Bayesian Estimation of Panel Models under Potentially Sparse Heterogeneity

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide*
Univ. of Pennsylvania
CEPR, NBER, and PIER

Boyuan Zhang
Amazon.com

October 16, 2026

Abstract

We incorporate a version of a spike-and-slab prior, comprising a point mass at zero ("spike") and a Normal distribution around zero ("slab"), into a dynamic panel data framework to model coefficient heterogeneity. In addition to homogeneity and full heterogeneity, our specification can also capture sparse heterogeneity, that is, there is a core group of units that share common parameters and a set of deviators with idiosyncratic parameters. We fit a model with unobserved components to income data from the Panel Study of Income Dynamics. We find evidence for sparse heterogeneity for balanced panels composed of individuals with long employment histories.

JEL CLASSIFICATION: C11, C23, C53, E20

KEY WORDS: Bayesian Analysis, Forecasting, Income Dynamics, Panel Data Models, Sparsity, Spike-and-Slab Priors

*Correspondence: H.R. Moon: Department of Economics, Univ. of Southern California, KAP 300, Los Angeles, CA 90089. E-mail: moonr@usc.edu. F. Schorfheide: Department of Economics, Perelman Center for Political Science and Economics, University of Pennsylvania, 133 S. 36th St., Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu. B. Zhang: Amazon.com, Seattle, WA. Email: zhang.boyuan@hotmail.com.
We thank Dirk Krueger, Víctor Ríos-Rull, and participants of the Wharton Macro Lunch for helpful comments and suggestions. Moon and Schorfheide gratefully acknowledge financial support from the National Science Foundation under Grants SES 1625586 and SES 1851634, respectively. This paper and its contents are not related to Amazon and do not reflect the position of the company and its subsidiaries.

1 Introduction

Panel data models feature two types of coefficients: homogeneous coefficients that are common among all units $i = 1, \ldots, N$, and heterogeneous coefficients that differ across units. In recent years, there has been a growing literature that focuses on the estimation of heterogeneous coefficients, which is challenging in environments in which the time series dimension is small. A natural way of adding information to the estimation of unit-specific coefficients is the use of prior distributions. The key insight in panel data applications is that one can extract information from the cross-section and equate the prior distribution with the (unobserved) cross-sectional distribution of unit-specific coefficients. The time series information for each unit can then be combined with the prior distribution to generate more precise estimates of the heterogeneous coefficients. While in many models the cross-sectional coefficient distribution is in principle nonparametrically identified, in practice it is often desirable to impose parametric restrictions. In this paper, we consider a particular restriction: we assume that there is a core group of units that share a common parameter value, whereas the remaining units (deviators) have idiosyncratic values. If the core group contains a large number of units, then heterogeneity is sparse.

We assume that the heterogeneous coefficients are distributed according to a version of a spike-and-slab (S&S) prior: with probability $1 - q$ the coefficient for unit i equals the predominant value (spike), and with probability q it deviates from the predominant value. This setup nests two important special cases: coefficient homogeneity ($q = 0$) and full heterogeneity ($q = 1$). We will say that heterogeneity is sparse if q is close to zero and dense if q is close to one. The probability q and the parameters characterizing the distribution of the deviations are hyperparameters that are estimated from the panel data. Whether the concept of sparse heterogeneity is useful in practice is an empirical question. We use the proposed prior distribution to estimate a dynamic model for labor earnings obtained from the Panel Study of Income Dynamics (PSID). We construct two types of samples from the PSID data: balanced panels that include individuals who have been working uninterruptedly for a long period of time, and an unbalanced panel that includes individuals with a variety of earnings and employment histories. The balanced panels do feature sparse heterogeneity, whereas the more diverse unbalanced panel is better described by full heterogeneity.

The paper makes the following contributions: first, we incorporate a version of an S&S prior, comprising a point mass at zero ("spike") and a Normal distribution around zero ("slab"), into a dynamic panel data state-space model to capture coefficient heterogeneity.

Such a prior has been used by Geweke (1996) and, more recently, Giannone, Lenza, and Primiceri (2021), in the context of regressor selection problems. We extend the prior to handle heterogeneous variance parameters and show how to implement posterior inference using a Gibbs sampler.

Second, we conduct a Monte Carlo study that compares the point estimation accuracy under the proposed S&S prior to that obtained by (incorrectly) imposing coefficient homogeneity ($q = 0$) or full heterogeneity ($q = 1$). If q is close to zero or one, then estimates under the S&S prior are about as accurate as the estimates that impose the nearly-correct restriction. For intermediate values of q, the estimates under the S&S prior are generally more accurate than at least one of these two restricted alternatives and remain close to the better-performing benchmark. Third, we show in our application that the concept of sparse heterogeneity is empirically useful.

Finally, the paper contributes to the empirical literature on income dynamics. We fit the panel state-space model to earnings data from the PSID. There is a longstanding debate in the literature whether earnings profiles are homogeneous or heterogeneous across individuals. Guvenen (2007) labeled these specifications RIP (restricted income profiles) and HIP (heterogeneous income profiles), respectively. We offer a third option in our setup: sparsely heterogeneous income profiles, which could be abbreviated as SHIP. The literature on the estimation of idiosyncratic income processes has considered many different specifications. There are a few features that all of them share: the income process typically takes the form of an unobserved components model with deterministic and stochastic components. We consider three components: a linear function in experience, an autoregressive component with AR coefficient ρ_i, and an independently and identically distributed (iid) component.[1] There is typically time variation and cross-sectional variation in income risk, i.e., the stochastic components. In our model we allow for time-specific variance coefficients that also vary across individuals. While most of the literature assumes that the autocorrelation parameter of the persistent income component is common across units, we allow for heterogeneity.

The evidence for sparse heterogeneity depends on the construction of the panel data set that is used in the estimation. Using relatively small balanced samples, our posterior estimates imply that heterogeneity with respect to returns to experience and the autocorrelation of the persistent stochastic income component is sparse at best, with large core groups of individuals sharing identical coefficients and a small number of deviators. These estimates can

[1]The unobserved components model is fitted to residuals from a regression of earnings on time fixed effects and some demographic characteristics.

be interpreted as evidence against a HIP version of the model. We also generate one-step-ahead forecasts for the entire cross-section of individuals using our baseline specification, a RIP specification, and a HIP specification. Using log predictive scores, the HIP version is dominated in eight out of nine samples, and the performance of our sparsely heterogeneous baseline specification is close to that of the more restrictive RIP specification. Repeating the estimation on a large unbalanced panel, the parameter estimates lead to an almost fully heterogeneous specification that does not feature core groups of units with identical parameter values, which is evidence in favor of the HIP conjecture.

Moreover, we examine unit-level interval forecasts conditional on information from the full panel $Y_{1:N,1:T}$ accounting for shock and parameter uncertainty, the full panel ignoring parameter uncertainty, and the individual history $Y_{i,1:T}$ accounting for shock and parameter uncertainty, distinguishing between core group members and deviators. This comparison has an important message for the macroeconomic life-cycle model literature: subtle assumptions about the agents' information set can have large effects on the income uncertainty that agents face and hence on their precautionary savings motive.

Our paper is related to several strands of literature. We are using a variant of an S&S prior, which was originally proposed for a regressor selection problem. The prior proposed by Mitchell and Beauchamp (1988) consisted of a point mass at zero ("spike") and a uniform prior ("slab"). For computational reasons George and McCulloch (1993, 1997) replaced the point mass by a Normal distribution with very small variance, and the uniform distribution by a Normal with large variance. Geweke (1996) combined a point mass at zero with a Normal distribution, which is the specification that was recently used by Giannone, Lenza, and Primiceri (2021) to examine whether regression models in economics tend to be sparse or dense. The regressor selection problem can be viewed as a special case of the fundamental problem of estimating the vector of means of a multivariate Normal distribution. Johnstone and Silverman (2004) examined the large sample properties of Bayesian posteriors under a prior that is a mixture of a point mass at zero and a continuous distribution that has fatter tails than a Normal distribution. The authors are particularly interested in sparse settings with a large number of zeros. The asymptotic concentration of posterior distributions in this environment is studied by Castillo and van der Vaart (2012), albeit under a somewhat different prior distribution.

Our paper is concerned with the estimation of potentially heterogeneous slopes and variance parameters in dynamic panel data models, which involves more complicated versions of the vector of means problem. We are building on recent Bayesian (or empirical Bayes)

approaches by Chamberlain and Hirano (1999), Gu and Koenker (2017b,a), Liu, Moon, and Schorfheide (2020, 2023), and Liu (2023). Our key point of departure is the use of an S&S type prior that allows the data to determine whether heterogeneity is sparse or dense, thereby bridging homogeneity ($q = 0$) and full heterogeneity ($q = 1$). This perspective is closely related to Bayesian work on parsimonious random-effects structures, including covariance restrictions in Frühwirth-Schnatter and Tüchler (2008), random-intercept selection in Frühwirth-Schnatter and Wagner (2011), and spike-and-slab random-effects selection in Rodriguez, Williams, and Rast (2024). Relative to these papers, we study a dynamic panel unobserved-components model in which shrinkage and selection jointly operate on intercept and slope heterogeneity, persistence heterogeneity, and variance heterogeneity.

Bonhomme and Manresa (2015) popularized panel data models with group heterogeneity, meaning that there exist a finite number of latent groups with homogeneous parameters, and each unit-level coefficient belongs to one of these groups.[2] Compared to group-heterogeneity specifications, our sparse-heterogeneity setup assumes one predominant value for each heterogeneous parameter and then allows unit-specific deviations around that core, yielding a tractable intermediate benchmark between RIP-style homogeneity and fully flexible group structures. At the same time, our setup can be viewed as a building block for richer specifications with multiple core groups. We leave this extension for future research.

There exists a large literature at the intersection of labor economics and macroeconomics on the estimation of idiosyncratic income processes. A longstanding debate in this literature is whether income profiles are homogeneous or heterogeneous. The heterogeneity typically refers to the return to experience. Among others, MaCurdy (1982), Abowd and Card (1989), and Guvenen (2007, 2009) provide evidence in favor of HIP, whereas Lillard and Weiss (1979), Hause (1980), Baker (1997), Haider (2001), Hryshko (2012), and Hoffmann (2019) favor the homogeneous RIP specification, which often involves a unit-root or highly persistent stochastic component. The importance of heterogeneity in income profiles, persistence of income processes, and the variance of income shocks is also emphasized by Browning, Ejrnæs, and Alvarez (2010) and Browning and Ejrnæs (2013). Coefficient heterogeneity in this literature is typically modeled through random effects. The latter two papers allow for these random effects to be correlated with each other, assuming they are determined by some underlying factors. While we do not allow for conditional heteroskedasticity in our model specification as, for instance, in Meghir and Pistaferri (2004) or Hospido (2012), we do allow

[2]For a Bayesian implementation of group heterogeneity that allows users to formulate priors about group memberships see Zhang (2023).

for cross-sectional heterogeneity as in Gu and Koenker (2017b) and find it to be dense.

In terms of empirical focus and findings, our application differs from the aforementioned papers, which do not focus on decomposing earnings dynamics into persistent and transitory components or on comparing RIP, HIP, and sparse-heterogeneity specifications. In the income-dynamics literature, our paper is complementary to work estimating flexible earnings processes from large administrative panels, such as Guvenen, Karahan, Ozkan, and Song (2021), and to models with additional state structure, such as Gordon, Jones, Neelakantan, and Athreya (2023). Our empirical question is whether the cross section supports sparse versus dense unit-level parameter heterogeneity in a dynamic panel state-space model, and how that conclusion changes with sample construction.

The relationship between pre-processing data and modeling heterogeneity is central for our empirical analysis. Following standard practice in the income-dynamics literature, we combine sample-selection rules (for instance, balancing requirements and related restrictions) with a first-stage regression that residualizes log income with respect to year effects and observed demographics; see Section 5.1 and, for example, Meghir and Pistaferri (2004). Joint estimation of both stages is feasible, but we follow the standard two-step design used in the income-dynamics literature and leave joint estimation for future work. These choices affect both the cross-sectional composition and the variation passed to the second-stage state-space model, and therefore the degree of heterogeneity that must be modeled. A broader and more heterogeneous sample can reveal richer unit-level dispersion but also raises the risk of misspecification in a parsimonious model, whereas tighter sample definitions reduce misspecification risk but may attenuate measured heterogeneity.

An important branch of the earnings dynamics literature focuses on the estimation of (correlated) random effects distributions, characterizing the degree of coefficient heterogeneity in the population. Our Bayesian estimation approach generates such estimates along with estimates of the unit-level coefficients, under the assumption of potentially sparse heterogeneity. Whether or not the interest lies in the random effects distribution or the unit-level coefficients depends on the particular application. One novel contribution of our application below is to also assess the fitted income-dynamics model by evaluating the pseudo-out-of-sample forecast accuracy for the individuals included in the panel data set, rather than a hypothetical population implied by the estimated model.

The remainder of the paper is organized as follows. Our approach of capturing potentially sparse heterogeneity through an S&S type prior distribution is described in the context of a

vector of means estimation in Section 2. This prior is then incorporated into a panel state-space model with heterogeneous intercept, slope coefficients, and shock standard deviations in Section 3. Results from a small Monte Carlo study are summarized in Section 4 and the empirical analysis is presented in Section 5. Finally, Section 6 concludes. Details on the posterior samplers for the Monte Carlo study and the empirical analysis are provided in Online Appendix A. Convergence diagnostics are presented in Online Appendix B, and additional empirical results in Online Appendix C.

2 Sparse Heterogeneity

Throughout this paper, we pursue a Bayesian approach, place a prior distribution over the parameters, and generate draws from the posterior distribution. In the context of estimating a vector of means of a multivariate Normal distribution, our approach to sparse heterogeneity takes the following form. Consider the model

$$y_i = \delta_i + u_i, \quad u_i \overset{iid}{\sim} \mathcal{N}(0,1), \quad i = 1, \ldots, N \tag{1}$$

equipped with a prior distribution that is a mixture of a point mass at zero and a Normal distribution:

$$\delta_i | (q, v_\delta) \sim \begin{cases} \mathcal{N}(0, v_\delta) & \text{with probability } q \\ 0 & \text{with probability } 1-q \end{cases}. \tag{2}$$

This prior distribution has been used by Geweke (1996) and, more recently, by Giannone, Lenza, and Primiceri (2021) in the context of regressor selection problems. The sparsity of the $\delta = [\delta_1, \ldots, \delta_N]'$ vector is determined by the hyperparameter q, which controls the probability of the coefficient for unit i deviating from the predominant value. Thus, $1-q$ is the probability assigned to the point mass at zero and hence the height of the "spike." The second hyperparameter, v_δ, controls the width of the "slab." The larger v_δ, the larger the deviations from the common parameter are on average. We will refer to units for which $\delta_i = 0$ as the core group, whereas the other units are deviators.[3] The prior can be generalized in a straightforward manner by replacing the $\mathcal{N}(0, v_\delta)$ distribution with a $\mathcal{N}(\underline{\delta}, v_\delta)$ distribution and treating the $\underline{\delta}$ as a third hyperparameter.

Let $Y_{1:N}$ denote $\{y_1, \ldots, y_N\}$. Conditional on the hyperparameters (q, v_δ), the posterior

[3] Alternatively, one could think of two groups: a $\delta_i = 0$ group and a $\delta_i \neq 0$ group.

distribution is also a mixture of a point mass at zero and a Normal component:

$$\delta_i|(Y_{1:N}, q, v_\delta) \sim \begin{cases} \mathcal{N}\big(\delta_i^*(v_\delta, y_i),\ v_\delta^*(v_\delta)\big) & \text{with prob. } q_i^*(q, v_\delta, y_i) \\ 0 & \text{with prob. } 1 - q_i^*(q, v_\delta, y_i) \end{cases}, \tag{3}$$

where

$$v_\delta^*(\cdot) = \frac{v_\delta}{v_\delta + 1}, \quad \delta_i^*(\cdot) = v_\delta^*(\cdot) y_i, \quad \frac{q_i^*(\cdot)}{1 - q_i^*(\cdot)} = \frac{q}{1-q}(v_\delta + 1)^{-1/2} \exp\left\{\frac{1}{2} v_\delta^*(\cdot) y_i^2\right\}. \tag{4}$$

The odds in favor of a Normal component are an increasing function of the squared deviations from zero, y_i^2. The posterior mean is given by $\mathbb{E}[\delta_i|Y_{1:N}] = q_i^* \delta_i^*$, which is never exactly equal to zero. However, it can be shown that the posterior median, denoted by $\text{med}(\delta_i|Y_{1:N})$, has the property that it can be exactly equal to zero for small values of y_i^2; see, for instance, Johnstone and Silverman (2004).

The hyperparameters are determined by the marginal likelihood function

$$\begin{aligned} & p(Y_{1:N}|q, v_\delta) \\ &= (2\pi)^{-N/2} \prod_{i=1}^N \left(q(1+v_\delta)^{-1/2} \exp\left\{-\frac{1}{2(1+v_\delta)} y_i^2\right\} + (1-q)\exp\left\{-\frac{1}{2} y_i^2\right\}\right). \end{aligned} \tag{5}$$

We combine this marginal likelihood with a prior $p(q, v_\delta)$ for the hyperparameters and conduct posterior inference. For the posterior computations it is convenient to introduce the auxiliary variables z_i such that $z_i = 0$ if $\delta_i = 0$ and $z_i = 1$ if $\delta_i \neq 0$. We say that unit i is a member of the core group if $z_i = 0$ and it is a deviator if $z_i = 1$. Then,

$$\begin{aligned} & p(Y_{1:N}, Z_{1:N}|q, v_\delta) \\ &= p(Y_{1:N}|Z_{1:N}, q, v_\delta) p(Z_{1:N}|q, v_\delta) \\ &= (2\pi)^{-N/2} \left(\prod_{i\,|\,z_i=1} (1+v_\delta)^{-1/2} \exp\left\{-\frac{1}{2(1+v_\delta)} y_i^2\right\}\right) \left(\prod_{i\,|\,z_i=0} \exp\left\{-\frac{1}{2} y_i^2\right\}\right) \\ &\quad \times \prod_{i=1}^N (1-q)^{1-z_i} q^{z_i}. \end{aligned} \tag{6}$$

Given a prior for (q, v_δ), draws from the posterior distribution can be obtained via Gibbs sampling, iterating over the conditional posteriors $Z_{1:N}|(Y_{1:N}, q, v_\delta)$, $q|(Y_{1:N}, Z_{1:N}, v_\delta)$, and $v_\delta|(Y_{1:N}, Z_{1:N}, q)$.

The prior distribution in (2) nests coefficient homogeneity ($q = 0$) and full coefficient heterogeneity ($q = 1$) as special cases. In practice, whether q is close to zero or one may depend on how the estimation sample is constructed from the raw data. For instance, in our empirical application we use the raw panel data set to construct balanced panels in which units are fairly homogeneous and an unbalanced panel that tries to include as many units as possible. The former samples will feature sparse heterogeneity in some dimensions, whereas the latter sample will lead to q estimates that are close to one.

3 A Panel State-Space Model with Sparse Heterogeneity

While sparse heterogeneity can be incorporated in many different types of panel data models, we consider an unobserved components model that is tailored toward the empirical application in Section 5:

$$y_{it} = x_{it}'(\alpha + \delta_i^{\alpha}) + s_{it} + \sigma_{u,t}\sqrt{\delta_i^{\sigma_u}}u_{it}, \tag{7}$$

$$s_{it} = (\rho + \delta_i^{\rho})s_{it-1} + \sigma_{\epsilon,t}\sqrt{\delta_i^{\sigma_\epsilon}}\epsilon_{it}. \tag{8}$$

In the application, y_{it} is the income of unit i in period t, which consists of a persistent component s_{it} and a transitory iid component u_{it}. More generally, (7) can be interpreted as the measurement equation of the state-space model, and (8) is the state transition equation. The parameter α is now a k-dimensional vector that interacts with the vector of regressors x_{it}. In the application, x_{it} comprises a constant and a measure of experience. We also let the shock standard deviations $\sigma_{u,t}$ and $\sigma_{\epsilon,t}$ vary over time. This specification allows for cross-sectional heterogeneity in the coefficients α, the autoregressive parameter ρ, and the innovation standard deviations through the discrepancies δ_i^{α}, δ_i^{ρ}, $\delta_i^{\sigma_u}$, and $\delta_i^{\sigma_\epsilon}$. Units i for which the discrepancies are equal to zero (the core group) share common parameter values, whereas the remaining units deviate from the parameters of the core groups and have heterogeneous coefficients.

Prior for the Homogeneous Coefficients. We use the following priors for the common parameters α, ρ, $\sigma_{u,t}^2$, and $\sigma_{\epsilon,t}^2$ of the core groups:

$$\alpha \sim \mathcal{N}(\underline{\mu}_{\alpha}, \underline{v}_{\alpha}), \quad \rho \sim \mathcal{N}(\underline{\mu}_{\rho}, \underline{v}_{\rho}), \quad \sigma_{u,t}^2 \sim IG\left(\frac{\underline{\nu}_{\sigma_u}}{2}, \frac{\underline{\tau}_{\sigma_u}}{2}\right), \quad \sigma_{\epsilon,t}^2 \sim IG\left(\frac{\underline{\nu}_{\sigma_\epsilon}}{2}, \frac{\underline{\tau}_{\sigma_\epsilon}}{2}\right), \tag{9}$$

where $\sigma^2_{u,t}$ and $\sigma^2_{\epsilon,t}$ are independent of each other and over time $t = 1, \ldots, T$.

Prior for the Heterogeneous Coefficients. To induce sparsity, we use the prior in (2) for δ_i^{α} and δ_i^{ρ}, equipping q and v_δ with superscripts α and ρ. For $k > 1$, v_{δ^α} is a covariance matrix. Because of the non-negativity of variances, we use slightly different priors for $\delta_i^{\sigma_u}$ and $\delta_i^{\sigma_\epsilon}$:

$$\delta_i^{\sigma_u} | (q^{\sigma_u}, v_{\delta^{\sigma_u}}) \sim \begin{cases} IG\big(\nu_{\delta^{\sigma_u}}(v_{\delta^{\sigma_u}})/2,\, \tau_{\delta^{\sigma_u}}(v_{\delta^{\sigma_u}})/2\big) & \text{with prob. } q^{\sigma_u} \\ 1 & \text{with prob. } 1 - q^{\sigma_u} \end{cases}, \tag{10}$$

where IG is the Inverse Gamma distribution. We impose that $\mathbb{E}[\delta_i^{\sigma_u}] = 1$ and reparameterize the prior distribution in terms of its variance $v_{\delta^{\sigma_u}}$, which leads to the transformation

$$\nu_{\delta^{\sigma_u}}(v_{\delta^{\sigma_u}}) = 2v_{\delta^{\sigma_u}}^{-1} + 4, \quad \tau_{\delta^{\sigma_u}}(v_{\delta^{\sigma_u}}) = 2v_{\delta^{\sigma_u}}^{-1} + 2. \tag{11}$$

The prior for $\delta_i^{\sigma_\epsilon}$ is identical, except that all u subscripts are replaced by ϵ subscripts.

Prior for the Distribution of Heterogeneous Coefficients. For the parameters of the S&S priors – see (2) and (10) – we use

$$q^l \sim B(a, b),\ l \in \{\alpha, \rho, \sigma_u, \sigma_\epsilon\}; \quad v_l \sim IG\left(\frac{\underline{\nu}_{\delta^l}}{2}, \frac{\underline{\tau}_{\delta^l}}{2}\right),\ l \in \{\rho, \sigma_u, \sigma_\epsilon\}, \tag{12}$$

where $B(a, b)$ is the Beta distribution. Because α is a vector for $k > 1$, we use an Inverse Wishart distribution for v_{δ^α}:

$$v_{\delta^\alpha} \sim IW(\underline{\nu}_{\delta^\alpha}, \underline{\Psi}_{\delta^\alpha}). \tag{13}$$

Prior for the Initial States. Finally, we need a prior distribution for the initial state s_{i0}. We assume that $s_{i0} \overset{iid}{\sim} \mathcal{N}(\mu_{s_0}, v_{s_0})$, where

$$\mu_{s_0} \sim \mathcal{N}(\underline{\mu}_{s_0}, \underline{v}_{s_0}), \quad v_{s_0} \sim IG\left(\frac{\underline{\nu}_{s_0}}{2}, \frac{\underline{\tau}_{s_0}}{2}\right). \tag{14}$$

The state-transition equation (8) induces the distribution of $s_{i1}, \ldots, s_{iT}$ conditional on s_{i0}.

