

Almuzara, Martín; Arellano, Manuel; Blundell, Richard W.; Bonhomme, Stéphane

Working Paper

Nonlinear micro income processes with macro shocks

cemmap working paper, No. CWP17/25

Provided in Cooperation with:

The Institute for Fiscal Studies (IFS), London

Suggested Citation: Almuzara, Martín; Arellano, Manuel; Blundell, Richard W.; Bonhomme, Stéphane (2025) : Nonlinear micro income processes with macro shocks, cemmap working paper, No. CWP17/25, Centre for Microdata Methods and Practice (cemmap), The Institute for Fiscal Studies (IFS), London,
<https://doi.org/10.47004/wp.cem.2025.1725>

This Version is available at:

<https://hdl.handle.net/10419/324455>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

SUPPLEMENTAL MATERIAL

Nonlinear Micro Income Processes with Macro Shocks

MARTÍN ALMUZARA*

MANUEL ARELLANO[‡]

RICHARD BLUNDELL[†]

STÉPHANE BONHOMME[§]

August 2025

[\[Link to the paper\]](#)

A Stylized model

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

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

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

[†]UCL and IFS: r.blundell@ucl.ac.uk

[‡]CEMFI: arellano@cemfi.es

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

B Identification

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

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

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

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

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

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

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

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

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

$$\begin{aligned} E[F_{\eta_r, t}(\tilde{\eta}|\eta, x) \mid W_t^S = W] &= E[P(\eta_{ir} \leq \tilde{\eta} \mid \eta_{i, r-1} = \eta, x_{ir} = x, \omega_t) \mid W_t^S = W] \\ &= E[P(\eta_{ir} \leq \tilde{\eta} \mid \eta_{i, r-1} = \eta, x_{ir} = x, Z_r, Z_{r-1}, G_r) \mid W_t^S = W] \\ &= E[P(\eta_{ir} \leq \tilde{\eta} \mid \eta_{i, r-1} = \eta, x_{ir} = x, Z_r, Z_{r-1}) \mid W_t^S = W], \end{aligned}$$

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

The previous equation can be written more explicitly as

$$E[F_{\eta_r, t}(\tilde{\eta}|\eta, x) \mid W_t^S = W] = \int_{\mathcal{Z}} F_{\eta_r|\eta_{r-1}, x_r, z_r}(\tilde{\eta}|\eta, x, Z_t, Z_{t-1}) f_{z_r|W_t^S}(Z_t, Z_{t-1}|\mathbf{W}) d(Z_t, Z_{t-1}).$$

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

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

$$\begin{aligned} [L_{\eta_r|\eta_{r-1}, x_r, W_t^S} h_1](\tilde{\eta}, \eta, x) &= \int_{\mathcal{W}^S} E[F_{\eta_r, t}(\tilde{\eta}|\eta, x) \mid W_t^S = W] h_1(W) dW, \\ [L_{z_r|W_t^S} h_1](Z_t, Z_{t-1}) &= \int_{\mathcal{W}^S} f_{z_r|W_t^S}(Z_t, Z_{t-1}|\mathbf{W}) h_1(W) dW, \end{aligned}$$

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

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

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

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

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

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

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

C Estimation

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

C.1 Moment conditions

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

Write $v_u(\omega) = u - 1\{\omega < 0\}$. For nodes $u = \bar{u}_1, \dots, \bar{u}_L$, we will use the orthogonality conditions from quantile regression ([Koenker and Bassett, 1978](#)),

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

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

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

$$\begin{aligned}
& \times \left[\psi(\eta_{i,\tau-1}, x_{i\tau})' \Theta(\bar{u}_1) \varphi(Z_\tau, Z_{\tau-1}) - \eta_{i\tau} - \exp\left(\psi_{\text{lo}}(\eta_{i,\tau-1}, Z_\tau, Z_{\tau-1}, x_{i\tau})' \theta_{\text{lo}}\right) \right], \\
m_{it}^{\text{up}}(\theta) &= \sum_{\tau=t+1}^{t+S-1} \psi_{\text{up}}(\eta_{i,\tau-1}, Z_\tau, Z_{\tau-1}, x_{i\tau}) \times 1\left\{ \eta_{i\tau} > \psi(\eta_{i,\tau-1}, x_{i\tau})' \Theta(\bar{u}_L) \varphi(Z_\tau, Z_{\tau-1}) \right\} \\
& \times \left[\eta_{i\tau} - \psi(\eta_{i,\tau-1}, x_{i\tau})' \Theta(\bar{u}_L) \varphi(Z_\tau, Z_{\tau-1}) - \exp\left(\psi_{\text{up}}(\eta_{i,\tau-1}, Z_\tau, Z_{\tau-1}, x_{i\tau})' \theta_{\text{up}}\right) \right].
\end{aligned}$$

