paper I focus on their cumulant-based counterparts for the purpose of quantifying the contribution of state-dependence (η_{it}) and heterogeneity (θ_i) to income risk.

3 Identification

In this section I discuss the nonparametric identification of model (1)–(5). The key result is that if $T \geq 2\dim(\theta_i) + 3$ and a normalization condition holds that allows the researcher to pin down the heterogeneity from the distribution of ε_{it} (Assumption 1 below), then f_θ and $\{f_{\eta_t|\eta_{t-1},\theta}, f_{\varepsilon_t|\theta}\}_{1<t<T}$ are identified from $f_{y_1,\ldots,y_T}$.

Assumptions. All units $i = 1, \ldots, N$ are i.i.d. draws from a common distribution. Hence, take any i and fix s such that both $s - 1 \geq \dim(\theta_i) + 1$ and $T - s \geq \dim(\theta_i) + 1$. Also let Θ be the support of θ_i. The key identifying condition is the following:

Assumption 1 (Normalization). *$E\left[\varepsilon_{it} \middle| \theta_i = \theta\right] = 0$ for each $t = 1, \ldots, T$ and there is a known functional M such that $M\left[f_{\varepsilon_s|\theta}(\cdot|\theta)\right] = \theta$ for each $\theta \in \Theta$.*

Assumption 1 is a mild requirement, typically determined by the interpretation of θ_i itself. As an example consider specification (4′) together with $\sigma^2_{e_s} = \mathrm{Var}\,(e_s) = 1$ for some s. Assuming $\Theta \subseteq \mathbb{R}_{\geq 0}$, we then have

$$M\left[f_{\varepsilon_s|\theta}(\cdot|\theta)\right] = \sqrt{\mathrm{Var}\,(\varepsilon_{is} \middle| \theta_i = \theta)} = \theta,$$

for each $\theta \in \Theta$. This makes θ_i the heterogeneous standard deviation of transitory shocks in period s. Moreover, it is clear that a condition of this type is necessary for identification for otherwise one can always multiply e_{is} and divide θ_i by some constant $c > 0$ without modifying the data. Alternatively, one could use quantiles or higher-order moments of $f_{\varepsilon_s|\theta}$ to pin down θ_i.[11]

I will also adopt the following:

Assumption 2 (Bounded densities). *The joint density $f_{y_1,\ldots,y_T,\eta_1,\ldots,\eta_T,\theta}$ is bounded and so are all the corresponding marginal and conditional densities.*

Assumption 3 (Completeness). *The families of densities*

$$\mathcal{F}_- = \left\{f_{\eta_s,\theta|y_1,\ldots,y_{s-1}}(\cdot|y_-) : y_- \in \mathcal{Y}_-\right\},$$

[11]For example, let Q_{e_t} and $Q_{\varepsilon_t|\theta}$ be the quantile functions for e_{it} and ε_{it} conditional on θ_i, respectively, in specification (4′). If we assume $Q_{e_s}(\tau) - Q_{e_s}(1-\tau) = 1$ for some $\tau > 1/2$, then

$$M\left[f_{\varepsilon_s|\theta}(\cdot|\theta)\right] = Q_{\varepsilon_s|\theta}(\tau|\theta) - Q_{\varepsilon_s|\theta}(1-\tau|\theta) = \theta,$$

for all $\theta \in \Theta$.

$$\mathcal{F}_+ = \left\{ f_{\eta_s,\theta|y_{s+1},\dots,y_T}(\cdot|y_+) : y_+ \in \mathcal{Y}_+ \right\},$$

where $\mathcal{Y}_-$ and $\mathcal{Y}_+$ are the supports of $(y_{i1}, \dots, y_{i,s-1})$ and $(y_{i,s+1}, \dots, y_{iT})$, are complete.

The requirement of bounded densities in Assumption 2 is typically satisfied in applications but rules out distributions for which significant probability mass concentrates on a small region of the support.[12] The completeness requirement in Assumption 3 is standard in nonparametric problems (e.g., Newey and Powell (2003), Chernozhukov and Hansen (2005) and Blundell, Chen, and Kristensen (2007)), and primitive conditions have been studied in the literature (e.g., D'Haultfoeuille (2011)). In linear IV, completeness is akin to the relevance condition that asks for nonzero correlation between endogenous variable and instrument.

One important observation is that for Assumption 3 to hold, $s-1$ and $T-s$ cannot be less than $\dim(\theta_i)+1$. This is linked to the fact that there can be no less instruments than endogenous variables in linear IV if identification is to be achieved. Hence, the requirement that $T \geq 2\dim(\theta_i)+3$. Conversely, if $T < 2\dim(\theta_i)+3$, examples can be constructed in which Assumptions 1 and 2 hold but identification fails. The restriction on T is therefore essential and implies a trade-off between allowing for rich forms of heterogeneity and fixed-T identification.

Finally, it is important to emphasize that Assumptions 1, 2 and 3 are very general and that it is easy to specify models that satisfy all of them. With the normalization discussed above, the HTR model of Section 2.1 where (a) v_{it} and e_{it} are normally distributed and (b) θ_i is gamma distributed with shape parameter exceeding unity is one example. This covers the model in Chamberlain and Hirano (1999) and the one I used in the simulations of Section 2.2.

Nonparametric identification. I can now state the main identification result.

Proposition 1. *Under Assumptions 1, 2 and 3, if $T \geq 2\dim(\theta_i)+3$, then the densities $\{f_{\eta_t|\eta_{t-1},\theta}, f_{\varepsilon_t|\theta}\}_{1<t<T}$ and f_θ are nonparametrically identified.*

Proof. See Appendix A. □

[12] For example, the gamma distribution with shape parameter below unity has an unbounded density near zero. This is close to a mixed distribution with a point mass at zero. Nonparametric identification with mixed distributions is not covered by Proposition 1 but it is an interesting open question.

Remark 1. The idea behind the proof is to view the problem as a measurement error model where $Y = y_{is}$ is the regressand, $Z = (\eta_{is}, \theta_i)'$ is the unobservable regressor, and $Y_- = (y_{i1}, \dots, y_{i,s-1})'$ and $Y_+ = (y_{i,s+1}, \dots, y_{iT})'$ are noisy measurements. Given Z, the elements Y_-, Y and Y_+ are mutually independent, which leads to

$$f_{Y_-,Y,Y_+}(y_-, y, y_+) = \int_{\mathcal{Z}} f_{Y_-|Z}(y_-|z) f_{Y,Z}(y, z) f_{Y_+|Z}(y_+|z)\, dz \tag{8}$$

for all $y_- \in \mathcal{Y}_-$, $y \in \mathcal{Y}$, $y_+ \in \mathcal{Y}_+$ where $\mathcal{Z}$ and $\mathcal{Y}$ are the supports of Z and Y. In this setup, nonparametric identification can be obtained by diagonalizing a certain linear operator, as in Hu and Schennach (2008). Assumptions 2 and 3 ensure that the linear operator is well defined and the diagonalization can be performed.

Assumption 1 is invoked to guarantee uniqueness of the spectral decomposition from where a unique set of densities $\{f_{Y_-|Z}, f_{Y,Z}, f_{Y_+|Z}\}$ emerges that solves (8). A key difference with the measurement error setup is the type of normalization. To pin down the latent Z, I normalize $f_{Y|Z}$ instead of $f_{Y_-|Z}$ or $f_{Y_+|Z}$ (cf. Hu and Schennach (2008, Assumption 5)). In measurement error problems, $f_{Y|Z}$ is the main structural equation while $f_{Y_-|Z}$ and $f_{Y_+|Z}$ embody the distributions of measurement error. Thus, restricting $f_{Y|Z}$ is not very appealing. In contrast, in an income process, $f_{Y|Z}$ is the measurement equation that maps latent variables to observables and restricting it is usually a natural step.

Identifying $f_{Y_-|Z}$, $f_{Y,Z}$ and $f_{Y_+|Z}$ delivers f_Z and $f_{Y_-,Y,Y_+|Z}$, and further manipulations (as in Arellano et al. (2017)) deliver the remaining latent variable distributions. See Appendix A for the formal argument.

Remark 2. Some implications of Proposition 1 are as follows. First, the model that assumes θ_i is the volatility of transitory income shocks (4′) is identified if $T \geq 5$. Second, in the presence of covariates x_i, $f_{\theta|x}$ and $\{f_{\eta_t|\eta_{t-1},\theta,x}, f_{\varepsilon_t|\theta,x}\}_{1<t<T}$ are identified from $f_{y_1,\dots,y_T|x}$ if Assumptions 1, 2 and 3 hold conditioning on x_i. Third, in the HTR model, $\{f_{v_t}, f_{e_t}\}_{1<t<T}$ are identified too.

Also notice that Proposition 1 does not provide identification of the initial- and end-period densities $f_{\eta_1|\theta}, f_{\eta_T|\eta_{T-1},\theta}, f_{\varepsilon_1|\theta}, f_{\varepsilon_T|\theta}$. Identification of them can be achieved under an additional stationarity assumption $f_{\varepsilon_t|\theta} = f_{\varepsilon|\theta}$ for all $t = 1, \dots, T$.

Although the identification argument is tailored to heterogeneity in transitory risk, it is possible to adapt the approach to cover other setups, such as the HIP model of Guvenen (2007) and some restricted versions of the "lots of heterogeneity" framework of Browning et al. (2010).[13]

Remark 3. Proposition 1 also implies identification of other objects, such as $\mathrm{Var}\left(\theta_i^2\right)$. In the simpler HTR model of Section 2.1, however, we can write those objects more explicitly as functions of the data. For example, for $\tau > t+1$,

$$\mathrm{Var}\left(\theta_i^2\right) = \frac{\mathrm{Cov}\left((\Delta y_{it})^2, (\Delta y_{i\tau})^2\right)}{\mathrm{Var}\left(\Delta \varepsilon_{it}\right)\mathrm{Var}\left(\Delta \varepsilon_{i\tau}\right)},$$
$$\mathrm{Cov}\left(\eta_{i1}, \theta_i^2\right) = \frac{\mathrm{Cov}\left(y_{it}, (\Delta y_{i\tau})^2\right)}{\mathrm{Var}\left(\Delta \varepsilon_{i\tau}\right)}.$$

Since, as in the canonical income process, second moments of η_{it} and ε_{it} are identified by the covariance structure of y_{it}, the HTR model imposes a joint covariance structure on y_{it} and $(\Delta y_{it})^2$, with overidentifying restrictions if $T \geq 5$.

4 Estimation and inference

Many objects of empirical interest (e.g., measures of persistence and skewness, or income risk decompositions) are smooth functionals of the densities $f_{\eta_t|\eta_{t-1},\theta}$, $f_{\varepsilon_t|\theta}$ and f_θ. In this section, I develop estimation and inference theory for such functionals. The approach I analyze relies on series approximation of quantile functions and builds on previous contributions to the estimation of nonlinear panel data models, such as Arellano and Bonhomme (2016) and Arellano et al. (2017).

Section 4.1 presents the estimator and Section 4.2 discusses a simulation-based algorithm that implements it. Large-sample properties of the estimator are given in Section 4.3 with more technical assumptions and proofs deferred to Supplemental Appendix B.

