Nonlinear Micro Income Processes with Macro Shocks*

Martín Almuzara† Manuel Arellano§

Richard Blundell‡ Stéphane Bonhomme¶

March 2025

Abstract

We propose a nonlinear framework to study the dynamic transmission of idiosyncratic and aggregate shocks to household income that relies on 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 macro and micro shocks propagate over time for households in different macro and micro states? (c) How do they shape the cost of business-cycle risk? We develop novel identification, estimation and inference results, and provide a complete empirical analysis combining macro time series from the US and household panels from the PSID.

Keywords: Income process, business cycle, nonlinear persistence, skewness.

*We have greatly benefited from comments and discussion with Frank Schorfheide, Mariacristina De Nardi and Jonathan Heathcote, and seminar participants at Princeton and Minnesota. All errors are our sole responsibility. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

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

‡University College London: 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 idiosyncratic and aggregate shocks to household income that relies on macro and micro data. The approach we develop allows us to examine empirically 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 households in different macro and micro states, or 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 consequences of 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, ABB); 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 household-level outcomes (Holm, Paul, and Tischbirek (2021), Andersen, Johannesen, Jørgensen, and Peydró (2023), Amberg, Jansson, Klein, and Rogantini Picco (2022), Almuzara and Sancibrián (2024)). However, a methodology for modeling the interaction between micro and macro shocks capable of integrating non-linearities in the life and business cycle dynamics of household income is still lacking.

That is our main contribution. We consider a nonlinear Markovian micro income process with a macro business cycle state of the form:

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

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 latent variables to observed micro and macro data—in our leading case, a persistent-transitory formulation for the micro states and a dynamic factor model for the macro states.

We study the nonparametric identification of this model bringing together macroe-

conometric 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 subpanel is representative of the economy in that period. Thus, identification relies on three relevant 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, which can be implemented by means of 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 or distribution regressions based on those draws. Inference is robust to some forms of cross-sectional dependence in micro shocks u_{it} arising from omitted macro shocks.

We also develop a methodology for impulse response function analysis in our nonlinear context where we want to measure the importance of macro/micro shocks and their interactions. We start by considering an experiment in which we directly perturb a state variable at some point in time; 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 alternative rules that map perturbations to a comparable system of measurement, 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 US data 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.

As a motivating piece of evidence, in Figure 1 we show, for each of three income measures (male earnings, household labor income, and household disposable income)

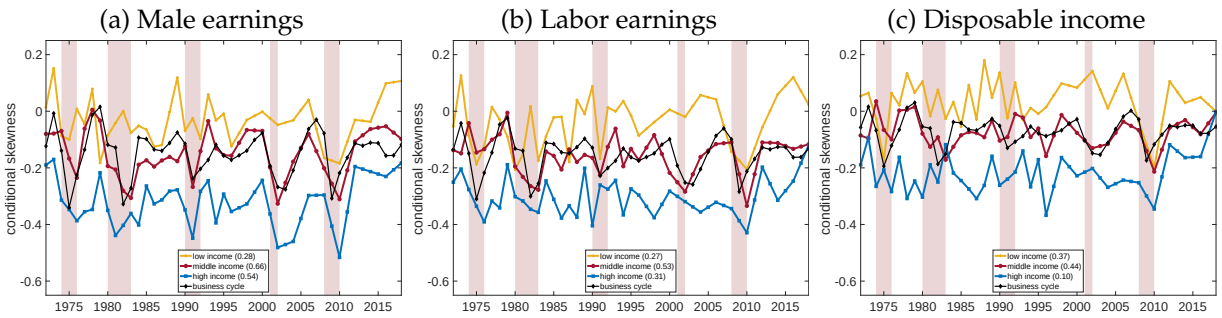


FIGURE 1. Conditional skewness and business cycles.

Note: We report the time series of conditional skewness measures from first-order quantile autoregressions of log income on a third-order Hermite polynomial of log income two years before. Low-, middle- and high-income refer to the 10-th, 50-th and 90-th percentile of the income distribution. The construction of the business cycle indicator is explained in Section 5. In the figure, it is standardized to have the same mean and variance as the middle-income skewness.

and three different cross-sectional percentiles (low, middle and high), a Bowley-Kelley measure of conditional skewness obtained from flexible quantile autoregressions. The two takeaways from this figure are, first, that left skewness is countercyclical (consistently with the findings in Guvenen et al. (2014)), and, second, that skewness markedly decreases with income at any point in time (consistently with the findings in ABB), but differentially so depending on the phase of the business cycle. Business cycle variability (measured by normalized NBER indicators) is most pronounced for male income and, to a lesser extent, for labor and disposable income, in that order—that is, from the measure with the least insurance to the one with the most, but the general results are the same for all three measures. This tale of two (conditional) skewnesses is consistent with a nonlinear transmission of micro and macro shocks. Characterizing such transmission patterns is one of the key goals of our empirical analysis.

Our empirical results show an asymmetric impact of the business cycle on income persistence. Specifically, income persistence increases for low-income households and decreases for high-income households during recessions. That is, low persistent income earners experience higher persistence of the low-income state during recessions, whereas high persistent income earners experience lower persistence of the high-income state during recessions. The ABB-like persistence measure we use provides the weight of current income in the process that produces future income when a household is hit by a micro shock for given macro conditions. Thus, it is a suitable summary of the cyclical evolution of nonlinear state dependence.

Concerning impulse response functions with respect to macro shocks (using a fixed unit shock rule), we find a large direct impact followed by a persistent indirect impact driven by the high persistence of the micro states. Moreover, both the effects on impact and the persistence are heterogeneous across households in different parts of the persistent income distribution. Differences in persistence along the persistent income distribution are also noticeable in the IRFs with respect to micro shocks. In fact, we show below that the micro impulse response (under a unit-shock rule) is intimately related to the ABB measure of persistence.

Finally, we use our estimated model to quantify the cost of macro and micro risk. For macro risk, we calculate the (per cent) compensating variation that equalizes the expected discounted utility of income with and without removing macro shocks. Expectations are calculated numerically with respect to the income process we estimate. We find a large but heterogeneous cost of business cycles that depends on the nonlinearity of the transmission of macro shocks to household income. In the presence of empirically plausible amplification effects, the cost of risk can be as high as 7 or 8 percent, 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, different from curvature in the utility function. As expected, the cost of micro risk is much higher, ranging from less than 10 percent to more than 20 percent, depending on household's age and income.

Selected literature. Our paper contributes to several strands of the literature. First, it builds on research on income dynamics, both in the presence and absence of business cycles. The literature on income process estimation is vast: Meghir and Pistaferri (2004), Blundell, Pistaferri, and Preston (2008), Gottschalk and Moffitt (2009), Browning, Ejrnæs, and Álvarez (2010), Altonji, Smith, and Vindangos (2013); see Blundell, Bollinger, Hokayem, and Ziliak (2024) for a complete review. Within that, our paper is more closely related to Storesletten, Telmer, and Yaron (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 work on the GRID project.

Second, we connect to the literature on estimating heterogeneous agent models using micro data, with contributions by Arellano and Bonhomme (2017), Liu and Plagborg-Møller (2023), Fernández-Villaverde, Hurtado, and Nuño (2023)), and Chang, Chen, and

Schorfheide (2024). Our results also connect to the macro literature on inequality and business cycles (Krusell and Smith (1998), Reiter (2009), Bhandari et al. (2021), among others) and to the vast literature that has dealt with the welfare cost of business cycles (Lucas (1987, 2003), Otrok (2001), Galí, Gertler, and López-Salido (2007), Krebs (2007), Barlevy (2005)).

Third, we draw on methodological advances in estimating models with aggregate shocks, as exemplified by Hahn, Kuersteiner, and Mazzocco (2020), and in analyzing nonlinear impulse response functions, pioneered by Gallant, Rossi, and Tauchen (1993). We also relate to an early literature on combining household survey data with time series aggregate data (Tobin (1950), Chetty (1968), Maddala (1971)). In summary, our paper integrates these strands by proposing a semi-structural framework that incorporates nonlinear microeconomic dynamics and aggregate shocks, offering a unified approach to studying the interplay between individual behavior and macroeconomic fluctuations.

Outline. The outline of the paper is as follows. In Section 2 we describe our framework of analysis, including the model and objects of empirical interest. Section 3 introduces our population as a time series of panels and studies identification. Section 4 explains the empirical specification and estimation strategy that we follow. In Section 5 we present the empirical results on nonlinear persistence and skewness over the business cycle. Section 6 develops a methodology for nonlinear impulse response functions of macro and micro shocks and provides empirical results. Section 7 describes our methodology for calculating the idiosyncratic and aggregate income risk faced by households and reports empirical results. Section 8 concludes. The appendices contain additional results and technical material.

2 Framework

In this section we describe our framework of analysis that combines time series aggregates with longitudinal information on individuals from survey data. Our emphasis is on the cyclical patterns of nonlinear income risk. But as we argue in Remark 1, our approach serves as 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.[1]

Following Arellano et al. (2017, ABB), the persistent component is a flexible first-order Markov process, the transitory component is serially independent, and both processes may depend on covariates x_{it} (e.g., the age of the household head). However, a difference with ABB is that we are interested in 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.

To be more precise, we specify the persistent component as

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

and the transitory component as

$$\varepsilon_{it} = Q_\varepsilon(Z_t, Z_{t-1}, x_{it}, v_{it}). \tag{2}$$

In (1), the function Q_η is strictly increasing in its last argument and the shock u_{it} is i.i.d. uniform on $(0,1)$, independent of all arguments in $X_{it} = (\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it})$. 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} given X_{it}. This model

[1] 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.

allows for a general nonlinear relationship between the persistent income component η_{it} and the elements of X_{it}: the dependence on the past persistent component can capture nonlinearities in income persistence; the dependence on age allows for rich life-cycle patterns; finally, the dependence on the aggregate factor Z_t and its lag allows for a flexible relationship between macro variables and micro income responses.

In addition, our modeling approach in (1) allows for nonlinear dependence on the shocks u_{it} and for interactions between shocks and the elements of X_{it}. In this model, as emphasized in ABB, different shocks to the persistent component can be associated with different degrees of persistence. Moreover, the impact of the aggregate factor Z_t can vary with the level of the idiosyncratic shock u_{it}. As shown below, this is key to flexibly capture the form of individual impulse responses to aggregate shocks.

Turning to the transitory component in (2), Q_ε is strictly increasing in its last argument, and the shock v_{it} is i.i.d. uniform on $(0, 1)$, independent of the arguments in X_{it} and of u_{it}. This specification captures general forms of non-Gaussianity in the conditional density of transitory innovations given age and aggregate conditions. In practice, ε_{it} will potentially be a mix of substantive transitory shocks and measurement error, and our approach will not allow us to distinguish between the two. For this reason, our empirical analysis will mostly focus on the properties of the persistent component.

Macro state variable. We estimate the aggregate factor Z_t from time series macro data W_t on GDP, consumption, investment, the unemployment rate, and total hours worked. As is common in the macroeconomic literature, we rely on a linear factor structure,[2]

$$W_t = \Lambda Z_t + E_t.$$

The aggregate factor Z_t and the aggregate error E_t are specified as

$$Z_t = \Phi Z_{t-1} + \Sigma_V^{1/2} V_t = Q_Z(Z_{t-1}, V_t), \tag{3}$$
$$E_t = \Sigma_E^{1/2} e_t = Q_E(e_t), \tag{4}$$

where V_t and e_t are mutually independent, i.i.d. standard normal random vectors.

Linearity and Gaussianity are not essential. Our framework can accommodate a richer

[2]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.

process with higher-order autoregressive dynamics, heteroskedasticity and nonlinearities, provided that these features be identified from the time series W_t (see Section 3). Nonetheless, and despite its parsimony, we find that (3) produces a good summary of aggregate variation in our empirical analysis, correctly indicating every recession and expansion in the sample and giving a suitable measure of their depth.

What is the interpretation of the macro state Z_t? Angeletos, Collard, and Dellas (2020) find that a single serially uncorrelated process, which they deem the *main business cycle* shock, accounts for the largest share of the unpredictable variation in each of the indicators W_t at business cycle frequencies. This main business cycle shock is not a specific structural shock (e.g., TFP or demand) but possibly a mix of structural shocks that produce consistent impulse-response patterns across cyclical variables. We think of the shock V_t in (3) as an approximation to the main business cycle shock and of Z_t as a *main business cycle state*: a process that reflects the underlying persistence of recessions and expansions, and whose stochastic nature is given precisely by the main business cycle shock.[3]

Macro-micro feedback. We can view (1), (2), (3) and (4) as a restricted nonlinear VAR(1) model for $\{\eta_{it}, \varepsilon_{it}\}_i, Z_t, E_t$, where i indexes the relevant time-t cross-sectional population—the precise notion is given 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, given Z_t, aggregate data W_t carry no extra information about the micro variables η_{it} and ε_{it}. This is similar to Liu and Plagborg-Møller (2023, Assumption 1) and implies that the aggregate factor is a sufficient statistic of the macro data to analyze the micro data.

Second, we impose that individual past income shocks do not affect Z_t. This does not necessarily rule out feedback from the micro side 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 should augment W_t with moments from the time-t earnings distribution and redefine Z_t accordingly. Instead, what we ask is that no single unit in the sample has weight in the aggregate—i.e., that units are atomistic.

To summarize, our framework allows for rich aggregate dynamics and macro-micro interactions. The limit is not conceptual but statistical: as argued in Section 4, estimation precision is determined by the time series length T, and a large Z_t may quickly exhaust

[3] One slight difference is that Angeletos et al. (2020) recover their 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 state Z_t) by first applying a band-pass filter to W_t (tailored to the same business-cycle frequencies) and then using linear factor techniques.

the degrees of freedom offered by a short time series.

Remark 1 (Reduced forms for heterogeneous agents models.). Recursive equilibria typically imply a Markovian law of motion for micro and macro states of the same form as that satisfied by η_{it} and Z_t. Hence, our framework can be interpreted as providing reduced forms for heterogeneous agents models. As illustration, suppose consumer i maximizes

$$\mathbb{E}\left[\sum_{t=0}^{\infty}\beta^t u(c_{it})\right]$$

choosing stochastic consumption sequences $\{c_{it}\}_{t=0}^{\infty}$ subject to the budget constraints

$$b(c_{it}, \eta_{it} + \varepsilon_{it}, a_{i,t+1}, a_{it}, R_t) = 0,$$

where a_{it} denotes assets and R_t is the interest rate. Under bounded rationality (à la Krusell and Smith (1998)), the optimal policy rules for consumption and assets take the form

$$\begin{aligned} c_{it} &= g_c(\eta_{it}, \varepsilon_{it}, a_{it}, \overline{\mathbf{Z}}_t), \\ a_{i,t+1} &= g_a(\eta_{it}, \varepsilon_{it}, a_{it}, \overline{\mathbf{Z}}_t), \end{aligned}$$

where $\overline{\mathbf{Z}}_t$ is a vector of state variables containing Z_t (and possibly its lag), R_t, and moments of the assets distribution. In that model, $\overline{\boldsymbol{\eta}}_{it} = (\eta_{it}, \varepsilon_{it}, c_{it}, a_{it})$ and $\overline{\mathbf{Z}}_t$ follow a multivariate version of (Q_η, Q_Z); see Arellano and Bonhomme (2017) for a related point.