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

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

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

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

C.2 Techniques for posterior sampling

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

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

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

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

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

The Gibbs sampling for our linear state-space model is a standard technique that builds

on the following conditional distributions:

- (a) Given $\{W_t, Z_t\}$, the distribution of parameters can be written in terms of easy-to-draw multivariate normal and inverse gamma random variables.
- (b) Given parameters, the distribution of $\{Z_t, \{e_{jt}\}_{j=1}^{n_w}\}$ is multivariate normal and can be efficiently sampled from using the algorithm of [Durbin and Koopman \(2002\)](#).

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

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

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

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

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

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

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

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

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

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

- ($s > 0$) ◦ If $y_{i,t+s}$ is missing:

- * Draw particles $\{\eta_{i,t+s}^k\}_{k=1}^{\bar{K}}$ from the conditional density $f_{\eta}(\cdot | \eta_{i,t+s-1}^k, X_{i,t+s})$.
- * Set the weights $\{w_{i,t+s}^k\}_{k=1}^{\bar{K}}$ to $w_{i,t+s}^k = w_{i,t+s-1}^k$.

- If $y_{i,t+s}$ is not missing:

- * Draw particles $\{\eta_{i,t+s}^k\}_{k=1}^{\bar{K}}$ from the proposal, $\eta_{i,t+s}^k \sim q_{\eta}(\cdot | \eta_{i,t+s-1}^k, y_{i,t+s}, X_{i,t+s})$.
- * Set the weights $\{w_{i,t+s}^k\}_{k=1}^{\bar{K}}$ to

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

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

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

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

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

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

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

$$\omega_{\text{init}}^2 = (1/s_{\text{init}}^2 + 1/s_{\varepsilon}^2)^{-1}.$$

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

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

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

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

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

C.3 Asymptotic approximations

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

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

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

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

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \bar{M}_{\theta,t}(\widehat{\theta}, \widehat{\delta}_t, \widehat{\lambda}) &= 0, \\ \bar{M}_{\delta,t}(\widehat{\theta}, \widehat{\delta}_t, \widehat{\lambda}) &= 0, \quad t = 1, \dots, T. \end{aligned}$$

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

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

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

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

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T (\bar{D}_{\theta\delta,t} \bar{D}_{\delta\delta,t}^{-1} \bar{D}_{\delta\theta,t} + \bar{D}_{\theta\theta,t}) &\xrightarrow{P} D_{\theta\theta,0}, \\ \frac{1}{T} \sum_{t=1}^T (\bar{D}_{\theta\delta,t} \bar{D}_{\delta\delta,t}^{-1} \bar{D}_{\delta\lambda,t} + \bar{D}_{\theta\lambda,t}) &\xrightarrow{P} D_{\theta\lambda,0}, \end{aligned}$$

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

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

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

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

C.4 Bootstrap approach

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

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

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

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

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

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

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

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

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

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

- 1) Simulate the time series of aggregate factors $\{G_{\eta,t}, G_{\varepsilon,t}, G_{\text{init},t}\}_{t=1}^T$.
- 2) For each unit i determine the first (t_0) and last (t_1) period in the dataset. Next,
 - (i) draw the path of idiosyncratic shocks $\{\tilde{u}_{it}, \tilde{v}_{it}, \tilde{v}_{it'}\}_{t_0 \leq t \leq t_1}$ imposing the correlations d_η , d_ε and d_{init} across consecutive periods;
 - (ii) combine aggregate and idiosyncratic factors to obtain $\{u_{it}, v_{it}, v_{it'}\}_{t_0 \leq t \leq t_1}$ imposing the cross-sectional dependence implied by c_η , c_ε and c_{init} ;
 - (iii) for the first two periods, use $Q_{\text{init},t}$ and v_{it} to generate η_{it} ;
 - (iv) for every other period, use Q_η and u_{it} to generate η_{it} ;
 - (v) for all periods, use $Q_{\varepsilon,t}$ and v_{it} to generate ε_{it} ;
 - (vi) form $y_{it} = \eta_{it} + \varepsilon_{it}$ for all $t_0 \leq t \leq t_1$.
- 3) Assign the data to the appropriate unit and time cell.

