centre for microdata methods and practice

Nonlinear micro income processes with macro shocks

Martín Almuzara
Manuel Arellano
Richard Blundell
Stéphane Bonhomme

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP17/25

Economic and Social Research Council

Nonlinear Micro Income Processes with Macro Shocks*

Martín Almuzara[†] Manuel Arellano[§]

Richard Blundell[‡] Stéphane Bonhomme[¶]

August 2025

[Link to supplemental appendix]

Abstract

We propose a nonlinear framework to study the dynamic transmission of aggregate and idiosyncratic shocks to household income that exploits both macro and micro data. Our approach allows us to examine empirically the following questions: (a) How do business-cycle fluctuations modulate the persistence of heterogeneous individual histories and the risk faced by households? (b) How do aggregate and idiosyncratic shocks propagate over time for households in different macro and micro states? (c) How do these shocks shape the cost of business-cycle risk? We develop new identification and estimation techniques, and provide a detailed empirical analysis combining macro time series for the U.S. and a time series of household panels from the PSID.

Keywords: Income process, business cycle, persistence, exposure to aggregate shocks.

*We have greatly benefited from discussions with Anmol Bhandari, Mariacristina De Nardi, Gianluca Violante, Jonathan Heathcote, and Frank Schorfheide. We also thank Greg Kaplan, Giuseppe Moscarini, Larry Schmidt, Christian Wolf, seminar participants at CEMFI, Minnesota and Princeton, and audiences at the Workshop on Income, Employment and Consumption Dynamics (Madrid), the NBER Summer Institute, and the 2025 World Congress of the Econometric Society. Blundell acknowledges support from the ESRC Centre for the Microeconomic Analysis of Public Policy (CPP) at IFS, grant number ES/M010147/1. The views expressed in this paper do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

[†]FRBNY: martin.almuzara@ny.frb.org

[‡]UCL and IFS: r.blundell@ucl.ac.uk

[§]CEMFI: arellano@cemfi.es

[¶]University of Chicago: sbonhomme@chicago.edu

1 Introduction

In this paper, we propose a nonlinear framework to study the dynamic transmission of aggregate and idiosyncratic shocks to income by leveraging both macro and micro data. Our approach makes it possible to empirically examine how business-cycle fluctuations modulate the persistence of heterogeneous individual histories and the risk faced by households. We also consider questions such as how aggregate and idiosyncratic shocks propagate over time for units in different macro and micro states, and how these shocks contribute to the cost of business-cycle risk. Answering these questions is important. They are essential to documenting the dynamics of income inequality over the business cycle. Furthermore, how the incomes of heterogeneous agents respond to macro and micro shocks is key for consumer and firm behavior, and for the design of optimal monetary and fiscal policies (Bhandari, Evans, Golosov, and Sargent, 2021).

The literature on income risk has uncovered significant nonlinearities in the dynamics of individual incomes (Arellano, Blundell, and Bonhomme, 2017; Guvenen, Karahan, Özkan, and Song, 2021) and in their variation over the business cycle (Guvenen, Ozkan, and Song, 2014). Moreover, a growing recent literature investigates the heterogeneous effects of monetary policy shocks on individual-level outcomes (Holm, Paul, and Tischbirek, 2021; Andersen, Johannesen, Jørgensen, and Peydró, 2023; Amberg, Jansson, Klein, and Rogantini Picco, 2022). Yet, a methodology for modeling the interaction between micro and macro shocks capable of integrating non-linearities in the life-cycle and business-cycle dynamics of income is still lacking. This is our main contribution.

We consider a nonlinear Markovian micro income process with a macro state variable of the following form:

$$\eta_{it} = Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, u_{it}),$$
$$Z_t = Q_Z(Z_{t-1}, V_t),$$

where u_{it} and V_t are micro and macro shocks, and η_{it} and Z_t are potentially unobserved. A measurement system connects these two latent variables to observed micro and macro data, specifically, a flexible persistent-transitory model for the micro states and a dynamic factor model for the macro states. Our triangular formulation has the potential to allow for feedback from the micro to the macro level, as Z_t can incorporate distributional characteristics of the micro data. The assumption underlying the triangular structure is atomicity, meaning that no single individual unit influences the aggregate state.

Based on our income process we will highlight two key quantities. The first one is the elasticity of individual persistent earnings to the aggregate business-cycle state Z_t:

$$\beta_{it} = \frac{\partial \eta_{it}}{\partial Z_t}.$$

In our setup, β_{it} is a measure of a household's exposure to shocks to the aggregate state that is heterogeneous along both the income distribution and business-cycle conditions. In addition, β_{it} may vary with the idiosyncratic shock u_{it}, and so the impact of an aggregate shock may differ depending on idiosyncratic events (such as a job loss or a promotion).[1]

The second quantity is income persistence:

$$\rho_{it} = \frac{\partial \eta_{it}}{\partial \eta_{i,t-1}}.$$

Here, ρ_{it} is a measure of nonlinear persistence (Arellano et al., 2017, ABB) that may vary depending on the position in the income distribution and the idiosyncratic shocks hitting the household. Moreover, unlike in ABB, our model allows for the aggregate state to affect income and, thus, for the shape of persistence to be different in good or bad times.

Documenting how β_{it}s and ρ_{it}s vary across income histories and over the business cycle allows us to paint a rich picture of the interaction between individual and aggregate income dynamics. Our model also allows us to flexibly measure how features of the income process such as (conditional) dispersion and skewness depend on business-cycle conditions. Since the model completely specifies the law of motion of individual income η_{it} and the aggregate state Z_t, it can be used for impulse response analysis and to quantify the cost of both micro and macro sources of income risk, as we illustrate empirically.

We study the nonparametric identification of this model bringing together macroeconometric and microeconometric techniques. The micro side builds on ABB. The macro side relies on identification results for factor models estimated from time series aggregates (Stock and Watson, 2016). To combine the two parts, we create a time series of short panels, such that each panel is representative of the economy in that period. Thus, identification

[1]Our measure is different from, but related to, others in the growing empirical literature on the heterogeneous effects of recessions. For example, Guvenen, Schulhofer-Wohl, Song, and Yogo (2017) document variation in individual income exposure to aggregate income—referred to as 'worker betas'—across age and income groups; while Patterson (2023) examines how these heterogeneous exposures relate to differences in marginal propensities to consume across demographic groups.

relies on three dimensions of our combined dataset: large cross-sections, short individual panels, and long time series spanning sufficient cyclical fluctuations.

We propose an approach to estimation and inference that uses a flexible parametric version of the model and can be implemented with stable simulation-based algorithms. This approach was first introduced in Arellano and Bonhomme (2016) and was adapted to a setup with time-varying latent variables in ABB. Here it is further extended to a long time series of short panels involving both micro and macro latent variables. Our stochastic EM algorithm iterates between draws of the latent variables from their posterior distributions evaluated at current parameter values, and updates of the parameters from quantile regressions based on those draws. Furthermore, the method of inference we develop is robust to forms of cross-sectional and unit-level dependence in micro shocks u_{it} that arise from unobserved common factors and realistic sampling designs.

We also develop a methodology for impulse response function analysis in our nonlinear context where we want to measure the importance of macro and micro shocks and their interactions. We start by considering an experiment in which we directly perturb a state variable at some point in time. We then compare the trajectory of the system following the perturbation with a baseline trajectory in the absence of perturbation. Since we wish to obtain comparability of impulse responses across households with different processes, we need to consider ways of introducing comparable perturbations. We do so via a set of *rules* that map perturbations to a common system of units, and we show that different perturbation experiments can be associated with different formulations of local shocks.

We take our model to quarterly macro time series data for the U.S. and a time series of panels that we construct from the Panel Study of Income Dynamics (PSID) spanning the period 1970-2019, thus covering seven recessions. PSID waves are annual up to 1997 and biennial afterwards. For consistency of the microdata, we then form sequences of biennial subpanels of four waves each covering all the years available, in the spirit of Storesletten, Telmer, and Yaron (2004). Compared to standard long panel approaches, this *time series of panels* approach has the advantage of mitigating concerns over the representativeness of the data. In this paper, our primary focus is on disposable household income net of taxes and transfers. However, for comparability with other studies, we also present estimates based on male earnings and household earnings before taxes and transfers.[2]

[2]The last results are limited to units with positive earnings. The share of households with zero earnings stays below 2% for most of the period, rising slightly during the Great Recession. For men, zero earners remain under 6% until the Great Recession, when the rate nears 8%. See Figure 2 and its discussion.

We leave to future work the estimation of the transmission of aggregate and idiosyncratic income shocks to consumption, although this is a natural extension of the analysis of nonlinear income and consumption dynamics in the PSID developed in Arellano et al. (2017) and Arellano, Blundell, Bonhomme, and Light (2023).

Empirical results. Our analysis yields a number of novel empirical insights.

To begin, our results illustrate the asymmetric impact of the business cycle on income persistence. Specifically, income persistence ρ_{it} increases for low-income households and decreases for high-income households during recessions. That is, during a downturn, it is harder for low-income earners to leave the low-income state, whereas for high-income earners remaining high-income becomes more difficult. The cyclical variation we find in the persistence of past income histories coexists with the ample variability along income and micro-shock distributions uncovered by ABB.

Our results also highlight the presence of heterogeneous exposures to aggregate shocks. The coefficients β_{it} tend to be higher when associated with bad idiosyncratic shocks. Moreover, we find that β_{it}s are countercyclical: they are higher in recessions and lower in expansions. This is a key finding because the cyclical behavior of income elasticities to the macro state—particularly the self-amplifying nature of negative aggregate shocks—has major implications for the cost of aggregate income risk, as we argue below.

In addition, we document two main facts about income skewness. First, we find that left skewness is countercyclical, consistent with the findings in Guvenen et al. (2014). Second, we find that skewness decreases with income at any point in time, consistent with the findings in ABB, but differentially so depending on the phase of the business cycle. This *tale of two skewnesses* is a clear reflection of the nonlinear transmission of micro and macro shocks. In this and other dimensions, business-cycle variability is most pronounced for male earnings and, to a lesser extent, for household earnings and disposable income, in that order—that is, from the income measure with the least insurance to the one with the most—but the patterns are qualitatively similar for all three measures.

Concerning impulse response functions, we find a large direct impact of macro shocks but with generally short-lived effects. These impacts, however, are highly heterogeneous across income measures (larger for male earnings, smaller for disposable income) and along the persistent income distribution (larger at the bottom, smaller in the middle), and they interact with idiosyncratic uncertainty by compounding the negative consequences of bad micro shocks. In contrast, micro impulse responses decay slowly, with different degrees of persistence depending on the initial level of income. In fact, as we show below, there is a tight link between impulse responses and β_{it}s and ρ_{it}s.

Finally, we use our framework to quantify the cost of both macro and micro sources of risk. For macro risk, we compute the (per cent) compensating variation that equalizes the expected lifetime utility of income with and without macro shocks (as in Lucas, 1987, 2003), where expectations come from our estimated income process. In this exercise, the cyclical behavior of the nonlinear income exposure to macro shocks β_{it} is a key determinant of the cost of business cycles. In the presence of empirically plausible amplification effects, the yearly cost of macro risk can be as high as 5 percent of income, whereas it is negligible under a linear transmission of macro shocks. This is a novel channel through which macro uncertainty can lead to welfare losses at the household level, distinct from curvature in the utility function. Although as expected the cost of micro risk is higher, for many units (mainly young and low-income) macro shocks explain a large fraction of the total cost of risk. It is worth emphasizing that an income process linear in aggregate shocks would miss all of the rich cyclical patterns documented here. Thus, a core lesson from our paper is the importance of accounting for nonlinearities in β_{it} when the goal is to study the consequences of aggregate fluctuations.

Selected literature. Our paper contributes to several strands of the literature. First, we build on the vast literature on income dynamics, both with and without business cycles, e.g., Gottschalk and Moffitt (1994, 2009), Meghir and Pistaferri (2004), Blundell, Pistaferri, and Preston (2008), Browning, Ejrnæs, and Álvarez (2010), Altonji, Smith, and Vindangos (2013); see Blundell, Bollinger, Hokayem, and Ziliak (2024) for a comprehensive review. Within that body of work, our paper is most closely related to Storesletten et al. (2004), Guvenen et al. (2014), Arellano et al. (2017), Guvenen, McKay, and Ryan (2023), Halvorsen, Holter, Ozkan, and Storesletten (2024), Guvenen, Pistaferri, and Violante (2022), and the multi-country GRID project. Relative to this work, we are the first to develop a framework capable of integrating aggregate shocks and rich nonlinear dynamics at the micro level.

Second, we contribute to the literature on estimating heterogeneous agents models using micro data (Arellano and Bonhomme, 2017; Liu and Plagborg-Møller, 2023; Fernández-Villaverde, Hurtado, and Nuño, 2023). Compared to them, we offer a principled approach to building nonlinear reduced forms for that class of models when agents face potentially latent macro-level uncertainty. More generally, we add to an early literature on the combination of household survey data with time series data (Tobin, 1950; Chetty, 1968; Maddala,

1971) and to recent work on the econometrics of models with aggregate shocks (Hahn, Kuersteiner, and Mazzocco, 2020; Almuzara and Sancibrián, 2024) by developing novel tools for identification analysis and estimation in a time series of panels framework.

An important reference is Chang, Chen, and Schorfheide (2024), who propose functional vector autoregression methods to combine macro aggregates with repeated cross-sections. In a setup with both macro and micro data, the effects of macro shocks at the unit level and their impact on cross-sectional distributions are distinct but related empirical objects. While we focus on the former, Chang et al. (2024) aim at the latter. From this perspective, our papers are complementary. Another recent related reference is Sargent and Selvakumar (2025), who propose a dynamic mode decomposition method to study distributional dynamics of income and consumption. Compared to both papers, a distinctive feature of our work is the use of panel data to analyze household dynamics.

Lastly, our empirical results speak to a vast macro literature on inequality and aggregate fluctuations (e.g., Krusell and Smith, 1998; Krueger, Mitman, and Perri, 2016; Ahn, Kaplan, Moll, Winberry, and Wolf, 2018; Bhandari et al., 2021), and to work that seeks to quantify the welfare cost of business cycles (Lucas, 1987, 2003; Storesletten, Telmer, and Yaron, 2001; Otrok, 2001; Barlevy, 2004; Galí, Gertler, and López-Salido, 2007; Krebs, 2007). In particular, we document new empirical patterns about the interaction between cyclical variation and nonlinearities in the income processes that have the potential to amplify the welfare consequences of recessions and expansions.

In summary, our paper integrates these different strands by building a semi-structural framework with nonlinear micro dynamics and macro shocks, offering a unified approach to study the interplay between individual behavior and aggregate fluctuations.

Outline. The paper is organized as follows. Section 2 outlines our framework of analysis and the key objects of empirical interest. Section 3 introduces the statistical population by means of a time series of panels and studies identification. Section 4 details our estimation strategy. In Section 5, we present empirical results on nonlinear persistence, exposure to aggregate shocks, and skewness over the business cycle. Section 6 develops methodology for nonlinear impulse response functions of macro and micro shocks, while Section 7 focuses on quantifying the idiosyncratic and aggregate components of income risk. Section 8 concludes. Additional material can be found in the Supplemental Appendix.

2 Framework

In this section we describe our framework of analysis that combines time series aggregates with longitudinal micro-level survey data. While our interest is on the impact of macro shocks on nonlinear income risk, our approach provides a basis to build rich nonlinear reduced forms for heterogeneous agents models with both macro and micro uncertainty.

2.1 Model

We model log income y_{it} of household i at time t as the sum of a persistent component η_{it} and a transitory component ε_{it},

$$y_{it} = \eta_{it} + \varepsilon_{it}.$$

While the researcher observes log income y_{it}, the two components η_{it} and ε_{it} are latent.[3]

Following Arellano et al. (2017, ABB), the persistent component is a flexible first-order Markov process, the transitory component is serially independent (using biennial observations), and both processes may depend on covariates x_{it} such as the age of the household head. However, unlike ABB, our focus is on understanding how aggregate conditions impact income trajectories. For this purpose, we introduce a time series aggregate Z_t, a macro state variable that affects η_{it} and ε_{it}, and that we will infer from aggregate time series data.

Persistent component. We specify the persistent component as

$$\eta_{it} = Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, u_{it}). \tag{1}$$

In (1), Q_η is strictly increasing in its last argument and the shock u_{it} is i.i.d. uniform on $(0,1)$, independent of $(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it})$. Hence, for all $\tau \in (0,1)$, $Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, \tau)$ is the conditional τ-quantile of η_{it}. This model allows for a general nonlinear relationship between the persistent income component η_{it} and its various determinants: lagged income $\eta_{i,t-1}$, the aggregate factor Z_t, covariates x_{it}, and idiosyncratic shocks u_{it}.

[3] We build on a long tradition that uses panel data to decompose income into persistent and transitory components (Hause, 1977; Lillard and Willis, 1978; MaCurdy, 1982); see Arellano (2014) for a survey.

To aid interpretation, it is useful to consider the following special case of Equation (1):

$$\eta_{it} = \rho\eta_{i,t-1} + \gamma\eta_{i,t-1}Z_t + \delta Z_t + \underbrace{g(Z_t, u_{it})}_{=\zeta_{it}}, \tag{1'}$$

where for simplicity we have omitted the dependence on Z_{t-1}. The quantity $\zeta_{it} = g(Z_t, u_{it})$ is a *composite income shock*, measured in log income units, which exhibits aggregate variation due to the presence of Z_t and idiosyncratic variation driven by u_{it}.

In model (1′), the impact of a marginal change in the aggregate state Z_t on the persistent component of income (the household's aggregate exposure β_{it}) can be decomposed as

$$\frac{\partial\eta_{it}}{\partial Z_t} = \underbrace{\gamma\eta_{i,t-1} + \delta}_{\text{income heterogeneity}} + \underbrace{\frac{\partial g(Z_t, u_{it})}{\partial Z_t}}_{\text{shock distribution}}.$$

The first term shows that the impact of an aggregate shock can vary along the income distribution, a point highlighted in an extensive literature in macroeconomics; the second shows that aggregate shocks may alter the distribution of income shocks, e.g., by affecting their variance (as in Storesletten et al., 2004) or skewness (as in Guvenen et al., 2014).[4]

Compared to (1′), which permits interactions between Z_t and past income $\eta_{i,t-1}$ and between Z_t and the rank u_{it}, model (1) allows for a third type of interaction between past income $\eta_{i,t-1}$ and the shock u_{it}. Model (1) is therefore better able to capture nonlinearities in income persistence, including the observation from ABB that different shocks to the persistent component may be associated with different degrees of persistence. This is in addition to the rich life-cycle patterns and the unrestricted relationship between macro and micro responses that model (1) can generate. The latter plays a key role in empirically capturing the shape of individual impulse responses to aggregate shocks.

Transitory component and initial condition. We specify the transitory component as

$$\varepsilon_{it} = Q_{\varepsilon,t}(x_{it}, v_{it}), \tag{2}$$

[4]For a special case consider $g(Z_t, u_{it}) = \sigma(Z_t)q(u_{it})$, where $\frac{\partial g(Z_t,u_{it})}{\partial Z_t} = \frac{\partial\sigma(Z_t)}{\partial Z_t}q(u_{it})$ reflects the impact of Z_t on the variance of the composite shock distribution $\sigma^2(Z_t)$.

with $Q_{\varepsilon,t}$ strictly increasing in its last argument, and v_{it} an i.i.d. shock, uniform on $(0,1)$ and independent of x_{it} and u_{it}. The function $Q_{\varepsilon,t}$ may vary over time in unrestricted ways reflecting general aggregate effects (coming from Z_t or other factors) on transitory income risk. Our specification can also accommodate non-Gaussianity in the conditional density of transitory shocks. In practice, ε_{it} will potentially be a mix of substantive transitory shocks and measurement error, and without further assumptions our approach will not allow us to distinguish between the two. For this reason, in our empirical analysis we will mostly focus on interpreting the properties of the persistent component.

Lastly, we specify the initial condition of the persistent income process as

$$\eta_{i,t_0} = Q_{\text{init},t_0}(x_{i,t_0}, \nu_{i,t_0}),$$

with Q_{init,t_0} strictly increasing in its last argument, and ν_{i,t_0} uniformly distributed on $(0,1)$, and independent of x_{i,t_0}, u_{it} and v_{it}. With a similar rationale as for $Q_{\varepsilon,t}$, we let Q_{init,t_0} depend flexibly on the initial time period t_0 (which may differ across individuals).

Since we permit general time-variation in both the transitory component and the initial condition, other factors beyond the business-cycle state Z_t can influence the evolution of income in our model. Thus, as $Q_{\varepsilon,t}$ and Q_{init,t_0} reflect a mix of business-cycle and other factors, we will interpret them as rich time-varying "controls", our chief goal being instead to document the dynamics of the persistent component captured by Q_η.

Macro state variable. The unobservable aggregate state Z_t is estimated from observables W_t by relying on a linear dynamic factor structure:

$$\begin{aligned} W_t &= \Lambda Z_t + e_t, \\ Z_t &= \Phi Z_{t-1} + \Sigma^{1/2} V_t = Q_Z(Z_{t-1}, V_t), \end{aligned} \tag{3}$$

with V_t an i.i.d. standard normal vector of shocks independent of e_t at all lags and leads, and with the elements of e_t specified as mutually independent Gaussian autoregressive processes. In this formulation, the aggregate state Z_t will typically display persistence, reflecting the dynamic impact of current and past aggregate shocks $\{V_{t-\ell}\}_{\ell \geq 0}$.[5]

In our empirical analysis, W_t consists of variables that are informative about aggregate

[5] Since Z_t can be a vector, the restriction to a first-order process in (3) is without loss of generality—any VAR(p) can be cast as a VAR(1) in companion form. Our setting also accommodates richer processes with heteroskedasticity and nonlinearities, provided Q_Z can be identified from the time series W_t (see Section 3).

fluctuations (e.g., GDP and the unemployment rate) while Z_t is a parsimonious summary indicator of the state of the business cycle.[6] One interpretation of Z_t draws on Angeletos, Collard, and Dellas (2020) who find that a single serially uncorrelated process, what they call the *main business cycle* (MBC) shock, accounts for the largest share of the unpredictable variation in each of the variables W_t at business cycle frequencies. This MBC shock is not a specific structural shock (e.g., TFP or demand), but possibly a mix of shocks that produce consistent impulse-response patterns across cyclical variables. In that light, we can view the shock V_t in (3) as an approximation to the MBC shock, and the business cycle state Z_t as a persistent process that captures the cumulative effect of MBC shocks over time.[7]

Macro-micro feedback. We can view Equations (1) and (3) as a nonlinear VAR(1) model for $\{\eta_{it}\}_i$ and Z_t, where i indexes the relevant time-t cross-sectional population—the precise notion is discussed in Section 3. What remains is to specify the link between the macro and micro sides. We impose two restrictions formalized in Assumption 3 below.

First, conditional on Z_t, the aggregate data W_t carry no extra information about shocks u_{it} to the dynamics of η_{it}. This is analogous to Liu and Plagborg-Møller (2023, Assumption 1) and implies that Z_t is a sufficient statistic of the macro data for micro-level persistent dynamics. Second, we impose that individual past income shocks do not affect Z_t. It is important to highlight here that our model does not rule out feedback from the micro to the macro side. We can accommodate such feedback provided it is channeled through aggregate summaries of the micro state distribution. In that case, one could augment W_t with moments from the time-t earnings distribution and redefine Z_t accordingly. Instead, what we require is that no single unit in the sample has aggregate effects, that is, that units are *atomistic*. In Supplemental Appendix A, we use a stylized example to illustrate how the equilibrium conditions from standard heterogeneous agents models with aggregate shocks can be cast in the form of (1) and (3), and how atomicity is to be interpreted.

To summarize, our framework can capture rich aggregate dynamics and macro-micro interactions. The limit is statistical rather than conceptual: as shown in Section 4, estima-

[6]The idea of extracting a low-dimensional summary of the covariation of multiple economic time series is at the core of the concept of business cycle (Burns and Mitchell, 1946). The use of linear factor models for that purpose also has a long history in economics; see Stock and Watson (2016) for a survey.

[7]One difference is that Angeletos et al. (2020) recover the MBC shock by maximizing the forecast error variance decomposition of W_t over a band of business-cycle frequencies, whereas we recover our shock V_t (and the state Z_t) by first applying a band-pass filter to W_t (tailored to business-cycle frequencies) and then using dynamic factor techniques. Empirically, IRFs of W_t to V_t are quantitatively similar to IRFs to the MBC shock, albeit slightly less persistent. See Supplemental Appendix E.3 for a comparison.

tion precision depends on the time series length T, and a high-dimensional Z_t can quickly exhaust the degrees of freedom offered by a short time series.

2.2 Objects of empirical interest

A primary goal of the empirical analysis is to quantify the effect of aggregate shocks on nonlinear income risk. Our framework allows us to recover several quantities of interest. A first quantity is the following measure of *nonlinear persistence* which extends the ABB persistence to a setup with aggregate shocks:

$$\rho\left(u_{it}, \eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}\right) = \frac{\partial Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, u_{it})}{\partial \eta_{i,t-1}}. \tag{4}$$

That is, persistence is measured by how a change in the micro state $\eta_{i,t-1}$ affects its next-period value. In a linear autoregressive model, $\rho(u_{it}, \eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}) = \rho$ is constant and equal to the autoregressive root of the process. In contrast, in our nonlinear process, this measure is state-dependent: persistence may vary with the past position in the income distribution $\eta_{i,t-1}$, with aggregate conditions Z_t and Z_{t-1}, and with covariates (e.g., age) x_{it}.

A key feature of the measure $\rho(u_{it}, \eta_{i.t-1}, Z_t, Z_{t-1}, x_{it})$ is that it captures how persistence varies with the micro shock u_{it}, and how u_{it} interacts with the remaining determinants of persistent income. In micro panels, a robust finding (first documented in ABB) is that persistence decreases for good-shocks/low-η and bad-shocks/high-η combinations. This reflects the fact that a good shock arriving in a low-income state has sometimes the power to erase a bad income history; the unlucky reverse holds for high-η households reached by a bad u. This feature is absent from linear income processes. But the novel framework of this paper adds an extra layer. In (4), the entire shape of ρ as a function of u_{it} and $\eta_{i,t-1}$—the intensity with which past income histories are wiped out by big shocks—can change with the aggregate state of the economy Z_t. Our empirical analysis in Section 5 presents evidence of business cycle variation in nonlinear persistence.