2.2 Objects of empirical interest

This framework allows us to recover several quantities of interest. A first quantity will be the following measure of *nonlinear persistence* which extends the ABB persistence measure to a setup with aggregate shocks:

$$\rho\left(u, \eta, \widetilde{Z}, Z, x\right) = \frac{\partial Q_\eta(\eta, \widetilde{Z}, Z, x, u)}{\partial \eta}. \tag{5}$$

That is, persistence is measured by how a change in the micro state $\eta_{i,t-1}$ affects its value in the next period. In a linear autoregressive model, $\rho(u, \eta, \widetilde{Z}, Z, x)$ is constant and equal to the autoregressive root of the process. However, in our nonlinear process, the measure in (5) is state-dependent: persistence may vary with the past position in the income distribution

($\eta = \eta_{i,t-1}$), with aggregate conditions ($\widetilde{Z} = Z_t$ and $Z = Z_{t-1}$), and with age ($x = x_{it}$).

An important feature of the measure $\rho\left(u, \eta, \widetilde{Z}, Z, x\right)$ is that it captures how persistence varies with the micro shock $u = u_{it}$, 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, freeing the household from the trap of poverty—the unlucky reverse holding for high-η households reached by a bad u. This phenomena is absent from linear income processes. But the novel framework of this paper adds an extra layer. In (5), the entire shape of ρ as a function of u and η—and the intensity with which past income histories are wiped out by big shocks—may change with the state of the economy $\widetilde{Z}$. Our empirical analysis in Section 5 presents evidence of such macro state-dependence.

A more complete picture of the cyclical variation in income risk nonlinearities emerges by considering the following measures of *conditional dispersion* and *conditional skewness*:

$$\begin{aligned}
\text{disp}\left(\eta, \widetilde{Z}, Z, x\right) &= Q_\eta(\eta, \widetilde{Z}, Z, x, u_0) - Q_\eta(\eta, \widetilde{Z}, Z, x, 1-u_0), \\
\text{sk}\left(\eta, \widetilde{Z}, Z, x\right) &= \frac{Q_\eta(\eta, \widetilde{Z}, Z, x, u_0) + Q_\eta(\eta, \widetilde{Z}, Z, 1-u_0) - 2Q_\eta\left(\eta, \widetilde{Z}, Z, x, \frac{1}{2}\right)}{Q_\eta(\eta, \widetilde{Z}, Z, x, u_0) - Q_\eta(\eta, \widetilde{Z}, Z, x, 1-u_0)},
\end{aligned}$$

with $1/2 < u_0 < 1$. These are respectively the interquantile range and the Kelley skewness of the predictive distribution of η_{it}. In micro panels, dispersion is U-shaped while skewness is decreasing in past η. That is, low- and high-η units face more risk than those in the middle of the distribution, but the risk is tilted to the upside for the low-η and to the downside for the high-η.

How dispersion and skewness vary with the aggregate factor $\widetilde{Z} = Z_t$ is a paramount empirical question. Guvenen et al. (2014), for example, find that recessions tend to have a small impact on the variance but 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 long-range 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 its stage in the life cycle. Our approach can thus provide a unified treatment of the two types of income dispersion and skewness— across the income distribution (indexed by η) and over the business cycle (indexed by $\widetilde{Z}$)—that have been

highlighted in the literature. Again, we develop this in Section 5.

Impulse responses and risk decompositions. Beyond exploring the cyclical behavior of nonlinear income risk, our framework allows us to recover impulse response functions (IRF) with respect 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). The presence of nonlinearities and of a cross-sectional dimension makes impulse response analysis more subtle than in linear time series models, a discussion we develop in Section 6.

Finally, our framework allows us to perform a macro/micro decomposition of income risk that generalizes the notion of forecast error variance decomposition (FEVD) from the macroeconometric literature. This is 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 statistical population (or, more precisely, 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 provides conditions under which our objects of interest remain interpretable. We leave estimation and inference for 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 aggregates $\{W_t\}_{t=0}^{T+S}$ and a sequence of panels $\{\{\{y_{i,t+s-1}\}_{s=1}^{S}\}_{i\in\mathcal{I}_t}\}_{t=1}^{T}$ where $\mathcal{I}_t$ is a set of cross-sectional indexes such that each i indicates the same individual over time.

The idea is that $\mathcal{I}_t$ forms a representative sample for the time-t population of units that interacts with the aggregates W_t and the state Z_t. We observe each unit $i \in \mathcal{I}_t$ for at least S periods (more than S if $i \in \mathcal{I}_\tau$ with $\tau \neq t$ which is permitted) and we observe T collections of cross-sectional units $\{\mathcal{I}_t\}_{t=1}^{T}$ distributed over time. Let $N_t \equiv \#\mathcal{I}_t$ be the size of the time-t subpanel. In this context, our approach relies on large cross-sections, short subpanels and a long time series spanning sufficient variation in aggregate conditions coexisting with an evolving population of micro units—in summary, large N_t and T with a small, fixed S.

Remark 2 (Extensions). It is straightforward to accommodate mixed frequencies, ragged edges and other forms of missing data (e.g., observing W_t over a longer timespan than y_{it}) within our framework. We abstract from them in this section for clarity of exposition.

Remark 3 (Special cases). Our structure nests several important cases. First, if the data follow a rotating panel scheme, a share $1/S$ of individuals is refreshed every period. This is covered by our setup, since an individual may appear in several subpanels; i.e., $i \in \mathcal{I}_t$ does not preclude (nor imply) $i \in \mathcal{I}_\tau$ for $\tau \neq t$. Second, if the data are made of repeated cross-sections, this amounts to $S = 1$ with T equal to the number of cross-sections. Third, if the data form a single long panel, we have $T = 1$ and S equal to the number of time periods—as an alternative, a long panel can always be split into an overlapping sequence of multiple shorter panels.

It is important to note that even though these structures can be cast in the format of a time series of panels, not all of them will satisfy the conditions needed for identification of a model like the one in Section 2.

As a final piece of notation, let $\overline{\mathcal{I}}$ be the infinite set of all possible cross-sectional indexes from where $\mathcal{I}_t$ (for any t) is drawn and let $R_{it} = \mathbb{1}\{i \in \mathcal{I}_t\}$ be the observation indicator for all $i \in \overline{\mathcal{I}}$ and t. We are now ready to describe the statistical population.

Assumption 1 (Macro states). *There is a macro stochastic process $\{Z_t, E_t, F_t\}_{t=-\infty}^{\infty}$ such that*

(a) $\{Z_t, E_t\}_{t=-\infty}^{\infty}$ satisfies (3) and (4) where V_t and e_t are mutually independent, i.i.d. over t, and standard normal; and

(b) $\{F_t\}_{t=-\infty}^{\infty}$ is i.i.d. over t and independent of $\{Z_t, E_t\}_{t=-\infty}^{\infty}$.

Assumption 2 (Micro process with macro states). *There is a cross-section of micro stochastic processes $\{\{\eta_{it}, \varepsilon_{it}, x_{it}\}_{t=-\infty}^{\infty}\}_{i\in\overline{\mathcal{I}}}$ such that*

(a) $\{\{\eta_{it}, \varepsilon_{it}, x_{it}\}_{t=-\infty}^{\infty}\}_{i\in\overline{\mathcal{I}}}$ is i.i.d. over i given $\{Z_t, E_t, F_t\}_{t=-\infty}^{\infty}$, and

(b) $\{\eta_{it}, \varepsilon_{it}\}_{t=-\infty}^{\infty}$ given $\{Z_t, x_{it}\}_{t=-\infty}^{\infty}$ satisfies (1) and (2) where u_{it} and v_{it} are mutually independent, i.i.d. over t, and uniformly distributed on $(0, 1)$.

Assumption 3 (Atomicity). *Conditional on covariates $\{\{x_{it}\}_{t=-\infty}^{\infty}\}_{i\in\overline{\mathcal{I}}}$, persistent and transitory micro shocks u_{it} and v_{it} are independent of macro shocks $\{V_\tau, e_\tau\}_{\tau=-\infty}^{\infty}$ and of $\{F_\tau\}_{\tau\neq t}$.*

Assumption 4 (Representative samples). *For each t, there is a sampling mechanism $\{R_{it}\}_{i\in\overline{\mathcal{I}}}$ independent of $\{\{\eta_{i,t+s-1}, \varepsilon_{i,t+s-1}, x_{i,t+s-1}\}_{i\in\overline{\mathcal{I}}}\}_{s=1}^{S}$ and $\{Z_{t+s-1}, E_{t+s-1}, F_{t+s-1}\}_{s=1}^{S}$.*

Assumption 1 deals with the macro side of the model. We already discussed Z_t and E_t. The novelty is the time series F_t which captures potentially omitted aggregate variables, for example, shocks orthogonal to the business cycle or survey-specific factors. Part (b) normalizes the serial dependence of F_t. In particular, omitted persistent macro variables can be accommodated by simply taking F_t to be their surprise (or shock) component. Part (b) also implies that W_t is uninformative about F_t and that Z_t (the aggregate of interest) is unaffected by F_t. The latter is simply the generalization to our nonlinear setup of requiring shocks to be orthogonal in linear macroeconometric models.

Even though F_t is a nuisance from a substantive point of view, it is important to account for it since it is a potential source of cross-sectional dependence at the micro level. By part 2, the micro variables η_{it}, y_{it} and x_{it} are exchangeable over i given Z_t and W_t—i.e., they are identically distributed but not necessarily independent—and they only become i.i.d. when we further condition on F_t. To see what this entails, write

$$\eta_{it} = \overline{Q}_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, \{F_\tau\}_{\tau=-\infty}^{\infty}, \tilde{u}_{it}),$$

where $\overline{Q}_\eta$ is the quantile function of η_{it} given $X_{it} = (\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it})$ and $\{F_\tau\}_{\tau=-\infty}^{\infty}$, and $\tilde{u}_{it}$ is i.i.d. over t, uniformly distributed on $(0,1)$, and independent of X_{it} and $\{F_\tau\}_{\tau=-\infty}^{\infty}$. Now,

$$\begin{aligned} u_{it} &= Q_\eta^{-1}(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, \overline{Q}_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, \{F_\tau\}_{\tau=-\infty}^{\infty}, \tilde{u}_{it})) \\ &= \Upsilon(F_t, \tilde{u}_{it}), \end{aligned} \tag{6}$$

since u_{it} is independent of X_{it} by Assumption 2(b) and $\{F_\tau\}_{\tau \neq t}$ by Assumption 3. Equation (6) imposes restrictions on the shape of the impulse response function of η_{it} to F_t, and this is key to preserve the interpretation of Z_t—see Section 6 for further discussion.

A final note is that $\tilde{u}_{it}$ is also i.i.d. over i by Assumption 2(a), allowing us to interpret it as a purely idiosyncratic innovation, and that a similar derivation applies to ε_{it}.

To sum up, to guarantee the interpretability of the objects introduced in Section 2 we need unpredictability of the observed aggregate Z_t relative to the omitted one F_t (part 1), and invariance of the dynamic responses of η_{it} to Z_t with respect to hypothetical inclusion of F_t in our model (parts 2 and 3). In our Markovian environment, the latter requires the dynamic responses of η_{it} to F_t and $\tilde{u}_{it}$ to be proportional to each other in a certain sense.[4]

[4] Again see Section 6 for further discussion. This restriction is reminiscent of that imposed in panel VAR models with unobserved macro and micro shocks; see, e.g., Almuzara and Sancibrián (2024, Section 3.3).

Representativeness. Assumption 4 clarifies the role of selectivity. It asks the probability of sampling a unit into subpanel t to not depend on observable and unobservable determinants of income, but only for the duration of the subpanel in question—i.e., from periods t to $t+S-1$. This justifies the application of standard short-panel thinking to each set $\mathcal{I}_t$, which essentially treats them as random samples of short time series, while permitting general selection patterns with periodicities higher than S.

An advantage of this, compared to standard long panel approaches, is that it mitigates concerns over the representativeness of the data when 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 overrepresent more stable units with higher incomes compared to an actually representative sample in any given year.[5]

3.2 Identification argument

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

$$P_t^S(\boldsymbol{y}|\boldsymbol{x}) = P\left((y_{it},...,y_{i,t+S-1}) \leq \boldsymbol{y} \middle| (x_{it},...,x_{i,t+S-1}) = \boldsymbol{x}, \{W_{t+s-1}, Z_{t+s-1}, F_{t+s-1}\}_{s=0}^S, R_{it}=1\right).$$

That is, $P_t^S(\boldsymbol{y}|\boldsymbol{x})$ is the cumulative distribution function (CDF) of $\{y_{i,t+s-1}\}_{s=1}^S$ given $\{x_{i,t+s-1}\}_{s=1}^S$ conditioning on "time effects" and sampling: It is a random function, measurable with respect to $\{W_{t+s-1}, Z_{t+s-1}, F_{t+s-1}\}_{s=0}^S$, that encapsulates information about the cross-sectional object $\{\{y_{i,t+s-1}, x_{i,t+s-1}\}_{s=1}^S\}_{i\in\mathcal{I}}$. This distribution depends on aggregate conditions prevalent at time t throughout $t+S-1$ and these are mostly unknown to the researcher—hence, the time effects jargon. Also to note: The sampling condition $R_{it}=1$ does not affect probability calculations due to Assumption 4 and the time series of panels' focus on short S.[6]

We next outline our identification analysis which brings together techniques from the

[5] A precursor of the time series of panels is the analysis of cyclical income risk in the US by Storesletten et al. (2004). An additional advantage of this approach is that it allows us to handle covariates that would otherwise be colinear 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.

[6] In contrast, probabilistic events about the full income history of an individual participating in multiple subpanels would require accounting for selectivity.

macro- and microeconometric literatures, linked by the time series of panels.

Identification of the macro process. Knowledge of P^S implies knowledge of the distribution of the stochastic process $\{W_t\}_{t=-\infty}^{\infty}$ and, in particular, of the autocovariance function $\ell \mapsto \text{Cov}(W_t, W_{t-\ell})$. In our linear Gaussian environment, this is enough to identify Q_Z and Q_E under suitable restrictions on Λ, Φ and Σ_E.

Proposition 1. *If (a) the upper block of Λ is the identity matrix, (b) all eigenvalues of Φ lie within the unit circle, and (c) Σ_E is diagonal, then Q_Z and Q_E are identified.*

This follows from standard results for linear state-space models. In more complicated nonlinear and non-Gaussian settings, a different technique will be needed. But the point is that we can often learn about Q_Z using external time series data alone.[7]

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

Proposition 2. *Suppose $S \geq 4$ and almost surely over realizations of $\{W_t, Z_t, F_t\}_{t=-\infty}^{\infty}$, for each t,*

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

(b) with $f_{\eta_\tau|y_r,x,t}$ the density of $\eta_{i\tau}$ given y_{ir}, $\{x_{i,t+s-1}\}_{s=1}^{S}$ and $\{W_{t+s-1}, Z_{t+s-1}, F_{t+s-1}\}_{s=0}^{S}$, the families

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

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

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

$$P_{\eta,\tau}(\widetilde{\eta}|\eta, x) = P\left(\eta_{i\tau} \leq \widetilde{\eta} \middle| \eta_{i,\tau-1} = \eta, x_{i\tau} = x, Z_\tau, Z_{\tau-1}, F_\tau\right),$$
$$P_{\varepsilon,\tau}(\varepsilon|x) = P\left(\varepsilon_{i\tau} \leq \varepsilon \middle| x_{i\tau} = x, Z_\tau, Z_{\tau-1}, F_\tau\right).$$

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

[7] When Z_t is directly observed (e.g., Storesletten et al. (2004) where Z_t is the NBER recession indicator), this step can be skipped.