D Additional empirical results

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

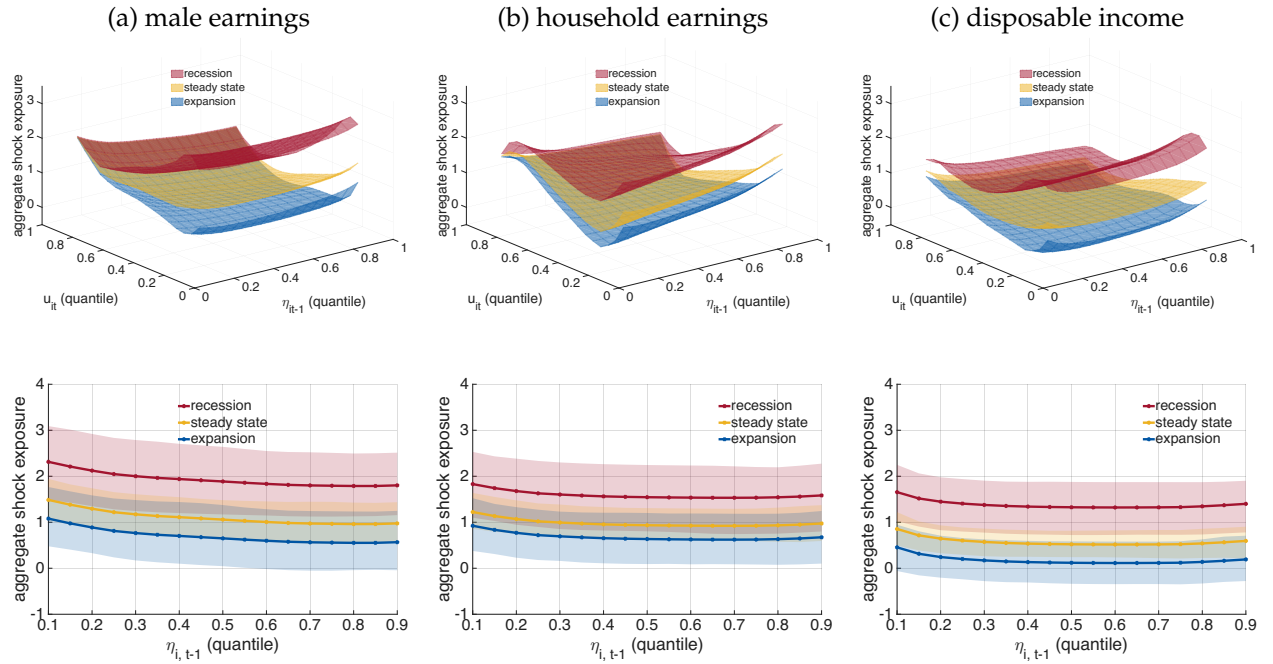


FIGURE D.1. Nonlinear exposure to aggregate shocks.