Grouping the Parameters. We define the parameter vector θ as

$$\theta = [\alpha', \rho, \sigma^2_{u,1:T}, \sigma^2_{\epsilon,1:T}, q^\alpha, q^\rho, q^{\sigma_u}, q^{\sigma_\epsilon}, v_{\delta^\alpha}, v_{\delta^\rho}, v_{\delta^{\sigma_u}}, v_{\delta^{\sigma_\epsilon}}, \mu_{s_0}, v_{s_0}]'. \tag{15}$$

This vector contains the (homogeneous) parameters of the core groups, the parameters that characterize the distribution of heterogeneous coefficients through the S&S specification, and

the mean and variance of the initial state. The hyperparameters associated with the prior distributions in (9), (10), (12), (13), and (14) are collected in the vector

$$\begin{aligned}\lambda &= [\underline{\mu}_{\alpha}, \underline{v}_{\alpha}, \underline{\mu}_{\rho}, \underline{v}_{\rho}, \underline{\nu}_{\sigma_u}, \underline{\tau}_{\sigma_u}, \underline{\nu}_{\sigma_\epsilon}, \underline{\tau}_{\sigma_\epsilon}, a, b, \underline{\nu}_{\delta^\alpha}, \underline{\Psi}_{\delta^\alpha}, \underline{\nu}_{\delta^\rho}, \underline{\tau}_{\delta^\rho}, \underline{\nu}_{\delta^{\sigma_u}}, \underline{\tau}_{\delta^{\sigma_u}}, \underline{\nu}_{\delta^{\sigma_\epsilon}}, \underline{\tau}_{\delta^{\sigma_\epsilon}}, \\ &\quad \underline{\mu}_{s_0}, \underline{v}_{s_0}, \underline{\nu}_{s_0}, \underline{\tau}_{s_0}]' . \end{aligned} \tag{16}$$

Posterior Sampling. We use a Gibbs sampler to generate draws from the posterior distribution of the parameters $\left(\theta, \left\{\delta_i^\alpha, \delta_i^\rho, \delta_i^{\sigma_u}, \delta_i^{\sigma_\epsilon}\right\}_{i=1}^N\right)$ and the latent states $S_{1:N,0:T} = \left\{s_{i0}, \ldots, s_{iT}\right\}_{i=1}^N$. We subsequently outline two important aspects of our implementation of the Gibbs sampler. A detailed description of the various steps of the Gibbs sampler is provided in Online Appendix A.[4] First, we use indicator variables for sparsity:

$$z_i^l = \begin{cases} 0 & \text{if } \delta_i^l = 0 \\ 1 & \text{otherwise} \end{cases}, \; l \in \{\alpha, \rho\}, \quad \text{and} \quad z_i^{\sigma_m} = \begin{cases} 0 & \text{if } \delta_i^{\sigma_m} = 1 \\ 1 & \text{otherwise} \end{cases}, \; m \in \{u, \epsilon\}. \tag{17}$$

Second, for posterior inference on $S_{i,0:T}$ we write the model as

$$Y_i = X_i\alpha + W_i\beta_i + D_{i,u}U_i, \tag{18}$$

where Y_i and U_i are vectors, and X_i are matrices that stack y_{it}, u_{it}, and x_{it}', respectively. $D_{i,u}$ is a $T \times T$ diagonal matrix with elements $\sigma_{u,t}\sqrt{\delta_i^{\sigma_u}}$ and the $T \times (k+T)$ design matrix W_i comprises the regressor matrix X_i and time dummies. For instance, if $k = 1$ and $x_{it} = 1$ such that α is an intercept,

$$W_i = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} X_i & I \end{bmatrix} \tag{19}$$

and $\beta_i' = [\delta_i^\alpha, \; s_{i1}, \ldots, s_{iT}]$ stacks the α discrepancy and the latent states. The prior for β_i combines the prior for δ_i^α with the prior for the latent states generated by the state-transition

[4]Posterior computations for the restricted $q = 0$ and $q = 1$ versions of the panel state-space model have been conducted by Nakata and Tonetti (2015). Their paper shows through a simulation study that Bayes estimates are generally more accurate than the GMM/minimum distance estimates that are widely used in the empirical literature.

equation (8). In the case $k = 1$, this leads to

$$\begin{aligned} \beta_i|(\rho, \delta_i^\rho, z_i^\alpha, q^\alpha, v_{\delta^\alpha}, D_{i,\epsilon}) &\sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0_{T\times 1} \end{bmatrix}, \begin{bmatrix} z_i^\alpha v_{\delta^\alpha} & 0_{1\times T} \\ 0_{T\times 1} & \underline{V}_{s_i} \end{bmatrix}\right) \\ z_i^\alpha|q^\alpha &= \begin{cases} 0 & \text{with prob. } 1-q^\alpha \\ 1 & \text{with prob. } q^\alpha \end{cases}, \end{aligned} \tag{20}$$

where $D_{i,\epsilon}$ is a $T \times T$ diagonal matrix with elements $\sigma_{\epsilon,t}\sqrt{\delta_i^{\sigma_\epsilon}}$. The advantage of this representation is that we can jointly sample the discrepancies δ_i^α and the latent states, which have a strong *a posteriori* correlation.

The (t, τ) elements of the $T \times T$ prior covariance matrix $\underline{V}_{s_i}$ can be calculated as follows:

$$\begin{aligned} \underline{V}_{s_i}(t,t) &= (\rho + \delta_i^\rho)^2 \underline{V}_{s_i}(t-1, t-1) + \sigma_{\epsilon,t}^2 \delta_i^{\sigma_\epsilon}, \quad t = 1, \ldots, T \\ \underline{V}_{s_i}(t,\tau) &= (\rho + \delta_i^\rho)^{|t-\tau|} \underline{V}_{s_i}\big(\min(t,\tau), \min(t,\tau)\big), \quad t = 1, \ldots, T \text{ and } \tau \neq t, \end{aligned}$$

with the understanding that $\underline{V}_{s_i}(0,0) = v_{s_0}$. The conditional *posterior* distribution of β_i is also a mixture of Normals.

In the simulation we let $x_{it} = [1\ t]'$, and in the application we set $x_{it} = [1\ h_{it}/10]'$, where h_{it} is the number of years of experience. In both settings, the design matrix W_i has $T + 2$ columns and β_i stacks $\delta_i^{\alpha_0}$, $\delta_i^{\alpha_1}$, and the latent states. The prior for β_i combines the S&S prior for δ_i^α with the state-transition equation (8) for the latent states.

MCMC Convergence Diagnostics. The posterior sampler generates draws of three parameter groups and the latent states: (i) the homogeneous parameters θ, (ii) the discrepancies δ_i^α, δ_i^ρ, $\delta_i^{\sigma_u}$, and $\delta_i^{\sigma_\epsilon}$, (iii) the sparsity indicators z_i^l, $l \in \{\alpha, \rho, \sigma_u, \sigma_\epsilon\}$, and (iv) the latent state sequences $S_{i,0:T}$. The convergence of the posterior sampler can be assessed by running multiple MCMC chains from overdispersed starting values. For parameter group (i) the results can be visualized by creating figures that overlay trace plots or recursive means generated from the multiple runs. For the other three groups it is impractical to inspect trace plots and recursive means because one would typically have a large number of idiosyncratic parameters. As an alternative, we recommend reporting histograms of Gelman–Rubin $\hat{R}$ statistics (Gelman and Rubin, 1992), which pool the convergence information across units i. We provide an illustration in Online Appendix B.

Bundling. In our baseline specification we assume that membership in the core group versus the deviator group is independent across α, ρ, σ_u, and σ_ϵ, which is the specification

that we use in the Monte Carlo study in Section 4 and the empirical analysis of Section 5. It is possible to generalize the setup to allow for correlation of group memberships across parameter groups. We refer to a perfect correlation as bundling. For instance, one could replace the four sparsity indicators in (17) by two indicators, one for the measurement equation (ME) and one for the state (transition) equation (SE):

$$z_i^{\text{ME}} \in \{0,1\} \text{ for } \left(\delta_i^{\alpha}, \delta_i^{\sigma_u}\right), \qquad z_i^{\text{SE}} \in \{0,1\} \text{ for } \left(\delta_i^{\rho}, \delta_i^{\sigma_\epsilon}\right). \tag{21}$$

Conditional on the bundle indicators, we let

$$\left.\begin{array}{l} \delta_i^{\alpha} = 0, \;\; \delta_i^{\sigma_u} = 1 \end{array}\right\} \text{ if } z_i^{\text{ME}} = 0, \qquad \left.\begin{array}{l} \delta_i^{\alpha} \sim \mathcal{N}(0, v_{\delta^\alpha}) \\ \delta_i^{\sigma_u} \sim IG\left(\frac{\nu_{\delta^{\sigma_u}}}{2}, \frac{\tau_{\delta^{\sigma_u}}}{2}\right) \end{array}\right\} \text{ if } z_i^{\text{ME}} = 1, \tag{22}$$

and symmetrically for z_i^{SE} with $(\delta_i^{\rho}, \delta_i^{\sigma_\epsilon})$. A unit therefore cannot deviate in the mean while remaining in the core group for the variance of the same equation. The four slab widths $v_{\delta^\alpha}, v_{\delta^{\sigma_u}}, v_{\delta^\rho}, v_{\delta^{\sigma_\epsilon}}$ could be kept distinct; only the sparsity probabilities collapse.

We modified the posterior sampler to account for bundling – it becomes substantially more difficult to sample from the conditional posterior distribution $(z_i^{\text{ME}}, \delta_i^{\alpha}, \delta_i^{\sigma_u}, s_{i,1:T})$ – and studied it in two unreported experiments. First, we extended the Monte Carlo analysis of Section 4 to include simulations in which the data are generated from the bundled specification, and we estimated both the bundled and the baseline specification. As one would expect, if the bundled specification is correct, its estimation leads to lower RMSEs for the unit-level coefficients than the baseline specification. However, the RMSE reductions tend to be small: in most cases they are less than 2%, and the largest reduction is 15%.

In the empirical application of Section 5 bundling of mean and variance parameter groups was not helpful. Our data set has a lot of heteroskedasticity, which means that the variance core groups are essentially empty. On the other hand, for the persistence parameter ρ heterogeneity is very sparse. Because δ_i^{ρ} and $\delta_i^{\sigma_\epsilon}$ share an indicator under bundling, almost every unit becomes a persistence deviator, which led to a deterioration of the forecasting performance.

4 Monte Carlo Simulations

In the Monte Carlo experiment we generate panels of observations $Y_{1:N,1:T}$ from the panel state-space model described in Section 3 and compute estimates

$$\hat{l}_i(Y_{1:N,1:T}) \quad \text{of} \quad l_i = l + \delta_i^l, \tag{23}$$

where $l \in \{\alpha_0, \alpha_1, \rho\}$, with $\delta_i^\alpha = [\delta_i^{\alpha_0}, \delta_i^{\alpha_1}]'$. We evaluate the estimates based on the quadratic compound loss function

$$\frac{1}{N}\sum_{i=1}^{N}\left(\hat{l}_i(Y_{1:N,1:T}) - l_i\right)^2. \tag{24}$$

The Monte Carlo experiment is based on the following algorithm:

Algorithm 1 (Monte Carlo Simulation)
Given $[\alpha', \rho, q^\alpha, q^\rho, q^{\sigma_u}, q^{\sigma_\epsilon}, v_{\delta^\alpha}, v_{\delta^\rho}, v_{\delta^{\sigma_u}}, v_{\delta^{\sigma_\epsilon}}, \mu_{s_0}, v_{s_0}]'$ *and* λ*, the following three steps are repeated* N_{sim} *times:*

1. **Parameter Draws.** *Draw* $\left\{\delta_i^\alpha, \delta_i^\rho, \delta_i^{\sigma_u}, \delta_i^{\sigma_\epsilon}\right\}_{i=1}^N$ *and* $\left\{\sigma_{u,t}^2, \sigma_{\epsilon,t}^2\right\}_{t=1}^T$ *from the priors in (2), (9), and (10).*

2. **Data Simulation.** *Draw a panel of iid error terms* u_{it} *and* ϵ_{it} *for* $i = 1, \ldots, N$ *and* $t = 1, \ldots, T$ *from the* $\mathcal{N}(0,1)$ *distribution and generate* $y_{i1:T}$ *and* $s_{i1:T}$ *based on (7) and (8), starting from the initial values* $s_{i0} = 0$*,* $i = 1, \ldots, N$ *and using the parameter values generated in Step 1. The regressor vector is specified as* $x_{it} = [1\ t]'$*, where* 1 *denotes the intercept and* t *is the time period.*[5]

3. **Estimation and Loss Calculation.** *Compute the estimators* $\hat{l}_i(Y_{1:N,1:T})$ *for* $l \in \{\alpha_0, \alpha_1, \rho\}$ *and evaluate the compound loss (24).*

Estimators. We compare the performance of four posterior mean estimators. The first one is the spike-and-slab (S&S) estimator defined as

$$\hat{l}_i^* = \mathbb{E}[l + \delta_i^l \mid Y_{1:N,1:T}, \lambda].$$

This estimator treats the slab probabilities as unknown, but conditions on the hyperparameters λ. Second, the estimator $\hat{l}_i^0$ imposes $q^\alpha = q^\rho = 0$ and thereby imposes homogeneity in

[5] In the application of Section 5 we will use the number of years of work experience.

α and ρ. Third, we compute a fully heterogeneous estimator $\hat{l}_i^1$ by imposing $q^\alpha = q^\rho = 1$. Finally, we report the performance of the oracle estimator, which is based on knowledge of the "true" parameter vector θ:

$$\hat{l}_i^o = \mathbb{E}[l + \delta_i^l \mid Y_{1:N,1:T}, \theta].$$

Calibration of the Monte Carlo Experiment. In Section 5 we will fit the panel state-space model to PSID income data. Thus, we parameterize the panel model so that it resembles income dynamics in the U.S. Baseline choices for the parameter vector θ and the hyperparameters λ are summarized in Table 1. The values for θ and λ are derived from the estimation results of the HIP process presented in Table 1 of Guvenen (2009). In the Monte Carlo DGP, random draws are generated from calibrated distributions indexed by θ, whereas λ only indexes the priors used at the estimation step. Starting with the elements of θ, we fix $\alpha_0 = 0.5$, $\alpha_1 = 0.05$, and set $\rho = .821$. In the experiments below we consider various values for $q^\alpha = q^\rho = q^{\sigma_u} = q^{\sigma_\epsilon} = q$ and for a scalar c in

$$v_{\delta^\alpha} = c \cdot \begin{bmatrix} .2200 & -.0067 \\ -.0067 & .0038 \end{bmatrix}. \tag{25}$$

Note that $\text{corr}(\delta_i^{\alpha_0}, \delta_i^{\alpha_1}) = -0.23$. These choices control the height of the spike and the width of the slab of the δ_i^α distribution. The parameter v_{δ^ρ} is set to 0.04.[6] We let $v_{\delta^{\sigma_u}} = \mathbb{V}[\delta_i^{\sigma_u}] = 1$ and $v_{\delta^{\sigma_\epsilon}} = \mathbb{V}[\delta_i^{\sigma_\epsilon}] = 1$.

Turning to the elements of λ, we set $\underline{\mu}_\alpha = 0_{2\times 1}$ and $\underline{\mu}_\rho = 0.8$. The variance of α, denoted by $\underline{v}_\alpha$, is equal to an identity matrix and we let $\underline{v}_\rho = 1$. The hyperparameters for the prior for $\sigma_{u,t}^2$ and $\sigma_{\epsilon,t}^2$ imply that $\mathbb{E}[\sigma_{u,t}^2] = \mathbb{E}[\sigma_{\epsilon,t}^2] = 0.05$ and $\mathbb{V}[\sigma_{u,t}^2] = \mathbb{V}[\sigma_{\epsilon,t}^2] = 0.05^2$. We choose (a, b) such that $q \sim U[0, 1]$. Our choice of $\underline{\nu}_{\delta^\alpha}$ and $\underline{\Psi}_{\delta^\alpha}$ implies

$$\mathbb{E}[v_{\delta^\alpha}] = \begin{bmatrix} 0.24 & 0 \\ 0 & .048 \end{bmatrix}, \quad \mathbb{V}[v_{\delta^\alpha}] = \begin{bmatrix} 2.38 & 0.16 \\ 0.16 & 0.10 \end{bmatrix}.$$

This distribution covers the range of v_{δ^α} values considered in the rows of Table 2. Our choice of $\nu_{\delta^\rho} = 4.125$ and $\tau_{\delta^\rho} = .531$ implies $\mathbb{E}[v_{\delta^\rho}] = 1/4$ and $\mathbb{V}[v_{\delta^\rho}] = 1$. The true value of $v_{\delta^\rho} = 0.04$ can be viewed as a draw from the left tail of this distribution. According to the

[6] In the DGP, draws of δ_i^ρ are truncated at 0.18, i.e., $\delta_i^\rho = \min\{\delta_i^\rho, 0.18\}$, to avoid explosive persistence in simulated state dynamics.

Table 1: Parameter Values for Monte Carlo Simulation

Parameter Vector θ

α_0	α_1	ρ	q	v_{δ^α}	v_{δ^ρ}	$v_{\delta^{\sigma_u}}$	$v_{\delta^{\sigma_\epsilon}}$
0.5	0.05	.821	various values	0.04	1	1

Hyperparameter Vector λ

$\underline{\mu}_\alpha$	$\underline{v}_\alpha$	$\underline{\mu}_\rho$	$\underline{v}_\rho$	$(\underline{\nu}_{\sigma_u}, \underline{\tau}_{\sigma_u})$	$(\underline{\nu}_{\sigma_\epsilon}, \underline{\tau}_{\sigma_\epsilon})$	(a, b)	$(\underline{\nu}_{\delta^\alpha}, \underline{\Psi}_{\delta^\alpha})$
$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$	0.8	1	(6, 0.2)	(6, 0.2)	(1.0, 1.0)	$\left(5.05, \begin{bmatrix}0.5 & 0\\0 & 0.1\end{bmatrix}\right)$

$(\underline{\nu}_{\delta^\rho}, \underline{\tau}_{\delta^\rho})$	$(\underline{\nu}_{\delta^{\sigma_u}}, \underline{\tau}_{\delta^{\sigma_u}})$	$(\underline{\nu}_{\delta^{\sigma_\epsilon}}, \underline{\tau}_{\delta^{\sigma_\epsilon}})$	$(\underline{\mu}_{s_0}, \underline{v}_{s_0})$	$(\underline{\nu}_{s_0}, \underline{\tau}_{s_0})$
(4.125, .531)	(12, 10)	(12, 10)	(0, 0.05)	(6, 0.2)

hyperparameters for the $v_{\delta^{\sigma_u}}$ and $v_{\delta^{\sigma_\epsilon}}$ priors, $\mathbb{E}[v_{\delta^{\sigma_u}}] = \mathbb{E}[v_{\delta^{\sigma_\epsilon}}] = 1$ and $\mathbb{V}[v_{\delta^{\sigma_u}}] = \mathbb{V}[v_{\delta^{\sigma_\epsilon}}] = 1/4$. The hyperparameters for the prior for v_{s_0} imply that $\mathbb{E}[v_{s_0}] = 0.05$ and $\mathbb{V}[v_{s_0}] = 0.05^2$.

Monte Carlo Results. Using Algorithm 1, we generate $N_{\text{sim}} = 200$ panel data sets of size $N = 500$ and $T = 10$. For each panel data set we generate 5,000 parameter draws from the posterior distribution, discard the first 2,500 draws, and use the remaining draws to compute posterior mean estimates $\hat{l}_i(Y_{1:N,1:T})$. Simulation results are summarized in Table 2. The top panel reports the root mean squared errors (RMSEs) for the intercept estimator $\hat{\alpha}_{i0}$ (where $\alpha_{i0} = \alpha_0 + \delta_i^{\alpha_0}$) across various combinations of (v_{δ^α}, q). We are averaging the compound loss in (24) across the N_{sim} Monte Carlo repetitions, before taking the square root. The middle and bottom panels report the corresponding RMSEs for the time trend coefficient $\hat{\alpha}_{i1}$ and the autoregressive parameter $\hat{\rho}_i$, respectively. During estimation we do not restrict q^α to be equal to q^ρ, q^{σ_u}, or q^{σ_ϵ}. Thus, members of the core α and ρ groups are not required to be the same. Formally, we do not impose the restriction $z_i^\alpha = z_i^\rho = z_i^{\sigma_u} = z_i^{\sigma_\epsilon}$ for all i.

For $q = 0$ the Bayes estimator with S&S prior delivers RMSEs that are very close to those of the estimator that imposes coefficient homogeneity ($q^\alpha = q^\rho = 0$), and for $q = 0.8$ and $q = 1$ its RMSEs are generally close to those of the estimator that treats l_i as fully heterogeneous ($q^\alpha = q^\rho = 1$). For values of $q \in \{0.2, 0.4, 0.6\}$ the S&S estimator typically

Table 2: Monte Carlo Results

	Estimators	$q^\alpha = q^\rho = q^{\sigma_u} = q^{\sigma_\epsilon} = q$					
		0	0.2	0.4	0.6	0.8	1.0
$c(v_{\delta^\alpha})$	RMSE for Estimates $\hat{\alpha}_{i0}$ of $\alpha_{i0} = \alpha_0 + \delta_i^{\alpha_0}$						
0.1	Oracle	0.000	0.642	0.872	1.036	1.158	1.248
	S&S	0.136	0.677	0.931	1.136	1.312	1.452
	$q^\alpha = q^\rho = 0$	0.127	0.681	0.946	1.160	1.337	1.491
	$q^\alpha = q^\rho = 1$	1.150	1.186	1.217	1.259	1.304	1.345
0.5	Oracle	0.000	1.178	1.533	1.759	1.905	1.991
	S&S	0.136	1.292	1.706	1.768	1.929	2.347
	$q^\alpha = q^\rho = 0$	0.127	1.501	2.102	2.584	2.981	3.325
	$q^\alpha = q^\rho = 1$	1.150	1.450	1.674	1.837	1.945	2.015
1.0	Oracle	0.000	1.370	1.772	2.018	2.173	2.252
	S&S	0.136	1.418	1.772	2.118	2.227	2.736
	$q^\alpha = q^\rho = 0$	0.127	2.120	2.970	3.652	4.214	4.701
	$q^\alpha = q^\rho = 1$	1.150	1.742	2.065	2.170	2.235	2.281
$c(v_{\delta^\alpha})$	RMSE for Estimates $\hat{\alpha}_{i1}$ of $\alpha_{i1} = \alpha_1 + \delta_i^{\alpha_1}$						
0.1	Oracle	0.000	0.086	0.119	0.144	0.164	0.181
	S&S	0.029	0.103	0.118	0.146	0.173	0.304
	$q^\alpha = q^\rho = 0$	0.024	0.091	0.125	0.152	0.176	0.320
	$q^\alpha = q^\rho = 1$	0.264	0.266	0.270	0.275	0.280	0.285
0.5	Oracle	0.000	0.177	0.238	0.282	0.314	0.337
	S&S	0.029	0.191	0.258	0.312	0.341	0.412
	$q^\alpha = q^\rho = 0$	0.024	0.197	0.275	0.338	0.391	0.436
	$q^\alpha = q^\rho = 1$	0.264	0.279	0.301	0.324	0.345	0.363
1.0	Oracle	0.000	0.224	0.299	0.349	0.386	0.409
	S&S	0.029	0.231	0.304	0.358	0.389	0.472
	$q^\alpha = q^\rho = 0$	0.024	0.278	0.389	0.477	0.552	0.617
	$q^\alpha = q^\rho = 1$	0.264	0.301	0.341	0.372	0.397	0.417
$c(v_{\delta^\alpha})$	RMSE for Estimates $\hat{\rho}_i$ of $\rho_i = \rho + \delta_i^\rho$						
0.1	Oracle	0.000	0.746	1.042	1.262	1.445	1.607
	S&S	0.145	0.807	1.138	1.547	1.811	2.322
	$q^\alpha = q^\rho = 0$	0.151	0.865	1.268	1.595	1.881	2.354
	$q^\alpha = q^\rho = 1$	2.186	2.165	2.168	2.188	2.219	2.255
0.5	Oracle	0.000	0.754	1.060	1.292	1.490	1.664
	S&S	0.145	0.856	1.218	1.876	1.953	2.241
	$q^\alpha = q^\rho = 0$	0.151	1.059	1.565	1.939	2.238	2.488
	$q^\alpha = q^\rho = 1$	2.186	1.972	1.900	1.917	1.976	2.059
1.0	Oracle	0.000	0.758	1.067	1.304	1.505	1.683
	S&S	0.145	0.801	1.111	1.708	1.846	2.173
	$q^\alpha = q^\rho = 0$	0.151	1.233	1.771	2.132	2.404	2.626
	$q^\alpha = q^\rho = 1$	2.186	1.859	1.706	1.752	1.864	1.996

Notes: The simulation is based on $N = 500$, $T = 10$, and $N_{\text{sim}} = 200$. The choices for (θ, λ) are summarized in Table 1. α_0 is the intercept and α_1 is the time trend coefficient. The constant c scales the covariance matrix v_{δ^α}; see (25).

achieves lower RMSEs than the estimators that impose full homogeneity or full heterogeneity. Overall, $\hat{l}_i^*$ is competitive with the better of $\hat{l}_i^0$ and $\hat{l}_i^1$ across configurations, although it does not uniformly dominate both. The risk differentials between the homogeneous and fully heterogeneous estimators are increasing in v_{δ^α}, and the risk patterns are similar for the estimation of α_{i0}, α_{i1}, and ρ_i. As a benchmark, we also report the risk of the oracle estimator that is obtained by conditioning on the "true" θ. By construction the oracle estimator dominates the Bayes estimator $\hat{l}_i^*$, but overall the risk differentials are relatively small, which means that θ is generally well estimated from the cross-sectional information.