Under the conditions of Proposition 2, given a path for the aggregate shocks, there is an injective mapping from the joint CDF of observables P_t^S and the CDF of latent variables $\{P_{\eta,t+s-1}, P_{\varepsilon,t+s-1}\}_{1<s<S}$. This is established as in ABB by the spectral decomposition approach of Hu and Schennach (2008) and Wilhelm (2015).[8] In consequence, $P_{\eta,\tau}$ and $P_{\varepsilon,\tau}$ are random functions measurable with respect to $\{W_t, Z_t, F_t\}_{t=-\infty}^{\infty}$ (and, by Assumptions 2 and 3, with respect to $Z_\tau, Z_{\tau-1}, F_\tau$). Proposition 2 simply asserts that the researcher can determine from P^S the distribution $\overline{P}^S$ of these CDFs concurrently with the aggregate data $\{W_{t+s-1}\}_{s=0}^{S}$.

Alternative models for the micro process will lead to different requirements on S and regularity conditions.[9] But the point is that we can typically learn about latent variable distributions that are specific to a given macroeconomic regime from the microdata alone. The challenge is to link them to the macro states of interest. That is the final step.

Proposition 3. *Let Assumptions 1, 2, 3 and 4, and the conditions of Propositions 1 and 2 hold. With $f_{z_\tau|w_t^S}$ the density of $(Z_\tau, Z_{\tau-1})$ given $w_t^S = (W_t, ..., W_{t+S-1})$ suppose that the family*

$$\mathcal{F}_{z_\tau|w_t^S} \equiv \left\{ f_{z_\tau|w_t^S}(\cdot|\boldsymbol{w}) : \boldsymbol{w} \in \mathbb{R}^{\dim(W_t)S} \right\}$$

is complete for some $t < \tau < t + S - 1$. Then, Q_η and Q_ε are identified.

Proof. See Appendix A. □

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

[8] In fact, there is overidentification as $P_{\eta,\tau}$ and $P_{\varepsilon,\tau}$ must coincide across different, overlapping subpanels.

[9] For example, if conditional on macro shocks, y_{it} follows the canonical income process that assumes η_{it} is a random walk, a version of Proposition 2 can be stated that demands $S \geq 3$. If, instead, y_{it} features heterogeneity in transitory risk (as in, e.g., Almuzara (2020)), we need $S \geq 5$. In general, it will not be possible to separate permanent from transitory components when S is too small, let alone when $S = 1$. That is a major limitation of repeated cross-sections versus time series of panels.

4 Empirical specification and estimation

This section specifies a flexible class of parametric models for our framework, and develops estimation and inference theory for that class. The empirical strategy we recommend uses series approximations of the unknown functions Q_η and Q_ε (as in Arellano and Bonhomme (2016)) to create rich, highly nonlinear descriptions of income risk.

Although the identification argument of Section 3 covers nonparametric models, the precision with which our 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. The approach we outline offers a good compromise between flexibility and statistical power.

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

$$\begin{aligned} Q_\eta(\eta, \widetilde{Z}, Z, x, u) &= \sum_{j=1}^{J} \sum_{k=1}^{K} \psi_j(\eta, x)\theta_{jk}(u)\varphi_k(\widetilde{Z}, Z) \\ &= \psi(\eta, x)'\Theta(u)\varphi(\widetilde{Z}, Z). \end{aligned}$$

where ψ, φ are vectors of known basis functions (such as orthogonal polynomials) and Θ is a matrix of linear splines with nodes $(\overline{u}_1, ..., \overline{u}_L)$ that depends on the parameter vector $\theta = \text{vec}(\{\theta_{jk}(\overline{u}_\ell)\}_{1\le j\le J, 1\le k\le K, 1\le \ell\le L})$. Of course, θ determines (through Q_η) the measures of persistence, dispersion and skewness we introduced in Section 2, and the IRFs and risk decompositions we will introduce in Sections 6 and 7. So our goal is to estimate θ.

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

$$\begin{aligned} Q_{\text{init},t}(x, u) &= \psi_{\text{init}}(x)'\Delta_{\text{init},t}(u), \\ Q_{\varepsilon,t}(x, u) &= \psi_\varepsilon(x)'\Delta_{\varepsilon,t}(u), \end{aligned}$$

where $\psi_{\text{init}}, \psi_\varepsilon$ are vectors of known basis functions and, for each t, $\Delta_{\text{init},t}, \Delta_{\varepsilon,t}$ are vectors of linear splines with nodes $(\overline{u}_1, ..., \overline{u}_L)$ which depend on the parameter vectors $\delta_{\text{init},t}, \delta_{\varepsilon,t}$.

The base-period quantile function $Q_{\text{init},t}$ delivers the distribution of η_{it} at the beginning of each subpanel conditioned on time effects. We need to specify this function in order to develop a likelihood-based method to simulate latent variable posteriors, a key ingredient of our implementation. Moreover, conditioning $Q_{\text{init},t}$ on time effects allows us to absorb long-term trends and other unmodeled sources of non-stationarities we see in the data.

For analogous reasons we model $Q_{\varepsilon,t}$ as functions of time effects rather than of $(\widetilde{Z}, Z)$.

4.1 Estimation approach and algorithm

Our estimation approach builds on the pseudo-likelihood EM method of Arcidiacono and Jones (2003) and the stochastic version proposed by Arellano and Bonhomme (2016).[10]

Our model implies moment conditions involving the latent data that identify θ together with $\{\delta_{\text{init},t}, \delta_{\varepsilon,t}\}$. Latent data are of course not directly available but our model also provides a set of pseudo-likelihood functions that we can use to obtain integrated moments in terms of the observable data.

Let $\nu_\tau(\omega) = (\tau - \mathbb{1}\{\omega < 0\})$ be the check-function derivative and define

$$
\begin{aligned}
m_{\eta,it}(\theta, u) &= \Big[\psi(\eta_{i,t-1}, x_{it}) \otimes \varphi(Z_t, Z_{t-1})\Big] \cdot \nu_u\Big(\eta_{it} - \psi(\eta_{i,t-1}, x_{it})'\Theta(u)\varphi(Z_t, Z_{t-1})\Big), \\
m_{\text{init},it}(\delta_{\text{init},t}, u) &= \psi_{\text{init}}(x_{it}) \cdot \nu_u\Big(\eta_{it} - \psi_{\text{init}}(x_{it})'\Delta_{\text{init},t}(u)\Big), \\
m_{\varepsilon,it}(\delta_{\varepsilon,t}, u) &= \psi_\varepsilon(x_{it}) \cdot \nu_u\Big(y_{it} - \eta_{it} - \psi_\varepsilon(x_{it})'\Delta_{\varepsilon,t}(u)\Big).
\end{aligned}
$$

Let $m_{\eta,it}(\theta)$, $m_{\text{init},it}(\delta_{\text{init},t})$ and $m_{\varepsilon,it}(\delta_{\varepsilon,t})$ be the vertical concatenation of these functions over $u = \bar{u}_1, \ldots, \bar{u}_L$. Our model implies the complete-data moments

$$
\begin{aligned}
E\Big[m_{\eta,it}(\theta)\Big] &= 0_{JKL\times 1}, \\
E\Big[m_{\text{init},it}(\delta_{\text{init},t})\Big] &= 0_{J_{\text{init}}L\times 1}, \\
E\Big[m_{\varepsilon,it}(\delta_{\varepsilon,t})\Big] &= 0_{J_\varepsilon L\times 1}.
\end{aligned}
$$

Now, suppose $\{Z_t\}_{1\le t\le T}$ is observed but η is not. Define $\boldsymbol{\eta}_t = \{\eta_{i,t+s-1}\}_{1\le s\le S,\, i\in\mathcal{I}_t}$ and $\boldsymbol{y}_t = \{y_{i,t+s-1}\}_{1\le s\le S,\, i\in\mathcal{I}_t}$. Our first algorithm deals with estimation of the nonlinear model parameters in that setup.

Algorithm 1. *Start with $\widehat{\theta}^{(0)}$ and for $m = 1, ..., M$:*

1) Stochastic E step:

- *Draw $\{\boldsymbol{\eta}_t^{(m)}\}_{1\le t\le T}$ from $f\left(\{\boldsymbol{\eta}_t\}_{1\le t\le T}\middle|\{\boldsymbol{y}_t, Z_t\}_{1\le t\le T}, \widehat{\theta}^{(m-1)}\right)$.*

2) Pseudo M step:

- *Update $\widehat{\theta}^{(m)}$ by linear quantile regressions using $\{\boldsymbol{\eta}_t^{(m)}, Z_t\}_{1\le t\le T}$.*

[10] This method is widely used in the context of nonlinear panel data models and permanent-transitory models of income (e.g., Arellano et al. (2017) and Arellano, Blundell, Bonhomme, and Light (2023)).

Once finished, set $\widehat{\theta} = (\mu M)^{-1} \sum_{m=(1-\mu)M}^{M} \widehat{\theta}^{(m)}$ *for some* $0 < \mu < 1$.

Next, suppose we can obtain macro model parameter estimates $\widehat{\lambda}$ from the macro data. The following algorithm allows for the presence of macro and micro latent variables.

Algorithm 2. *Start with* $\widehat{\theta}^{(0)}$ *and for* $m = 1, ..., M$*:*

1) Stochastic E step:

- *Draw* $\{Z_t^{(m)}\}_{1 \le t \le T}$ *from* $f(\{Z_t\}_{1 \le t \le T} | \{W_t\}_{1 \le t \le T}, \widehat{\lambda})$.
- *Draw* $\{\boldsymbol{\eta}_t^{(m)}\}_{1 \le t \le T}$ *from* $f\left(\{\boldsymbol{\eta}_t\}_{1 \le t \le T} \middle| \{\boldsymbol{y}_t, Z_t^{(m)}\}_{1 \le t \le T}, \widehat{\theta}^{(m-1)}\right)$.

2) Pseudo M step:

- *Update* $\widehat{\theta}^{(m)}$ *by linear quantile regressions using* $\{\boldsymbol{\eta}_t^{(m)}, Z_t^{(m)}\}_{1 \le t \le T}$.

Once finished, set $\widehat{\theta} = (\mu M)^{-1} \sum_{m=(1-\mu)M}^{M} \widehat{\theta}^{(m)}$ *for some* $0 < \mu < 1$.

4.2 Sampling properties

One could think of this setup as a nonparametric sieve model in which we let $J, K, L \to \infty$; see, e.g., Newey (1997) and Chen (2007). Instead, we adopt here a parametric perspective in which the tuning parameters are fixed as the sample size grows.

This section discusses the sampling properties of the estimator $\widehat{\theta}$ and functions of $\widehat{\theta}$—e.g., the plug-in estimators of summary measures introduced in section 2.

To describe those sampling properties we rely on asymptotic approximations that take $N, T \to \infty$ with S fixed. In other words, we let the number of individuals in each sub-panel and the number of subpanels to be large while we keep the length of each sub-panel fixed. This reflects the empirical reality of a long time series of short panels, which is key to identifying the impact of aggregate state variables on individual outcomes as argued in Section 3.

Under regularity conditions, we can establish the following:

(1) Consistency: $\widehat{\theta} \xrightarrow{\text{p}} \theta$.

(2) Asymptotic normality: $\sqrt{T}\left(\widehat{\theta} - \theta\right) \xrightarrow{\text{d}} N(0_{d \times 1}, \Omega)$ for some positive definite matrix Ω.

Asymptotic analysis is challenging because of the filtering and omitted aggregates. There is also the non-smoothness of the complete data objective functions but the filtering step makes the integrated moments differentiable in our setup.

Next, consider any smooth functional γ of θ so that $\widehat{\gamma} = \gamma(\widehat{\theta})$ is the plug-in estimator of $\gamma(\theta)$. This includes most of the summary measures discussed above such as the nonlinear persistence, skewness, impulse responses and risk calculations. Again, under standard regularity conditions we obtain

(1) Consistency: $\widehat{\gamma} \xrightarrow{\text{p}} \gamma(\theta)$.

(2) Asymptotic normality: $\sqrt{T}\left(\widehat{\gamma} - \gamma(\theta)\right) \xrightarrow{\text{d}} N(0_{d\times 1}, \Gamma\Omega\Gamma')$ where Γ is the Jacobian of $\gamma(\cdot)$ evaluated at θ.

We compute standard error and confidence intervals by means of a parametric bootstrap approach. To address potential cross-sectional dependence in u_{it} (for example, due to omitted aggregate shocks) we specify and estimate a factor model for the ranks:

$$u_{it} = \Phi\left(\sqrt{1-\gamma^2}U_{it} + \gamma F_t\right),$$

where U_{it} is i.i.d. over i, t, F_t is i.i.d. over t, and both are mutually independent $N(0,1)$ conditional on $\{Z_\tau\}$. As stated, F_t captures omitted aggregate factors, such as some forms of measurement error in the survey, or other aggregate shocks conditionally mean independent of Z_t.

4.3 Implementation

In our empirical analysis, we implement the estimation algorithm and inference procedure described above using the following specification.

The function Q_η is specified using a flexible polynomial representation. The function ψ is modeled as a third-order Hermite polynomial in η combined with a second-order Hermite polynomial in age h. Meanwhile, φ is constructed as a second-order Hermite polynomial in the aggregate states (Z_t, Z_{t-1}), but with specific restrictions: the linear term is excluded from interactions involving η and h, and the quadratic term is only included in the intercept. The rank-space grid is set at $L = 11$, and the tails for $u < \overline{u}_1$ and $u > \overline{u}_L$ are modeled as exponential. As a result, Q_η contains 288 parameters, calculated as $26 \times 11 + 2$, with an approximate observation count of 200,000. Despite the large sample size, the

sampling theory suggests that the cross-sectional dimension effectively identifies δ_t, but that θ instead relies on the time series variation.

For the initial distribution Q_{init} (the base-period η) and the shock distribution Q_ε, we adopt a second-order Hermite polynomial in age, combined with a rank-space grid of $L = 11$ and exponential tails. Unlike Q_η, which depends on functions of (Z_t, Z_{t-1}), here we instead incorporate time effects directly. This specification ensures flexibility while maintaining a manageable number of parameters.

5 Nonlinearities over the business cycle

In this section we present the empirical results on nonlinear persistence and skewness over the business cycle. We start by describing our main data source, the Panel Study of Income Dynamics (PSID), and the income measurements we use. We then discuss the aggregate measurements that enter the aggregate business cycle index. These data are the key inputs into our empirical analysis of nonlinearities over the business cycle. The first set of results we present highlight the variation in nonlinear persistence. We then show the estimates of skewness. We find a strong impact of the business cycle on both measurements.

5.1 Data and descriptives

Our data is drawn from the long history of panel data on income and labor earnings in the PSID. The PSID interviewed an initial sample representative of US households in 1968. Thereafter, kept track of initial households and their offspring: e.g. if the daughter moves out to form her own household, added as a new unit. With refresher and immigrant samples in an attempt to preserve representativeness. Interviews were conducted annually between 1968 and 1997, switching to a biennial framework between 1999 and 2019.[11] We select households that consist of married couples where the head is male and aged between 25 and 60. The time series of panels is constructed as described in Section 3. Each panel has $S = 4$ and made biennial for comparability, although we use all years.

We look at three measures of male earnings, household earnings, and household disposable income. As we look at couples, the household labor earnings is for the head

[11] We do not make use the low-income sample (Survey of Economic Opportunities).

(male) and spouse. Disposable income is household labor income minus taxes paid and plus transfers received. We obtained federal taxes by applying the time-varying tax functions estimated by Borella, De Nardi, Pak, Russo, and Yang (2023).[12] Each of these incomes is net of demographic characteristics through the regression of log income on education, race, family size, state of residence, as standard in the earnings dynamic literature—see below.

We construct y_{it} by regressing the logarithm of income on the following set of demographic covariates: 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. Given the large number of regressors that result from categorical variables, we restrict the coefficients on the covariates to be constant over time. We also exclude age from this operation but we include it as an additional conditioning variable in our quantile specification.