A second key quantity is the persistent income *nonlinear exposure to aggregate shocks*:

$$\beta\left(u_{it}, \eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}\right) = \frac{\partial Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, u_{it})}{\partial Z_t}. \tag{5}$$

This measure captures how the exposure of a household to the aggregate state varies with the determinants of persistent income η_{it}. In particular, the aggregate exposure may vary along the persistent income distribution $\eta_{i,t-1}$, over the life-cycle x_{it}, or as a function of the idiosyncratic shock the household is hit with u_{it}. It may also display cyclical patterns as it depends on the aggregate states Z_t and Z_{t-1}. It is worth noting that this measure is not available in models of income risk that exclude macro shocks, and in most specifications that do include aggregate uncertainty, it is typically restricted to be constant, thus ruling out cyclical exposures and interactions with micro-level states and shocks.

A third quantity of interest is the following measure of *conditional skewness*:

$$\begin{aligned}\mathrm{sk}\left(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}\right) = &\left[Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.9) - Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.1)\right]^{-1} \\ &\times \left[Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.9) + Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.1) - 2Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.5)\right].\end{aligned} \tag{6}$$

This is the Kelley skewness of the predictive distribution of η_{it}. In micro panels, skewness tends to be decreasing in past $\eta_{i,t-1}$, that is, income risk is tilted to the upside for the low-η and to the downside for the high-η. As for persistence, the novelty of our paper is to allow us to trace how skewness changes with the aggregate state of the economy Z_t (and Z_{t-1}). How skewness varies with Z_t is a paramount empirical question. Guvenen et al. (2014), for example, find that recessions have a sizable negative impact on the skewness of one-year income growth. Our framework enhances previous analyses in various directions. First, it allows us to quantify the cyclical patterns of persistent income risk directly, without the need to proxy it by income changes that complicate imputing the effect of specific macro shocks. Second, it allows us to explore the role of the household's position in the income distribution and in the life cycle. Thus, our approach provides a unified treatment of the two types of income skewness that have been highlighted in the literature: across the income distribution (indexed by $\eta_{i,t-1}$) and over the business cycle (indexed by Z_t).[8]

Beyond exploring the cyclical behavior of nonlinear income risk, our setup allows us to recover impulse response functions (IRF) to both macro shocks (i.e., a shock to Z_t such as a recession or boom) and micro shocks (i.e., a shock u_{it} such as a promotion or demotion), as we discuss in Section 6. Finally, our framework allows us to quantify the contribution

[8] Similarly, we can define measures of conditional dispersion and kurtosis:

$$\begin{aligned}\mathrm{disp}\left(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}\right) &= Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.9) - Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.1), \\ \mathrm{kurt}\left(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}\right) &= \left[Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.75) - Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.25)\right]^{-1} \\ &\quad \times \left[Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.95) - Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.05)\right].\end{aligned}$$

We report estimates of these quantities in Supplemental Appendix D.

of macro and micro shocks to the cost of income risk, the topic of Section 7.

3 Statistical population and identification

Here we describe the statistical population concept for our framework which is based on what we refer to as a *time series of panels*. Specifying the data generating process involves two challenges: allowing for potentially omitted aggregate shocks and accounting for the fact that the cross-sectional population under observation changes over time. This section focuses on identification and we defer estimation and inference until Section 4.

3.1 Time series of panels

We begin by introducing the notion of time series of panels, a data structure whose infinite-sample counterpart is the statistical population of our setting. The researcher observes a time series of aggregate indicators

$$W_t, \qquad t = 1, \ldots, T + S,$$

which we use to measure the aggregate factor Z_t, and a sequence of panel datasets

$$Y_t, \qquad t = 1, \ldots, T,$$

where each $Y_t = \{y_{i,t+s} : i \in \mathcal{I}_t,\ 0 \leq s \leq S - 1\}$ contains individual observations on units in a sample $\mathcal{I}_t$ of size N_t, representative of the population of interest at time t. The length of each panel, S, aims at preserving the representativeness of the samples $\mathcal{I}_t$ while ensuring identification of the parameters of the model.[9] Our approach relies on large cross-sections (large N_t), short subpanels (small and fixed S) and a long time series spanning sufficient aggregate variation coexisting with an evolving population of micro units (large T).

An advantage of our time series of panels approach, compared to standard long panel approaches, is that it mitigates concerns over the representativeness of the data when

[9] This structure nests several important cases. First, a rotating panel scheme, where a share $1/S$ of individuals is refreshed every period, is covered by our setup, since an individual may appear in several subpanels; i.e., $i \in \mathcal{I}_t$ neither precludes nor implies $i \in \mathcal{I}_\tau$ for $\tau \neq t$. Second, repeated cross-sectional data corresponds to the case $S = 1$ with T equal to the number of cross-sections. Third, a single long panel corresponds to $T = 1$ and S equal to the number of time periods—alternatively, a long panel can always be split into an overlapping sequence of shorter panels.

attrition depends on income shocks or covariates (such as age). For example, a long panel of households spanning the years 1970 to 2019 (as in our empirical analysis) will over-represent more stable units with higher incomes compared to an actually representative sample in any given year. A precursor of the time series of panels is the analysis of cyclical income risk in the U.S. by Storesletten et al. (2004).[10]

We are now ready to describe the statistical population. For conciseness we use $\{A_t\}$ as shorthand for the time series process $\{A_t : -\infty < t < \infty\}$.

Assumption 1 (Macro states). *There is a macro stochastic process* $\{Z_t, e_t, G_t, \omega_t\}$ *such that*

(a) $\{Z_t\}$ *is stationary and satisfies* (3) *with* V_t *i.i.d. over* t *and standard normal;*

(b) the entries of $\{e_t\}$ *and the possibly unobserved aggregate process* $\{G_t\}$ *are mutually independent and independent of* $\{Z_t\}$, *with* e_t *stationary and Gaussian, and with* G_t *i.i.d. over* t*; and*

(c) $\{\omega_t\}$ *is a process of potentially arbitrary dimension that encompasses* Z_t, e_t *and* G_t, *as well as any other time series factors that may affect the evolution of the micro data (e.g., through* ε_{it}*).*

Assumption 2 (Micro processes with macro states). *There is a cross-section of micro stochastic processes* $\{\eta_{it}, \varepsilon_{it}, x_{it}\}$ *such that*

(a) $\{\eta_{it}, \varepsilon_{it}, x_{it}\}$ *is i.i.d. over* i *given* $\{\omega_t\}$*; and*

(b) $\{\eta_{it}, \varepsilon_{it}\}$ *given* $\{\omega_t, x_{it}\}$ *satisfies* (1) *and* (2) *with* u_{it} *and* v_{it} *mutually independent, i.i.d. over* t, *and uniformly distributed on* $(0, 1)$.

Assumption 3 (Atomicity). *Conditional on* x_{it} *and* G_t, *for each unit* i, *the persistent micro shock* u_{it} *is independent of the macro shocks* V_t *and the errors* e_t.

In Assumptions 1, 2 and 3, the process G_t represents potentially unobserved or omitted aggregate variables, such as shocks orthogonal to the business cycle or survey-specific factors. Part 1(b) normalizes the serial dependence of G_t. In particular, omitted persistent macro variables can be accommodated by simply taking G_t to be their surprise (or shock) component. We also assume that W_t is uninformative about G_t and that Z_t (the aggregate of interest) is unaffected by G_t. This extends to our nonlinear setup the standard shock orthogonality idea from linear macroeconometric models (e.g., Ramey, 2016).

[10] An additional advantage of this approach is that it allows us to handle covariates that would otherwise be collinear with time, e.g., age. Moreover, it may be expensive (sometimes impossible) to keep track of the same units (e.g., companies) for very long periods.

The remaining parts of Assumptions 1 and 2 impose (1), (2) and (3), while Assumption 3 requires that no single unit has influence on the aggregate. Nonetheless, as discussed in Section 2, our framework permits feedback from the micro to the macro side of the model, as summaries of micro distributions may be included in Z_t.

3.2 Identification

The identification problem is to determine the unknown functions Q_η and Q_Z from the probability distribution of observables. As anticipated, our argument applies to $N_t, T \to \infty$ with S fixed. In our setup, the distribution of observables is the distribution P^S that assigns probabilities to events involving finite segments of the vector/function-valued stochastic process $\{\xi_t^S\}$ where $\xi_t^S = (\{W_{t+s-1}\}_{s=0}^S, F_t^S)$ with

$$F_t^S(\boldsymbol{y}|\boldsymbol{x}) = P\Big((y_{it}, ..., y_{i,t+S-1}) \leq \boldsymbol{y} \Big| (x_{it}, ..., x_{i,t+S-1}) = \boldsymbol{x}, \omega_t \Big).$$

Here, $F_t^S(\boldsymbol{y}|\boldsymbol{x})$ is the cumulative distribution function (CDF) of $\{y_{i,t+s}\}_{s=0}^{S-1}$ given $\{x_{i,t+s}\}_{s=0}^{S-1}$ at time t. It is a random function that describes cross-sectional probabilities of $\{y_{i,t+s}, x_{i,t+s}\}_{s=0}^{S-1}$. The randomness of $F_t^S(\boldsymbol{y}|\boldsymbol{x})$ comes from the time series process $\{\omega_t\}$ of potentially arbitrary dimension and unknown to the researcher introduced in Assumption 1 (c).

We next outline our identification analysis which brings together techniques from the macro- and microeconometric literatures, linked by the time series of panels.

Identification of the macro process. Knowing P^S implies knowing the distribution of the stochastic process $\{W_t\}$ and, in particular, the autocovariance function $\ell \mapsto \mathrm{Cov}(W_t, W_{t-\ell})$. In our linear Gaussian environment, this suffices to identify Q_Z provided there are enough measurements and a factor normalization holds.

Proposition 1. *Under Assumption 1, if (a)* $(\dim(W_t) - \dim(Z_t))^2 > \dim(W_t) + \dim(Z_t)$ *and (b) the upper block of* Λ *is the identity matrix, then* Q_Z *is identified.*

Proof. See Supplemental Appendix B. □

This follows from standard results for linear state-space models. We note that condition (b) could be replaced by alternative factorr normalizations (Stock and Watson, 2016). Although in more general nonlinear and non-Gaussian settings a different technique will be needed, it is often possible to learn about Q_Z using external time series data alone. On the other hand, when Z_t is directly observed (as in, e.g., Storesletten et al., 2004 where Z_t is the NBER recession indicator), this step is not needed as Q_Z is automatically identified.

Identification of the micro process. For a given t, knowledge of F_t^S allows the researcher to pin down the latent variable distributions that condition on the unknown macro shocks coincidental with the time-t panel. This is a microeconometric identification problem.

Proposition 2. *Suppose $S \geq 4$ and almost surely over realized paths of $\{\omega_\tau\}$, for each t,*

(a) the density of $\{y_{i,t+s}, \eta_{i,t+s}\}_{s=0}^{S-1}$ given $\{x_{i,t+s}\}_{s=0}^{S-1}$ and ω_t is bounded away from zero and infinity and so are all corresponding marginals and conditionals, and

(b) with $f_{\eta_r|y_{\tilde{r}},x,t}$ the density of η_{ir} given $y_{i\tilde{r}}$, $\{x_{i,t+s}\}_{s=0}^{S-1}$ and ω_t, the families

$$\mathcal{F}_-^{r,t}(\boldsymbol{x}) \equiv \left\{f_{\eta_r|y_{r-1},x,t}(\cdot|y,\boldsymbol{x}) : y \in \mathbb{R}\right\} \text{ and } \mathcal{F}_+^{r,t}(\boldsymbol{x}) \equiv \left\{f_{\eta_r|y_{r+1},x,t}(\cdot|y,\boldsymbol{x}) : y \in \mathbb{R}\right\}$$

are complete for $t < r < t+S-1$ and for all $\boldsymbol{x}$.

Let $\overline{P}^S$ be the probability distribution of the vector/function-valued stochastic process $\{\overline{\xi}_t^S\}$ where $\overline{\xi}_t^S = (\{W_{t+s-1}\}_{s=0}^{S}, \{F_{\eta,t+s,t}\}_{s=2}^{S-2})$ with

$$F_{\eta,r,t}(\widetilde{\eta}|\eta,x) = P\Big(\eta_{ir} \leq \widetilde{\eta}\Big|\eta_{i,r-1} = \eta, x_{ir} = x, \omega_t\Big).$$

Then, under Assumption 2, $\overline{P}^S$ is identified.

Proof. See Supplemental Appendix B. □

Under the conditions of Proposition 2, given a path for the aggregate shocks, there is a known injective mapping from the joint CDF of observables F_t^S to the latent variable CDFs $\{F_{\eta,t+s,t}\}_{s=2}^{S-2}$. This is established as in ABB using the spectral decomposition techniques of Hu and Schennach (2008) and Wilhelm (2015). The *completeness* assumption stated in part (b), a nonparametric analog of the rank conditions often used in identification arguments, is instrumental for this result. Proposition 2 then simply asserts that the researcher can determine from P^S the joint distribution $\overline{P}^S$ of those CDFs and the macro data $\{W_{t+s-1}\}_{s=0}^{S}$.

A by-product of this operation is the identification of the distribution of the transitory component CDF, $F_{\varepsilon,r,t}(\varepsilon|x) = P(\varepsilon_{ir} \leq \varepsilon\big|x_{ir} = x, \omega_t)$ for $1 \leq s \leq S-2$. Under the cross-panel consistency requirement $F_{\varepsilon,t,t} = F_{\varepsilon,t,t-1}$, the distribution of $F_{\text{init},t}(\eta|x) = P(\eta_{it} \leq \eta\big|x_{it} = x, \omega_t)$ is also identified. This is a mild condition asking the transitory income distribution of the populations of two consecutive panels to agree *within* a given period, and it is compatible with unrestricted nonstationarity *across* periods.

It is important to note that Proposition 2 gives sufficient conditions for identification, but it may still be possible to recover the object of interest under a different set of assumptions and a different value of S.[11] There are many other models of micro-level dynamics for which a large enough S enables identification of time-specific latent variable CDFs using micro data alone.[12] The challenge is linking them to the macro-level states of interest.

Proposition 3. *Let Assumptions 1, 2, 3, and the conditions of Propositions 1 and 2 hold. With* $f_{z_r|\boldsymbol{W}_t^S}$ *the density of* (Z_r, Z_{r-1}) *given* $\boldsymbol{W}_t^S = (W_t, ..., W_{t+S-1})$ *suppose that the family*

$$\mathcal{F}_{z_r|\boldsymbol{W}_t^S} \equiv \left\{ f_{z_r|\boldsymbol{W}_t^S}(\cdot|\boldsymbol{W}) : \boldsymbol{W} \in \mathbb{R}^{\dim(W_t)S} \right\}$$

is complete for some $1 < r - t < S - 1$*. Then,* Q_η *is identified.*

Proof. See Supplemental Appendix B. □

Our identification analysis clarifies the role played by each element in our framework: large cross-sections help recover the time-t short-panel distribution F_t^S; the subpanel length S balances the ability to identify latent variable distributions $\{F_{\eta,r,t}\}$ with the credibility of sample representativeness; finally, long time series of both external macro data and micro data allow the researcher to link the time-t cross-sections to the latent aggregate of interest while averaging out the impact of omitted macro factors.

4 Empirical specification and estimation

In this section we specify a flexible class of parametric models for our framework and outline a suitable estimation technique. Our empirical strategy relies on series expansions of the functions Q_η, $Q_{\varepsilon,t}$ and $Q_{\text{init},t}$ to create rich, highly nonlinear descriptions of income risk. We build on the stochastic pseudo-likelihood EM method of Arellano and Bonhomme

[11] Identification simplifies, for example, when Z_t is observable and binary.

[12] For example, if given macro shocks, y_{it} follows a canonical income process where η_{it} is a random walk, one needs $S \geq 3$. If, instead, y_{it} features heterogeneity in transitory risk (as in Chamberlain and Hirano, 1999 or Almuzara, 2020), one needs $S \geq 5$. This highlights a key advantage of time series of panels over time series of cross-sections: Observing the same unit over a number of periods makes it possible to separate permanent from transitory components directly by exploiting information on individual transitions.

(2016) and its extension to time-varying latent variables in ABB, further extending it to a long time series of short panels involving both macro and micro latent variables.

Empirical specification. For the conditional quantile function of η_{it} we set

$$Q_{\eta}(\eta, \widetilde{Z}, Z, x, u) = \psi(\eta, x)'\Theta(u)\varphi(\widetilde{Z}, Z), \tag{7}$$

where ψ, φ are vectors of known basis functions (such as orthogonal polynomials) and Θ is a matrix of linear splines with nodes $(\overline{u}_1, ..., \overline{u}_L)$ determined by the parameter vector θ. The vector θ shapes (through Q_η) the measures of persistence, aggregate exposure, dispersion and skewness we introduced in Section 2, and the IRFs and risk decompositions we will introduce in Sections 6 and 7. The task is to estimate θ.

For the transitory component and for the base-period η_{it} we set

$$Q_{\varepsilon,t}(x, u) = \psi_{\varepsilon}(x)'\Delta_{\varepsilon,t}(u), \tag{8}$$

$$Q_{\text{init},t}(x, u) = \psi_{\text{init}}(x)'\Delta_{\text{init},t}(u), \tag{9}$$

where $\psi_\varepsilon, \psi_{\text{init}}$ are vectors of known basis functions and, for each t, $\Delta_{\varepsilon,t}, \Delta_{\text{init},t}$ are vectors of linear splines with nodes $(\overline{u}_1, ..., \overline{u}_L)$ determined by the parameter vectors $\delta_{\varepsilon,t}, \delta_{\text{init},t}$. As stated in Section 2, the quantile functions $Q_{\varepsilon,t}$ and $Q_{\text{init},t}$ deliver the distribution of ε_{it} and of η_{it} at the beginning of each subpanel conditioned on time effects. Specifying $Q_{\varepsilon,t}$ and $Q_{\text{init},t}$ as flexible functions of time allows us to absorb long-term trends and other unmodeled sources of non-stationarity in the data.

Finally, to ensure the distributions implied by Q_η, $Q_{\varepsilon,t}$ and $Q_{\text{init},t}$ are supported on the real line, we propose to model the tails as conditionally exponential. For Q_η we set

$$Q_{\eta}(\eta, \widetilde{Z}, Z, x, u) = \begin{cases} Q_{\eta}(\eta, \widetilde{Z}, Z, x, \overline{u}_1) - \exp\left(\psi_{\text{lo}}(\eta, \widetilde{Z}, Z, x)'\theta_{\text{lo}}\right)\ln\left(\frac{\overline{u}_1}{u}\right) & \textit{if } u < \overline{u}_1 \\ Q_{\eta}(\eta, \widetilde{Z}, Z, x, \overline{u}_L) + \exp\left(\psi_{\text{up}}(\eta, \widetilde{Z}, Z, x)'\theta_{\text{up}}\right)\ln\left(\frac{1-\overline{u}_L}{1-u}\right) & \textit{if } u > \overline{u}_L \end{cases}$$

where $\psi_{\text{lo}}, \psi_{\text{up}}$ are vectors of known basis functions and $\theta_{\text{lo}}, \theta_{\text{up}}$ are unknown parameters included in θ. We adopt a similar specification for $Q_{\varepsilon,t}$ and $Q_{\text{init},t}$.

An arbitrarily good approximation to the true Q_η, $Q_{\varepsilon,t}$ and $Q_{\text{init},t}$ can be obtained by a suitable choice of basis functions for sufficiently regular classes as in nonparametric series methods. Instead, we adopt a flexible parametric perspective in which θ, $\delta_{\varepsilon,t}$ and $\delta_{\text{init},t}$ are finite-dimensional. While the identification argument of Section 3 covers nonparametric models, the precision with which the objects of interest can be estimated is limited by the time series length T, or more concretely, by the scarcity of recessions and booms in the sample. In that context, our approach balances flexibility with statistical power.

Moments. Our model implies complete-data moments involving $\bar{y}_{it}^S = \{y_{i,t+s}, x_{i,t+s}\}_{s=0}^{S-1}$, $\bar{\eta}_{it}^S = \{\eta_{i,t+s}\}_{s=0}^{S-1}$ and $\bar{Z}_t^S = \{Z_{t+s}\}_{s=0}^{S-1}$ that pin down the true θ and $\delta_t = (\{\delta_{\varepsilon,t+s}\}_{s=0}^{S-1}, \delta_{\text{init},t})$,

$$E\Big[m_\theta \left(\theta_0; \bar{y}_{it}^S, \bar{\eta}_{it}^S, \bar{Z}_t^S\right) \Big] = 0, \qquad E\Big[m_\delta \left(\delta_{0t}; \bar{y}_{it}^S, \bar{\eta}_{it}^S\right) \Big] = 0. \tag{10}$$

These are moments from quantile and exponential regressions (see Supplemental Appendix C.1), and they are unfeasible as they depend on macro and micro latent variables.

Pseudo posteriors. To operationalize our approach, we need to transform the unfeasible moments (10) into feasible moments that depend only on observable data. We do so in two steps by sequentially integrating out the unobserved latent variables against a convenient choice of pseudo posterior distributions. Specifically, in a first step we integrate η_{it} out with respect to its (unit-level) posterior density given the micro data y_{it} and x_{it} and the aggregate state Z_t. In a second step, we integrate Z_t out with respect to its aggregate posterior density given W_t. Proceeding in this way has the advantage of clarifying the role of micro and macro components for estimation, in analogy to our identification analysis. It also leads to a tractable numerical implementation that draws on known algorithms from microeconometric and macroeconometric traditions.

Let f be a generic probability density function (PDF). In a first step, we define the (still unfeasible) partial-data moments, for hypothetical values $\theta, \theta', \delta_t, \delta_t'$,

$$\overline{m}_\theta\left(\theta; \theta', \delta_t', \bar{y}_{it}^S, \bar{Z}_t^S\right) = \int m_\theta\left(\theta; \bar{y}_{it}^S, \bar{\eta}^S, \bar{Z}_t^S\right) f\left(\bar{\eta}^S \middle| \bar{y}_{it}^S, \bar{Z}_t^S, \theta', \delta_t'\right) d\bar{\eta}^S,$$
$$\overline{m}_\delta\left(\delta_t; \theta', \delta_t', \bar{y}_{it}^S, \bar{Z}_t^S\right) = \int m_\delta\left(\delta_t; \bar{y}_{it}^S, \bar{\eta}^S\right) f\left(\bar{\eta}^S \middle| \bar{y}_{it}^S, \bar{Z}_t^S, \theta', \delta_t'\right) d\bar{\eta}^S.$$

The conditional density $f\left(\bar{\eta}^S \middle| \bar{y}_{it}^S, \bar{Z}_t^S, \theta, \delta_t\right)$ is the individual micro-level posterior from ABB but conditioning on aggregate states and time effects. It is fully determined via Bayes rule through our parametric model, and it can be efficiently sampled from using Sequential Monte Carlo techniques (Arellano et al., 2023, Section 4.3).

By (10) and the law of iterated expectations,

$$E\left[\overline{m}_\theta\left(\theta_0;\theta_0,\delta_{0t},\bar{y}_{it}^S,\bar{Z}_t^S\right)\right]=0,\qquad E\left[\overline{m}_\delta\left(\delta_{0t};\theta_0,\delta_{0t},\bar{y}_{it}^S,\bar{Z}_t^S\right)\right]=0,\tag{11}$$

which indicates the partial-data moments (11) provide valid restrictions on θ and δ_t.

Let λ be a vector containing the macro parameters. Write $\overline{Y}_t^S=\{\bar{y}_{it}^S\}_{i\in\mathcal{I}_t}$ and $\overline{W}=\{W_t\}_{t=1}^{T+S}$. In a second step, we define the aggregated observed-data moments,

$$\overline{M}_\theta\left(\theta;\theta',\delta_t',\lambda',\overline{Y}_t^S,\overline{W}\right)=\int\left[\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}\overline{m}_\theta\left(\theta;\theta',\delta_t',\bar{y}_{it}^S,\bar{Z}_t^S\right)\right]f\left(\bar{Z}^S\middle|\overline{W},\lambda'\right)\,d\bar{Z}^S,$$
$$\overline{M}_\delta\left(\delta_t;\theta',\delta_t',\lambda',\overline{Y}_t^S,\overline{W}_t\right)=\int\left[\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}\overline{m}_\delta\left(\delta_t;\theta',\delta_t',\bar{y}_{it}^S,\bar{Z}^S\right)\right]f\left(\bar{Z}^S\middle|\overline{W},\lambda'\right)\,d\bar{Z}^S.$$

The conditional density $f\left(\bar{Z}^S\middle|\overline{W},\lambda\right)$ is the smoothing posterior from model (3), and it can be efficiently sampled from using the Kalman filter. Again, by (11) and iterated expectations,

$$E\left[\overline{M}_\theta\left(\theta_0;\theta_0,\delta_{0t},\lambda_0,\overline{Y}_t^S,\overline{W}\right)\right]=0,\qquad E\left[\overline{M}_\delta\left(\delta_{0t};\theta_0,\delta_{0t},\lambda_0,\overline{Y}_t^S,\overline{W}\right)\right]=0.\tag{12}$$

Functions $\overline{M}_\theta$ and $\overline{M}_\delta$ give restrictions on θ and δ_t (given λ) that only depend on observables.[13] We rely on the sample counterpart of the moments in (12) for estimation.

Stochastic EM implementation. The main challenge in exploiting (12) for estimation is integrating the primitive moments against the pseudo posteriors of η_{it} and Z_t. We follow Arellano and Bonhomme (2016) and adopt a simulation-based approach. Let $\widehat{\lambda}$ be macro parameter estimates obtained from the macro data $\overline{W}$ alone (e.g., maximum likelihood). We rely on the following stochastic EM algorithm to estimate θ and $\{\delta_t\}_{t=1}^T$, which iterates between simulation smoothing of macro and micro latent variables and running quantile and exponential regressions.

Algorithm 1. *Initialize parameters* $\widehat{\theta}^{(0)}$ *and* $\{\widehat{\delta}_t^{(0)}\}_{t=1}^T$. *For* $j=1,\ldots,J$, *iterate between the following:*

1) Pseudo-Stochastic E step:

[13] This is in the spirit of the unbiased likelihood estimate used by Liu and Plagborg-Møller (2023, Section 3.1) in a full-information Bayesian setup. Our approach is based on pseudo likelihood functions but allows for panel data, micro-level latent variables and potentially omitted aggregate shocks.