In the top three rows of Figure 1 we plot histograms of the cross-sectional distribution of the posterior mean estimates $\hat{\alpha}_{i0}^*$, $\hat{\rho}_i^*$, and $\hat{\delta}_i^{\sigma_u}$ obtained conditional on data sets simulated with $q = 0.2$ and $q = 0.8$, respectively. Unlike the results reported in Table 2, the histograms are generated from a single Monte Carlo iteration and not averaged across N_{sim} runs. For $q = 0.2$ the histogram of posterior means inherits the spike-and-slab shape of the prior distribution with a distinct spike near the "true" values of α_0, ρ, and $\delta_i^{\sigma_u}$, which are 0.5, .821, and 1, respectively; see Table 1. For $q = 0.8$ the spike in the distribution of posterior means is much less pronounced and there is a lot more mass away from $\alpha_0 = 0.5$, $\rho = .821$, and $\delta_i^{\sigma_u} = 1$. In the last row we plot the posterior densities for the slab probabilities q^α, q^ρ, q^{σ_u}, and q^{σ_ϵ}. While there is considerable uncertainty about q^l, the densities peak near the "true" values of 0.2 and 0.8, respectively. We do not show the histograms for $\hat{\alpha}_{i1}^*$ and $\hat{\delta}_i^{\sigma_\epsilon}$ as they are similar to those for $\hat{\alpha}_{i0}^*$ and $\hat{\delta}_i^{\sigma_u}$, respectively.

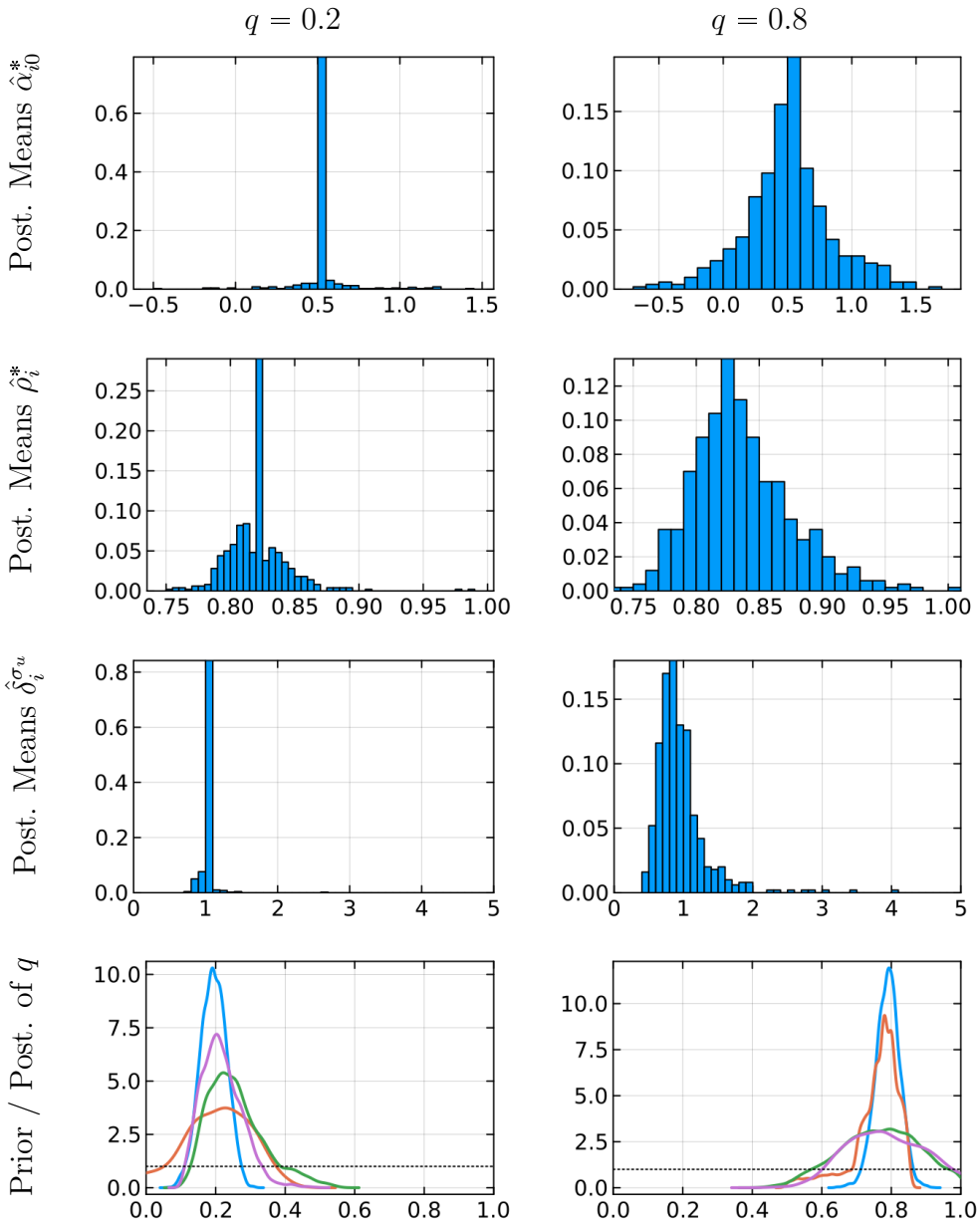
5 Empirical Analysis

We now use the panel state-space model, comprising (7) and (8), to estimate income profiles from the PSID data. The model captures some of the key features of the empirical specifications considered in the literature over the past four decades; see for instance the model specification discussion in the survey paper Browning and Ejrnæs (2013). These features include a deterministic income component $x_{it}'\alpha$, and persistent and transitory stochastic components, s_{it} and u_{it}, respectively. We allow for potential heterogeneity in the α vector, the autocorrelation ρ of the persistent component, and the shock standard deviations.

More specifically, we use h_{it} to denote experience of individual i in period t, defined as age minus years of education minus six. We rescale experience, let

$$x_{it} = [1, h_{it}/10]', \tag{26}$$

Figure 1: Histograms of Posterior Means $\hat{\alpha}^*_{i0}$, $\hat{\rho}^*_i$, and $\hat{\delta}^{\sigma_u}_i$, and Posterior Densities of q



Notes: We adjust the posterior estimates for δ^l_i in the following way: if more than 80% of posterior draws for δ^l_i are 0 ($l = \alpha, \rho$) or 1 ($l = \sigma_u, \sigma_\epsilon$), then we set $\hat{\delta}^l_i = 0$ ($l = \alpha, \rho$) or 1 ($l = \sigma_u, \sigma_\epsilon$). The plots in the last row overlay a $B(1,1)$ prior density for q^l (black, dotted) and posterior densities for q^α (blue, solid), q^ρ (red, solid), q^{σ_u} (green, solid), and q^{σ_ϵ} (violet, solid). The choices for the hyperparameter vector are summarized in Table 1.

and partition the 2×1 vector $\alpha = [\alpha_0, \alpha_1]'$. If $\delta_i^\alpha = 0$ for all i, then income profiles are homogeneous or restricted, which we abbreviate as RIP in the introduction. This has been the preferred specification of Lillard and Weiss (1979), Hause (1980), Baker (1997), Haider (2001), Hryshko (2012), and Hoffmann (2019). The case of $\delta_i^\alpha \neq 0$ for all individuals corresponds to the HIP specification considered in the literature and is favored by, for instance, MaCurdy (1982), Abowd and Card (1989), and Guvenen (2007, 2009). Unlike the previous literature, our model considers a third option, under which heterogeneity is potentially sparse: $\delta_i^\alpha \neq 0$ for some, but not all, individuals.

Income has a persistent and a transitory stochastic component. The persistent component is given by an AR(1) process s_{it} with autocorrelation coefficient ρ and the transitory component is represented by the iid process u_{it}. Through $\delta_i^{\sigma_u}$, $\delta_i^{\sigma_\epsilon} \neq 1$ we allow income risk to be heterogeneous across agents. In addition, the standard deviations $\sigma_{u,t}$ and $\sigma_{\epsilon,t}$ are time dependent, which can capture business cycle fluctuations, e.g., higher income risk during recessions, and trends in idiosyncratic risk. Finally, we allow for idiosyncratic autocorrelations ρ_i, something that is typically not done in the literature; exceptions are Browning, Ejrnæs, and Alvarez (2010) and Browning and Ejrnæs (2013).

There are a few model features that have been explored in the literature that we do not include in our model specification to preserve parsimony: a third stochastic component that features a unit root, idiosyncratic autoregressive conditional heteroskedasticity (see, for instance, Meghir and Pistaferri (2004) for both features), time variation in the returns to experience (e.g., Haider (2001)), or age effects in the variances of the idiosyncratic shocks (e.g., Hoffmann (2019)). We consider three restricted versions of our model: one without heteroskedasticity (Homosk), and the RIP and HIP specifications mentioned above. A summary is provided in Table 3. Some authors have emphasized cohort effects on earnings, education effects on earnings profiles, or age-dependent persistence in earnings. While we are not explicitly controlling for this type of heterogeneity through sample splitting or interaction terms, it will be reflected in the posterior estimates of δ_i^l.

Much of the previous literature has used (simulated) minimum-distance estimation of a subset of the parameters that we collect in the vector θ, which involves minimizing the discrepancy between sample moments and model-implied moments. Typically, these moments are time-dependent autocovariances that are computed by cross-sectional averaging across all individuals in the panel. In a few instances, researchers have also used moments computed from unit-specific time series; see Browning and Ejrnæs (2013). Our likelihood-based estimation approach aims at tracking individual earnings histories, which implicitly

Table 3: Restricted Model Specifications

Label	Restrictions
Homosk	$q^{\sigma_u} = q^{\sigma_\epsilon} = 0$; $\delta_i^{\sigma_u} = \delta_i^{\sigma_\epsilon} = 1$ for $i = 1, \ldots, N$
RIP	$q^{\alpha} = q^{\rho} = 0$; $\delta_i^{\alpha} = \delta_i^{\rho} = 0$ for $i = 1, \ldots, N$
HIP	$q^{\alpha} = q^{\rho} = 1$

encompasses both time series and cross-sectional moments. This approach has several advantages. In particular, we are easily able to handle non-Normalities introduced by the S&S prior and the cross-sectional heteroskedasticity. Moreover, in addition to parameters that characterize the distribution of coefficients across agents, we can also make inference about the unit-specific coefficients themselves, and subsequently generate predictions about future income for individual units.

The remainder of this section is organized as follows. The construction of the data set is described in Section 5.1 and the prior distribution for the Bayesian analysis is summarized in Section 5.2. The estimation results are presented in Section 5.3. Using balanced panels of individuals with a long employment history, we document that heterogeneity with respect to returns to experience and the autocorrelation of the persistent stochastic income component is sparse, with large core groups of individuals sharing identical coefficients and a small number of deviators. For a large unbalanced panel, we provide evidence for full coefficient heterogeneity, with the caveat that the specification with homogeneous coefficients delivers slightly more accurate one-step-ahead point forecasts, while its density forecasts are less accurate. Section 5.4 examines earnings forecasts for specific individuals conditional on information from the full panel $Y_{1:N,1:T}$ accounting for shock and parameter uncertainty, the full panel ignoring parameter uncertainty, and the individual history $Y_{i,1:T}$ accounting for shock and parameter uncertainty. Parameter heterogeneity affects the extent to which an individual can exploit information from others to improve their earnings forecast. Additional empirical results are relegated to Online Appendix C.

5.1 Data Set

The empirical analysis is based on the PSID. We use raw data from 1968 to 1997, starting from 32,465 observations for 2,052 individuals. Up to 1997 the PSID data are annual, whereas

subsequently the survey was conducted only biennially. The data set used for estimation is constructed following the conventions of the literature, e.g., Meghir and Pistaferri (2004). The dependent variable is log real income from all sources, including: wages and salaries; bonuses, overtime, and/or commissions; income from professional practice and trade; the labor part of farm income and unincorporated business income; the labor portion of income from market gardening, roomers, and boarders. Nominal wages are converted into real wages using the PCE deflator (base = 1993). We take account of the fact that the measure of income refers to the previous calendar year. We eliminate observations with outlying earnings records, defined as a change in log real earnings greater than two or less than minus one.

As is common in the literature, we restrict the sample to male heads of household and keep individuals between ages 25 and 55.[7] To avoid oversampling low-income households, members from the Survey of Economic Opportunity (SEO) are removed. We also drop the members of the Latino sample added in 1990 and self-employed individuals. We keep individuals who are in the labor force, have positive labor income, and have been in the sample for at least nine years. We chose the nine-year cutoff so that we can construct unit-specific time-series estimates that carry some information about the coefficients and about the decomposition of income into a persistent and a transitory component. We remove individuals with missing race, marital status, region, state, and education information. In a first-stage regression we project log income on year dummies and indicators for race, educational group,[8] region of residence, and marital status. The variable y_{it} is the residual from this regression. We refer to the resulting panel data set as the full sample.

The estimated heterogeneity and the prevalence of sparse heterogeneity depend on the construction of the data set. We begin with the construction of balanced panels, denoted by $\mathcal{B}_{T,\tau}$. The balanced panels are studied in the first part of Section 5.3 and in Section 5.4. They have a time dimension of $T+1$ and include only individuals for whom $T+1$ observations are available. We use the first T time periods in these samples for estimation and reserve the last time period, $t = T+1$, for one-step-ahead forecast evaluation. The index τ refers to the year that corresponds to period $T+1$. For instance, the sample $\mathcal{B}_{20,1991}$ starts in 1971 ($t = 1$) and ends in 1991 ($t = 20+1$). For the balanced panel analysis we choose $T = 20$

[7] If the household head's age is inconsistent across years, we compute the mode of the differences between reported age and year across periods. Age is then defined as the year plus the mode.

[8] High School dropouts (those with less than 12 grades of schooling); High School graduates (those with at least a High School diploma, but no College degree); College graduates (those with a College degree or more).

Table 4: Prior Distribution for θ

Para	Distr.	a	b	5%	95%
α	$\mathcal{N}(a,b)$	$0_{2\times1}$	$\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$	-1.64 -1.64	1.64 1.64
ρ	$\mathcal{N}(a,b)$	0.8	1	-0.84	2.44
$\sigma^2_{u,t}$	$IG\big(a/2, b/2\big)$	6	0.2	0.02	0.12
$\sigma^2_{\epsilon,t}$	$IG\big(a/2, b/2\big)$	6	0.2	0.02	0.12
$q^l,\ l \in \{\alpha, \rho, \sigma_u, \sigma_\epsilon\}$	$B(a,b)$	1	1	0.05	0.95
v_{δ^α}	$IW(a,b)$	5.05	$\begin{bmatrix}0.5 & 0\\ 0 & 0.1\end{bmatrix}$	0.05 0.01	0.68 0.14
v_{δ^ρ}	$IG\big(a/2, b/2\big)$	16.5	3.625	0.13	0.44
$v_{\delta^{\sigma_m}},\ m \in \{u, \epsilon\}$	$IG\big(a/2, b/2\big)$	12	10	0.48	1.91
μ_{s_0}	$\mathcal{N}(a,b)$	0	0.05	-0.37	0.37
v_{s_0}	$IG\big(a/2, b/2\big)$	6	0.2	0.02	0.12

Notes: $\mathcal{N}(a,b)$ is the Normal distribution, $IG(a,b)$ is the Inverse Gamma distribution, $B(a,b)$ is the Beta distribution, and $IW(a,b)$ is the Inverse Wishart distribution. The last two columns contain the 5th and 95th percentile of the marginal prior distributions. For IG priors: $(a,b) = (6, 0.2)$ implies a mean of 0.05 and a standard deviation of 0.05; $(a,b) = (16.5, 3.625)$ implies a mean of 0.25 and a standard deviation of 0.1; $(a,b) = (12, 10)$ implies a mean of 1 and a standard deviation of 0.5.

to focus on individuals who have fairly long employment histories. *A priori* we expect that this group of individuals is fairly homogeneous.

In the last part of Section 5.3 we re-estimate the panel state-space model based on an unbalanced panel $\mathcal{U}$, which by construction has a lot more observational heterogeneity. Individuals are required to have at least nine *consecutive observations*, but are not required to have observations in each period t. As in the balanced panels, the last observation of each individual is held out for forecast evaluation. Thus, the shortest unit-specific estimation sample consists of eight observations. Our unbalanced panel covers approximately 500 units in 1968. The number of units rises above 1,100 in 1988 and then declines to 800 in 1995. The modal duration in the panel is 10 to 12 years. About 200 units stay in the panel for the minimum of nine years and approximately 330 units stay for more than 20 years. *A priori*, we expect coefficient heterogeneity to be stronger in the unbalanced sample $\mathcal{U}$ than in the balanced samples $\mathcal{B}_{T,\tau}$.

5.2 Prior Distribution

Table 4 summarizes the marginal prior distributions for the elements of θ. We are assuming that the elements are *a priori* independent, such that the joint distribution is the product of the marginals. In addition to the (a, b) parameterization of the prior, we also report the 5th and 95th percentiles, which delimit the 90% equal-tail probability credible intervals. In some dimensions, the prior is specified relative to the scale of the data. If y_{it} or x_{it} were rescaled, the prior distribution would have to be rescaled accordingly.

The prior for α is centered at zero with a variance of one. Recall that we divided experience by ten, which means that a value of $\alpha_1 = 0.2$ implies a return to a year of experience of $0.2/10 = 0.02$, or two percent. Thus, the 90% *a priori* credible interval covers all reasonable and some unreasonable values for the return to experience. The marginal distribution of ρ is centered at 0.8, representing an *a priori* belief that the idiosyncratic income process s_{it} is fairly persistent. The variance of one implies that the prior does not have a lot of curvature in the range of 0.6 to 1.0, which covers the estimates obtained in previous studies. We do allow for explosive dynamics *a priori*, but such dynamics turn out to be extremely unlikely under the posterior.[9]

The 90% credible interval for the innovation variances reaches from 0.02 to 0.12. Note that our prior implies that they are independent across time periods, because in our panel data set we want to identify the variances from the cross-sectional information. Our prior for μ_{s_0} and v_{s_0} implies that s_{i0} has mean zero, which is the long-run mean of s_{it} provided that ρ is less than one. Moreover, s_{i0} falls in the interval -1.0 to 1.0 with high probability.

The prior for the probabilities of not belonging to the core group, q^l, is uniform on the unit interval. The prior for the deviation variance v_{δ^α} implies that the expected values of the diagonal elements of the α deviations covariance matrix are 0.24 and .048, respectively. At .048, the 90% credible interval for $\delta_i^{\alpha_1}/10$ ranges from roughly -.036 to .036.[10] The prior for $v_{\delta^{\sigma_m}}$, $m \in \{u, \epsilon\}$, implies that an individual at the 5th percentile under the prior has an idiosyncratic shock variance that is one fourth of that of someone at the 95th percentile.

[9] As a robustness exercise, we re-estimated the model under the prior $\rho \sim \mathcal{N}(0.7, 0.1^2)$, which assigns a probability of only 0.001 to $\rho > 1$. We comment on the posterior estimates that it generates in Section 5.3.

[10] The credible interval for the off-diagonal element of v_{δ^α} ranges from -0.1 to 0.1.

5.3 Parameter Estimates

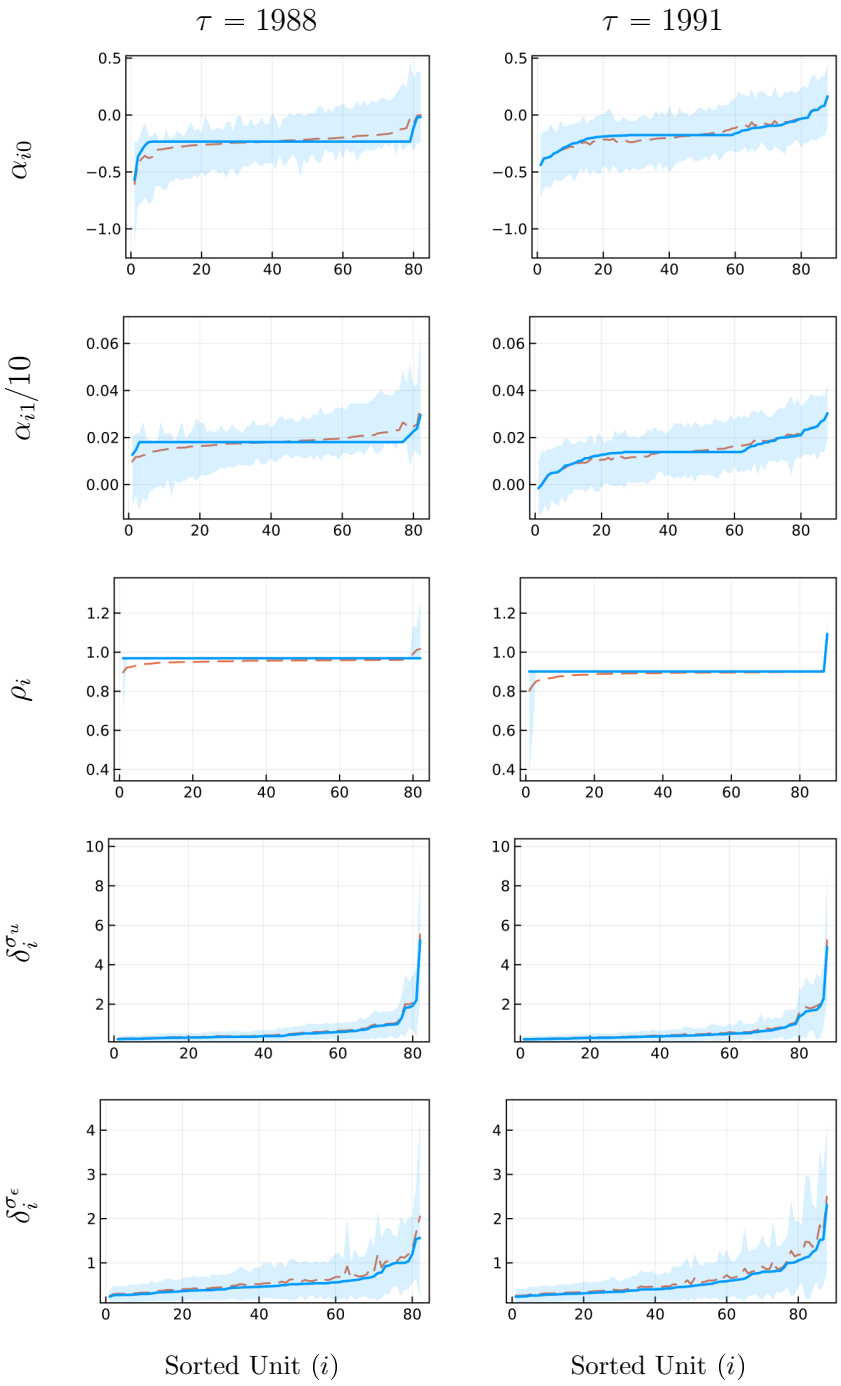
For each of the balanced panels and for the unbalanced panel, we generate 10,000 draws from the posterior distribution and discard the first 5,000. The chains are longer than in the Monte Carlo study of Section 4 (5,000 draws), where the computational cost of repeating the estimation for many simulated panels is the binding constraint.