Descriptive statistics for the income measures across the full data period from 1968 through to 2019 are presented in Figure 2. The shaded years represent recessions. Figure 2(a) shows that for most of the period under analysis there are less than 2% of the sample with zero annual household labor income (blue) even in the recessions of the early 1980s and early 1990s. This fraction went a little over 2% in the great recession. The story is different for male labor earnings. Here the fraction of zeros remains below 6% until the Great Recession and then jumps to nearly 8%.

Figures 2(b) through 2(d), display the times series of the 10th, 50th and 90th percentiles of the three incomes measures. The secular decline in real male earnings at the 10th percentile is striking and is evident even at the 50th percentile. The asymmetric nature of recessions is also clear at lower percentiles with the real decline in annual male earnings at recessions failing to make up the losses in the expansion periods. Household labor earnings tell a different story. Although the recessions are clear, there is real growth over the period, reflecting the growth of married women's labor supply and real wages.

The redistributive impact of taxation and transfers is also very clear in these figures, with the disposable incomes lying everywhere above the labor earnings lines at 10th percentile, while lying everywhere below at the 90th percentile. The insurance value of the tax and transfer system is also evident with the recessions generally less severe in disposable income, at least for low earning households.

[12] We do not explicitly account for differences in state taxes but one possibility is to use the tax functions estimated by Fleck, Heathcote, Storesletten, and Violante (2025).

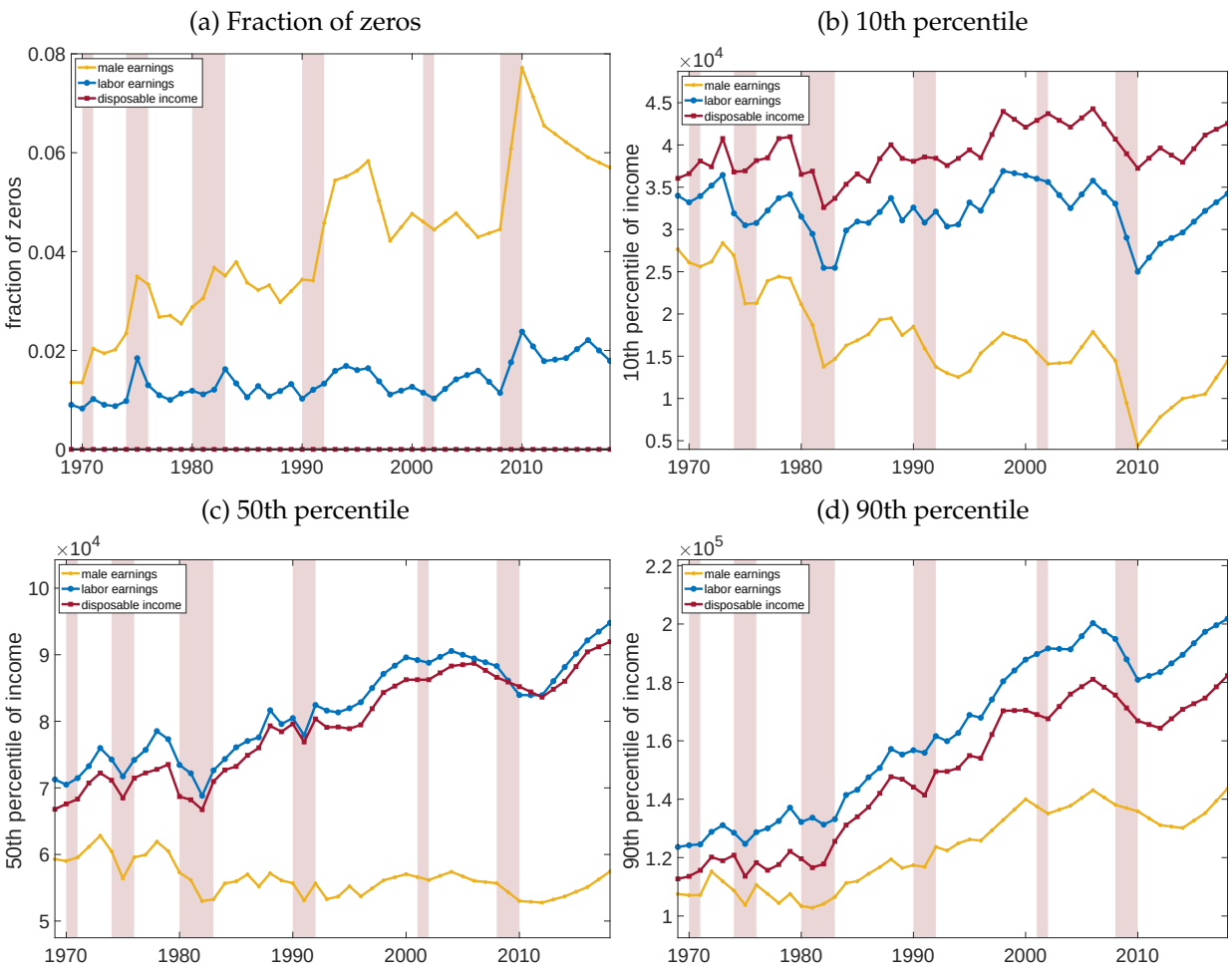


FIGURE 2. The distributional characteristics of the panel income measurements over time

Note: (a) shows the fraction of zeros. (b), (c) and (d) display the times series of the 10th, 50th and 90th percentiles of the three incomes measures respectively. The shaded areas document the periods of recession.

5.2 The business cycle indicators and the aggregate index

To construct our index of aggregate behavior Z_t we use five economic measures. The vector of aggregate observables W_t contains GDP per capita, consumption per capita, investment per capita, hours worked and the unemployment rate. The time series behavior of these variables is described in Figure 3.

We restrict the loading on the aggregate factor for GDP per capita to be equal to one, while leaving the other loadings free. We then model the aggregate factor Z_t as an AR(2) process and the idiosyncratic errors in each observable as mutually and serially orthogonal AR(2) processes. Note that the panel data measures of income are biennial while Z_t is

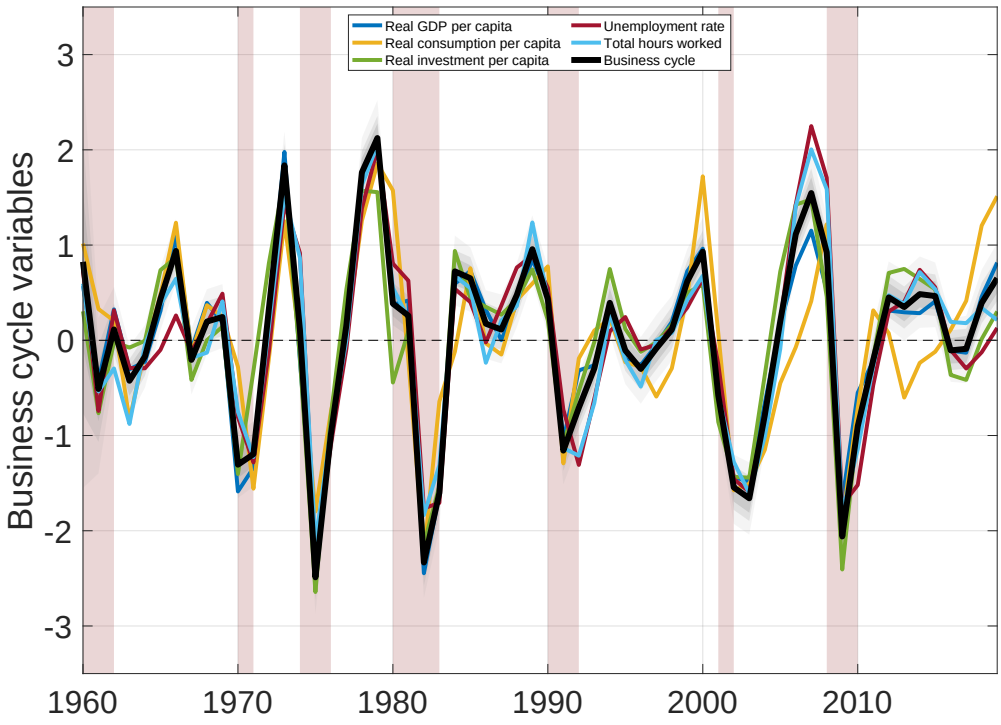


FIGURE 3. Business cycle indicators

Note: Time series of real consumption per capita; real investment per capita; unemployment rate; total hours worked; and the business cycle. Shaded areas represent recessions.

quarterly.

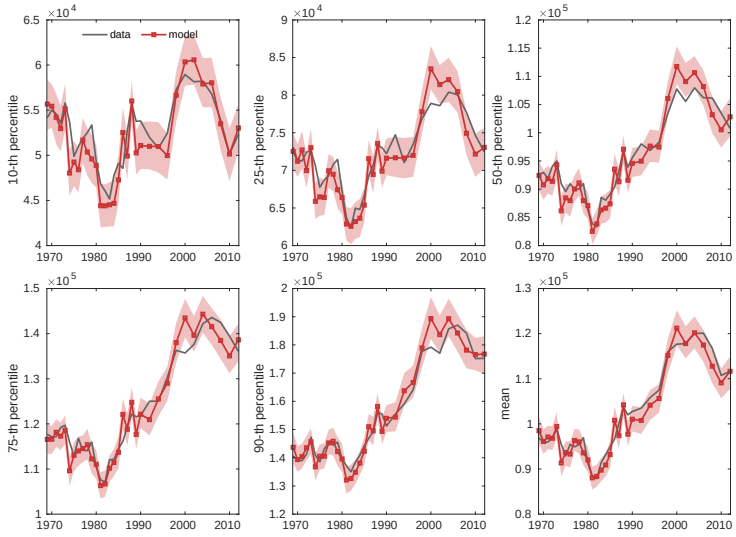
5.3 Model fit

We implement estimation and filtering by assuming normality of the aggregate factor and idiosyncratic errors. We estimate the parameters and latent variables by formulating a flat prior on the former and exploiting a Gibbs sampler algorithm.

An analysis of the model is provided in Figures 4 (a) and (b). The first set of figures in panel (a) compares the data and model output in terms of their implications for the level of income $exp(y_{it})$. We report the time series evolution of various income percentiles and the mean. Panel (b) does the same for log differences Δy_{it}. Here we report the time series of selected percentiles together with persistence, dispersion and skewness measures.

We now turn to the estimates of nonlinear income persistence and skewness, and importantly, how they change over the business cycle.

(a) Levels



(b) Log differences

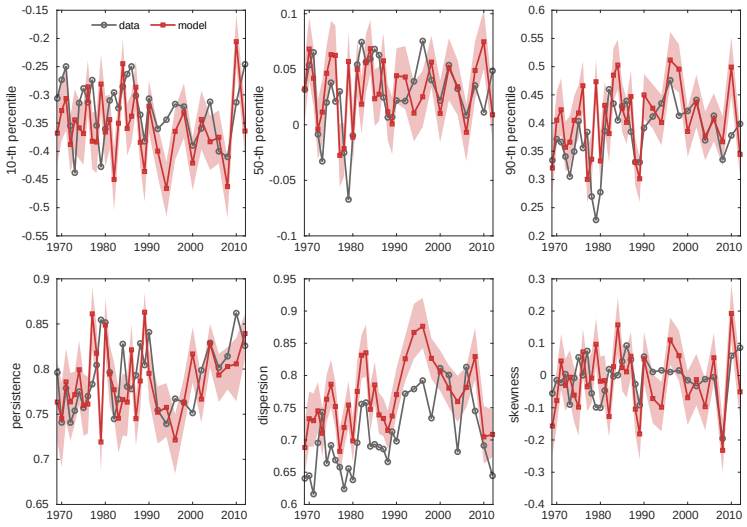


FIGURE 4. Measures of model fit (disposable income).

Note: Panel (a) compares the data and model output in terms of their implications for the level of income $\exp(y_{it})$. We report the time series evolution of various income percentiles and the mean. Panel (b) does the same for log differences Δy_{it}. Here we report the time series of selected percentiles together with persistence, dispersion and skewness measures.

5.4 Persistence: the asymmetric role of recessions

As discussed in Section 2, the measure of nonlinear persistence is given by

$$\rho(u, \eta, \widetilde{Z}, Z) = \frac{\partial Q_\eta(\eta, \widetilde{Z}, Z, u)}{\partial \eta}.$$

Figures 5, 6 and 7 present our estimates for the three measures of income: male earnings, household labor earnings, and household disposable income respectively. In each panel, Figure (a) presents the measure of persistence by quantile of the current shock to persistent income u_{it}, and last period's persistent income component η_{it-1}. The three surfaces distinguish between recession, steady state, and expansion. In Figure (b) the difference between recession and expansion is plotted with the corresponding 95% interval surfaces.

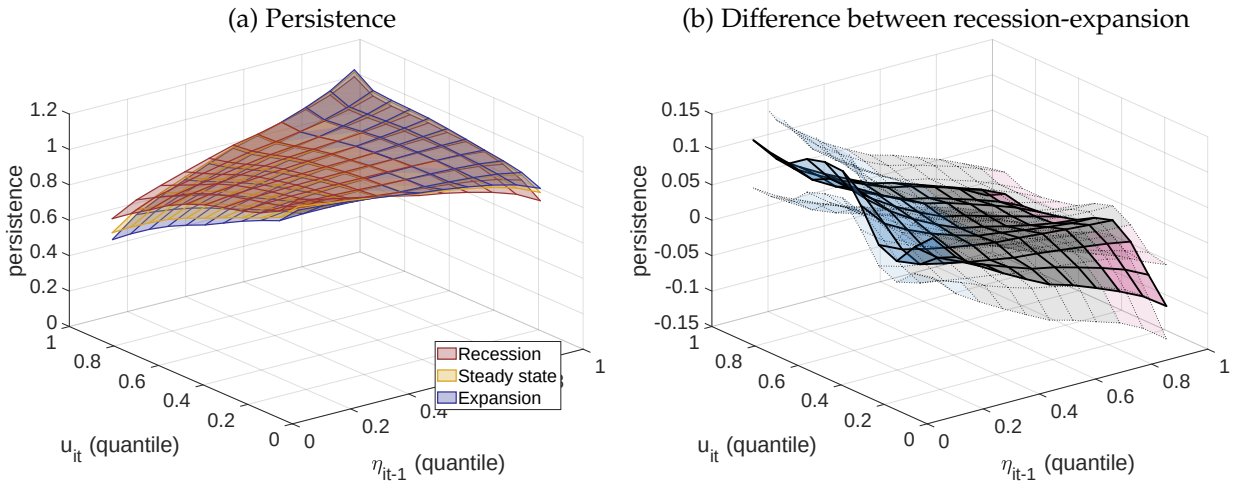


FIGURE 5. Nonlinear persistence in male earnings.

Note: (a) presents graphs of the measure of persistence by quantile of the current shock to persistent income u_{it}, and last period's persistent income component η_{it-1}. The three surfaces distinguish between recession, steady state, and expansion. (b) shows the difference between recession and expansion, with the corresponding 95% interval surfaces.

These Figures confirm the nonlinear persistence uncovered in Arellano et al. (2017). They do this for the longer time series from the complete PSID income data, and they do this across three measures of income.