4.1 Estimation method

The researcher observes panel data $\{y_{it}\}_{1\leq i\leq N, 1\leq t\leq T}$ so that for each unit i the vector $y_i = (y_{i1}, ..., y_{iT})'$ is an i.i.d. draw from model (1)–(5). To simplify the exposition, I focus on a model with scalar θ_i, and time-invariant $f_{\eta_t|\eta_{t-1},\theta}$ and $f_{\varepsilon_t|\theta}$. Thus, under the conditions

[13] For the HIP model, the outline of the proof in Remark 1 applies if we redefine $Y_- = (y_{i1}, ..., y_{i,t-2})'$ and $Y = (y_{i,t-1}, y_{it})'$. For the "lots of heterogeneity" model, one obstacle is that nonparametric identification is not possible if there is heterogeneity in AR coefficients (see, e.g., Lee (2022)).

of Section 3, all the densities f_θ, $f_{\eta_1|\theta}$, $\{f_{\eta_t|\eta_{t-1},\theta}\}_{2\le t\le T}$ and $\{f_{\varepsilon_t|\theta}\}_{1\le t\le T}$ are nonparametrically identified if $T \ge 5$.[14]

To begin, I parameterize model (1)–(5) using quantiles. Let Q_1, Q_η, Q_ε and Q_θ be the quantile functions associated with $f_{\eta_1|\theta}$, $f_{\eta_t|\eta_{t-1},\theta}$, $f_{\varepsilon_t|\theta}$ and f_θ. Then,

$$\begin{aligned}\theta_i &= Q_\theta(\nu_i),\\ \eta_{i1} &= Q_1(\theta_i, u_{i1}),\\ \eta_{it} &= Q_\eta(\eta_{i,t-1}, \theta_i, u_{it}), \quad t = 2, \dots, T,\\ \varepsilon_{it} &= Q_\varepsilon(\theta_i, v_{it}), \quad t = 1, \dots, T,\end{aligned}$$

where ν_i, $\{u_{it}\}_{1\le t\le T}$, $\{v_{it}\}_{1\le t\le T}$ are cross-sectionally, mutually and (when applicable) serially independent Uniform$(0,1)$ random variables.

Fitted model. I assume the researcher specifies a flexible finite-dimensional linear approximation to the unknown functions Q_θ, Q_1, Q_η and Q_ε. For known vectors of basis functions $\varphi_1, \varphi_\eta, \varphi_\varepsilon$ of dimensions $K_1, K_\eta, K_\varepsilon$ and unknown parameter β,

$$\begin{aligned}Q_\theta(\nu;\beta) &= s(\nu, \beta_\theta),\\ Q_1(\theta, u;\beta) &= \varphi_1(\theta)' s(u, \beta_1),\\ Q_\eta(\eta, \theta, u;\beta) &= \varphi_\eta(\eta, \theta)' s(u, \beta_\eta),\\ Q_\varepsilon(\theta, v;\beta) &= \varphi_\varepsilon(\theta)' s(v, \beta_\varepsilon),\end{aligned}$$

where $s(\tau, b)$ is the column-wise linear spline interpolant of the $K \times L$-array b with nodes $\{\tau_\ell\}_{\ell=1}^L$ on $(0,1)$; i.e., for each ℓ, $s(\tau_\ell, b) = b_{\bullet\ell}$ (the ℓ-th column of b). The nodes are $\tau_\ell = \ell/(L+1)$ but this can be easily relaxed. Moreover, a special model for the tails is frequently used. For example, following Arellano et al. (2017),

$$\begin{aligned}Q_\theta(\nu;\beta) &= Q_\theta(\tau_1) + \beta_{\theta,\text{lo}} \ln(\nu/\tau_1), && \text{if } \nu < \tau_1,\\ Q_\theta(\nu;\beta) &= Q_\theta(\tau_L) - \beta_{\theta,\text{up}} \ln((1-\nu)/(1-\tau_L)), && \text{if } \nu > \tau_L,\end{aligned}$$

[14] To deal with a multivariate $\theta_i = (\theta_{i,1}, \dots, \theta_{i,D})'$ one can specify the conditional densities (or associated quantile functions) $\{f_{\theta_d|\theta_{d-1},\dots,\theta_1}\}_{1\le d\le D}$ sequentially. Additionally, one can accommodate some forms of time-variation in $f_{\eta_t|\eta_{t-1},\theta}$ and $f_{\varepsilon_t|\theta}$ by including time effects as covariates (see Section 4.2).

with a similar treatment of Q_1, Q_η and Q_ε. In sum, this is a parametric model with $KL + 8$ parameters β with $K = 1 + K_1 + K_\eta + K_\varepsilon$ the number of basis functions and 8 the number of tail coefficients. The basis functions are selected to provide good approximation to the unknown quantile functions in the spirit of series estimation (typical choices are orthogonal polynomials, splines and wavelets; see Chen (2007) for a complete discussion).

Let $\xi_i = (\theta_i, \eta_i')'$ with $\eta_i = (\eta_{i1}, ..., \eta_{iT})'$. The joint distribution of ξ_i and y_i is known up to β since

$$f_{\xi,y}(\xi_i, y_i; \beta) = f_\theta(\theta_i; \beta) f_{\eta_1|\theta}(\eta_{i1}|\theta_i; \beta) \prod_{t=2}^{T} f_{\eta_t|\eta_{t-1},\theta}(\eta_{it}|\eta_{i,t-1}, \theta_i; \beta) \prod_{t=1}^{T} f_{\varepsilon_t|\theta}(y_{it} - \eta_{it}|\theta_i; \beta)$$

and each density is the inverse of the derivative with respect to the rank argument of either Q_θ, Q_1, Q_η or Q_ε. In principle, by integrating ξ_i out, one could obtain the likelihood $f_y(y_i; \beta)$ and estimate β by maximizing $\sum_{i=1}^N \ln f_y(y_i; \beta)$ but the integration and maximization can be computationally challenging, particularly when K is large.

Moment conditions and estimator. Define for each τ,

$$\Psi_\tau(\xi_i, y_i, \beta) = \begin{pmatrix} \varrho_\tau'\left(\theta_i - s(\tau, \beta_\theta)\right) \\ \varphi_1(\theta_i) \cdot \varrho_\tau'\left(\eta_{i1} - \varphi_1(\theta_i)' s(\tau, \beta_1)\right) \\ \sum_{t=2}^T \varphi_\eta(\eta_{i,t-1}, \theta_i) \cdot \varrho_\tau'\left(\eta_{it} - \varphi_\eta(\eta_{i,t-1}, \theta_i)' s(\tau, \beta_\eta)\right) \\ \sum_{t=1}^T \varphi_\varepsilon(\theta_i) \cdot \varrho_\tau'\left(y_{it} - \eta_{it} - \varphi_\varepsilon(\theta_i)' s(\tau, \beta_\varepsilon)\right) \end{pmatrix} \tag{9}$$

with $\varrho_\tau'(x) = (\tau - \mathbb{1}\{x < 0\})$ the derivative of the so-called check function from linear quantile regression (Koenker and Bassett (1978)). Let $\Psi_{\text{tail}}(\xi_i, y_i, \beta)$ be the moment functions related to the tail parameters (e.g., $\mathbb{1}\{\theta_i < s(\tau_1, \beta_\theta)\}(s(\tau_1, \beta_\theta) - \theta_i - \beta_{\theta,\text{lo}})$ and $\mathbb{1}\{\theta_i > s(\tau_L, \beta_\theta)\}(\theta_i - s(\tau_L, \beta_\theta) - \beta_{\theta,\text{up}})$ for Q_θ and similar moments for Q_1, Q_η and Q_ε). Finally, stack $\{\Psi_{\tau_\ell}(\xi_i, y_i, \beta)\}_{\ell=1}^L$ and $\Psi_{\text{tail}}(\xi_i, y_i, \beta)$ into $\Psi(\xi_i, y_i, \beta)$.

The estimator $\hat{\beta}$ solves the integrated (sample) moment condition:

$$\sum_{i=1}^{N} \int \Psi(\xi_i, y_i, \hat{\beta}) f_{\xi|y}(\xi_i|y_i; \hat{\beta}) d\xi_i = 0_{(KL+8)\times 1}. \tag{10}$$

Direct calculation of $\hat{\beta}$ can be difficult but one can get a good approximation to it by means of a convenient Stochastic EM algorithm that involves simulating the latent variables ξ_i from their distribution conditional on y_i and solving (fairly simple) quantile regressions and moment equations. Section 4.2 presents the algorithm.

Objects of interest. Researchers are often interested in estimating $g(\beta)$ where g is a smooth functional of β. This applies to the measures of persistence and skewness defined in Section 2.2 and to the income risk decompositions in Section 2.3. The natural estimator for $g(\beta)$ is the plug-in estimator $g(\hat{\beta})$. The large-sample properties (for $N \to \infty$ and fixed T) of $g(\hat{\beta})$ are the topic of Section 4.3 below.

Covariates. Let x_i denote a set of exogenous covariates (e.g., age), which could be either time-invariant or time-varying. Within the quantile-based approach above, a natural way to include those covariates is through the basis functions. For example, for the quantile function for θ_i,

$$Q_\theta(v, x; \beta) = \varphi_\theta(x)' s(v, \beta_\theta),$$

where φ_θ is a known vector of K_θ basis functions. Similarly, the remaining basis functions become $\varphi_1(\theta, x)$, $\varphi_\eta(\eta, \theta, x)$ and $\varphi_\varepsilon(\theta, x)$.[15] The estimation method goes though by adapting the complete-data moments Ψ in (9) accordingly.

Estimation of ABB and HTR models. The estimation method described in this section generalizes the one developed in Arellano et al. (2017) for the ABB model by accommodating the heterogeneous quantity θ_i. It is also possible to deal with the HTR model (Section 2.1) in a similar way. Allowing for a non-unit autoregressive root in the dynamics of η_{it} and an intercept that depends on θ_i, the quantile function for η_{it} specializes to

$$Q_\eta(\eta, \theta, u) = \beta_{\eta,1}\eta + \varphi_\eta(\theta)'\beta_{\eta,2} + s(u, \beta_{\eta,3}), \tag{11}$$

for some basis functions φ_η. Here, $s(u, \beta_{\eta,3})$ captures the quantiles of the persistent-income shock v_{it} and it is normalized so that $E\left[v_{it} \middle| \eta_{i,t-1}, \theta_i\right] = 0$. Then, in (9), the complete-data

[15] Standard ways of generating multivariate basis functions is by tensor products of univariate basis functions (again, see Chen (2007) for further discussion).

moments for $\beta_{\eta,1}$, $\beta_{\eta,2}$ and $\beta_{\eta,3}$ should be replaced by

$$\begin{pmatrix} \sum_{t=2}^{T} \eta_{i,t-1} \cdot \left(\eta_{it} - \beta_{\eta,1}\eta_{i,t-1} - \varphi_\eta(\theta_{it})'\beta_{\eta,2}\right) \\ \sum_{t=2}^{T} \varphi_\eta(\theta_i) \cdot \left(\eta_{it} - \beta_{\eta,1}\eta_{i,t-1} - \varphi_\eta(\theta_i)'\beta_{\eta,2}\right) \\ \sum_{t=2}^{T} \varrho_\tau' \left(\eta_{it} - \beta_{\eta,1}\eta_{i,t-1} - \varphi_\eta(\theta_i)'\beta_{\eta,2} - s(\tau, \beta_{\eta,3})\right) \end{pmatrix}.$$

4.2 Stochastic EM algorithm

Computing $\hat{\beta}$ by solving the sample integrated moment conditions (10) can be difficult, but an iterative approach based on the EM algorithm of Dempster, Laird, and Rubin (1977) offers a convenient alternative. The approach extends Arellano and Bonhomme (2016), and relies on latent-variable simulation for the E-step (as in Nielsen (2000)) and solving moment equations instead of maximizing the likelihood of latent data for the M-step (as in Arcidiacono and Jones (2003)).