Idiosyncratic Parameters (Balanced Panels). We denote the idiosyncratic parameters by $\alpha_i = \alpha + \delta_i^{\alpha} = [\alpha_{i0}, \alpha_{i1}]'$, where $\alpha_{i0} = \alpha_0 + \delta_i^{\alpha_0}$ and $\alpha_{i1} = \alpha_1 + \delta_i^{\alpha_1}$, and by $\rho_i = \rho + \delta_i^{\rho}$. The unit-specific shock variances are $\sigma_{u,t}^2 \delta_i^{\sigma_u}$ and $\sigma_{\epsilon,t}^2 \delta_i^{\sigma_\epsilon}$. Of particular interest are the estimates of the idiosyncratic parameters and the degree of heterogeneity they exhibit. The results are summarized in Figure 2. The figure shows posterior estimates for α_{i0}, α_{i1}, ρ_i, $\delta_i^{\sigma_u}$, and $\delta_i^{\sigma_\epsilon}$ for two samples $\mathcal{B}_{T,\tau}$. The panels in the left column correspond to $\tau = 1988$ and the panels in the right column to $\tau = 1991$. In both cases $T = 20$. We depict posterior medians (solid blue lines), means (dashed red lines), and 90% credible intervals (light blue bands). Units are sorted based on posterior median estimates. As discussed in Section 2, posterior median estimates of δ_i^l can deliver exact zeros. To highlight the potential sparsity implied by the posterior medians of δ_i^l, we plot $\text{med}(l|Y_{1:N,1:T}) + \text{med}(\delta_i^l|Y_{1:N,1:T})$ instead of $\text{med}(l + \delta_i^l|Y_{1:N,1:T})$ in the first three rows for $l \in \{\alpha_0, \alpha_1, \rho\}$.

Sparse heterogeneity manifests itself through a long flat section of the posterior median line. For instance, the $\tau = 1988$ sample consists of 82 individuals. The number of units with $\text{med}(\delta_i^{\alpha}|Y_{1:N,1:T}) = 0$ is 74 for α. Thus, the α_is feature sparse heterogeneity. From examining the posterior distribution of the α_{i1}s, which are divided by 10 to undo the rescaling of experience in (26), we find that some units have a return to experience that is up to 0.5 percentage points less than that of the core group, whereas others have a return that is up to 1.5 percentage points above that of the core group. The $\tau = 1991$ sample has more heterogeneity in the α dimension, with only 32 out of 88 units having $\text{med}(\delta_i^{\alpha}|Y_{1:N,1:T}) = 0$. Unlike the posterior median curves, the posterior mean curves generally have no flat segments, because the posterior means of δ_i^{α} are never exactly equal to zero.

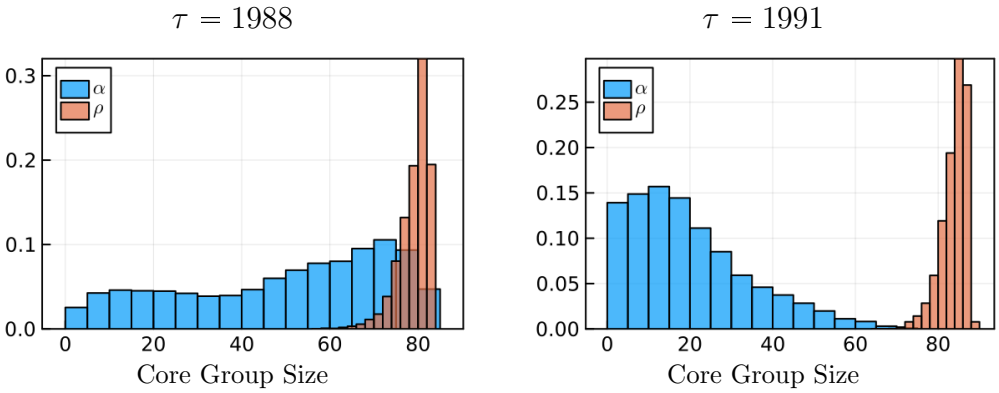
For the autocorrelation and the two shock variances a different picture emerges. For ρ essentially all units have $\text{med}(\delta_i^{\rho}|Y_{1:N,1:T}) = 0$, meaning that the autocorrelations are homogeneous, whereas for the variances essentially all of the units have $\text{med}(\delta_i^{\sigma_m}|Y_{1:N,1:T}) \doteq 1$, $m \in \{u, \epsilon\}$ (recall the different centering of the variance discrepancies). Thus, the shock variances $\sigma_{u,t}^2 \delta_i^{\sigma_u}$ and $\sigma_{\epsilon,t}^2 \delta_i^{\sigma_\epsilon}$ are fully heterogeneous, meaning there is strong cross-sectional heteroskedasticity, a point that is also emphasized, for instance, by Gu and Koenker (2017b).

Figure 2: Posterior of Idiosyncratic Parameters, $T = 20$



Notes: Solid blue lines in first three rows are $\text{med}(l|Y_{1:N,1:T}) + \text{med}(\delta_i^l|Y_{1:N,1:T})$. In the bottom two rows they are $\text{med}(\delta_i^{\sigma_u}|Y_{1:N,1:T})$ and $\text{med}(\delta_i^{\sigma_\epsilon}|Y_{1:N,1:T})$. The dashed red lines are posterior means and the light blue bands are 90% credible bands. Units are sorted based on posterior median estimates.

Figure 3: Posterior Histograms of the Sizes of α and ρ Core Groups, $T = 20$

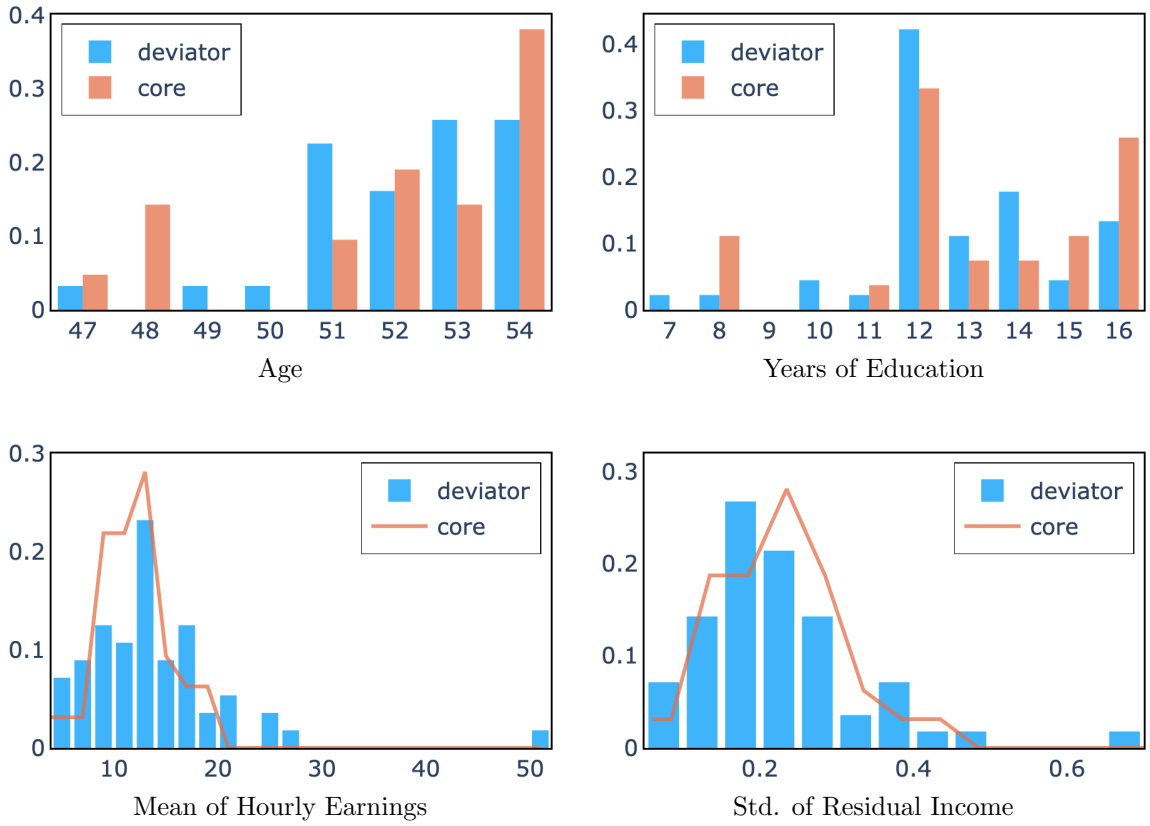


Notes: The number of units is $N = 82$ for the 1988 sample and $N = 88$ for the 1991 sample.

Figure 3 shows posterior histograms for the sizes of the α and ρ core groups. Recall that the indicator variable z_i^l, defined in (17), is 1 if unit i belongs to the group of deviators. Thus, $N - \sum_{i=1}^{N} z_i^l$ is the size of the core group for parameter $l \in \{\alpha, \rho, \sigma_u, \sigma_\epsilon\}$. Each draw from the posterior generates sequences $Z_{1:N}^l$ that allow us to evaluate the core group sizes and to generate the histograms. The posterior modes of the ρ core group sizes are 80 (out of 82 units in total) and 82 (out of 88 units in total), respectively. For the 1988 sample the posterior histogram for the size of the α group peaks at around 70, but the probability mass is spread out between zero (full heterogeneity) and 82 (homogeneity). In the 1991 sample, the probability mass for the α core group size shifts to the left and much of it lies between 0 and 30. We do not plot the distribution of the σ group sizes, as it is clear from Figure 2 that there are no core groups and the coefficients are fully heterogeneous.

How Are Deviators Different (Balanced Panels)? In Figure 4 we compare characteristics of the core group members and the deviators. The top panels show two histograms and the bottom panels show a histogram and a density estimate. In terms of the age distribution and years of education the core group and deviators are quite similar. The core group contains slightly fewer individuals below the age of 50 and not as many individuals of age 54, but if the information is aggregated to a bin width of two or three years, many of these differences would vanish. The same is true for years of education. The distribution of time series means (computed prior to projecting onto demographics) for the deviators and the core group also looks quite similar, except for the presence of a few high(er) income individuals among the deviators. A similar pattern emerges for the time series standard deviation of the residual income after projecting it onto the demographics. Overall, in terms of the distribution of age, years of education, mean earnings, and standard deviation of earnings

Figure 4: Histograms / Densities of Core Group vs. Deviator Characteristics, $\mathcal{B}_{20,1991}$



Notes: Top panels: bar charts represent histograms for core group and deviators. Bottom panels: histograms for deviators and density estimates for core group.

residuals the core group members and deviators look quite similar.

Distribution of Heterogeneous Coefficients (Balanced Panels). Table 5 summarizes the posterior median and 90% highest posterior density (HPD) estimates for various elements of the θ vector. Coefficient heterogeneity is controlled by the parameters q^l and v_{δ^l}, $l \in \{\alpha, \rho, \sigma_u, \sigma_\epsilon\}$. The point estimates and HPD intervals for q^l line up with the information in Figures 2 and 3. The posterior medians of q^{σ_u} and q^{σ_ϵ} are close to one and those for q^ρ are close to zero. The posteriors for q^α are more spread out. For the 1988 sample the 90% credible set ranges from 0 to 0.84 and for 1991 from 0.53 to 1.

The size of the discrepancy δ_i^α from the α value of the core group is determined by the matrix v_{δ^α}. We report posterior summary statistics for the diagonal elements. For the 1988 sample, the HPD interval for $v_{\delta^{\alpha_0}}$ ranges from 0.03 to 0.21. The interval for $v_{\delta^{\alpha_1}}$ (divided by 100 to undo the division of experience by 10 in the definition of x_{it} in (26)) extends from .0001 to .0004. To put these variances into perspective, note that the posterior median

Table 5: Posterior Distribution of Elements of θ, $T = 20$

	$\tau = 1988$		$\tau = 1991$	
Para	Median	HPD	Median	HPD
Distribution of Heterogeneous Coefficients				
q^{α}	0.34	[0.00, 0.84]	0.81	[0.53, 1.00]
q^{ρ}	0.04	[0.00, 0.12]	0.05	[0.00, 0.12]
q^{σ_u}	0.97	[0.94, 1.00]	0.98	[0.95, 1.00]
q^{σ_ϵ}	0.97	[0.93, 1.00]	0.97	[0.92, 1.00]
$v_{\delta^{\alpha_0}}$	0.08	[0.03, 0.21]	0.05	[0.02, 0.07]
$v_{\delta^{\alpha_1}}/100$	.0002	[.0001, .0004]	.0001	[.0001, .0002]
$v_{\delta^{\rho}}$	0.21	[0.11, 0.35]	0.20	[0.10, 0.33]
$v_{\delta^{\sigma_u}}$	6.97	[2.76, 12.59]	8.36	[3.40, 16.27]
$v_{\delta^{\sigma_\epsilon}}$	4.58	[1.63, 8.87]	5.95	[2.26, 11.23]
Core Group				
α_0	-0.23	[-0.30, -0.18]	-0.18	[-0.25, -0.11]
$\alpha_1/10$	.018	[.015, .021]	.014	[.011, .017]
ρ	0.97	[0.91, 1.00]	0.90	[0.84, 0.95]

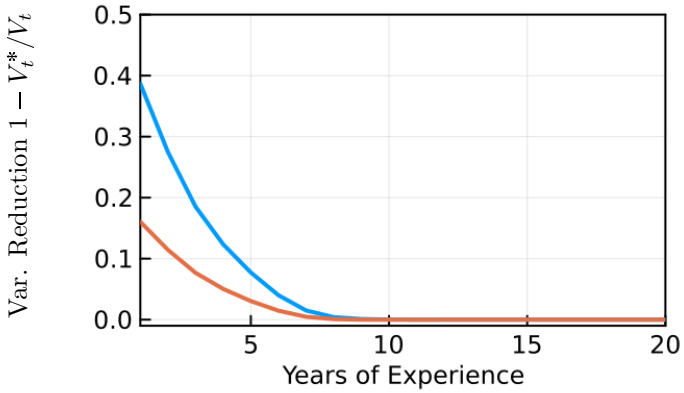
Notes: We report 90% posterior HPD intervals. α_1 and $v_{\delta^{\alpha_1}}$ are rescaled to undo the division of experience by 10 in the definition of x_{it}, see (26).

estimate of the return to an additional year of experience for members of the core group is 1.8% in the 1988 sample and 1.4% in the 1991 sample.[11] Our estimates can be compared to Guvenen (2009), who reports a $v_{\delta^{\alpha_0}}$ estimate of .022 with a standard error of .074, and a $v_{\delta^{\alpha_1}}$ estimate of .00038 with a standard error of .0008. Hryshko (2012) reports $\hat{v}_{\delta^{\alpha_1}} = .0004$. Their estimates are based on a different subset of the PSID data and obtained from a different estimation procedure. Our $v_{\delta^{\alpha_0}}$ estimate is slightly higher, and the $v_{\delta^{\alpha_1}}$ estimate is slightly lower than those obtained in the two earlier papers, but the intervals overlap.

We now turn to the autocorrelations. Because the estimated q^{ρ} is close to zero, the sample contains hardly any information about $v_{\delta^{\rho}}$ and the posterior is close to the prior. The HPD interval for the autocorrelation parameter ρ of the core group ranges from 0.91 to 1.00 for the 1988 sample and from 0.84 to 0.95 for the 1991 sample. Guvenen (2009) obtains $\hat{\rho} = .988$ for the RIP specification with homogeneous coefficients, and 0.82 for the

[11] Both the time-series and the cross-sectional variation contribute to the identification of the return to experience. Examining the cross-sectional distribution of experience at the beginning of the 1988 sample, we find that the 25th percentile is seven years of experience, the median is ten years, and the 75th percentile is thirteen years. For the 1991 sample these numbers are very similar.

Figure 5: Effect of α_i Heterogeneity and Idiosyncratic Shocks on Cohort-Level Income Inequality, $\mathcal{B}_{20,1988}$ Sample



Notes: The counterfactual variance V_t^* is either $V_t(\delta_i^\alpha = 0)$ (no return-to-experience heterogeneity, solid blue line) or $V_t(\sigma_{u,t}^2 = 0)$ (no transitory shocks, solid red line).

HIP model with fully heterogeneous coefficients. Our posterior medians of 0.97 and 0.90 are in between those estimates. The $\tau = 1988$ sample features less heterogeneity in the experience profiles and is therefore closer to Guvenen's RIP model, whereas the 1991 sample exhibits more heterogeneity in the α_is. Thus, we are able to reproduce the finding that homogeneous coefficient specifications are associated with more persistence in the stochastic income component.

Under the prior $\rho \sim \mathcal{N}(0.7, 0.1^2)$ mentioned in Footnote 9, the posterior median of the core-group persistence for the 1991 sample shifts from 0.90 down to 0.88, and the 90% HPD interval shifts from $[0.84, 0.95]$ to $[0.82, 0.94]$. The heterogeneity in persistence is unaffected: the posterior median of q^ρ is 0.06 compared to 0.05, and the slab variance v_{δ^ρ} is essentially unchanged, so the ρ core group remains large and the conclusion that persistence is sparsely heterogeneous carries over. The posterior median of q^α rises from 0.81 to 0.88. This reflects the trade-off between persistence and heterogeneity in experience profiles: when less of the earnings dynamics is attributed to the persistent component, more of it is attributed to heterogeneous income profiles. For all parameters reported in Table 5, the posterior medians under the alternative prior lie inside the 90% HPD intervals obtained for the 1991 sample under the baseline prior.

Parameter Heterogeneity and Income Inequality (Balanced Panels). Using our estimates, we can decompose income inequality within a cohort into the contribution of return-to-experience heterogeneity and heterogeneity in the *ex post* realization of the per-

sistent stochastic process s_{it}. To construct the decomposition, we simulate a cohort of $N = 10,000$ individuals for $T = 20$ periods from the panel state-space model using the posterior mean estimates presented in Figure 2 and Table 5. The cohort starts with an experience of $h_{i1} = 1$. We subsequently conduct two more simulations in which we set either the return-to-experience heterogeneity δ_i^α or the transitory shock variances $\sigma_{u,t}^2$ to zero, but use the same innovations as in the baseline simulation. We measure inequality as the cross-sectional variance of log income. For the baseline simulation, this variance is denoted by V_t, and for the counterfactual simulations they are denoted by $V_t^* = V_t(\delta_i^\alpha = 0)$ and $V_t^* = V_t(\sigma_{u,t}^2 = 0)$, respectively.

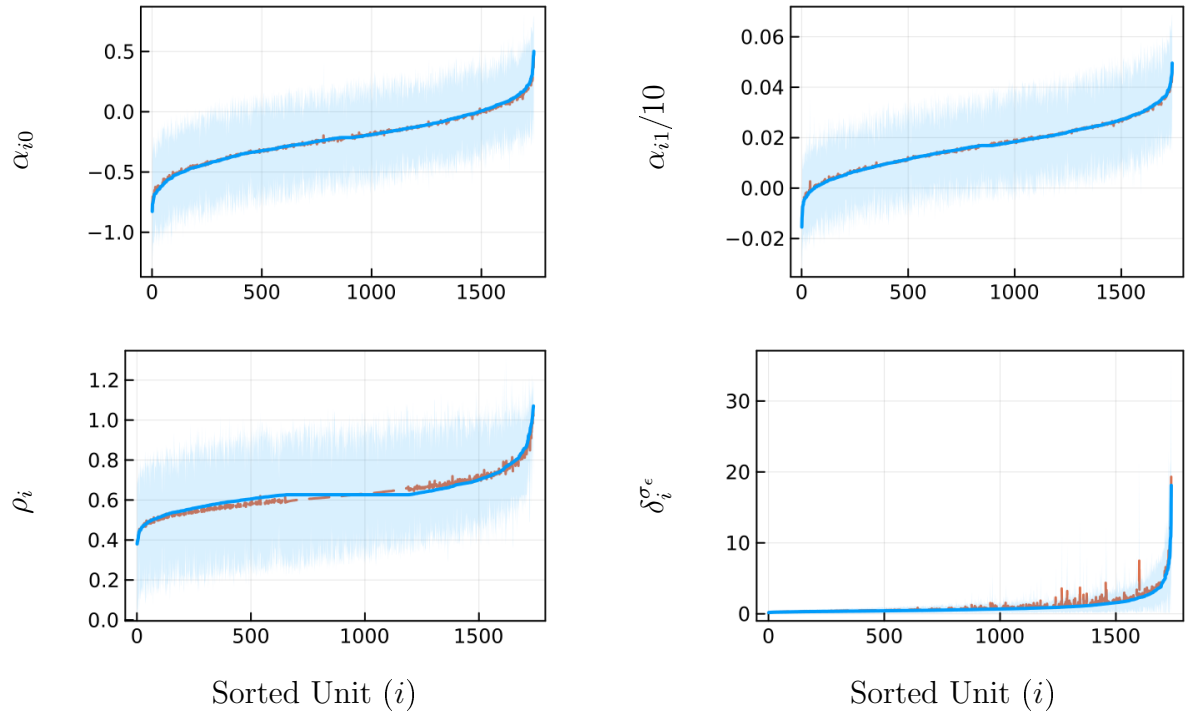
The ratio $[V_t - V_t(\delta_i^\alpha = 0)]/V_t$ for the $\mathcal{B}_{20,1988}$ sample is plotted in Figure 5 as a solid blue line. Recall that $\hat{q}^\alpha = 0.34$. Initially, return-to-experience heterogeneity, due to the individuals deviating from the core group, explains about 40% of income inequality in the cohort. However, the share drops quickly and after seven years α heterogeneity is irrelevant and income inequality is solely due to the stochastic income components. The red line corresponds to the relative effect of the transitory shock on inequality. Initially, it is about 15% but it also quickly converges to zero. Thus, after seven years, inequality is essentially solely due to the persistent idiosyncratic income shocks.

Idiosyncratic Parameters (Unbalanced Panel). We now re-estimate the panel state-space model on the unbalanced panel $\mathcal{U}$ described in Section 5.1. Posterior estimates of the idiosyncratic parameters α_{i0}, $\alpha_{i1}/10$, ρ_i, and $\delta_i^{\sigma_\epsilon}$ are provided in Figure 6. In each panel, we sort the units by the posterior median estimates. As before, heteroskedasticity, i.e., full heterogeneity of $\delta_i^{\sigma_\epsilon}$, is a key feature of the panel. However, unlike in the case of the balanced panels in Figure 2, there is no evidence of sparsity in the δ_i^α estimates in this larger panel data set. The sorted posterior median estimates of α_{i0} and $\alpha_{i1}/10$ do not exhibit any flat segments. The only parameter that does have a flat segment in the posterior median function is ρ_i, but the flat segment is not as pronounced as under the $\mathcal{B}_{20,1988}$ and $\mathcal{B}_{20,1991}$ samples. Thus, as expected, expanding the cross-sectional dimension of the panel by including more individuals pushes the estimates toward full heterogeneity.

5.4 Prediction (Balanced Panels)

We now examine the predictive properties of our estimated models. We consider two types of analyses. In the first part, we use the predictive performance as a way of comparing the benchmark S&S specification to the restricted versions of the panel state-space model

Figure 6: Posterior of Idiosyncratic Parameters, Unbalanced Panel



Notes: Solid blue lines in the α_{i0}, $\alpha_{i1}/10$, and ρ_i panels are $\text{med}(l|Y_{1:N,1:T}) + \text{med}(\delta_i^l|Y_{1:N,1:T})$. In the bottom-right panel they are $\text{med}(\delta_i^{\sigma_\epsilon}|Y_{1:N,1:T})$. The dashed red lines are posterior means and the light blue bands are 90% credible bands. Units are sorted based on posterior median estimates.

listed in Table 3: Homosk, RIP, and HIP. For a given sample $\mathcal{B}_{T,\tau}$, we estimate the model specifications based on data up to period T and then we forecast income in period $T+1$ for the individuals $i = 1, \ldots, N$ and evaluate the accuracy of the predictions, averaging across all individuals. In the second part of this section we examine the prediction problem from the perspective of an individual who would like to forecast their income to make economic decisions, e.g., a consumption/savings decision. The cross-sectional heterogeneity determines the extent to which this individual could benefit from analyzing the income profiles of other individuals.