In terms of the aim of this paper, we find recessions tilt the estimated persistence surface. Income persistence increases for low-η and decreases for high-η during recessions. This is particularly striking for male earnings, but remains clear and significant

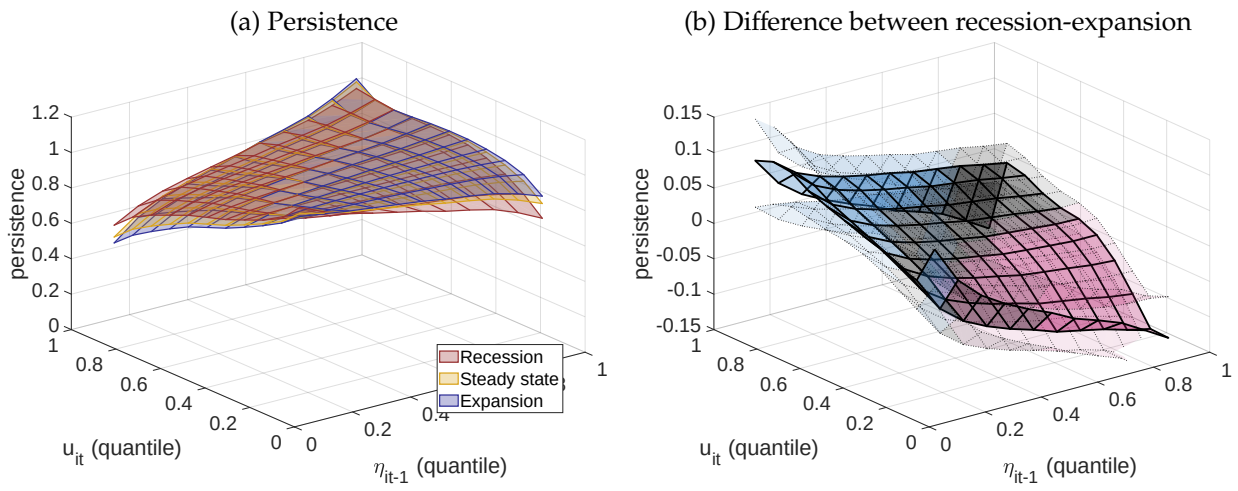


FIGURE 6. Nonlinear persistence in labor earnings.

Note: See Figure 5.

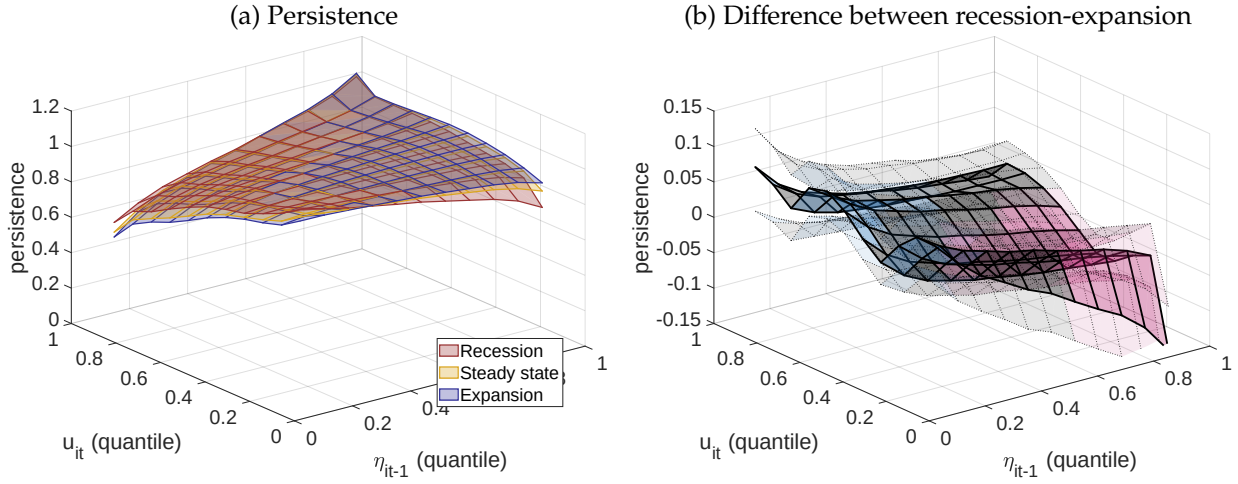


FIGURE 7. Nonlinear persistence in disposable income.

Note: See Figure 5.

for household labor earnings. Taxes and transfers attenuate the impact of the business cycle on income persistence. There is a simple interpretation. For earned income, those on low earnings experience more persistence of the low earning state during recessions. For high earners the story is reversed, persistence declines in recessions. In other words, for high earners, their higher persistent earnings is less persistent in recessions. They have an increased chance of moving to a lower percentile of persistent income during recessions. In contrast, low earners have a reduced chance of escaping low persistent income in recessions.

5.5 Dispersion

As discussed in Section 2, the measure of dispersion for the persistent income component we use is given by

$$\text{disp}(\eta, \widetilde{Z}, Z) = Q_\eta(\eta, \widetilde{Z}, Z, u_0) - Q_\eta(\eta, \widetilde{Z}, Z, 1-u_0).$$

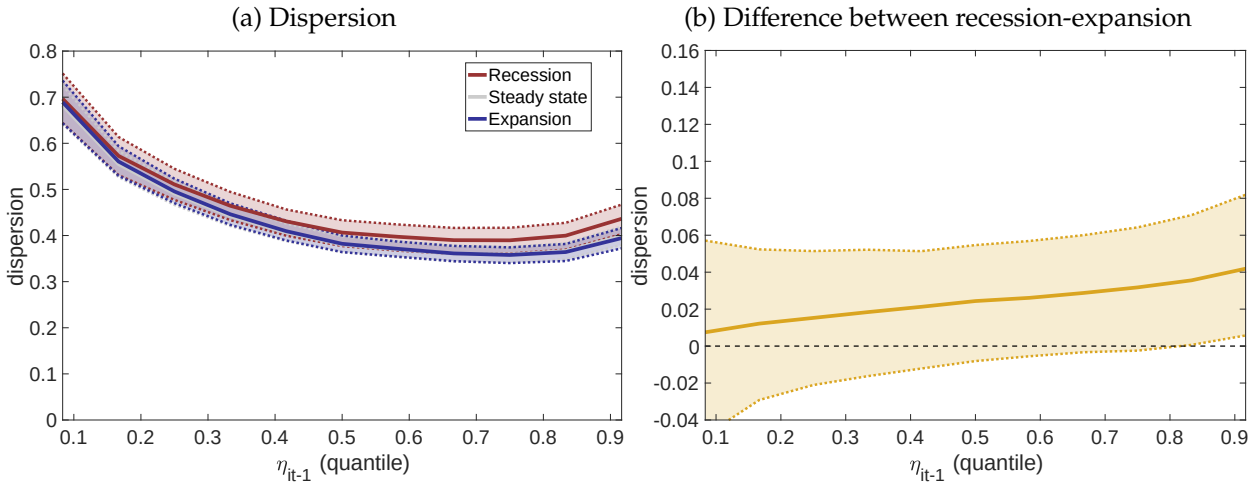


FIGURE 8. Conditional dispersion in male earnings.

Note: Panel (a) plots the estimates of the dispersion measure and (b) the estimated difference between recessions and expansion. The shaded areas show the 95% confidence intervals.

Figures 8, 9 and 10 show the estimates of this dispersion measure and the estimated difference between recessions and expansion for the three income measures across quantiles of the persistence income component $\eta_{i,t-1}$.

5.6 Skewness: a tale of two skewnesses

As discussed in Section 2, the measure of skewness for the persistent income component we use is given by

$$\text{sk}(\eta, \widetilde{Z}, Z) = \frac{Q_\eta(\eta, \widetilde{Z}, Z, u_0) + Q_\eta(\eta, \widetilde{Z}, Z, 1-u_0) - 2Q_\eta\left(\eta, \widetilde{Z}, Z, \frac{1}{2}\right)}{Q_\eta(\eta, \widetilde{Z}, Z, u_0) - Q_\eta(\eta, \widetilde{Z}, Z, 1-u_0)}.$$

Figures 11, 12 and 13 show the estimates of this skewness measure and the estimated difference between recessions and expansion for the three income measures across quan-

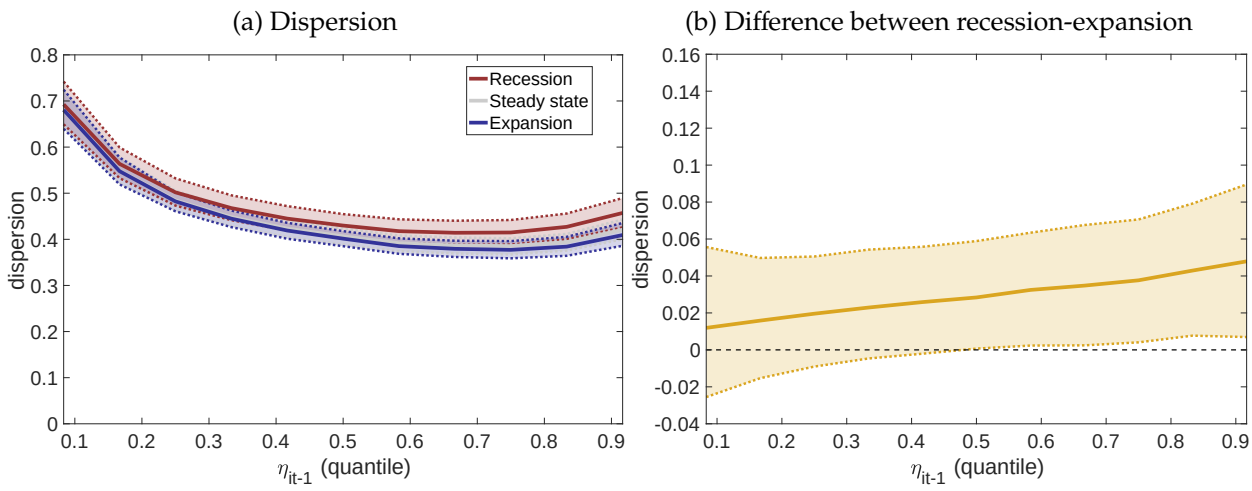


FIGURE 9. Conditional dispersion in labor earnings.

Note: See Figure 8.

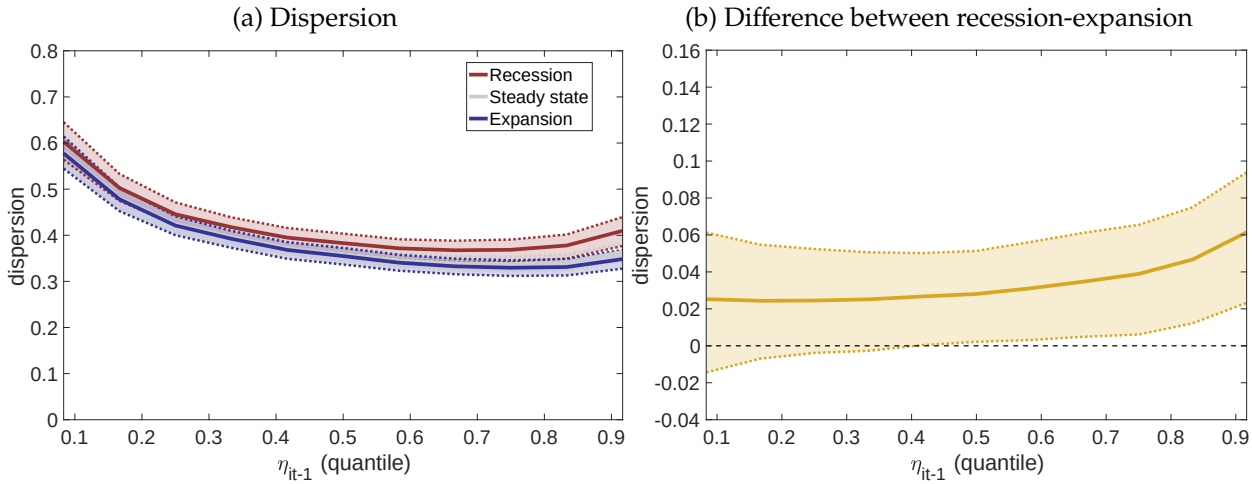


FIGURE 10. Conditional dispersion in disposable income.

Note: See Figure 8.

tiles of the persistence income component $\eta_{i,t-1}$.

The estimates show a decreasing skewness as well as a shift in the skewness in recessions. Recessions are found to result in negative conditional skewness of η at most income levels, particularly at higher quantiles. Negative skewness in recessions is common across all three incomes measures. Mirroring the results for nonlinear persistence, the shift in skewness in recessions is strongest for male earnings and weakest for household disposable income, that is, after other household earnings and the tax-transfer system is taken into account.

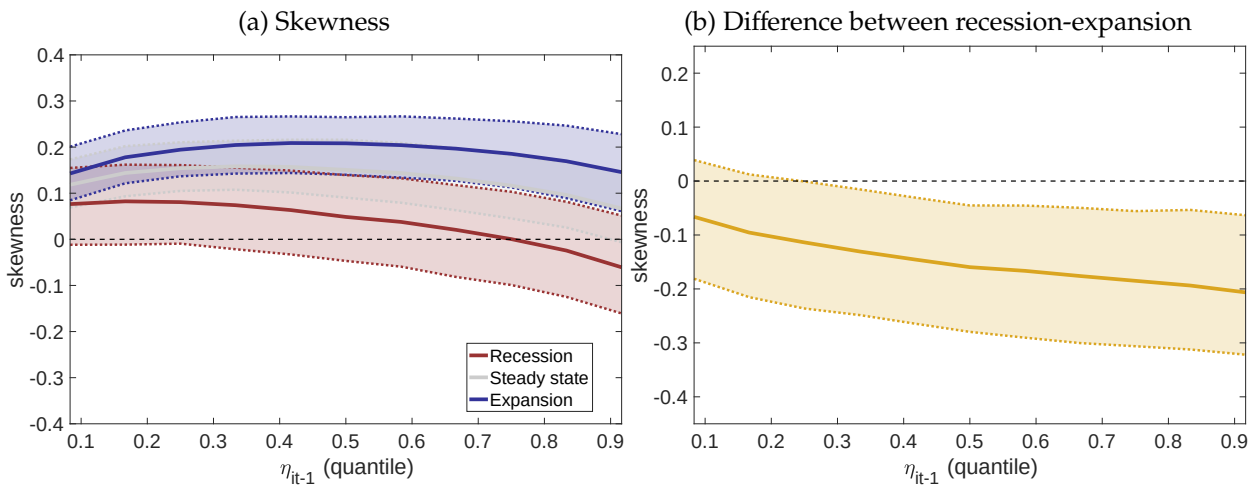


FIGURE 11. Conditional skewness in male earnings.

Note: Panel (a) plots the estimates of the skewness measure and (b) the estimated difference between recessions and expansion. The shaded areas show the 95% confidence intervals.