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

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

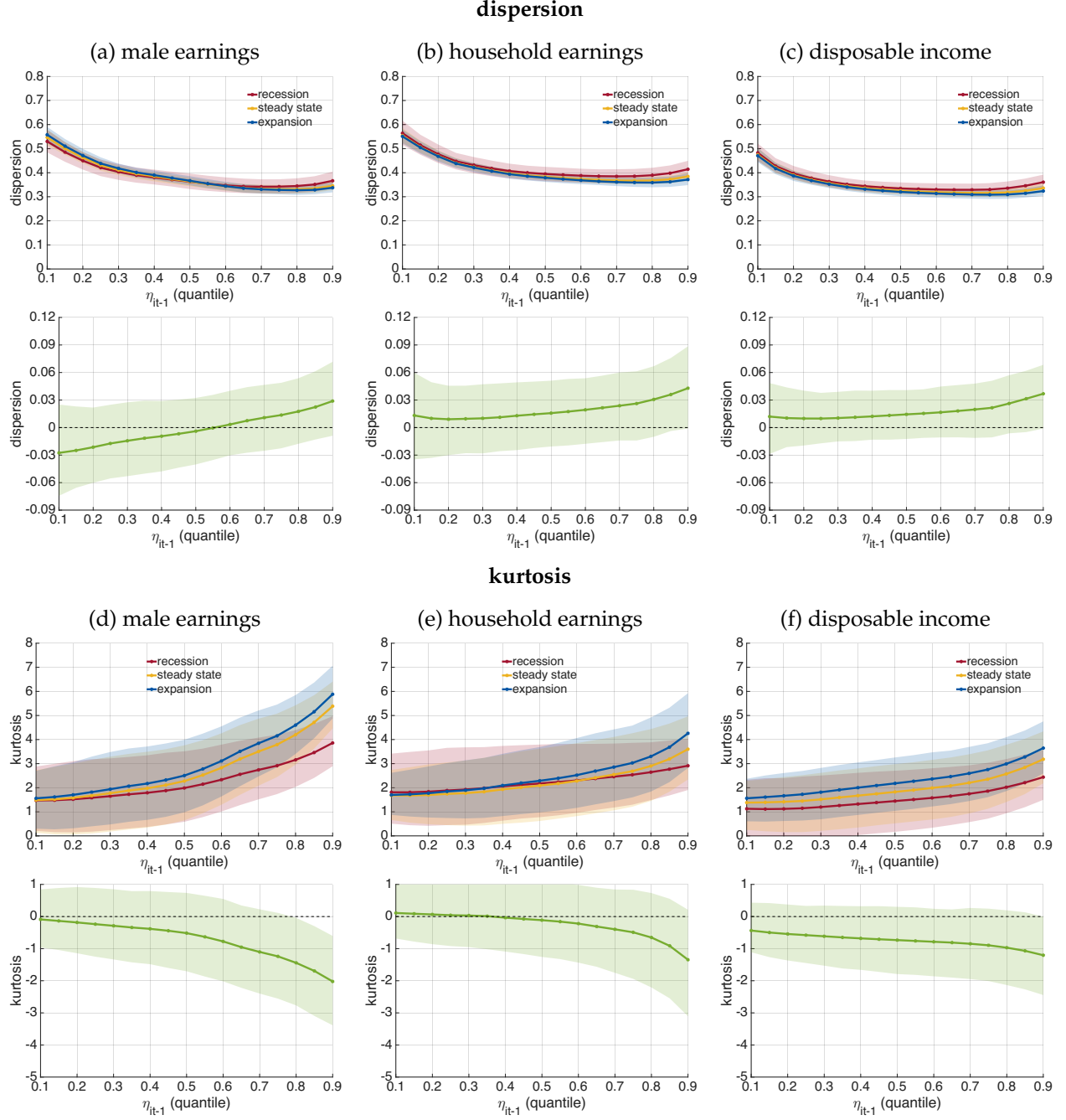


FIGURE D.2. Measures of dispersion and kurtosis.

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

E Additional material on impulse response analysis

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

E.1 Perturbations and shocks

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

$$\begin{aligned}\text{IRF}_{\eta Z}(h; \delta) &= \frac{E[\eta_{i,t+h} \mid \eta_{i,t-1}, \tilde{V}_t = \delta, Z_{t-1}] - E[\eta_{i,t+h} \mid \eta_{i,t-1}, \tilde{V}_t = 0, Z_{t-1}]}{\delta}, \\ \text{IRF}_{\eta\eta}(h, \delta) &= \frac{E[\eta_{i,t+h-1} \mid \tilde{u}_{i,t-1} = \delta, \eta_{i,t-2}, Z_t, Z_{t-1}] - E[\eta_{i,t+h-1} \mid \tilde{u}_{i,t-1} = 0, \eta_{i,t-2}, Z_t, Z_{t-1}]}{\delta}\end{aligned}$$

and, for infinitesimal changes,

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

The implied shocks are given by

$$\begin{aligned}\tilde{V}_t &= g(Q_Z(Z_{t-1}, V_t)) - g(Z^b), \\ \tilde{u}_{i,t-1} &= g(Q_\eta(\eta_{i,t-2}, Z_{t-1}, Z_{t-2}, u_{i,t-1})) - g(\eta^b),\end{aligned}$$