(i) *draw* $\overline{Z}(j) = \{Z_t(j)\}_{t=0}^{T+S}$ *from the macro posterior* $f\left(\overline{Z}\middle|\overline{W}, \widehat{\lambda}\right)$,

(ii) *independently over units i and subpanels t, draw* $\bar{\eta}_{it}^{S}(j) = \{\eta_{i,t+s}(j)\}_{s=0}^{S-1}$ *from the micro posterior* $f\left(\bar{\eta}_{it}^{S}\middle|\bar{y}_{it}^{S}, \bar{Z}^{S}(j), \widehat{\theta}^{(j-1)}, \widehat{\delta}_t^{(j-1)}\right)$.

2) *Pseudo M step:*

(i) *update the parameters to* $\widehat{\theta}^{(j)}$ *and* $\{\widehat{\delta}_t^{(j)}\}_{t=1}^{T}$ *by quantile and exponential regressions treating* $\left\{\{\{\eta_{i,t+s}(j), y_{i,t+s}, x_{i,t+s}, Z_{t+s}(j)\}_{s=0}^{S-1}\}_{i\in\mathcal{I}_t}\right\}_{t=1}^{T}$ *as data.*

For some $\mu \in (0,1)$, *set* $\widehat{\theta} = (\mu J)^{-1}\sum_{j=(1-\mu)J}^{J}\widehat{\theta}^{(j)}$ *and* $\widehat{\delta}_t = \sum_{j=(1-\mu)J}^{J}\widehat{\delta}_t^{(j)}$.

We present algorithms to perform steps 1(i) and 1(ii) in Supplemental Appendix C.2.

Our approach departs from full-information likelihood estimation in two directions. First, we use complete-data moments from quantile and exponential regressions instead of solving the score equations from the complete-data likelihood. This carries a significant computational simplification as quantile and exponential regressions are fast and stable to run, compared to the mostly intractable score equations of the model. Second, rather than smoothing latent variables $\{\{\bar{\eta}_{it}^{S}\}_{i\in\mathcal{I}_t}, Z_t\}_{t=1}^{T}$ using their full joint posterior given $\{\overline{Y}_t^{S}\}_{t=1}^{T}$ and $\overline{W}$, we use only certain slices of the posterior (akin to composite-likelihood methods): e.g., we integrate unit-i micro latent variables $\bar{\eta}_{it}^{S}$ conditioning on unit-i data $\bar{y}_{it}^{S}$ rather than the full microdata $\{\overline{Y}_t^{S}\}_{t=1}^{T}$. This comes potentially at a cost in terms of asymptotic efficiency but it has the advantage that the micro-level pseudo posteriors do not require us to model the cross-sectional dependence induced by the omitted aggregate factors G_t.

Even though our focus is on income risk, it is important to highlight that this method offers a flexible, general-purpose way to estimate rich models of micro-level latent variables subject to macro shocks, and can be applied far more generally.

Large-sample properties. We next summarize statistical properties of $\widehat{\theta}$ and functions of $\widehat{\theta}$ which cover the objects in Section 2.2, and the IRFs and risk calculations we discuss later. See Supplemental Appendix C.3 for further discussion. Our asymptotic analysis assumes $N_t, T \to \infty$ with S fixed, and $T/N_t \to 0$. That is, the number of individuals in each subpanel and the number of subpanels is large while the length of each subpanel is small, reflecting our identification argument. We also assume that the time-series dimension is small relative to the cross-sectional dimension, as in our empirical setting.

Given the macro parameter estimate $\widehat{\lambda}$, Algorithm 1 delivers $\widehat{\theta}$ and the sequence $\{\widehat{\delta}_t\}_{t=1}^{T}$. In our empirical analysis, we are interested in objects that can be expressed as a smooth function $\gamma(\theta)$ of θ, for which it is natural to use the plug-in estimator $\widehat{\gamma} = \gamma(\widehat{\theta})$. Recall that θ_0 denotes the true value of θ and let $\gamma_0 = \gamma(\theta_0)$. Our main result is as follows:

Proposition 4. *Let Assumptions 1, 2, 3 hold with (7), (8), (9). As $N_t, T \to \infty$ with $T/N_t \to 0$,*

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{\text{d}} N\left(0, \Sigma_{\theta_0}\right) \text{ and } \sqrt{T}\left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{\text{d}} N\left(0, J_{\theta_0}\Sigma_{\theta_0}J'_{\theta_0}\right),$$

where Σ_{θ_0} is symmetric and positive semi-definite, and J_{θ_0} is the Jacobian of γ evaluated at θ_0.

Proof. See Supplemental Appendix C.3. □

Proposition 4 implies that the precision of the estimator is determined by T, not by N_t. The main reason is that identification of θ relies on time series variation to average out the effect of the unobserved factors G_t.[14] Instead, the role of cross-sectional variation is to help control the error coming from the estimation of the time-varying parameters δ_t.

Empirical implementation. For Q_η in (7), we model ψ as a third-order polynomial in η combined with a second-order polynomial in x, and we construct φ as a restricted second-order polynomial in $(\widetilde{Z}, Z)$: we exclude the linear term from interactions between η and x, and include the quadratic term only in the intercept. We set ψ_{lo} and ψ_{up} to second-order polynomials in η, x, $\widetilde{Z}$ and Z without interactions. Regarding $Q_{\varepsilon,t}$ in (8) and $Q_{\text{init},t}$ in (9) we allow their intercepts and tail coefficients to change over time. We also include in $Q_{\text{init},t}$ a time-invariant second-order polynomial in x. The rank-space grid size is set to $L = 11$.

Inference. To gauge the statistical precision of our estimates below we report confidence intervals computed by a parametric bootstrap technique that exploits our model structure while accounting for (i) cross-sectional dependence induced by omitted aggregate factors G_t and (ii) unit-level dependence induced by the sampling design of the PSID. The specifics are discussed in Supplemental Appendix C.4.

5 Macro shocks and nonlinear income risk

We now present our empirical analysis of the effects of macro shocks on nonlinear income risk. We discuss the macro data and business cycle state in Section 5.1. This is followed by a description of our main source of micro data, the PSID, in Section 5.2. We then turn to

[14]The convergence rate would be the same if Z_t were observable (e.g., the NBER recession indicator).

the quantification of aggregate effects on measures of nonlinear persistence (Section 5.3), exposure to aggregate shocks (Section 5.4), and conditional skewness (Section 5.5).

5.1 Macro data and the aggregate state

We extract the aggregate state Z_t from a vector of macro observables W_t informative about the business cycle consisting of GDP, consumption, investment, the unemployment rate and hours worked. The data are quarterly and cover 1960Q1–2019Q4. GDP, consumption, investment and hours are measured in log per-capita terms while the unemployment rate is multiplied by -1 to make it procyclical as the other entries of W_t. All series are detrended by removing a two-sided 40-quarter low-pass filter from them.[15]

For estimation, we normalize the loading on GDP to one. Therefore, the business cycle process Z_t is measured as a log deviation from its per-capita growth trend. We specify Z_t and each entry of e_t as independent AR(2) processes and estimate the macro parameters λ by Gibbs sampling using a flat prior. Supplemental Appendix C.2 discusses the details. Figure 1 shows the time series evolution of W_t and estimates of Z_t.

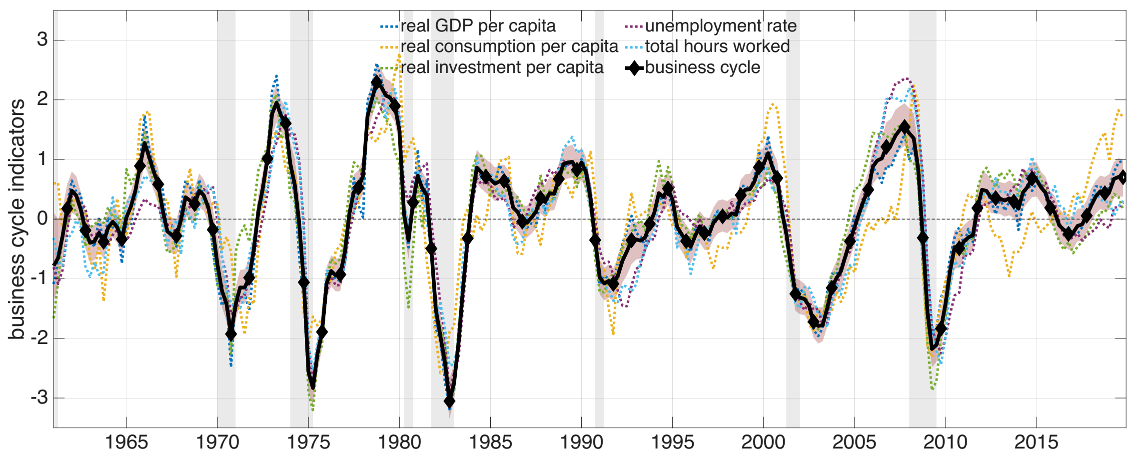

FIGURE 1. Business cycle indicators and estimates of Z_t

Note: We plot real GDP, consumption and investment per capita, the unemployment rate and hours worked alongside our estimate of the business cycle factor Z_t. We standardize them to mean zero and unit variance. Red shaded areas are 90% pointwise probability bands and gray areas indicate NBER-dated recessions.

Three takeaways from the figure are as follows. First, and not surprisingly, the variables included in W_t show strong comovements. As a result, filtering out Z_t is statistically easy,

[15]The data come from Federal Reserve Economic Data (FRED). To construct W_t we took from Angeletos et al. (2020) the subset of variables most informative about the MBC shock, although our results are robust to alternative specifications that expand W_t with other macro indicators.

as evidenced by the tight 90% probability bands drawn around the estimate. Second, there is substantial cyclical variation in Z_t, and this remains true when we consider temporally aggregated versions of it. For example, Figure 1 marks with a diamond the estimates of Z_t for the fourth quarter of each year. They clearly retain a meaningful share of the aggregate variation we see at the quarterly frequency. This is important because income data at the micro level typically refer to annual income. Third, there is evidence of some asymmetry between recessions and expansions, particularly in the last 40 years: Recessions are deep and occur suddenly, while expansions are mild and unfold gradually. It would be possible to enrich model (3) to accommodate that feature, although this is unlikely to impact our estimates of Z_t given their precision. Nonetheless, in computing summaries that condition on the aggregate state we will adopt the notion of a typical recession, $\tilde{Z}_r = -2\sigma_Z$, a steady state, $\tilde{Z}_{ss} = 0$, and a typical expansion, $\tilde{Z}_e = \sigma_Z$, where $\sigma_Z^2 = \text{Var}(Z_t)$.

In this context, a central object is the IRF of the macro state variable Z_t to the shock V_t. We will return to it when discussing IRFs more generally in Section 6.2.[16]

5.2 Micro data and cyclical patterns in the PSID

As we outlined in Section 3, the micro data for our study takes the form of a time series of panels. We draw to this end from the long history of panel data on income and earnings available in the Panel Study of Income Dynamics (PSID).

Established in 1968, the PSID initially surveyed a nationally representative set of U.S. households. Ever since, it has followed those families and, as their children come of age and form independent households, incorporated those new units into the panel. With the periodic addition of refresher and immigrant samples, the PSID design aims at capturing the process of household formation and dissolution in the U.S. economy over time. It also accords with our notion of time series of panels in Section 3 as it allows us to form a long sequence of short subpanels, each reflecting representativeness at a point in time, which taken together span a rich history of aggregate fluctuations.

Interviews were conducted annually between 1968 and 1997, and biennially after 1999. Whether annual or biennial, the year-k interview (which typically occurs between March and November) asks the household to report annual income for year $k-1$. We make use of all available waves beginning in 1970 and ending in 2019: we exclude 1968 and 1969 as

[16] We also discuss IRFs of W_t to V_t and compare them to the IRFs estimated in Angeletos et al. (2020) for the MBC shock in Supplemental Appendix E.3.

some income and demographic variables were not collected in those waves, and we stop at 2019 because of the COVID-19 pandemic. We then form a time series of panels where each subpanel has $S = 4$ biennial waves, although, crucially, we exploit all years. We use biennial panels because we are constrained by the change in data collection frequency after 1999. This also has the advantage of making the assumption of serially uncorrelated ε_{it} more plausible. But nonetheless, prior to 1999, the panels we construct have base years which cover both even and odd years, so we make use of all available annual observations. Given the biennial nature of the subpanels, $t - 1$ below is understood as two years before t, and with a slight abuse of notation we denote by Z_t the value of the macro state in the fourth quarter of t. This is relevant to interpret the income process and summaries.

In our data selection, a household is so that the representative person is male, married and aged 25 to 60, and so that income is positive.[17] We consider three different measures of income: male earnings, household earnings, and disposable income. Male earnings are the labor income of the household head. Household earnings add in the labor income of the spouse. Disposable income adds transfers and subtracts federal taxes. We compute taxes with the tax functions estimated by Borella, De Nardi, Pak, Russo, and Yang (2023) and deflate everything to 2016 dollars using the consumer price index (CPI-U).[18]

We construct y_{it} by residualizing log income against the following covariates interacted with a quadratic time trend: education, race, family size, number of children, an indicator for the presence of dependents, and state of residence. Except for family size and number of children, the other variables are treated as categorical. We then add back to all units the same constant representing a household with typical covariates.[19] We also exclude age from this operation which, as anticipated, enters Q_η and Q_{init,t_0} in the form of x_{it}.[20]

This is the data for our analysis. Before we delve into it, we give a descriptive account of cyclical patterns in the PSID using similarly constructed datasets with $S = 1$ and $S = 2$.

Descriptives. Figure 2 offers a first look at the micro data using repeated cross-sections ($S = 1$). The period covers seven NBER recessions (the two downturns of the 80s merged

[17] In addition to the main sample representative of the U.S. population, the PSID maintains a low-income sample known as the Survey of Economic Opportunities. Our analysis makes use only of the former.

[18] Beyond partialling out state-of-residence dummies, we do not explicitly account for differences in state taxes but one possibility is to use the tax functions in Fleck, Heathcote, Storesletten, and Violante (2025).

[19] This is a white college graduate with family size four, two children, no dependents, who resides in the state of New York. This level shift has no impact on the dynamics of η_{it} or any of the summaries of interest.

[20] Including x_{it} as an argument of $Q_{\varepsilon,t}$ does not affect our empirical results.

into a single bar). Panel (a) shows the fraction of zeros that would be dropped to compute y_{it}. It shows that most of the time less than 2% of the sample reports zero annual household earnings (blue). This applies to recessions too, except for the Great Recession in 2007 when the fraction briefly exceeded 2%. The story is different for male earnings, for which zeros constitute a larger fraction and there seems to be a steeper upward trend over time.

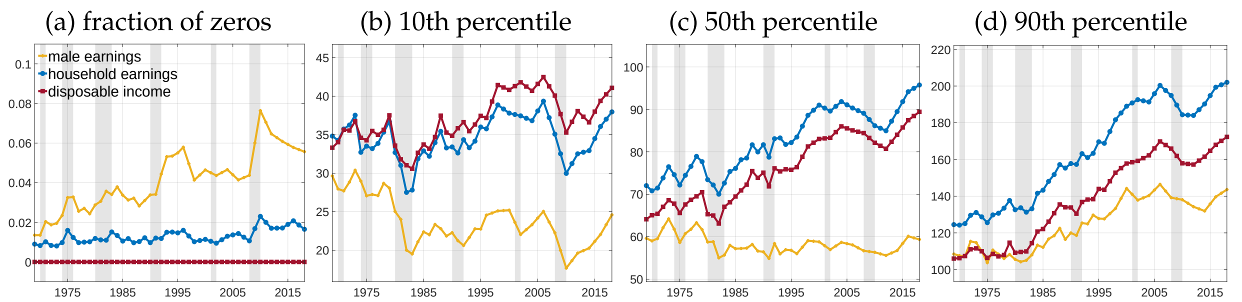

FIGURE 2. Distributional characteristics of income measures over time

Note: We show the fraction of zeros and the 10th, 50th and 90th percentiles (conditional on positive income) for our income measures in thousands of 2016 dollars. Shaded areas are NBER recessions.

Given the demographic criteria in our data selection, practically no unit reports zero disposable income for a whole year. Since this is also the relevant measure for studying the effects of aggregate risk on consumption and welfare, disposable income will be our primary focus below. We report results for male and household earnings for comparability with the caveat that doing so abstracts from the extensive margin. To the extent that this margin is cyclical and recessions imply larger drops in labor income than what we observe, our results for male and household earnings should be taken as a lower bound on the impact of negative macro shocks along those margins.

Panels (b) to (d) in Figure 2 illustrate the evolution of low, middle and high incomes over time. One highlight is that there are divergent secular trends in the income distribution, with fast growth at the top and stagnation (or even decline) at the bottom. These trends have been extensively studied in the literature, usually linking them to changes in female labor force participation, structural transformation, reforms to the tax and social security systems, and other phenomena.[21] But importantly for us, these trends coexist with sizable cyclical variation. A quick glance reveals that incomes fall in recessions in ways that are heterogeneous over the income distribution (more at the bottom than in the middle), across

[21]In our framework of analysis, part of these low-frequency changes are controlled for by netting out the effect of covariates interacted with time trends in the construction of y_{it}, while another part will be absorbed by the time-varying functions Q_{init,t_0} and $Q_{\varepsilon,t}$.

income measures (more for male earnings than for disposable income) and depending on the severity of the downturn (the double-dip recession of the 80s and the Great Recession being the worse). Figure 2 is also indicative of the redistributive role of taxes and transfers, as going from household earnings to disposable income pushes low incomes up and higher incomes down, and of the insurance role of both spousal income and taxation which tend to mitigate income losses during recessions.

Next, Figure 3 uses a time series of biennial panels ($S = 2$) to measure the evolution of conditional skewness from flexible quantile autoregressions of y_{it}. Skewness is measured as in (6) but period-by-period (instead of conditioning on Z_t) and for y_{it}, not for η_{it}.

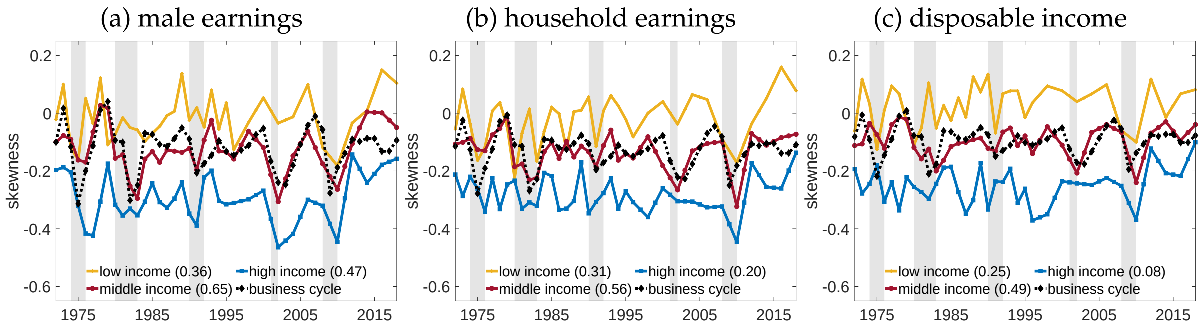

FIGURE 3. Conditional skewness over the business cycles.

Note: We report conditional skewness from quantile regressions of y_{it} on a third-order polynomial of past income. Low, middle and high income refer to the 10th, 50th and 90th percentile of the income distribution. The business cycle indicator is normalized to have the mean and variance of middle-income skewness.

The figure tells a tale of two skewnesses. On the one hand, skewness is countercyclical (as in Guvenen et al., 2014). On the other, it markedly decreases with income at any point in time (as in ABB), and differentially so depending on the phase of the business cycle. We see these patterns at play in all three income measures, although the cyclical movements in skewness become less pronounced when we add spousal income and taxation, as seen in the correlation with Z_t reported in parenthesis in the legend of each panel.

The descriptive analysis presented so far points to salient cyclical features in income data. These are consistent with heterogeneous nonlinear transmission of micro and macro shocks which our framework will allow us to quantify. We undertake this task next.

Model fit. We fit our model to the macro and micro data with the empirical specification of Section 4, where we also explain how estimation and inference are implemented.

To assess goodness of fit, Figure 4 compares a selection of data summaries (red) with their counterparts in our model obtained by simulation (blue) for (a) the level of income,

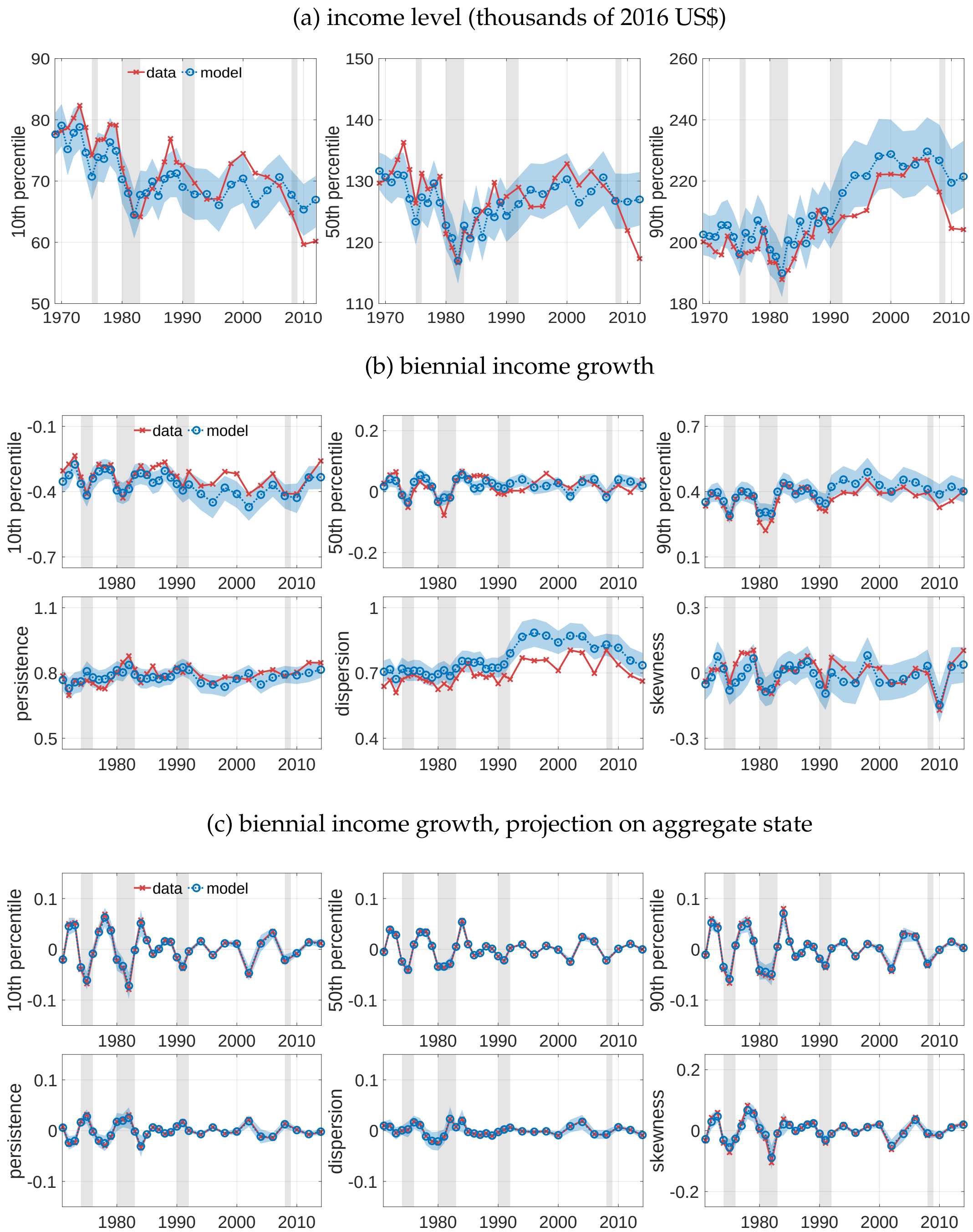

FIGURE 4. Model fit assessment (disposable income).

Note: Panel (a) compares data (red) and model (blue) implications for the percentiles of the level of income in thousands of 2016 dollars ($e^{y_{it}}/1000$). Panel (b) compares percentiles and measures of persistence, dispersion and skewness for income growth Δy_{it} while panel (c) reports those objects projected on (Z_t, Z_{t-1}) net of an intercept and time trend. Model outputs are obtained from 1,000 simulated samples where we draw shocks accounting for cross-sectional and unit-level dependence; shaded areas are 90% probability bands.

(b) income growth, and (c) their cyclical component. Although there are clear trends in the distributions of the level and growth rate of income which are not related to the business cycle, our model tracks their evolution over time closely. The model is especially good at matching the dynamics of income growth persistence and skewness in panel (b) although it slightly overstates the dispersion.[22] Reassuringly, when we project these summaries on the business-cycle state in panel (c), data and model coincide, even for the dispersion. The main takeaways from this comparison, done here for disposable income, also apply to male and household earnings, and they extend to income growth over longer horizons (omitted for economy of space).[23]

5.3 Nonlinear persistence: ρ tilts with the aggregate state

We now turn to quantifying empirically the effect of aggregate shocks on different aspects of nonlinear income risk. We begin by discussing the persistence of past income histories and how it changes over the business cycle. To that end, we use the measure of nonlinear persistence $\rho(u, \eta, Z_t, Z_{t-1}, x)$ defined in Equation (4) of Section 2.

The left panels of Figure 5 show our estimates of $\rho(u, \eta, Z_t, Z_{t-1}, x)$ for the three measures of income we consider. We plot persistence by quantiles of the current shock $u = u_{it}$ and past persistent income $\eta = \eta_{i,t-1}$ where age $x = x_{it}$ is averaged out, $Z_{t-1} = \tilde{Z}_{ss}$ is held at its steady state value and Z_t takes on different values representing a typical recession $\tilde{Z}_r$ (red), the stead state $\tilde{Z}_{ss}$ (yellow) or a typical mild expansion $\tilde{Z}_e$ (blue). On the right panels, we report the recession minus the expansion persistence. If the recession persistence is above (below) its expansion counterpart according to a one-sided pointwise test at the 5% significance level, we indicate that with magenta (cyan).

We highlight two takeaways from Figure 5. First, it confirms the nonlinear persistence pattern uncovered in ABB over a much longer history and across income measures. For example, in the steady state surface for disposable income, the average persistence is 0.92. It falls to 0.85 for a unit in the 90th η-percentile hit with a bad shock $u = 0.1$ and to 0.60 for a unit in the 10th η-percentile hit with a good shock $u = 0.9$. As already discussed, since ρ is a measure of how closely related current and past incomes are, this captures the fact

[22] We measure persistence by the coefficient on past income in a quantile regression of y_{it} on $y_{i,t-\ell}$ including an intercept. This corresponds to the measure (4) in a linear quantile autoregression.

