Supplemental Material
Nonlinear Micro Income Processes with Macro Shocks

Martín Almuzara Manuel Arellano

Richard Blundell Stéphane Bonhomme

June 2026

[Link to the paper]

A Identification

As highlighted in Section 3, our identification strategy consists of two steps. For the first one (the "micro" step), a key tool is the ability to recover conditional quantiles of the persistent component $Q_\eta(\cdot, z_{t+s}, z_{t+s-1}, \cdot)$ *at particular realizations of the aggregate state* $Z_\tau = z_\tau$ from F_t, the joint CDF of income $(y_{it}, \ldots, y_{i,t+S-1})$, available in short panels. We achieve this by relying on the identification argument in Arellano, Blundell, and Bonhomme (2017) who, in turn, build on Hu and Schennach (2008) and Wilhelm (2015).

Proposition 1. *Suppose $S \geq 4$ and almost surely over realized paths of $\{z_\tau\}$, for each t,*

(a) the density of $\{y_{i,t+s}, \eta_{i,t+s}\}_{s=0}^{S-1}$ conditional on $\{z_{t+s}\}_{s=0}^{S-1}$ is bounded away from zero and infinity on $\mathbb{R}^S \times \mathcal{A}^S$, and so are all corresponding marginals and conditionals, and

(b) letting $f_{\eta_r|y_{\tilde{r}};t}$ be the density of η_{ir} conditional on $y_{i\tilde{r}}$ and $\{z_{t+s}\}_{s=0}^{S-1}$, the classes of densities $\left\{f_{\eta_r|y_{r-1};t}(\cdot|y) : y \in \mathbb{R}\right\}$ and $\left\{f_{\eta_r|y_{r+1};t}(\cdot|y) : y \in \mathbb{R}\right\}$ are complete in L^1 for $t < r < t+S-1$. In addition, the joint characteristic function of $(\varepsilon_{ir}, \varepsilon_{i,r+1})$ does not vanish on $\mathbb{R}^2$.

Let Assumption 2 hold. Then, the mapping

$$(\eta, u) \mapsto Q_\eta(\eta, z_t, z_{t-1}, u), \qquad \eta \in \mathcal{A}, \quad u \in (0,1), \tag{A.1}$$

is identified for all t, almost surely in $\{z_\tau\}$.

Proof. Fix t and r such that $t < r < t + S - 1$ and consider

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

where the subindex t indicates that each density conditions on the realized path $\{z_{t+s}\}_{s=0}^{S-1}$ of the macro state in subpanel t. The density on the left-hand side is known to the researcher as it involves observable income only; the goal is to recover the densities on the right.

The decomposition defines an integral operator equation that can be solved by applying the diagonalization method of Hu and Schennach (2008). By (a) and (b), the operator equation and its spectral decomposition are well defined. Moreover, uniqueness of the decomposition is ensured by the normalization $E_t[\, y_{ir} \mid \eta_{ir}\,] = \eta_{ir}$ where the subindex t indicates that the expectation is an integral against the subpanel-specific density $f_{y_r|\eta_r;t}$.[1] This analysis delivers identification of $f_{y_{r-1}|\eta_r;t}$, $f_{y_r,\eta_r;t}$, and $f_{y_{r+1}|\eta_r;t}$. In particular, $f_{y_r|\eta_r;t}$ is identified.

Since $S \geq 4$, we can replicate this strategy replacing r by $r+1$, delivering identification of $f_{y_{r+1}|\eta_{r+1};t}$. Now, by independence between persistent and transitory components and serial independence of ε_{it} conditional on $\{z_{t+s}\}_{s=0}^{S-1}$ (Assumption 2), the joint density of transitory shocks $(e_0, e_1) \mapsto f_{\varepsilon_r,\varepsilon_{r+1};t}(e_0, e_1) = f_{y_r - \eta_r;t}(e_0) f_{y_{r+1}-\eta_{r+1};t}(e_1)$ is identified.

Lastly, under assumption (b), a standard deconvolution argument delivers identification of $f_{\eta_r,\eta_{r+1};t}$, from where it follows that $Q_\eta(\cdot, z_{r+1}, z_r, \cdot)$ is identified. □

Proposition 1 states that, given a realized path for the aggregate states, there is a known injective mapping from the joint CDF of observables F_t to the law of motion of the latent variables $Q_\eta(\cdot, z_t, z_{t-1}, \cdot)$ evaluated at the aggregate path. The *completeness* assumption in part (b), a nonparametric analog of the rank conditions often used in identification arguments for instrumental variable models, plays a central role in the result. In our context, it requires that η_{it} displays sufficient persistence over time to ensure dependence between η_{it} and both

[1] This follows from a simplified version of Almuzara (2020, Proposition 1) without heterogeneity. See, in particular, Remark 1 in that paper for further discussion on the role of the normalization. In our model, we assume that $E_t[\, \varepsilon_{ir}\,] = 0$ for all r and we impose it in the estimation algorithm by renormalizing the mixture parameters at the end of every iteration.

future and past income. Intuitively, if η_{it} had no persistence at all, it would be impossible to distinguish it from the transitory shocks ε_{it}.

A by-product of the argument is the identification of the transitory component density, $f_{\varepsilon_{t+s};t}$ for $1 \leq s \leq S-2$. Under the cross-panel consistency requirement $f_{\varepsilon_{t+s};t} = f_{\varepsilon_{t+s};t-1}$, the quantile function of the initial condition Q_{init,t_0} is also identified. This is a mild condition asking the transitory shock distribution of populations in two consecutive panels to agree *within* a given period, and it is compatible with unrestricted nonstationarity *across* periods.

Finally, note that Proposition 1 gives sufficient conditions for identification, but it may still be possible to recover the object of interest under a different set of assumptions and a different value of S. There are many other models of micro-level dynamics for which a large enough S enables identification of subpanel-specific latent variable distributions using micro data alone.[2] The question is then how to link them to the macro states of interest (the "macro" step), a problem to which we turn next.

Proposition 2. *Let Assumptions 1, 2 and 3, and the conditions of Proposition 1 hold. Suppose Q_Z has full conditional support, that is, for every $\tilde{z} \in \mathcal{Z}$ the conditional distribution of Z_t given $Z_{t-1} = \tilde{z}$ has support $\mathcal{Z}$. Then, the mapping*

$$(\eta, \tilde{z}, z, u) \mapsto Q_\eta(\eta, \tilde{z}, z, u), \qquad \eta \in \mathcal{A}, \quad \tilde{z}, z \in \mathcal{Z}, \quad u \in (0,1),$$

is identified.

Proof. Let $H_t = (Z_t, Z_{t-1})$. Since Z_t is first-order Markov, stationary, and has full conditional support, the support of H_t is $\mathcal{Z} \times \mathcal{Z}$. Now, Proposition 1 implies that (A.1) is identified for all $(z, \tilde{z})$ in the support of H_t. The conclusion follows. $\square$

Note that this identification argument is for observed Z_t. The case where Z_t is unobservable can be dealt with using related arguments as done formally in the working paper version (Almuzara, Arellano, Blundell, and Bonhomme, 2025). In particular, identification of the process for Z_t follows from standard dynamic factor techniques.

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

B Large-sample theory

Throughout the section, we provide intuition about the setup and assumptions by using a simple linear model as an illustrative example.

B.1 Model and estimator

Setup. For $t = 1, \ldots, T$, we observe macro data Z_t and micro data $\{Y_{it}\}_{i \in \mathcal{I}_t}$ where $\mathcal{I}_t$ is a collection of cross-sectional indexes of size $N_t = \#(\mathcal{I}_t)$. We are interested in the parameter vector $\theta \in \Theta \subseteq \mathbb{R}^k$ which—together with the time effects $\{\delta_t\}_{t=1}^T$ where $\delta_t \in \Delta \subseteq \mathbb{R}^{k_d}$—are determined by conditional moment restrictions:

$$E\left[\frac{1}{T} \sum_{t=1}^{T} \psi_t(Y_{it}, Z_t, \theta, \delta_t) \;\middle|\; \mathcal{Z}_T \right] = 0_{k \times 1}, \tag{B.1}$$

where $\mathcal{Z}_T = \sigma\left(\{Z_t\}_{t=1}^T\right)$, and

$$E[\, \varphi_t(Y_{it}, Z_t, \theta, \delta_t) \mid \mathcal{Z}_T \,] = 0_{k_d \times 1}, \qquad t = 1, \ldots, T. \tag{B.2}$$

To map this notation to our time series of panels setting, let t denote a subpanel, Y_{it} denote all the micro observations (i.e., sequences of income and covariates) for household i in subpanel t, and Z_t denote all the aggregate observations (that is, sequences of aggregate states) in that subpanel. The moment restrictions in (B.1)-(B.2) are *integrated* moment restrictions in the spirit of ABB, where the latent persistent income components have been integrated out with respect to their posterior distribution under the model.

The setup is parametric, with finite-dimensional spaces Θ and Δ. As in ABB, we abstract from the impact of simulation noise on the estimates in the theory, yet we account for the variability induced by the simulation by relying on the bootstrap for inference. See Nielsen (2000) and Arellano and Bonhomme (2016) for theoretical analyses that incorporate the impact of simulation noise.

The functions $\psi_t(\cdot)$ and $\varphi_t(\cdot)$ are known up to θ and $\{\delta_t\}_{t=1}^T$, and we focus on the just-identified case. For $c \in \Theta$ and $d \in \Delta$, we denote

$$\overline{\psi}_t(c, d) \equiv \frac{1}{N_t} \sum_{i \in \mathcal{I}_t} \psi_t(Y_{it}, Z_t, c, d), \qquad \overline{\varphi}_t(c, d) \equiv \frac{1}{N_t} \sum_{i \in \mathcal{I}_t} \varphi_t(Y_{it}, Z_t, c, d).$$

For the derivatives of $\overline{\psi}_t(c,d)$ we use

$$\nabla_\theta \overline{\psi}_t(c,d) = \frac{\partial \overline{\psi}_t(c,d)}{\partial \theta'}, \qquad \nabla_\delta \overline{\psi}_t(c,d) = \frac{\partial \overline{\psi}_t(c,d)}{\partial \delta_t'},$$

with similar notation for the derivatives of $\overline{\varphi}_t(c,d)$.

To describe the limiting objective function, we introduce the quantities

$$m_{\psi,t}(c,d) = E\left[\, \overline{\psi}_t(c,d) \;\middle|\; \mathcal{Z}_T \right], \qquad m_{\varphi,t}(c,d) = E[\, \overline{\varphi}_t(c,d) \mid \mathcal{Z}_T\,],$$

and, for $j = \theta, \delta$, we use $\nabla_j m_{\psi,t}(c,d)$ and $\nabla_j m_{\varphi,t}(c,d)$ for their derivatives.

We will omit the arguments (c,d) from the functions $\overline{\psi}_t(\cdot)$, $\overline{\varphi}_t(\cdot)$, $m_{\psi,t}(\cdot)$, $m_{\varphi,t}(\cdot)$ and their derivatives when they are evaluated at true values $c = \theta$ and $d = \delta_t$, i.e., we use the shorthand notation $\overline{\psi}_t = \overline{\psi}_t(\theta, \delta_t)$, $m_{\psi,t} = m_{\psi,t}(\theta, \delta_t)$, and so on.

Finally, in what follows, $\|V\|$ is the Euclidean norm of a vector V, and $\|A\|$ and $s_{\min}(A)$ denote the Frobenius norm and smallest singular value of a matrix A, respectively.

Estimator. We study GMM estimation under just identification based on the unconditional counterparts to moments (B.1) and (B.2). Specifically, $\widehat{\theta}$ and $\widehat{\delta}_1, ..., \widehat{\delta}_T$ satisfy the sample moments

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \overline{\psi}_t(\widehat{\theta}, \widehat{\delta}_t) &= 0_{k\times 1}, \\ \overline{\varphi}_t(\widehat{\theta}, \widehat{\delta}_t) &= 0_{k_d \times 1}, \qquad t = 1, \ldots, T. \end{aligned}$$

We will derive our asymptotic approximations using the following assumption regarding the evolution of cross-sectional sample sizes N_t. Specifically, we consider cases where no cross section is too small or too big compared to the others, that is, ruling out extreme forms of unbalancedness.

Assumption B.1 (Unbalancedness). *There is a sequence N and two fixed constants $\underline{\gamma}, \overline{\gamma}$ such that, for all t, $0 < \underline{\gamma} \le N_t/N \le \overline{\gamma} < \infty$.*

Linear model example. As an illustration, consider the following regression model:

$$y_{it} = \theta Z_t + \delta_t x_{it} + u_{it}, \qquad u_{it} = \lambda_N G_t + \varepsilon_{it}, \tag{B.3}$$

where the aggregate and idiosyncratic disturbances satisfy

$$E\Big[\, G_t \,\Big|\, \{Z_t, \{x_{it}\}_{i\in\mathcal{I}_t}\}_{t=1}^T \Big] = 0, \qquad E\Big[\, \varepsilon_{it} \,\Big|\, \{Z_t, \{x_{jt}\}_{j\in\mathcal{I}_t}\}_{t=1}^T \Big] = 0.$$

The researcher observes micro and macro data $Y_{it} = (y_{it}, x_{it})'$ and Z_t for units $i \in \mathcal{I}_t$ and periods $t = 1, \ldots, T$. Note the presence of a common shock G_t, scaled by the sample-size-dependent constant λ_N, in u_{it}. Our setup nests situations without common shocks (i.e., $\lambda_N = 0$) as well as various degrees of cross-sectional dependence. We will comment on this model of dependence below.

Model (B.3) shares several similarities with the model we focus on in the main text. θ is a common parameter that captures the effect of the aggregate conditions Z_t on the outcome. In our earnings model, we similarly estimate effects of aggregate conditions on earnings persistence or skewness, say. In turn, δ_t are time-varying parameters that are identified from cross-sectional variation. Similarly, our permanent-transitory model contains time-varying parameters indexing the distributions of transitory shocks and initial conditions. However, the stylized model (B.3) we focus on here for illustration is linear and does not contain latent variables, unlike our nonlinear permanent-transitory model.