and lead to the representations

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

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

E.2 Impulse responses, nonlinear persistence and aggregate exposures

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

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

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

$$q_{\eta,h}(\mathbf{u}_{it}^h, \mathbf{V}_{t+1}^{h-1}, \eta_{i,t-1}, Z_t, Z_{t-1}) = Q_\eta \left(q_{\eta,h-1}(\mathbf{u}_{it}^{h-1}, \mathbf{V}_{t+1}^{h-2}, \eta_{i,t-1}, Z_t, Z_{t-1}), \dots, q_{Z,h}(\mathbf{V}_{t+1}^{h-1}, Z_t), q_{Z,h-1}(\mathbf{V}_{t+1}^{h-2}, Z_t), u_{i,t+h} \right),$$

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

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

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

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

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

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

The impulse responses of the macro state using our methodology is

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

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

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

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

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

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

by introducing dependence on the potential history of future shocks.

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

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

$$\gamma_{i,t+1} = -\rho_{i,t+1}\beta_{it}, \quad (\text{E.2})$$

then

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

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

E.3 Additional IRF figures: comparison to MBC shocks

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

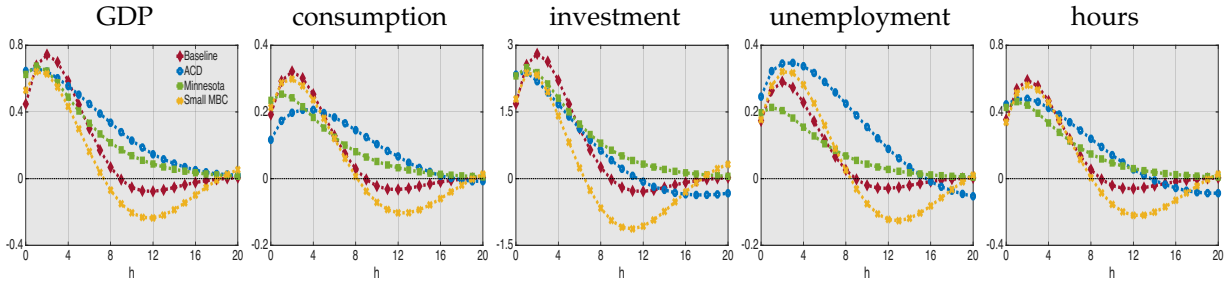


FIGURE E.1. IRFs of W_t to V_t and MBC shock

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

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

While the discrepancy is small relative to the statistical uncertainty around the IRFs, part of it can be attributed to the choice of prior. In our case, the dynamic factor structure already achieves, without further penalization, adequate dimension reduction. Instead, the 10-variable VAR(2) underlying the original MBC shock IRFs is based on a Minnesota prior. This choice is understandable, but penalizes deviations from unit roots that may bias the estimated persistence upward. To explore the issue, Figure E.1 shows two additional

estimates: IRFs obtained from a dynamic factor model under a Minnesota prior (green, squares), and IRFs for an MBC-type shock from a small 5-variable VAR(2) on W_t under a flat prior (gold, crosses).⁷ Consistent with our claim, the former mimics the persistence of the original MBC responses while the latter matches our baseline closely. But reassuringly, the small-model MBC shock and our V_t shock are highly correlated as seen in Figure E.2.