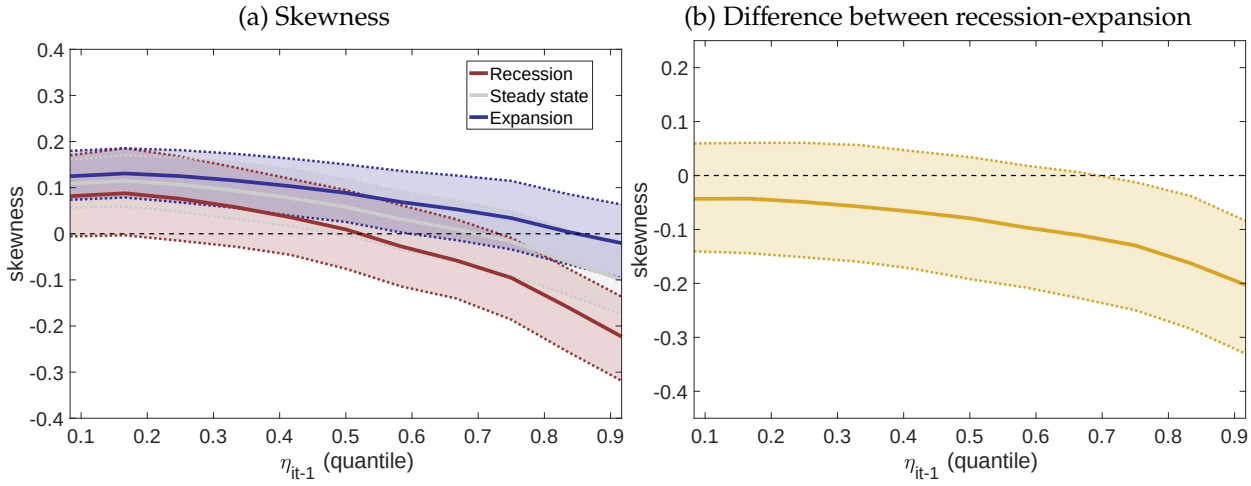


FIGURE 12. Conditional skewness in labor earnings.

Note: See Figure 11.

6 Impulse response analysis

This section develops methodology for quantifying the dynamic propagation of macro and micro shocks in our nonlinear setup. We introduce a novel approach to nonlinear impulse response functions (IRFs) in Section 6.1 and we provide empirical calculations in Section 6.2.

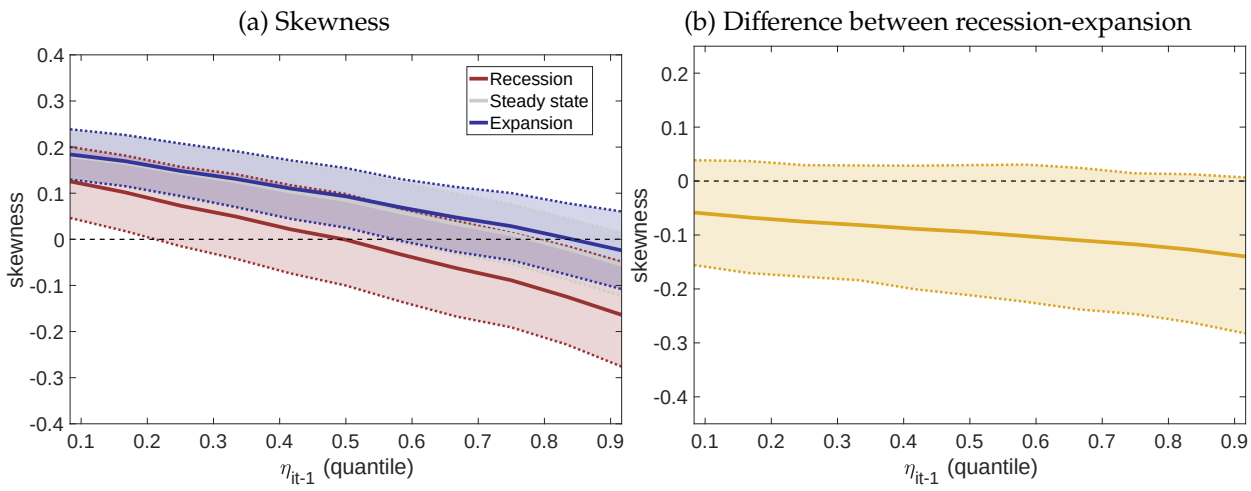


FIGURE 13. Conditional skewness in disposable income.

Note: See Figure 11.

6.1 The dynamic propagation of macro and micro shocks

Three features of our framework make the analysis of impulse responses subtle: its nonlinear nature, the macro-micro nature of income risk, and the presence of a cross-sectional dimension.

In linear models, IRFs are derivatives with respect to macro and micro shocks, in our case u_{it} and V_t. But u_{it} and V_t are affected by normalizations that were adopted for convenience, not for economic reasons: u_{it} and V_t are not in comparable units; what is more, a change of a given magnitude in u_{it} need not be comparable across units. To address these issues, we build on Gallant et al. (1993) and define impulse responses by means of perturbations of the state variables around a benchmark value. We explain the methodology in the context of our nonlinear income process with macro shocks.

Defining IRFs. Consider the state transition equations from our model,

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

From the perspective of unit i, there is a micro state variable $\eta_{i,t-1}$ and two macro state variables Z_t and Z_{t-1}. The idea we pursue consists of fixing them at a benchmark initial value $(\eta_{i0}, Z_1, Z_0) = (\eta_0^b, Z_1^b, Z_0^b)$. We then perturb one of the states by a potentially individual-

specific amount κ_i (in the units of the state variable) and we track the evolution of η_{it} in both the benchmark and perturbed scenarios. We define impulse responses as suitably scaled differences in the expected trajectory of η_{it} for marginal perturbations κ_i.

This raises the important question of how to select κ_i. To this end, we introduce a rule that maps κ_i to a system of comparable units δ. The rule is given by a function g_i such that if ξ^b is the benchmark value of the state,

$$g_i(\xi^b) = g(\xi^b + \kappa_i(\delta)) - \delta.$$

Examples of such rules are the following:

- The unit-perturbation rule sets $g_i(\cdot)$ to the identity function,

  $$\eta_{i0}^b = (\eta_{i0}^b + \kappa_i) - \delta,$$

  or simply $\kappa_i = \delta$.

- The rank-perturbation rule sets $g_i(\cdot)$ to the cumulative distribution function (CDF) of η_{i0} conditioned on $(Z_1, Z_0) = (Z_1^b, Z_0^b)$, denoted $F_{\eta_0}(\cdot|Z_1^b, Z_0^b)$,

  $$F_{\eta_0}(\eta_{i0}^b|Z_1^b, Z_0^b) = F_{eta_0}(\eta_{i0}^b + \kappa_i|Z_1^b, Z_0^b) - \delta,$$

  or $\kappa_i = Q_{\eta_0}(F_{\eta_0}(\eta_{i0}^b|Z_1^b, Z_0^b) + \delta|Z_1^b, Z_0^b) - \eta_{i0}^b$ where $Q_{\eta_0}(\cdot|Z_1^b, Z_0^b)$ denotes the conditional quantile function of η_{i0} given $(Z_1, Z_0) = (Z_1^b, Z_0^b)$.

Of course, the context will dictate what rule is appropriate. For example, the rank-perturbation rule may be attractive if η represents a concept with no natural unit of measurement, such as welfare. The unit-perturbation rule, on the other hand, will be attractive when η is measured in money (or real) terms or when it is the logarithm of a quantity (in which case the impulse consists of the same approximate $100 \cdot \delta$ percent change applied to all individuals). If η represents income, κ_i can be thought of as the result of a hypothetical transfer program implemented by a central planner with a particular social utility function, g being the rule that maximizes social utility.

Macro IRFs. We define the macro impulse response of η as follows. For a given benchmark value $Z_1 = Z^b$ and perturbation $g(Z^b) = g(Z^b + \kappa(\delta)) - \delta$:

$$\overline{\text{IRF}}_{it}(Z^b, \eta_{i0}, Z_0) = \lim_{\delta \to 0} \frac{E\left[\eta_{it} \middle| Z_{i1} = Z^b + \kappa(\delta), \eta_{i0}, Z_0\right] - E\left[\eta_{it} \middle| Z_{i1} = Z^b, \eta_{i0}, Z_0\right]}{\delta}.$$

Note that the expectations can be computed by simulation because

$$f\left(\eta_{i1}, ..., \eta_{iH}, Z_2, ..., Z_H \middle| Z_1, \eta_{i0}, Z_0\right) = \prod_{t=1}^{H} f(\eta_{it} | \eta_{i,t-1}, Z_t, Z_{t-1}, \theta) \prod_{t=2}^{H} f(Z_t | Z_{t-1}, \lambda)$$

and we know (or have estimated) θ and λ.

This leads to an implied macro shock:

$$\widetilde{V}_1 = g\left(Q_Z(Z_1 | Z_0)\right) - g(Z^b),$$

which reduces to $\widetilde{V}_1 = V_1$ when g is the unit perturbation rule.

Micro IRFs. We also define the micro impulse response of η in a similar way. For a benchmark $\eta_{i1} = \eta^b$ and perturbation $g_i(\eta^b) = g_i(\eta^b + \kappa_i(\delta)) - \delta$:

$$\text{IRF}_{it}(\eta^b, Z_1, Z_0) = \lim_{\delta \to 0} \frac{E\left[\eta_{it} \middle| \eta_{i1} = \eta^b + \kappa_i(\delta), Z_1, Z_0\right] - E\left[\eta_{it} \middle| \eta_{i1} = \eta^b, Z_1, Z_0\right]}{\delta}.$$

Again, expectations can be easily computed by simulation.

This leads to an implied micro shock:

$$\widetilde{u}_{i1} = g_i\left(Q_\eta(\eta_{i1} | \eta_{i0}, Z_1, Z_0)\right) - g_i(\eta^b),$$

which reduces to $\widetilde{u}_{i1} = u_{i1}$ when g is the rank perturbation rule.

We note that with this definition of shock, the IRF equals the derivative of the expectation with respect to the implied innovation. But the definition of shock is only valid locally in the sense that it depends on the benchmark values. If we were to compare expectations for values of δ which are not infinitesimal, one could also define innovations which will depend on the size of the perturbation δ.

The derivation of implied shocks is useful for interpretation. In our setup, the innovations $\widetilde{u}_{i1}, \widetilde{V}_1$ generalize recursive identification (i.e., Cholesky orderings) in linear systems.

In our case, the recursive order places the macro state first (with conditional independence generalizing orthogonality).

Link between IRFs and nonlinear persistence. One can show that

$$\text{IRF}_{it}(\eta^b, Z_1, Z_0) = E\left[\prod_{s=2}^{t} \rho\left(u_{is}, \eta_{i,s-1}, Z_t, Z_{t-1}\right) \middle| \eta_{i1} = \eta^b, Z_1, Z_0\right] \times \left(g_i'(\eta^b)\right)^{-1}$$

The first part does not depend on the rule g, while the second part does not depend on the horizon t. Hence the first part determines the dynamic propagation of uncertainty and is fully described by composing the nonlinear persistence measure we introduced before. For instance, it collapses to ρ^t in the non-state dependent linear homogeneous AR process.

The second part determines the scale of the IRF: it's 1 for the unit-perturbation and the inverse of the density at η^b for the rank-perturbation. The unit-perturbation always produces a unit response on impact by construction, so using a different rule shifts the IRFs curves permanently by a scaling factor given by the inverse of the derivative of the rule. Put differently, all IRFs are scaled versions of the unit-perturbation rule IRF, which is a reflection of the ABB measure of persistence.

6.2 Empirical results

Figures 14, 15 and 16 report both macro and micro impulse responses for male earnings, labor earnings and disposable income, respectively. We present both negative and positive shocks and we multiply the IRF by -1 in the case of a negative shock.

The unit-normalization rule is used in all cases. This means that macro impulse responses correspond to changes in income per unit of change in the aggregate factor (which is in the same scale as GDP growth per capita). On the other hand, micro impulse responses correspond to changes in income per unit of change in income itself, thereby ensuring a unit response on impact.

The impulse response of the business cycle indicator is measured on the right axis: it corresponds to a change of one unit (or minus one when the shock is negative) for the macro IRFs, while it is unchanged for micro IRFs in line with our no-feedback assumption.

The impact of aggregate shocks (panels (a) and (b) in Figures 14, 15 and 16) shares some common features across income measures. First, the contemporaneous impact of the shock implies roughly a 2% expected drop or jump in income depending on whether the

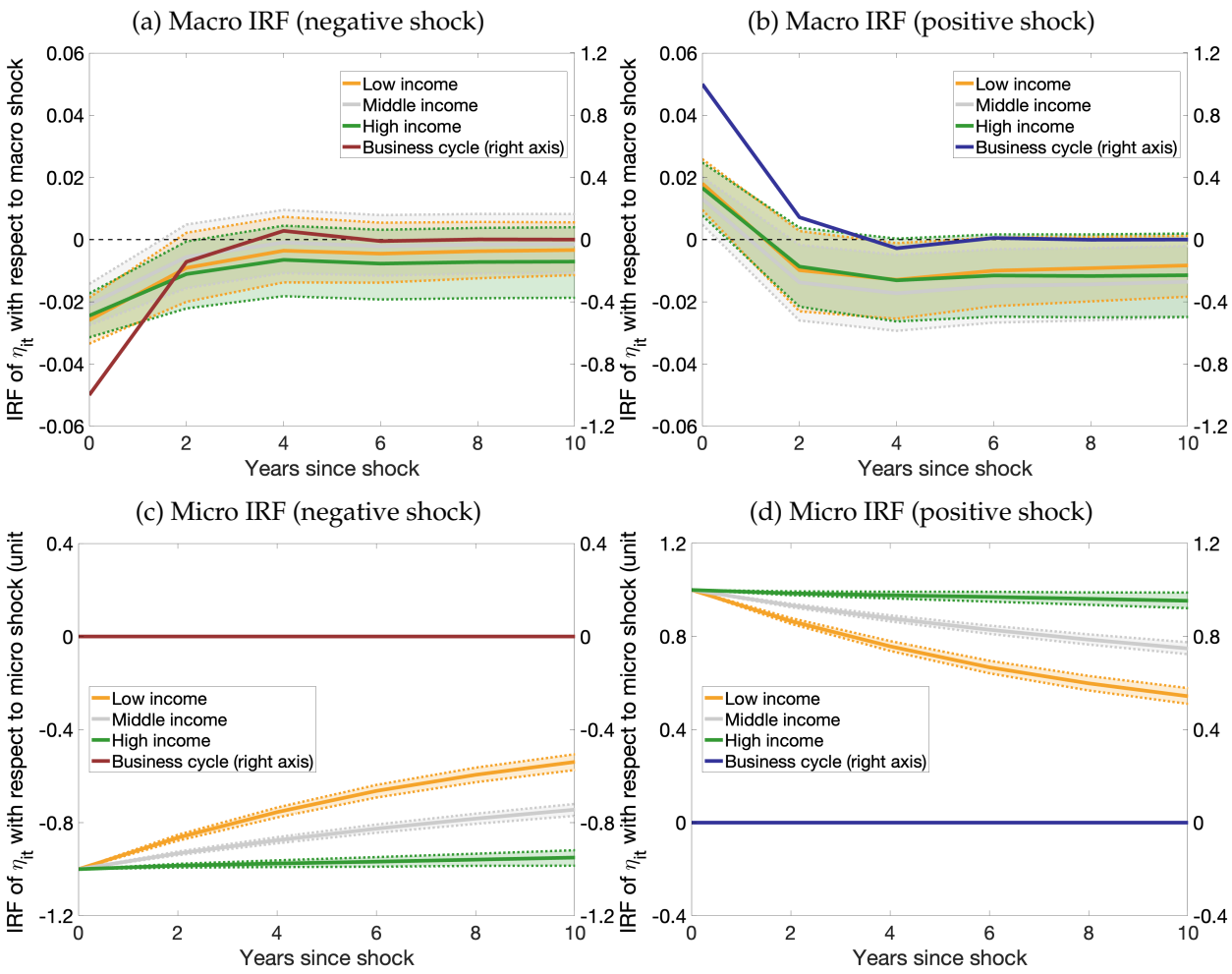


FIGURE 14. Impulse responses for male earnings.