[23] Our calculations show that the permanent-transitory specification for y_{it} is key to fit the persistence of long-run income growth. A model with no transitory component understates the persistence at long horizons.

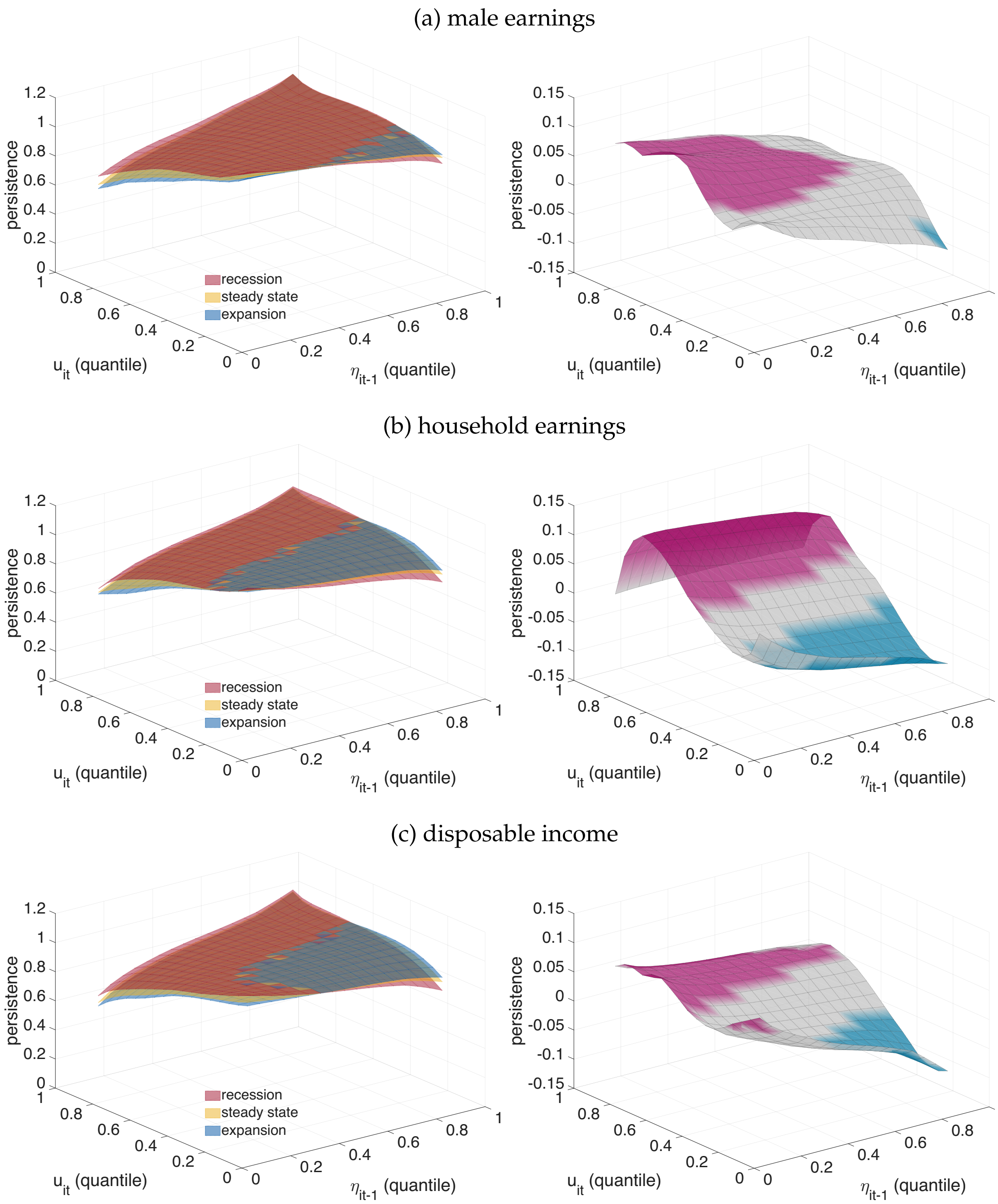

FIGURE 5. Nonlinear persistence.

Note: The left panels report the persistence measure $\rho(u, \eta, Z_t, Z_{t-1}, x)$ by quantile of the shock $u = u_{it}$ and past persistent income $\eta = \eta_{i,t-1}$. Here, age $x = x_{it}$ is averaged out, $Z_{t-1} = \tilde{Z}_{ss}$ and Z_t is a recession $\tilde{Z}_r$, the steady state $\tilde{Z}_{ss}$ or an expansion $\tilde{Z}_e$ (see Section 5.1). The right panels show the gap in persistence between recession and expansion, $\rho(u, \eta, \tilde{Z}_r, \tilde{Z}_{ss}, x) - \rho(u, \eta, \tilde{Z}_e, \tilde{Z}_{ss}, x)$. The difference is painted magenta (cyan) if it is statistically positive (negative) at the 5% significance level (one-sided test).

that a big shock of a given direction may sometimes erase an entire income history.

The second takeaway is that the ABB persistence pattern is itself macro state-dependent, as aggregate shocks *tilt* the estimated persistence surface. Comparing recessions and expansions, ρ increases for low-η and decreases for high-η units (particularly those affected by bad shocks) during a downturn. In words, a low-η household has it more difficult to leave the low-income state in the midst of a contraction, while a high-η household finds it harder to remain high-income. These effects are large both statistically and economically. For disposable income, ρ decreases by 0.09 for the 90th η-percentile with $u = 0.1$ and increases by 0.07 for the 10th η-percentile with $u = 0.9$ as we move from $Z_t = \tilde{Z}_e$ to $Z_t = \tilde{Z}_r$.

Some additional insights are obtained by comparing disposable income with male and household earnings. For our earnings measures, the increase in persistence at low incomes during recessions tends to be bigger than for disposable income. This fact is suggestive of the insurance role of taxes and transfers in attenuating the impact of negative aggregate shocks for the left tail of the income distribution, consistent with the descriptive account in Section 5.2—although the comparison abstracts from extensive margin effects.

5.4 Exposure to aggregate shocks: β is countercyclical

A second key aspect of nonlinear income risk is the persistent income exposure to macro shocks, measured by the nonlinear aggregate exposure coefficient $\beta(u, \eta, Z_t, Z_{t-1}, x)$ given in Equation (5) of Section 2. Figure 6 presents our estimates.

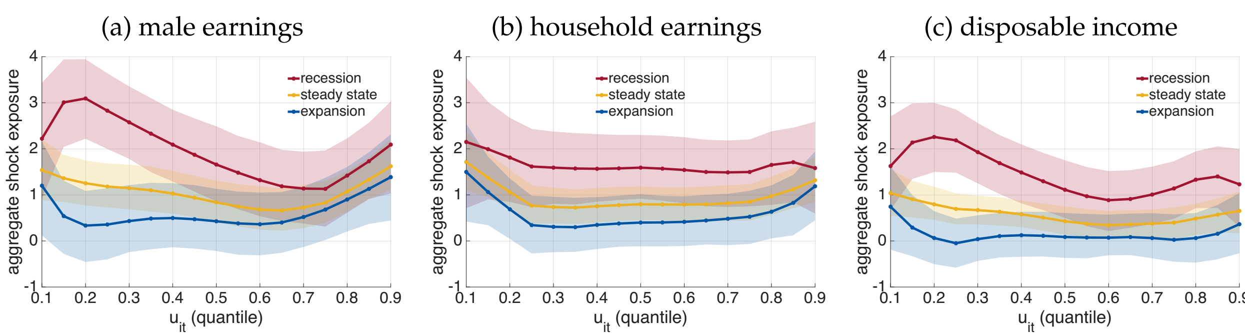

FIGURE 6. Nonlinear exposure to aggregate shocks.

Note: We report the aggregate risk exposure $\beta(u, \eta, Z_t, Z_{t-1}, x)$ by quantile of the shock $u = u_{it}$ averaged over persistent income $\eta = \eta_{i,t-1}$. Age $x = x_{it}$ is averaged out, $Z_{t-1} = \tilde{Z}_{ss}$ and Z_t is a recession $\tilde{Z}_r$, the steady state $\tilde{Z}_{ss}$ or an expansion $\tilde{Z}_e$ (see Section 5.1). Shaded areas represent 90% confidence bands.

In Figure 6, we show $\beta(u, \eta, Z_t, Z_{t-1}, x)$ as a function of the micro-level shock $u = u_{it}$ that occurs at the time of the macro shock. We average past income $\eta = \eta_{i,t-1}$ and age $x = x_{it}$, and fix $Z_{t-1} = \tilde{Z}_{ss}$, allowing Z_t to take values compatible with a recession $\tilde{Z}_r$, steady state $\tilde{Z}_{ss}$ or expansion $\tilde{Z}_e$ (discussed in Section 5.1).

Three patterns stand out. First, the exposure to aggregate shocks depends on Z_t and is countercyclical. Averaged over u and η, in our disposable income calculations, β is 1.3 in a recession, 0.6 in steady state and 0.2 in an expansion. Given the normalization of Z_t in Section 5.1, these numbers can be interpreted as the elasticity of persistent income to an aggregate shock that implies a one percentage point change in GDP per capita relative to its trend. Accordingly, a negative macro shock leads to an aggregate reduction in the persistent component of disposable income that is more or less than one-for-one the fall in GDP depending on whether the economy is already in a recession when the shock hits. This is a major form of aggregate state-dependence and one that is ruled out by models in which aggregate shocks enter additively. What is more, we argue in Section 7 that this nonlinearity plays a key role in the cost of business cycle risk.

Second, the aggregate exposure coefficient varies across income measures, with male earnings the most and disposable income the least sensitive to Z_t. For example, the recession average β is 1.8 for male earnings, 1.6 for household earnings and 1.3 for disposable income. The same applies to the steady state and expansion β, and at different quantiles of $u = u_{it}$ and $\eta = \eta_{i,t-1}$, as seen in additional results in Supplemental Appendix D.

The last pattern to analyze is the interaction between macro and micro shocks. Units affected by bad micro shocks $u = u_{it}$ at the time when the macro shock occurs are relatively more exposed during recessions (and generally less during expansions) than units subject to neutral or good shocks. In other words, the consequences of recessions are not evenly distributed but are borne mostly by those who experience bad micro luck as the downturn unfolds. This is another feature ruled out by linear models with aggregate shocks.

5.5 A tale of two skewnesses

We conclude by discussing the conditional skewness measure introduced in Equation (6) of Section 2. A full picture of aggregate effects on income risk is completed by measures of dispersion and kurtosis (footnote 8). They appear to be less cyclical than the skewness (in line with Guvenen et al., 2014), and we report them in Supplemental Appendix D.

The upper panels of Figure 7 show $\text{sk}(\eta, Z_t, Z_{t-1}, x)$ for different quantiles of $\eta = \eta_{i,t-1}$ with $Z_{t-1} = \tilde{Z}_{ss}$ when Z_t ranges from recession $\tilde{Z}_r$, through steady state $\tilde{Z}_{ss}$ to expansion $\tilde{Z}_e$. Age $x = x_{it}$ is averaged out. The lower panels display the difference in skewness between recession and expansion including 90% pointwise confidence bands.

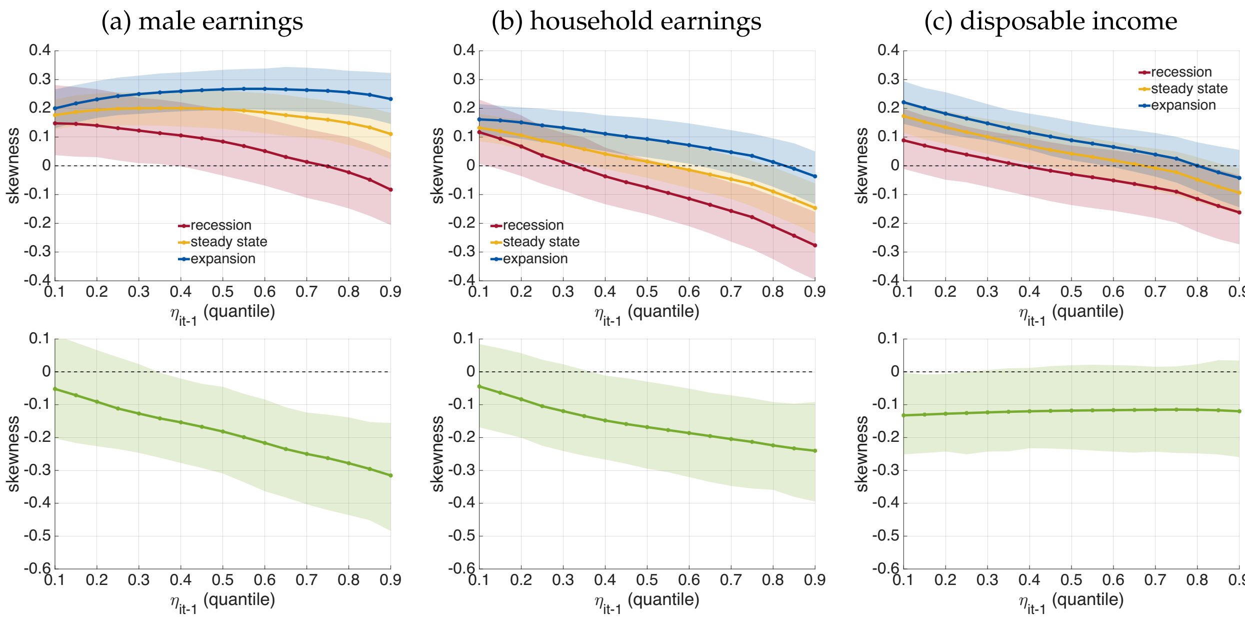

FIGURE 7. Conditional skewness.

Note: The upper panels report skewness $\text{sk}(\eta, Z_t, Z_{t-1}, x)$ by past persistent income $\eta = \eta_{i,t-1}$ where age $x = x_{it}$ is averaged out, $Z_{t-1} = \tilde{Z}_{ss}$ and Z_t is a recession $\tilde{Z}_r$, the steady state $\tilde{Z}_{ss}$ or an expansion $\tilde{Z}_e$ (see Section 5.1). The lower panels show the gap in skewness between recession and expansion, $\text{sk}(\eta, \tilde{Z}_r, \tilde{Z}_{ss}, x) - \text{sk}(\eta, \tilde{Z}_e, \tilde{Z}_{ss}, x)$. Shaded areas represent 90% confidence bands.

Figure 7 provides empirical support for the tale of two skewnesses suggested by our descriptive analysis (Figure 3 in Section 5.2). The ABB skewness pattern, by which $\text{sk}(\eta, Z_t, Z_{t-1}, x)$ decreases for higher η, is accompanied here by a cyclical skewness pattern: Recessions tend to shift $\text{sk}(\eta, Z_t, Z_{t-1}, x)$ toward the negative plane at all levels of η, often by a large margin. Mirroring nonlinear persistence and exposure to aggregate shocks, cyclical shifts in skewness are strongest for male earnings and weakest for disposable income, that is, after spousal earnings and the tax-transfer system are taken into account.

Since $\text{sk}(\eta, Z_t, Z_{t-1}, x)$ is a measure of the relative strength of upside and downside income risks, our results characterize recessions as periods when downside risk becomes more prevalent in a generalized but heterogeneous way across the income distribution.

6 Impulse response analysis

In this section, we develop new methodology for measuring impulse responses to macro and micro shocks. We first present our approach and then discuss empirical estimates.

6.1 Measuring the propagation of macro and micro shocks

To develop the idea, we focus on the persistent component η and omit covariates x. From Equations (1) and (3) we get by recursive substitution the representation

$$\eta_{i,t+h} = q_{\eta,h}\left(\boldsymbol{u}_{it}^{h}, \boldsymbol{V}_{t+1}^{h-1}, \eta_{i,t-1}, Z_t, Z_{t-1}\right), \quad h = 0, 1, \ldots, \tag{13}$$

where $\boldsymbol{u}_{it}^{\ell} = (u_{it}, \ldots, u_{i,t+\ell})$ and $\boldsymbol{V}_{t}^{\ell} = (V_t, \ldots, V_{t+\ell})$. With the distribution of micro and macro shocks, Equation (13) determines the predictive distribution of $\eta_{i,t+h}$ given initial states $(\eta_{i,t-1}, Z_t, Z_{t-1})$. In what follows we assume Z_t is scalar, although it is straightforward to generalize the derivation to the multivariate case.

The macroeconometric tradition typically defines macro and micro impulse responses as $E\left[\eta_{i,t+h} \mid V_t = 1\right] - E\left[\eta_{i,t+h} \mid V_t = 0\right]$ and $E\left[\eta_{i,t+h} \mid u_{it} = 1\right] - E\left[\eta_{i,t+h} \mid u_{it} = 0\right]$, respectively. In linear models, they coincide with the derivatives of $q_{\eta,h}$ with respect to shocks V_t and u_{it}. In our setup, however, this approach suffers from various shortcomings. First, our model features significant nonlinearities in the persistence and interactions between macro and micro states and shocks; impulse responses should account for those. Second, in our model, the shocks V_t and u_{it} are pinned down by normalizations adopted for mathematical convenience, not for economic reasons. Thus, changes in one unit of V_t or u_{it} need not be comparable. Third, in our panel data setup, a change of given size in u_{it} may have different impacts for different individuals.

To address these issues, we define impulse responses as the impact on the predictive distribution of $\eta_{i,t+h}$ of perturbations to *past states* (as opposed to shocks), extending Gallant, Rossi, and Tauchen (1993) beyond the time series setup. For this purpose, we introduce a *rule* (denoted g below) that maps perturbations onto a consistent system of measurement. As a result, impulse responses will be indexed by past states—allowing us to document nonlinearities—and depend on the normalization rule—allowing us to achieve comparability across shocks and units.

IRFs via perturbations. Let us define a benchmark value for one of the state variables: Z^b for Z_t for the macro impulse response; η^b for $\eta_{i,t-1}$ for the micro impulse response. Let us then define the new state value Z^p (resp., η^p) by means of the perturbation $\Delta = Z^p - Z^b$ (resp., $\Delta = \eta^p - \eta^b$). Considering experiments that perturb a single state at a time, we define impulse responses as suitably scaled differences in the expected trajectory of $\eta_{i,t+h}$ for marginal perturbations Δ, holding every other past and current state constant.

To select the perturbation Δ, we introduce a rule that maps it to a system of comparable units of change δ. The rule is given by a function g such that

$$g(Z^b) = g(Z^b + \Delta(\delta)) - \delta \text{ or } g(\eta^b) = g(\eta^b + \Delta(\delta)) - \delta.$$

The rule g may depend on the benchmark value and the reference value for the remaining states but we omit the dependence from the notation. We will focus on the *unit rule*:

$$g(z) = z, \text{ leading to } \Delta(\delta) = \delta.$$

The unit rule is natural if the perturbed state is measured in money terms, or if it is in logs in which case the impulse consists of the same approximate $100 \times \delta\%$ change applied to all individuals. Other choices are possible, depending on the empirical context. For example, under a *rank rule* $g(z)$ is the CDF of the perturbed state given the other state and past states $\{\eta_{i,t-1-\ell}, Z_{t-\ell}\}_{\ell>1}$.[24] The rank rule is appropriate when the state is a concept with no natural unit of measurement, such as welfare. For the micro IRF, g can be thought of as an income transfer program implemented by a social planner or policymaker.

Given a rule g, the macro and micro impulse responses are, respectively,

$$\text{IRF}_{\eta Z}(h, \delta) = \frac{E\Big[\eta_{i,t+h} \,\Big|\, \eta_{i,t-1}, Z_t = Z^b + \Delta(\delta), Z_{t-1}\Big] - E\Big[\eta_{i,t+h} \,\Big|\, \eta_{i,t-1}, Z_t = Z^b, Z_{t-1}\Big]}{\delta},$$

$$\text{IRF}_{\eta\eta}(h, \delta) = \frac{E\Big[\eta_{i,t+h-1} \,\Big|\, \eta_{i,t-1} = \eta^b + \Delta(\delta), Z_t, Z_{t-1}\Big] - E\Big[\eta_{i,t+h-1} \,\Big|\, \eta_{i,t-1} = \eta^b, Z_t, Z_{t-1}\Big]}{\delta},$$

where we leave the dependence on $\eta_{i,t-1}$, Z_t and and Z_{t-1} implicit. For infinitesimal changes, $\text{IRF}_{\eta Z}(h) = \lim_{\delta \to 0} \text{IRF}_{\eta Z}(h, \delta)$ and $\text{IRF}_{\eta\eta}(h) = \lim_{\delta \to 0} \text{IRF}_{\eta\eta}(h, \delta)$.

Impulse responses are functions of parameters θ and λ, and can thus be estimated by the plug-in estimators whose properties we covered in Section 4. In addition, it is possible to link impulse responses to the ρ and β measures of Section 5, and to derivatives with respect to locally-defined shocks. We explore those links in Supplemental Appendix E.

[24] In the macro case, if F_V is the CDF of V_t and Q_Z^{-1} is the inverse of Q_Z with respect to its last argument, $\Delta = Q_Z(Z_{t-1}, F_V^{-1}[F_V(Q_Z^{-1}[Z_{t-1}, Z^b]) + \delta]) - Z^b$. Similarly, in the micro case, if Q_η^{-1} is the inverse of Q_η with respect to its last argument, $\Delta = Q_\eta(\eta_{i,t-2}, Z_{t-1}, Z_{t-2}, Q_\eta^{-1}[\eta_{i,t-2}, Z_{t-1}, Z_{t-2}, \eta^b] + \delta) - \eta^b$.

6.2 Empirical estimates of impulse responses

Macro impulse responses. We begin with estimates of $\text{IRF}_{\eta Z}$ for a negative perturbation to Z_t around the steady state benchmark $Z^b = \tilde{Z}_{ss}$. This emulates an aggregate shock that tips the economy into a recession. We calibrate $\delta = -2\sigma_V$ with $\sigma_V^2 = \text{Var}(\,Z_t \mid Z_{t-1}\,)$ to match a shock comparable to the Great Recession and, to facilitate interpretation, we multiply impulse responses by -1. Figure 8 shows the results.

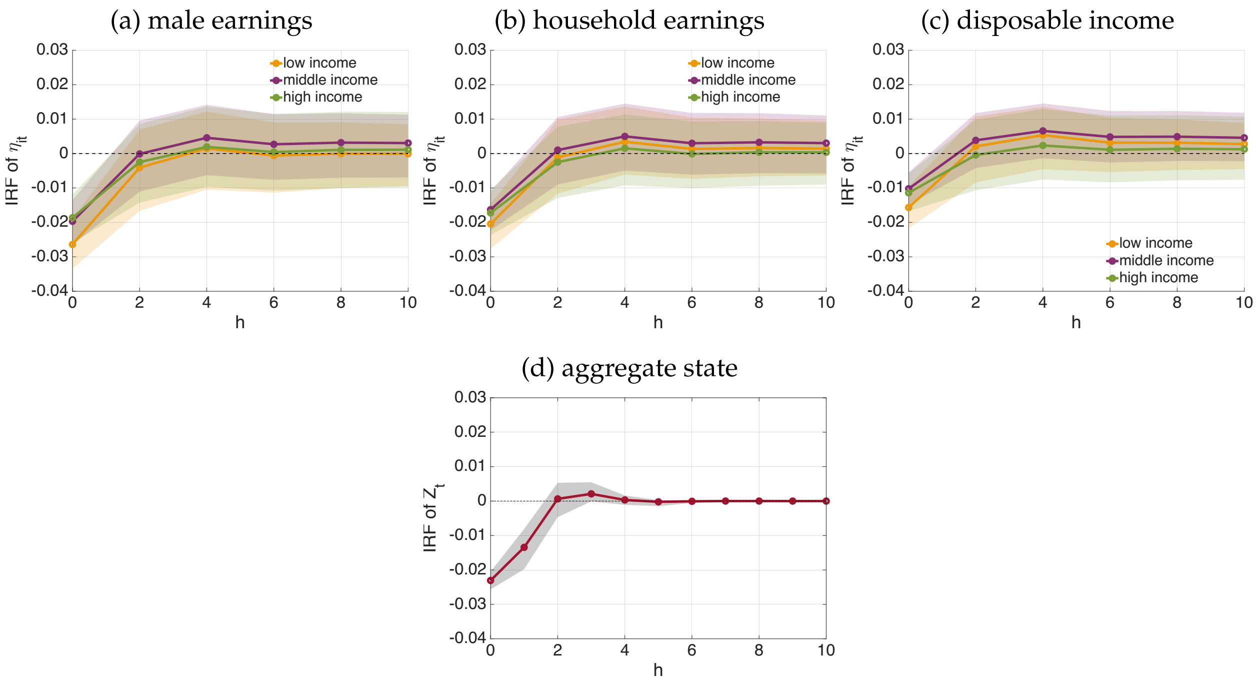

FIGURE 8. Macro impulse responses

Note: Panels (a), (b) and (c) show IRFs of η_{it} to a negative macro shock for different income measures with $Z_t^b = Z_{t-1} = \tilde{Z}_{ss}$ and $\eta_{i,t-1}$ set to the 10th (low), 50th (middle) and 90th (high) percentiles of the persistent income distribution. Panel (d) shows the IRF of Z_t annualized and scaled to detrended log GDP per capita. IRFs are multiplied by -1. Shaded areas are 90% pointwise confidence bands.

It is instructive to consider first the trajectory of Z_t in panel (d), which is annualized by averaging the quarterly responses and scaled by the standard deviation of log GDP. The normalization translates responses into log deviations from the GDP per capita trend. By this account, our underlying experiment implies a GDP roughly 2.3 and 1.3 percentage points (pp) below trend in years zero and one after the shock, respectively, returning to trend afterwards with a slight overshoot in year three.

Next, panels (a), (b) and (c) report macro impulse responses of η_{it} for our three income measures and three initial levels of income: the 10th, 50th and 90th percentiles of the $\eta_{i,t-1}$-

distribution. We highlight the following takeaways. First, the responses are quantitatively consistent with the dynamics of GDP described above. For example, averaging over the distribution of $\eta_{i,t-1}$, male earnings fall 2.1 pp on impact following the macro shock, while household earnings and disposable income fall 1.7 and 1.1 pp.[25] Furthermore, all income measures are near their pre-shock trends after two years; i.e., responses are short-lived.[26]

Second, household earnings appear as less cyclically sensitive than male earnings but more than disposable income. This mimics the discussion of Section 5 and is suggestive of the role of spousal income and the tax-transfer system as potential sources of insurance against aggregate shocks. Third, the macro IRF is U-shaped in $\eta_{i,t-1}$, with lower (but still significant) responses in the middle of the income distribution.