The moment functions implied by model (B.3) are

$$\psi_t(Y_{it}, Z_t, c, d) = Z_t\,(y_{it} - cZ_t - dx_{it})\,, \qquad \varphi_t(Y_{it}, Z_t, c, d) = x_{it}\,(y_{it} - cZ_t - dx_{it})\,,$$

so that, at true values, by mean independence and iterated expectations,

$$E\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(Y_{it}, Z_t, \theta, \delta_t)\,\middle|\, \mathcal{Z}_T\right] = E\left[\frac{1}{T}\sum_{t=1}^{T} Z_t\,(\lambda_N G_t + \varepsilon_{it})\,\middle|\, \mathcal{Z}_T\right] = 0$$

and

$$E[\,\varphi_t(Y_{it}, Z_t, \theta, \delta_t) \mid \mathcal{Z}_T\,] = E[\,x_{it}\,(\lambda_N G_t + \varepsilon_{it}) \mid \mathcal{Z}_T\,] = 0, \quad t = 1, \ldots, T.$$

There are closed-form expressions for the estimators $\widehat{\theta}$ and $\{\widehat{\delta}_t\}_{t=1}^T$. Let

$$\widetilde{Z}_{it} \equiv Z_t\left(1 - \frac{\sum_{j\in\mathcal{I}_t} x_{jt}}{\sum_{j\in\mathcal{I}_t} x_{jt}^2}\, x_{it}\right).$$

We then have

$$\widehat{\theta} = \frac{\sum_{t=1}^T \sum_{i\in\mathcal{I}_t} \widetilde{Z}_{it} y_{it}}{\sum_{t=1}^T \sum_{i\in\mathcal{I}_t} \widetilde{Z}_{it}^2},$$

$$\widehat{\delta}_t = \frac{\sum_{i\in\mathcal{I}_t} x_{it}\left(y_{it} - \widehat{\theta} Z_t\right)}{\sum_{i\in\mathcal{I}_t} x_{it}^2}, \quad t = 1, \ldots, T.$$

We will model cross-sectional dependence by assuming that

$$\lambda_N = \frac{\overline{\lambda}}{\sqrt{N^\alpha}},$$

for some constants $\overline{\lambda} > 0$ and $\alpha > 0$. This captures a situation in which there may be (weak) cross-sectional dependence across individual observations induced by the factor G_t. This will be important in our discussion of asymptotic normality. In the context of our analysis of income dynamics, G_t may reflect the presence of a survey-specific factor whose influence decays in larger samples.

B.2 Consistency

Assumption B.2 (Conditions for consistency). *The following holds:*

(i) The parameter spaces Θ and Δ are compact.

(ii) There exist some random variables $A_{\psi,t}, B_{\psi,t}, A_{\varphi,t}, B_{\varphi,t}$, $t = 1, ..., T$, such that, for all $c, c' \in \Theta$ and $d, d' \in \Delta$ and almost surely,

$$\|m_{\psi,t}(c', d') - m_{\psi,t}(c, d)\| \leq A_{\psi,t}\|c' - c\| + B_{\psi,t}\|d' - d\|,$$

$$\|m_{\varphi,t}(c', d') - m_{\varphi,t}(c, d)\| \leq A_{\varphi,t}\|c' - c\| + B_{\varphi,t}\|d' - d\|,$$

and

$$\frac{1}{T}\sum_{t=1}^T \|A_{\psi,t}\| = O_p(1),\ \frac{1}{T}\sum_{t=1}^T \|B_{\psi,t}\| = O_p(1),\ \max_{t=1,...,T} \|A_{\varphi,t}\| = O_p(1),\ \max_{t=1,...,T} \|B_{\varphi,t}\| = O_p(1).$$

(iii) For all $c \in \Theta$ and $t = 1, \ldots, T$, there is a unique $\delta_t(c) \in \Delta$ (a random quantity) such that

$$m_{\varphi,t}(c, \delta_t(c)) = 0_{k_d \times 1}.$$

Moreover, for all $\epsilon > 0$, there is $\eta_\epsilon > 0$ such that, with probability approaching one as T, N tend to infinity,

$$\inf_{c\in\Theta} \min_{1\le t\le T} \inf_{\|d-\delta_t(c)\|>\epsilon} \|m_{\varphi,t}(c,d)\| \ge \eta_\epsilon, \qquad \inf_{\|c-\theta\|>\epsilon} \left\| \frac{1}{T}\sum_{t=1}^{T} m_{\psi,t}(c,\delta_t(c)) \right\| \ge \eta_\epsilon.$$

(iv) As $T, N \to \infty$,

$$\sup_{c\in\Theta} \max_{1\le t\le T} \sup_{d\in\Delta} \left\| \overline{\varphi}_t(c,d) - m_{\varphi,t}(c,d) \right\| \xrightarrow{\ \text{p}\ } 0,$$
$$\sup_{c\in\Theta} \frac{1}{T}\sum_{t=1}^{T} \sup_{d\in\Delta} \left\| \overline{\psi}_t(c,d) - m_{\psi,t}(c,d) \right\| \xrightarrow{\ \text{p}\ } 0.$$

Linear model example (continued). In the linear model, we have, for all $t = 1, ..., T$,

$$\begin{aligned} m_{\psi,t}(c,d) &= E[\, Z_t\,(y_{it} - cZ_t - dx_{it}) \mid \mathcal{Z}_T\,] \\ &= E[\, Z_t\,(\theta Z_t + \delta_t x_{it} + \lambda_N G_t + \varepsilon_{it} - cZ_t - dx_{it}) \mid \mathcal{Z}_T\,] \\ &= (\theta - c)Z_t^2 + (\delta_t - d)Z_t E[\, x_{it} \mid \mathcal{Z}_T\,], \end{aligned}$$

and

$$\begin{aligned} m_{\varphi,t}(c,d) &= E[\, x_{it}\,(y_{it} - cZ_t - dx_{it}) \mid \mathcal{Z}_T\,] \\ &= E[\, x_{it}\,(\theta Z_t + \delta_t x_{it} + \lambda_N G_t + \varepsilon_{it} - cZ_t - dx_{it}) \mid \mathcal{Z}_T\,] \\ &= (\theta - c)Z_t E[\, x_{it} \mid \mathcal{Z}_T\,] + (\delta_t - d)E\left[\, x_{it}^2 \,\middle|\, \mathcal{Z}_T\,\right]. \end{aligned}$$

It follows that Assumption B.2(ii) is satisfied provided

$$\max_{t=1,\ldots,T} Z_t^2 = O_p(1), \quad \max_{t=1,\ldots,T} \left(E[\, x_{it} \mid \mathcal{Z}_T\,]\right)^2 = O_p(1), \quad \max_{t=1,\ldots,T} E\left[\, x_{it}^2 \,\middle|\, \mathcal{Z}_T\,\right] = O_p(1).$$

Next, we have

$$\delta_t(c) = \delta_t + (\theta - c)\left(E\left[\, x_{it}^2 \,\middle|\, \mathcal{Z}_T\,\right]\right)^{-1} E[\, x_{it} \mid \mathcal{Z}_T\,]\, Z_t,$$

and we have

$$\|m_{\varphi,t}(c,d)\| = \|m_{\varphi,t}(c,d) - m_{\varphi,t}(c,\delta_t(c))\| = |\delta_t(c) - d| \times E\left[\, x_{it}^2 \,\middle|\, \mathcal{Z}_T\,\right],$$

and

$$\left\| \frac{1}{T} \sum_{t=1}^{T} m_{\psi,t}(c, \delta_t(c)) \right\| = \left\| \frac{1}{T} \sum_{t=1}^{T} (\theta - c) Z_t^2 + (\delta_t - \delta_t(c)) Z_t E[\, x_{it} \mid \mathcal{Z}_T \,] \right\|$$
$$= |\theta - c| \times \frac{1}{T} \sum_{t=1}^{T} Z_t^2 \left(1 - \left(E\left[\, x_{it}^2 \,\middle|\, \mathcal{Z}_T \,\right] \right)^{-1} (E[\, x_{it} \mid \mathcal{Z}_T \,])^2 \right),$$

so Assumption B.2(iii) follows provided

$$\min_{1 \le t \le T} E\left[\, x_{it}^2 \,\middle|\, \mathcal{Z}_T \,\right] \text{ and } \frac{1}{T} \sum_{t=1}^{T} Z_t^2 \left(1 - \left(E\left[\, x_{it}^2 \,\middle|\, \mathcal{Z}_T \,\right] \right)^{-1} (E[\, x_{it} \mid \mathcal{Z}_T \,])^2 \right)$$

are both bounded away from zero with probability approaching one.

Lastly, for all $t = 1, ..., T$ we have

$$\begin{aligned} \overline{\varphi}_t(c, d) - m_{\varphi,t}(c, d) =& (\theta - c) Z_t \left(\frac{1}{N_t} \sum_{i \in \mathcal{I}_t} x_{it} - E[\, x_{it} \mid \mathcal{Z}_T \,] \right) \\ &+ (\delta_t - d) \left(\frac{1}{N_t} \sum_{i \in \mathcal{I}_t} x_{it}^2 - E\left[\, x_{it}^2 \,\middle|\, \mathcal{Z}_T \,\right] \right) \\ &+ \lambda_N \left(\frac{1}{N_t} \sum_{i \in \mathcal{I}_t} x_{it} \right) G_t + \left(\frac{1}{N_t} \sum_{i \in \mathcal{I}_t} x_{it} \varepsilon_{it} \right), \end{aligned}$$

and

$$\begin{aligned} \overline{\psi}_t(c, d) - m_{\psi,t}(c, d) =& (\delta_t - d) Z_t \left(\frac{1}{N_t} \sum_{i \in \mathcal{I}_t} x_{it} - E[\, x_{it} \mid \mathcal{Z}_T \,] \right) \\ &+ \lambda_N (Z_t G_t) + Z_t \left(\frac{1}{N_t} \sum_{i \in \mathcal{I}_t} \varepsilon_{it} \right). \end{aligned}$$

It follows that Assumption B.2(iv) holds provided x_{it}, G_t, Z_t and ε_{it} satisfy appropriate weak dependence conditions. A sufficient condition is that x_{it}, ε_{it} are i.i.d. across i and G_t are i.i.d. across t (with bounded moments).

Proposition 3 (Consistency). *As $T, N \to \infty$, under Assumptions B.1 and B.2,*

$$\widehat{\theta} \xrightarrow{\ \mathrm{p}\ } \theta \quad \text{and} \quad \max_{1 \le t \le T} \|\widehat{\delta}_t - \delta_t\| \xrightarrow{\ \mathrm{p}\ } 0.$$

Proof. For each $c \in \Theta$ and $t = 1, \dots, T$, let $\widehat{\delta}_t(c) \in \Delta$ be any solution to the sample equation

$$\overline{\varphi}_t(c, \widehat{\delta}_t(c)) = 0_{k_d \times 1}.$$

Since $(\widehat{\theta}, \{\widehat{\delta}_t\}_{t=1}^T)$ satisfies the sample moment equations, we can always find a solution such that $\widehat{\delta}_t = \widehat{\delta}_t(\widehat{\theta})$ for all $t = 1, \dots, T$. In what follows, we adopt that specific choice.

Next, we define the profiled population and sample moments

$$M(c) \equiv \frac{1}{T}\sum_{t=1}^{T} m_{\psi,t}(c, \delta_t(c)), \qquad \widehat{M}(c) \equiv \frac{1}{T}\sum_{t=1}^{T} \overline{\psi}_t(c, \widehat{\delta}_t(c)).$$

Since $(\widehat{\theta}, \{\widehat{\delta}_t\}_{t=1}^T)$ solves (B.1) and (B.2), $\widehat{M}(\widehat{\theta}) = 0_{k \times 1}$. We prove consistency in three steps.

Step 1 (Uniform consistency of profiled time effects). Fix $\epsilon > 0$ and for the value $\eta_\epsilon > 0$ in Assumption B.2(iii) define the event

$$\mathcal{A}_{T,N}(\epsilon) \equiv \left\{ \sup_{c \in \Theta} \max_{1 \le t \le T} \sup_{d \in \Delta} \left\| \overline{\varphi}_t(c, d) - m_{\varphi,t}(c, d) \right\| < \eta_\epsilon \right\}.$$

By Assumption B.2(iv), $P\Big(\mathcal{A}_{T,N}(\epsilon)\Big) \longrightarrow 1$ as $T, N \to \infty$. Now, on that same event $\mathcal{A}_{T,N}(\epsilon)$, any root $\widehat{\delta}_t(c)$ of the equation $\overline{\varphi}_t(c, d) = 0_{k_d \times 1}$ must satisfy $\sup_{c \in \Theta} \|\widehat{\delta}_t(c) - \delta_t(c)\| \le \epsilon$. Suppose towards a contradiction that for some c and t we had $\|\widehat{\delta}_t(c) - \delta_t(c)\| > \epsilon$. Then,

$$\left\| m_{\varphi,t}(c, \widehat{\delta}_t(c)) \right\| \ge \eta_\epsilon$$

by the separation condition in Assumption B.2(iii). But since $\overline{\varphi}_t(c, \widehat{\delta}_t(c)) = 0_{k_d \times 1}$, we get

$$\left\| m_{\varphi,t}(c, \widehat{\delta}_t(c)) \right\| = \left\| m_{\varphi,t}(c, \widehat{\delta}_t(c)) - \overline{\varphi}_t(c, \widehat{\delta}_t(c)) \right\| < \eta_\epsilon,$$

a contradiction. Hence, as $T, N \to \infty$,

$$\sup_{c \in \Theta} \max_{1 \le t \le T} \|\widehat{\delta}_t(c) - \delta_t(c)\| \xrightarrow{\ \mathrm{p}\ } 0.$$

Step 2 (Uniform convergence of profiled sample moment). For each $c \in \Theta$, adding and subtracting $m_{\psi,t}(c, \widehat{\delta}_t(c))$,

$$\widehat{M}(c) - M(c) = \frac{1}{T}\sum_{t=1}^{T}\left[\overline{\psi}_t(c, \widehat{\delta}_t(c)) - m_{\psi,t}(c, \widehat{\delta}_t(c))\right] + \frac{1}{T}\sum_{t=1}^{T}\left[m_{\psi,t}(c, \widehat{\delta}_t(c)) - m_{\psi,t}(c, \delta_t(c))\right].$$

Consequently,

$$\begin{aligned}\sup_{c\in\Theta}\|\widehat{M}(c) - M(c)\| \leq \sup_{c\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\sup_{d\in\Delta}\left\|\overline{\psi}_t(c,d) - m_{\psi,t}(c,d)\right\| \\ + \sup_{c\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\left\|m_{\psi,t}(c, \widehat{\delta}_t(c)) - m_{\psi,t}(c, \delta_t(c))\right\|.\end{aligned}$$

The first term is $o_p(1)$ by Assumption B.2(iv), while the second term is $o_p(1)$ by Assumption B.2(ii) and Step 1. Hence, as $T, N \to \infty$,

$$\sup_{c\in\Theta}\|\widehat{M}(c) - M(c)\| \xrightarrow{\text{p}} 0.$$

Step 3 (Consistency of $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\delta}}_t$). By definition of $\delta_t(\theta)$ and the population moment restrictions at true values θ and $\{\delta_t\}_{t=1}^{T}$,

$$M(\theta) = \frac{1}{T}\sum_{t=1}^{T} m_{\psi,t}(\theta, \delta_t(\theta)) = \frac{1}{T}\sum_{t=1}^{T} m_{\psi,t}(\theta, \delta_t) = 0_{k\times 1}.$$

Fix $\epsilon > 0$ and for the $\eta_\epsilon > 0$ in Assumption B.2(iii) consider the event

$$\mathcal{B}_{T,N}(\epsilon) \equiv \left\{\sup_{c\in\Theta}\left\|\widehat{M}(c) - M(c)\right\| < \eta_\epsilon\right\}.$$