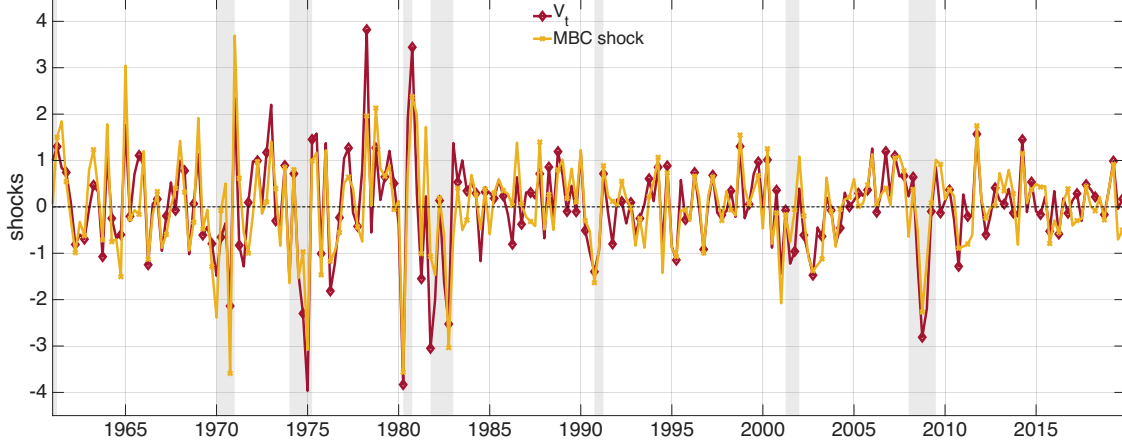


FIGURE E.2. V_t and MBC shocks

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

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

E.4 Additional IRF figures: local projection estimates

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

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

⁸These robustness checks are available in our replication package.