A key advantage of our framework is the possibility to measure the interplay between macro and micro uncertainty. This is illustrated in Figure 9 where we compute a modified version of $\text{IRF}_{\eta Z}(h, \delta)$ that further conditions on the micro shock u_{it} contemporaneous to the macro perturbation, averaged over the distribution of $\eta_{i,t-1}$. Unlike in linear models, we cannot separate the contributions of macro and micro shocks additively but we can quantify the macro state-dependence of micro quantile treatment effects.

FIGURE 9. Interaction between macro impulse responses and micro shocks

Note: Panels (a), (b) and (c) show macro IRFs conditional on the micro shock u_{it} averaged over the distribution of $\eta_{i,t-1}$ for male earnings, household earnings and disposable income with $Z_t^b = Z_{t-1} = \tilde{Z}_{ss}$. A magenta (cyan) area indicates the response is statistically negative (positive) at the 5% significance level.

Figure 9 shows sizable heterogeneity in the impact of macro shocks along the micro-rank distribution. For disposable income, an individual subject to a bad shock $u_{it} = 0.1$ suffers an expected income loss of 1.6 pp, well above the average of 1.1 pp. The number is

[25] In line with the literature, this suggests a mildly procyclical labor share (the ratio of household earnings to total income), although our estimates exclude a small fraction of zero-earnings units.

[26] Supplemental Appendix E.4 presents local projection estimates of income responses to V_t that uncover similar patterns. They also point to a significant response at $h = 1$ that our biennial setup cannot measure.

0.9 and 1.1 pp for $u_{it} = 0.5$ and $u_{it} = 0.9$. We find similar (but more pronounced) U-shaped patterns in our household and male earnings estimates; see panels (a) and (b).

Micro impulse responses. We conclude this section with estimates of the micro impulse responses $\mathrm{IRF}_{\eta\eta}$ for a negative perturbation δ that implies a 10% reduction in $\eta_{i,t-1}$. We hold Z_t and Z_{t-1} at their steady state value $\tilde{Z}_{ss}$ and multiply responses by -0.1 for ease of interpretation. See Figure 10.

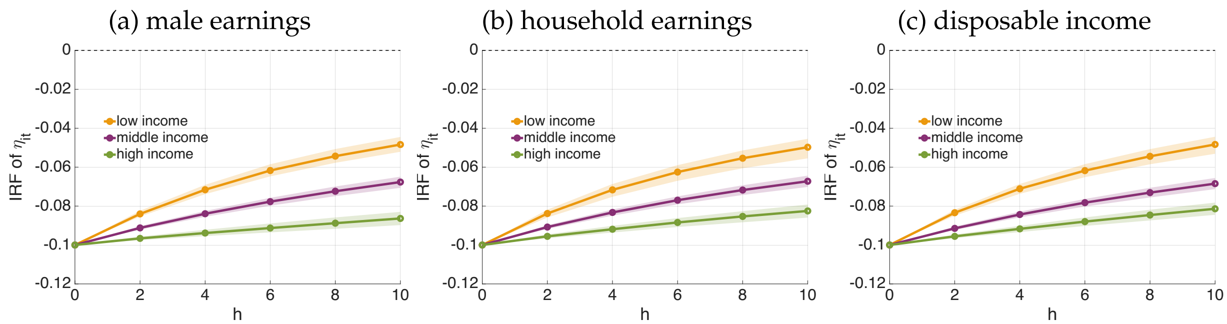

FIGURE 10. Micro impulse responses

Note: Panels (a), (b) and (c) display IRFs of η_{it} to a negative micro shock for different income measures with $Z_t = Z_{t-1} = \tilde{Z}_{ss}$ and $\eta^b_{i,t-1}$ set to the 10th (low), 50th (middle) and 90th (high) percentiles of the distribution of $\eta_{i,t-1}$. Shaded areas are 90% pointwise confidence bands.

The main takeaway, consistent across income measures, is that micro responses decay slowly and at different rates that depend on the initial level of income, with less (more) persistence for low-η (high-η) units. This reflects an intrinsic connection between nonlinear persistence ρ and $\mathrm{IRF}_{\eta\eta}$ that we discuss in Supplemental Appendix E.

Positive shocks. Supplemental Appendix E.5 reports IRFs to positive perturbations.

7 Risk quantification

In this last section, we develop our methodology to quantify aggregate and idiosyncratic contributions to income risk. We conclude by discussing empirical estimates.

7.1 Measuring the contribution to risk of macro and micro shocks

Our approach to risk quantification relies on the indirect utility of persistent income. Let $U(e^{\eta_{it}})$ be the period utility of a household whose persistent log income component at time

t is η_{it}. We focus on indirect utility since we do not have consistent data on consumption, and abstract away from the transitory component ε_{it} as it may contain measurement error. Following the macroeconomics tradition after Lucas (1987, 2003), we will measure the risk contribution of shocks by compensating variation: the value CV such that

$$\mathbb{E}^{\star}\left[\sum_{h=1}^{H}\delta^{t}U\Big((1-\mathrm{CV})\,e^{\eta_{i,t+h}}\Big)\,\middle|\,\eta_{it},Z_t\right]=\mathbb{E}\left[\sum_{h=1}^{H}\delta^{t}U\Big(e^{\eta_{i,t+h}}\Big)\,\middle|\,\eta_{it},Z_t\right]. \tag{14}$$

In (14), $\mathbb{E}$ denotes expectations over the actual persistent income process, while $\mathbb{E}^{\star}$ denotes expectations under a counterfactual income process where the macro or the micro shocks have been removed. Here CV may depend on η_{it} and Z_t, although we leave the dependence implicit.

The quantity CV measures the fraction of income the agent would be willing to forego in every period in order to eliminate a source (macro or micro) of income risk. Much of the literature emphasizes the role of curvature in preferences for the cost of risk. A common finding is that log-utility with an exponential income process implies little aggregate risk, and that high risk-aversion is needed to obtain even moderate costs of business cycles.

Here we highlight another channel that we find matters greatly to infer the magnitude of aggregate income risk: the nonlinear relationship between the income process and the aggregate factor. Indeed, our model allows for a general nonlinear relationship between η_{it} and (Z_t, Z_{t-1}); see (1). These nonlinearities, and more specifically a countercyclical β (as documented in Section 5.4), can generate higher amounts of risk than usually found.

To make our point formally, we consider a second-order small-noise expansion of CV around a no-shock baseline. Given (η_{it}, Z_t), let $\eta_{i,t+h}=\tilde{\eta}^{h}_{it}(\boldsymbol{u}^{h-1}_{i,t+1},\boldsymbol{V}^{h-1}_{t+1})$ where $\boldsymbol{u}^{h-1}_{i,t+1}=(u_{i,t+\ell})^{h}_{\ell=1}$ and $\boldsymbol{V}^{h-1}_{t+1}=(V_{t+\ell})^{h}_{\ell=1}$ are the histories of micro and macro shock. We also let $\widetilde{U}(\eta)=U(e^{\eta})$ and (purely for notational convenience) transform u_{it} so that it has zero mean—say, by applying the Gaussian inverse CDF to the original u_{it}. We multiply $\boldsymbol{u}^{h-1}_{i,t+1}$ by ς_u and $\boldsymbol{V}^{h-1}_{t+1}$ by ς_V, so that $\tilde{\eta}^{h}_{it}(\boldsymbol{u}^{h-1}_{i,t+1},\boldsymbol{V}^{h-1}_{t+1})=\eta_{i,t,h}(\varsigma_u,\varsigma_V)|_{(\varsigma_u,\varsigma_V)=(1,1)}$ for some function $\eta_{i,t,h}(\cdot)$. Similarly, we get $\mathrm{CV}=\mathrm{CV}(\varsigma_u,\varsigma_V)|_{(\varsigma_u,\varsigma_V)=(1,1)}$ which we expand to second-order around $(\varsigma_u,\varsigma_V)=(0,0)$.

We focus on the macro risk measure, denoted $\mathrm{CV}_{\mathrm{macro}}$, where the experiment eliminates only the macro shocks; analogous insights apply to its micro risk counterpart $\mathrm{CV}_{\mathrm{micro}}$ that eliminates only micro shocks. The small-noise expansion ($\varsigma_u = 0$ and $\varsigma_V \to 0$) delivers

$$\mathrm{CV}_{\mathrm{macro}} \approx -\frac{\sum_{h=1}^{H} \delta^h \sum_{\ell=1}^{h} \left(\widetilde{U}''(\tilde{\eta}_{it}^h(\mathbf{0},\mathbf{0})) \left[\frac{\partial \tilde{\eta}_{it}^h(\mathbf{0},\mathbf{0})}{\partial V_{t+\ell}} \right]^2 + \widetilde{U}'(\tilde{\eta}_{it}^h(\mathbf{0},\mathbf{0})) \left[\frac{\partial^2 \tilde{\eta}_{it}^h(\mathbf{0},\mathbf{0})}{\partial V_{t+\ell}^2} \right] \right)}{\sum_{h=1}^{H} \delta^h \widetilde{U}'(\tilde{\eta}_{it}^h(\mathbf{0},\mathbf{0}))}. \tag{15}$$

In the case of log-utility $U(e^{\eta_{it}}) = \eta_{it}$, this reduces to

$$\mathrm{CV}_{\mathrm{macro}} \approx -\sum_{h=1}^{H} \underbrace{\frac{\delta^{h-1}(1-\delta)}{(1-\delta^H)}}_{>0,\text{ sum to }1} \sum_{\ell=1}^{h} \underbrace{\left[\frac{\partial^2 \tilde{\eta}_{it}^h(\mathbf{0},\mathbf{0})}{\partial V_{t+\ell}^2} \right]}_{<0\text{ for countercyclical }\beta} . \tag{16}$$

Equation (16) shows that the cost of business cycle risk depends crucially on how nonlinear the income process is with respect to the aggregate shock V_t. For log-utility and a linear income process, $\mathrm{CV}_{\mathrm{macro}}$ is approximately zero. In contrast, it can be large when aggregate shock exposures β are countercyclical, that is, when $\partial^2 \tilde{\eta}_{it}^h(\mathbf{0},\mathbf{0})/\partial V_{t+h}^2 = \partial \beta_{it}/\partial Z_t < 0$. This is in line with our findings in Section 5.4 and reflects the self-amplifying nature of recessions.

More generally, when the utility function is not logarithmic, (15) shows that $\mathrm{CV}_{\mathrm{macro}}$ captures the combination of two effects: nonlinear income exposures to aggregate shocks, and the curvature of the utility function. To the extent that, empirically, the coefficient of relative risk aversion is often found to be low, the former effect will dominate. We verify this in our empirical calculations next.

7.2 Empirical estimates of macro and micro costs of risk

We present below our compensating variation estimates against macro and micro shocks for disposable income.[27] To investigate the role of nonlinearities in the aggregate shocks, in addition to computing risk measures from the full model estimated in Section 5, we also consider a version in which $\beta(u, \eta, Z_t, Z_{t-1}, x)$ is constant in (Z_t, Z_{t-1}), even though it can depend on (u, η, x).[28] In the restricted model, β is acylical in contrast to the countercyclical β we find in our full model.

[27] In a model without endogenous labor supply, this is the relevant income measure to study consumption and welfare. In that context, the indirect utility of persistent income $U(e^{\eta_{it}})$ can be rationalized by a variety of consumption functions: e.g., hand-to-mouth households if ε_{it} is mostly measurement error or unconstrained permanent-income households if $\mathbb{E}_t[e^{y_{i,t+h}} \mid \eta_{it}, \varepsilon_{it}, Z_t] \approx e^{\eta_{it}}$.

[28] We achieve this by restricting φ to a first-order polynomial in $(\widetilde{Z}, Z)$ in the specification described at the end of Section 4. We fit this model to the data using the same approach as for our full model.

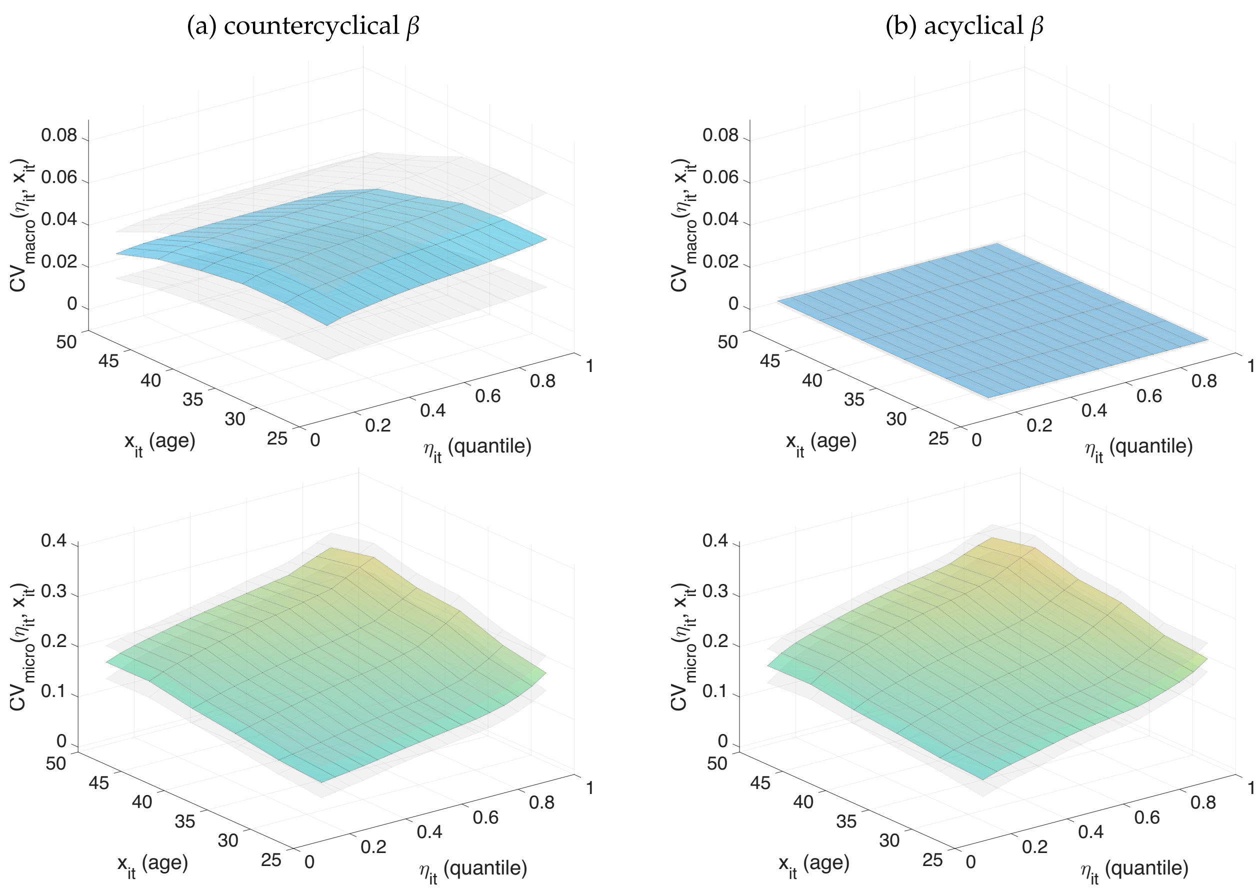

FIGURE 11. Risk quantification (disposable income).

Note: We show compensating variation for aggregate (upper panels) and idiosyncratic risk (lower panels) for various ages $x = x_{it}$ and initial incomes η_{it}. Our full model (where β is countercyclical) is on the left and a restricted acyclical-β model on the right. Lifetime utility is $\sum_{h=1}^{(65-x_{it})/2} \delta^h \frac{e^{(1-\gamma)\eta_{i,t+h}}}{(1-\gamma)}$ with $\delta = (0.96)^2$ and $\gamma = 3$. Gray shaded areas are 90% confidence bands.

Figure 11 illustrates the results. Left panels show risk calculations from our full model while right panels show the restricted model, with macro risk above and micro risk below. Period utility is $U(c) = c^{1-\gamma}/(1-\gamma)$ with risk aversion coefficient $\gamma = 3$. We set the biennial discount factor to $\delta = (0.96)^2$ and we calibrate the horizon for lifetime utility to the number of biennial periods until a notional retirement age of 65, i.e., $H = (65 - x_{it})/2$.

The main highlight from Figure 11 is the striking difference in the cost of business cycle risk between linear and nonlinear income exposures to aggregate shocks. In line with our analytical derivation, with an acyclical β (right), CV_{macro} is virtually zero, supporting the conclusion in Lucas (2003) that eliminating business cycle fluctuations is second-order from a welfare point of view. In stark contrast, a countercyclical β (left) produces CV_{macro}

orders of magnitude higher, with households willing to give up between 2.2% and 5% of their income each period in order to avoid cyclical fluctuations. Since our full model nests the restricted one and CV_{macro} is statistically non-zero in the former, the evidence favors the view that aggregate shocks have large welfare costs.[29]

Turning to the cost of idiosyncratic risk in Figure 11, perhaps unsurprisingly, CV_{micro} is higher than CV_{macro} even in the countercyclical-β case, but there is ample variation over age and across the income distribution. To give a sense, the cost of macro risk is less than a tenth the cost of micro risk for older and richer households; it is more than a third for young low-income households. This aligns with the tale of two skewnesses documented in Section 5.5: For low-η units, micro risk is primarily upside risk while aggregate shocks (given their self-amplifying effects) carry downside risks. As η increases (or as households age), conditional skewness becomes more negative and the cost of micro risk increases.

Interestingly, the restricted model and our full model have similar implications for CV_{micro}. The nonlinearities that underlie the aggregate and idiosyncratic components of income risk are separate features, targeted by different parts of our framework. A natural question is whether agents have different amounts of information about the macro- and micro-relevant features, and how to account for it within the risk quantification exercise. We leave this important question for future research.

8 Conclusion

In this paper we propose a nonlinear semi-structural framework for heterogeneous-agent models with aggregate shocks. The nonlinear reduced form relationships our approach can uncover are useful to assess the fit of such models and their implications. Allowing for general nonlinear dynamics and rich macro-micro feedback, we study identification, estimation and inference tools that leverage both macro and micro data.

In our empirical analysis of U.S. panel data on income, we find that the nonlinear persistence and conditional skewness patterns documented in previous work are affected by business cycle fluctuations. The persistence surface tilts in recessions, decreasing for high-income households hit by a bad shock and increasing for low-income households,

[29] This conclusion changes very little for larger values of the risk aversion coefficient γ. Instead, the cost of risk under countercyclical β is higher for larger persistence and variance of aggregate shocks, confirming that the interaction between $\partial\beta_{it}/\partial Z_t$ and marginal utility is, empirically, the relevant dimension of the income process for risk quantification.

while skewness declines throughout the income distribution. We also find evidence of nonlinear exposures to aggregate shocks, with higher sensitivity of income to macro shocks during recessions and nontrivial interactions between macro and micro shocks. Our results suggest that nonlinearities with respect to aggregate shocks matter for risk quantification. One avenue to explore is how to account for model uncertainty in our risk measurement approach.

Three natural extensions come to mind which we leave to future work. The first is to examine the pass-through of aggregate and idiosyncratic income shocks to consumption. This would follow Arellano et al. (2017) and Arellano et al. (2023), where consumption data from the PSID was combined with the nonlinear income process to examine the impact of income shocks on consumption. The results in those studies found a differential response according to initial household assets and household heterogeneity. The framework developed in this paper would enable estimation of the transmission of aggregate shocks through income to consumption as well as the differential impact of idiosyncratic income shocks by aggregate state. Thus allowing an assessment of the extent by which insurance to persistent income shocks varies across the cycle for different types of households, and providing a more complete analysis of partial insurance and the welfare costs of income risk across the business cycle.

The second extension relates to the primary components of household income: male and female earnings. These two earnings measures embody the labor supply responses of the spouses. In a linear partial insurance framework, Blundell, Pistaferri, and Saporta-Eksten (2016) investigate the responses to wage shocks of spousal hours of work and consumption in the PSID, finding significant responses to permanent and transitory wage shocks. However, they do not consider non-linear wage processes and ignore the impact of aggregate factors. Arellano et al. (2017) found that the permanent component of male wages in the PSID follows a nonlinear persistent process. However, they did not examine the joint process of male and female wages, nor the impact of aggregate factors. The potential for uncovering important wage dynamics and effects of aggregate factors for couples makes this an exciting area for future research.

A third extension, related to the second, concerns the zeros in the income data. Figure 2 showed a small but growing, and cyclical, proportion of zeros for male earnings in our couples sample, a more stable and much smaller proportion of zeros for household earnings and no zeros for disposable income. This is one reason why we have focused attention on disposable income. Nonetheless, we have also drawn conclusions in comparison with male and household earnings. Modeling the extensive margin requires some assumptions on the process of selection. A natural framework is to assume missingness conditional on history of the income process. But it is likely that aggregate shocks also play a role here. Braxton, Herkenhoff, Rothbaum, and Schmidt (2021) develop a Kalman filter and EM algorithm approach to incorporate observations with zero (or missing) earnings. By specifying a law of motion for persistent earnings they show how assuming a model for earnings upon re-entry can deliver identification under selection. In another recent paper, Gobillon, Magnac, and Roux (2022) study earnings dynamics in French administrative data assuming zeros are missing at random conditional on a set of factors drawn from earnings history. We believe the nonlinear framework with aggregate factors developed in our paper provides an ideal setting for incorporating the extensive margin into the study of income dynamics over the cycle. We leave this extension to future work.

References

Ahn, S., G. Kaplan, B. Moll, T. Winberry, and C. Wolf (2018): "When Inequality Matters for Macro and Macro Matters for Inequality." *NBER Macroeconomics Annual*, 32, 1–75.

Almuzara, M. (2020): "Heterogeneity in Transitory Income Risk." Working paper.

Almuzara, M. and V. Sancibrián (2024): "Micro Responses to Macro Shocks." Federal Reserve Bank of New York Staff Report 1090.

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

Amberg, N., T. Jansson, M. Klein, and A. Rogantini Picco (2022): "Five Facts about the Distributional Income Effects of Monetary Policy Shocks." *AER: Insights*, 4, 289–304.

Andersen, A. L., N. Johannesen, M. Jørgensen, and J.-L. Peydró (2023): "Monetary Policy and Inequality." *Journal of Finance*, 78, 2945–2989.

Angeletos, G.-M., F. Collard, and H. Dellas (2020): "Business-Cycle Anatomy." *American Economic Review*, 110, 3030–3070.

Arellano, M. (2014): "Uncertainty, persistence and heterogeneity: A panel data perspective." *Journal of the European Economic Association*, 12, 1127–1153.

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

——— (2017): "Nonlinear panel data methods for dynamic heterogeneous agent models." *Annual Review of Economics*, 9, 471–496.

Barlevy, G. (2004): "The Cost of Business Cycles Under Endogenous Growth." *American Economic Review*, 94, 964–990.

Bhandari, A., D. Evans, M. Golosov, and T. J. Sargent (2021): "Inequality, Business Cycles, and Monetary-Fiscal Policy." *Econometrica*, 89, 2559–2599.

Blundell, R., C. R. Bollinger, C. Hokayem, and J. P. Ziliak (2024): "Interpreting Cohort Profiles of Lifecycle Earnings Volatility." Working Paper CES-24-21.

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.

Borella, M., M. De Nardi, M. Pak, N. Russo, and F. Yang (2023): "The Importance of Modeling Income Taxes Over Time." *Journal of the European Economic Association*, 21, 2237–2286.

Braxton, J. C., K. F. Herkenhoff, J. L. Rothbaum, and L. D. W. Schmidt (2021): "Changing Income Risk across the US Skill Distribution: Evidence from a Generalized Kalman Filter." NBER Working Paper 29567.

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

Burns, A. F. and W. C. Mitchell (1946): *Measuring Business Cycles*, NBER.

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

Chang, M., X. Chen, and F. Schorfheide (2024): "Heterogeneity and Aggregate Fluctuations." *Journal of Political Economy*, 132.

Chetty, V. K. (1968): "Pooling of Time Series and Cross Section Data." *Econometrica*, 36, 279–290.

Fernández-Villaverde, J., S. Hurtado, and G. Nuño (2023): "Financial Frictions and the Wealth Distribution." *Econometrica*, 91, 869–901.

Fleck, J., J. Heathcote, K. Storesletten, and G. L. Violante (2025): "Fiscal Progressivity of the U.S. Federal and State Governments." NBER Working Paper 33385.

Gallant, A., P. Rossi, and G. Tauchen (1993): "Nonlinear dynamic structures," *Econometrica*, 61, 871–907.

Galí, J., M. Gertler, and J. D. López-Salido (2007): "Markups, Gaps, and the Welfare Costs of Business Fluctuations." *Review of Economics and Statistics*, 89, 44–59.

Gobillon, L., T. Magnac, and S. Roux (2022): "Lifecycle Wages and Human Capital Investments: Selection and Missing Data." Working Paper TSE Working Paper 22-1299.

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., F. Karahan, S. Özkan, and J. Song (2021): "What Do Data on Millions of U.S. Workers Reveal About Life-Cycle Earnings Dynamics?" *Econometrica*, 89, 2303–2339.

Guvenen, F., A. McKay, and C. B. Ryan (2023): "A Tractable Income Process for Business Cycle Analysis." NBER Working Paper 30959.

Guvenen, F., S. Ozkan, and J. Song (2014): "The nature of countercyclical income risk." *Journal of Political Economy*, 122, 621–660.

Guvenen, F., L. Pistaferri, and G. L. Violante (2022): "Global trends in income inequality and income dynamics: New insights from GRID." *Quantitative Economics*, 13, 1321–1360.

Guvenen, F., S. Schulhofer-Wohl, J. Song, and M. Yogo (2017): "Worker Betas: Five Facts about Systematic Earnings Risk." *American Economic Review*, 107, 398–403.

Hahn, J., G. Kuersteiner, and M. Mazzocco (2020): "Estimation with Aggregate Shocks." *Review of Economic Studies*, 87, 1365–1398.

Halvorsen, E., H. A. Holter, S. Ozkan, and K. Storesletten (2024): "Dissecting Idiosyncratic Earnings Risk." *Journal of the European Economic Association*, 22, 617–668.

Hause, J. C. (1977): "The Covariance Structure of Earnings and the On-The-Job Training Hypothesis." in *Annals of Economic and Social Measurement, Volume 6, number 4*, NBER, 335–365.

Holm, M. B., P. Paul, and A. Tischbirek (2021): "The Transmission of Monetary Policy under the Microscope." *Journal of Political Economy*, 129, 2861–2904.