By Step 2, $P\Big(\mathcal{B}_{T,N}(\epsilon)\Big) \longrightarrow 1$. Mimicking the argument in Step 1, on the event $\mathcal{B}_{T,N}(\epsilon)$, any solution $\widehat{\theta}$ to $\widehat{M}(\widehat{\theta}) = 0_{k\times 1}$ must satisfy $\|\widehat{\theta} - \theta\| \leq \epsilon$. Suppose towards a contradiction that $\|\widehat{\theta} - \theta\| > \epsilon$. Then,

$$\left\|M(\widehat{\theta})\right\| \geq \eta_\epsilon,$$

but because $\widehat{M}(\widehat{\theta}) = 0_{k\times 1}$,

$$\left\|M(\widehat{\theta})\right\| = \left\|M(\widehat{\theta}) - \widehat{M}(\widehat{\theta})\right\| < \eta_\epsilon,$$

which contradicts the separation condition in Assumption B.2(iii). Thus, as $T, N \to \infty$,

$$\widehat{\theta} \xrightarrow{\text{p}} \theta.$$

Finally,

$$\max_{1 \le t \le T} \left\| \widehat{\delta}_t - \delta_t \right\| \le \max_{1 \le t \le T} \left\| \widehat{\delta}_t(\widehat{\theta}) - \delta_t(\widehat{\theta}) \right\| + \max_{1 \le t \le T} \left\| \delta_t(\widehat{\theta}) - \delta_t(\theta) \right\|.$$

The first term is $o_p(1)$ by Step 1 at $c = \widehat{\theta}$. For the second term, fix $\epsilon > 0$ and let $\eta_\epsilon > 0$ be as in Assumption B.2(iii). Suppose that $\|\delta_t(\widehat{\theta}) - \delta_t(\theta)\| > \epsilon$ for some t. Since $m_{\varphi,t}(\widehat{\theta}, \delta_t(\widehat{\theta})) = 0_{k_d \times 1}$, the separation condition in Assumption B.2(iii) applied at $c = \widehat{\theta}$ yields

$$\left\| m_{\varphi,t}(\widehat{\theta}, \delta_t(\theta)) \right\| \ge \eta_\epsilon.$$

On the other hand, since $m_{\varphi,t}(\theta, \delta_t(\theta)) = 0_{k_d \times 1}$, Assumption B.2(ii) gives

$$\left\| m_{\varphi,t}(\widehat{\theta}, \delta_t(\theta)) \right\| = \left\| m_{\varphi,t}(\widehat{\theta}, \delta_t(\theta)) - m_{\varphi,t}(\theta, \delta_t(\theta)) \right\| \le A_{\varphi,t} \|\widehat{\theta} - \theta\| \le \left(\max_{1 \le t \le T} A_{\varphi,t} \right) \|\widehat{\theta} - \theta\|.$$

Hence, $(\max_{1 \le t \le T} A_{\varphi,t}) \|\widehat{\theta} - \theta\| \ge \eta_\epsilon$. Now, since $\max_{1 \le t \le T} A_{\varphi,t} = O_p(1)$ by Assumption B.2(ii) and $\widehat{\theta} \xrightarrow{\text{p}} \theta$ as shown above, the probability that this happens tends to zero as $T, N \to \infty$. Hence, as $T, N \to \infty$,

$$\max_{1 \le t \le T} \left\| \delta_t(\widehat{\theta}) - \delta_t(\theta) \right\| \xrightarrow{\text{p}} 0,$$

and therefore

$$\max_{1 \le t \le T} \left\| \widehat{\delta}_t - \delta_t \right\| \xrightarrow{\text{p}} 0,$$

and the proof is complete. □

B.3 Asymptotic normality

The constant α plays a key role in our discussion of asymptotic normality. It is a device to introduce dependence across units in a high-level manner. In essence, $\alpha \ge 1$ corresponds to weak dependence (for example, units being i.i.d. conditional on $\mathcal{Z}_T$), while $\alpha \to 0$ captures a situation where there is strong dependence because of the presence of common factors.

Letting

$$\underline{\alpha} = \min(\alpha, 1),$$

dependence across units reduces the effective cross-section sample size to $N^{\underline{\alpha}}$. As a result, we require larger cross-sections to deal with the estimation noise in the time effects δ_t. In fact, we will require that $T/N^{\widetilde{\alpha}} \to 0$ where $0 < \widetilde{\alpha} < \underline{\alpha}$. Under this condition, $\widehat{\theta}$ will be $\sqrt{TN^{\underline{\alpha}}}$–consistent and asymptotically normal (in Proposition 4 below). A special case covered by our setup is when observations are independent across units, in which case $\underline{\alpha} = 1$ and $\widehat{\theta}$ is $\sqrt{TN}$–consistent.

Assumption B.3 (Conditions for asymptotic normality). *The following holds:*

(i) The true parameters $\theta \in \Theta$ and $\delta_t \in \Delta$ (for all t) are interior points.

(ii) The functions $\psi_t(Y, Z, c, d)$ and $\varphi_t(Y, Z, c, d)$ are twice continuously differentiable with respect to (c, d) on $\Theta \times \Delta$ for each Y and Z and all t.

(iii) As $T, N \to \infty$ and for $j = \theta, \delta$,

$$\sup_{c\in\Theta} \max_{1\leq t\leq T} \sup_{d\in\Delta} \left\| \nabla_j \overline{\varphi}_t(c, d) - \nabla_j m_{\varphi,t}(c, d) \right\| \xrightarrow{\;p\;} 0,$$

$$\sup_{c\in\Theta} \frac{1}{T} \sum_{t=1}^{T} \sup_{d\in\Delta} \left\| \nabla_j \overline{\psi}_t(c, d) - \nabla_j m_{\psi,t}(c, d) \right\| \xrightarrow{\;p\;} 0.$$

(iv) There exists a neighborhood $\mathcal{N} \subseteq \Theta \times \Delta$ of $\{(\theta, \delta_t)\}_{t=1}^{T}$ and a constant $\underline{s} > 0$ such that, for all sufficiently large T, N,

$$\inf_{(c,d)\in\mathcal{N}} \min_{1\leq t\leq T} s_{\min}\Big(\nabla_\delta m_{\varphi,t}(c, d)\Big) \geq \underline{s},$$

Moreover, using ∇^2 for vectorized second derivatives,

$$\sup_{(c,d)\in\mathcal{N}} \max_{1\leq t\leq T} \|\nabla^2 \overline{\varphi}_t(c, d)\| = O_p(1)\,, \qquad \sup_{(c,d)\in\mathcal{N}} \frac{1}{T} \sum_{t=1}^{T} \|\nabla^2 \overline{\psi}_t(c, d)\| = O_p(1)\,,$$

and there is $0 < \nu < \underline{\alpha}$ such that

$$\max_{1\leq t\leq T} \|\overline{\varphi}_t\| = O_p\Big(N^{-\frac{\underline{\alpha}-\nu}{2}}\Big)\,.$$

(v) For an invertible matrix D_0,

$$\frac{1}{T} \sum_{t=1}^{T} \Big(\nabla_\theta m_{\psi,t} - \nabla_\delta m_{\psi,t} \left(\nabla_\delta m_{\varphi,t}\right)^{-1} \nabla_\theta m_{\varphi,t} \Big) \xrightarrow{\;p\;} D_0.$$

(vi) For a positive definite matrix V_0,

$$\sqrt{TN^{\underline{\alpha}}} \times \frac{1}{T}\sum_{t=1}^{T}\left(\overline{\psi}_t - \nabla_\delta m_{\psi,t}\left(\nabla_\delta m_{\varphi,t}\right)^{-1}\overline{\varphi}_t\right) \xrightarrow{\ \mathrm{d}\ } N(0_{k\times 1}, V_0)\,.$$

Linear model example (continued). In the linear model, we have for all $a > 0$

$$\begin{aligned}\Pr\left(\max_{1\le t\le T}\left|\nabla_\theta\overline{\varphi}_t(c,d) - \nabla_\theta m_{\varphi,t}(c,d)\right| > a\right) &= \Pr\left(\max_{1\le t\le T}\left|Z_t\left(\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}x_{it} - E[\,x_{it}\mid\mathcal{Z}_T\,]\right)\right| > a\right)\\ &\le T\max_{1\le t\le T}\Pr\left(\left|Z_t\left(\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}x_{it} - E[\,x_{it}\mid\mathcal{Z}_T\,]\right)\right| > a\right),\end{aligned}$$

which can be shown to tend to zero if $Z_t = O_p(1)$ and the x_{it} satisfy weak dependence conditions across i as well as suitable tail conditions, under the relative rate condition $T/N^{\widetilde{\alpha}} \to 0$ for $0 < \widetilde{\alpha} < \underline{\alpha}$. In this way, one can verify Assumption B.3(iii).

Next, for Assumption B.3(iv) we have, for all $a > 0$,

$$\begin{aligned}&\Pr\left(\max_{1\le t\le T}\|\overline{\varphi}_t\| > aN^{-\frac{\underline{\alpha}-\nu}{2}}\right) \le T\max_{1\le t\le T}\Pr\left(\|\overline{\varphi}_t\| > aN^{-\frac{\underline{\alpha}-\nu}{2}}\right)\\ &= T\max_{1\le t\le T}\Pr\left(\left|\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}x_{it}(\lambda_N G_t + \varepsilon_{it})\right| > aN^{-\frac{\underline{\alpha}-\nu}{2}}\right)\\ &\le T\max_{1\le t\le T}\left(\Pr\left(\left|\overline{\lambda}G_t\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}x_{it}\right| > \frac{1}{2}aN^{\frac{\alpha-\underline{\alpha}+\nu}{2}}\right) + \Pr\left(\left|\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}x_{it}\varepsilon_{it}\right| > \frac{1}{2}aN^{-\frac{\underline{\alpha}-\nu}{2}}\right)\right),\end{aligned}$$

which tends to zero provided $G_t\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}x_{it}$ has sufficiently thin tails and $x_{it}\varepsilon_{it}$ is weakly dependent across i with suitable tail conditions, under the relative rate condition $T/N^{\widetilde{\alpha}} \to 0$ for $0 < \widetilde{\alpha} < \underline{\alpha}$.

Next, to verify Assumption B.3(v), note that

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\left(\nabla_\theta m_{\psi,t} - \nabla_\delta m_{\psi,t}\left(\nabla_\delta m_{\varphi,t}\right)^{-1}\nabla_\theta m_{\varphi,t}\right)\\ &= \frac{1}{T}\sum_{t=1}^{T}Z_t^2\left(1 - \left(E\left[\,x_{it}^2\;\middle|\;\mathcal{Z}_T\,\right]\right)^{-1}\left(E[\,x_{it}\mid\mathcal{Z}_T\,]\right)^2\right),\end{aligned}$$

admits a non-singular probability limit D_0 under weak dependence conditions in the time series.

Lastly, to verify Assumption B.3(vi), note that

$$\sqrt{TN^{\underline{\alpha}}} \times \frac{1}{T}\sum_{t=1}^{T}\left(\overline{\psi}_t - \nabla_\delta m_{\psi,t}\left(\nabla_\delta m_{\varphi,t}\right)^{-1}\overline{\varphi}_t\right)$$

$$= \sqrt{TN^{\underline{\alpha}}} \times \frac{1}{T}\sum_{t=1}^{T}\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}\left(Z_t(\lambda_N G_t + \varepsilon_{it}) - Z_t E[\, x_{it} \mid \mathcal{Z}_T\,]\left(E\left[\, x_{it}^2 \;\middle|\; \mathcal{Z}_T\,\right]\right)^{-1} x_{it}(\lambda_N G_t + \varepsilon_{it})\right)$$

$$= \sqrt{N^{\underline{\alpha}-\alpha}}\sqrt{T} \times \frac{1}{T}\sum_{t=1}^{T}\overline{\lambda} Z_t G_t\left(1 - E[\, x_{it} \mid \mathcal{Z}_T\,]\left(E\left[\, x_{it}^2 \;\middle|\; \mathcal{Z}_T\,\right]\right)^{-1}\frac{1}{N_t}\sum_{i\in\mathcal{I}_t} x_{it}\right)$$
$$+ \sqrt{TN^{\underline{\alpha}}} \times \frac{1}{T}\sum_{t=1}^{T}\left(Z_t\frac{1}{N_t}\sum_{i\in\mathcal{I}_t}\varepsilon_{it} - Z_t E[\, x_{it} \mid \mathcal{Z}_T\,]\left(E\left[\, x_{it}^2 \;\middle|\; \mathcal{Z}_T\,\right]\right)^{-1}\frac{1}{N_t}\sum_{i\in\mathcal{I}_t} x_{it}\varepsilon_{it}\right).$$

The first term is normally distributed as $T \to \infty$ under weak dependence conditions in the time series (if $\alpha \leq 1$, otherwise it converges to zero). The second term is $o_p(1)$ when $\alpha < 1$, and contributes an additional normally distributed component when $\alpha \geq 1$, provided ε_{it}, x_{it} are weakly dependent across i and t.

Proposition 4 (Asymptotic normality). *Suppose Assumptions B.1, B.2 and B.3 hold. Then, as $T, N \to \infty$ with $T/N^{\underline{\alpha}-2\nu} \to 0$,*

$$\sqrt{TN^{\underline{\alpha}}}(\widehat{\theta} - \theta) \xrightarrow{\ \mathrm{d}\ } N\left(0_{k\times 1}, D_0^{-1}V_0(D_0^{-1})'\right).$$

Proof. For notational convenience, define the matrices

$$A_t \equiv \nabla_\delta m_{\varphi,t}, \quad B_t \equiv \nabla_\theta m_{\varphi,t}, \quad C_t \equiv \nabla_\theta m_{\psi,t}, \quad H_t \equiv \nabla_\delta m_{\psi,t},$$

all evaluated at true values (θ, δ_t), and let

$$K_t \equiv H_t A_t^{-1}, \qquad J_t \equiv C_t - K_t B_t.$$

For $j = \theta, \delta$, denote by $\nabla_j \overline{\psi}_t^{\mathrm{mv}}$ and $\nabla_j \overline{\varphi}_t^{\mathrm{mv}}$ derivative matrices evaluated at mean values. Let

$$A_t^{\mathrm{mv}} \equiv \nabla_\delta \overline{\varphi}_t^{\mathrm{mv}}, \quad B_t^{\mathrm{mv}} \equiv \nabla_\theta \overline{\varphi}_t^{\mathrm{mv}}, \quad C_t^{\mathrm{mv}} \equiv \nabla_\theta \overline{\psi}_t^{\mathrm{mv}}, \quad H_t^{\mathrm{mv}} \equiv \nabla_\delta \overline{\psi}_t^{\mathrm{mv}},$$

and define

$$K_t^{\mathrm{mv}} \equiv H_t^{\mathrm{mv}}(A_t^{\mathrm{mv}})^{-1}, \qquad J_t^{\mathrm{mv}} \equiv C_t^{\mathrm{mv}} - K_t^{\mathrm{mv}} B_t^{\mathrm{mv}}.$$

We prove the result in four steps.