Assessing Model Specifications Using Predictive Performance. The subsequent pseudo-out-of-sample approach to compare competing model specifications is different from what is typically done in the income-dynamics literature. Because a large number of authors have estimated the income models using minimum-distance techniques, e.g., by minimizing the discrepancy between sample autocovariances $\widehat{\text{Cov}}[y_{it}, y_{i,t-j}]$ and the model-implied population analogues, it is natural to consider tests of the hypothesis $v_{\delta^\alpha} = 0$. In our Bayesian

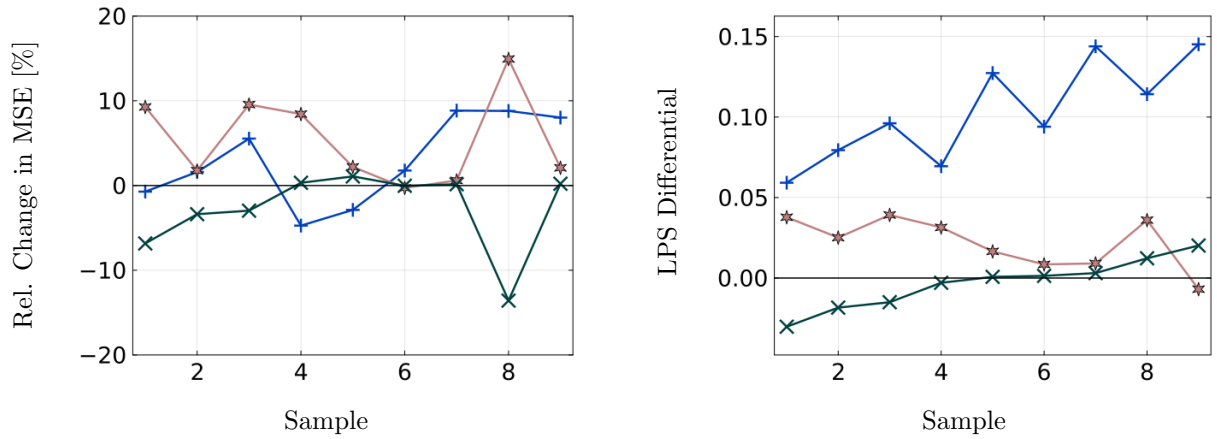
setting we could in principle use posterior odds tests of $q^l = 0$ versus $q^l > 0$, but because the income processes are often embedded in life-cycle models in which agents have to predict their future income to make optimal decisions, we directly focus on the pseudo-out-of-sample performance.

We estimate the model specifications for a sequence of samples $\mathcal{B}_{T,\tau}$, starting with $\tau = 1988$ and ending with $\tau = 1996$. Throughout, we consider $T = 20$. The forecast evaluation approach is essentially the same as in Liu, Moon, and Schorfheide (2020, 2023) and Liu (2023). As evaluation criteria we consider mean-squared-error (MSE) statistics and log predictive scores (LPS). The MSEs are computed as in Step 3 of Algorithm 1 and the LPS are defined as $\frac{1}{N}\sum_{i=1}^{N} \ln \hat{p}(y_{i,T+1}|Y_{1:N,1:T})$, where $\hat{p}(\cdot)$ is the predictive density with a particular model specification. We evaluate $\hat{p}(y_{i,T+1}|Y_{1:N,1:T})$ by averaging, over the posterior draws, the normal density with mean $x'_{i,T+1}(\alpha + \delta_i^{\alpha}) + (\rho + \delta_i^{\rho})s_{iT}$ and variance $\delta_i^{\sigma_u}\sigma^2_{u,T+1} + \delta_i^{\sigma_\epsilon}\sigma^2_{\epsilon,T+1}$. Because our prior treats the time-specific variances as independent across periods, the estimation sample is not informative about $\sigma^2_{u,T+1}$ and $\sigma^2_{\epsilon,T+1}$, and we replace them by $\sigma^2_{u,T}$ and $\sigma^2_{\epsilon,T}$.

The results are summarized in Figure 7. In the left panel of the figure we show relative changes in MSEs [%] defined as $100\cdot\big(\text{MSE(xxx)}-\text{MSE(S\&S)}\big)/\text{MSE(S\&S)}$, where (xxx) is one of the alternative specifications in Table 3. Thus, a negative value indicates that model (xxx) attains a lower MSE than the benchmark model S&S. The LPS differentials are defined such that a negative value is an improvement, meaning that we report $\text{LPS(S\&S)} - \text{LPS(xxx)}$. To convert them into predictive log odds, they would have to be multiplied by the cross-sectional dimension N, which varies across samples but is between 82 and 127.

We draw the following conclusions from Figure 7. First, the fully heterogeneous HIP specification has a larger MSE than the baseline S&S model for eight out of nine samples. Moreover, S&S attains a higher LPS than HIP in eight out of nine samples, with predictive log odds of up to 4 in favor of S&S. Second, there are only three samples for which Homosk attains a lower MSE than S&S. In the other six samples S&S delivers more accurate point forecasts. More importantly, Homosk has the lowest LPS of all specifications in every sample, and the predictive log odds in favor of S&S range from 5 to 18. This confirms that the cross-sectional heteroskedasticity is an important feature of the data. Finally, RIP delivers a similar point forecast performance to S&S in five out of nine samples and attains an MSE reduction between approximately 3% and 13% in the remaining four samples. The LPS differentials between RIP and S&S are small: the predictive log odds are below 3 in absolute value in every sample. RIP attains a higher LPS in three samples and S&S in two, while

Figure 7: One-Step-Ahead Predictive Performance Relative to S&S, $T = 20$, Various Samples



Notes: Samples τ are sorted by the LPS differential for RIP in both panels. Negative values indicate an improvement over the benchmark S&S specification. Homosk, blue plus; RIP, dark green cross; HIP, brown star. The 1988 and 1991 samples are ranked 9th and 1st, respectively.

Table 6: Prior Distribution for Individual-Level Estimation (Based on $Y_{i,1:T}$ Only)

Para	Distr.	a	b
α_i	$\mathcal{N}(a,b)$	$0_{2\times 1}$	$\begin{bmatrix} 0.24 & 0 \\ 0 & .048 \end{bmatrix}$
ρ_i	$\mathcal{N}(a,b)$	0.8	0.25
$(\sigma_i^u)^2$	$IG\big(a/2, b/2\big)$	4.02	.101
$(\sigma_i^\epsilon)^2$	$IG\big(a/2, b/2\big)$	4.02	.101
s_{i0}	$\mathcal{N}(a,b)$	0	0.05

Notes: $\mathcal{N}(a,b)$ is the Normal distribution, $IG(a,b)$ is the Inverse Gamma distribution, $B(a,b)$ is the Beta distribution, and $IW(a,b)$ is the Inverse Wishart distribution. For IG priors: $(a,b) = (4.02, .101)$ implies a mean of 0.05 and a standard deviation of 0.5.

in the remaining four samples the log odds are below 0.5 in absolute value. This is not surprising because, as we have seen in Section 5.3, regression coefficient heterogeneity is generally quite sparse and imposing a restriction that is not quite correct often favorably trades off a slight increase in bias against a variance reduction.

Individual-Level Forecasts Using Three Different Information Sets. We previously generated simultaneous forecasts for all units to compare the predictive ability of competing model specifications. Now we consider the forecasting problem from the perspective of a particular individual i. We consider three different interval forecasts. The baseline forecast for individual i conditions on the information $Y_{1:N,1:T}$ and accounts for parameter uncertainty.

Chamberlain and Hirano (1999) interpret this scenario[12] as "an individual seeking advice from a financial planner. The individual provides data on his earnings history and on various personal characteristics such as age and education. The planner has access to longitudinal data sets that provide data on earnings histories and personal characteristics for samples of individuals." (Abstract, Page 211). The second set of interval forecasts is also based on the $Y_{1:N,1:T}$ information, but we eliminate the parameter uncertainty by fixing (δ, θ) at the posterior mean estimates.

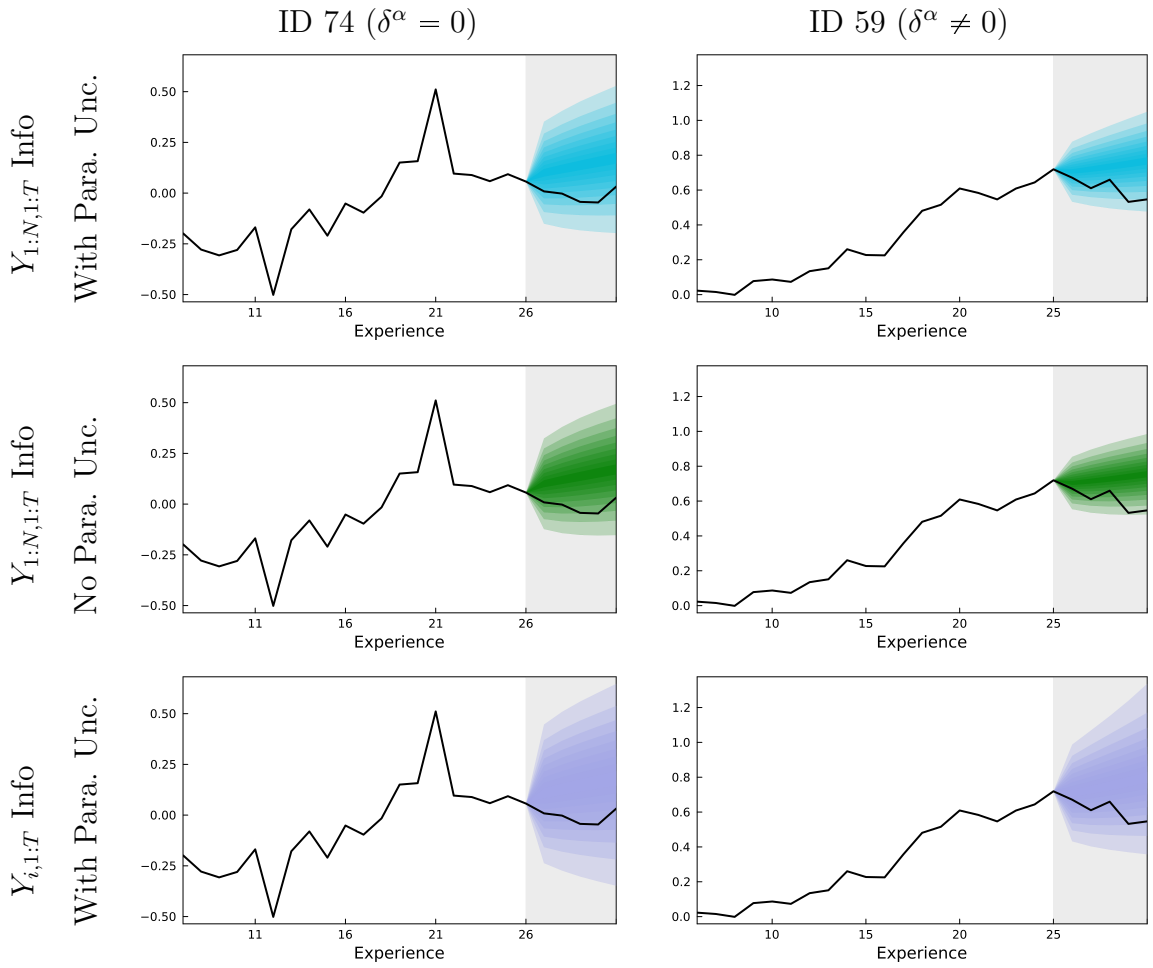
Third, we generate forecasts based on the time series information of individual i: $Y_{i,1:T}$. To obtain forecasts based on $Y_{i,1:T}$, Equations (7) and (8) are re-estimated based on data from individual i only. In the individual-level model the parameters are unit-specific levels, and the shock variances are constant over time. The prior distributions are summarized in Table 6. The hyperparameters for the priors are selected based on the prior specification outlined in Table 4. Specifically, the prior variances of α_i and ρ_i are set to be equal to the prior means of v_{δ^α} and v_{δ^ρ}, respectively. The prior means of $(\sigma_i^u)^2$ and $(\sigma_i^\epsilon)^2$ are aligned with the prior means of $\sigma_{u,t}^2$ and $\sigma_{\epsilon,t}^2$. Both the prior mean and variance of s_{i0} are matched with the prior mean of μ_{s_0} and v_{s_0}.

Figure 8 depicts fan charts to summarize interval forecasts under the three scenarios for a member of the core group, the unit with ID 74, and one of the deviators, the unit with ID 59. We show their actual income for periods $t = 1, \ldots, T = 20$ and then interval forecasts for the next five years. A comparison of the panels in the first and second row of the figure suggests that shutting down parameter uncertainty shortens the predictive intervals somewhat. But, overall, the uncertainty appears to be dominated by future shocks. The forecasts in the third row of the figure are obtained by conditioning on the smaller information set $Y_{i,1:T}$, which widens the intervals relative to the baseline forecasts in the first row. The reason is that the cross-section provides information about the distribution of the idiosyncratic parameters that can be used as a prior to generate sharper time series estimates for unit i.

To conduct a more systematic analysis of how parameter uncertainty and the information set used for generating the forecasts affect interval and density forecasts, we plot 90% interval width ratios for all individuals i in the $\mathcal{B}_{20,1991}$ sample in Figure 9 as a function of years of experience. The denominators of the ratios are obtained from the forecasts based on $Y_{1:N,1:T}$ information with parameter uncertainty (baseline). The numerators for the left panels are the interval widths associated with the no-parameter-uncertainty forecasts, whereas

[12]This setup is also considered by Gu and Koenker (2017b).

Figure 8: Individual Earnings Forecasts, $\mathcal{B}_{20,1991}$ Sample

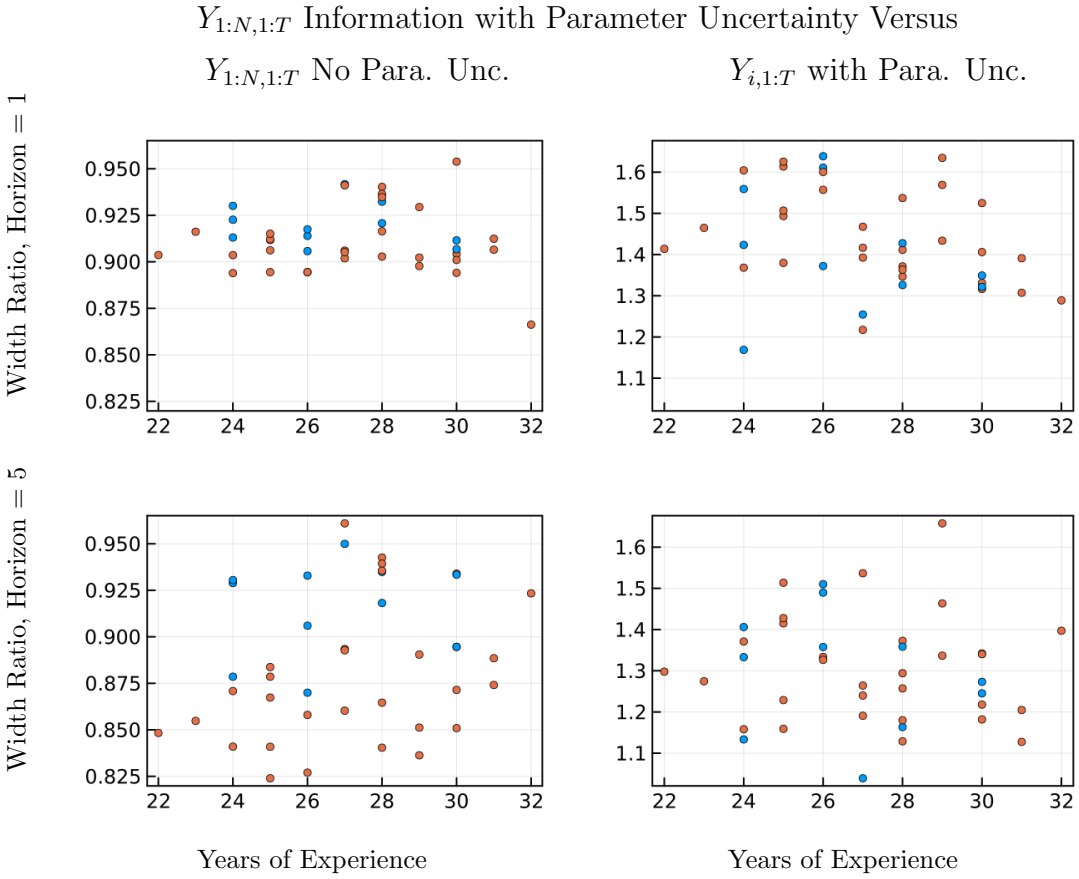


Notes: Forecast origin is period $t = T = 20$. Pointwise prediction bands are shown with gradual shading indicating bands from the 0.05 to 0.95 quantiles.

the numerators for the right panels are constructed from the $Y_{i,1:T}$ forecasts. As suggested by Figure 8, the ratios on the left are below one, whereas the ratios on the right are above one.

Ratios for the core group are indicated through blue dots and values for the deviators by red dots. For the no-parameter-uncertainty-to-baseline ratios in the left panels we expect the average ratio for the deviators to be smaller than for the core group. Because the parameter estimates for the deviators do not benefit from pooling observations across units, parameter uncertainty should play a more important role for the predictive distribution. For the one-step-ahead forecasts the average for the deviators is 0.91 and for the core group is 0.92. For the five-step-ahead forecasts the discrepancy is more pronounced as the average of the

Figure 9: Ratio of Predictive Interval Widths, $\mathcal{B}_{20,1991}$ Sample



Notes: Ratio of width of 90% predictive interval based on (i) $Y_{1:N,1:T}$ information without parameter uncertainty (numerator) and (ii) $Y_{i,1:T}$ information with parameter uncertainty (numerator), respectively, versus $Y_{1:N,1:T}$ information with parameter uncertainty (denominator). x-Axis: years of experience. Core group: blue dots. Deviators: red dots.

deviators drops from 0.91 to 0.88, whereas the average for the core group falls only from 0.92 to 0.91. However, none of these differences are statistically significant.

For the $Y_{i,1:T}$-to-$Y_{1:N,1:T}$ information ratios plotted in the right panel, we expect the core group members to benefit more from including the information from other individuals. How much more depends on the size of the deviations, captured by the v_δ parameters. In principle, this translates into higher ratios for core group members than for deviators. It turns out that this effect is not reflected in the average ratios. For one-step-ahead (five-step-ahead) forecasts the average core group ratio is 1.4 (1.3), whereas the average deviator ratio is 1.45 (1.31). As for the ratios in the left panels, the differences are not statistically significant. One reason for the insignificance could be that the classification of units into core members and deviators is not sharp. In general there is posterior uncertainty about the memberships,

which might explain the insignificance.

Uncertainty about future income plays an important role in life-cycle models because it determines the magnitude of precautionary savings. Many models assume full information rational expectations, which in the context of our analysis means that there is no uncertainty about the parameters of the income process. The only source of uncertainty is the realization of future shocks. Some authors, e.g., Guvenen (2007), explicitly model the Bayesian learning about heterogeneous parameters, and account for parameter uncertainty in addition to shock uncertainty. Our estimates highlight that the degree of uncertainty varies considerably based on the assumptions about how the income forecasts are generated. The no-parameter-uncertainty case could be interpreted as rational expectations. Accounting for parameter uncertainty increases the width of the 90% predictive intervals by approximately 10%. Using a sparser information set that only includes individual i's own income trajectory raises the width of the intervals by an additional 40%.

6 Conclusion

We incorporated a version of a spike-and-slab prior, comprising a point mass at zero ("spike") and a Normal distribution around zero ("slab"), into a dynamic panel data framework to model coefficient heterogeneity. Our framework nests coefficient homogeneity and full heterogeneity as special cases, but can also capture sparse heterogeneity, meaning that there is a core group of units that share the same coefficients. The remaining units deviate from the core group.

Using balanced panels of individuals with a long employment history, we found that heterogeneity with respect to returns to experience and the autocorrelation of the persistent stochastic income component is sparse, with large core groups of individuals sharing identical coefficients and a small number of deviators. For a large unbalanced panel, we found evidence for full coefficient heterogeneity, with the caveat that the specification with homogeneous coefficients delivers slightly more accurate one-step-ahead point forecasts, while its density forecasts are less accurate. We also examined forecasts for specific individuals conditional on information from the full panel $Y_{1:N,1:T}$ accounting for shock and parameter uncertainty, the full panel ignoring parameter uncertainty, and the individual history $Y_{i,1:T}$ accounting for shock and parameter uncertainty. This comparison has an important message for the macroeconomic life-cycle model literature: subtle assumptions about the agents' information

set can have large effects on the income uncertainty that agents face and hence on their precautionary savings motive.

There is a straightforward extension of our framework to correlated random effects and one could also treat the slab part of the δ_i^l distribution nonparametrically as in Liu (2023). However, in many empirical applications simple parametric specifications tend to work well and we hope that the proposed framework will prove to be useful beyond the application presented in this paper.

References

Abowd, J. M., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Change," *Econometrica*, 57(2), 411–445.

Atchadé, Y. F., and J. S. Rosenthal (2005): "On Adaptive Markov Chain Monte Carlo Algorithms," *Bernoulli*, 11(5), 815–828.

Baker, M. (1997): "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics*, 15(2), 338–375.

Bonhomme, S., and E. Manresa (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83(3), 1147–1184.

Browning, M., and M. Ejrnæs (2013): "Heterogeneity in the Dynamics of Labor Earnings," *Annual Review of Economics*, 5, 219–245.

Browning, M., M. Ejrnæs, and J. Alvarez (2010): "Modelling Income Processes with Lots of Heterogeneity," *Review of Economic Studies*, 77(4), 1353–1381.

Castillo, I., and A. van der Vaart (2012): "Needles and Straw in a Haystack: Posterior Concentration for Possibly Sparse Sequences," *Annals of Statistics*, 40(4), 2069–2101.

Chamberlain, G., and K. Hirano (1999): "Predictive Distributions Based on Longitudinal Earnings Data," *Annales d'Économie et de Statistique*, 55/56, 211–242.

Frühwirth-Schnatter, S., and R. Tüchler (2008): "Bayesian Parsimonious Covariance Estimation for Hierarchical Linear Mixed Models," *Statistics and Computing*, 18(1), 1–13.

Frühwirth-Schnatter, S., and H. Wagner (2011): "Bayesian Variable Selection for Random Intercept Modelling of Gaussian and Non-Gaussian Data," in *Bayesian Statistics 9*, ed. by J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, pp. 165–200. Oxford University Press.

Gelman, A., and D. B. Rubin (1992): "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7(4), 457–472.

George, E. I., and R. E. McCulloch (1993): "Variable Selection Via Gibbs Sampling," *Journal of the American Statistical Association*, 88(423), 881–889.

——— (1997): "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7(2), 339–373.

Geweke, J. (1996): "Variable Selection and Model Comparison in Regression," in *Bayesian Statistics*, ed. by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, vol. 5, pp. 169–194. Oxford University Press.

Giannone, D., M. Lenza, and G. Primiceri (2021): "Economic Predictions with Big Data: The Illusion of Sparsity," *Econometrica*, 89(5), 2409–2437.

Gordon, G., J. B. Jones, U. Neelakantan, and K. Athreya (2023): "Incarceration, Employment and Earnings: Dynamics and Differences," *Review of Economic Dynamics*, 51, 677–697.

Griffin, J. E. (2016): "An Adaptive Truncation Method for Inference in Bayesian Nonparametric Models," *Statistics and Computing*, 26(1), 423–441.

Gu, J., and R. Koenker (2017a): "Empirical Bayesball Remixed: Empirical Bayes Methods for Longitudinal Data," *Journal of Applied Econometrics*, 32(3), 575–599.

——— (2017b): "Unobserved Heterogeneity in Income Dynamics: An Empirical Bayes Perspective," *Journal of Business & Economic Statistics*, 35(1), 1–16.

Guvenen, F. (2007): "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?," *American Economic Review*, 97(3), 687–712.

——— (2009): "An Empirical Investigation of Labor Income Processes," *Review of Economic Dynamics*, 12(1), 58–79.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021): "What Do Data on Millions of U.S. Workers Reveal About Life-Cycle Earnings Dynamics?," *Econometrica*, 89(5), 2303–2339.

Haider, S. J. (2001): "Earnings Instability and Earnings Inequality of Males in the United States: 1967–1991," *Journal of Labor Economics*, 19(4), 799–836.

Hause, J. C. (1980): "The Fine Structure of Earnings and the On-the-Job Training Hypothesis," *Econometrica*, 48(4), 1013–1029.

Hoffmann, F. (2019): "HIP, RIP, and the Robustness of Empirical Earnings Processes," *Quantitative Economics*, 10, 1279–1315.

Hospido, L. (2012): "Modelling Heterogeneity and Dynamics in the Volatility of Individual Wages," *Journal of Applied Econometrics*, 27, 386–414.

Hryshko, D. (2012): "Labor Income Profiles are Not Heterogeneous: Evidence from Income Growth Rates," *Quantitative Economics*, 3, 177–209.

Johnstone, I. M., and B. W. Silverman (2004): "Needles and Straw in Haystacks: Empirical Bayes Estimates of Possibly Sparse Sequences," *Annals of Statistics*, 32(4), 1594–1649.

Lillard, L. A., and Y. Weiss (1979): "Components of Variation in Panel Earnings Data: American Scientists 1960–70," *Econometrica*, 47(2), 437–454.

Liu, L. (2023): "Density Forecasts in Panel Data Models: A Semiparametric Bayesian Perspective," *Journal of Business & Economic Statistics*, 41(2), 349–363.