Algorithm 1 describes the implementation for the full model (1)–(5) of the paper. Specializing the algorithm to the ABB and HTR models, and including exogenous covariates x_i are straightforward variations.

Algorithm 1 (Stochastic EM). *Starting from initial values $\tilde{\beta}^{(0)}$ and $\{\tilde{\theta}_i^{(0)}\}_{1\le i\le N}$, for all $m = 1, \dots, M$, alternate between steps (E) and (M) below:*

(E) For each unit $i = 1, \dots, N$, draw the latent variables $\xi_i = (\theta_i, \eta_i')'$ for given observables y_i and parameters $\tilde{\beta}^{(m-1)}$ in two sub-steps:

(i) Draw $\tilde{\eta}_i^{(m)}$ by Sequential Monte Carlo (SMC) targeting the distribution

$$\tilde{\eta}_i^{(m)} \sim f_{\eta|y,\theta}(\eta|\tilde{\theta}^{(m-1)}, y_i; \tilde{\beta}^{(m-1)}).$$

(ii) Draw $\tilde{\theta}_i^{(m)}$ by Markov Chain Monte Carlo (MCMC) targeting the distribution

$$\tilde{\theta}_i^{(m)} \sim f_{\theta|\eta,y}(\theta|\tilde{\eta}_i^{(m)}, y_i; \tilde{\beta}^{(m-1)}).$$

(M) Update the parameters to $\tilde{\beta}^{(m)}$ by solving the moment equations

$$\sum_{i=1}^{N} \Psi(\tilde{\xi}_i^{(m)}, y_i, \tilde{\beta}^{(m)}) = 0_{(KL+8)\times 1}.$$

This is achieved by the following:

(i) *Quantile regressions of* θ_i *on an intercept to update* $\tilde{\beta}_{\theta}^{(m)}$.

(ii) *Quantile regressions of* η_{i1} *on* $\varphi_1(\theta_i)$ *to update* $\tilde{\beta}_1^{(m)}$.

(iii) *Quantile regressions of* η_{it} *on* $\varphi_\eta(\eta_{i,t-1}, \theta_i)$ *to update* $\tilde{\beta}_\eta^{(m)}$.

(iv) *Quantile regressions of* $y_{it} - \eta_{it}$ *on* $\varphi_\varepsilon(\theta_i)$ *to update* $\tilde{\beta}_\varepsilon^{(m)}$.

(v) *Sample averages to update* $\tilde{\beta}_{\theta,\text{lo}}^{(m)}, \tilde{\beta}_{\theta,\text{up}}^{(m)}, \tilde{\beta}_{1,\text{lo}}^{(m)}, \tilde{\beta}_{1,\text{up}}^{(m)}, \tilde{\beta}_{\eta,\text{lo}}^{(m)}, \tilde{\beta}_{\eta,\text{up}}^{(m)}, \tilde{\beta}_{\varepsilon,\text{lo}}^{(m)}, \tilde{\beta}_{\varepsilon,\text{up}}^{(m)}$.

Finally, for some $0 < \mu < 1$, *compute* $\tilde{\beta} = (\mu M)^{-1} \sum_{m=(1-\mu)M}^{M} \tilde{\beta}^{(m-1)}$.

Conditioning on θ_i, the model reduces to that of Arellano et al. (2017). For that reason, I can adopt for the SMC in step (E)(i) the particle filter technique proposed in Arellano et al. (2023) for that model.[16] On the other hand, for the MCMC in step (E)(ii) I use the slice sampling technique of Neal (2003). The idea behind sub-steps (E)(i) and (E)(ii) is to alternate between $f_{\eta|\theta,y}$ and $f_{\theta|\eta,y}$ in order to approximate the joint distribution of latent variables given data $f_{\theta,\eta|y}$. If the parameters were held fixed at β, this would essentially be a Gibbs sampler that has $f_{\theta,\eta|y}(\cdot,\cdot|y_i;\beta)$ as invariant distribution.

Because of the (M) step, Algorithm 1 makes $\left(\{\tilde{\xi}_i^{(m)}\}_{1\leq i\leq N}, \tilde{\beta}^{(m)}\right)$ (and not just $\tilde{\xi}_i^{(m)}$) a Markov chain. The convergence properties of this ensemble have been studied by Nielsen (2000) in a (parametric) likelihood context. With similar arguments, in the pseudo-likelihood setup of this paper, one can establish $\sqrt{M}(\tilde{\beta} - \hat{\beta}) = o_p(1)$ under general regularity conditions. From that point of view, Algorithm 1 is a convenient way of approximating the estimator $\hat{\beta}$ defined by (10).

4.3 Asymptotic properties

Under correct specification of the parametric model, the sampling properties of $\hat{\beta}$ as $N \to \infty$ are well known. If β_0 denotes the true value of β, under certain regularity conditions,

$$\sqrt{N}\left(g(\hat{\beta}) - g(\beta_0)\right) \xrightarrow{\text{d}} N(0, V_0),$$

for any smooth functional of interest g and some variance V_0. Moreover, confidence intervals based on a nonparametric bootstrap method that resamples units $\{y_i\}_{1\leq i\leq N}$ at random with replacement provides valid inference for $g(\beta_0)$.

[16]See Creal (2012) for a survey of Sequential Monte Carlo methods.

Supplemental Appendix B develops nonparametric asymptotics for $g(\hat{\beta})$ under the assumption of model misspecification of the form

$$
\begin{aligned}
Q_\theta(\nu) &= Q_\theta(\nu;\beta_\theta) + r_\theta(\nu), \\
Q_1(\theta,u) &= Q_1(\theta,u;\beta_1) + r_1(\theta,u), \\
Q_\eta(\eta,\theta,u) &= Q_\eta(\eta,\theta,u;\beta_\eta) + r_\eta(\eta,\theta,u), \\
Q_\varepsilon(\theta,v) &= Q_\varepsilon(\eta,\theta,u;\beta_\varepsilon) + r_\varepsilon(\theta,v),
\end{aligned}
$$

where $r_\theta(\nu)$, $r_1(\theta,u)$, $r_\eta(\eta,\theta,u)$ and $r_\varepsilon(\theta,u)$ vanish as $L, K_1, K_\eta, K_\varepsilon \to \infty$.[17]

5 Empirical analysis

The empirical exercise of this section has four main parts. First, I assess the ability of the full model (1)–(5) and of the ABB and HTR models to fit the nonlinearities found in income data (Section 5.1). Second, I use them to document state-dependence in persistent income risk and heterogeneity in transitory income risk (Section 5.2). Third, I analyze the implications of heterogeneity for the predictive distribution of future income (Section 5.3). And fourth, I quantify the consumption passthrough of transitory shocks (Section 5.4).

Background and data. The analysis relies on the Panel Study of Income Dynamics (PSID), a panel of US households collected by the University of Michigan. In 1968, the PSID interviewed a representative sample of US families, gathering data on their income and demographics. Since then, every year until 1997 and every other year thereafter, it has kept track of the same families and the families formed by their offspring. This strategy together with the periodic inclusion of refresher and immigrant samples aims at keeping the PSID representative of the US economy in every period, and allows me to build panels in which units face different economic regimes and aggregate conditions.[18][19]

I study various datasets. The first is the panel of households analyzed by Blundell,

[17] If η_i and θ_i were observable, this would be the setup studied by Belloni, Chernozhukov, Chetverikov, and Fernández-Val (2019).

[18] The PSID also collected data on a sample that over-represented low-income families known as the Survey of Economic Opportunities. This sample is not used in the empirical analysis below.

[19] The motivation to consider different periods is to assess the stability over time of the heterogeneity and state-dependence patterns I document in this section. Explicitly measuring the interaction between aggregate shocks and micro-level variation is beyond the scope of this paper; for work on that direction see Almuzara, Arellano, Blundell, and Bonhomme (2025).

Pistaferri, and Saporta-Eksten (2016) and Arellano et al. (2017). This is a panel for the years 1999-2009 (i.e., six waves) in which the head of the household is male and married, and income is the sum of pre-tax labor earnings of head and spouse, and transfers. The sample selection excludes units outside the 25-to-60 age range and those with zero income or earnings that are too low a fraction of transfers.[20] Thus, y_{it} is the residual from regressing log income on a set of demographic variables (education, family size, number of children, state of residence, etc.). Age is not used in the residualization but as a covariate within the model (x_{it} in the notation above).

In this section, I focus on results from the dataset just described, the goal being to ensure comparability with previous studies (in particular, Arellano et al. (2017)). However, Supplemental Appendix E reports estimates of my model obtained from alternative datasets where (a) I expand the sample selection, (b) I consider periods more and less recent than 1999-2009, and (c) I use the net log wage rate (i.e., earnings per hour of the head of the household net of the effect of demographics) as y_{it}.

Trends in volatility and measurement error. One advantage of the PSID relative to the administrative datasets available in the US is the ability to study households as opposed to individuals. This is important because the household is the relevant unit for analyzing consumption decisions, one of the goals of the income dynamics literature. Nevertheless, it is worth mentioning two caveats.

First, it is a matter of debate whether the PSID correctly captures trends in the volatility of earnings growth for male workers. Gottschalk and Moffitt (1994, 2009) find in the PSID a raising volatility trend during the 70s and 80s followed by a flat or perhaps mildly increasing trend thereafter. In contrast, Salgado, Bloom, Guvenen, Pistaferri, and Sabelhaus (2023) document a decreasing trend since the late 80s in data from the Social Security Administration. One interpretation of the divergence is as evidence of the PSID (despite its best efforts) losing representativity over time. How severe this is for households (as opposed to male workers) cannot be easily ascertained since social security data does not allow researchers to link individuals into family units. But from this point of view, the distribution of θ_i that I estimate below conditions on a specific subpopulation of US households determined by the attrition and refresher/immigration mechanisms of the PSID. How different is that subpopulation from the one in administrative data? A persuasive analysis by Moffitt, Abowd, Bollinger, Carr, Hokayem, McKinney, Wiemers,

[20] More information about the data construction can be found in Arellano et al. (2017, Appendix C).

Zhang, and Ziliak (2023) suggests that the trend divergence can be explained by the PSID missing a big part of the left tail of earnings. To the extent that individuals in the left tail also experience higher volatility than the rest, one should expect more heterogeneity in risk (transitory risk, in particular) in administrative data than what I find below.