Step 1 (Local behavior of mean-value Jacobians). By Proposition 3,

$$\widehat{\theta} \xrightarrow{\ \mathrm{p}\ } \theta, \qquad \max_{1\le t\le T} \left\| \widehat{\delta}_t - \delta_t \right\| \xrightarrow{\ \mathrm{p}\ } 0.$$

Hence, with probability approaching one, all mean-value points used to define the matrices $A_t^{\mathrm{mv}}, B_t^{\mathrm{mv}}, C_t^{\mathrm{mv}}, H_t^{\mathrm{mv}}$ lie in the neighborhood $\mathcal{N}$ from Assumption B.3(iv).

Next, using Assumption B.3(iii), consistency, and the local equicontinuity property implied by the second part of Assumption B.3(iv),

$$\max_{1\le t\le T} \|A_t^{\mathrm{mv}} - A_t\| \xrightarrow{\ \mathrm{p}\ } 0, \qquad \max_{1\le t\le T} \|B_t^{\mathrm{mv}} - B_t\| \xrightarrow{\ \mathrm{p}\ } 0,$$
$$\frac{1}{T}\sum_{t=1}^{T} \|C_t^{\mathrm{mv}} - C_t\| \xrightarrow{\ \mathrm{p}\ } 0, \qquad \frac{1}{T}\sum_{t=1}^{T} \|H_t^{\mathrm{mv}} - H_t\| \xrightarrow{\ \mathrm{p}\ } 0.$$

Also, by Assumption B.3(iv), A_t is uniformly nonsingular in a neighborhood of the truth, so

$$\max_{1\le t\le T} \left\| (A_t^{\mathrm{mv}})^{-1} \right\| = O_p(1)\,, \qquad \max_{1\le t\le T} \left\| A_t^{-1} \right\| = O_p(1)\,.$$

Therefore,

$$\frac{1}{T}\sum_{t=1}^{T} \|K_t^{\mathrm{mv}} - K_t\| \xrightarrow{\ \mathrm{p}\ } 0, \qquad \frac{1}{T}\sum_{t=1}^{T} \|J_t^{\mathrm{mv}} - J_t\| \xrightarrow{\ \mathrm{p}\ } 0.$$

Moreover, the local Lipschitz property from Assumption B.3(iv) yields

$$\frac{1}{T}\sum_{t=1}^{T} \|K_t^{\mathrm{mv}} - K_t\| = O_p\!\left(\left\| \widehat{\theta} - \theta \right\| + \max_{1\le t\le T} \left\| \widehat{\delta}_t - \delta_t \right\| \right). \tag{B.4}$$

Step 2 (Profiled linearization). A first-order mean-value expansion of the estimating equations around (θ, δ_t) delivers

$$0_{k_d\times 1} = \overline{\varphi}_t + B_t^{\mathrm{mv}}(\widehat{\theta} - \theta) + A_t^{\mathrm{mv}}(\widehat{\delta}_t - \delta_t), \qquad t = 1, \ldots, T, \tag{B.5}$$
$$0_{k\times 1} = \frac{1}{T}\sum_{t=1}^{T} \left[\overline{\psi}_t + C_t^{\mathrm{mv}}(\widehat{\theta} - \theta) + H_t^{\mathrm{mv}}(\widehat{\delta}_t - \delta_t) \right]. \tag{B.6}$$

Solving (B.5) for $\widehat{\delta}_t - \delta_t$,

$$\widehat{\delta}_t - \delta_t = -(A_t^{\mathrm{mv}})^{-1}\overline{\varphi}_t - (A_t^{\mathrm{mv}})^{-1}B_t^{\mathrm{mv}}(\widehat{\theta} - \theta). \tag{B.7}$$

Substituting (B.7) into (B.6), and solving for $\widehat{\theta} - \theta$,

$$\widehat{\theta} - \theta = -(\overline{J}_{T,N}^{\mathrm{mv}})^{-1}\overline{S}_{T,N}^{\mathrm{mv}}. \tag{B.8}$$

where

$$\overline{S}_{T,N}^{\mathrm{mv}} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\overline{\psi}_t - K_t^{\mathrm{mv}}\overline{\varphi}_t\right), \qquad \overline{J}_{T,N}^{\mathrm{mv}} \equiv \frac{1}{T}\sum_{t=1}^{T} J_t^{\mathrm{mv}}.$$

Step 3 (Rates of convergence). Define the infeasible profiled score

$$\overline{S}_{T,N} \equiv \frac{1}{T}\sum_{t=1}^{T}\left(\overline{\psi}_t - K_t\overline{\varphi}_t\right).$$

By Assumption B.3(vi),

$$\overline{S}_{T,N} = O_p\left(\frac{1}{\sqrt{TN^{\underline{\alpha}}}}\right). \tag{B.9}$$

Also, by the triangle and Hölder's inequality,

$$\left\|\overline{S}_{T,N}^{\mathrm{mv}} - \overline{S}_{T,N}\right\| \leq \frac{1}{T}\sum_{t=1}^{T}\|(K_t^{\mathrm{mv}} - K_t)\overline{\varphi}_t\| \leq \max_{1\leq t\leq T}\|\overline{\varphi}_t\| \cdot \left(\frac{1}{T}\sum_{t=1}^{T}\|K_t^{\mathrm{mv}} - K_t\|\right). \tag{B.10}$$

Next, from (B.7),

$$\begin{aligned}\max_{1\leq t\leq T}\left\|\widehat{\delta}_t - \delta_t\right\| &\leq \max_{1\leq t\leq T}\left\|(A_t^{\mathrm{mv}})^{-1}\right\|\max_{1\leq t\leq T}\|\overline{\varphi}_t\| + \max_{1\leq t\leq T}\left\|(A_t^{\mathrm{mv}})^{-1}B_t^{\mathrm{mv}}\right\|\left\|\widehat{\theta} - \theta\right\| \\ &= O_p\left(N^{-\frac{\underline{\alpha}-\nu}{2}}\right) + O_p\left(\left\|\widehat{\theta} - \theta\right\|\right).\end{aligned} \tag{B.11}$$

Combining (B.4), (B.10), (B.11), and Assumption B.3(iv),

$$\begin{aligned}\left\|\overline{S}_{T,N}^{\mathrm{mv}} - \overline{S}_{T,N}\right\| &= O_p\left(N^{-\frac{\underline{\alpha}-\nu}{2}}\left[\left\|\widehat{\theta} - \theta\right\| + \max_{1\leq t\leq T}\left\|\widehat{\delta}_t - \delta_t\right\|\right]\right) \\ &= O_p\left(N^{-\frac{\underline{\alpha}-\nu}{2}}\left\|\widehat{\theta} - \theta\right\|\right) + O_p\left(N^{-(\underline{\alpha}-\nu)}\right).\end{aligned} \tag{B.12}$$

By Step 1, Assumption B.3(v) and the triangle inequality,

$$\overline{J}^{\mathrm{mv}}_{T,N} \xrightarrow{\ \mathrm{p}\ } D_0,$$

so $(\overline{J}^{\mathrm{mv}}_{T,N})^{-1} = O_p(1)$. Then (B.8), (B.9), and (B.12) imply

$$\left\|\widehat{\theta} - \theta\right\| = O_p\left(\frac{1}{\sqrt{TN^{\underline{\alpha}}}}\right) + O_p\left(N^{-\frac{\underline{\alpha}-\nu}{2}}\|\widehat{\theta} - \theta\|\right) + O_p\left(N^{-(\underline{\alpha}-\nu)}\right).$$

The first term is coming from the variability of the infeasible profiled score, the second from the effect of estimating θ itself and the third from the estimation error in the time effects δ_t. Since $\nu < \underline{\alpha}$, $N^{-\frac{\underline{\alpha}-\nu}{2}} = o(1)$, and since $T/N^{\underline{\alpha}-2\nu} \to 0$, $\sqrt{TN^{\underline{\alpha}}}N^{-(\underline{\alpha}-\nu)} = o(1)$. Therefore,

$$\widehat{\theta} - \theta = O_p\left(\frac{1}{\sqrt{TN^{\underline{\alpha}}}}\right). \tag{B.13}$$

Substituting (B.13) into (B.11), we also obtain

$$\max_{1\le t\le T} \|\widehat{\delta}_t - \delta_t\| = O_p\left(N^{-\frac{\underline{\alpha}-\nu}{2}}\right). \tag{B.14}$$

Step 4 (Asymptotic linear representation and limit). From (B.12), (B.13), and (B.14),

$$\begin{aligned}\sqrt{TN^{\underline{\alpha}}}\|\overline{S}^{\mathrm{mv}}_{T,N} - \overline{S}_{T,N}\| &= O_p\left(\sqrt{TN^{\underline{\alpha}}}N^{-(\underline{\alpha}-\nu)}\right) + O_p\left(\sqrt{TN^{\underline{\alpha}}}N^{-\frac{\underline{\alpha}-\nu}{2}}\|\widehat{\theta} - \theta\|\right) \\ &= O_p\left(\sqrt{\frac{T}{N^{\underline{\alpha}-2\nu}}}\right) + O_p\left(\frac{1}{N^{\frac{\underline{\alpha}-\nu}{2}}}\right) = o_p(1).\end{aligned}$$

Hence,

$$\sqrt{TN^{\underline{\alpha}}}\,\overline{S}^{\mathrm{mv}}_{T,N} = \sqrt{TN^{\underline{\alpha}}}\,\overline{S}_{T,N} + o_p(1). \tag{B.15}$$

By Step 1,

$$\overline{J}^{\mathrm{mv}}_{T,N} \xrightarrow{\ \mathrm{p}\ } D_0. \tag{B.16}$$

Multiplying (B.8) by $\sqrt{TN^{\underline{\alpha}}}$ and using (B.15)–(B.16),

$$\sqrt{TN^{\underline{\alpha}}}(\widehat{\theta}-\theta) = -(\overline{J}^{\text{mv}}_{T,N})^{-1}\sqrt{TN^{\underline{\alpha}}}\,\overline{S}^{\text{mv}}_{T,N} = -D_0^{-1}\sqrt{TN^{\underline{\alpha}}}\frac{1}{T}\sum_{t=1}^{T}\Big(\overline{\psi}_t - K_t\overline{\varphi}_t\Big) + o_p(1)\,.$$

Finally, by Assumption B.3(vi) and Slutsky's theorem,

$$\sqrt{TN^{\underline{\alpha}}}(\widehat{\theta}-\theta) \xrightarrow{\ \text{d}\ } N\Big(0_{k\times 1}, D_0^{-1}V_0(D_0^{-1})'\Big)\,,$$

as claimed. □

B.4 Inference

In the absence of cross-sectional dependence (i.e., $\alpha \to \infty$ and $\underline{\alpha} = 1$), Proposition 4 justifies the use of conventional GMM inference. In practice, inference can be based on empirical estimates of the matrices D_0 and V_0 and normal critical values. Alternatively, one can use versions of the bootstrap to construct standard errors and confidence intervals, the validity of which has been formally established in a variety of related settings (e.g., Horowitz (2001)). In our application, we rely on a parametric bootstrap method.

C Estimation

Below we provide additional information about the estimation strategy outlined in Section 4. We focus on the case where Z_t is estimated based on aggregate measures W_t, which is the case that we rely on in practice. Sections C.1 and C.2 spell out the moments implied by our model. Sections C.3 and C.4 describe various aspects of the stochastic EM algorithm. Finally, Section C.5 discusses our bootstrap approach to inference.

C.1 Infeasible moment conditions

Our model implies two types of infeasible complete-data moments for θ and δ_t. We specify them explicitly for the parameter vector θ which contains $\text{vec}\{\Theta(\overline{u}_\ell)\}$ for $\ell = 1,\ldots,L$ together with θ_{lo} and θ_{up} below.

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

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

which is a rewrite of (8) in a more compact form.

Following ABB we model tails as conditionally exponential, such as

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

where $\psi_{\text{lo}}, \psi_{\text{up}}$ are vectors of known basis functions and $\theta_{\text{lo}}, \theta_{\text{up}}$ are unknown parameters included in θ.

For these tail parameters we use the orthogonality conditions from exponential regression,

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

We similarly derive moment functions for quantile parameters in the initial condition η_{i,t_0} (including tail parameters based on an exponential specification), which take similar forms to $m_{it}^{\text{qr}}(\theta, u)$, $m_{it}^{\text{lo}}(\theta)$, and $m_{it}^{\text{up}}(\theta)$, and we do not report here for brevity.

Lastly, in our main specification we model the transitory shocks ε_{it} as following a finite mixture of Gaussians with L_ε components. The implied infeasible moment function based on the score equations is

$$m_{it,\varepsilon}(\delta_t) = \sum_{\tau=t}^{t+S-1} \frac{\partial \ln f_\varepsilon}{\partial \delta_{t,\tau,\varepsilon}} \left(Y_{i\tau} - \eta_{i\tau}; \delta_{t,\tau,\varepsilon}\right),$$

where $\delta_{t,\tau,\varepsilon}$ contains all the parameters in the Gaussian mixture density f_ε for period τ in subpanel t: means, variances, and probabilities of the components. Our main estimates correspond to $L_\varepsilon = 4$, but we have performed robustness checks for other values. In addition, we have estimated specifications where transitory shocks ε_{it} follow a quantile specification similar to the one for η_{it}, also allowing for dependence of the distribution on age, and found similar results to the ones reported in the text (see the working paper version Almuzara et al., 2025).

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

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

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 (corresponding to the initial conditions) and score equations (corresponding to transitory shocks). At the true value δ_t,

$$E\Big[\, m_\delta\left(\delta_t; \bar{y}_{it}^S, \bar{\eta}_{it}^S\right)\Big] = 0_{\dim(\delta_t)\times 1}. \tag{C.2}$$

C.2 Feasible moment conditions

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

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

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

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

By (C.1)-(C.2) and the law of iterated expectations,

$$E\Big[\,\overline{m}_\theta\left(\theta;\theta,\delta_t,\bar{y}_{it}^S,\bar{Z}_t^S\right)\Big]=0,\qquad E\Big[\,\overline{m}_\delta\left(\delta_t;\theta,\delta_t,\bar{y}_{it}^S,\bar{Z}_t^S\right)\Big]=0,\tag{C.3}$$

which indicates the partial-data moments (C.3) provide valid restrictions on θ and δ_t.

We now describe how we construct feasible moments by conditioning on the aggregate measurements W_t. In the approach we implement, we rely on the measurement system described at the end of Section 4 to estimate a process for Z_t based on a sequence of aggregate measures W_t, $t=1,...,T+S$. We denote as λ the vector containing the parameters governing the joint process of Z_t and W_t.