Liu, L., H. R. Moon, and F. Schorfheide (2020): "Forecasting With Dynamic Panel Data Models," *Econometrica*, 88(1), 171–201.

——— (2023): "Forecasting With a Panel Tobit Model," *Quantitative Economics*, 14(1), 117–159.

MaCurdy, T. E. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 18, 83–114.

Meghir, C., and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

Mitchell, T. J., and J. J. Beauchamp (1988): "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association*, 83(404), 1023–1032.

Nakata, T., and C. Tonetti (2015): "Small Sample Properties of Bayesian Estimators of Labor Income Processes," *Journal of Applied Economics*, 18(1), 121–148.

Rodriguez, J. E., D. R. Williams, and P. Rast (2024): "Who Is and Is Not "Average"? Random Effects Selection with Spike-and-Slab Priors," *Psychological Methods*, 29(1), 117–136.

Rosenthal, J. S. (2011): "Optimal Proposal Distributions and Adaptive MCMC," in *Handbook of Markov Chain Monte Carlo*, ed. by S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, pp. 93–112. Chapman & Hall/CRC, Boca Raton, FL.

Zhang, B. (2023): "Incorporating Prior Knowledge of Latent Group Structure in Panel Data," *arXiv Working Paper*, 2211.16714.

Online Appendix: Bayesian Estimation of Panel Models under Potentially Sparse Heterogeneity

Hyungsik Roger Moon, Frank Schorfheide, and Boyuan Zhang

The Online Appendix consists of the following parts:

A. Posterior Sampling for the Panel State-Space Model

B. Convergence Diagnostics

C. Additional Empirical Results

A Posterior Sampling for the Panel State-Space Model

The income process in the panel state-space model is

$$\begin{aligned} y_{it} &= x_{it}'(\alpha + \delta_i^{\alpha}) + s_{it} + \sigma_{u,t}\sqrt{\delta_i^{\sigma_u}}u_{it} \\ s_{it} &= (\rho + \delta_i^{\rho})s_{it-1} + \sigma_{\epsilon,t}\sqrt{\delta_i^{\sigma_\epsilon}}\epsilon_{it}, \end{aligned} \tag{A.1}$$

where $\alpha = [\alpha_0 \ \alpha_1]'$, $\delta_i^{\alpha} = [\delta_i^{\alpha_0} \ \delta_i^{\alpha_1}]'$, and $x_{it} = [1 \ h_{it}/10]'$. All posterior distributions in this appendix are understood to be conditional on the observables $Y_{1:N,1:T}$ and $X_{1:N,1:T}$, and we suppress this conditioning set to simplify notation. Throughout this section, proportionality statements denote kernels of conditional posterior densities, and terms that do not depend on the parameter being updated are omitted.

Priors for Heterogeneous Parameters. For $(\delta^{\alpha}, \delta^{\rho})$ we use a (multivariate) Normal distribution for the continuous part and let

$$\delta_i^l | q^l, v_{\delta^l} \sim \begin{cases} \mathcal{N}(0, v_{\delta^l}) & \text{with prob. } q^l \\ 0 & \text{with prob. } 1-q^l \end{cases} \qquad l \in \{\alpha, \rho\}. \tag{A.2}$$

Note that v_{δ^α} is a 2×2 matrix. For $(\delta_i^{\sigma_u}, \delta_i^{\sigma_\epsilon})$ we use an Inverse Gamma (IG) distribution instead of a Normal distribution to capture the discrepancies:

$$\delta_i^{\sigma_m} | q^{\sigma_m}, \nu_{\delta^{\sigma_m}}, \tau_{\delta^{\sigma_m}} \sim \begin{cases} IG\left(\frac{\nu_{\delta^{\sigma_m}}}{2}, \frac{\tau_{\delta^{\sigma_m}}}{2}\right) & \text{with prob. } q^{\sigma_m} \\ 1 & \text{with prob. } 1-q^{\sigma_m} \end{cases} \qquad m \in \{u, \epsilon\}. \tag{A.3}$$

We impose that $\mathbb{E}[\delta_i^{\sigma_m}] = 1$ and reparameterize the prior distribution in terms of its variance $v_{\delta^{\sigma_m}}$, which leads to

$$\nu_{\delta^{\sigma_m}} = 2v_{\delta^{\sigma_m}}^{-1} + 4, \quad \tau_{\delta^{\sigma_m}} = 2v_{\delta^{\sigma_m}}^{-1} + 2. \tag{A.4}$$

Priors for Homogeneous Parameters. We consider the following prior distributions for the homogeneous parameters

$$\begin{aligned}
\alpha &\sim \mathcal{N}(\underline{\mu}_\alpha, \underline{v}_\alpha), \quad \rho \sim \mathcal{N}(\underline{\mu}_\rho, \underline{v}_\rho), \quad \sigma_{u,t}^2 \sim IG\left(\frac{\underline{\nu}_{\sigma_u}}{2}, \frac{\underline{\tau}_{\sigma_u}}{2}\right), \quad \sigma_{\epsilon,t}^2 \sim IG\left(\frac{\underline{\nu}_{\sigma_\epsilon}}{2}, \frac{\underline{\tau}_{\sigma_\epsilon}}{2}\right), \\
q^l &\sim B(a,b),\ l \in \{\alpha, \rho, \sigma_u, \sigma_\epsilon\}, \\
v_{\delta^\alpha} &\sim IW\left(\underline{\nu}_{\delta^\alpha}, \underline{\Psi}_{\delta^\alpha}\right), \quad v_{\delta^\rho} \sim IG\left(\frac{\underline{\nu}_{\delta^\rho}}{2}, \frac{\underline{\tau}_{\delta^\rho}}{2}\right), \\
v_{\delta^{\sigma_u}} &\sim IG\left(\frac{\underline{\nu}_{\delta^{\sigma_u}}}{2}, \frac{\underline{\tau}_{\delta^{\sigma_u}}}{2}\right), \quad v_{\delta^{\sigma_\epsilon}} \sim IG\left(\frac{\underline{\nu}_{\delta^{\sigma_\epsilon}}}{2}, \frac{\underline{\tau}_{\delta^{\sigma_\epsilon}}}{2}\right) \\
\mu_{s_0} &\sim \mathcal{N}(\underline{\mu}_{s_0}, \underline{v}_{s_0}), \quad v_{s_0} \sim IG\left(\frac{\underline{\nu}_{s_0}}{2}, \frac{\underline{\tau}_{s_0}}{2}\right),
\end{aligned}$$

where $B(a,b)$ is the Beta distribution and $IW\left(\underline{\nu}_{\delta^\alpha}, \underline{\Psi}_{\delta^\alpha}\right)$ is the Inverse Wishart distribution. We collect the homogeneous parameters into the vector

$$\theta = [\alpha', \rho, \sigma_{u,1:T}^2, \sigma_{\epsilon,1:T}^2, q^\alpha, q^\rho, q^{\sigma_u}, q^{\sigma_\epsilon}, v_{\delta^\alpha}, v_{\delta^\rho}, v_{\delta^{\sigma_u}}, v_{\delta^{\sigma_\epsilon}}, \mu_{s_0}, v_{s_0}]'.$$

The hyperparameters that index the various prior distributions are collected in the vector

$$\begin{aligned}
\lambda = [&\underline{\mu}_\alpha, \underline{v}_\alpha, \underline{\mu}_\rho, \underline{v}_\rho, \underline{\nu}_{\sigma_u}, \underline{\tau}_{\sigma_u}, \underline{\nu}_{\sigma_\epsilon}, \underline{\tau}_{\sigma_\epsilon}, a, b, \underline{\nu}_{\delta^\alpha}, \underline{\Psi}_{\delta^\alpha}, \underline{\nu}_{\delta^\rho}, \underline{\tau}_{\delta^\rho}, \underline{\nu}_{\delta^{\sigma_u}}, \underline{\tau}_{\delta^{\sigma_u}}, \underline{\nu}_{\delta^{\sigma_\epsilon}}, \underline{\tau}_{\delta^{\sigma_\epsilon}}, \\
&\underline{\mu}_{s_0}, \underline{v}_{s_0}, \underline{\nu}_{s_0}, \underline{\tau}_{s_0}]'.
\end{aligned} \tag{A.5}$$

We can sample from the joint posterior of

$$\left(\alpha, \rho, q^\alpha, q^\rho, q^{\sigma_u}, q^{\sigma_\epsilon}, v_{\delta^\alpha}, v_{\delta^\rho}, v_{\delta^{\sigma_u}}, v_{\delta^{\sigma_\epsilon}}, \delta, z, \sigma_{u,1:T}^2, \sigma_{\epsilon,1:T}^2, s\right)$$

using a Gibbs sampling algorithm with eleven blocks. The block structure below follows a standard decomposition: common coefficients, sparsity probabilities, hyperparameters of heterogeneous coefficients, unit-level indicators and discrepancies, latent states, and shock variances.

Conditional Posterior of α. Let $\breve{y}_{it}^{\alpha} = y_{it} - x_{it}'\delta_i^{\alpha} - s_{it}$ and $\sigma_{it}^{u} = \sigma_{u,t}\sqrt{\delta_i^{\sigma_u}}$. Then

$$\begin{aligned} p\left(\alpha|s,\delta^{\alpha},\sigma_{u,1:T}^{2},\delta^{\sigma_u}\right) &\propto \exp\left[-\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\left(\breve{y}_{it}^{\alpha}-x_{it}'\alpha\right)^2}{2(\sigma_{it}^{u})^2}\right]\exp\left[-\frac{1}{2}\left(\alpha-\underline{\mu}_{\alpha}\right)'\underline{v}_{\alpha}^{-1}\left(\alpha-\underline{\mu}_{\alpha}\right)\right] \\ &\propto \exp\left[-\frac{1}{2}(\alpha-\bar{\alpha})'\bar{v}_{\alpha}^{-1}(\alpha-\bar{\alpha})\right], \end{aligned}$$

where

$$\bar{v}_{\alpha} = \left(\underline{v}_{\alpha}^{-1} + \sum_{i=1}^{N}\sum_{t=1}^{T}(\sigma_{it}^{u})^{-2}x_{it}x_{it}'\right)^{-1}, \quad \bar{\alpha} = \bar{v}_{\alpha}\left(\underline{v}_{\alpha}^{-1}\underline{\mu}_{\alpha} + \sum_{i=1}^{N}\sum_{t=1}^{T}(\sigma_{it}^{u})^{-2}x_{it}\breve{y}_{it}^{\alpha}\right).$$

This implies

$$\alpha|s,\delta^{\alpha},\sigma_{u,1:T}^{2},\delta^{\sigma_u} \sim \mathcal{N}\left(\bar{\alpha},\bar{v}_{\alpha}\right). \tag{A.6}$$

Conditional Posterior of ρ. Let $\sigma_{it}^{\epsilon} = \sigma_{\epsilon,t}\sqrt{\delta_i^{\sigma_\epsilon}}$ and $\tilde{s}_{it} = s_{it} - \delta_i^{\rho}s_{it-1}$. Then

$$\begin{aligned} p\left(\rho|s,\delta^{\rho},\sigma_{\epsilon,1:T}^{2},\delta^{\sigma_\epsilon}\right) &\propto \exp\left[-\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\left(\tilde{s}_{it}-\rho s_{it-1}\right)^2}{2(\sigma_{it}^{\epsilon})^2}\right]\exp\left[-\frac{\left(\rho-\underline{\mu}_{\rho}\right)^2}{2\underline{v}_{\rho}}\right] \\ &\propto \exp\left[-\frac{(\rho-\bar{\rho})^2}{2\bar{v}_{\rho}}\right], \end{aligned}$$

where

$$\bar{v}_{\rho} = \left(\underline{v}_{\rho}^{-1} + \sum_{i=1}^{N}\sum_{t=1}^{T}(\sigma_{it}^{\epsilon})^{-2}s_{it-1}^{2}\right)^{-1}, \quad \bar{\rho} = \bar{v}_{\rho}\left(\underline{v}_{\rho}^{-1}\underline{\mu}_{\rho} + \sum_{i=1}^{N}\sum_{t=1}^{T}(\sigma_{it}^{\epsilon})^{-2}s_{it-1}\tilde{s}_{it}\right).$$

This implies

$$\rho|\left(s,\delta^{\rho},\sigma_{\epsilon,1:T}^{2},\delta^{\sigma_\epsilon}\right) \sim \mathcal{N}\left(\bar{\rho},\bar{v}_{\rho}\right). \tag{A.7}$$

Conditional Posterior of $(q^{\alpha}, q^{\rho}, q^{\sigma_u}, q^{\sigma_\epsilon})$. For $l \in \{\alpha, \rho, \sigma_u, \sigma_\epsilon\}$,

$$p(q^{l}|z_{1:N}^{l}) \propto (q^{l})^{a+\psi(z^{l})-1}(1-q^{l})^{b+N-\psi(z^{l})-1},$$

where $\psi(z^{l}) = \sum_{i=1}^{N} z_i^{l}$. For $l \in \{\alpha, \rho\}$ this is the total number of non-zero elements in δ^{l}, and for $l \in \{\sigma_u, \sigma_\epsilon\}$ this is the total number of non-one elements in δ^{l}. This implies

$$q^{l}|z_{1:N}^{l} \sim B(a+\psi(z^{l}), b+N-\psi(z^{l})). \tag{A.8}$$

Conditional Posterior of $(v_{\delta^\alpha}, v_{\delta^\rho})$**.** For v_{δ^α},

$$\begin{aligned} p(v_{\delta^\alpha}|z^\alpha, \delta^\alpha) &\propto \prod_{i=1}^{N}\left[|v_{\delta^\alpha}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\delta_i^{\alpha\prime}v_{\delta^\alpha}^{-1}\delta_i^\alpha\right)\right]^{z_i^\alpha}|v_{\delta^\alpha}|^{-\frac{\underline{\nu}_{\delta^\alpha}+3}{2}}\exp\left[-\frac{1}{2}\text{tr}\left(\underline{\Psi}_{\delta^\alpha}v_{\delta^\alpha}^{-1}\right)\right] \\ &\propto |v_{\delta^\alpha}|^{-\frac{\underline{\nu}_{\delta^\alpha}+\psi(z^\alpha)+3}{2}}\exp\left\{-\frac{1}{2}\text{tr}\left[\left(\underline{\Psi}_{\delta^\alpha}+\sum_{i|z_i^\alpha=1}\delta_i^\alpha\delta_i^{\alpha\prime}\right)v_{\delta^\alpha}^{-1}\right]\right\}. \end{aligned}$$

This implies

$$v_{\delta^\alpha}|(z^\alpha, \delta^\alpha) \sim IW\left(\bar{\nu}_{\delta^\alpha}, \bar{\Psi}_{\delta^\alpha}\right), \tag{A.9}$$

where

$$\begin{aligned} \bar{\nu}_{\delta^\alpha} &= \underline{\nu}_{\delta^\alpha}+\psi(z^\alpha) \\ \bar{\Psi}_{\delta^\alpha} &= \underline{\Psi}_{\delta^\alpha}+\sum_{i|z_i^\alpha=1}\delta_i^\alpha\delta_i^{\alpha\prime}. \end{aligned}$$

Regarding the posterior distribution of v_{δ^ρ},

$$\begin{aligned} p(v_{\delta^\rho}|z^\rho, \delta^\rho) &\propto \prod_{i=1}^{N}\left[\left(\frac{1}{v_{\delta^\rho}}\right)^{\frac{1}{2}}\exp\left(-\frac{(\delta_i^\rho)^2}{2v_{\delta^\rho}}\right)\right]^{z_i^\rho}\left(\frac{1}{v_{\delta^\rho}}\right)^{\frac{\underline{\nu}_{\delta^\rho}}{2}+1}\exp\left(-\frac{\underline{\tau}_{\delta^\rho}}{2v_{\delta^\rho}}\right) \\ &\propto \left(\frac{1}{v_{\delta^\rho}}\right)^{\frac{\underline{\nu}_{\delta^\rho}+\psi(z^\rho)}{2}+1}\exp\left(-\frac{\underline{\tau}_{\delta^\rho}+\sum_{i|z_i^\rho=1}(\delta_i^\rho)^2}{2v_{\delta^\rho}}\right). \end{aligned}$$

This implies

$$v_{\delta^\rho}|(z^\rho, \delta^\rho) \sim IG\left(\frac{\bar{\nu}_{\delta^\rho}}{2}, \frac{\bar{\tau}_{\delta^\rho}}{2}\right), \tag{A.10}$$

where

$$\begin{aligned} \bar{\nu}_{\delta^\rho} &= \underline{\nu}_{\delta^\rho}+\psi(z^\rho) \\ \bar{\tau}_{\delta^\rho} &= \underline{\tau}_{\delta^\rho}+\sum_{i|z_i^\rho=1}(\delta_i^\rho)^2. \end{aligned}$$

Conditional Posterior of $(v_{\delta^{\sigma_u}}, v_{\delta^{\sigma_\epsilon}})$**.** To economize on notation, we represent $v_{\delta^{\sigma_m}}$ by

ω_m, where $m \in \{u, \epsilon\}$. The conditional posterior of ω_m is given by

$$\begin{aligned} & p(\omega_m | z^{\sigma_m}, \delta^{\sigma_m}) \\ \propto & \prod_{i|z_i^{\sigma_m}=1} \left[\frac{(\omega_m^{-1}+1)^{\left(\omega_m^{-1}+2\right)}}{\Gamma(\omega_m^{-1}+2)} \left(\frac{1}{\delta_i^{\sigma_m}}\right)^{\omega_m^{-1}+3} \exp\left(-\frac{\omega_m^{-1}+1}{\delta_i^{\sigma_m}}\right) \right] \left(\frac{1}{\omega_m}\right)^{\frac{\nu_{\delta^{\sigma_m}}}{2}+1} \exp\left(-\frac{\tau_{\delta^{\sigma_m}}}{2\omega_m}\right) \\ \propto & \frac{(\omega_m^{-1}+1)^{\psi(z^{\sigma_m})\left(\omega_m^{-1}+2\right)}}{\Gamma(\omega_m^{-1}+2)^{\psi(z^{\sigma_m})}} \left(\prod_{i|z_i^{\sigma_m}=1} \delta_i^{\sigma_m}\right)^{-\omega_m^{-1}} \exp\left(-\omega_m^{-1} \sum_{i|z_i^{\sigma_m}=1} \frac{1}{\delta_i^{\sigma_m}} - \frac{\tau_{\delta^{\sigma_m}}}{2\omega_m}\right) \omega_m^{-\left(\frac{\nu_{\delta^{\sigma_m}}}{2}+1\right)}, \end{aligned}$$

where $\psi(z^{\sigma_m}) = \sum_{i=1}^N z_i^{\sigma_m}$ is the total number of non-one elements in δ^{σ_m}. The log posterior of ω_m is given by

$$\begin{aligned} \ln p(\omega_m | z^{\sigma_m}, \delta^{\sigma_m}) \propto & \ \psi(z^{\sigma_m})\left(\omega_m^{-1}+2\right)\ln\left(\omega_m^{-1}+1\right) - \psi(z^{\sigma_m})\ln\Gamma(\omega_m^{-1}+2) \\ & - \left(\frac{\nu_{\delta^{\sigma_m}}}{2}+1\right)\ln\omega_m - \omega_m^{-1}\left[\sum_{i|z_i^{\sigma_m}=1}\left(\ln\delta_i^{\sigma_m} + \frac{1}{\delta_i^{\sigma_m}}\right) + \frac{\tau_{\delta^{\sigma_m}}}{2}\right]. \end{aligned}$$

We sample from this nonstandard posterior via a random walk Metropolis-Hastings (RWMH) algorithm. Notice that the posterior is well defined only for $\omega_m > 0$. As a result, we use a truncated Normal distribution for the proposal distribution. Let us denote by $\mathcal{N}_+(\mu, \mu^-, \sigma^2)$ the truncated Normal distribution with left truncation point μ^-. Then the proposal distribution $q(\vartheta|\omega_m^{j-1})$ follows $\mathcal{N}_+(\omega_m^{j-1}, 0, c^2)$, whose density is given by

$$f\left(\vartheta \mid \omega_m^{j-1}, 0, c^2\right) = \frac{\phi\left(\frac{\vartheta-\omega_m^{j-1}}{c}\right)}{\Phi\left(\frac{\omega_m^{j-1}}{c}\right)} \mathbf{1}_{\vartheta>0},$$

where ϕ and Φ are the PDF and CDF of the standard Normal distribution. This proposal distribution ensures the posterior draw of ω_m is strictly bounded below by 0.

The following adaptive RWMH algorithm is based on Atchadé and Rosenthal (2005) and Griffin (2016), who adaptively adjust the random walk step-size to keep the acceptance rate around a desired level.

Algorithm A.1 (Adaptive RWMH Algorithm for ω_m)
Set ω_m^0 to its prior mean 1 and initial random walk step-size $c^0 = 1$. For $j = 1, \ldots, N_{\text{MCMC}}$, conditional on the posterior draws of other parameters,

1. *Draw ϑ from the proposal distribution $\mathcal{N}_+(\omega_m^{j-1}, 0, (c^j)^2)$.*

2. *Set* $\omega_m^j = \vartheta$ *with probability*

$$\begin{aligned}\text{acc}\left(\vartheta|\omega_m^{j-1}\right) = \min\Big\{1, \exp\Big[&\left(\ln p(\vartheta|z^{\sigma_m}, \delta^{\sigma_m}) - \ln p(\omega_m^{j-1}|z^{\sigma_m}, \delta^{\sigma_m})\right) \\ &- \left(\ln \Phi\left(\vartheta/c^j\right) - \ln \Phi\left(\omega_m^{j-1}/c^j\right)\right)\Big]\Big\}\end{aligned}$$

and $\omega_m^j = \omega_m^{j-1}$ *otherwise.*

3. *Update the random walk step-size for the next iteration,*

$$\log c^{j+1} = g\left\{\log c^j + j^{-p}\left[\text{acc}\left(\vartheta|\omega_m^{j-1}\right) - \text{acc}^*\right]\right\}$$

where $0.5 < p \leqslant 1$, acc^* *is the target acceptance rate, and*

$$g(x) = \min(|x|, \bar{x}) \cdot \text{sgn}(x)$$

where $\bar{x} = 10$ *is a user-specified large number. Following Griffin (2016), we set* $p = 0.55$.

Because this posterior is nonstandard, it is difficult to derive a theoretically optimal acceptance rate for the RWMH algorithm. Instead, we set acc^* to 0.44, which is the optimal acceptance rate for univariate sampling suggested by Rosenthal (2011).

With a posterior draw of ω_m, we can find the values for the hyperparameters of $\delta_i^{\sigma_m}$ by (11), which will be used in the next block.

Conditional Posterior of s_{it} and $(z_i^\alpha, \delta_i^\alpha)$. We now turn to the unit-level block. To improve mixing, we jointly update the latent states and the heterogeneity in α by sampling s_{it} together with $(z_i^\alpha, \delta_i^\alpha)$. We choose to draw s_{it} and δ_i^α jointly with the following representation:

$$\breve{y}_{it}^\beta = y_{it} - x_{it}'\alpha = W_{it}'\beta_i + \sigma_{it}^u u_{it},$$

and its matrix form

$$\breve{Y}_i^\beta = Y_i - X_i\alpha = W_i\beta_i + D_{i,u}U_i,$$

where the design matrix $W_i = [X_i \;\; I]$ contains x_{it}' and dummies for s_{it}:

$$W_i = \begin{bmatrix} 1 & h_{i1}/10 & 1 & 0 & \cdots & 0 \\ 1 & h_{i2}/10 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & h_{iT}/10 & 0 & 0 & \cdots & 1 \end{bmatrix} \tag{A.11}$$

$\beta_i = [\delta_i^{\alpha_0},\ \delta_i^{\alpha_1},\ s_{i1}, \ldots, s_{iT}]'$ and $\Sigma_{i,u} = D_{i,u} D_{i,u}' = \delta_i^{\sigma_u} \operatorname{diag}\left(\sigma_{u,1}^2, \sigma_{u,2}^2, \ldots, \sigma_{u,T}^2\right)$, with $D_{i,u}$ as defined in Section 3.