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

Krebs, T. (2007): "Job Displacement Risk and the Cost of Business Cycles." *American Economic Review*, 97, 664–686.

Krueger, D., K. Mitman, and F. Perri (2016): "Macroeconomics and Household Heterogeneity." in *Handbook of Macroeconomics*, ed. by J. Taylor and H. Uhlig, North-Holland, vol. 2A.

Krusell, P. and A. A. Smith (1998): "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy*, 106, 867–896.

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

Liu, L. and M. Plagborg-Møller (2023): "Full-information estimation of heterogeneous agent models using macro and micro data." *Quantitative Economics*, 14, 1–35.

Lucas, Robert E., J. (1987): *Models of Business Cycles*, Basil Blackwell.

——— (2003): "Macroeconomic Priorities." *American Economic Review*, 93, 1–14.

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.

Maddala, G. S. (1971): "The Likelihood Approach to Pooling Cross-Section and Time-Series Data." *Econometrica*, 39, 939–953.

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

Otrok, C. (2001): "On Measuring the Welfare Cost of Business Cycles." *Journal of Monetary Economics*, 47, 61–92.

Patterson, C. (2023): "The Matching Multiplier and the Amplification of Recessions." *American Economic Review*, 113, 982–1012.

Ramey, V. (2016): "Macroeconomic shocks and their propagation." in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, chap. 2.

Sargent, T. J. and Y. J. Selvakumar (2025): "Aggregate Shocks and Cross-Section Dynamics: Quantifying Redistribution and Insurance in U.S. Household Data." Working paper.

Stock, J. H. and M. W. Watson (2016): "Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics." in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, Elsevier, vol. 2, chap. 8.

Storesletten, K., C. I. Telmer, and A. Yaron (2001): "The Welfare Cost of Business Cycles Revisited: Finite Lives and Cyclical Variation in Idiosyncratic Risk." *European Economic Review*, 45, 1311–1339.

——— (2004): "Cyclical dynamics in idiosyncratic labor market risk." *Journal of Political Economy*, 112, 695–717.

Tobin, J. (1950): "A Statistical Demand Function for Food in the U.S.A." *Journal of the Royal Statistical Society. Series A (General)*, 113, 113–141.

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

Supplemental Material

Nonlinear Micro Income Processes with Macro Shocks

Martín Almuzara* Manuel Arellano‡

Richard Blundell† Stéphane Bonhomme§

August 2025

[Link to the paper]

A Stylized model

Recursive equilibria typically imply a Markovian law of motion for the micro and macro states of the same form as that followed by η_{it} and Z_t in Equations (1) and (3). Hence, our framework can be interpreted as providing reduced forms for heterogeneous agents models with aggregate shocks. This section illustrates this with a stylized model.

Setup. Consider an overlapping generations version of Krusell and Smith (1998). The economy is populated by a continuum of households and firms. Households live for H periods with certainty. Let N_{ht} be the mass of households aged $h = 1, \dots, H$ during period t. Thus, N_{1t} is the mass of households who enter the economy in t, N_{Ht} is the mass who leave at the end of t, and $N_{ht} = N_{h-1,t-1}$. Firms live forever and have mass J.

*FRBNY: martin.almuzara@ny.frb.org

†UCL and IFS: r.blundell@ucl.ac.uk

‡CEMFI: arellano@cemfi.es

§University of Chicago: sbonhomme@chicago.edu

There is an exogenous aggregate state z_t that will drive both households' employment and firms' productivity—and therefore innovations to z_t will be a mix of labor supply and TFP shocks. The aggregate state z_t follows an AR(1) with serially independent innovations,

$$z_t = \Phi z_{t-1} + V_t, \qquad V_t \sim N(0, \sigma_V^2). \tag{A.1}$$

Households. Household i inelastically supplies skill $s_{it} = e^{\eta_{it}+\varepsilon_{it}}$ to firms in exchange for the market wage w_t. It begins period t with assets a_{it} which lends to firms in exchange for the market interest rate r_t, and it consumes c_{it}. If x_{it} is the age of household i at time t, its labor income is $\tilde{y}_{it} = w_t s_{it} = w_t e^{\eta_{it}+\varepsilon_{it}}$ with permanent and transitory components

$$\begin{aligned} &\eta_{it} = \rho\eta_{i,t-1} + \beta z_t + u_{it}, \qquad u_{it} \sim N(0, \sigma_u^2), \qquad \eta_{i,t-x_{it}+1} \sim N(\mu_{\text{init}}, \sigma_{\text{init}}^2), \\ &\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2). \end{aligned} \tag{A.2}$$

Micro shocks $(u_{it}, \varepsilon_{it})$ are i.i.d. across i and over t. The household's budget constraint is

$$c_{it} + a_{i,t+1} = \tilde{y}_{it} + (1 + r_t)a_{it},$$

with $c_{it}, a_{i,t+1} \geq 0$ and $a_{i,t-x_{it}+1} = 0$ (households are born with no wealth).

Then, individual state variables are $(a_{it}, \eta_{it}, \varepsilon_{it})$. As noted by Krusell and Smith (1998), however, because w_t and r_t are determined in equilibrium (see (A.4) below) the state vector for the household problem should also include the aggregate z_t and the distribution of individual states in the population. Let μ_{ht} be the time-t joint distribution of assets and skill components $(a_{it}, \eta_{it}, \varepsilon_{it})$ for households of age h and let us collect all the age-specific distributions in $\mu_t = \{\mu_{1t}, \ldots, \mu_{Ht}\}$.

Consumption and asset paths solve a finite-horizon sequential problem with value

$$v(a_{it}, \eta_{it}, \varepsilon_{it}, z_t, \mu_t, x_{it}) = \max E\left[\sum_{\ell=0}^{H-x_{it}} \delta^\ell U(c_{i,t+\ell}) \,\middle|\, \eta_{it}, \varepsilon_{it}, z_t, \mu_t, x_{it}\right],$$

where the maximization is over stochastic consumption and asset plans that satisfy the budget constraints. Optimal choices are given by two age-dependent policy functions,

$$\begin{aligned} c_{it} &= g_c(a_{it}, \eta_{it}, \varepsilon_{it}, z_t, \mu_t, x_{it}), \\ a_{i,t+1} &= g_a(a_{it}, \eta_{it}, \varepsilon_{it}, z_t, \mu_t, x_{it}) = w_t e^{\eta_{it}+\varepsilon_{it}} + (1 + r_t)a_{it} - g_c(a_{it}, \eta_{it}, \varepsilon_{it}, z_t, \mu_t, x_{it}). \end{aligned} \tag{A.3}$$

Firms. Firm j hires labor $L_t(j)$ and rents capital $K_t(j)$ in perfectly competitive markets, and produces final goods via the constant-returns-to-scale technology, $Y_t(j) = F(K_t(j), e^{z_t} L_t(j))$. Profit maximization leads to conditions $w_t = e^{z_t} F_l(K_t/e^{z_t} L_t, 1)$ and $r_t = F_k(K_t/e^{z_t} L_t, 1) - d_k$ with $L_t = \int_J L_t(j)\, dj$ and $K_t = \int_J K_t(j)\, dj$ the total amounts of labor and capital demanded by the firm sector, F_l and F_k the marginal products, and d_k the depreciation rate.

Market clearing. In equilibrium, firms' demand for labor and capital meets households' supply of skills and assets, that is,

$$L_t = S_t \equiv \sum_{h=1}^{H} N_{ht} \int e^{\eta+\varepsilon}\, d\mu_{ht}(a, \eta, \varepsilon), \qquad K_t = A_t \equiv \sum_{h=1}^{H} N_{ht} \int a\, d\mu_{ht}(a, \eta, \varepsilon),$$

Substituting into the conditions for profit maximization, we get the pricing functions

$$\begin{aligned} w_t &= e^{z_t} F_l(A_t/e^{z_t} S_t, 1) = w(z_t, \mu_t), \\ r_t &= F_k(A_t/e^{z_t} S_t, 1) - d_k = r(z_t, \mu_t). \end{aligned} \tag{A.4}$$

A recursive equilibrium for this economy is given by the policy functions in (A.3) and the pricing functions in (A.4), reflecting optimal behavior of households and firms, and market clearing. Given the laws of motion for exogenous macro and micro states (A.1) and (A.2), the recursive equilibrium implies a functional law of motion for μ_t:

$$\mu_{t+1} = \Pi(\mu_t, z_{t+1}, z_t).$$

Semi-structural reduced form and atomicity. Denote $\overline{\boldsymbol{\eta}}_{it} = (\eta_{it}, \varepsilon_{it}, c_{it}, a_{it})$ and $\overline{\boldsymbol{Z}}_t = (z_t, \mu_t)$. The equilibrium from the stylized model can be represented as

$$\begin{aligned} \overline{\boldsymbol{\eta}}_{it} &= Q_\eta(\overline{\boldsymbol{\eta}}_{i,t-1}, \overline{\boldsymbol{Z}}_t, \overline{\boldsymbol{Z}}_{t-1}, \overline{\boldsymbol{u}}_{it}), \\ \overline{\boldsymbol{Z}}_t &= Q_Z(\overline{\boldsymbol{Z}}_{t-1}, V_t), \end{aligned} \tag{A.5}$$

for some functions Q_η and Q_Z, where the micro shocks are $\overline{\boldsymbol{u}}_{it} = (u_{it}, \varepsilon_{it})$ and the macro shock is the innovation V_t in (A.1). This is a multivariate (or, more precisely, a functional) version of Equations (1) and (3).[1]

A key insight from representation (A.5) is that it embodies the atomicity assumption;

[1]See Arellano and Bonhomme (2017) for a related point.

see Assumption 3. Specifically, $\overline{\boldsymbol{Z}}_t$ is independent of the micro shock $\overline{\boldsymbol{u}}_{it}$, for each individual i, even though $\overline{\boldsymbol{Z}}_t$ itself contains the full distribution of micro states driven by those shocks. The explanation lies in the iidness of $\overline{\boldsymbol{u}}_{it}$ over i coupled with the household population being a continuum, which allows the law of large numbers to operate.

A difficulty with the macro side of (A.5) is that it is high-dimensional. Many structural approaches rely on approximating μ_t by a finite-dimensional summary, often given by a small collection of moments or quantiles.[2] In our semi-structural reduced-form approach, enriching the macro state variable Z_t (and the macro measurement system W_t) along those lines appears as a promising avenue to account for the role of general equilibrium effects from the dynamics of micro distributions. This is in addition to including in Z_t variables that are informative about additional shocks and shifts in policy. In that sense, the class of models that can be represented as (A.5) is wide and the main appeal of our approach is that it may be possible to identify and estimate economically-relevant parameters without the need to fully specify preferences, expectations formation, technology, frictions, etc.

B Identification

Proof of Proposition 1. Let s_W, s_Z and s_e be the spectral density matrices of W_t, Z_t and e_t—all well-defined by Assumption 1. By Gaussianity, the distribution of $\{W_t\}$ is identified if and only if s_W is identified. For all $\omega \in [-\pi, \pi]$, the equation $s_W(\omega) = \Lambda s_Z(\omega)\Lambda' + s_E(\omega)$ has a unique solution $\{\Lambda, s_Z(\omega), s_E(\omega)\}$ under (a) and (b) by the steps in the proof of Geweke and Singleton (1981, Proposition 2). Hence, s_Z is identified and by Gaussianity so is Q_Z. □

Proof of Proposition 2. The argument follows from a simplified version of Almuzara (2020, Proposition 1) without heterogeneity. Fix t and r such that $t < r < t + S - 1$ and consider

$$f_{y_{r-1},y_r,y_{r+1}|x,t}(y_-, y, y_+|\boldsymbol{x}) = \int f_{y_{r-1}|\eta_r,x,t}(y_-|\eta, \boldsymbol{x}) f_{y_r,\eta_r|x,t}(y, \eta|\boldsymbol{x}) f_{y_{r+1}|\eta_r,x,t}(y_+|\eta, \boldsymbol{x})\, d\eta.$$

[2] Examples are Krusell and Smith (1998), Reiter (2009), Winberry (2018) and Bayer and Luetticke (2020). In our stylized model, although $\overline{\boldsymbol{Z}}_t$ is infinite-dimensional, it is stochastically singular as the only source of randomness is V_t. Moreover, in the income process, μ_t only enters through the market wage w_t:

$$y_{it} = \ln(\tilde{y}_{it}) = \ln w(z_t, \mu_t) + \eta_{it} + \varepsilon_{it}.$$

Under stationarity, we can write $\ln w(z_t, \mu_t) = \overline{w}(\{V_{t-\ell}\}_{\ell \geq 0})$ for some transfer function $\overline{w}(\cdot)$. Subsuming this term into η_{it}, one can think of the empirical specification in our paper as capturing in a parsimonious way the composite effect (both direct and through equilibrium) of shocks $\{V_t\}$.

This defines an integral operator equation that can be solved applying the diagonalization method of Hu and Schennach (2008). By (a) and (b), the operator equation and its spectral decomposition are well defined. Moreover, by the reasoning in Almuzara (2020, Remark 1), uniqueness of the decomposition is ensured by the normalization $E_t[\, y_{ir} \mid \eta_{ir}, x_{ir} \,] = \eta_{ir}$ where the subindex t indicates the expectation is an integral against the subpanel-specific density $f_{y_r|\eta_r,x,t}$.[3] This analysis delivers identification of $f_{y_{r-1}|\eta_r,x,t}$, $f_{y_r,\eta_r|x,t}$ and $f_{y_{r+1}|\eta_r,x,t}$ from where the CDF $F_{\eta,r}$ can be pinned down.

It follows that there is a known injective mapping from the observables $(\{W_{t+s}\}_{s=0}^{S-1}, F_t^S)$ to $(\{W_{t+s}\}_{s=0}^{S-1}, \{F_{\eta,t+s,t}\}_{s=2}^{S-2})$ for each t. Hence, the latter is measurable with respect to the former and $\overline{P}^S$ is identified from P^S. □

Proof of Proposition 3. Take r such that $\mathcal{F}_{z_r|\boldsymbol{W}_t^S}$ is complete. For any $\widetilde{\eta}$, η, x, and $\boldsymbol{W}$,

$$
\begin{aligned}
E\Big[\, F_{\eta,r,t}(\widetilde{\eta}|\eta, x) \;\Big|\; \boldsymbol{W}_t^S = \boldsymbol{W} \,\Big] &= E\Big[\, P\Big(\, \eta_{ir} \le \widetilde{\eta} \Big| \eta_{i,r-1} = \eta, x_{ir} = x, \omega_t \Big) \;\Big|\; \boldsymbol{W}_t^S = \boldsymbol{W} \,\Big] \\
&= E\Big[\, P\Big(\, \eta_{ir} \le \widetilde{\eta} \Big| \eta_{i,r-1} = \eta, x_{ir} = x, Z_r, Z_{r-1}, G_r \Big) \;\Big|\; \boldsymbol{W}_t^S = \boldsymbol{W} \,\Big] \\
&= E\Big[\, P\Big(\, \eta_{ir} \le \widetilde{\eta} \Big| \eta_{i,r-1} = \eta, x_{ir} = x, Z_r, Z_{r-1} \Big) \;\Big|\; \boldsymbol{W}_t^S = \boldsymbol{W} \,\Big],
\end{aligned}
$$

where the second line uses the fact that ω_t encompasses (Z_r, Z_{r-1}, G_τ) and Assumption 2(b), while the third uses independence between (Z_r, Z_{r-1}) and G_r given $\boldsymbol{W}_t^S = (W_t, ..., W_{t+S-1})$, which comes from Assumption 1(b).

The previous equation can be written more explicitly as

$$
E\Big[\, F_{\eta,r,t}(\widetilde{\eta}|\eta, x) \;\Big|\; \boldsymbol{W}_t^S = \boldsymbol{W} \,\Big] = \int_{\mathcal{Z}^2} F_{\eta_r|\eta_{r-1},x_r,z_r}(\widetilde{\eta}|\eta, x, Z_t, Z_{t-1})\, f_{z_r|\boldsymbol{W}_t^S}(Z_t, Z_{t-1}|\boldsymbol{W})\, d(Z_t, Z_{t-1}).
$$

Here $\mathcal{Z}$ denotes the support of Z_t and $F_{\eta_r|\eta_{r-1},x_r,z_r}$ is the CDF of η_{ir} given $(\eta_{i,r-1}, x_{ir}, Z_r, Z_{r-1})$. Now, the object on the left is identified by Proposition 2 and the density $f_{z_r|\boldsymbol{W}_t^S}$ is identified under Proposition 1. The only unknown in the equation above is $F_{\eta_r|\eta_{r-1},x_r,z_r}$.

Let $\mathcal{W}$ be the support of W_t. Define the integral operators

$$
\begin{aligned}
\Big[L_{\eta_r|\eta_{r-1},x_r,\boldsymbol{W}_t^S} h_1 \Big](\widetilde{\eta}, \eta, x) &= \int_{\mathcal{W}^S} E\Big[\, F_{\eta,r}(\widetilde{\eta}|\eta, x) \;\Big|\; \boldsymbol{W}_t^S = \boldsymbol{W} \,\Big]\, h_1(\boldsymbol{W})\, d\boldsymbol{W}, \\
\Big[L_{z_r|\boldsymbol{W}_t^S} h_1 \Big](Z_t, Z_{t-1}) &= \int_{\mathcal{W}^S} f_{z_r|\boldsymbol{W}_t^S}(Z_t, Z_{t-1}|\boldsymbol{W})\, h_1(\boldsymbol{W})\, d\boldsymbol{W},
\end{aligned}
$$

[3] The completeness condition needed for this to work is assumed to hold relative to the space of absolutely integrable functions on the relevant domain, as in Hu and Schennach (2008).

$$\left[L_{\eta_r|\eta_{r-1},x_r,z_r} h_2\right](\widetilde{\eta}, \eta, x) = \int_{\mathcal{Z}^2} F_{\eta_r|\eta_{r-1},x_r,z_r}(\widetilde{\eta}|\eta, x, Z_t, Z_{t-1})\, h_2(Z_t, Z_{t-1})\, d(Z_t, Z_{t-1}),$$

so that our main equation is equivalent to (see Carrasco, Florens, and Renault, 2007)

$$L_{\eta_r|\eta_{r-1},x_r,\mathbf{W}_t^S} = L_{\eta_r|\eta_{r-1},x_r,z_r} L_{z_r|\mathbf{W}_t^S}. \tag{B.1}$$

By our previous discussion, $L_{\eta_r|\eta_{r-1},x_r,\mathbf{W}_t^S}$ and $L_{z_r|\mathbf{W}_t^S}$ are known to the researcher. Since $\mathcal{F}_{z_r|\mathbf{W}_t^S}$ is complete, $L_{z_r|\mathbf{W}_t^S}$ has a right inverse and Equation (B.1) has solution

$$L_{\eta_r|\eta_{r-1},x_r,z_r} = L_{\eta_r|\eta_{r-1},x_r,\mathbf{W}_t^S} L^{-1}_{z_r|\mathbf{W}_t^S}.$$

which uniquely determines $F_{\eta_r|\eta_{r-1},x_r,z_r}$. It follows that Q_η in (1) is identified. □

C Estimation

Below we provide additional information about the estimation strategy outlined in Section 4. Section C.1 spells out the moments implied by our model. Section C.2 summarizes the simulation-based techniques used in the E step of Algorithm 1. Section C.3 develops the asymptotic analysis. Finally, Section C.4 discusses our bootstrap approach to inference.

C.1 Moment conditions

Our model implies two types of infeasible complete-data moments that pin down θ and $\delta_t = (\delta_{\text{init},t}, \{\delta_{\varepsilon,t+s}\}_{s=0}^{S-1})$. We specify them explicitly for the parameter vector θ which contains $\text{vec}\{\Theta(\overline{u}_\ell)\}$ for $\ell = 1, \ldots, L$ together with θ_{lo} and θ_{up}.

Write $\nu_u(\omega) = u - 1\{\omega < 0\}$. For nodes $u = \overline{u}_1, \ldots, \overline{u}_L$, we will use the orthogonality conditions from quantile regression (Koenker and Bassett, 1978),

$$m_{it}^{\text{qr}}(\theta, u) = \sum_{\tau=t+1}^{t+S-1} \left[\psi(\eta_{i,\tau-1}, x_{i\tau}) \otimes \varphi(Z_\tau, Z_{\tau-1})\right] \times \nu_u\Big(\eta_{i\tau} - \psi(\eta_{i,\tau-1}, x_{i\tau})'\Theta(u)\varphi(Z_\tau, Z_{\tau-1})\Big).$$

For the tails we will use the orthogonality conditions from exponential regression,

$$m_{it}^{\text{lo}}(\theta) = \sum_{\tau=t+1}^{t+S-1} \psi_{\text{lo}}(\eta_{i,\tau-1}, Z_\tau, Z_{\tau-1}, x_{i\tau}) \times 1\Big\{\eta_{i\tau} < \psi(\eta_{i,\tau-1}, x_{i\tau})'\Theta(\overline{u}_1)\varphi(Z_\tau, Z_{\tau-1})\Big\}$$

$$\times \Big[\psi(\eta_{i,\tau-1}, x_{i\tau})'\Theta(\overline{u}_1)\varphi(Z_\tau, Z_{\tau-1}) - \eta_{i\tau} - \exp\Big(\psi_{\mathrm{lo}}(\eta_{i,\tau-1}, Z_\tau, Z_{\tau-1}, x_{i\tau})'\theta_{\mathrm{lo}}\Big)\Big],$$

$$m_{it}^{\mathrm{up}}(\theta) = \sum_{\tau=t+1}^{t+S-1} \psi_{\mathrm{up}}(\eta_{i,\tau-1}, Z_\tau, Z_{\tau-1}, x_{i\tau}) \times 1\Big\{\eta_{i\tau} > \psi(\eta_{i,\tau-1}, x_{i\tau})'\Theta(\overline{u}_L)\varphi(Z_\tau, Z_{\tau-1})\Big\}$$
$$\times \Big[\eta_{i\tau} - \psi(\eta_{i,\tau-1}, x_{i\tau})'\Theta(\overline{u}_L)\varphi(Z_\tau, Z_{\tau-1}) - \exp\Big(\psi_{\mathrm{up}}(\eta_{i,\tau-1}, Z_\tau, Z_{\tau-1}, x_{i\tau})'\theta_{\mathrm{up}}\Big)\Big].$$

Thus, letting $\bar{y}_{it}^S = \{y_{i,t+s}, x_{i,t+s}\}_{s=0}^{S-1}$, $\bar{\eta}_{it}^S = \{\eta_{i,t+s}\}_{s=0}^{S-1}$ and $\bar{Z}_t^S = \{Z_{t+s}\}_{s=0}^{S-1}$, the moment conditions $m_\theta(\theta; \bar{y}_{it}^S, \bar{\eta}_{it}^S, \bar{Z}_t^S)$ arise from stacking the conditions $m_{it}^{\mathrm{qr}}(\theta, \overline{u}_\ell)$ for $\ell = 1, \dots, L$ together with $m_{it}^{\mathrm{lo}}(\theta)$ and $m_{it}^{\mathrm{up}}(\theta)$. At the true parameter value θ_0, we obtain

$$E\Big[\, m_\theta\Big(\theta_0; \bar{y}_{it}^S, \bar{\eta}_{it}^S, \bar{Z}_t^S\Big)\Big] = 0_{\dim(\theta)\times 1}.$$

The moments $m_\delta(\delta_t; \bar{y}_{it}^S, \bar{\eta}_{it}^S)$ associated to δ_t are also a combination of quantile and exponential regression orthogonality conditions. At the true value δ_{0t},

$$E\Big[\, m_\delta\Big(\delta_{0t}; \bar{y}_{it}^S, \bar{\eta}_{it}^S\Big)\Big] = 0_{\dim(\delta_t)\times 1}.$$

C.2 Techniques for posterior sampling

Macro posterior: Kalman recursions. Our analysis relies on the macro linear state-space model (3) where the observable vector $W_t = \Lambda Z_t + e_t$ has $n_W = 5$ entries: GDP, consumption, investment, the unemployment rate and hours worked, all transformed and dentrended as explained in Section 5.1. The data are quarterly and span the period 1960Q1-2019Q4.

We model the univariate state Z_t and each entry in E_t as AR(2) processes:

$$Z_t = \Phi_1 Z_{t-1} + \Phi_2 Z_{t-2} + \sigma_V V_t,$$
$$e_{jt} = \phi_{j1} e_{j,t-1} + \phi_{j2} e_{j,t-2} + \sigma_{E,j} \nu_{jt}, \quad j = 1, \dots, n_W,$$

where $V_t, \nu_{1t}, \dots, \nu_{n_w,t}$ are i.i.d. standard normal and mutually independent. Moreover, as stated in the text, we normalize the entry of Λ that corresponds to GDP to unity so that Z_t is measured in units of GDP per capita relative to its low-frequency trend.

We perform estimation of parameters $\lambda = (\Lambda, \Phi_1, \Phi_2, \sigma_V, \{\phi_{j1}, \phi_{j2}, \sigma_{E,j}\}_{j=1}^{n_W})$ and filtering of latent variables Z_t, $\{e_{jt}\}_{j=1}^{n_W}$ jointly via Gibbs sampling using (i) a flat prior on the parameters and (ii) a diffuse prior on the initial conditions of the latent variables.

The Gibbs sampling for our linear state-space model is a standard technique that builds on the following conditional distributions:

(a) Given $\{W_t, Z_t\}$, the distribution of parameters can be written in terms of easy-to-draw multivariate normal and inverse gamma random variables.

(b) Given parameters, the distribution of $\{Z_t, \{e_{jt}\}_{j=1}^{n_W}\}$ is multivariate normal and can be efficiently sampled from using the algorithm of Durbin and Koopman (2002).

We alternate between (a) and (b) for a total of 12,000 draws, burning in the first 2,000. We then retain 1 in 2 parameter draws (5,000 in total) and 1 in 20 latent variable draws (500 in total). We set $\widehat{\lambda}$ to the median of the parameter draws and we use each latent variable draw in a different iteration of Algorithm 1 for Step 1(i). Inspection of parameter and latent variable paths (available in our replication package) suggests very good convergence.