Second, the presence of measurement error in the PSID raises the question of how it modifies the interpretation of the latent variable estimates from my model. To account for this, let me re-write the observation equation (1) as $y_{it} = \eta_{it} + \varepsilon_{it} + m_{it}$ where m_{it} denotes measurement error. Validation studies, such as Bound, Brown, Duncan, and Rodgers (1994) and Pischke (1995), find evidence that m_{it} has a stable variance over time and is weakly negatively correlated with true income $\eta_{it} + \varepsilon_{it}$. In fact, Pischke (1995) suggests that a reasonable model is (omitting a small persistent component) $m_{it} = \alpha\varepsilon_{it} + \zeta_{it}$ where $\alpha < 0$ and ζ_{it} is white noise, independent of $\eta_{it}, \varepsilon_{it}$. Adopting (4′), I can write $y_{it} = \eta_{it} + \tilde{\varepsilon}_{it}$ with $\tilde{\varepsilon}_{it} = (1-\alpha)\theta_i e_{it} + \zeta_{it}$. Suppose ζ_{it} is independent of θ_i and that I estimate the model ignoring measurement error with the normalization $\text{Var}\left(\tilde{\varepsilon}_{is} \middle| \tilde{\theta}_i = \theta\right) = \tilde{\theta}^2$ when the DGP satisfies $\text{Var}\left(\varepsilon_{is} \middle| \theta_i = \theta\right) = \theta^2$ (recall Assumption 1).[21] With obvious notation, I will recover the distribution of

$$\tilde{\theta}_i^2 = (1-\alpha)^2\theta_i^2 + \sigma_\zeta^2. \tag{12}$$

Let me focus on $\tilde{\theta}_i^2/E[\tilde{\theta}_i^2]$ (as in, e.g., Figure 4). If all measurement error came from underreporting transitory income ($\alpha < 0$, $\sigma_\zeta^2 = 0$), then $\tilde{\theta}_i^2/E[\tilde{\theta}_i^2] = \theta_i^2/E[\theta_i^2]$, that is, underreporting affects the scale but not the relative dispersion of transitory risks. More generally, if some measurement error came from the classical term ζ_{it} ($\sigma_\zeta^2 > 0$), $\tilde{\theta}_i^2/E[\tilde{\theta}_i^2] = a + b\theta_i^2/E[\theta_i^2]$ with $b < 1$ and I would tend to understate the dispersion of transitory risk. All of this suggests that the heterogeneity in θ_i I document below is most likely a lower bound on the true extent of heterogeneity in transitory risk.[22]

Implementation. I estimate model (1)–(5), as well as the ABB and the HTR model using Algorithm 1. In all cases, I perform $M = 1,000$ iterations and average the last $\mu = 20\%$ of them. I then run the nonparametric bootstrap approach mentioned in Section 4.3 with 200 bootstrap replicas for each model.

Let x_{it} be the household head's age. For the full model (1)–(5), I specify Q_1 as a second-

[21] Note that estimates of objects related to η_{it} are unaffected under this model of measurement error.

[22] For the 1982-1986 period, Pischke (1995) estimates $\sigma_\zeta^2 \approx E[\theta_i^2]$ and $\alpha = -0.25$ (with a big standard error). It implies $\tilde{\theta}_i^2/E[\tilde{\theta}_i^2] \approx 0.72 + 0.36 \times \theta_i^2/E[\theta_i^2]$, which can be used to re-calibrate my results.

order Hermite polynomial in $\ln(\theta_i)$ and x_{i1} (without interactions) and Q_η as a third-order Hermite polynomial in $\eta_{i,t-1}$ and a second-order Hermite polynomial in $\ln(\theta_i)$ and x_{it} (with interactions between $\eta_{i,t-1}$ and $\ln(\theta_i)$ and between $\eta_{i,t-1}$ and x_{it}). I also adopt the representation (4′), specifying the quantile function Q_e of e_{it} as a second-order Hermite polynomial in x_{it}, taking Q_θ to be independent of x_{it}.

For the ABB model, Q_1, Q_η and Q_ε are specified in the same way as above except that I omit the terms involving θ_i (and so, Q_ε reduces to Q_e). Finally, for the HTR model, Q_η is linear in $\eta_{i,t-1}$ and a second-order polynomial in $\ln(\theta_i)$ and x_{it} (with interactions between $\eta_{i,t-1}$ and x_{it}) but with rank-independent coefficients as in (11).

5.1 Nonlinearities in income data

To begin, I resume the discussion of Section 2.2. How well can models with different amounts of state-dependence and heterogeneity fit the nonlinearities often found in income data? Figures 2 and 3 compare estimates of dispersion, skewness and persistence obtained directly from PSID with those implied by each model.

Dispersion is measured by the difference between quantiles τ_0 and $(1-\tau_0)$ while skewness and persistence are measured by the reduced-form quantities $\tilde{\text{sk}}_t$ and $\tilde{\rho}_t$ described in Section 2.2. Following Arellano et al. (2017), I set $\tau_0 = 11/12$ and map $y_{i,t-1}$ to its unconditional quantile τ_{init}. In all figures, age x_{it} is averaged out.

Three salient patterns are to be noted: (i) a U-shaped conditional dispersion curve implying higher variance for low- and high-income households, (ii) a decreasing conditional skewness curve implying more upside (downside) risk for low-income (high-income) households, and (iii) lower persistence for low-income households with good shocks and high-income households with bad shocks. All three models capture them. If anything, the HTR model fits the dispersion a bit better than ABB and the opposite holds for skewness, but the differences are small.

I do not pursue a nonparametric test of the restrictions imposed by the ABB and HTR models on the full model (1)–(5), but Figures 2 and 3 make clear that it would be difficult to reject either on the basis of patterns (i), (ii) and (iii), at least in the PSID data. What follows next, however, suggests that both models should be rejected as estimates of the full model show strong evidence of heterogeneity in θ_i (which goes against ABB) and of nonlinear persistence in η_{it} (against the HTR setup).

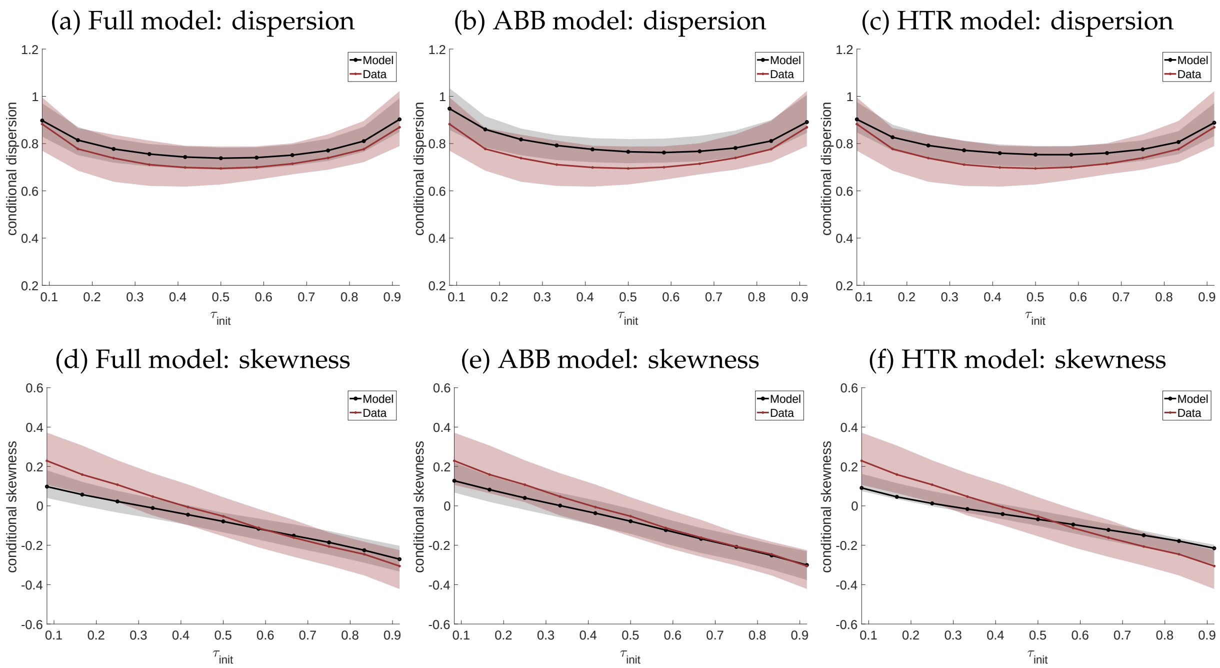

FIGURE 2. Conditional dispersion and skewness in y_{it}.

Note: Panels (a), (b) and (c) report measures of conditional dispersion while panels (d), (e) and (f) report measures of conditional skewness. Red lines indicate estimates from PSID data using the same quantile autoregression approach of Figure 3. Black lines indicate estimates from simulated data for each model. Shaded areas correspond to 95% confidence sets pointwise in τ_{init}.

5.2 Heterogeneity in transitory risk and nonlinear persistence

I now turn to quantifying the extent of transitory risk heterogeneity and nonlinear persistence. Figure 4 deals with the former, showing that the HTR model recovers sizable cross-sectional differences in θ_i. Reading from the point estimates, almost 70% of household units have less than half the average transitory income variance (normalized to one in the figure), whereas 20% of them have twice the average variance or more. This points to a dimension of inequality in income risk that is (a) first-order and (b) missed by the vast majority of models in the literature.

Interestingly, the full model estimates, seen in panel (a), imply virtually the same amount of transitory risk heterogeneity. Or put another way, allowing for flexible state-dependence does not attenuate the evidence of inequality in transitory risk. This is one of the main takeaways from my empirical analysis.

Next, figure 5 examines the converse: does allowing for heterogeneity attenuate the evidence of nonlinearities in the persistent income component η_{it}?