Letting $\overline{Y}_t^S=\{\bar{y}_{it}^S\}_{i\in\mathcal{I}_t}$ and $\overline{W}=\{W_t\}_{t=1}^{T+S}$, we then define the aggregated observed-data moments,

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

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

$$E\Big[\,\overline{M}_\theta\left(\theta;\theta,\delta_t,\lambda,\overline{Y}_t^S,\overline{W}\right)\Big]=0,\qquad E\Big[\,\overline{M}_\delta\left(\delta_t;\theta,\delta_t,\lambda,\overline{Y}_t^S,\overline{W}\right)\Big]=0.\tag{C.4}$$

The functions $\overline{M}_\theta$ and $\overline{M}_\delta$ give restrictions on θ and δ_t (given λ) that only depend on observables.[3] We rely on the sample counterpart of the moments in (C.4) for estimation.

C.3 Stochastic EM implementation

The main challenge in exploiting (C.4) for estimation is to integrate the primitive moments against the pseudo posteriors of η_{it} and Z_t. We follow Arellano and Bonhomme (2016)

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

and adopt a simulation-based approach. Let $\widehat{\lambda}$ be parameter estimates obtained from the aggregate measures $\overline{W}$ alone (e.g., maximum likelihood). We rely on the following stochastic EM algorithm to estimate θ and $\{\delta_t\}_{t=1}^T$, which iterates between simulation smoothing of macro and micro latent variables and running quantile and exponential regressions.

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

1) Pseudo-Stochastic E step:

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

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

2) Pseudo M step:

(i) update the parameters to $\widehat{\theta}^{(j)}$ and $\{\widehat{\delta}_t^{(j)}\}_{t=1}^T$ by quantile and exponential regressions treating $\left\{\{\{\eta_{i,t+s}(j), y_{i,t+s}, x_{i,t+s}, Z_{t+s}(j)\}_{s=0}^{S-1}\}_{i\in\mathcal{I}_t}\right\}_{t=1}^T$ as data, based on (C.4).

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

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

C.4 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:

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

where $V_t, \nu_{1t}, \ldots, \nu_{n_W,t}$ are i.i.d. standard normal and mutually independent. Moreover, as stated in the text, we normalize the entry of Λ that corresponds to GDP to unity so that Z_t is measured in units of GDP per capita relative to its low-frequency trend. Note that this process is first-order Markov in companion form when stacking (Z_t, Z_{t-1}).

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

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

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

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

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

Micro posterior: Sequential Monte Carlo. Step 1(ii) in Algorithm 1 requires sampling, for each i and t, the distribution of $\{\eta_{i,t+s}\}_{s=0}^{S-1}$ conditional on $\{y_{i,t+s}, x_{i,t+s}, Z_{t+s}\}_{s=0}^{S-1}$

taking Q_η, $Q_{\varepsilon,t}$ and $Q_{\text{init},t}$ (evaluated at certain parameter values θ, $\delta_{\varepsilon,t}$ and $\delta_{\text{init},t}$) as given. We do so by Sequential Monte Carlo.[4]

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

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

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

$(s=0)$

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

$$w^k_{it} = \frac{f_{\text{init},t}(\eta^k_{it}|x_{it}) \cdot f_{\varepsilon,t}(y_{it} - \eta^k_{it}|x_{it})}{q_{\text{init},t}(\eta^k_{it}|y_{it}, x_{it})}.$$

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

$(s>0)$

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

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

[4]See Creal (2012) for a review of Sequential Monte Carlo methods and Arellano et al. (2023) for an application to models of income and consumption.

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

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

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

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

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

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

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

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

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

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

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.5 Bootstrap approach

Unit overlap. 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. The 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:

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

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.

Cross-sectional dependence. In a robustness exercise, we compute measures of statistical uncertainty that account for the presence of cross-sectional dependence, which we model as follows. Let $G_t = (G_{\eta,t}, G_{\varepsilon,t}, G_{\text{init},t})'$ where entries are i.i.d. uniformly distributed on $(0,1)$ and mutually independent. Then, we assume the micro-level errors in our model are

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

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

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 $\{G_{\eta,t}, G_{\varepsilon,t}, G_{\text{init},t}\}_{t=1}^{T}$. [This step can be skipped in the absence of cross-sectional dependence]

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

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

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

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

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

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

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

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

C.6 Empirical specification and functional form

For Q_η, we model ψ as a third-order polynomial in η combined with a second-order polynomial in x, and we construct φ as a restricted second-order polynomial in $(\widetilde{Z}, Z)$: we exclude the linear term from interactions between η and x, and include the quadratic term only in the intercept. We set ψ_{lo} and ψ_{up} to second-order polynomials in η, x, $\widetilde{Z}$ and Z without interactions. The rank-space grid size is set to $L = 11$.

We allow the quantile and tail parameters of $Q_{\text{init},t}$ to depend on time in an unrestricted way. In addition, we include a time-invariant second-order polynomial in age x. We specify $Q_{\varepsilon,t}$ as a 4-component Gaussian mixture where all mean, variance and probability parameters vary over time in an unrestricted way. For estimation, we run Algorithm 1 for 500 iterations, averaging the last 100 of them.[5] Both the draws in the E-step and the parameter sequences produced by the algorithm appear to have converged after the first 100 iterations in all datasets and specifications.

D Model fit assessment

Figure D.1 compares a selection of data summaries (red) with their model counterparts obtained by simulation (blue) for (a) the level of income, (b) income growth, and (c) their cyclical component. While there are clear trends in the distributions of income levels and growth which are not related to the business cycle, our model tracks their evolution over time closely. The model is especially good at matching the dynamics of the persistence and skewness of income growth in panel (b) although it slightly overstates the dispersion.[6] Reassuringly, when we project these summaries onto the business-cycle state in (c), data and model coincide, even for the dispersion. The main takeaways from the comparison, done here for disposable income, also apply to male and household earnings, and they extend to income growth over longer horizons (omitted to economize space).[7]

[5] To average the mixture model for ε_{it}, we take the parameters for each of the last 100 iterations and we simulate a large number of draws (500 draws for each iteration). We pool those draws and fit the mixture model on them one final time. This completely bypasses labeling issues for the components.

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

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

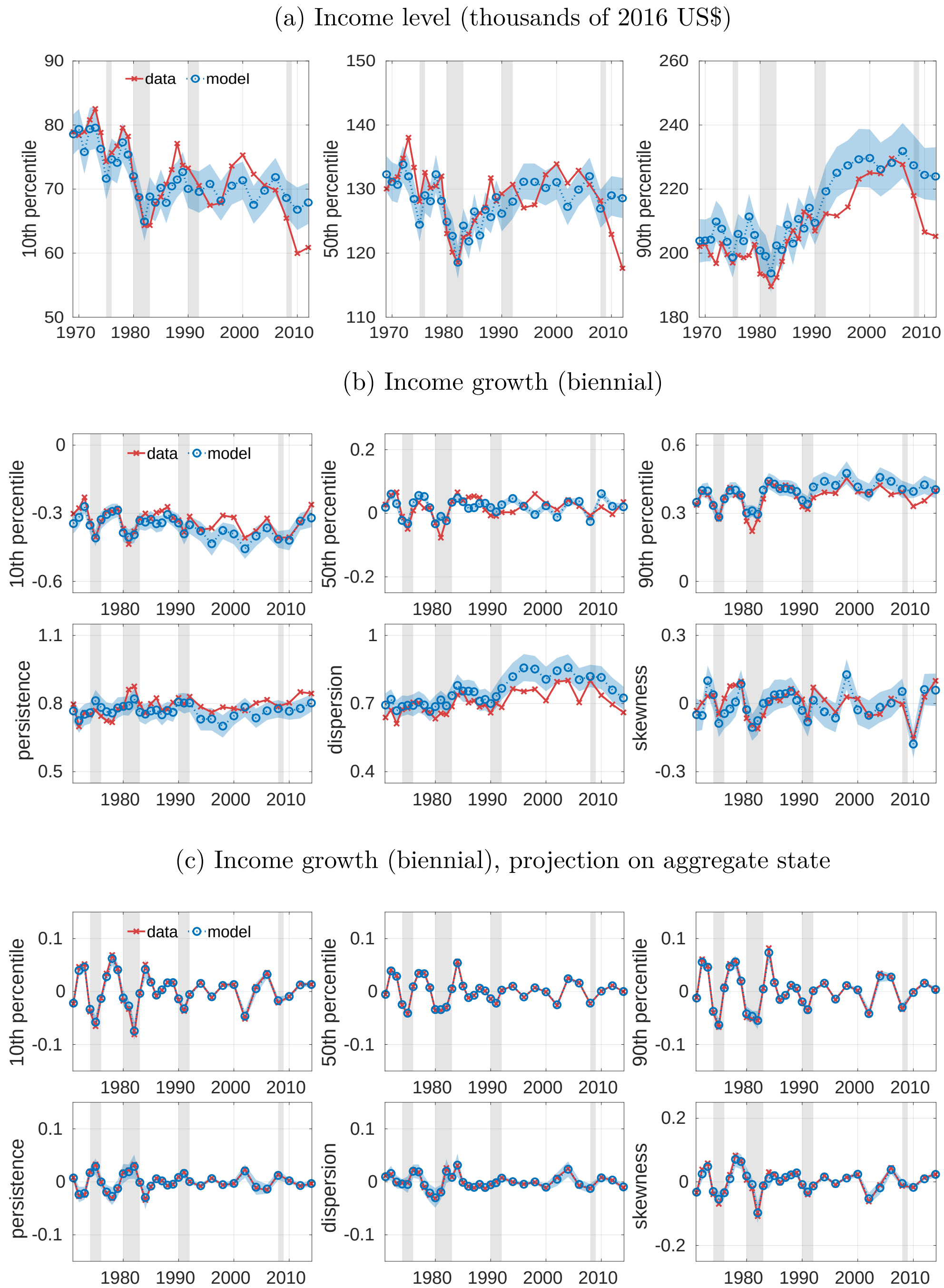


FIGURE D.1. Model fit assessment (disposable income).

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

E Inference comparison with cross-sectional dependence

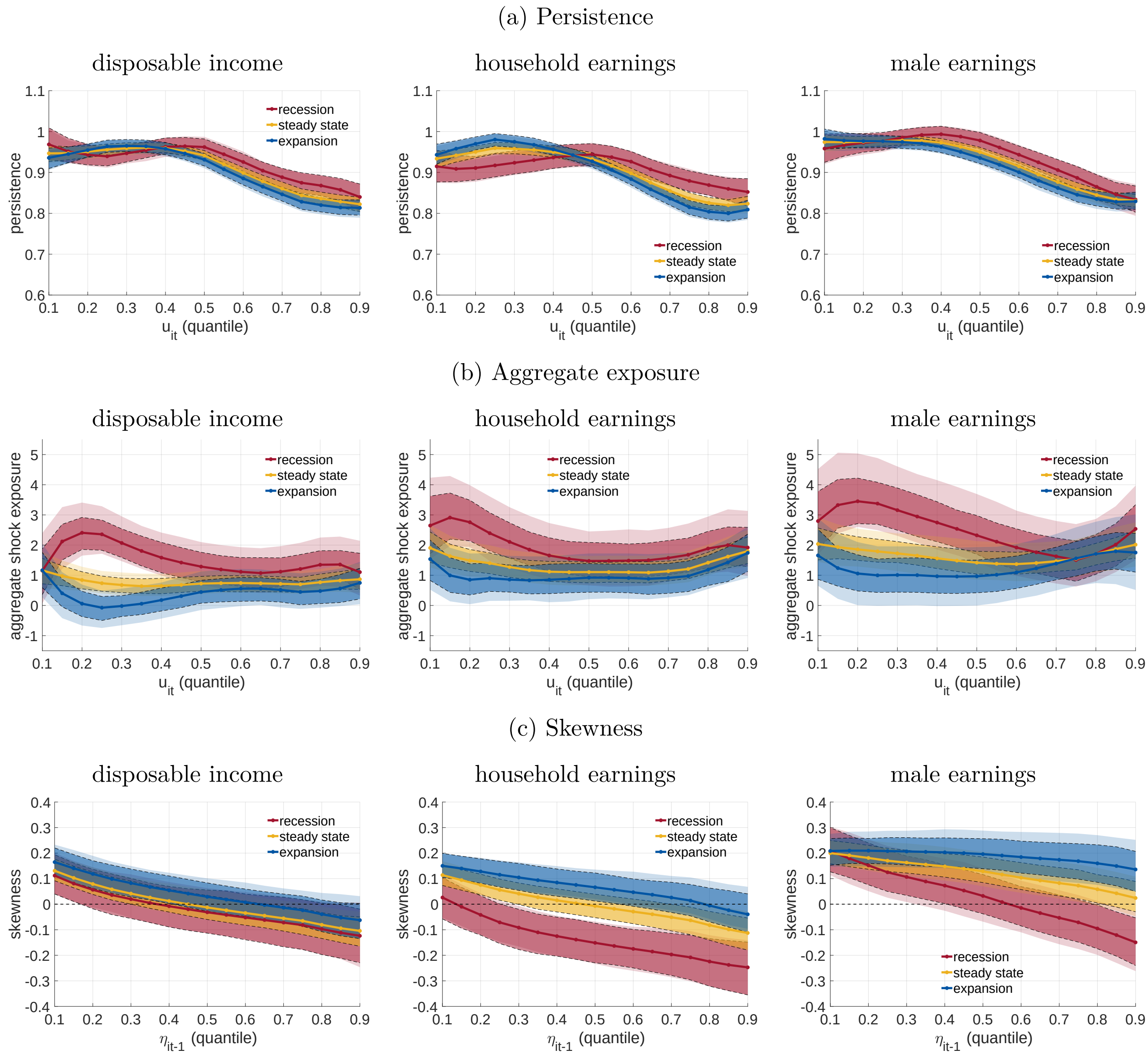


FIGURE E.1. Comparison of confidence bands with and without cross-sectional dependence. *Note*: We report persistence (top) and aggregate exposure (middle) by quantile of the shock $u = u_{it}$ averaging past persistent income $\eta = \eta_{i,t-1}$. We also present conditional skewness (bottom) by quantile of $\eta = \eta_{i,t-1}$. These are presented for disposable income (left), household earnings (center) and male earnings (right). 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). Dark shaded areas correspond to our baseline bootstrap while a lighter shade indicates the confidence bands from a bootstrap that accounts for cross-sectional dependence.