Note: Panels (a) and (b) report macro IRFs to a negative and a positive shock. Panels (c) and (d) instead report micro IRFs to a negative and a positive shock, respectively. The benchmark values are set to $Z_1^b = Z_0^b = 0$ while we set η^b to the 10th (low, orange line), 50th (middle, gray line) and 90th (high, green line) percentile of the unconditional distribution of η_{i0}.

shock is negative or positive. Second, the effect dies out over time, with more persistence at higher levels of income. Third, the response to a positive shock indicates some level of overshooting, with the level of income lower than in the absence of the shock for many years before reverting back to zero (not seen in the figure).

Micro shocks also show some robust patterns across income measures (panels (c) and (d) in Figures 14, 15 and 16). In particular, we find more persistence at higher levels of income. Given the connection between impulse responses and nonlinear persistence, this is not surprising: the average persistence is lower for low η's than for higher η's.

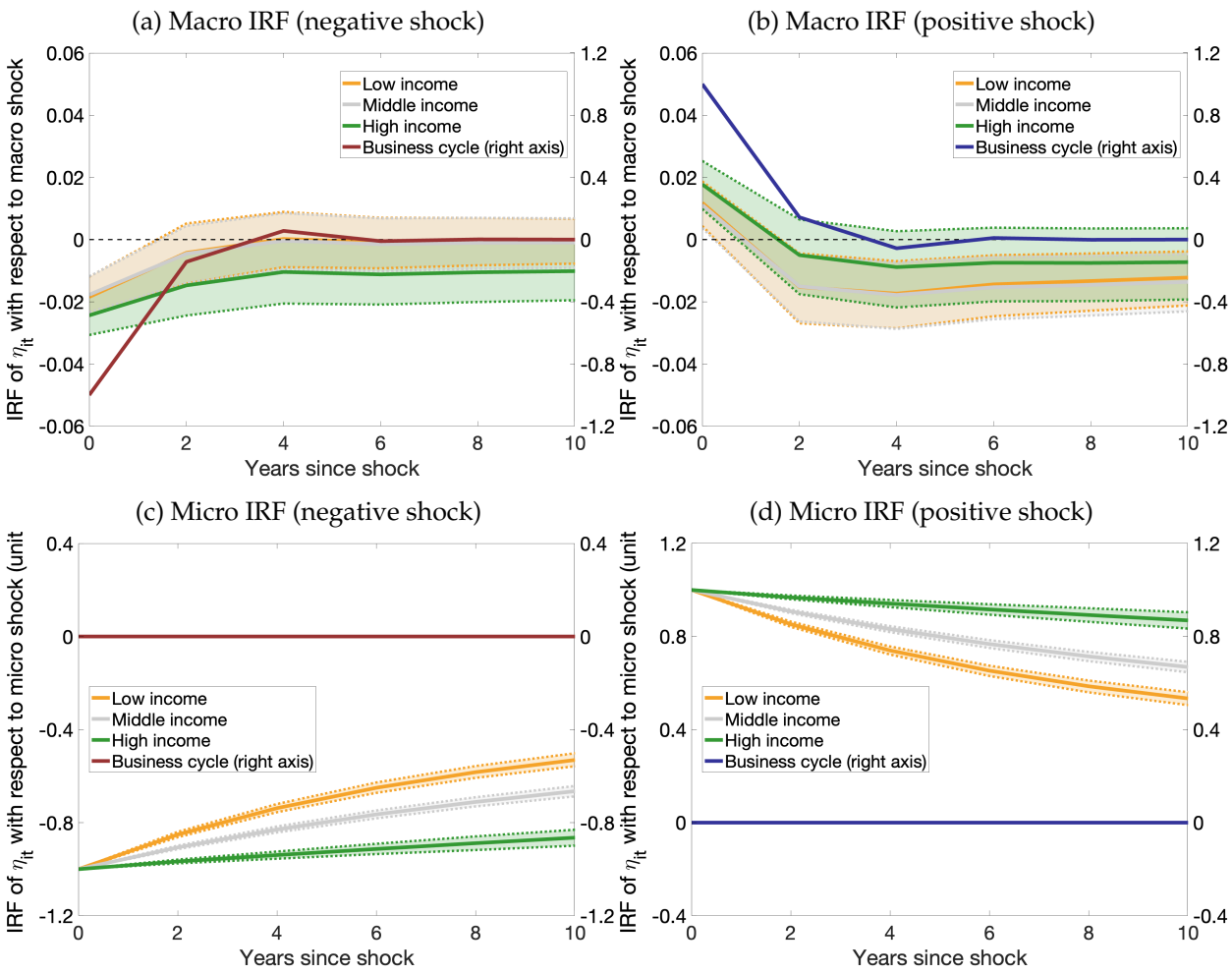


FIGURE 15. Impulse responses for labor earnings.

Note: See notes to Figure 14.

7 Risk quantification

In this section we describe our methodology to quantify the idiosyncratic and aggregate income risk faced by agents, and we report empirical estimates.

7.1 How income nonlinearities affect income risk

Let $U(\exp(\eta_{it}))$ denote the utility function of a household whose persistent income component is η_{it}. We focus on indirect utility, since we do not have consistent data on consumption, and abstract away from the presence of the transitory component ε_{it} since the latter may contain measurement error.

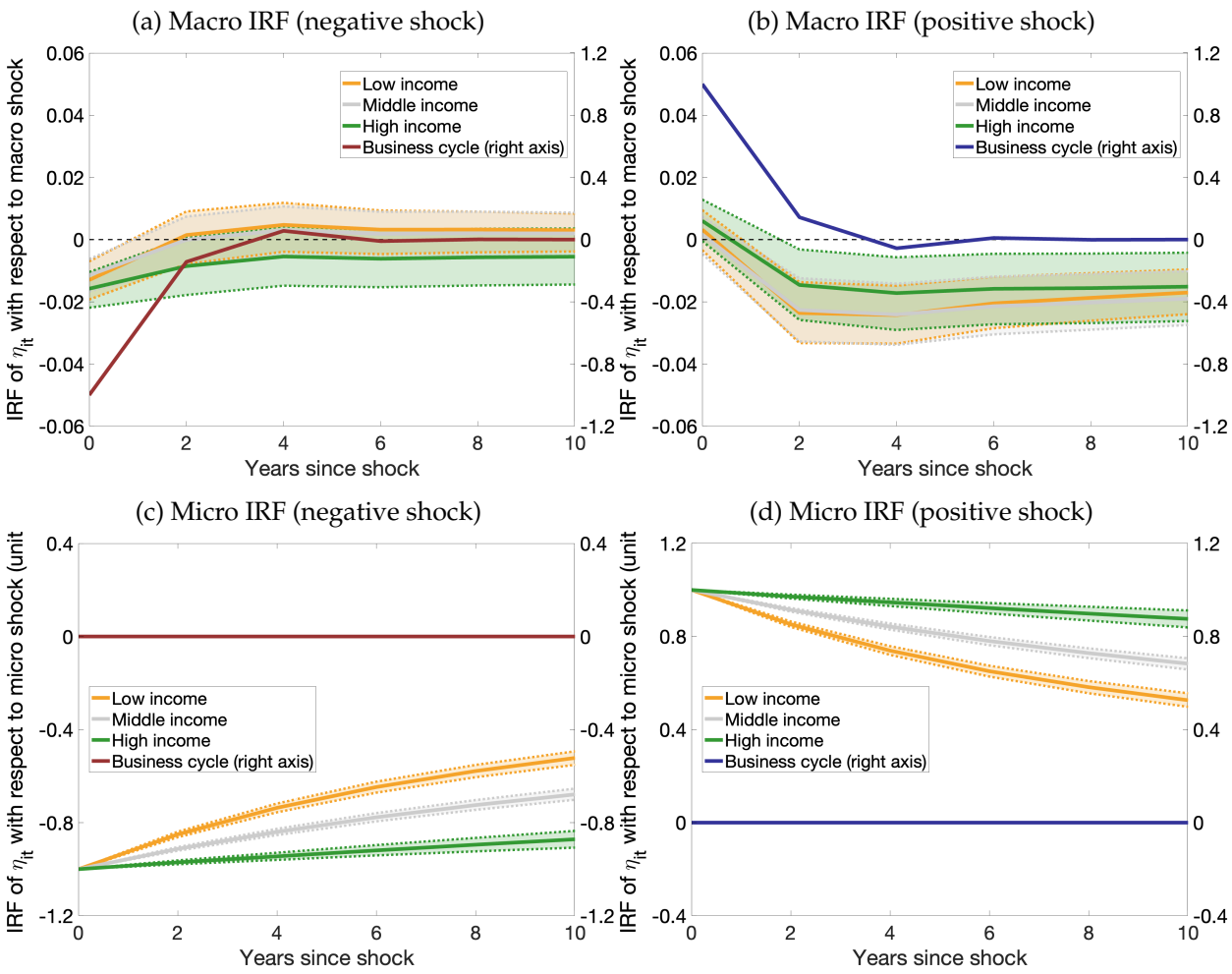


FIGURE 16. Impulse responses for disposable income.

Note: See notes to Figure 14.

Following a standard approach in macroeconomics after Lucas (1987, 2003), we compute the compensating variation find CV such that

$$\mathbb{E}^{\star}\left[\sum_{t=1}^{H}\beta^{t}U\Big((1-\text{CV})\exp(\eta_{it})\Big)\,\middle|\,\eta_{i0},Z_{0}\right]=\mathbb{E}\left[\sum_{t=1}^{H}\beta^{t}U\Big(\exp(\eta_{it})\Big)\,\middle|\,\eta_{i0},Z_{0}\right], \tag{7}$$

where, while $\mathbb{E}$ denotes the agent's expectation over the actual process of income, $\mathbb{E}^{\star}$ denotes the expectation under a counterfactual income process where the aggregate or idiosyncratic sources of variation have been removed. As is common in such calculations we assume that households have rational expectations.

The quantity CV in (7) measures the proportion of current (persistent) income the

household would be willing to forego in order to eliminate a source of income risk, either idiosyncratic or aggregate. Part of the literature concerned with income risk quantification focuses on curvature in preferences. A typical 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). The presence of nonlinearities in this relationship can generate higher amounts of risk than the literature often finds.

To make this point formally, for given η_{i0}, Z_0, write $\eta_{it} = \eta_t(\mathbf{u}_i, \mathbf{V})$ with $\mathbf{u}_i, \mathbf{V}$ the histories of idiosyncratic shocks u_{it} and aggregate shocks V_t, respectively. Consider adding a small amount of aggregate uncertainty to the income process through a small-variance V_t. Defining $\widetilde{U}(x) = U(\exp(x))$, the compensating variation associated with this change can be approximated as

$$\text{CV} \approx -\frac{\sum_{t=1}^{H} \beta^t \sum_{\ell=0}^{t-1} \left[\widetilde{U}''(\eta_t(\mathbf{0},\mathbf{0})) \left(\frac{\partial \eta_t(\mathbf{0},\mathbf{0})}{\partial V_{t-\ell}} \right)^2 + \widetilde{U}'(\eta_t(\mathbf{0},\mathbf{0})) \left(\frac{\partial^2 \eta_t(\mathbf{0},\mathbf{0})}{\partial V_{t-\ell}^2} \right) \right]}{\sum_{t=1}^{H} \beta^t \widetilde{U}'(\eta_t(\mathbf{0},\mathbf{0}))}. \tag{8}$$

A similar approximation holds in the case of idiosyncratic risk, although here we focus on the case of aggregate risk for the presentation.

For example, in the case of log-utility, we obtain

$$\text{CV} \approx -\frac{\sum_{t=1}^{H} \beta^t \sum_{\ell=0}^{t-1} \left(\frac{\partial^2 \eta_t(\mathbf{0},\mathbf{0})}{\partial V_{t-\ell}^2} \right)}{\sum_{t=1}^{H} \beta^t}. \tag{9}$$

Equation (9) illustrates that the willingness to pay to avoid aggregate income risk depends crucially on how nonlinear the income process is in the aggregate shock V_t. For a linear income process, and log utility, the compensating variation is approximately zero. However, it can be large when income depends nonlinearly on V_t. Empirically, nonlinearities may reflect the fact that different recessions, say, impact household income differently or the presence of amplification effects. More generally, when the utility function is not logarithmic, (8) shows that the compensating variation reflects the combination of two effects: nonlinearities between income and aggregate shocks, and the curvature of the utility function.

7.2 Empirical quantification

To quantify income risk empirically, we assume the agent faces three layers of uncertainty. The first one is economic uncertainty, which stems from the idiosyncratic and aggregate shocks u_{it} and V_t. This first kind of uncertainty is the one the macroeconomic literature often focuses on, and it has been our focus in this section so far. The second layer of uncertainty comes from model uncertainty coming from not knowing some aspect of the processes Q_η and Q_Z which govern the evolution of income. The third layer is sampling uncertainty coming from the fact that we, i.e., the econometricans, only observe a finite sample of macro and micro data, $\{Z_t\}$ and $\{y_{it}\}$.

We assume that agents are well informed about the idiosyncratic features of Q_η and Q_Z. However, we allow them to know less about aggregate features than the econometrician. In that context, we are interested in obtaining point estimates and confidence intervals for measures of risk that embody the first two layers of uncertainty, while at the same time suitably reflect the third layer, i.e., sampling uncertainty. We provide more details on risk quantification in Appendix B. In this version of the paper, we will compare several possible income models in order to shed light on the possible impact of model uncertainty (from the agent's perspective) on risk calculations.

In Figure 17 we report estimates of the compensating variation in the case of aggregate risk. That is, CV measures the household's willingness to pay to avoid aggregate income risk. Since we believe households may not be aware of the exact nonlinear relationship between their income and the aggregate shocks, we present the calculations for two cases: on the left we show estimates based on a process where η_{it} depends linearly on Z_t and Z_{t-1}, while on the right we show estimates based on a model that allows for a quadratic impact.

Figure 17 shows strikingly different magnitudes for the aggregate risk households face, depending on their perceived income process. In the linear case (left) the CV coefficient is very close to zero. This is in line with the original calculations in Lucas (2003). In contrast, in the quadratic case (right), magnitudes are much higher, suggesting that households would be willing to give up between 4% and 10% of their income each period in order to avoid cyclical fluctuations, with somewhat larger CV's for the young. This finding based on a nonlinear model suggests a different magnitude of the cost of business cycles compared to the classic findings in the literature.

Lastly, in Figure 18 we report a similar calculation of the compensating variation, in this case for removing idiosyncratic risk. We see magnitudes are large, especially for the

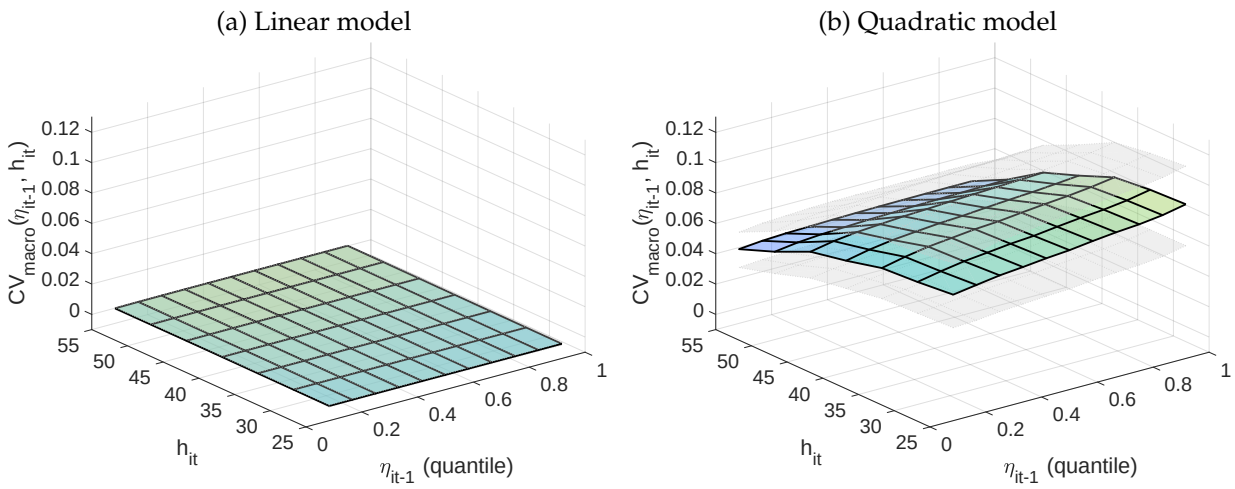


FIGURE 17. Macro risk quantification for disposable income.