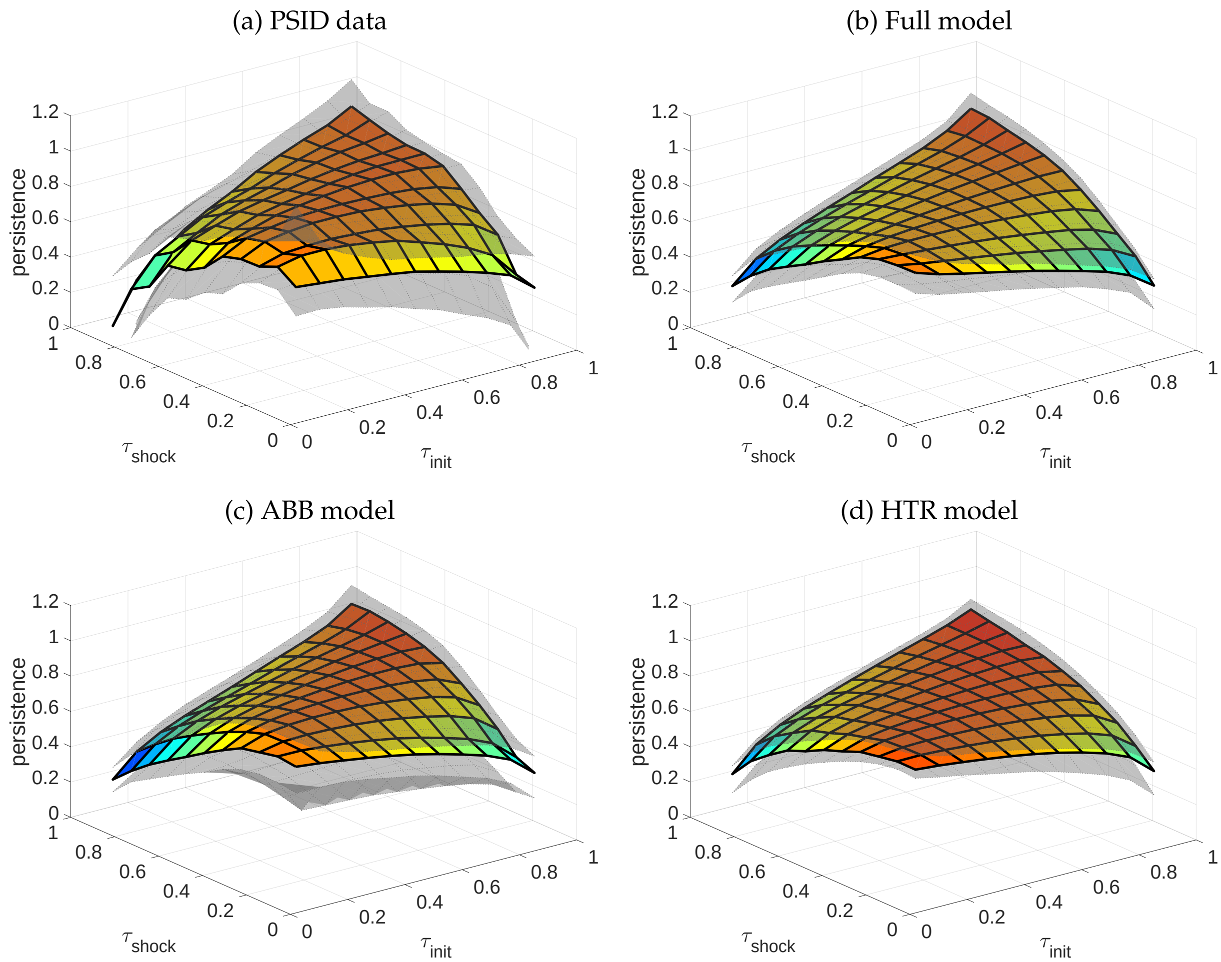

FIGURE 3. Nonlinear persistence in y_{it}.

Note: All panels report the measure $\tilde{\rho}_t$ of nonlinear persistence defined in Section 2.2. Panel (a) reports estimates from PSID data using a quantile autoregression of y_{it} on a third-order Hermite polynomial in $y_{i,t-1}$ and a second-order polynomial in x_{it} (age). Panels (b), (c) and (d) apply the same procedure to simulated data from the full model, the ABB model and the HTR model, respectively. Shaded areas correspond to 95% confidence sets pointwise in τ_{shock} and τ_{init}.

Comparing the ABB and the full model, it appears that the U-shaped conditional dispersion, the decreasing conditional skewness and the persistence surface are broadly consistent. There is at most a slight reduction in the slope of the skewness curve, albeit small compared to the estimation uncertainty. Since by construction the HTR model cannot reproduce any of these patterns, one has to conclude neither the ABB nor the HTR model offers a complete description of income dynamics.

Lastly, what about the interplay between θ_i and η_{it}? Panels (a) and (d) in Figure 5

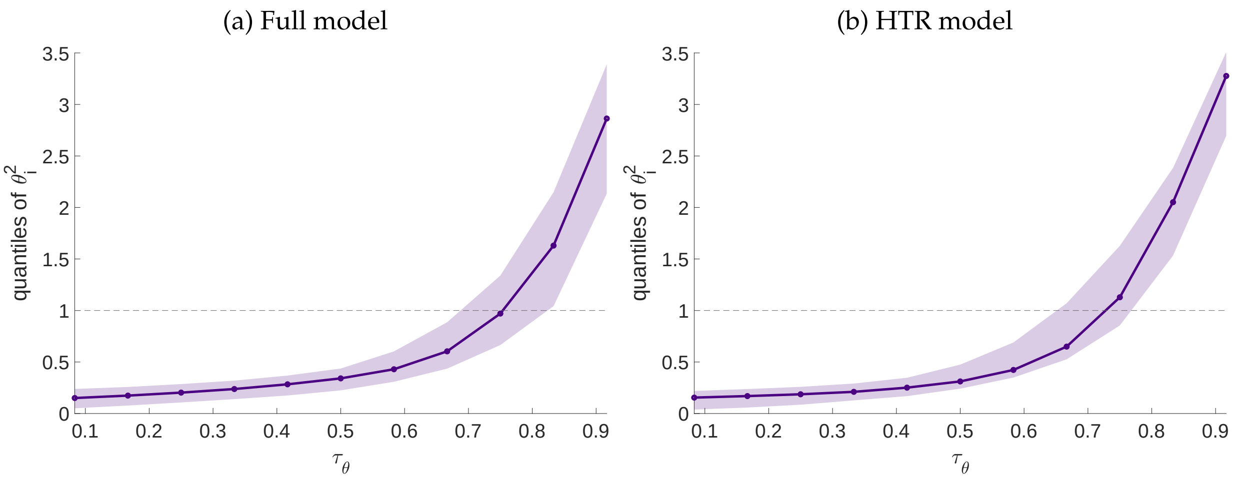

FIGURE 4. Heterogeneity in transitory risk.
Note: Panels (a) and (b) report estimates of the quantiles of $\theta_i^2/E[\theta_i^2]$ from the full model and the HTR model respectively. Shaded areas correspond to 95% confidence sets pointwise in the θ_i-rank τ_θ.

document an interesting association between transitory risk and persistent risk: units with lower θ_i tend to display lower conditional dispersion and more positive conditional skewness in η_{it} compared to units with higher θ_i. In words, transitory and persistent income risks (in the form of larger, more negatively skewed shocks) go hand in hand.

5.3 Predictive distributions of future income

Let us now return to the discussion of Section 2.3 with Figures 6 and 7.

The preceding analysis supports a two-dimensional model of income inequality, where households are persistently unequal in both income level (captured by η_{it}) and income variability (captured by θ_i). But how unequal are income risks? And which dimension matters most? To get a sense, I report moments from the predictive distribution of $y_{i,t+h}$ conditioning on η_{it} and θ_i for $h = 1$ (i.e., 2 years ahead) and $h = 5$ (10 years ahead); see Section 2.3 for their interpretation.

I obtain the estimates by simulation. I first compute the τ_θ-quantile of θ_i (given by Q_θ) and the τ_η-quantile of η_{it} conditional on θ_i (given by Q_1). I then simulate 50,000 paths for $\{\eta_{i,t+\ell}, \varepsilon_{i,t+\ell}\}_{\ell=1}^{h}$ which I use to approximate $\text{Var}\left(y_{i,t+h} \middle| \eta_{it}, \theta_i\right)$ and $\text{Skew}\left(y_{i,t+h} \middle| \eta_{it}, \theta_i\right)$. These are displayed in Figures 6 and 7 as a function of (τ_θ, τ_η).

The main takeaway is this: There are large, persistent differences in income risk across households in the PSID, and most of those differences are explained by θ_i.

Consider first the short horizon $h = 1$. For the median η_{it}, going from $\tau_\theta = 1/12$ to

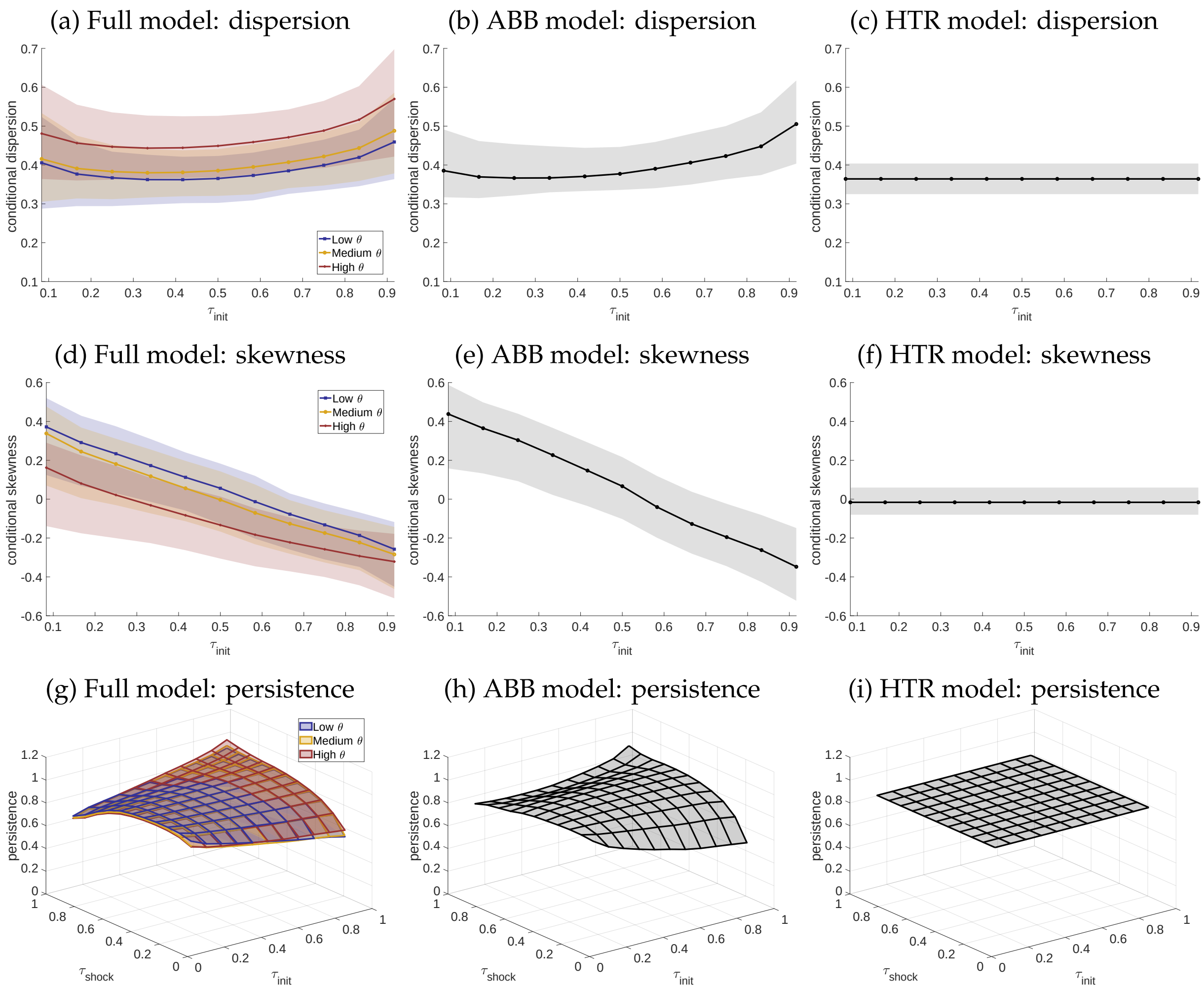

FIGURE 5. Nonlinearities in the distribution of η_{it} given $\eta_{i,t-1}$ and θ_i.
Note: Panels (a), (b) and (c) show conditional dispersion measures. Panels (d), (e) and (f) show conditional skewness measures. Finally, panels (g), (h) and (i) display measures of persistence. When applicable, blue, yellow and red lines indicate a θ-value in the 1/12, 1/2 and 11/12 quantiles. Shaded areas in panels (a)-to-(f) correspond to 95% confidence sets pointwise in the $\eta_{i,t-1}$-rank τ_{init}.