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

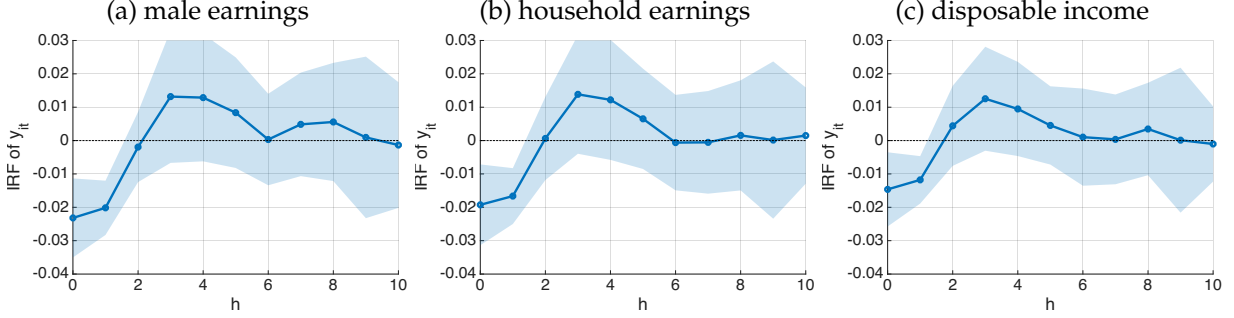


FIGURE E.3. Local projection estimates of macro impulse responses

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

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

E.5 Additional IRF figures: positive shocks

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

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

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

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

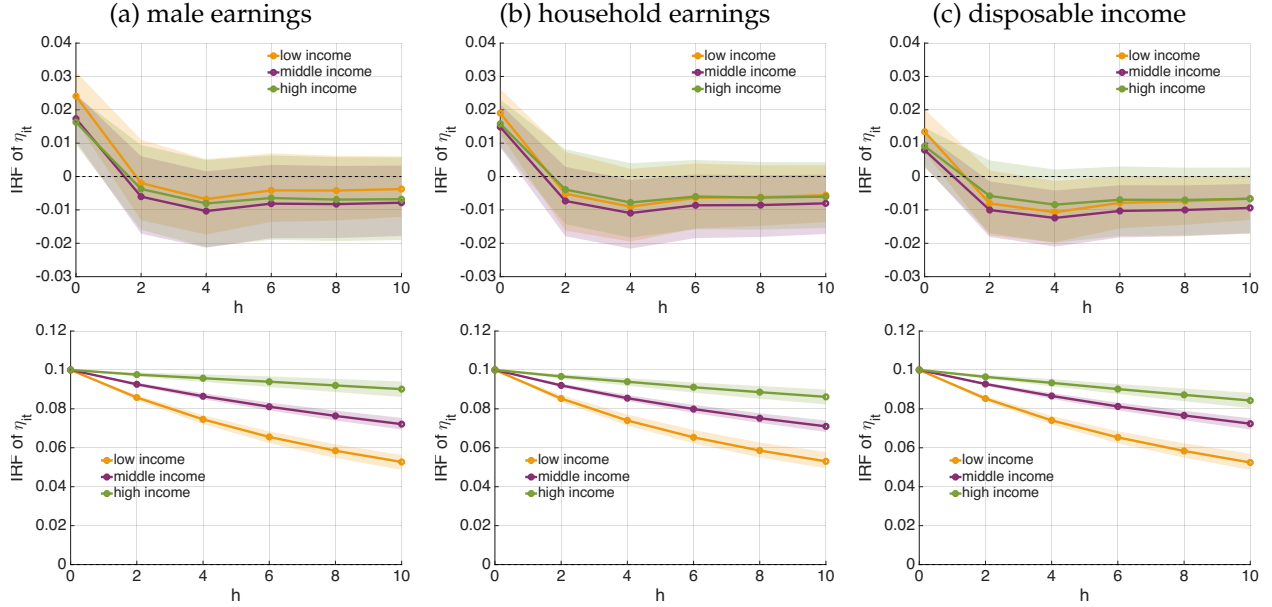


FIGURE E.4. Macro and micro impulse responses to positive shocks

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

References

- ALMUZARA, M. (2020): “Heterogeneity in Transitory Income Risk.” Working paper.
- ALMUZARA, M. AND V. SANCIBRIÁN (2024): “Micro Responses to Macro Shocks.” Federal Reserve Bank of New York Staff Report 1090.
- ANGELETOS, G.-M., F. COLLARD, AND H. DELLAS (2020): “Business-Cycle Anatomy.” *American Economic Review*, 110, 3030–3070.
- ARELLANO, M., R. BLUNDELL, S. BONHOMME, AND J. LIGHT (2023): “Heterogeneity of Consumption Responses to Income Shocks in the Presence of Nonlinear Persistence,” *Journal of Econometrics*, 1–45.

- ARELLANO, M. AND S. BONHOMME (2016): "Nonlinear panel data estimation via quantile regressions." *Econometrics Journal*, 19, C61–C94.
- (2017): "Nonlinear panel data methods for dynamic heterogeneous agent models." *Annual Review of Economics*, 9, 471–496.
- BAYER, C. AND R. LUETTICKE (2020): "Solving discrete time heterogeneous agent models with aggregate risk and many idiosyncratic states by perturbation." *Quantitative Economics*, 11, 1253–1288.
- CARRASCO, M., J. P. FLORENS, AND E. RENAULT (2007): "Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization." in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, Elsevier, vol. 6B, chap. 77.
- CHEN, X. (2007): "Large sample sieve estimation of semi-nonparametric models." in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, Elsevier, vol. 6B, chap. 76.
- CREAL, D. (2012): "A Survey of Sequential Monte Carlo Methods for Economics and Finance." *Econometric Reviews*, 31, 245–296.
- DURBIN, J. AND S. J. KOOPMAN (2002): "A simple and efficient simulation smoother for state space time series analysis." *Biometrika*, 89, 603–615.
- GEWEKE, J. F. AND K. J. SINGLETON (1981): "Maximum Likelihood "Confirmatory" Factor Analysis of Economic Time Series." *International Economic Review*, 22, 37–54.
- GRILICHES, Z. (1961): "A note on serial correlation bias in estimates of distributed lags." *Econometrica*, 29, 65–73.
- (1967): "Distributed Lags: A Survey." *Econometrica*, 35, 16–49.
- GUVENEN, F., S. OZKAN, AND J. SONG (2014): "The nature of countercyclical income risk." *Journal of Political Economy*, 122, 621–660.
- HU, Y. AND S. M. SCHENNACH (2008): "Instrumental variable treatment of nonclassical measurement error models." *Econometrica*, 76, 195–216.
- KOENKER, R. AND G. BASSETT (1978): "Regression quantiles." *Econometrica*, 46, 33–50.
- KRUSELL, P. AND A. A. SMITH (1998): "Income and wealth heterogeneity in the macroeconomy." *Journal of Political Economy*, 106, 867–896.

- NEWKEY, W. K. (1997): "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics*, 79, 147–168.
- NIELSEN, S. F. (2000): "The stochastic EM algorithm: estimation and asymptotic results." *Bernoulli*, 6, 457–489.
- REITER, M. (2009): "Solving heterogeneous-agent models by projection and perturbation." *Journal of Economic Dynamics and Control*, 33, 649–665.
- SARGAN, J. D. (1964): "Wages and Prices in the United Kingdom: A Study in Econometric Methodology." in *Econometric Analysis for National Economic Planning*, ed. by P. E. Hart, G. Mills, and J. N. Whitaker, Butterworths, vol. 16, 25–54.
- (1980): "Some Tests of Dynamic Specification for a Single Equation." *Econometrica*, 48, 879–897.
- WINBERRY, T. (2018): "A method for solving and estimating heterogeneous agent macro models." *Quantitative Economics*, 9, 1123–1151.