We construct the prior for β_i to account for the sparsity assumption in δ^α, which is given by

$$\begin{aligned} \beta_i | (\rho, \delta_i^\rho, z_i^\alpha, v_{\delta^\alpha}, \sigma_{\epsilon,1:T}^2, \delta_i^{\sigma_\epsilon}, v_{s_0}) &\sim \mathcal{N}\left(\begin{bmatrix} 0_{2\times 1} \\ 0_{T\times 1} \end{bmatrix}, \begin{bmatrix} z_i^\alpha v_{\delta^\alpha} & 0_{2\times T} \\ 0_{T\times 2} & \underline{V}_{s_i} \end{bmatrix} \right) \\ z_i^\alpha | q^\alpha &= \begin{cases} 0 & \text{with prob. } 1-q^\alpha \\ 1 & \text{with prob. } q^\alpha \end{cases}, \end{aligned} \tag{A.12}$$

The (t,τ) elements of the $T \times T$ prior covariance matrix $\underline{V}_{s_i}$ can be calculated as follows:

$$\begin{aligned} \underline{V}_{s_i}(t,t) &= (\rho + \delta_i^\rho)^2 \underline{V}_{s_i}(t-1, t-1) + \sigma_{\epsilon,t}^2 \delta_i^{\sigma_\epsilon}, \quad t = 1, \ldots, T \\ \underline{V}_{s_i}(t,\tau) &= (\rho + \delta_i^\rho)^{|t-\tau|} \underline{V}_{s_i}\big(\min(t,\tau), \min(t,\tau)\big), \quad t = 1, \ldots, T \text{ and } \tau \neq t, \end{aligned}$$

with the understanding that $\underline{V}_{s_i}(0,0) = v_{s_0}$.

The posterior odds of $z_i^\alpha = 1$ versus $z_i^\alpha = 0$ are given by

$$\begin{aligned} K_i^\alpha &= \frac{\mathbb{P}(z_i^\alpha = 1 | Y_i, \alpha, \rho, \delta^\rho, q^\alpha, v_{\delta^\alpha}, \sigma_{u,1:T}^2, \sigma_{\epsilon,1:T}^2, \delta_i^{\sigma_u}, \delta_i^{\sigma_\epsilon}, v_{s_0})}{\mathbb{P}(z_i^\alpha = 0 | Y_i, \alpha, \rho, \delta^\rho, q^\alpha, v_{\delta^\alpha}, \sigma_{u,1:T}^2, \sigma_{\epsilon,1:T}^2, \delta_i^{\sigma_u}, \delta_i^{\sigma_\epsilon}, v_{s_0})} \\ &= \frac{q^\alpha}{1-q^\alpha} \frac{p(Y_i | \alpha, \rho, \delta^\rho, v_{\delta^\alpha}, \sigma_{u,1:T}^2, \sigma_{\epsilon,1:T}^2, \delta_i^{\sigma_u}, \delta_i^{\sigma_\epsilon}, v_{s_0}, z_i^\alpha = 1)}{p(Y_i | \alpha, \rho, \delta^\rho, v_{\delta^\alpha}, \sigma_{u,1:T}^2, \sigma_{\epsilon,1:T}^2, \delta_i^{\sigma_u}, \delta_i^{\sigma_\epsilon}, v_{s_0}, z_i^\alpha = 0)}. \end{aligned}$$

Under $z_i^\alpha = 1$, β_i is set to jointly follow a Normal distribution. The marginalized likelihood of $Y_i | (\alpha, \rho, \delta^\rho, v_{\delta^\alpha}, \sigma_{u,1:T}^2, \sigma_{\epsilon,1:T}^2, \delta_i^{\sigma_u}, \delta_i^{\sigma_\epsilon}, v_{s_0}, z_i^\alpha = 1)$ can be obtained by rearranging

Bayes Theorem:

$$\begin{aligned} & p(Y_i|\alpha,\rho,\delta^\rho,v_{\delta^\alpha},\sigma^2_{u,1:T},\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_u},\delta_i^{\sigma_\epsilon},v_{s_0},z_i^\alpha=1) \\ = \; & p(Y_i|\beta_i,\alpha,\rho,\delta^\rho,v_{\delta^\alpha},\sigma^2_{u,1:T},\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_u},\delta_i^{\sigma_\epsilon},v_{s_0},z_i^\alpha=1) \\ & \times\frac{p(\beta_i|\alpha,\rho,\delta^\rho,v_{\delta^\alpha},\sigma^2_{u,1:T},\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_u},\delta_i^{\sigma_\epsilon},v_{s_0},z_i^\alpha=1)}{p(\beta_i|Y_i,\alpha,\rho,\delta^\rho,v_{\delta^\alpha},\sigma^2_{u,1:T},\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_u},\delta_i^{\sigma_\epsilon},v_{s_0},z_i^\alpha=1)}. \end{aligned}$$

It is given by

$$\begin{aligned} & p(Y_i|\alpha,\rho,\delta^\rho,v_{\delta^\alpha},\sigma^2_{u,1:T},\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_u},\delta_i^{\sigma_\epsilon},v_{s_0},z_i^\alpha=1) \\ = \; & \frac{(2\pi)^{-\frac{T}{2}}|\Sigma_{i,u}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\check{Y}_i^{\beta\prime}\Sigma_{i,u}^{-1}\check{Y}_i^{\beta}\right)(2\pi)^{-\frac{T+2}{2}}\,|\underline{v}_{\beta_i}|^{-\frac{1}{2}}}{(2\pi)^{-\frac{T+2}{2}}\,|\bar{v}_{\beta_i}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\bar{\beta}_i'\bar{v}_{\beta_i}^{-1}\bar{\beta}_i\right)} \\ = \; & (2\pi)^{-\frac{T}{2}}|\Sigma_{i,u}|^{-\frac{1}{2}}\left(\frac{|\underline{v}_{\beta_i}|}{|\bar{v}_{\beta_i}|}\right)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\check{Y}_i^{\beta\prime}\Sigma_{i,u}^{-1}\check{Y}_i^{\beta}+\frac{1}{2}\bar{\beta}_i'\bar{v}_{\beta_i}^{-1}\bar{\beta}_i\right), \end{aligned}$$

where $\check{Y}_i^{\beta}=Y_i-X_i\alpha$. $\bar{\beta}_i$ and $\bar{v}_{\beta_i}$ are posterior mean and variance of β_i:

$$\bar{v}_{\beta_i}=\left(\underline{v}_{\beta_i}^{-1}+W_i'\Sigma_{i,u}^{-1}W_i\right)^{-1},\quad \bar{\beta}_i=\bar{v}_{\beta_i}\left(W_i'\Sigma_{i,u}^{-1}\check{Y}_i^{\beta}\right).$$

For $z_i^\alpha=0$ we have $\beta_i'=[0,\ 0,\ s_{i1},\ldots,s_{iT}]$. Then,

$$\begin{aligned} & p(Y_i|\alpha,\rho,\delta^\rho,v_{\delta^\alpha},\sigma^2_{u,1:T},\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_u},\delta_i^{\sigma_\epsilon},v_{s_0},z_i^\alpha=0) \\ = \; & \frac{(2\pi)^{-\frac{T}{2}}|\Sigma_{i,u}|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\check{Y}_i^{\beta\prime}\Sigma_{i,u}^{-1}\check{Y}_i^{\beta}\right)(2\pi)^{-\frac{T}{2}}\,\left|\underline{V}_{s_i}\right|^{-\frac{1}{2}}}{(2\pi)^{-\frac{T}{2}}\,\left|\overline{V}_{s_i}\right|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\bar{s}_i'\overline{V}_{s_i}^{-1}\bar{s}_i\right)} \\ = \; & (2\pi)^{-\frac{T}{2}}|\Sigma_{i,u}|^{-\frac{1}{2}}\left(\frac{|\underline{V}_{s_i}|}{|\overline{V}_{s_i}|}\right)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\check{Y}_i^{\beta\prime}\Sigma_{i,u}^{-1}\check{Y}_i^{\beta}+\frac{1}{2}\bar{s}_i'\overline{V}_{s_i}^{-1}\bar{s}_i\right), \end{aligned}$$

where $\bar{s}_i$ and $\overline{V}_{s_i}$ are posterior mean and variance of s_i when $\delta_i^\alpha=0$:

$$\overline{V}_{s_i}=\left(\underline{V}_{s_i}^{-1}+\Sigma_{i,u}^{-1}\right)^{-1},\quad \bar{s}_i=\overline{V}_{s_i}\left(\Sigma_{i,u}^{-1}\check{Y}_i^{\beta}\right).$$

We conclude that the posterior odds of $z_i^\alpha=1$ vs. $z_i^\alpha=0$ are

$$K_i^\alpha=\frac{q^\alpha}{1-q^\alpha}\left(\frac{|\underline{v}_{\beta_i}||\overline{V}_{s_i}|}{|\bar{v}_{\beta_i}||\underline{V}_{s_i}|}\right)^{-\frac{1}{2}}\exp\left(\frac{1}{2}\bar{\beta}_i'\bar{v}_{\beta_i}^{-1}\bar{\beta}_i-\frac{1}{2}\bar{s}_i'\overline{V}_{s_i}^{-1}\bar{s}_i\right). \tag{A.13}$$

We can draw z_i^α as follows:

$$z_i^\alpha|(Y_i,\alpha,\rho,\delta^\rho,q^\alpha,v_{\delta^\alpha},\sigma^2_{u,1:T},\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_u},\delta_i^{\sigma_\epsilon},v_{s_0}) = \begin{cases} 1 & \text{with prob. } \frac{K_i^\alpha}{K_i^\alpha+1} \\ 0 & \text{with prob. } \frac{1}{K_i^\alpha+1} \end{cases}. \tag{A.14}$$

Moreover,

$$\beta_i|(Y_i,\alpha,\rho,\delta^\rho,v_{\delta^\alpha},\sigma^2_{u,1:T},\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_u},\delta_i^{\sigma_\epsilon},v_{s_0},z_i^\alpha) \sim \begin{cases} \mathcal{N}\left(\bar\beta_i,\bar v_{\beta_i}\right) & \text{if } z_i^\alpha=1 \\ \delta_i^\alpha=0 \text{ and } s_i\sim\mathcal{N}\left(\bar s_i,\overline{V}_{s_i}\right) & \text{if } z_i^\alpha=0 \end{cases}. \tag{A.15}$$

Conditional Posterior of (z_i^ρ,δ_i^ρ). Conditional on the previous block draws, updating (z_i^ρ,δ_i^ρ) is univariate in δ_i^ρ and follows the same spike-and-slab logic as above. Sampling from the conditional distribution $z_i^\rho,\delta_i^\rho|(s_{i,0:T},q^\rho,\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_\epsilon},v_{\delta^\rho})$ proceeds in the same manner. We define $\breve{s}_{it}^\rho = s_{it}-\rho s_{it-1}$ and $\sigma_{it}^\epsilon = \sigma_{\epsilon,t}\sqrt{\delta_i^{\sigma_\epsilon}}$ and let $\bar\delta_i^\rho$ and $\bar v_{\delta_i^\rho}$ be the posterior mean and variance of δ_i^ρ:

$$\bar v_{\delta_i^\rho} = \left(v_{\delta^\rho}^{-1}+\sum_{t=1}^T(\sigma_{it}^\epsilon)^{-2}s_{it-1}^2\right)^{-1}, \quad \bar\delta_i^\rho = \bar v_{\delta_i^\rho}\sum_{t=1}^T(\sigma_{it}^\epsilon)^{-2}s_{it-1}\breve{s}_{it}^\rho.$$

The posterior odds of $z_i^\rho=1$ vs. $z_i^\rho=0$ are

$$K_i^\rho = \frac{q^\rho}{1-q^\rho}\left(\frac{v_{\delta^\rho}}{\bar v_{\delta_i^\rho}}\right)^{-\frac{1}{2}}\exp\left(\frac{(\bar\delta_i^\rho)^2}{2\bar v_{\delta_i^\rho}}\right). \tag{A.16}$$

We can draw z_i^ρ as follows:

$$z_i^\rho|(s_{i,0:T},q^\rho,\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_\epsilon},v_{\delta^\rho}) = \begin{cases} 1 & \text{with prob. } \frac{K_i^\rho}{K_i^\rho+1} \\ 0 & \text{with prob. } \frac{1}{K_i^\rho+1} \end{cases}. \tag{A.17}$$

Moreover,

$$\delta_i^\rho|(s_{i,0:T},z_i^\rho,\sigma^2_{\epsilon,1:T},\delta_i^{\sigma_\epsilon},v_{\delta^\rho}) \sim \begin{cases} \mathcal{N}\left(\bar\delta_i^\rho,\bar v_{\delta_i^\rho}\right) & \text{if } z_i^\rho=1 \\ 0 & \text{if } z_i^\rho=0 \end{cases}. \tag{A.18}$$

Conditional Posterior of $(z_i^{\sigma_u},\delta_i^{\sigma_u})$ and $(z_i^{\sigma_\epsilon},\delta_i^{\sigma_\epsilon})$. Next, we update heterogeneity in the shock variances. Define

$$\breve{y}_{it}^{\sigma_u} = y_{it}-x_{it}'(\alpha+\delta_i^\alpha)-s_{it} \quad \text{and} \quad \breve{y}_{it}^{\sigma_\epsilon} = s_{it}-(\rho+\delta_i^\rho)s_{it-1}.$$

We show the derivation for $(z_i^{\sigma_m}, \delta_i^{\sigma_m})$ for $m \in \{u, \epsilon\}$ as they are identical except for subscripts. If $z_i^{\sigma_m} = 1$, then $\delta_i^{\sigma_m}$ has an Inverse Gamma distribution. The marginalized likelihood of $\breve{Y}_i^{\sigma_m} | (\alpha, \rho, \delta_i^{\alpha}, \delta_i^{\rho}, \sigma^2_{m,1:T}, s_{i,1:T}, v_{\delta^{\sigma_m}}, z_i^{\sigma_m})$ is:

$$
\begin{aligned}
& p(\breve{Y}_i^{\sigma_m} | \alpha, \rho, \delta_i^{\alpha}, \delta_i^{\rho}, \sigma^2_{m,1:T}, s_{i,1:T}, v_{\delta^{\sigma_m}}, z_i^{\sigma_m} = 1) \\
= & \int_0^{\infty} p(\breve{Y}_i^{\sigma_m} | \alpha, \rho, \delta_i^{\alpha}, \delta_i^{\rho}, \sigma^2_{m,1:T}, s_{i,1:T}, v_{\delta^{\sigma_m}}, \delta_i^{\sigma_m}, z_i^{\sigma_m} = 1) p(\delta_i^{\sigma_m} | v_{\delta^{\sigma_m}}, z_i^{\sigma_m} = 1) d\delta_i^{\sigma_m} \\
= & \int_0^{\infty} \prod_{t=1}^{T} \sigma_{m,t}^{-1} \left(2\pi\delta_i^{\sigma_m}\right)^{-\frac{T}{2}} \exp\left(-\frac{\sum_{t=1}^{T} (\breve{y}_{it}^{\sigma_m})^2 / \sigma^2_{m,t}}{2\delta_i^{\sigma_m}}\right) \frac{(v_{\delta^{\sigma_m}}^{-1} + 1)^{\left(v_{\delta^{\sigma_m}}^{-1} + 2\right)}}{\Gamma(v_{\delta^{\sigma_m}}^{-1} + 2)} \\
& \times \left(\frac{1}{\delta_i^{\sigma_m}}\right)^{v_{\delta^{\sigma_m}}^{-1} + 3} \exp\left(-\frac{v_{\delta^{\sigma_m}}^{-1} + 1}{\delta_i^{\sigma_m}}\right) d\delta_i^{\sigma_m} \\
= & \prod_{t=1}^{T} \left(2\pi\sigma^2_{m,t}\right)^{-\frac{1}{2}} \frac{\Gamma\left(\frac{\bar{\nu}_{\delta^{\sigma_m}}}{2}\right)}{\Gamma\left(v_{\delta^{\sigma_m}}^{-1} + 2\right)} \frac{\left(v_{\delta^{\sigma_m}}^{-1} + 1\right)^{v_{\delta^{\sigma_m}}^{-1} + 2}}{\left(\frac{\bar{\tau}_{\delta_i^{\sigma_m}}}{2}\right)^{\frac{\bar{\nu}_{\delta^{\sigma_m}}}{2}}}.
\end{aligned}
$$

where $\bar{\nu}_{\delta^{\sigma_m}} = 2v_{\delta^{\sigma_m}}^{-1} + 4 + T$ and $\bar{\tau}_{\delta_i^{\sigma_m}} = 2v_{\delta^{\sigma_m}}^{-1} + 2 + \sum_{t=1}^{T} \sigma_{m,t}^{-2} (\breve{y}_{it}^{\sigma_m})^2$.

For $z_i^{\sigma_m} = 0$, we have $\delta_i^{\sigma_m} = 1$ and

$$
p(\breve{Y}_i^{\sigma_m} | \alpha, \rho, \delta_i^{\alpha}, \delta_i^{\rho}, \sigma^2_{m,1:T}, s_{i,1:T}, v_{\delta^{\sigma_m}}, z_i^{\sigma_m} = 0) = \prod_{t=1}^{T} \left(2\pi\sigma^2_{m,t}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \sum_{t=1}^{T} \sigma_{m,t}^{-2} (\breve{y}_{it}^{\sigma_m})^2\right). \tag{A.19}
$$

For $m \in \{u, \epsilon\}$, the posterior odds of $z_i^{\sigma_m} = 1$ vs. $z_i^{\sigma_m} = 0$ are

$$
K_{i,m}^{\sigma} = \frac{q^{\sigma_m}}{1 - q^{\sigma_m}} \frac{\Gamma\left(\frac{\bar{\nu}_{\delta^{\sigma_m}}}{2}\right)}{\Gamma\left(v_{\delta^{\sigma_m}}^{-1} + 2\right)} \frac{\left(v_{\delta^{\sigma_m}}^{-1} + 1\right)^{v_{\delta^{\sigma_m}}^{-1} + 2}}{\left(\frac{\bar{\tau}_{\delta_i^{\sigma_m}}}{2}\right)^{\frac{\bar{\nu}_{\delta^{\sigma_m}}}{2}}} \exp\left(\frac{1}{2} \sum_{t=1}^{T} \sigma_{m,t}^{-2} (\breve{y}_{it}^{\sigma_m})^2\right). \tag{A.20}
$$

Based on the posterior odds $K_{i,m}^{\sigma}$ we can draw $z_i^{\sigma_m}$ as follows:

$$
z_i^{\sigma_m} | (Y_i, \alpha, \rho, \delta_i^{\alpha}, \delta_i^{\rho}, \sigma^2_{m,1:T}, s_{i,1:T}, q^{\sigma_m}, v_{\delta^{\sigma_m}}) = \begin{cases} 1 & \text{with prob. } \frac{K_{i,m}^{\sigma}}{K_{i,m}^{\sigma} + 1} \\ 0 & \text{with prob. } \frac{1}{K_{i,m}^{\sigma} + 1} \end{cases}. \tag{A.21}
$$

If $z_i^{\sigma_m} = 1$, the conditional posterior of $\delta_i^{\sigma_m}$ is given by

$$\begin{aligned} & p(\delta_i^{\sigma_m}|Y_i, \alpha, \rho, \delta_i^{\alpha}, \delta_i^{\rho}, \sigma^2_{m,1:T}, s_{i,1:T}, v_{\delta^{\sigma_m}}, z_i^{\sigma_m} = 1) \\ \propto & \left(\frac{1}{\delta_i^{\sigma_m}}\right)^{\frac{T}{2}} \exp\left(-\frac{\sum_{t=1}^{T} \sigma_{m,t}^{-2}(\breve{y}_{it}^{\sigma_m})^2}{2\delta_i^{\sigma_m}}\right)\left(\frac{1}{\delta_i^{\sigma_m}}\right)^{v_{\delta^{\sigma_m}}^{-1}+3} \exp\left(-\frac{v_{\delta^{\sigma_m}}^{-1}+1}{\delta_i^{\sigma_m}}\right) \\ \propto & \left(\frac{1}{\delta_i^{\sigma_m}}\right)^{\frac{T}{2}+v_{\delta^{\sigma_m}}^{-1}+3} \exp\left(-\frac{\sum_{t=1}^{T} \sigma_{m,t}^{-2}(\breve{y}_{it}^{\sigma_m})^2 + 2v_{\delta^{\sigma_m}}^{-1}+2}{2\delta_i^{\sigma_m}}\right). \end{aligned}$$

If $z_i^{\sigma_m} = 0$ then $\delta_i^{\sigma_m} = 1$. Overall, we obtain

$$\delta_i^{\sigma_m}|(Y_i, \alpha, \rho, \delta_i^{\alpha}, \delta_i^{\rho}, \sigma^2_{m,1:T}, s_{i,1:T}, v_{\delta^{\sigma_m}}, z_i^{\sigma_m}) \sim \begin{cases} IG\left(\frac{\bar{\nu}_{\delta^{\sigma_m}}}{2}, \frac{\bar{\tau}_{\delta_i^{\sigma_m}}}{2}\right) & \text{if } z_i^{\sigma_m} = 1 \\ 1 & \text{if } z_i^{\sigma_m} = 0 \end{cases}. \tag{A.22}$$

Conditional Posterior of μ_{s_0} and v_{s_0}. Given the updated latent state path, the initial-state moments are conjugate and can be sampled directly.

$$\mu_{s_0}|(s_{i0}, v_{s_0}) \sim \mathcal{N}(\bar{\mu}_{s_0}, \bar{v}_{s_0}), \tag{A.23}$$

where

$$\begin{aligned} \bar{v}_{s_0} &= \left(\underline{v}_{s_0}^{-1} + N v_{s_0}^{-1}\right)^{-1} \\ \bar{\mu}_{s_0} &= \bar{v}_{s_0}\left(\underline{v}_{s_0}^{-1}\underline{\mu}_{s_0} + v_{s_0}^{-1}\sum_{i=1}^{N} s_{i0}\right). \end{aligned}$$

$$v_{s_0}|(s_{i0}, \mu_{s_0}) \sim IG\left(\frac{\bar{\nu}_{s_0}}{2}, \frac{\bar{\tau}_{s_0}}{2}\right), \tag{A.24}$$

where

$$\begin{aligned} \bar{\nu}_{s_0} &= \underline{\nu}_{s_0} + N \\ \bar{\tau}_{s_0} &= \underline{\tau}_{s_0} + \sum_{i=1}^{N}\left(s_{i0} - \mu_{s_0}\right)^2. \end{aligned}$$

Conditional Posterior of s_{i0}. We sample s_{i0} using a simulation smoother conditional on the posterior draws of s_{i1} and other parameters:

$$s_{i0}|(s_{i1}, \mu_{s_0}, v_{s_0}, \rho, \delta_i^{\rho}, \sigma^2_{\epsilon,1}, \delta_i^{\sigma_\epsilon}) \sim \mathcal{N}(\bar{s}_{i,0|1}, P_{i,0|1}), \tag{A.25}$$

where

$$\begin{aligned}
\bar{s}_{i,0|1} &= \bar{s}_{i,0|0} + P_{i,0|0}\Phi_{i1}' P_{i,1|0}^{-1}\left(s_{i1} - \Phi_{i1}\bar{s}_{i,0|0}\right), \\
P_{i,0|1} &= P_{i,0|0} - P_{i,0|0}\Phi_{i1}' P_{i,1|0}^{-1}\Phi_{i1}P_{i,0|0}, \\
P_{i,1|0} &= \Phi_{i1}' P_{i,0|0}\Phi_{i1} + \sigma_{\epsilon,1}^2\delta_i^{\sigma_\epsilon}, \\
\Phi_{i1} &= \rho + \delta_i^{\rho}, \\
\bar{s}_{i,0|0} &= \mu_{s_0}, \\
P_{i,0|0} &= v_{s_0}.
\end{aligned}$$

Conditional Posterior of $(\sigma_{u,1:T}^2, \sigma_{\epsilon,1:T}^2)$. Finally, conditional on current draws of coefficients, discrepancies, and latent states, each time-specific variance parameter has a standard Inverse Gamma conditional posterior. For $\sigma_{u,t}^2$,

$$\begin{aligned}
&p(\sigma_{u,t}^2|y_{1:N,t}, s_{1:N,t}, \alpha, \delta_{1:N}^{\alpha}, \delta_{1:N}^{\sigma_u}) \\
&\quad\propto \left(\frac{1}{\sigma_{u,t}^2}\right)^{\frac{\underline{\nu}_{\sigma_u}}{2}+1} \exp\left(-\frac{\underline{\tau}_{\sigma_u}}{2\sigma_{u,t}^2}\right)\left(\frac{1}{\sigma_{u,t}^2}\right)^{\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma_{u,t}^2}\sum_{i=1}^{N}\left[y_{it} - x_{it}'\left(\alpha + \delta_i^{\alpha}\right) - s_{it}\right]^2/\delta_i^{\sigma_u}\right\} \\
&\quad= \left(\frac{1}{\sigma_{u,t}^2}\right)^{\frac{\underline{\nu}_{\sigma_u}+N}{2}+1} \exp\left\{-\frac{1}{2\sigma_{u,t}^2}\left\{\underline{\tau}_{\sigma_u} + \sum_{i=1}^{N}\left[y_{it} - x_{it}'\left(\alpha + \delta_i^{\alpha}\right) - s_{it}\right]^2/\delta_i^{\sigma_u}\right\}\right\}.
\end{aligned}$$