$\tau_\theta = 11/12$ substantially magnifies the predictive variance (from 0.04 to 0.27) and skewness (from -0.01 to -0.17). Instead, for the median θ_i, $\tau_\eta = 1/12$ and $\tau_\eta = 11/12$ imply only minor differences in predictive variance (roughly constant at 0.06) and skewness (from -0.01 to -0.02). The pattern for the variance is not entirely surprising given (i) the sizable heterogeneity in θ_i (see Figure 4) and (ii) the limited impact of η_{it} on the variance of $\eta_{i,t+1}$ (see Figure 5(a)).

In contrast, η_{it} does affect the conditional skewness of $\eta_{i,t+1}$ quite a lot (see Figure 5(d)). So why does θ_i still dominate in Figure 7(a)? This is simply a matter of scale: at short

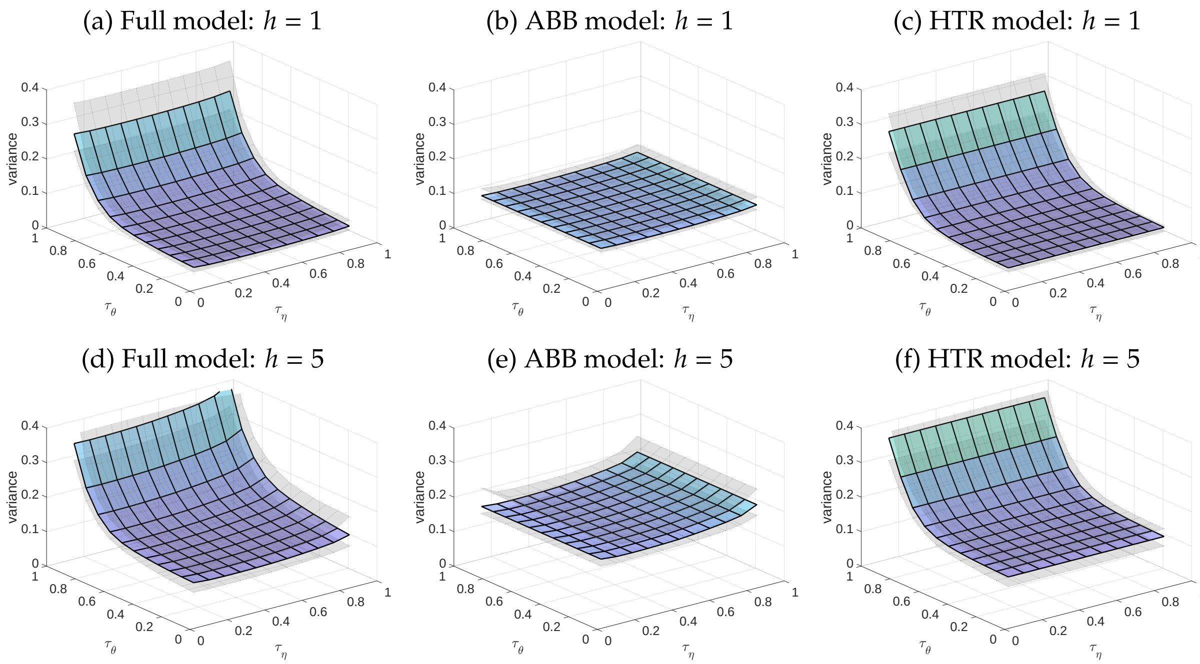

FIGURE 6. Variance of predictive distribution of $y_{i,t+h}$ given η_{it} and θ_i.

Note: Shaded areas are 95% confidence sets pointwise in the η_{it}-rank τ_η and the θ_i-rank τ_θ.

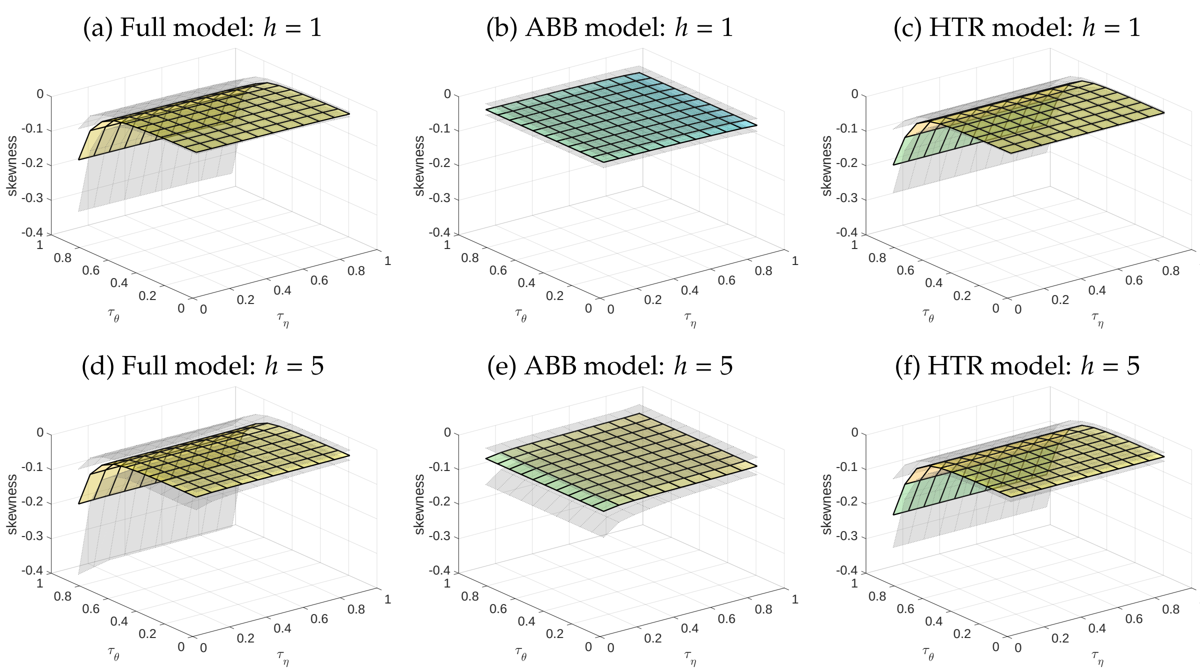

FIGURE 7. Skewness of predictive distribution of $y_{i,t+h}$ given η_{it} and θ_i.

Note: Shaded areas are 95% confidence sets pointwise in the η_{it}-rank τ_η and the θ_i-rank τ_θ.

horizons, most income changes are driven by transitory shocks and, as a consequence, most of the skewness in $y_{i,t+h}$ reflects skewness in $\varepsilon_{i,t+h}$, not in $\eta_{i,t+h}$. At horizon $h = 5$ there is a slightly more visible gradient with respect to τ_η, but the figures continue to indicate a dominant role for θ_i.

The last thing to highlight is that the HTR model comes much closer than ABB to reproducing the large disparities in income risk implied by the full model. Again, this is because most income risk is transitory income risk, most transitory risk is heterogeneous, and the HTR model recovers a big part of that heterogeneity.

5.4 Consumption passthrough of income shocks

Since 1999, PSID has collected rich data on household-level spending and wealth, thus allowing a complete empirical analysis of consumption responses to income shocks. I am particularly interested in measures of the transmission of transitory shocks ε_{it} that account for the role of heterogeneity θ_i.

For context, the permanent income hypothesis predicts no impact on consumption for transitory income shocks. Blundell et al. (2008) report evidence consistent with this, but quasi-experimental estimates tend to indicate a non-zero (often large) response.[23] Quantifying these impacts empirically is relevant to understanding the extent of insurance available to households and the welfare consequences of various policies and shocks.

For this exercise, I adopt (4′) which decomposes transitory shocks into a surprise component e_{it} and the unit-specific transitory risk exposure θ_i, which is presumably (better) known to the household. If e_{it} were observable, a natural approach would involve local projections (Jordà, 2005) of consumption log-changes Δc_{it}:

$$\Delta c_{i,t+h} = m_i(h) + \psi(h)e_{it} + \zeta_{it}(h), \qquad h = 0, 1, \ldots \tag{13}$$

where the error $\zeta_{it}(h)$ contains variation in consumption orthogonal to e_{it}. Under the random effects assumption $\text{Cov}(m_i(h), e_{it}) = 0$ (which is plausible if e_{it} is a shock), equation

[23] See, e.g., Souleles (1999), Parker et al. (2013) and Misra and Surico (2014) for estimates exploiting tax rebates and Fagereng, Holm, and Natvik (2021) for lottery wins. A review of the literature and a potential explanation of the disagreement between semi-structural and quasi-experimental approaches can be found in Commault (2022).

(13) could be estimated by OLS since then

$$\psi(h) = \frac{\text{Cov}\Big(\Delta c_{i,t+h}, e_{it}\Big)}{\text{Var}\,(e_{it})}.$$

Instead, when e_{it} is a latent variable, one must ask if $\psi(h)$ is identified from the joint distribution of total income and consumption. To answer this question, let $y_i = (y_{i1}, \ldots, y_{iT})'$ and $c_i = (\Delta c_{i2}, \ldots, \Delta c_{iT})'$. For each $\bar{c}$ outside a zero-probability set, suppose $f_{y|c}(y_i|\bar{c})$ satisfies Assumptions 1, 2 and 3. Proposition 1 then implies that the joint distribution of $(e_{it}, \Delta c_{i,t+h})$ is identified for $h < T$ which, in turn, delivers identification of $\psi(h)$. Moreover, valid estimates can be obtained by appending

$$E\Big[e_{it} \cdot \Big(\Delta c_{i,t+h} - \psi(h) e_{it}\Big)\Big] = 0$$

to the complete-data moment conditions (9) with straightforward modifications to the Stochastic EM algorithm and nonparametric bootstrap procedure.

Table 1 reports estimates of cumulative passthrough coefficients $\sum_{\ell=0}^{h} \psi(h)$ from the full model (1)–(5). The consumption data is the same as in Arellano et al. (2017) and covers about two-thirds of total spending on nondurables and services.[24]

TABLE 1. Estimates of cumulative passthrough coefficients.

	$h = 0$	$h = 1$	$h = 2$	$h = 3$	$h = 4$
$\boldsymbol{\sum_{\ell=0}^{h} \psi(h)}$	0.065	-0.008	0.026	0.011	-0.010
95% CI	[0.026, 0.101]	[-0.041, 0.032]	[-0.009, 0.067]	[-0.025, 0.052]	[-0.092, 0.060]

The main finding is a significantly positive contemporaneous response of spending to transitory shocks followed by no significant passthrough at longer horizons.

How do my estimates compare to the rest of the literature? The semi-structural robust estimator of Commault (2022) applied to the same dataset produces a point estimate of 0.023 and a 95% confidence interval of [-0.052, 0.098] for $\psi(0)$.[25] Using a longer biennial

[24]Supplemental Appendix E presents estimates using a more recent panel constructed from the PSID that contains more comprehensive spending data.