Micro posterior: Sequential Monte Carlo. Step 1(ii) in Algorithm 1 requires sampling, for each i and t, the distribution of $\{\eta_{i,t+s}\}_{s=0}^{S-1}$ conditional on $\{y_{i,t+s}, x_{i,t+s}, Z_{t+s}\}_{s=0}^{S-1}$ taking Q_η, $Q_{\varepsilon,t}$ and $Q_{\text{init},t}$ (evaluated at certain parameter values θ, $\delta_{\varepsilon,t}$ and $\delta_{\text{init},t}$) as given. We do so by Sequential Monte Carlo.[4]

The measurement equation for the problem is $y_{i,t+s} = \eta_{i,t+s} + \varepsilon_{i,t+s}$ for $s = 0, \ldots, S-1$ with state variable $\eta_{i,t+s}$, a first-order Markov process. Let $X_{i,t+s} = (x_{i,t+s}, Z_{t+s}, Z_{t+s-1})'$.

To implement Sequential Monte Carlo, we need two distinct proposal distributions with densities $q_{\text{init},t}(\eta_{it}|y_{it}, x_{it})$ and $q_\eta(\eta_{i,t+s}|\eta_{i,t+s-1}, y_{i,t+s}, X_{i,t+s})$ from which to draw particles. We discuss the calibration of $q_{\text{init},t}$ and q_η below. We also use f_η, $f_{\text{init},t}$ and $f_{\varepsilon,t}$ to denote the densities associated to the quantile functions Q_η, $Q_{\text{init},t}$ and $Q_{\varepsilon,t}$.

The Sequential Monte Carlo algorithm generates $\overline{K}$ particles $\{\{\eta^k_{i,t+s}\}_{k=1}^{\overline{K}}\}_{s=0}^{S-1}$ as follows:

$(s = 0)$
- If y_{it} is missing:
  - Draw independent particles $\{\eta^k_{it}\}_{k=1}^{\overline{K}}$ from the unconditional density $f_{\text{init},t}$.
  - Set the weights $\{w^k_{it}\}_{k=1}^{\overline{K}}$ to $w^k_{it} = 1$.
- If y_{it} is not missing:
  - Draw independent particles $\{\eta^k_{it}\}_{k=1}^{\overline{K}}$ from the proposal, $\eta^k_{it} \sim q_{\text{init},t}(\cdot|y_{it}, x_{it})$.

[4] See Creal (2012) for a review of Sequential Monte Carlo methods and Arellano, Blundell, Bonhomme, and Light (2023) for an application to models with time-varying latent variables.

* Set the weights $\{w_{it}^k\}_{k=1}^{\overline{K}}$ to

$$w_{it}^k = \frac{f_{\text{init},t}(\eta_{it}^k | x_{it}) \cdot f_{\varepsilon,t}(y_{it} - \eta_{it}^k | x_{it})}{q_{\text{init},t}(\eta_{it}^k | y_{it}, x_{it})}.$$

- If $\text{ESS}_t = 1/\sum_{k=1}^{\overline{K}} (w_{it}^k)^2 < \overline{\text{ESS}}$, resample particles from the discrete distribution supported on $\{\eta_{it}^k\}_{k=1}^{\overline{K}}$ with probabilities proportional to $\{w_{it}^k\}_{k=1}^{\overline{K}}$.

$(s > 0)$

- If $y_{i,t+s}$ is missing:
  * Draw particles $\{\eta_{i,t+s}^k\}_{k=1}^{\overline{K}}$ from the conditional density $f_\eta(\cdot | \eta_{i,t+s-1}^k, X_{i,t+s})$.
  * Set the weights $\{w_{i,t+s}^k\}_{k=1}^{\overline{K}}$ to $w_{i,t+s}^k = w_{i,t+s-1}^k$.
- If $y_{i,t+s}$ is not missing:
  * Draw particles $\{\eta_{i,t+s}^k\}_{k=1}^{\overline{K}}$ from the proposal, $\eta_{i,t+s}^k \sim q_\eta(\cdot | \eta_{i,t+s-1}^k, y_{i,t+s}, X_{i,t+s})$.
  * Set the weights $\{w_{i,t+s}^k\}_{k=1}^{\overline{K}}$ to

$$w_{i,t+s}^k = w_{i,t+s-1}^k \times \frac{f_\eta(\eta_{i,t+s}^k | \eta_{i,t+s-1}^k, X_{i,t+s}) \cdot f_{\varepsilon,t}(y_{i,t+s} - \eta_{i,t+s}^k | x_{i,t+s})}{q_\eta(\eta_{i,t+s}^k | \eta_{i,t+s-1}^k, y_{i,t+s}, X_{i,t+s})}.$$

- If $\text{ESS}_{t+s} = 1/\sum_{k=1}^{\overline{K}} (w_{i,t+s}^k)^2 < \overline{\text{ESS}}$, resample particles.

This algorithm can be efficiently vectorized over k and parallelized across units i. We use $\overline{K} = 5,000$ particles, choosing one of them at random (with weights $\{w_{i,t+S}^k\}_{k=1}^{\overline{K}}$) at the end of the algorithm as the draw $\bar{\eta}_{it}^S(j) = \{\eta_{i,t+s}(j)\}_{s=0}^{S-1}$ in Step 1(ii) of Algorithm 1. We also set $\overline{\text{ESS}} = \overline{K}/4$ as the threshold for resampling.

We calibrate the proposals as follows. We take $q_{\text{init},t}(\eta_{it} | y_{it}, x_{it})$ to be the density of η_{it} conditional on (y_{it}, x_{it}) implied by the model

$$\begin{aligned} y_{it} &= \eta_{it} + \varepsilon_{it}, & \varepsilon_{it} &\sim N(0, s_\varepsilon^2), \\ \eta_{it} &= \psi_{\text{init}}(x_{it})' b_{\text{init},t} + u_{it}, & u_{it} &\sim N(0, s_{\text{init}}^2), \end{aligned}$$

where ψ_{init} is the same vector of basis functions used for $Q_{\text{init},t}$ and we update $b_{\text{init},t}, s_{\text{init}}^2, s_\varepsilon^2$ by least squares in each iteration of Algorithm 1. The proposal then becomes

$$\begin{aligned} q_{\text{init},t}(\eta_{it} | y_{it}, x_{it}) &= N(\mu_{\text{init},t}(y_{it}, x_{it}), \omega_{\text{init}}^2), \\ \mu_{\text{init},t}(y_{it}, x_{it}) &= (1 - \phi_{\text{init}}) \psi_{\text{init}}(x_{it})' b_{\text{init},t} + \phi_{\text{init}} y_{it} \text{ with } \phi_{\text{init}} = s_{\text{init}}^2 / (s_{\text{init}}^2 + s_\varepsilon^2), \end{aligned}$$

$$\omega^2_{\text{init}} = (1/s^2_{\text{init}} + 1/s^2_{\varepsilon})^{-1}.$$

For $q_\eta(\eta_{i,t+s}|\eta_{i,t+s-1}, y_{i,t+s}, X_{i,t+s})$ we use the density of $\eta_{i,t+s}$ conditional on $(\eta_{i,t+s-1}, y_{i,t+s}, X_{it})$ implied by the model

$$\begin{aligned} y_{i,t+s} &= \eta_{i,t+s} + \varepsilon_{i,t+s}, & \varepsilon_{i,t+s} &\sim N(0, s^2_\varepsilon), \\ \eta_{i,t+s} &= \overline{\psi}_\eta(\eta_{i,t+s-1}, X_{i,t+s})' b_\eta + u_{i,t+s}, & u_{i,t+s} &\sim N(0, s^2_\eta), \end{aligned}$$

where $\overline{\psi}_\eta(\eta_{i,t+s-1}, X_{i,t+s}) = \psi(\eta_{i,t+s-1}, x_{i,t+s}) \otimes \varphi(Z_{t+s}, Z_{t+s-1})$ contains the basis functions used for Q_η and we update b_η, s^2_η by least squares in each iteration too. The proposal is then

$$\begin{aligned} q_\eta(\eta_{i,t+s}|\eta_{i,t+s-1}, y_{i,t+s}, X_{i,t+s}) &= N(\mu_\eta(\eta_{i,t+s-1}, y_{i,t+s}, X_{i,t+s}), \omega^2_\eta), \\ \mu_\eta(\eta_{i,t+s-1}, y_{i,t+s}, X_{i,t+s}) &= (1-\phi_\eta)\overline{\psi}_\eta(\eta_{i,t+s-1}, X_{i,t+s})' b_\eta + \phi_\eta y_{i,t+s} \text{ with } \phi_\eta = s^2_\eta/(s^2_\eta + s^2_\varepsilon), \\ \omega^2_\eta &= (1/s^2_\eta + 1/s^2_\varepsilon)^{-1}. \end{aligned}$$

As a practical matter, to ensure thorough exploration of the tails of the micro posterior, we switch from normal to Laplace (with the same location and scale) below the 2.5 and above the 97.5 percentiles of the proposal distributions.

C.3 Asymptotic approximations

We develop next the large sample properties of $\widehat{\theta}$ and the plug-in estimator $\widehat{\gamma} = \gamma(\widehat{\theta})$. Our asymptotic analysis assumes $N_t, T \to \infty$ with S fixed. The data generating process (DGP) is given by Assumptions 1, 2, and 3 with the flexible parametric specification in (7), (8) and (9), with regularity conditions.[5] In Algorithm 1, when J is fixed, $\widehat{\theta}$ depends not just on the data but on the realizations of latent variables drawn in the E step. In practice, J is set to a large number to reduce the influence of simulation noise and starting values. In light of that, here we focus on the limit case $J \to \infty$.[6]

[5] As discussed in the text, we hold the dimension of the basis functions (i.e., ψ, φ, ψ_{init}, ψ_ε, ψ_{lo}, ψ_{up}, etc.) and L fixed. Alternatively, these could be viewed as tuning parameters that grow with the sample size in a nonparametric sieve approach (Newey, 1997; Chen, 2007) but we leave that for future research.

[6] Analyses of the fixed-J case for cross-sectional and short-panel setups can be found in Nielsen (2000) and Arellano and Bonhomme (2016).

Thus, we view the estimator as the (approximate) solution to

$$\frac{1}{T}\sum_{t=1}^{T}\overline{M}_{\theta,t}(\widehat{\theta},\widehat{\delta_t},\widehat{\lambda})=0,$$
$$\overline{M}_{\delta,t}(\widehat{\theta},\widehat{\delta_t},\widehat{\lambda})=0,\quad t=1,\ldots,T.$$

where $\overline{M}_{\theta,t}(\theta,\delta_t,\lambda) = \int\left[N_t^{-1}\sum_{i\in\mathcal{I}_t}\int m_\theta(\theta,\bar{y}_{it}^S,\bar{\eta}^S,\bar{Z}^S)\,f(\bar{\eta}^S|\bar{y}_{it}^S,\bar{Z}^S,\theta,\delta_t)\,d\bar{\eta}^S\right]\,f(\bar{Z}^S|\overline{W},\lambda)\,d\bar{Z}^S$ and $\overline{M}_{\delta,t}(\theta,\delta_t,\lambda) = \int\left[N_t^{-1}\sum_{i\in\mathcal{I}_t}\int m_\delta(\delta,\bar{y}_{it}^S,\bar{\eta}^S)\,f(\bar{\eta}^S|\bar{y}_{it}^S,\bar{Z}^S,\theta,\delta_t)\,d\bar{\eta}^S\right]\,f(\bar{Z}^S|\overline{W},\lambda)\,d\bar{Z}^S$.

Doing a Taylor expansion to the two equations above and using $\overline{D}_{pq,t}$ to denote a matrix of first derivatives of $\overline{M}_{p,t}$ for $p=\theta,\delta$ with respect to $q=\theta,\delta,\lambda$ where each row is evaluated at a possibly different intermediate value between $(\widehat{\theta},\widehat{\delta_t},\widehat{\lambda})$ and the true value $(\theta_0,\delta_{0t},\lambda_0)$,

$$\sqrt{T}(\widehat{\theta}-\theta_0)=\left[\frac{1}{T}\sum_{t=1}^{T}(\overline{D}_{\theta\delta,t}\overline{D}_{\delta\delta,t}^{-1}\overline{D}_{\delta\theta,t}+\overline{D}_{\theta\theta,t})\right]^{-1}$$
$$\times\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{M}_{\theta,0t}+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\overline{D}_{\theta\delta,t}\overline{D}_{\delta\delta,t}^{-1}\overline{M}_{\delta,0t}+\left[\frac{1}{T}\sum_{t=1}^{T}(\overline{D}_{\theta\delta,t}\overline{D}_{\delta\delta,t}^{-1}\overline{D}_{\delta\lambda,t}+\overline{D}_{\theta\lambda,t})\right]\cdot\sqrt{T}(\widehat{\lambda}-\lambda_0)\right)$$

where $\overline{M}_{p,0t}=\overline{M}_{p,t}(\theta_0,\delta_{0t},\lambda_0)$ for $p=\theta,\delta$. Assuming that our parametric model is correctly specified and standard regularity conditions on $\widehat{\lambda}$, one can show that

$$\frac{1}{T}\sum_{t=1}^{T}(\overline{D}_{\theta\delta,t}\overline{D}_{\delta\delta,t}^{-1}\overline{D}_{\delta\theta,t}+\overline{D}_{\theta\theta,t})\ \xrightarrow{\text{p}}\ D_{\theta\theta,0},$$
$$\frac{1}{T}\sum_{t=1}^{T}(\overline{D}_{\theta\delta,t}\overline{D}_{\delta\delta,t}^{-1}\overline{D}_{\delta\lambda,t}+\overline{D}_{\theta\lambda,t})\ \xrightarrow{\text{p}}\ D_{\theta\lambda,0},$$

where $D_{\theta\theta,0}$ and $D_{\theta\lambda,0}$ are two fixed matrices and $D_{\theta\theta,0}$ is non-singular.

One can also apply a central limit theorem to the scaled averages to show that

$$\sqrt{T}\begin{pmatrix}T^{-1}\sum_{t=1}^{T}\overline{M}_{\theta,0t}\\ T^{-1}\sum_{t=1}^{T}\overline{D}_{\theta\delta,t}\overline{D}_{\delta\delta,t}^{-1}\overline{M}_{\delta,0t}\\ \widehat{\lambda}-\lambda_0\end{pmatrix}\ \xrightarrow{\text{d}}\ N(0,\Omega_0)$$

for some symmetric, positive semi-definite matrix Ω_0. Collecting all pieces, the asymptotic distribution of $\widehat{\theta}$ follows from Slutsky, whereas that of $\widehat{\gamma}$ follows from the delta method.

C.4 Bootstrap approach

The asymptotic analysis suggests that a parametric bootstrap approach can be justified for statistical uncertainty quantification. It also highlights the role of the omitted aggregate factor G_t and the need to account for the cross-sectional dependence that such factors may induce. In addition, some objects of interest are primarily identified by cross-sectional variation. A key advantage of the parametric bootstrap is that it allows us to replicate the unit-level dependence caused by sampling the same units into different subpanels, a natural feature in our time series of panels framework.

Omitted aggregate factors. We model the cross-sectional dependence as follows. We let $G_t = (G_{\eta,t}, G_{\varepsilon,t}, G_{\text{init},t})'$ where entries are i.i.d. uniformly distributed on $(0,1)$ and mutually independent. Then, we assume the micro-level errors in our model are

$$\begin{aligned} u_{it} &= \Phi\left(c_\eta \Phi^{-1}(G_{\eta,t}) + \sqrt{1-c_\eta^2}\Phi^{-1}(\tilde{u}_{it})\right), \\ v_{it} &= \Phi\left(c_\varepsilon \Phi^{-1}(G_{\varepsilon,t}) + \sqrt{1-c_\varepsilon^2}\Phi^{-1}(\tilde{v}_{it})\right), \\ \nu_{i,t_0} &= \Phi\left(c_{\text{init}} \Phi^{-1}(G_{\text{init},t_0}) + \sqrt{1-c_{\text{init}}^2}\Phi^{-1}(\tilde{\nu}_{i,t_0})\right), \end{aligned}$$

where $\tilde{u}_{it}, \tilde{v}_{it}, \tilde{\nu}_{i,t_0}$ are i.i.d. uniformly distributed on $(0,1)$ and mutually independent. The parameters c_η, c_ε and c_{init} are pinned down by the common variability in the micro-level errors—e.g., $\hat{c}_\eta = [T^{-1}\sum_{t=1}^T(\sum_{i\in\mathcal{I}_t}\Phi^{-1}(u_{it})/N_t)^2]^{1/2}$ consistently estimates c_η as $T, N_t \to \infty$. Given estimates $\widehat{\theta}$, $\{\widehat{\delta_t}\}_{t=1}^T$ and $\widehat{\lambda}$, we estimate c_η, c_ε and c_{init} by performing steps 1(i) and 1(ii) of Algorithm 1, computing the implied ranks u_{it}, v_{it} and ν_{i,t_0}, and using them as above (we repeat this for 100 iterations, averaging the parameter paths across iterations).

Unit overlap. The time series of panels data structure allows the same unit to be part of different subpanels. Because our model is biennial, it already specifies the cross-panel dependence if the year gap between two subpanels is even: apply Equation (1) recursively.

When the same unit i appears in consecutive odd- and even-year panels (denoted t and t') we assume the following for the micro-level errors net of their common component:

$$\left(\Phi^{-1}(\tilde{u}_{it}) \quad \Phi^{-1}(\tilde{u}_{it'})\right)' \sim N\left(0, \begin{pmatrix} 1 & d_\eta \\ d_\eta & 1 \end{pmatrix}\right),$$

$$\left(\Phi^{-1}(\tilde{v}_{it}) \quad \Phi^{-1}(\tilde{v}_{it'})\right)' \sim N\left(0, \begin{pmatrix} 1 & d_\varepsilon \\ d_\varepsilon & 1 \end{pmatrix}\right),$$

$$\left(\Phi^{-1}(\tilde{\nu}_{it}) \quad \Phi^{-1}(\tilde{\nu}_{it'})\right)' \sim N\left(0, \begin{pmatrix} 1 & d_{\text{init}} \\ d_{\text{init}} & 1 \end{pmatrix}\right).$$

We estimate the parameters d_η, d_ε and d_{init} within the same algorithm described above for c_η, c_ε and c_{init}. To this end, we use the correlation of the idiosyncratic components of the ranks across any two consecutive years.

Implementation. Given estimates of $(c_\eta, c_\varepsilon, c_{\text{init}}, d_\eta, d_\varepsilon, d_{\text{init}})$, it is easy to obtain bootstrap samples that reflect the estimated degrees of cross-sectional and unit-level dependence. The following procedure reproduces the repetition and overlap patterns in the data:

1) Simulate the time series of aggregate factors $\{G_{\eta,t}, G_{\varepsilon,t}, G_{\text{init},t}\}_{t=1}^{T}$.

2) For each unit i determine the first (t_0) and last (t_1) period in the dataset. Next,

   (i) draw the path of idiosyncratic shocks $\{\tilde{u}_{it}, \tilde{v}_{it}, \tilde{\nu}_{it}\}_{t_0 \le t \le t_1}$ imposing the correlations d_η, d_ε and d_{init} across consecutive periods;

   (ii) combine aggregate and idiosyncratic factors to obtain $\{u_{it}, v_{it}, \nu_{it}\}_{t_0 \le t \le t_1}$ imposing the cross-sectional dependence implied by c_η, c_ε and c_{init};

   (iii) for the first two periods, use $Q_{\text{init},t}$ and ν_{it} to generate η_{it};

   (iv) for every other period, use Q_η and u_{it} to generate η_{it};

   (v) for all periods, use $Q_{\varepsilon,t}$ and v_{it} to generate ε_{it};

   (vi) form $y_{it} = \eta_{it} + \varepsilon_{it}$ for all $t_0 \le t \le t_1$.

3) Assign the data to the appropriate unit and time cell.

D Additional empirical results

This appendix expands on three sets of empirical results. Figure D.1 reports our nonlinear measure of aggregate risk exposure $\beta(u, \eta, Z_t, Z_{t-1}, x)$ along quantiles of the rank u and past persistent income η, as well as averaged over η. This complements Figure 6 in the text. The main nonlinearity in the figure is the increase in exposure to aggregate shocks during recessions and its decline during expansions. This form of aggregate state dependence at the micro level is not captured by linear models and plays a paramount role in macro risk calculations, as discussed in Section 7.

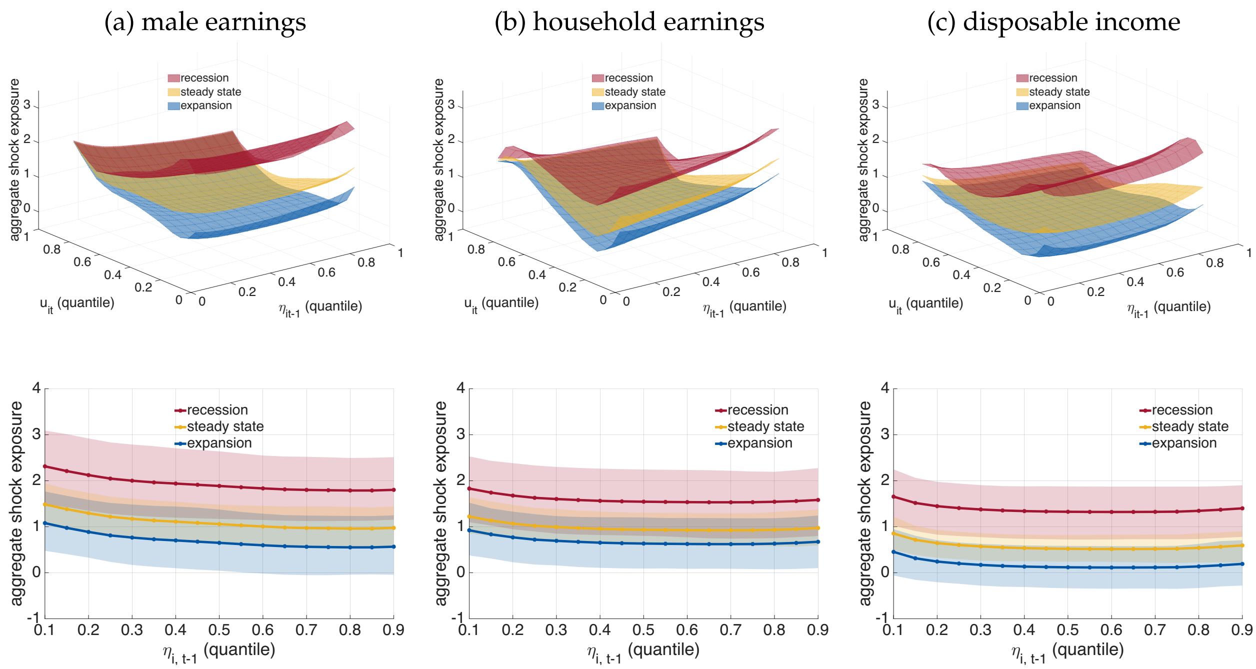

FIGURE D.1. Nonlinear exposure to aggregate shocks.
Note: We report the aggregate risk exposure $\beta(u, \eta, Z_t, Z_{t-1}, x)$ by quantile of the shock $u = u_{it}$ and of past persistent income $\eta = \eta_{i,t-1}$ (upper panels), or averaged across $\eta = \eta_{i,t-1}$ (lower panels). Here, age $x = x_{it}$ is averaged out, $Z_{t-1} = \tilde{Z}_{ss}$ and Z_t is a recession $\tilde{Z}_r$, the steady state $\tilde{Z}_{ss}$ or an expansion $\tilde{Z}_e$ (see Section 5.1). Shaded areas in the lower panels represent 90% pointwise confidence bands.

Figure D.2 displays estimates of dispersion and kurtosis, together with their differences between recessions and expansions. This complements Figure 7 in the text that documents the cyclical pattern of skewness. We find a slight increase in the dispersion and decrease in the kurtosis of persistent income shocks in recessions compared to expansions, but they are generally not statistically non-zero. Although different in methodology and data, our results are in line with the findings in Guvenen, Ozkan, and Song (2014).

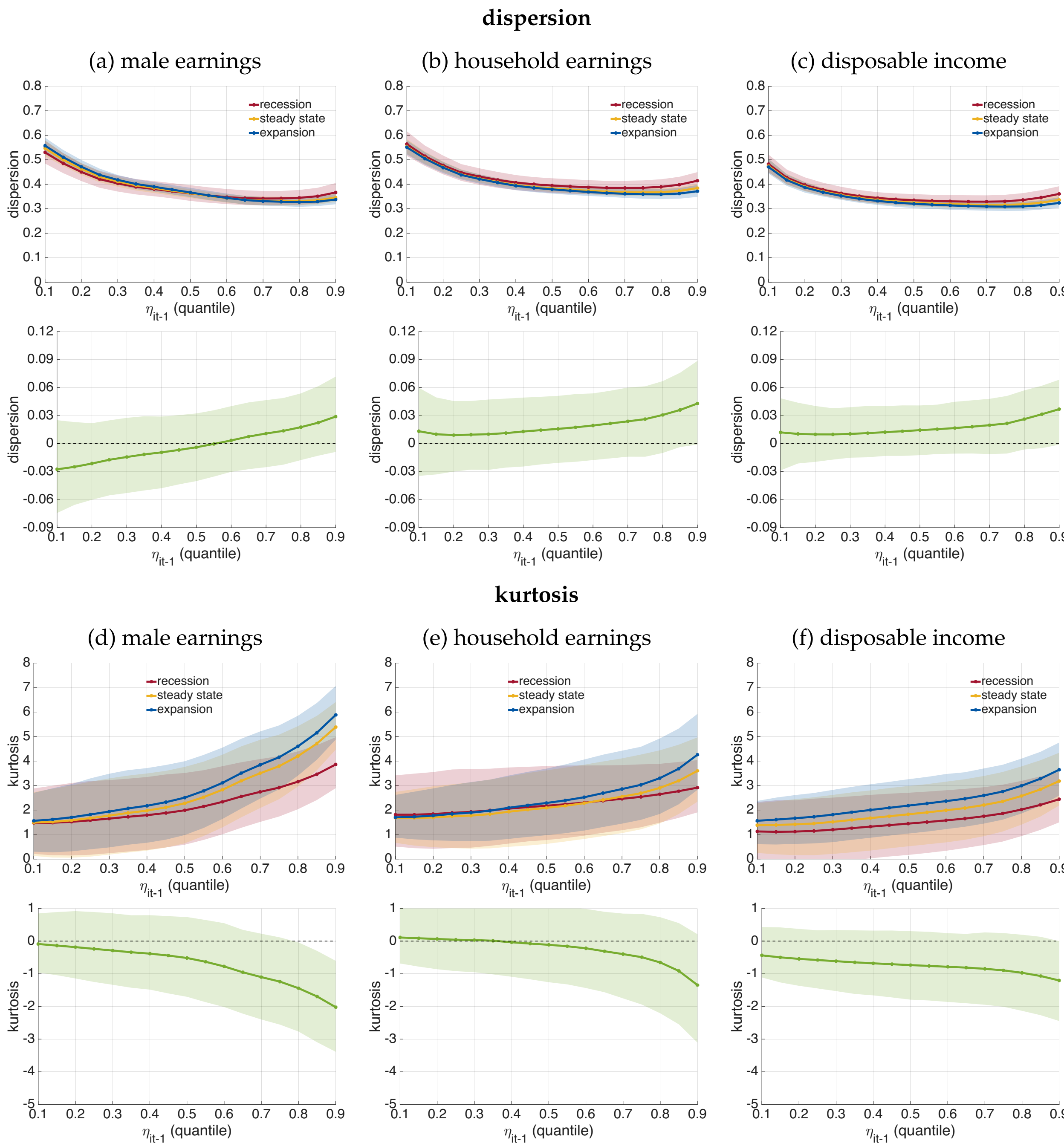

FIGURE D.2. Measures of dispersion and kurtosis.