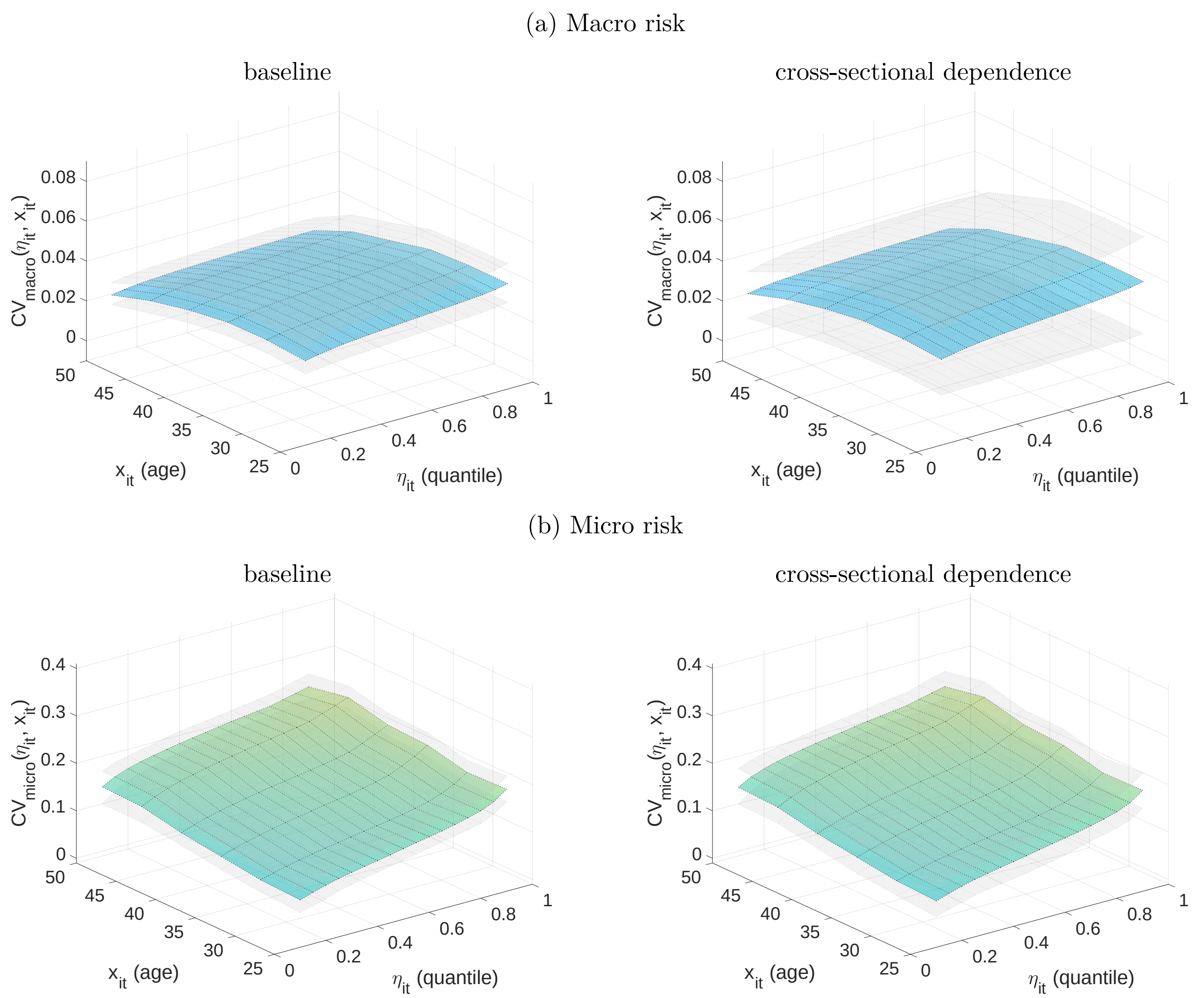


FIGURE E.2. Risk quantification (disposable income).

Note: We show compensating variation for aggregate (upper panels) and idiosyncratic risk (lower panels) for various ages $x = x_{it}$ and initial incomes η_{it}. Lifetime utility is $\sum_{h=1}^{(65-x_{it})/2} \delta^h \frac{e^{(1-\gamma)\eta_{i,t+h}}}{(1-\gamma)}$ with $\delta = (0.96)^2$ and $\gamma = 3$. Gray shaded areas are 90% confidence bands. The panels on the left show our baseline confidence bands while the ones on the right report bands from a bootstrap that accounts for cross-sectional dependence.

F Additional empirical results

This appendix expands on three sets of empirical results. Figure F.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 4 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.

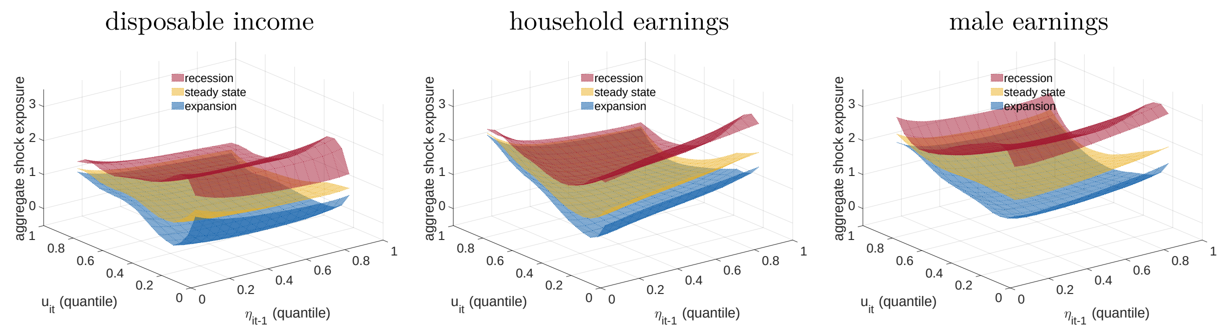
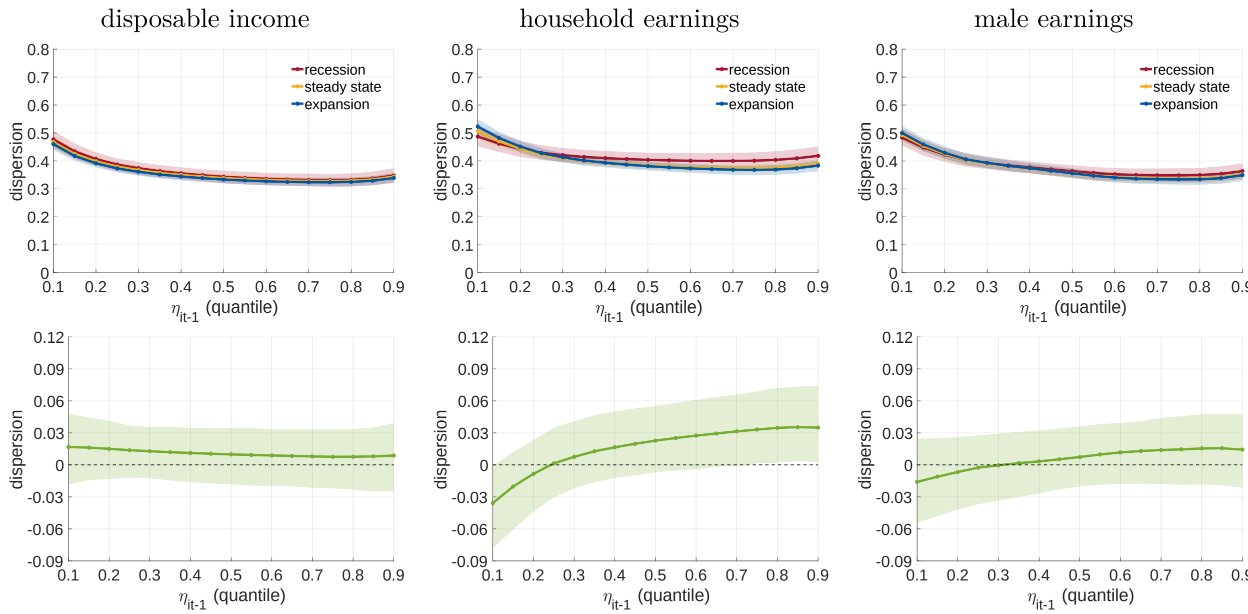


FIGURE F.1. Nonlinear exposure to aggregate shocks.

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 $\eta = \eta_{i,t-1}$. 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).

Figure F.2 displays estimates of dispersion and kurtosis, together with their differences between recessions and expansions. This complements Figure 5 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 insignificant. Though different in methodology and data, our results are in line with the findings in Guvenen, Ozkan, and Song (2014).

(a) Dispersion



(b) Kurtosis

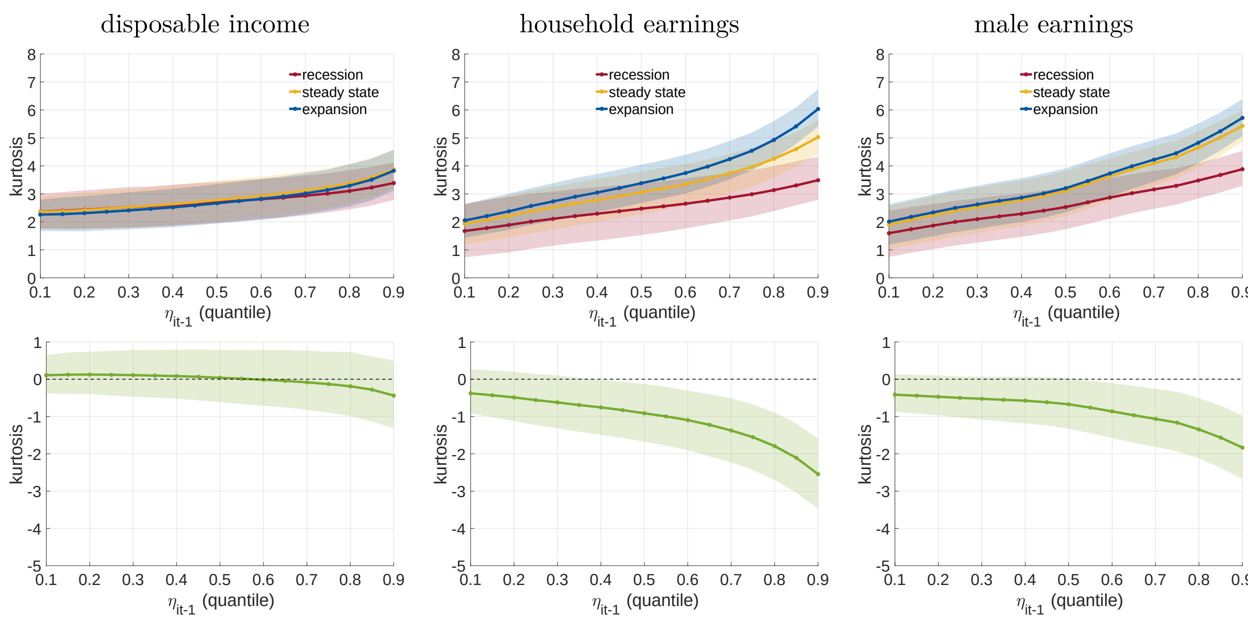


FIGURE F.2. Measures of dispersion and kurtosis.

Note: We report dispersion $\text{disp}\left(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}\right) = Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.9) - Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.1)$ (first row) and kurtosis $\text{kurt}\left(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}\right) = \frac{Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.95) - Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.05)}{Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.75) - Q_\eta(\eta_{i,t-1}, Z_t, Z_{t-1}, x_{it}, 0.25)}$ (third row), 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.

G 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 G.1. We characterize analytically the link between impulse responses, nonlinear persistence and exposure to aggregate shocks in Section G.2. Sections G.3, G.4 and G.5 contain additional results.

G.1 Perturbations and shocks

Having defined impulse responses using perturbations of state variables in the main text, we can next relate them to derivatives with respect to certain macro and micro shocks, which we will denote $\widetilde{V}_t$ and $\widetilde{u}_{it}$. 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,

$$\mathrm{IRF}_{\eta Z}(h;\pi) = \frac{E\Big[\,\eta_{i,t+h}\,\Big|\,\eta_{i,t-1}, \widetilde{V}_t = \pi, Z_{t-1}\Big] - E\Big[\,\eta_{i,t+h}\,\Big|\,\eta_{i,t-1}, \widetilde{V}_t = 0, Z_{t-1}\Big]}{\pi},$$
$$\mathrm{IRF}_{\eta\eta}(h,\pi) = \frac{E\Big[\,\eta_{i,t+h}\,\Big|\,\widetilde{u}_{it} = \pi, \eta_{i,t-1}, Z_{t+1}, Z_t\Big] - E\Big[\,\eta_{i,t+h}\,\Big|\,\widetilde{u}_{it} = 0, \eta_{i,t-1}, Z_{t+1}, Z_t\Big]}{\pi}$$

and, for infinitesimal changes,

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

The implied shocks are given by

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

and lead to the representations

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

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

G.2 IRFs, nonlinear persistence and aggregate exposures

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

$$Z_{t+h} = q_{Z,h}(\boldsymbol{V}_{t+1}^{h-1}, Z_t) = \sum_{\ell=0}^{h-1} \Phi^\ell \sigma_V V_{t+h-\ell} + \Phi^h Z_t, \quad h = 0, 1, \ldots \tag{G.1}$$

Combining Equations (11) with (G.1), we have for $h = 1, 2, \ldots$

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

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

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

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

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

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

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

The impulse responses of the macro state using our methodology is

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

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

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

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

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

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

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

$$\mathrm{IRF}_{\eta\eta}(h) = \rho^h \times \left\{g'(\eta^b)\right\}^{-1},$$

by introducing dependence on the potential history of future shocks.

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

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

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

then

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

Constructing a test of this functional restriction is beyond the scope of our paper, but a look at our estimates suggest that it is unlikely to hold in our sample. 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 1.4 in a typical recession, 0.8 in steady state and 0.5 in a mild expansion. The three quantities also vary substantially over the distribution of past persistent income and micro ranks. All of this suggests that the dynamic common factor restriction (G.2) is not satisfied.

G.3 Additional IRF figures: comparison to MBC shocks

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

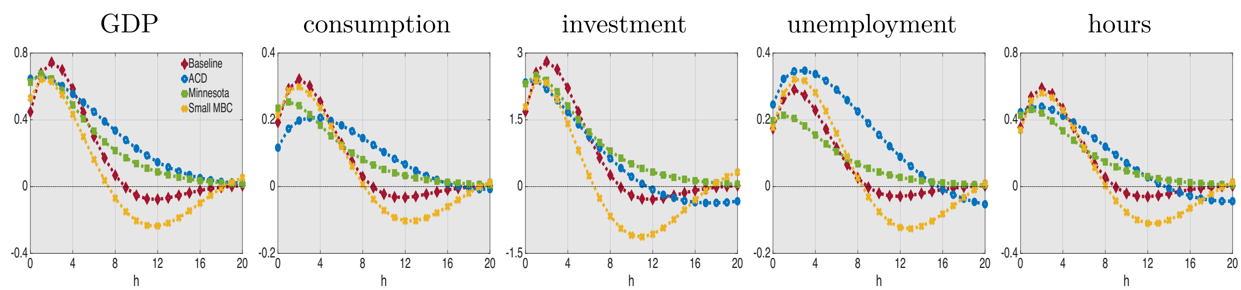


FIGURE G.1. IRFs of W_t to V_t and MBC shock

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 G.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 in 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.

Part of the discrepancy 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. While natural in this setting, this choice penalizes deviations from unit roots that may bias the estimated persistence upward. To explore the issue, Figure G.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).[8] 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 G.2.