[25]This estimator relies on assuming that at the annual frequency the persistent component is a random walk whereas the transitory component is an MA(1) process. As a result, $\psi(0)$ is the estimand of an IV regression of Δc_{it} on Δy_{it} using $\Delta y_{i,t+1}$ as an instrument where the time unit t is biennial. In this case, the estimator proposed by Commault (2022) coincides with the one used in Blundell et al. (2008).

panel for the years 1999-2017, and omitting her adjustment for measurement error, she obtains an estimate of 0.063 with a 95% confidence interval of [0.002, 0.124].[26] In either case, I recover stronger passthrough than semi-structural approaches.

Comparison with quasi-experimental estimates is more difficult. One issue is measurement error which introduces attenuation bias by inflating the variance of transitory shocks. For example, the measurement error model (12) calibrated to the estimates of Pischke (1995) implies that my estimate of $\psi(0)$ underestimates true passthrough by a factor of 2.08. Another issue is the biennial frequency: if annual transitory shocks have negative effects on next-year consumption growth, the biennial coefficient will tend to be much lower than the annual one. Commault (2022), for example, finds biennial coefficients to be around half their annual counterparts. All things considered, my results are not incompatible with the high annual and quarterly estimates of quasi-experimental studies, but more research is needed to achieve a quantitatively precise answer to this question.

6 Conclusion

I find robust evidence of a large amount of cross-sectional inequality in exposure to transitory income risks among US households. This inequality is well captured by permanent heterogeneity in the scale of transitory shocks that cannot be easily accounted for by observables, such as age, and implies much larger differences in total income risk than nonlinearities in the persistent income process.

All this naturally raises the question of how it affects consumption and welfare. I estimate a significantly positive (but short-lived) average passthrough of transitory shocks to consumption which suggests that part of the inequality in transitory risk should translate into inequality in consumption variability. But it is plausible that passthrough itself is highly heterogeneous across households. Work along the lines of Lewis, Melcangi, and Pilossoph (2024) appears promising. Adding heterogeneity to the semi-structural consumption framework of Arellano et al. (2017) and to fully-specified structural models can provide insights into the implications of transitory risk for wealth accumulation, a feature that most models fit poorly.

From an econometric point of view, the modeling approach, identification techniques

[26] The measurement error adjustment involves dividing the IV estimand of the previous footnote by the fraction of the transitory biennial variance explained by measurement error which she calibrates to 0.5 leading her to a point estimate of $0.063/0.5 \approx 0.125$; see Commault (2022, Table 4).

and inference theory developed in this paper are potentially relevant for a much wider class of problems and applications. It covers, for example, various flexible nonlinear models with latent variables. Unfortunately, though, the asymptotic approximations are not particularly transparent about the bias-variance tradeoff associated to the choice of tuning parameters $L, K_1, K_\eta, K_\varepsilon$. More research is needed on this important econometric question.

A Proofs

Proof of Proposition 1. With the notation of Remark 1, I begin by showing that

$$f_{Y,Y_+|Y_-}(y, y_+|y_-) = \int_{\mathcal{Z}} f_{Z|Y_-}(z|y_-) f_{Y|Z}(y|z) f_{Y_+|Z}(y_+|z)\, dz, \tag{14}$$

for all $y_- \in \mathcal{Y}_-$, $y \in \mathcal{Y}$ and $y_+ \in \mathcal{Y}_+$, admits a unique solution $\{f_{Z|Y_-}, f_{Y|Z}, f_{Y_+|Z}\}$. Let $\mathcal{L}^1(\mathcal{D})$ be the set of integrable real-valued functions with domain $\mathcal{D}$. Fix $y \in \mathcal{Y}$ and define the linear operators $L_{y,Y_+|Y_-} : \mathcal{L}^1(\mathcal{Y}_-) \to \mathcal{L}^1(\mathcal{Y}_+)$, $L_{Y_+|Y_-} : \mathcal{L}^1(\mathcal{Y}_-) \to \mathcal{L}^1(\mathcal{Y}_+)$, $L_{Z|Y_-} : \mathcal{L}^1(\mathcal{Y}_-) \to \mathcal{L}^1(\mathcal{Z})$, $\Delta_{y|Z} : \mathcal{L}^1(\mathcal{Z}) \to \mathcal{L}^1(\mathcal{Z})$ and $L_{Y_+|Z} : \mathcal{L}^1(\mathcal{Z}) \to \mathcal{L}^1(\mathcal{Y}_+)$ by

$$\begin{aligned}
[L_{y,Y_+|Y_-} g_1](y_+) &= \int_{\mathcal{Y}_-} f_{Y,Y_+|Y_-}(y, y_+|y_-) g_1(y_-)\, dy_-, \\
[L_{Y_+|Y_-} g_1](y_+) &= \int_{\mathcal{Y}_-} f_{Y_+|Y_-}(y_+|y_-) g_1(y_-)\, dy_-, \\
[L_{Z|Y_-} g_1](z) &= \int_{\mathcal{Y}_-} f_{Z|Y_-}(z|y_-) g_1(y_-)\, dy_-, \\
[\Delta_{y|Z} g_2](z) &= f_{Y|Z}(y|z) g_2(z), \\
[L_{Y_+|Z} g_2](y_+) &= \int_{\mathcal{Z}} f_{Y_+|Z}(y_+|z) g_2(z)\, dz,
\end{aligned}$$

for all $g_1 \in \mathcal{L}^1(\mathcal{Y}_-)$ and $g_2 \in \mathcal{L}^1(\mathcal{Z})$. From (14) follows the spectral decomposition

$$L_{y,Y_+|Y_-} L^{-1}_{Y_+|Y_-} = L_{Y_+|Z} \Delta_{y|Z} L^{-1}_{Y_+|Z},$$

with the inverses guaranteed to exist by Assumption 3. The operator $L_{y,Y_+|Y_-} L^{-1}_{Y_+|Y_-}$ is known because it is a function of known densities. Therefore, $f_{Y_+|Z}$ and $f_{Y|Z}$ are identified if $L_{Y_+|Z}$ and $\Delta_{y|Z}$ are the unique eigenfunction-eigenvalue decomposition of $L_{y,Y_+|Y_-} L^{-1}_{Y_+|Y_-}$

satisfying Assumption 1.

As in the proof of Hu and Schennach (2008, Theorem 1), note the following:

(i) The scale of eigenfunctions is pinned down by $\int_{\mathcal{Y}_+} f_{Y_+|Z}(y_+|z)\, dy_+ = 1$.

(ii) The intersection over $y \in \mathcal{Y}$ of the linear span of $\left\{f_{Y_+|Z}(\cdot|\tilde{z}) : f_{Y|Z}(y|z) = f_{Y|Z}(y|\tilde{z})\right\}$ is one dimensional. This rules out ambiguities in selecting eigenfunctions for repeated eigenvalues and it is implied by the fact that

$$\begin{aligned}\mathcal{A}(z, \tilde{z}) &\equiv \left\{y \in \mathcal{Y} : f_{Y|Z}(y|z) \neq f_{Y|Z}(y|\tilde{z})\right\} \\ &= \left\{y = \mathcal{Y} : f_{\varepsilon_t|\theta}(y - \eta_{it}|\theta_i) \neq f_{\varepsilon_t|\theta}(y - \tilde{\eta}_{it}|\tilde{\theta}_i)\right\}\end{aligned}$$

has positive probability for any $z = (\eta_{it}, \theta_i) \neq \tilde{z} = (\tilde{\eta}_{it}, \tilde{\theta}_i)$ under Assumption 1. To see this, take $\eta_{it} \neq \tilde{\eta}_{it}$. If $P\left(\mathcal{A}(z, \tilde{z})\right) = 0$, then

$$E\left[y_{it}\middle|\eta_{it}, \theta_i\right] = E\left[y_{it}\middle|\tilde{\eta}_{it}, \tilde{\theta}_i\right].$$

But by Assumption 1 this implies $\eta_{it} = \tilde{\eta}_{it}$, a contradiction. Next, take $\eta_{it} = \tilde{\eta}_{it}$ and $\theta_i \neq \tilde{\theta}_i$. If $P\left(\mathcal{A}(z, \tilde{z})\right) = 0$, then

$$M\left[f_{\varepsilon_t|\theta}(\cdot|\theta_i)\right] = M\left[f_{\varepsilon_t|\theta}(\cdot|\tilde{\theta}_i)\right].$$

But, again, by Assumption 1 this implies $\theta_i = \tilde{\theta}_i$, a contradiction. Therefore, $P\left(\mathcal{A}(z, \tilde{z})\right) > 0$ as claimed.

(iii) The order and indexing of eigenvalues is also determined by Assumption 1. Suppose $\tilde{Z} = (\tilde{\eta}_{it}, \tilde{\theta}_i)$ is another re-indexing of $Z = (\eta_{it}, \theta_i)' = (m_{\eta_t}(\tilde{Z}), m_\theta(\tilde{Z}))$ that satisfies the spectral decomposition and Assumption 1, then

$$\eta_{it} = E\left[y_{it}\middle|\eta_{it}, \theta_i\right] = E\left[y_{it}\middle|m_{\eta_t}(\tilde{Z}), m_\theta(\tilde{Z})\right] = m_{\eta_t}(\tilde{Z})$$

and

$$\theta_i = M\left[f_{\varepsilon_t|\theta}(\cdot|\theta_i)\right] = M\left[f_{\varepsilon_t|\tilde{\theta}}(\cdot|m_\theta(\tilde{Z}))\right] = m_\theta(\tilde{Z})$$

which means $m_{\eta_t}(\cdot)$ and $m_\theta(\cdot)$ are the identity mappings.

In sum, there is a unique pair $f_{Y_+|Z}$, $f_{Y|Z}$ (up to sets of measure zero) compatible with

(14). The density $f_{Z|Y_-}$ is then recovered from $L_{Z|Y_-} = L^{-1}_{Y_+|Z} L_{Y_+|Y_-}$. Integrating y_- delivers f_Z, Bayes rule delivers $f_{Y_-|Z}$ and conditional independence delivers the joint conditional density $f_{Y_-,Y,Y_+|Z}$. Also identification of f_Z implies f_θ is identified.

Knowledge of $f_{Y_-,Y,Y_+|Z}$ implies knowledge of $f_{y_1,\ldots,y_T|\theta}(\cdot|\theta)$ for any $\theta \in \Theta$, which is the density of the income process of Arellano et al. (2017) for a fixed θ. The next step is to solve, for each $t = 2, \ldots, T-1$ and $c \in \Theta$, the integral equation

$$f_{y_t,y_{t+1}|y_{t-1},\theta}(y, y_+|y_-, c) = \int_{\mathcal{H}_t} f_{\eta_t|y_{t-1},\theta}(h|y_-, c) f_{y_t|\eta_t,\theta}(y|h, c) f_{y_{t+1}|\eta_t,\theta}(y_+|h, c)\, dh$$

for all $y_- \in \mathcal{Y}_{t-1}$, $y \in \mathcal{Y}_t$ and $y_+ \in \mathcal{Y}_{t+1}$ where $\mathcal{H}_t$, $\mathcal{Y}_{t-1}$, $\mathcal{Y}_t$ and $\mathcal{Y}_{t+1}$ are the supports of η_{it}, $y_{i,t-1}$, y_{it} and $y_{i,t+1}$. The argument is identical to the one given for (14) and I therefore omit it but I note that Assumption 2 and Assumption 3 imply the required boundedness and completeness conditions while $E\left[y_{it}\middle|\eta_{it}, \theta_i = c\right] = \eta_{it}$ now plays the role of the normalization. This delivers identification of $f_{y_t|\eta_t,\theta}$ and, therefore, of $f_{\varepsilon_t|\theta}$ for each $t = 2, \ldots, T-1$.