Note: Panels (a) and (b) report compensating variation for aggregate risk for different age and levels of initial income. Panel (a) corresponds to estimates from a model that restrict the aggregate state to enter linearly, whereas panel (b) allows a quadratic aggregate term in the intercept.

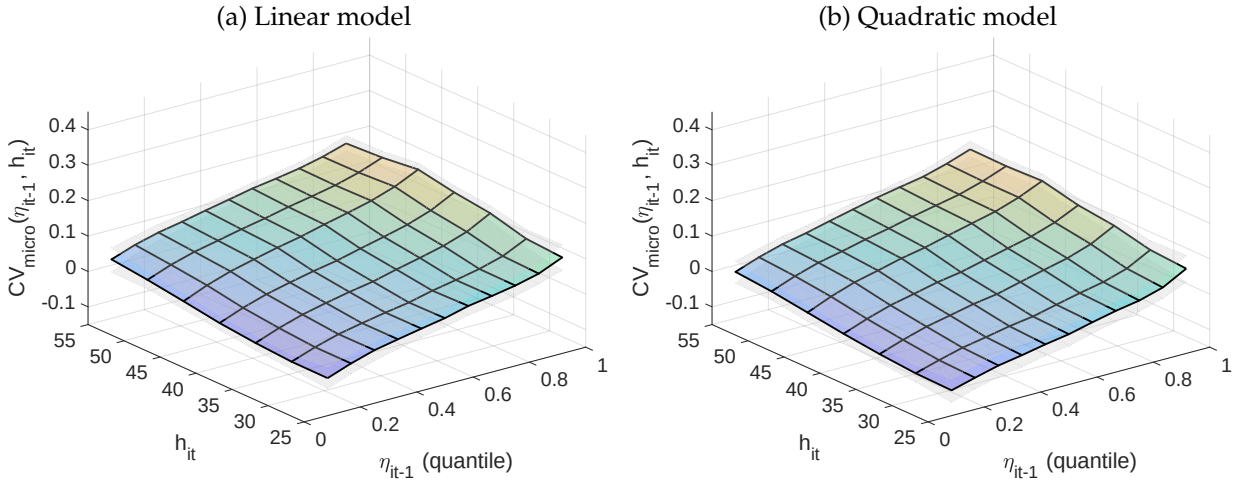


FIGURE 18. Micro risk quantification for disposable income.

older and richer households, suggesting households are willing to give up as much as 20% of their income each period in order to remove the sources of idiosyncratic risk.

8 Conclusion

In this paper we propose a nonlinear semi-structural framework for heterogeneous agents models with aggregate shocks. The nonlinear reduced form relationships are useful to

assess the fit of such models and their implications for micro data. The framework allows for general nonlinearities and dynamics. We study identification, estimation and inference tools for this purpose.

We find that the nonlinear persistence and conditional skewness of earnings documented in previous work are affected by business cycle fluctuations. The persistence surface tilts in recessions, with low persistence becoming more likely for high income households hit by a bad shock, and high persistence being more likely for low income households hit by a good shock. Skewness lowers in recessions throughout the income distribution, especially for the high-income households.

Lastly, we find preliminary empirical evidence that nonlinearities in the relationship between the income process and the aggregate shocks matters for risk calculations. One avenue we are exploring is how to account for both model and estimation uncertainty in the risk calculations.

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.

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.

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

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. (2005): "The cost of business cycles and the benefits of stabilization." *Economic Perspectives*, 29, 32–49.

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.

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.

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*, National Bureau of Economic Research.

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.

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

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.

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.

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.

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*, National Bureau of Economic Research, 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.

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.

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

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

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

Sims, C. A. (1980): "Macroeconomics and reality," *Econometrica*, 48, 1–48.

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 (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.

A Proofs

Proof of Proposition 3. Take τ such that $\mathcal{F}_{z_\tau|w_t^S}$ is complete. For any $\widetilde{\eta}$, η, x, and $\boldsymbol{w}$,

$$\begin{aligned} E\left[P_{\eta,\tau}(\widetilde{\eta}|\eta,x)\middle|w_t^S=\boldsymbol{w}\right] &= E\left[P\left(\eta_{i\tau}\le\widetilde{\eta}\middle|\eta_{i,\tau-1}=\eta,x_{i\tau}=x,Z_\tau,Z_{\tau-1},F_\tau\right)\middle|w_t^S=\boldsymbol{w}\right] \\ &= E\left[P\left(\eta_{i\tau}\le\widetilde{\eta}\middle|\eta_{i,\tau-1}=\eta,x_{i\tau}=x,Z_\tau,Z_{\tau-1}\right)\middle|w_t^S=\boldsymbol{w}\right], \end{aligned}$$

where the first line uses the definition of $P_{\eta,\tau}$ and the second uses independence between $(Z_\tau, Z_{\tau-1})$ and F_τ conditional on $w_t^S = (W_t, ..., W_{t+S-1})$, which comes from Assumption 1(b).

Differentiating with respect to $\widetilde{\eta}$ ($\eta_{i\tau}$ is continuously distributed by Assumption 2(b)),

$$\frac{\partial E\left[P_{\eta,\tau}(\widetilde{\eta}|\eta,x)\middle|w_t^S=\boldsymbol{w}\right]}{\partial\widetilde{\eta}} = \int_{\mathcal{Z}^2} f_{\eta_\tau|\eta_{\tau-1},x_\tau,z_\tau}(\widetilde{\eta}|\eta,x,\widetilde{Z},Z)\, f_{z_\tau|w_t^S}(\widetilde{Z},Z|\boldsymbol{w})\, d(\widetilde{Z},Z). \tag{10}$$

Here $\mathcal{Z}$ denotes the support of Z_t and $f_{\eta_\tau|\eta_{\tau-1},x_\tau,z_\tau}$ is the density of $\eta_{i\tau}$ given $(\eta_{i,\tau-1}, x_{i\tau}, Z_\tau, Z_{\tau-1})$. Now, the object on the left is identified by Proposition 2 and the density $f_{z_\tau|w_t^S}$ is identified by Proposition 1. The only unknown in Equation (10) is therefore $f_{\eta_\tau|\eta_{\tau-1},x_\tau,z_\tau}$.

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

$$\begin{aligned} \left[L_{\eta_\tau|\eta_{\tau-1},x_\tau,w_t^S}h_1\right](\widetilde{\eta},\eta,x) &= \int_{\mathcal{W}^S} \frac{\partial E\left[P_{\eta,\tau}(\widetilde{\eta}|\eta,x)\middle|w_t^S=\boldsymbol{w}\right]}{\partial\widetilde{\eta}}\, h_1(\boldsymbol{w})\, d\boldsymbol{w}, \\ \left[L_{z_\tau|w_t^S}h_1\right](\widetilde{Z},Z) &= \int_{\mathcal{W}^S} f_{z_\tau|w_t^S}(\widetilde{Z},Z|\boldsymbol{w})\, h_1(\boldsymbol{w})\, d\boldsymbol{w}, \\ \left[L_{\eta_\tau|\eta_{\tau-1},x_\tau,z_\tau}h_2\right](\widetilde{\eta},\eta,x) &= \int_{\mathcal{Z}^2} f_{\eta_\tau|\eta_{\tau-1},x_\tau,z_\tau}(\widetilde{\eta}|\eta,x,\widetilde{Z},Z)\, h_2(\widetilde{Z},Z)\, d(\widetilde{Z},Z), \end{aligned}$$

so that Equation (10) is equivalent to (see, e.g., Carrasco, Florens, and Renault (2007))

$$L_{\eta_\tau|\eta_{\tau-1},x_\tau,w_t^S} = L_{\eta_\tau|\eta_{\tau-1},x_\tau,z_\tau} L_{z_\tau|w_t^S}. \tag{11}$$

By our previous discussion, $L_{\eta_\tau|\eta_{\tau-1},x_\tau,w_t^S}$ and $L_{z_\tau|w_t^S}$ are known to the researcher. Since $\mathcal{F}_{z_\tau|w_t^S}$ is complete, $L_{z_\tau|w_t^S}$ has a right inverse and Equation (11) has solution

$$L_{\eta_\tau|\eta_{\tau-1},x_\tau,z_\tau} = L_{\eta_\tau|\eta_{\tau-1},x_\tau,w_t^S} L^{-1}_{z_\tau|w_t^S}.$$

which uniquely determines $f_{\eta_\tau|\eta_{\tau-1},x_\tau,z_\tau}$ (up to a set of zero Lebesgue measure). It follows

that Q_η in (1) is identified. The argument for Q_ε in (2) is entirely analogous. □

B Measuring risk via compensating variation

Let θ_0 denote the true value of θ. We suppose the agent has a prior distribution Π over θ — and, to begin, we abstract from the fact that θ may be infinite-dimensional.

We define compensating variations for model, macro, and micro risks as follows. We fix initial values $y_{i1} = \overline{y}$ and $Z_1 = \overline{Z}$. We also set a discount factor β, a period utility function $U(\cdot)$ and a maximum lifespan H. Below we let Φ_t be the CDF from the $N(0_{(t-1)\times 1}, I_{t-1})$ distribution.

For model risk we define C_{model} as the solution to

$$0 = \sum_{t=2}^{H} \beta^t \left[\int\int\int U\left(Y_t(\overline{y}, \overline{Z}, V_{2:t}, u_{i,2:t}, \theta) + C_{model}\right) d\Phi_t(V_{2:t}) d\Phi_t(u_{i,2:t}) d\Pi(\theta) \right.$$
$$\left. - \int\int U\left(Y_t(\overline{y}, \overline{Z}, V_{2:t}, u_{i,2:t}, \theta_0)\right) d\Phi_t(V_{2:t}) d\Phi_t(u_{i,2:t}) \right],$$

Similarly, for macro risk we define C_{macro} as

$$0 = \sum_{t=2}^{H} \beta^t \left[\int\int U\left(Y_t(\overline{y}, \overline{Z}, V_{2:t}, u_{i,2:t}, \theta_0) + C_{macro}\right) d\Phi_t(V_{2:t}) d\Phi_t(u_{i,2:t}) \right.$$
$$\left. - \int U\left(Y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, u_{i,2:t}, \theta_0)\right) d\Phi_t(u_{i,2:t}) \right],$$

and C_{micro} as

$$0 = \sum_{t=2}^{H} \beta^t \left[\int U\left(Y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, u_{i,2:t}, \theta_0) + C_{micro}\right) d\Phi_t(u_{i,2:t}) \right.$$
$$\left. - U\left(Y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)\right) \right].$$

Thus, C_{model}, C_{macro} and C_{micro} measure how much income the agent would require in order to be compensated for bearing the risk from not knowing θ_0, aggregate shocks and idiosyncratic shocks, respectively.[13]

[13] Because the model is nonlinear, there are in principle different ways of defining compensating variations that depend on the order in which we remove each source of risk.

Alternatively, given that we specify the income process in logs and we usually assume U to be power (or log) utility, it is convenient to measure risk as a proportion of income rather than an amount. In particular, let $\widetilde{U}(y) = U(e^y)$. Then, we define

$$0 = \sum_{t=2}^{H} \beta^t \left[\int\int\int \widetilde{U}\left(y_t(\overline{y}, \overline{Z}, V_{2:t}, u_{i,2:t}, \theta) + c_{model}\right) d\Phi_t(V_{2:t}) d\Phi_t(u_{i,2:t}) d\Pi(\theta) \right.$$
$$\left. - \int\int \widetilde{U}\left(y_t(\overline{y}, \overline{Z}, V_{2:t}, u_{i,2:t}, \theta_0)\right) d\Phi_t(V_{2:t}) d\Phi_t(u_{i,2:t}) \right],$$

for model risk, and we define c_{macro} and c_{micro} in the obvious way.

We can approximate c_{macro} and c_{micro} locally around a scenario with no risk by substituting $V_{2:t}$ and $u_{i,2:t}$ by $\gamma_V V_{2:t}$ and $\gamma_u u_{i,2:t}$ above, and computing second-order Taylor expansions around $\gamma = (\gamma_V, \gamma_u) = (0, 0)$. This results in the following:

$$c_{macro} \approx -\left[\frac{\sum_{t=2}^{H} \sum_{\ell=0}^{t-2} \beta^t \widetilde{U}''\left(y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)\right) \left(\partial y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)/\partial V_{t-\ell}\right)^2}{\sum_{t=2}^{H} \beta^t \widetilde{U}'\left(y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)\right)} \right.$$
$$\left. + \frac{\sum_{t=2}^{H} \sum_{\ell=0}^{t-2} \beta^t \widetilde{U}'\left(y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)\right) \left(\partial^2 y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)/\partial V_{t-\ell}^2\right)}{\sum_{t=2}^{H} \beta^t \widetilde{U}'\left(y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)\right)} \right],$$

$$c_{micro} \approx -\left[\frac{\sum_{t=2}^{H} \sum_{\ell=0}^{t-2} \beta^t \widetilde{U}''\left(y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)\right) \left(\partial y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)/\partial u_{i,t-\ell}\right)^2}{\sum_{t=2}^{H} \beta^t \widetilde{U}'\left(y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)\right)} \right.$$
$$\left. + \frac{\sum_{t=2}^{H} \sum_{\ell=0}^{t-2} \beta^t \widetilde{U}'\left(y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)\right) \left(\partial^2 y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)/\partial u_{i,t-\ell}^2\right)}{\sum_{t=2}^{H} \beta^t \widetilde{U}'\left(y_t(\overline{y}, \overline{Z}, 0_{(t-1)\times 1}, 0_{(t-1)\times 1}, \theta_0)\right)} \right].$$

For CRRA utility functions of the form $U(Y) = (Y^{1-\zeta} - 1)/(1 - \zeta)$ we get derivatives $\widetilde{U}'(y) = e^{(1-\zeta)y}$ and $\widetilde{U}''(y) = (1 - \zeta)e^{(1-\zeta)y}$. In the log-utility limit $\zeta = 1$, these reduce to $\widetilde{U}'(y) = 1$ and $\widetilde{U}''(y) = 0$ for all y. In that case, c_{macro} and c_{micro} are approximately present discounted values of cumulative second derivatives of log-income:

$$c_{macro} \approx -\left[\frac{\sum_{t=2}^{H}\sum_{\ell=0}^{t-2}\beta^t\left(\partial^2 y_t(\overline{y},\overline{Z},0_{(t-1)\times 1},0_{(t-1)\times 1},\theta_0)/\partial V_{t-\ell}^2\right)}{\sum_{t=2}^{H}\beta^t}\right]$$

$$c_{micro} \approx -\left[\frac{\sum_{t=2}^{H}\sum_{\ell=0}^{t-2}\beta^t\left(\partial^2 y_t(\overline{y},\overline{Z},0_{(t-1)\times 1},0_{(t-1)\times 1},\theta_0)/\partial u_{i,t-\ell}^2\right)}{\sum_{t=2}^{H}\beta^t}\right].$$

More generally, risk can originate in curvature in the utility function or nonlinear effects of shocks in the (log) income process.