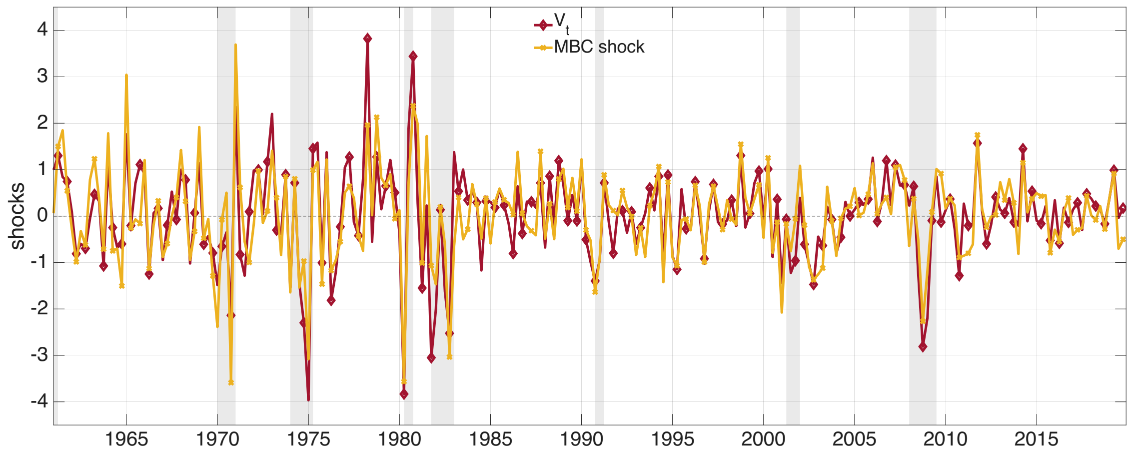


FIGURE G.2. V_t and MBC shocks

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.

The preceding discussion suggests that pinning down the persistence in macro IRFs is empirically difficult. However, our main results are robust to this feature. Because there is very little filtering uncertainty about Z_t, the choice of prior has practically no effect on the estimation of the income process, and objects such as $\rho(\cdot)$, $\text{sk}(\cdot)$ and $\beta(\cdot)$ remain the same. Higher persistence in Z_t produces slower decay in $\text{IRF}_{\eta Z}(h)$ compared to Figure 6 and slightly larger costs of aggregate risk compared to Figure 9, but these results cannot be distinguished statistically from our baseline.[9]

G.4 Additional IRF figures: local projection estimates

In Figure G.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 polynomial on age x_{it} and unit fixed effects. We compute the time-clustered lag-augmented

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

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

heteroskedasticity-robust (t-LAHR) confidence intervals proposed by Almuzara and Sancibrián (2024) to assess statistical precision.

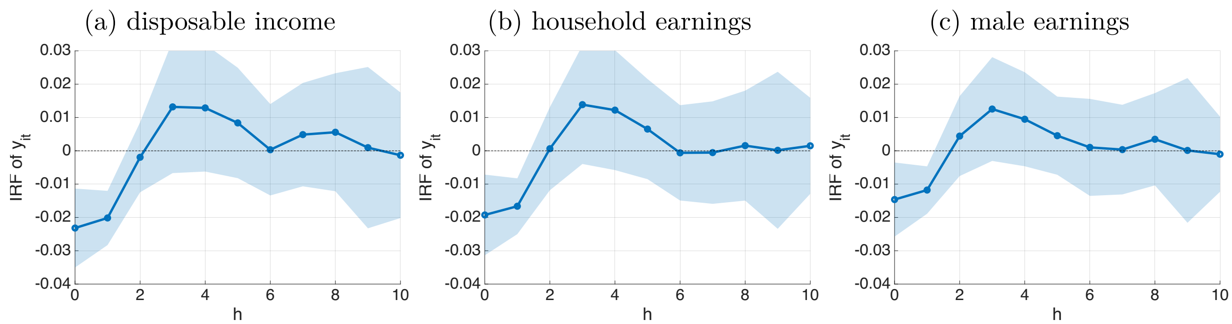


FIGURE G.3. Local projection estimates of macro impulse responses

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 ($h = 1$).[10] 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 6, with larger responses for male earnings compared to disposable income.

G.5 Additional IRF figures: positive shocks

Figure G.4 shows responses to positive macro and micro perturbations, complementing Figure 6 (panels (b) to (d)) and Figure 8. 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 $\pi = \sigma_V$ with $\sigma_V^2 = \text{Var}(\, Z_t \mid Z_{t-1}\,)$. This emulates a mild expansionary aggregate shock. The implied trajectory for Z_t (annualized and scaled to log GDP per capita) is the mirror image of panel (d) in Figure 6, 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 positive perturbation π that implies a 10% increase in $\eta_{i,t-1}$. Similar to Figure 8, 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.

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

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

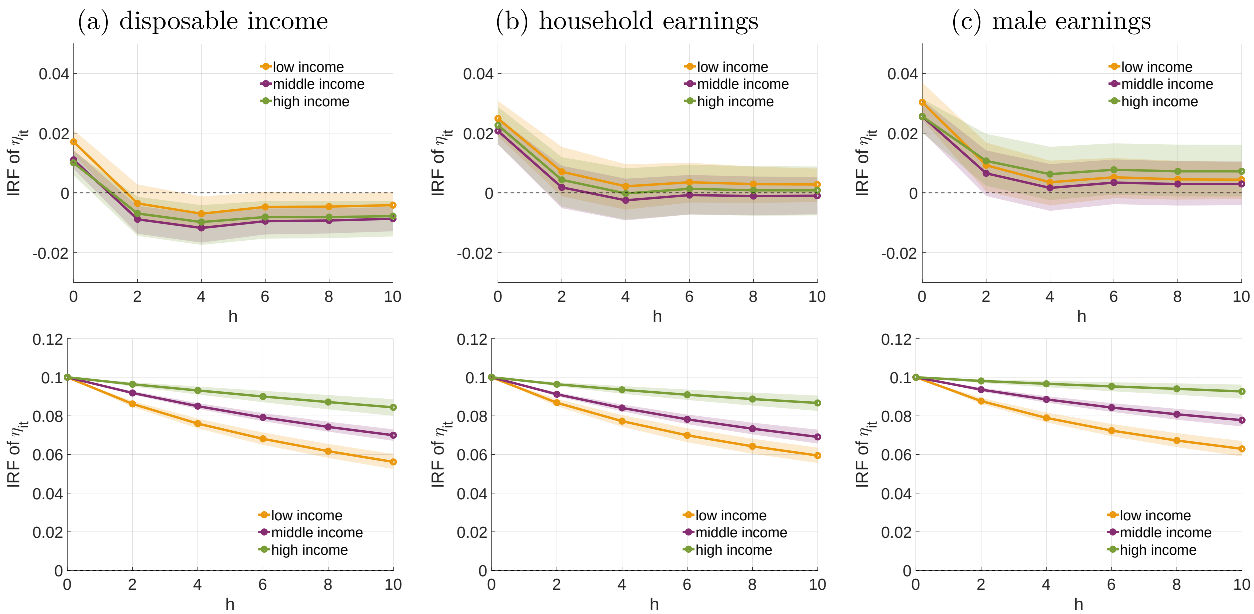


FIGURE G.4. Macro and micro impulse responses to positive shocks

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.

H Additional material on risk quantification

Here we develop a second-order small-noise expansion of CV around a no-shock baseline.

Given (η_{it}, Z_t), let $\eta_{i,t+h} = \tilde{\eta}_{it}^h(\boldsymbol{u}_{i,t+1}^{h-1}, \boldsymbol{V}_{t+1}^{h-1})$ where $\boldsymbol{u}_{i,t+1}^{h-1} = (u_{i,t+\ell})_{\ell=1}^h$ and $\boldsymbol{V}_{t+1}^{h-1} = (V_{t+\ell})_{\ell=1}^h$ are the histories of micro and macro shock. Also let $\widetilde{U}(\eta) = U(e^\eta)$ and transform u_{it} so that it has zero mean—say, by applying the Gaussian inverse CDF to the original u_{it}.

Multiply $\boldsymbol{u}_{i,t+1}^{h-1}$ by ς_u and $\boldsymbol{V}_{t+1}^{h-1}$ by ς_V, so that $\tilde{\eta}_{it}^h(\boldsymbol{u}_{i,t+1}^{h-1}, \boldsymbol{V}_{t+1}^{h-1}) = \eta_{i,t,h}(\varsigma_u, \varsigma_V)|_{(\varsigma_u,\varsigma_V)=(1,1)}$ for some function $\eta_{i,t,h}(\cdot)$. Similarly, we get $\mathrm{CV} = \mathrm{CV}(\varsigma_u, \varsigma_V)|_{(\varsigma_u,\varsigma_V)=(1,1)}$ which we expand to second-order around $(\varsigma_u, \varsigma_V) = (0, 0)$. We focus on the macro risk measure $\mathrm{CV}_{\text{macro}}$ where the experiment eliminates only the macro shocks; analogous insights apply to its micro risk counterpart $\mathrm{CV}_{\mathrm{micro}}$. The small-noise expansion ($\varsigma_u = 0$ and $\varsigma_V \to 0$) delivers

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

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

$$\mathrm{CV}_{\mathrm{macro}} \approx -\frac{\sigma_V^2}{2} \sum_{h=1}^{H} \frac{\delta^{h-1}(1-\delta)}{(1-\delta^H)} \sum_{\ell=1}^{h} \left(\frac{\partial^2 \tilde{\eta}_{it}^h(\mathbf{0},\mathbf{0})}{\partial V_{t+\ell}^2} \right).$$

I Additional results by age and education

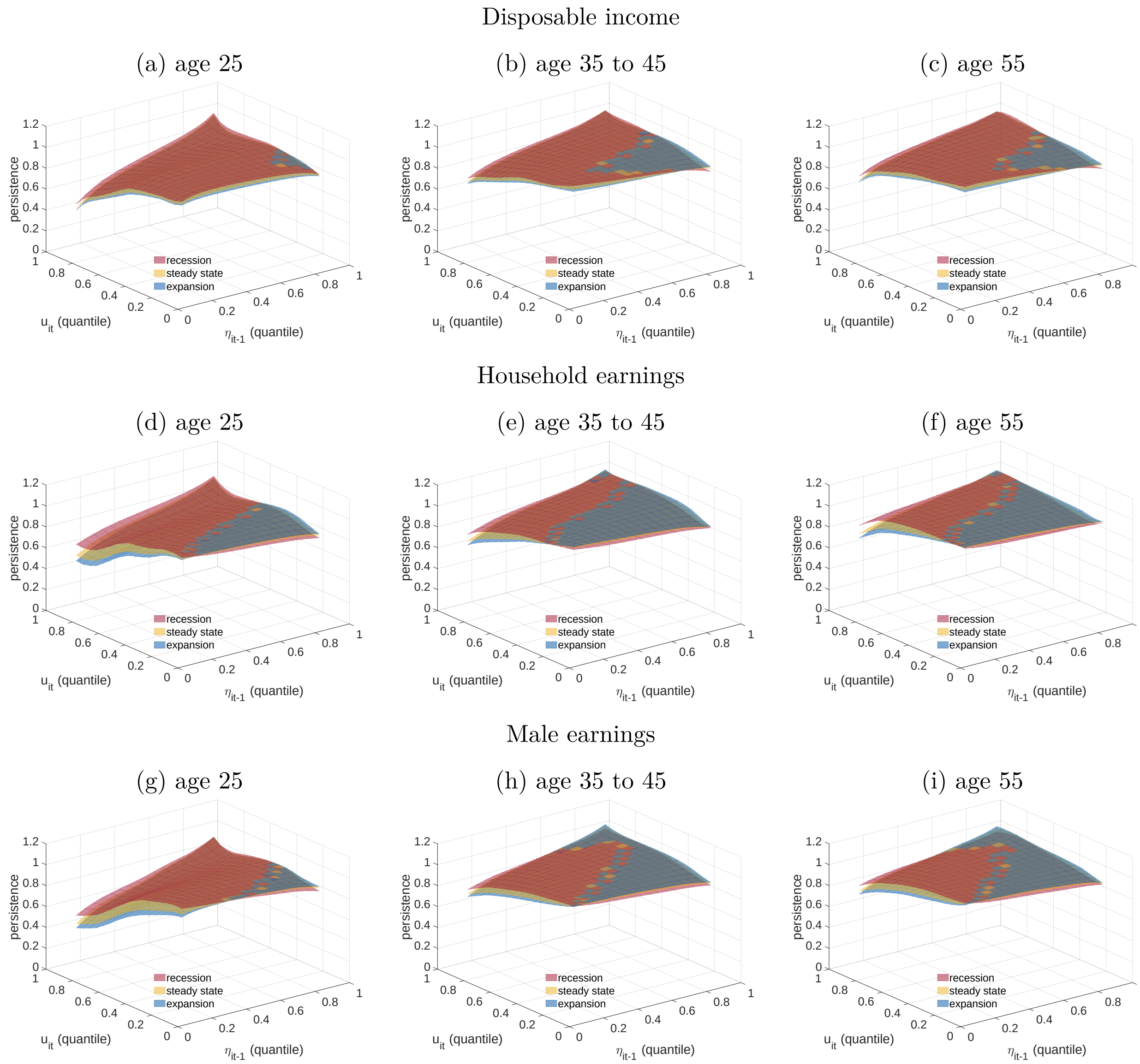


FIGURE I.1. Persistence by age.

Note: We report the persistence measure $\rho(u, \eta, Z_t, Z_{t-1}, x)$ by quantile of the shock $u = u_{it}$ and past persistent income $\eta = \eta_{i,t-1}$ for disposable income, household earnings and male earnings. Age $x = x_{it}$ is either 25, 35-to-45 or 55, $Z_{t-1} = \tilde{Z}_{ss}$ and Z_t is a recession $\tilde{Z}_r$, the steady state $\tilde{Z}_{ss}$ or an expansion $\tilde{Z}_e$.