By serial independence of ε_{it}, $f_{\varepsilon_2,\ldots,\varepsilon_{T-1}|\theta}$ is identified. Since $f_{y_2,\ldots,y_{T-1}|\theta}$ is known, by a deconvolution argument (possible under Assumption 3), $f_{\eta_2,\ldots,\eta_{T-1}|\theta}$ is identified and, in consequence, identification of $f_{\eta_t|\eta_{t-1},\theta}$ for $t = 3, \ldots, T-1$ follows. □

References

Abowd, J. M. and D. Card (1989): "On the covariance structure of earnings and hours changes." *Econometrica*, 57, 411–445.

Ai, C. and X. Chen (2003): "Efficient estimation of models with conditional moment restrictions containing unknown functions." *Econometrica*, 71, 1795–1843.

Alan, S., M. Browning, and M. Ejrnæs (2018): "Income and consumption: A micro semi-structural analysis with pervasive heterogeneity." *Journal of Political Economy*, 126, 1827–1864.

Almuzara, M., M. Arellano, R. Blundell, and S. Bonhomme (2025): "Nonlinear Micro Income Processes with Macro Shocks." Working paper.

Altonji, J. G., A. A. Smith, and I. Vindangos (2013): "Modeling earnings dynamics." *Econometrica*, 81, 1395–1454.

Arcidiacono, P. and J. B. Jones (2003): "Finite mixture distributions, sequential likelihood and the EM algorithm." *Econometrica*, 71, 933–946.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and consumption dynamics: A nonlinear panel data framework." *Econometrica*, 95, 693–734.

Arellano, M., R. Blundell, S. Bonhomme, and J. Light (2023): "Heterogeneity of Consumption Responses to Income Shocks in the Presence of Nonlinear Persistence," *Journal of Econometrics*, 1–45.

Arellano, M. and S. Bonhomme (2012): "Identifying distributional characteristics in random coefficients panel data models." *Review of Economic Studies*, 79, 987–1020.

——— (2016): "Nonlinear panel data estimation via quantile regressions." *Econometrics Journal*, 19, C61–C94.

Arellano, M., S. Bonhomme, M. De Vera, L. Hospido, and S. Wei (2022): "Income Risk Inequality: Evidence from Spanish Administrative Records." *Quantitative Economics*, 13, 1747–1801.

Belloni, A., V. Chernozhukov, D. Chetverikov, and I. Fernández-Val (2019): "Conditional quantile processes based on series or many regressors." *Journal of Econometrics*, 213, 4–29.

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-nonparametric IV estimation of shape-invariant engel curves." *Econometrica*, 75, 1613–1669.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption inequality and partial insurance." *American Economic Review*, 98, 1887–1921.

Blundell, R., L. Pistaferri, and I. Saporta-Eksten (2016): "Consumption inequality and family labor supply." *American Economic Review*, 106, 387–435.

Blundell, R. and I. Preston (1998): "Consumption inequality and income uncertainty." *Quarterly Journal of Economics*, 113, 603–640.

Bonhomme, S. and J. M. Robin (2010): "Generalized non-parametric deconvolution with an application to earnings dynamics." *Review of Economic Studies*, 77, 491–533.

Botosaru, I. (2023): "Time-Varying Unobserved Heterogeneity in Earnings Shocks." 235, 1378–1393.

Botosaru, I. and Y. Sasaki (2018): "Nonparametric heteroskedasticity in persistent panel processes: An application to earnings dynamics." *Journal of Econometrics*, 203, 283–296.

Bound, J., C. Brown, G. J. Duncan, and W. L. Rodgers (1994): "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data." *Journal of Labor Economics*, 12, 345–368.

Browning, M., M. Ejrnæs, and J. Álvarez (2010): "Modelling income processes with lots of heterogeneity." *Review of Economic Studies*, 77, 1353–1381.

Carrasco, M., J. P. Florens, and E. Renault (2007): "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization." in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, Elsevier, vol. 6B, chap. 77.

Chamberlain, G. and K. Hirano (1999): "Predictive distributions based on longitudinal earnings data." *Annales d'Économie et de Statistique*, 55/56, 211–242.

Chen, X. (2007): "Large sample sieve estimation of semi-nonparametric models." in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, Elsevier, vol. 6B, chap. 76.

Chen, X. and X. Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data." *Econometrica*, 66, 289–314.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects." *Econometrica*, 73, 245–261.

Cochrane, J. H. (1991): "A simple test of consumption insurance." *Journal of Political Economy*, 99, 957–976.

Commault, J. (2022): "Does Consumption Respond to Transitory Shocks? Reconciling Natural Experiments and Semistructural Methods." *American Economic Journal: Macroeconomics*, 14, 96–122.

Creal, D. (2012): "A Survey of Sequential Monte Carlo Methods for Economics and Finance." *Econometric Reviews*, 31, 245–296.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the technology of cognitive and noncognitive skill formation." *Econometrica*, 78, 883–931.

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

Dempster, A., N. Laird, and D. Rubin (1977): "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society, Series B*, 39, 1–38.

D'Haultfoeuille, X. (2011): "On the completeness condition in nonparametric instrumental problems." *Econometric Theory*, 27, 460–471.

Fagereng, A., M. B. Holm, and G. J. Natvik (2021): "MPC Heterogeneity and Household Balance Sheets." *American Economic Journal: Macroeconomics*, 13, 1–54.

FRYBERGER, J. (2018): "Non-parametric panel data models with interactive fixed effects." *Review of Economic Studies*, 85, 1824–1851.

GOTTSCHALK, P. AND R. MOFFITT (1994): "The Growth of Earnings Instability in the U.S. Labor Market." *Brookings Papers on Economic Activity*, 1994, 217–272.

GOTTSCHALK, P. AND R. A. MOFFITT (2009): "The rising instability of U.S. earnings." *Journal of Economic Perspectives*, 23, 3–24.

GUVENEN, F. (2007): "Learning your earning: Are labor income shocks really very persistent?" *American Economic Review*, 97, 687–712.

GUVENEN, F., S. OZKAN, AND J. SONG (2014): "The nature of countercyclical income risk." *Journal of Political Economy*, 122, 621–660.

HALL, R. E. AND F. S. MISHKIN (1982): "The sensitivity of consumption to transitory income: Estimates from panel data on households." *Econometrica*, 50, 461–481.

HOROWITZ, J. L. AND M. MARKATOU (1996): "Semiparametric estimation of regression models for panel data." *Review of Economic Studies*, 63, 145–168.

HOSPIDO, L. (2012): "Modelling heterogeneity and dynamics in the volatility of individual wages." *Journal of Applied Econometrics*, 27, 386–414.

HU, Y. (2008): "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution." *Journal of Econometrics*, 144, 27–61.

——— (2024): *The Econometrics of Unobservables: Latent Variable and Measurement Error Models and Their Applications in Empirical Industrial Organization and Labor Economics.*, Johns Hopkins University.

HU, Y. AND S. M. SCHENNACH (2008): "Instrumental variable treatment of nonclassical measurement error models." *Econometrica*, 76, 195–216.

HU, Y. AND M. SHUM (2012): "Nonparametric identification of dynamic models with unobserved state variables." *Journal of Econometrics*, 171, 32–44.

JORDÀ, O. (2005): "Estimation and inference of impulse responses by local projections." *American Economic Review*, 95, 161–182.

KAPLAN, G. AND G. L. VIOLANTE (2010): "How much consumption insurance beyond self-insurance?" *American Economic Journal: Macroeconomics*, 2, 53–87.

Karahan, F. and S. Ozkan (2013): "On the persistence of income shocks over the life cycle: Evidence, theory and implications." *Review of Economic Dynamics*, 16, 452–476.

Koenker, R. and G. Bassett (1978): "Regression quantiles." *Econometrica*, 46, 33–50.

Kotlarski, I. (1967): "On characterizing the gamma and the normal distribution." *Pacific Journal of Mathematics*, 20, 69–76.

Lee, W. (2022): "Identification and estimation of dynamic random coefficient models." Working paper.

Lewis, D. J., D. Melcangi, and L. Pilossoph (2024): "Latent Heterogeneity in the Marginal Propensity to Consume." NBER Working Paper 32523.

Lillard, L. A. and R. J. Willis (1978): "Dynamics aspects of earning mobility." *Econometrica*, 46, 985–1012.

MaCurdy, T. E. (1982): "The use of time series processes to model the error structure of earnings in longitudinal data analysis." *Journal of Econometrics*, 18, 83–114.

Malcomson, J. H. (1999): "Individual employment contracts." in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier, vol. 3, chap. 35.

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

Misra, K. and P. Surico (2014): "Consumption, Income Changes, and Heterogeneity: Evidence from Two Fiscal Stimulus Programs." *American Economic Journal: Macroeconomics*, 6, 84–106.

Moffitt, R., J. Abowd, C. Bollinger, M. Carr, C. Hokayem, K. McKinney, E. Wiemers, S. Zhang, and J. Ziliak (2023): "Reconciling Trends in U.S. Male Earnings Volatility: Results from Survey and Administrative Data." *Journal of Business and Economic Statistics*, 41, 1–11.

Neal, R. M. (2003): "Slice sampling." *Annals of Statistics*, 31, 705–741.

Newey, W. K. and J. L. Powell (2003): "Instrumental variable estimation of nonparametric models." *Econometrica*, 71, 1565–1578.

Nielsen, S. F. (2000): "The stochastic EM algorithm: estimation and asymptotic results." *Bernoulli*, 6, 457–489.

Parker, J. A., N. S. Souleles, D. S. Johnson, and R. McClelland (2013): "Consumer Spending and the Economic Stimulus Payments of 2008." *American Economic Review*, 103, 2530–2553.

Pischke, J.-S. (1995): "Measurement Error and Earnings Dynamics: Some Estimates from the PSID Validation Study." *Journal of Business and Economic Statistics*, 13, 305–314.

Salgado, S., N. Bloom, F. Guvenen, L. Pistaferri, and J. Sabelhaus (2023): "Evaluating the Great Micro Moderation." Working paper.

Schennach, S. M. (2020): "Mismeasured and Unobserved Variables." in *Handbook of Econometrics*, ed. by S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, Elsevier, vol. 7A.

——— (2022): "Measurement Systems." *Journal of Economic Literature*, 60, 1223–1263.

Shen, X. (1997): "On methods of sieves and penalization." *Annals of Statistics*, 25, 2555–2591.

Souleles, N. S. (1999): "The response of household consumption to income tax refunds." *American Economic Review*, 89, 947–958.

Székely, J. C. and C. R. Rao (2000): "Identifiability of distributions of independent random variables by linear combinations and moments." *Sankhyā*, 62, 193–202.

Wilhelm, D. (2015): "Identification and estimation of nonparametric panel data regressions with measurement error." CeMMAP Working Paper 34/15.