Note: The first and third rows report the dispersion disp(η, Z_t, Z_{t-1}, x) and kurtosis kurt(η, Z_t, Z_{t-1}, x) defined in footnote 8 by past persistent income $\eta = \eta_{i,t-1}$ where age $x = x_{it}$ is averaged out, $Z_{t-1} = \tilde{Z}_{ss}$ and Z_t is a recession $\tilde{Z}_r$, the steady state $\tilde{Z}_{ss}$ or an expansion $\tilde{Z}_e$ (see Section 5.1). The second and fourth rows show the gaps between recession and expansion. Shaded areas represent 90% confidence bands.

E Additional material on impulse response analysis

This appendix expands Section 6 in various directions. We relate impulse responses to derivatives with respect to some macro and micro shocks in Section E.1. We characterize analytically the link between impulse responses, nonlinear persistence and exposure to aggregate shocks in Section E.2. Sections E.3, E.4 and E.5 contain additional results.

E.1 Perturbations and shocks

Having defined impulse responses using perturbations of state variables in the main text, we can next relate them to derivatives with respect to certain macro and micro shocks, which we will denote $\widetilde{V}_t$ and $\widetilde{u}_{i,t-1}$. In other words, there is a duality relation between deterministic perturbations of state variables and the stochastic disturbances that embody macro and micro sources of income risk. More specifically,

$$\text{IRF}_{\eta Z}(h;\delta) = \frac{E\left[\eta_{i,t+h} \,\middle|\, \eta_{i,t-1}, \widetilde{V}_t = \delta, Z_{t-1}\right] - E\left[\eta_{i,t+h} \,\middle|\, \eta_{i,t-1}, \widetilde{V}_t = 0, Z_{t-1}\right]}{\delta},$$
$$\text{IRF}_{\eta\eta}(h,\delta) = \frac{E\left[\eta_{i,t+h-1} \,\middle|\, \widetilde{u}_{i,t-1} = \delta, \eta_{i,t-2}, Z_t, Z_{t-1}\right] - E\left[\eta_{i,t+h-1} \,\middle|\, \widetilde{u}_{i,t-1} = 0, \eta_{i,t-2}, Z_t, Z_{t-1}\right]}{\delta}$$

and, for infinitesimal changes,

$$\text{IRF}_{\eta Z}(h) = \frac{\partial E\left[\eta_{i,t+h} \,\middle|\, \eta_{i,t-1}, \widetilde{V}_t, Z_{t-1}\right]}{\partial \widetilde{V}_t}, \quad \text{IRF}_{\eta\eta}(h) = \frac{\partial E\left[\eta_{i,t+h-1} \,\middle|\, \widetilde{u}_{i,t-1}, \eta_{i,t-2}, Z_t, Z_{t-1}\right]}{\partial \widetilde{u}_{i,t-1}}.$$

The implied shocks are given by

$$\widetilde{V}_t = g\left(Q_Z(Z_{t-1}, V_t)\right) - g(Z^b),$$
$$\widetilde{u}_{i,t-1} = g\left(Q_\eta(\eta_{i,t-2}, Z_{t-1}, Z_{t-2}, u_{i,t-1})\right) - g(\eta^b),$$

and lead to the representations

$$Z_t = Q_Z(Z_{t-1}, Q_Z^{-1}[Z_{t-1}, g^{-1}(g(Z^b) + \widetilde{V}_t)]),$$
$$\eta_{i,t-1} = Q_\eta(\eta_{i,t-2}, Z_{t-1}, Z_{t-2}, Q_\eta^{-1}[\eta_{i,t-2}, Z_{t-1}, Z_{t-2}, g^{-1}(g(\eta^b) + \widetilde{u}_{i,t-1})]).$$

These representations are local to the benchmark values and to the normalization rule g.

E.2 Impulse responses, nonlinear persistence and aggregate exposures

To get some intuition on the role of nonlinearities in shaping impulse responses we look at the derivative-based definitions. First, by recursive substitution on Equation (3), let

$$Z_{t+h} = q_{Z,h}(\boldsymbol{V}_{t+1}^{h-1}, Z_t) = \sum_{\ell=0}^{h-1} \Phi^{\ell} \Sigma_V^{1/2} V_{t+h-\ell} + \Phi^h Z_t, \quad h = 0, 1, \dots \tag{E.1}$$

Combining Equations (13) with (E.1), we have for $h = 1, 2, \dots$

$$\begin{aligned} q_{\eta,h}(\boldsymbol{u}_{it}^h, \boldsymbol{V}_{t+1}^{h-1}, \eta_{i,t-1}, Z_t, Z_{t-1}) = Q_\eta\Big(& q_{\eta,h-1}(\boldsymbol{u}_{it}^{h-1}, \boldsymbol{V}_{t+1}^{h-2}, \eta_{i,t-1}, Z_t, Z_{t-1}), \dots \\ & q_{Z,h}(\boldsymbol{V}_{t+1}^{h-1}, Z_t), q_{Z,h-1}(\boldsymbol{V}_{t+1}^{h-2}, Z_t), u_{i,t+h}\Big), \end{aligned}$$

with the recursion beginning at $q_{\eta,0}(\boldsymbol{u}_{it}^h, \boldsymbol{V}_{t+1}^{h-1}, \eta_{i,t-1}, Z_t, Z_{t-1}) = Q_\eta\left(\eta_{i,t-1}, Z_t, Z_{t-1}, u_{it}\right)$.

It will also be useful to define the following random variables:

$$\rho_{it} = \rho(u_{it}, \eta_{i,t-1}, Z_t, Z_{t-1}), \quad \beta_{it} = \beta(u_{it}, \eta_{i,t-1}, Z_t, Z_{t-1}), \quad \gamma_{it} = \gamma(u_{it}, \eta_{i,t-1}, Z_t, Z_{t-1}),$$

where, similarly to ρ and β, the nonlinear measure γ is

$$\gamma(u_{it}, \eta_{i,t-1}, Z_t, Z_{t-1}) = \frac{\partial Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, u_{it})}{\partial Z_{t-1}}.$$

In particular, ρ_{it} and β_{it} are the values of the nonlinear persistence and household exposure to aggregate shocks defined in Section 2 for a given realization of micro and macro state variables and shocks, and γ_{it} measures the nonlinear exposure of the persistent component of income to the lagged macro variable Z_{t-1}.

The impulse responses of the macro state using our methodology is

$$\text{IRF}_{ZZ}(h) = \lim_{\delta \to 0} \frac{E\left[Z_{t+h} \,\middle|\, Z_t = Z^b + \Delta(\delta) \right] - E\left[Z_{t+h} \,\middle|\, Z_t = Z^b \right]}{\delta} = \Phi^h \times \left\{ g'(Z^b) \right\}^{-1}.$$

Next, taking derivatives and exchanging the order of differentiation and integration,

$$\text{IRF}_{\eta Z}(h) = E\left[\sum_{\ell=0}^{h} \beta_{i,t+h-\ell} \Phi^{h-\ell} \left(\prod_{j=0}^{\ell-1} \rho_{i,t+h-j} \right) \middle| \, \eta_{i,t-1} = \eta^b, Z_t, Z_{t-1} \right] \times \left\{ g'(Z^b) \right\}^{-1}$$

$$+ E\left[\sum_{\ell=0}^{h-1} \gamma_{i,t+h-\ell} \Phi^{h-\ell-1} \left(\prod_{j=0}^{\ell-1} \rho_{i,t+h-j} \right) \middle| \eta_{i,t-1} = \eta^b, Z_t, Z_{t-1} \right] \times \left\{ g'(Z^b) \right\}^{-1},$$

$$\mathrm{IRF}_{\eta\eta}(h) = E\left[\prod_{\ell=1}^{h} \rho_{i,t+h-\ell} \middle| \eta_{i,t-1} = \eta^b, Z_t, Z_{t-1} \right] \times \left\{ g'(\eta^b) \right\}^{-1}.$$

The expressions for $\mathrm{IRF}_{ZZ}(h)$, $\mathrm{IRF}_{\eta Z}(h)$ and $\mathrm{IRF}_{\eta\eta}(h)$ have two parts: The first is independent of the rule g, whereas the second part is independent of the horizon h. Hence the first part sets the dynamic propagation of uncertainty and is fully determined by the macro state persistence parameter Φ, the nonlinear persistence measure ρ_{it} and the micro elasticities to macro shocks β_{it} and γ_{it}. They generalize the dynamic transmission patterns from the linear homogeneous income process, $\eta_{it} = \rho\eta_{i,t-1} + \beta Z_t + \gamma Z_{t-1} + u_{it}$, for which

$$\mathrm{IRF}_{\eta Z}(h) = \left(\beta \sum_{\ell=0}^{h} \Phi^{h-\ell} \rho^{\ell} + \gamma \sum_{\ell=0}^{h-1} \Phi^{h-\ell-1} \rho^{\ell} \right) \times \left\{ g'(Z^b) \right\}^{-1},$$
$$\mathrm{IRF}_{\eta\eta}(h) = \rho^h \times \left\{ g'(\eta^b) \right\}^{-1},$$

by introducing dependence on the potential history of future shocks.

The second part fixes the scale of the IRF and is determined by the rule g. For example, $g'(z)$ is one for the unit rule and the conditional density of the state being perturbed at the benchmark value for the rank rule. It follows that, for infinitesimal perturbations, all IRFs are scaled versions of unit-rule IRFs, which in turn reflect nonlinear persistence and micro exposures to macro shocks.

The derivation offers insights into the relationship between the persistence of macro and micro shocks. Empirically, we find low persistence of macro shocks ($\mathrm{IRF}_{\eta Z}(h)$ roughly proportional to $\mathrm{IRF}_{ZZ}(h)$ indicating a short-lived response) but high persistence of micro shocks ($\mathrm{IRF}_{\eta\eta}(h)$ decays slowly). These patterns raise the question of whether a nonlinear dynamic common factor restriction analogous to that of linear partial adjustment models (Griliches, 1961, 1967; Sargan, 1964, 1980) holds. Specifically, if

$$\gamma_{i,t+1} = -\rho_{i,t+1}\beta_{it}, \tag{E.2}$$

then

$$\mathrm{IRF}_{\eta Z}(h) = E\left[\beta_{i,t+h} \,\middle|\, \eta_{i,t-1} = \eta^b, Z_t, Z_{t-1} \right] \mathrm{IRF}_{ZZ}(h).$$

Constructing a test of this functional restriction is beyond the scope of our paper, but a look at our estimates does not offer conclusive evidence in its favor. For example, according to the point estimates for disposable income, the average $\gamma_{i,t+1}$ is around -1, the average $\rho_{i,t+1}$ is around 0.92 and the average β_{it} is 1.3 in a typical recession, 0.6 in steady state and 0.2 in an mild expansion. The three quantities also vary substantially over the distribution of past persistent income and micro ranks. All of this suggests a departure from the dynamic common factor restriction (E.2), the size of which depends on macro and micro state variables.

E.3 Additional IRF figures: comparison to MBC shocks

Figure E.1 compares the IRF of each entry in W_t to shock V_t from our baseline specification (red, diamonds) against the IRFs to the MBC shock of Angeletos, Collard, and Dellas (2020) obtained by targeting the unemployment rate FEVD (blue, circles).

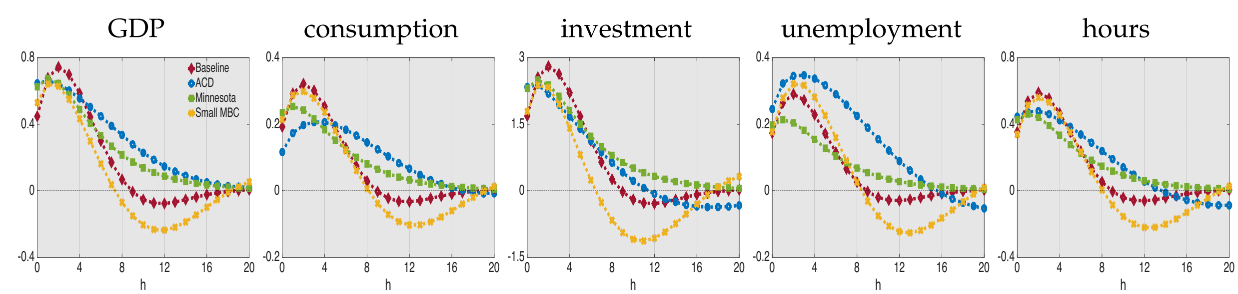

FIGURE E.1. IRFs of W_t to V_t and MBC shock

Note: We show IRFs of W_t to the following: the V_t shock from our baseline model (red, diamonds), the MBC shock from Angeletos et al. (2020) (blue, circles), a V_t shock from a dynamic factor model with a Minnesota prior (green, squares), and an MBC shock from a 5-variable VAR(2) with a flat prior (yellow, crosses).

The takeaway from Figure E.1 is that the two approaches generally agree on the relative impact among variables and the cumulative impact over the first two years, but they differ on their distribution over time. Specifically, our baseline specification places a larger share of the impact on the first year compared to the original MBC shock.

While the discrepancy is small relative to the statistical uncertainty around the IRFs, part of it can be attributed to the choice of prior. In our case, the dynamic factor structure already achieves, without further penalization, adequate dimension reduction. Instead, the 10-variable VAR(2) underlying the original MBC shock IRFs is based on a Minnesota prior. This choice is understandable, but penalizes deviations from unit roots that may bias the estimated persistence upward. To explore the issue, Figure E.1 shows two additional estimates: IRFs obtained from a dynamic factor model under a Minnesota prior (green, squares), and IRFs for an MBC-type shock from a small 5-variable VAR(2) on W_t under a flat prior (gold, crosses).[7] Consistent with our claim, the former mimics the persistence of the original MBC responses while the latter matches our baseline closely. But reassuringly, the small-model MBC shock and our V_t shock are highly correlated as seen in Figure E.2.

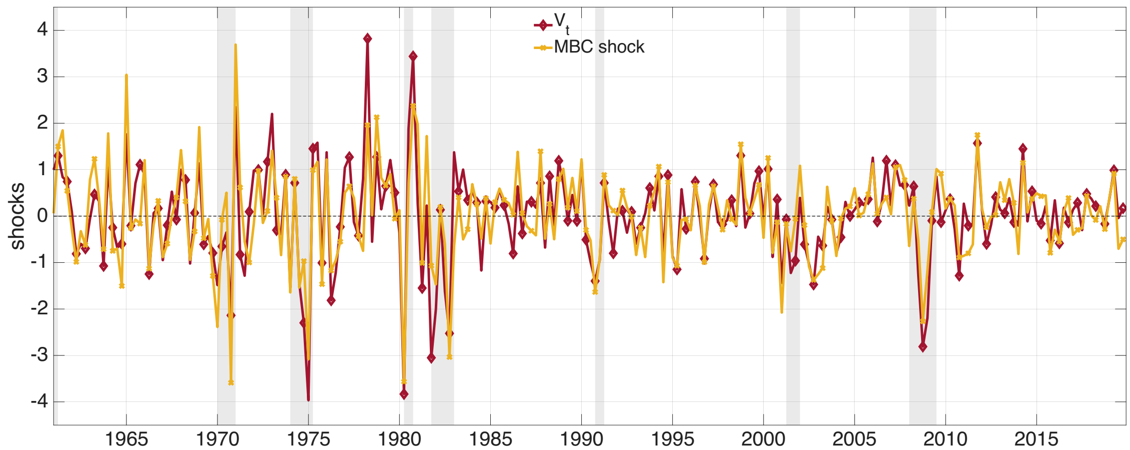

FIGURE E.2. V_t and MBC shocks

Note: We plot posterior median estimates of V_t from our baseline (red, diamonds) and the MBC shock from a 5-variable VAR(2) model with a flat prior (gold, crosses). Red areas indicate NBER-dated recessions.

It follows from the preceding discussion that pinning down the persistence in macro IRFs is empirically difficult. However, our main results are robust to this feature. Because there is very little filtering uncertainty about Z_t, the choice of prior has practically no effect on the estimation of the income process, and objects such as $\rho(\cdot)$, $\text{sk}(\cdot)$ and $\beta(\cdot)$ remain the same. Higher persistence in Z_t produces slower decay in $\text{IRF}_{\eta Z}(h)$ compared to Figure 8 and slightly larger costs of aggregate risk compared to Figure 11, but these results cannot be distinguished statistically from our baseline.[8]

E.4 Additional IRF figures: local projection estimates

In Figure E.3 we report estimates of macro impulse responses (multiplied by -1 to emulate the trajectory after a negative shock) obtained by panel local projections. To be concrete, for each horizon h, we regress $y_{i,t+h}$ on Z_t controlling for $y_{i,t-1}$, Z_{t-1}, a second-order Hermite

[7] For the factor model we set lag lengths to 4 and calibrate the prior to $E[\Phi_\ell] = E\left[\phi_{j\ell}\right] = 1\{\ell = 1\}$ and $\text{Var}(\Phi_\ell) = \text{Var}\left(\phi_{j\ell}\right) = 0.5/\ell^2$. For the MBC shock, we target the unemployment rate FEVD.

[8] These robustness checks are available in our replication package.

polynomial on age x_{it} and unit fixed effects. We compute the t-LAHR confidence intervals proposed by Almuzara and Sancibrián (2024) to assess statistical precision.

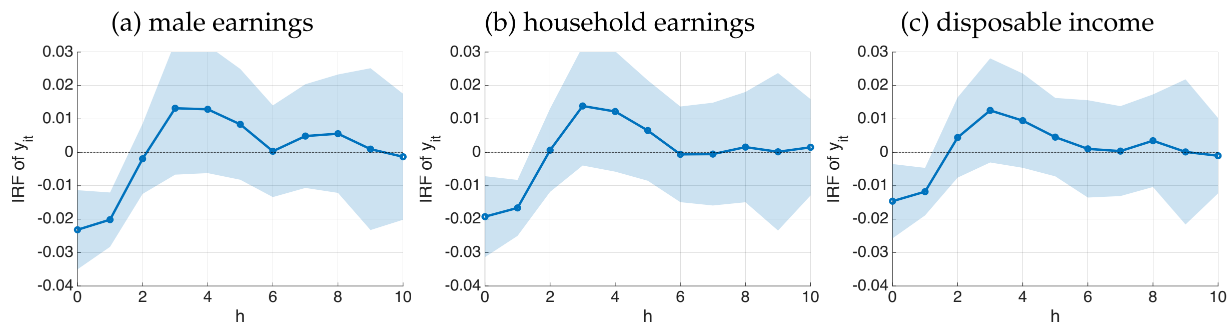

FIGURE E.3. Local projection estimates of macro impulse responses

Note: Panels (a), (b) and (c) display IRFs of y_{it} to a negative macro shock for different income definitions: Z_t is scaled by the standard deviation of log GDP per capita for comparability with the IRFs in the main text. Shaded areas are 90% t-LAHR pointwise confidence bands.

One advantage of this exercise is that pooling the household-level data from the time series of panels allows us to measure the average impulse responses at the annual (rather than biennial) frequency. This reveals a significant response to macro shocks on impact ($h = 0$) and in the first year following the shock ($h = 1$).[9] On the other hand, although these responses correspond to y_{it}, not to η_{it}, the estimates are quantitatively similar to the ones in Figure 8, with larger responses for male earnings compared to disposable income.

E.5 Additional IRF figures: positive shocks

Figure E.4 shows responses to positive macro and micro perturbations, complementing Figure 8 (panels (b) to (d)) and Figure 10. For the estimates of $\text{IRF}_{\eta Z}$ on the upper panels we apply a positive perturbation to Z_t around the steady state benchmark $Z^b = \tilde{Z}_{ss}$ calibrated to $\delta = \sigma_V$ with $\sigma_V^2 = \text{Var}(Z_t \mid Z_{t-1})$. This emulates a mild expansionary aggregate shock. The implied trajectory for Z_t (annualized and scaled to log GDP per capita) is the mirror image of panel (d) in Figure 8, and we refer the reader to the main text to get a sense of the macro implications of the underlying experiment.

For the estimates of $\text{IRF}_{\eta\eta}$ on the lower panels we apply a negative perturbation δ that implies a 10% increase in $\eta_{i,t-1}$. Similar to Figure 10, we hold Z_t and Z_{t-1} at their steady state value $\tilde{Z}_{ss}$ and multiply responses by 0.1 for ease of interpretation.

[9]The figure is also indicative of some overshooting for $h = 3, 4, 5$, albeit not statistically significant.

The main takeaway from the figure is that, as in our analysis of negative perturbations, macro responses are short-lived while micro responses are more persistent. The difference with the negative-shock case is that $\text{IRF}_{\eta Z}$ displays a stronger overshooting effect (i.e., the response crossing the zero line) after $h = 2$, particularly for disposable income.

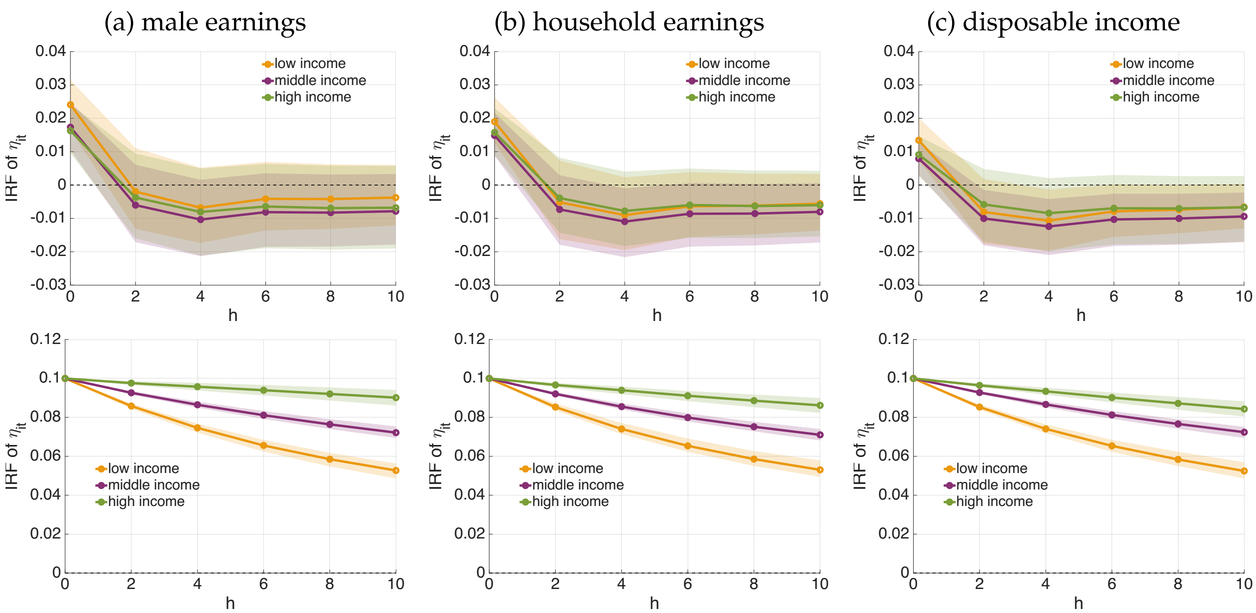

FIGURE E.4. Macro and micro impulse responses to positive shocks

Note: Panels (a), (b) and (c) display IRFs of η_{it} to positive macro (upper panel) and micro (lower panel) shocks for different income measures with $Z_t^b = Z_{t-1} = \tilde{Z}_{ss}$ and $\eta_{i,t-1}^b$ set to the 10th (low), 50th (middle) and 90th (high) percentiles of the persistent income distribution. Shaded areas are 90% confidence bands.

References

Almuzara, M. (2020): "Heterogeneity in Transitory Income Risk." Working paper.

Almuzara, M. and V. Sancibrián (2024): "Micro Responses to Macro Shocks." Federal Reserve Bank of New York Staff Report 1090.

Angeletos, G.-M., F. Collard, and H. Dellas (2020): "Business-Cycle Anatomy." *American Economic Review*, 110, 3030–3070.

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

——— (2017): "Nonlinear panel data methods for dynamic heterogeneous agent models." *Annual Review of Economics*, 9, 471–496.

Bayer, C. and R. Luetticke (2020): "Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation." *Quantitative Economics*, 11, 1253–1288.

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.

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.

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

Durbin, J. and S. J. Koopman (2002): "A simple and efficient simulation smoother for state space time series analysis." *Biometrika*, 89, 603–615.

Geweke, J. F. and K. J. Singleton (1981): "Maximum Likelihood "Confirmatory" Factor Analysis of Economic Time Series." *International Economic Review*, 22, 37–54.

Griliches, Z. (1961): "A note on serial correlation bias in estimates of distributed lags." *Econometrica*, 29, 65–73.

——— (1967): "Distributed Lags: A Survey." *Econometrica*, 35, 16–49.

Guvenen, F., S. Ozkan, and J. Song (2014): "The nature of countercyclical income risk." *Journal of Political Economy*, 122, 621–660.

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

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

Krusell, P. and A. A. Smith (1998): "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy*, 106, 867–896.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics*, 79, 147–168.

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

Reiter, M. (2009): "Solving heterogeneous-agent models by projection and perturbation." *Journal of Economic Dynamics and Control*, 33, 649–665.

Sargan, J. D. (1964): "Wages and Prices in the United Kingdom: A Study in Econometric Methodology." in *Econometric Analysis for National Economic Planning*, ed. by P. E. Hart, G. Mills, and J. N. Whitaker, Butterworths, vol. 16, 25–54.

——— (1980): "Some Tests of Dynamic Specification for a Single Equation." *Econometrica*, 48, 879–897.

Winberry, T. (2018): "A method for solving and estimating heterogeneous agent macro models." *Quantitative Economics*, 9, 1123–1151.