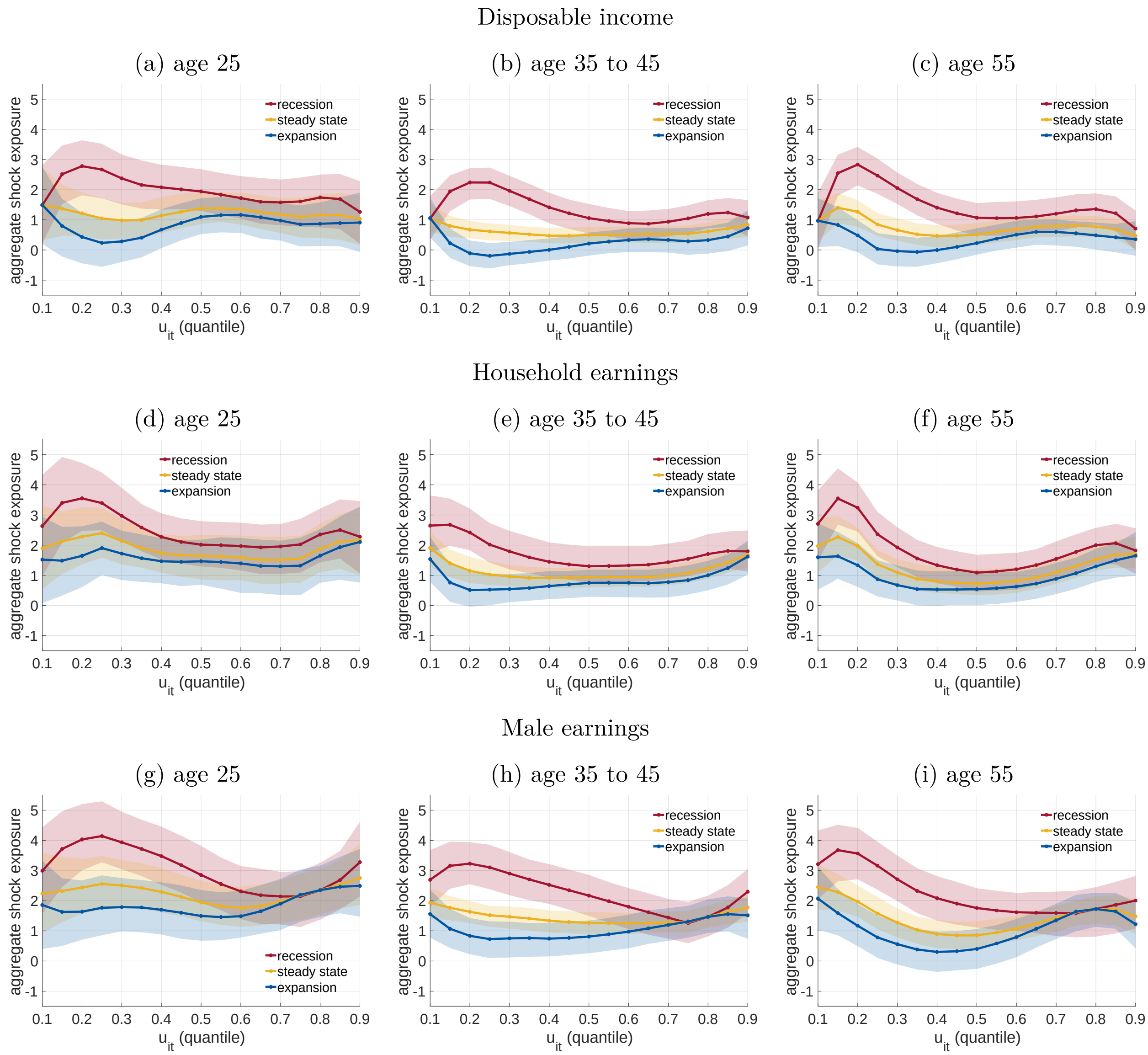


FIGURE I.2. Aggregate exposure by age.

Note: We report the exposure coefficient $\beta(u, \eta, Z_t, Z_{t-1}, x)$ by quantile of the shock $u = u_{it}$ averaged over persistent income $\eta = \eta_{i,t-1}$ for disposable income, household earnings and male earnings. Age $x = x_{it}$ is either 25, 35-to-45 or 55, $Z_{t-1} = \tilde{Z}_{ss}$ and Z_t is a recession $\tilde{Z}_r$, the steady state $\tilde{Z}_{ss}$ or an expansion $\tilde{Z}_e$. Shaded areas represent 90% confidence bands.

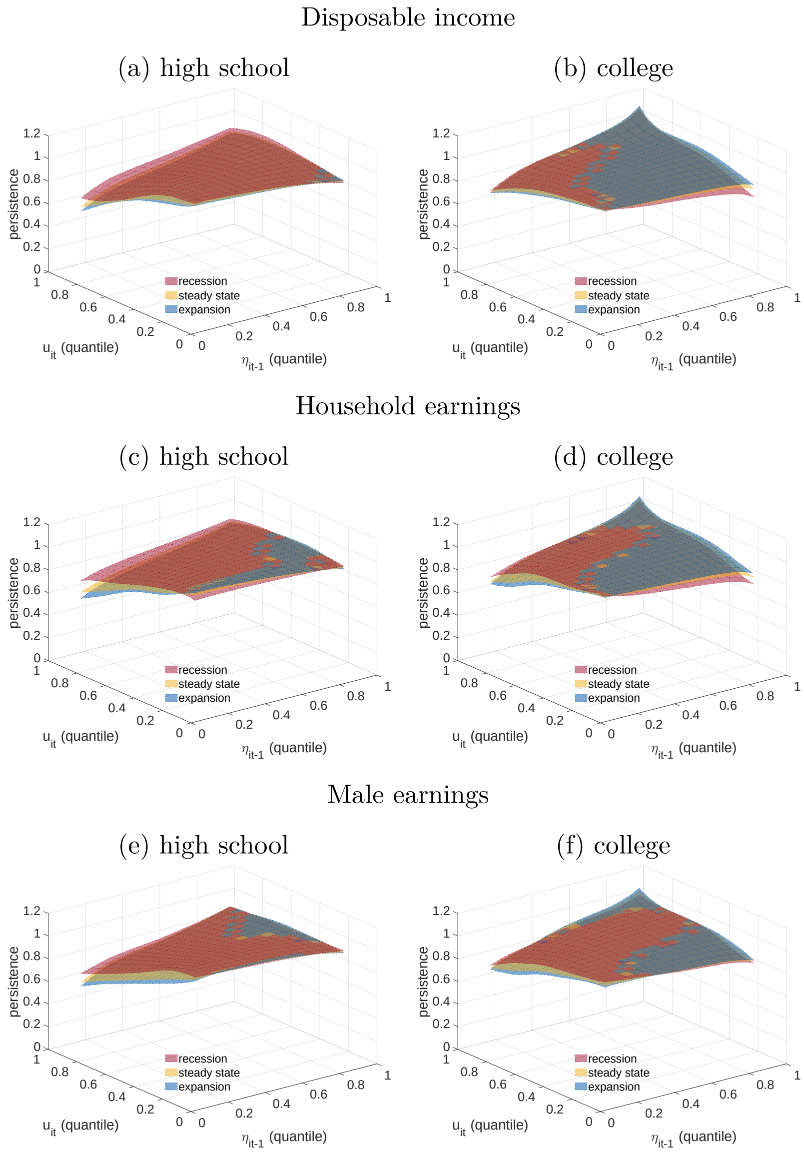


FIGURE I.3. Persistence by education.

Note: We report the persistence measure $\rho(u, \eta, Z_t, Z_{t-1}, x)$ by quantile of the shock $u = u_{it}$ and past persistent income $\eta = \eta_{i,t-1}$ for disposable income, household earnings and male earnings. They are reported for a subsample with up to high-school education (left) and at least some college education (right). 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 $\tilde{Z}_e$.

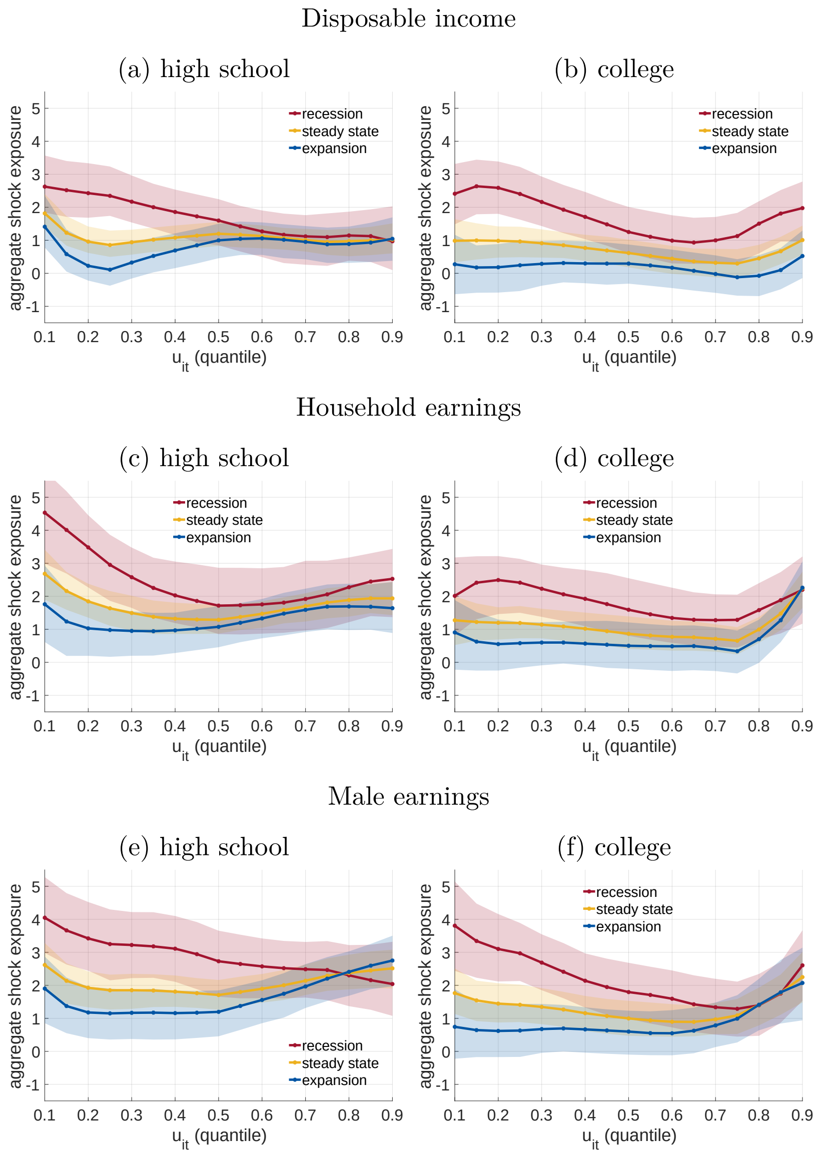


FIGURE I.4. Aggregate exposure by education.

Note: We report the exposure coefficient $\beta(u, \eta, Z_t, Z_{t-1}, x)$ by quantile of the shock $u = u_{it}$ averaged over persistent income $\eta = \eta_{i,t-1}$ for disposable income, household earnings and male earnings. They are reported for a subsample with up to high-school education (left) and at least some college education (right). 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 $\tilde{Z}_e$. Shaded areas represent 90% confidence bands.

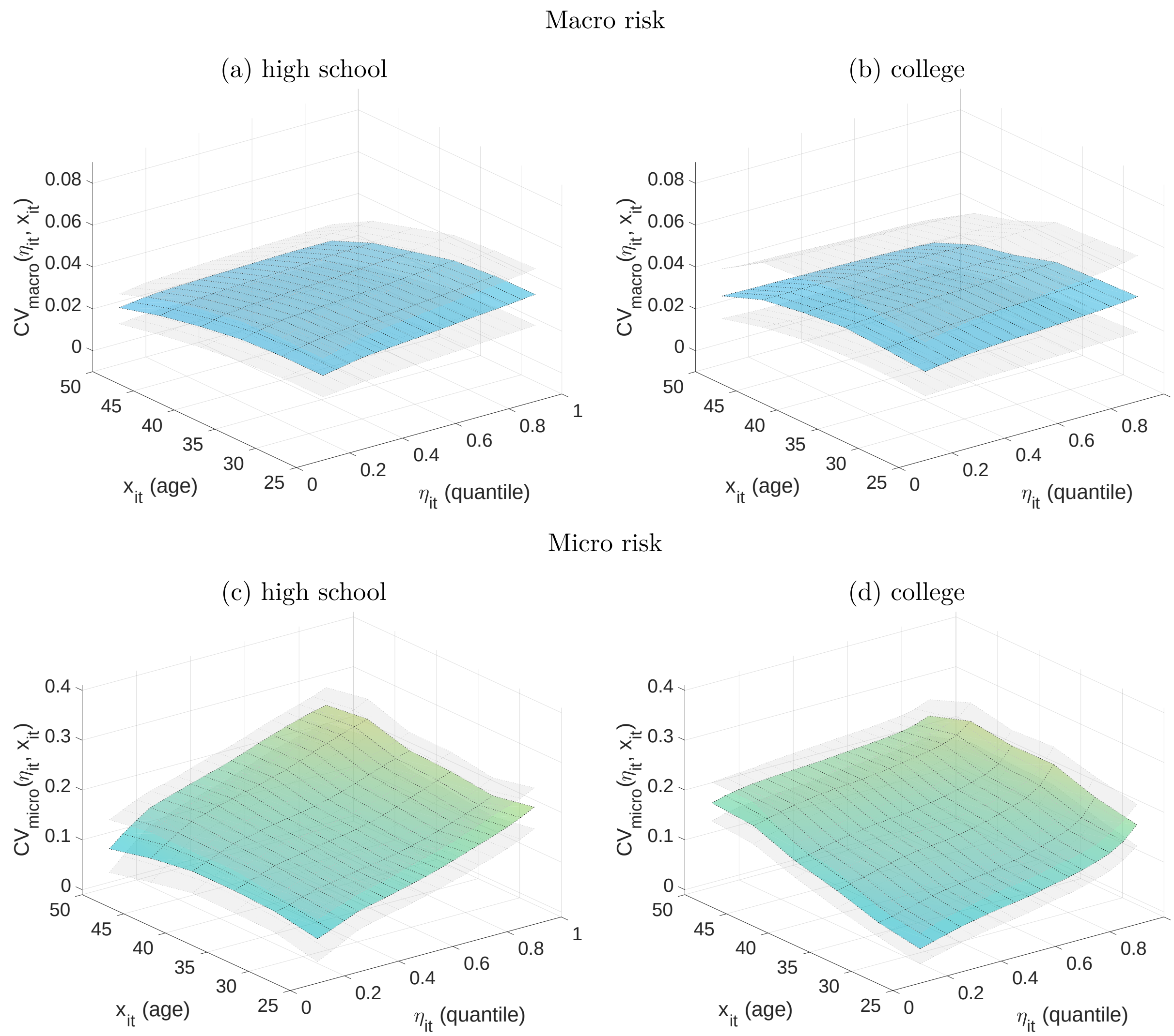


FIGURE I.5. Risk quantification by education.

Note: We show compensating variation for aggregate (upper panels) and idiosyncratic risk (lower panels) for various ages $x = x_{it}$ and initial incomes η_{it}. They are reported for a subsample with up to high-school education (left) and at least some college education (right). Lifetime utility is $\sum_{h=1}^{(65-x_{it})/2} \delta^h \frac{e^{(1-\gamma)\eta_{i,t+h}}}{(1-\gamma)}$ with $\delta = (0.96)^2$ and $\gamma = 3$. Gray shaded areas are 90% confidence bands.

References

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

Almuzara, M., M. Arellano, R. Blundell, and S. Bonhomme (2025): "Nonlinear Micro Income Processes with Macro Shocks." 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, and S. Bonhomme (2017): "Earnings and consumption dynamics: A nonlinear panel data framework." *Econometrica*, 85, 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.

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

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.

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.

Horowitz, J. L. (2001): "The bootstrap." in *Handbook of Econometrics*, ed. by J. J. Heckman and E. Leamer, Elsevier, vol. 5, chap. 52.

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.

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

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

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.

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