This implies

$$\sigma_{u,t}^2|(y_{1:N,t}, s_{1:N,t}, \alpha, \delta_{1:N}^{\alpha}, \delta_{1:N}^{\sigma_u}) \sim IG\left(\frac{\bar{\nu}_{\sigma_u}}{2}, \frac{\bar{\tau}_{\sigma_{u,t}}}{2}\right), \tag{A.26}$$

where

$$\begin{aligned}
\bar{\nu}_{\sigma_u} &= \underline{\nu}_{\sigma_u} + N, \\
\bar{\tau}_{\sigma_{u,t}} &= \underline{\tau}_{\sigma_u} + \sum_{i=1}^{N}\left[y_{it} - x_{it}'\left(\alpha + \delta_i^{\alpha}\right) - s_{it}\right]^2/\delta_i^{\sigma_u}.
\end{aligned}$$

For $\sigma^2_{\epsilon,t}$,

$$\begin{aligned}
&p(\sigma^2_{\epsilon,t}|s_{1:N,t}, s_{1:N,t-1}, \rho, \delta^{\rho}_{1:N}, \delta^{\sigma_\epsilon}_{1:N}) \\
&\quad \propto \left(\frac{1}{\sigma^2_{\epsilon,t}}\right)^{\frac{\underline{\nu}_{\sigma_\epsilon}}{2}+1} \exp\left(-\frac{\underline{\tau}_{\sigma_\epsilon}}{2\sigma^2_{\epsilon,t}}\right) \left(\frac{1}{\sigma^2_{\epsilon,t}}\right)^{\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2_{\epsilon,t}} \sum_{i=1}^{N} \left[s_{it} - (\rho + \delta^{\rho}_i)\, s_{it-1}\right]^2 / \delta^{\sigma_\epsilon}_i\right\} \\
&\quad = \left(\frac{1}{\sigma^2_{\epsilon,t}}\right)^{\frac{\underline{\nu}_{\sigma_\epsilon}+N}{2}+1} \exp\left\{-\frac{1}{2\sigma^2_{\epsilon,t}} \left\{\underline{\tau}_{\sigma_\epsilon} + \sum_{i=1}^{N} \left[s_{it} - (\rho + \delta^{\rho}_i)\, s_{it-1}\right]^2 / \delta^{\sigma_\epsilon}_i\right\}\right\}.
\end{aligned}$$

This implies

$$\sigma^2_{\epsilon,t}|(s_{1:N,t}, s_{1:N,t-1}, \rho, \delta^{\rho}_{1:N}, \delta^{\sigma_\epsilon}_{1:N}) \sim IG\left(\frac{\bar{\nu}_{\sigma_\epsilon}}{2}, \frac{\bar{\tau}_{\sigma_{\epsilon,t}}}{2}\right), \tag{A.27}$$

where

$$\begin{aligned}
\bar{\nu}_{\sigma_\epsilon} &= \underline{\nu}_{\sigma_\epsilon} + N, \\
\bar{\tau}_{\sigma_{\epsilon,t}} &= \underline{\tau}_{\sigma_\epsilon} + \sum_{i=1}^{N} \left[s_{it} - (\rho + \delta^{\rho}_i)\, s_{it-1}\right]^2 / \delta^{\sigma_\epsilon}_i.
\end{aligned}$$

B Convergence Diagnostics

The posterior sampler generates draws of three parameter groups and the latent states: (i) the homogeneous parameters θ, (ii) the discrepancies δ_i^{α}, δ_i^{ρ}, $\delta_i^{\sigma_u}$, and $\delta_i^{\sigma_\epsilon}$, (iii) the sparsity indicators z_i^l, $l \in \{\alpha, \rho, \sigma_u, \sigma_\epsilon\}$, and (iv) the latent state sequences $S_{i,0:T}$. To assess the convergence of the posterior sampler, we run multiple MCMC chains on a single data set simulated from the Monte Carlo design of Table 1, with $N = 500$, $T = 10$, $q^{\alpha} = q^{\rho} = q^{\sigma_u} = q^{\sigma_\epsilon} = 0.8$, and $c = 1$. The runs are initialized using an overdispersed set of starting values.

We show trace plots of the draws for ten group (i) parameters in Figure A-1 and recursive averages of the draws (without burn-in), defined as

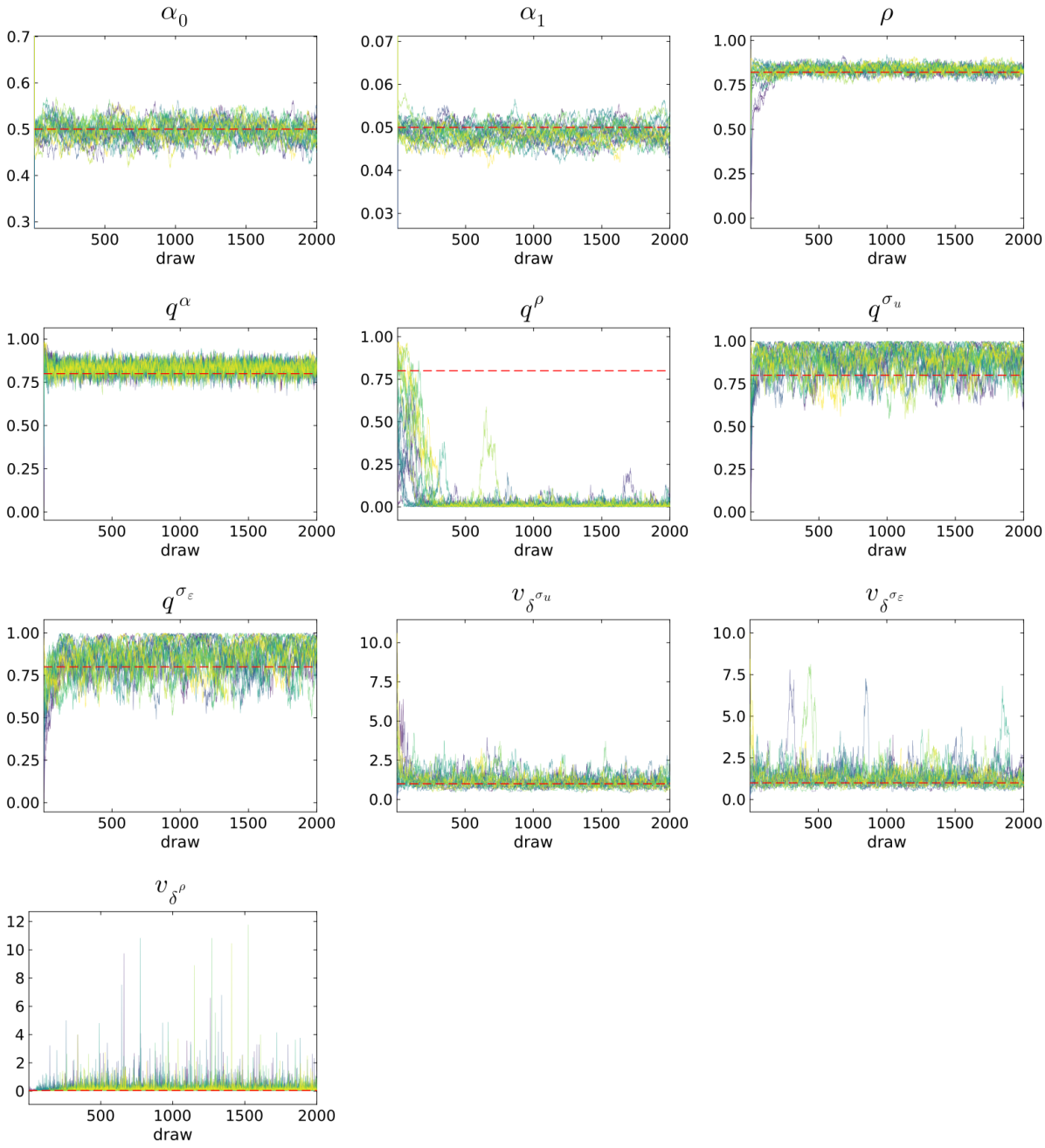
$$\bar{\theta}_k(N_{\text{MCMC}}) = \frac{1}{N_{\text{MCMC}}} \sum_{j=1}^{N_{\text{MCMC}}} \theta_k^j$$

in Figure A-2. Here θ_k^j denotes the kth element of θ in the jth draw. Each hairline corresponds to a different MCMC run. After 2,000 draws, the variation of the Monte Carlo averages is fairly small, indicating that even with a low number of draws the posterior means of the θ elements are well approximated. Because the chains have converged after 2,000 draws, the 5,000 draws with a burn-in of 2,500 used in Section 4 strike a good balance between accuracy and computational complexity.

For this simulated sample there are substantial discrepancies between the true $(q^{\rho}, v_{\delta^{\rho}})$ and the posterior mean estimates. However, an inspection of the plots suggests that this is a feature of the posterior rather than a deficiency of the sampler. The posterior estimates imply a low probability of fairly large outliers whereas the DGP features a higher probability of small outliers. We verified that in a simulation with $N = 2{,}500$ and $T = 20$ this discrepancy vanishes.

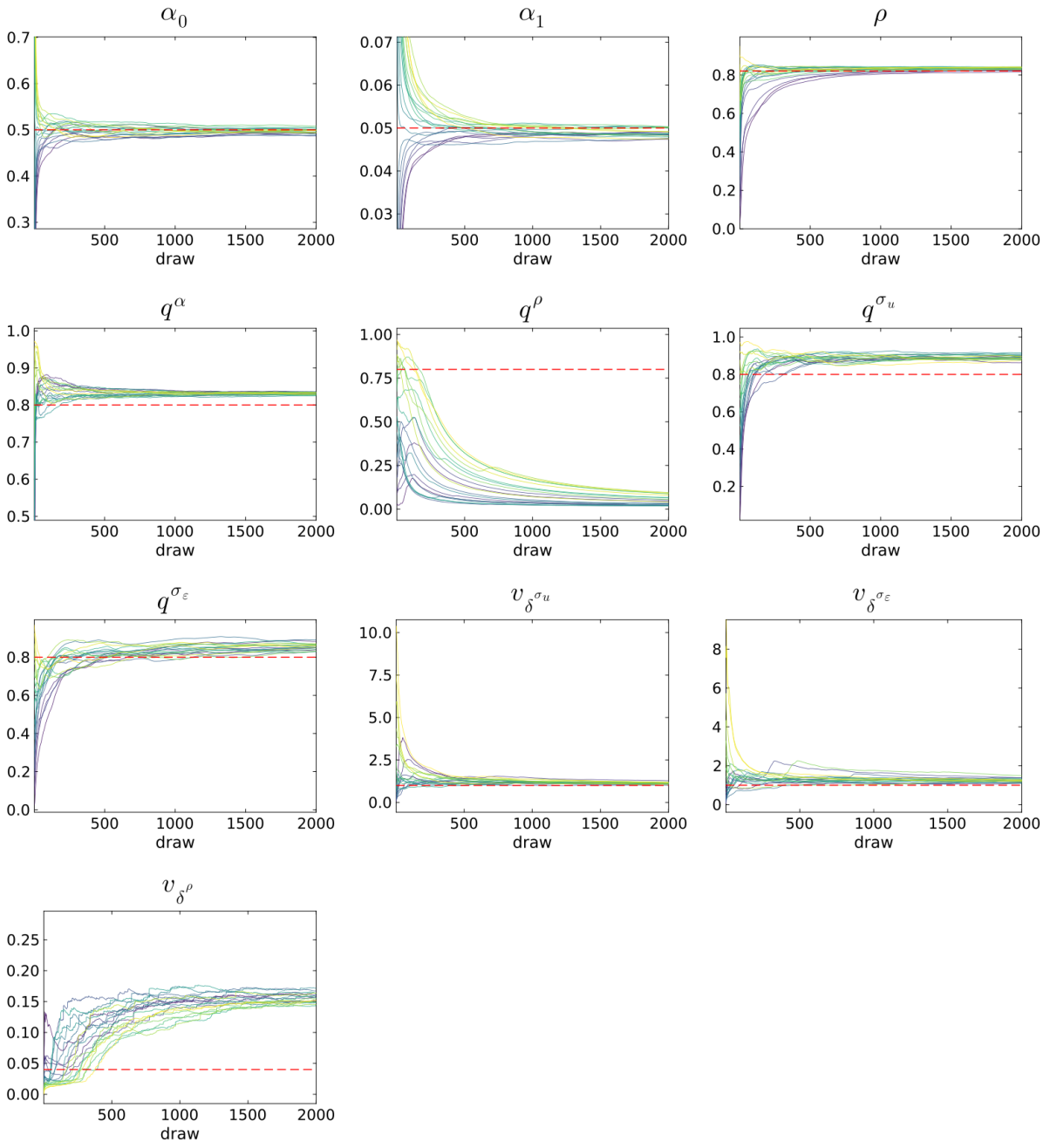
In addition to the homogeneous parameters we have a large number of idiosyncratic parameters for which it is not feasible to display trace plots and recursive means. Instead, we report histograms of Gelman–Rubin $\hat{R}$ statistics (Gelman and Rubin, 1992) based on the 20 MCMC chains in Figure A-3. For all parameters except ρ_i the cross-sectional distribution (across units i) is tightly concentrated around one. For ρ_i the distribution is centered at 1.02 and covers values between 1.00 and 1.04, all below the rule-of-thumb cutoff value of 1.1.

Figure A-1: MCMC Diagnostics: Trace Plots



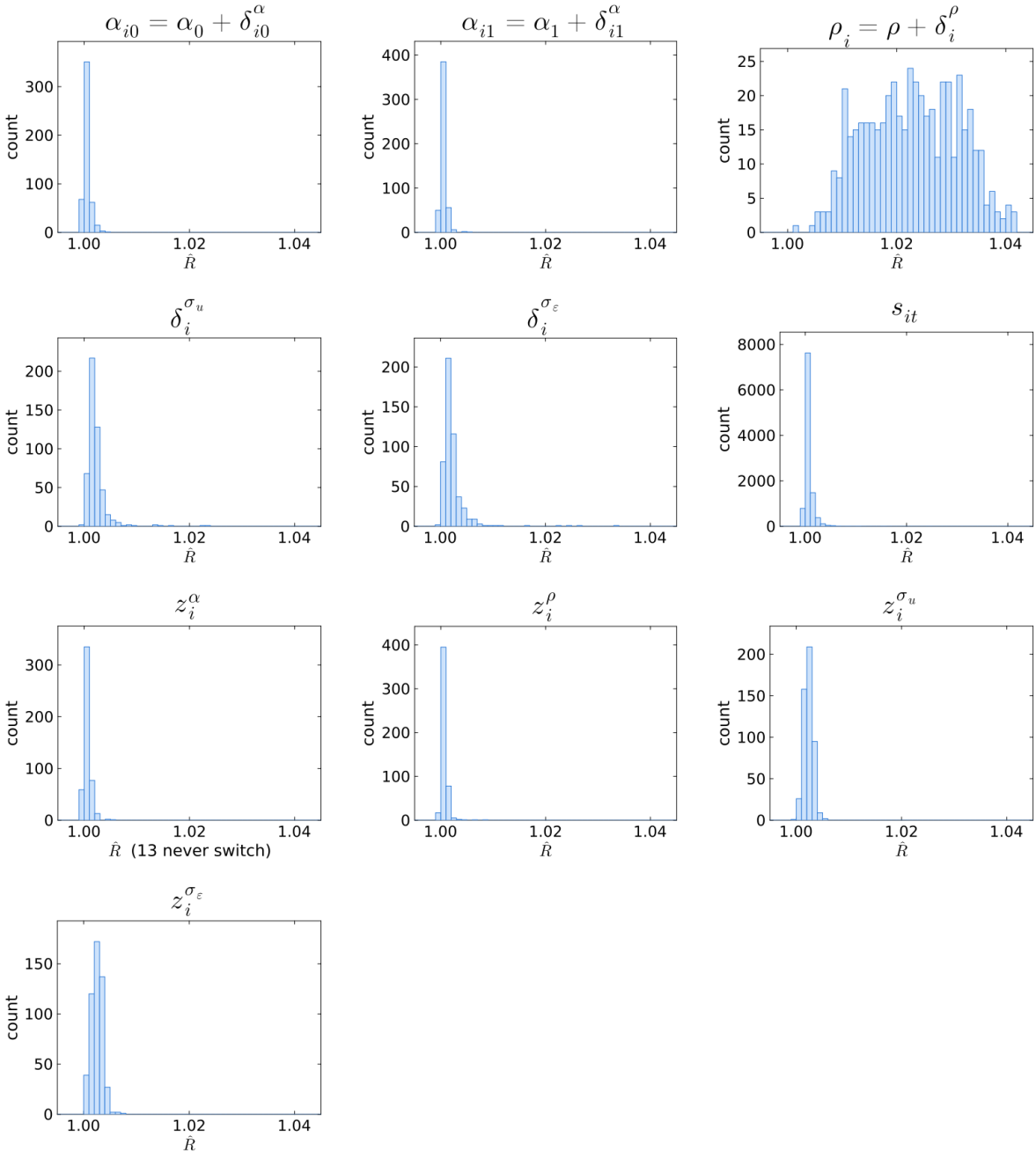
Notes: Each panel shows 2,000 draws of 20 MCMC chains for one sample generated from the ($N = 500$, $T = 10$, $q = 0.8$, $c = 1$) DGP; no draws are discarded. The dashed line marks the design value used to simulate the data. For α_0 and α_1 the vertical axis is zoomed, so that the starting values of some chains lie outside the plotted range. The parameters of the distribution of the initial state s_{i0} are omitted.

Figure A-2: MCMC Diagnostics: Recursive Means, No Burn-In



Notes: Each panel shows recursive means of 2,000 draws for each of 20 MCMC chains for one sample generated from the ($N = 500$, $T = 10$, $q = 0.8$, $c = 1$) DGP; no draws are discarded. The dashed line marks the design value used to simulate the data. For α_0 and α_1 the vertical axis is zoomed, so that the starting values of some chains lie outside the plotted range. The parameters of the distribution of the initial state s_{i0} are omitted.

Figure A-3: MCMC Diagnostics: Gelman–Rubin Statistics for the Unit-Level Parameters



Notes: Histograms of the Gelman–Rubin statistic $\hat{R}$, computed from the second half of each of the 20 chains of Figure A-1, over all 500 units and, for the latent states, over all periods $t = 0, \ldots, T$.

C Additional Empirical Results

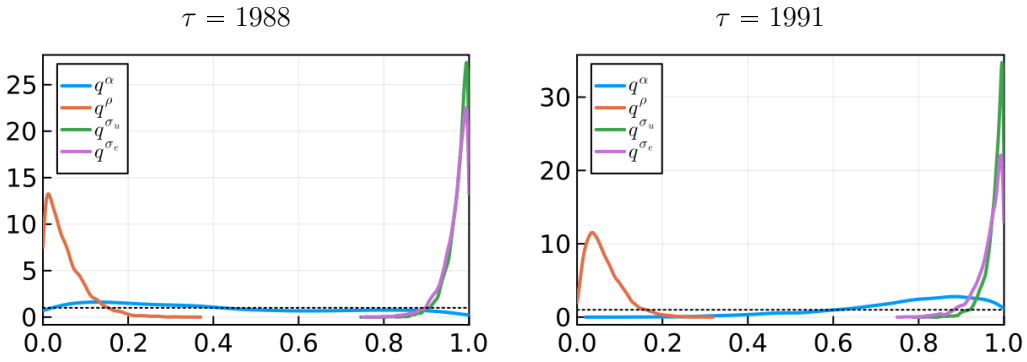
Further Results on Core Groups versus Deviators (Balanced Panels). Table A-1 shows the numbers of units that share the same posterior median, i.e., how long the flat segments in Figure 2 are.[13] N is the number of individuals in the subsample and N_l is the size of the core group for parameter $l \in \{\alpha, \rho, \sigma_u, \sigma_\epsilon\}$.

Table A-1: Size of Core Groups

τ	N	N_α	N_α/N	N_ρ	N_ρ/N	N_{σ_u}	N_{σ_u}/N	N_{σ_ϵ}	N_{σ_ϵ}/N
1988	82	74	90.24%	82	100%	0	0%	0	0%
1991	88	32	36.36%	87	98.86%	0	0%	0	0%

In Figure A-4 we plot prior densities, which we take to be uniform, and posterior densities for the q parameters that determine the probability of deviating from the core group. The posterior mode for q^ρ is close to zero for both samples, which is consistent with the ρ_i estimates being identical for most units, as could be seen in Figure 2. The posterior density for q^α stays close to the prior density and is fairly flat. It has a small mode near 0.1 for the 1988 sample, and a small mode near 0.9 for the 1991 sample.

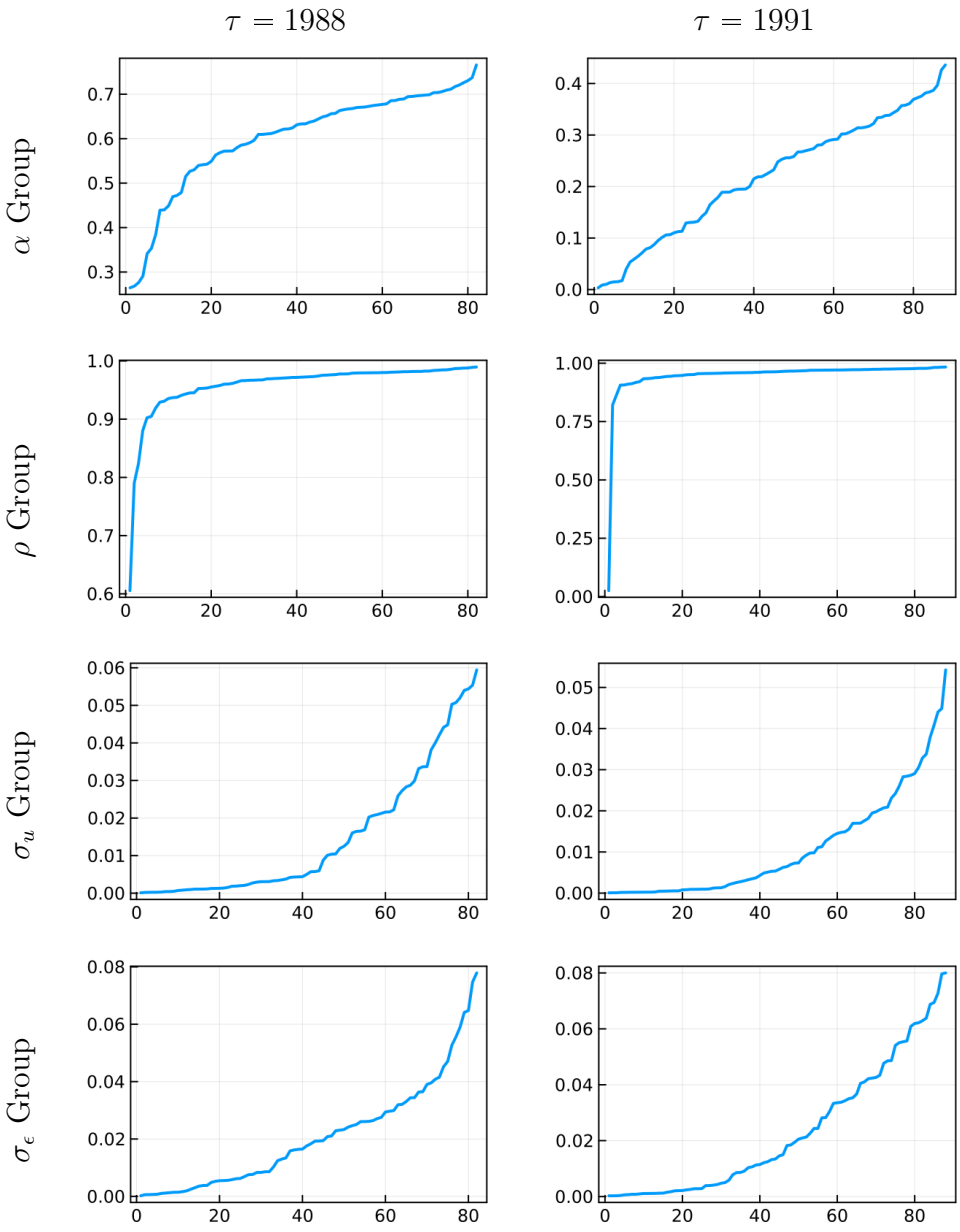
Figure A-4: Prior and Posterior Densities of q, $T = 20$



Notes: The black dotted line corresponds to the $B(1,1)$ prior density.

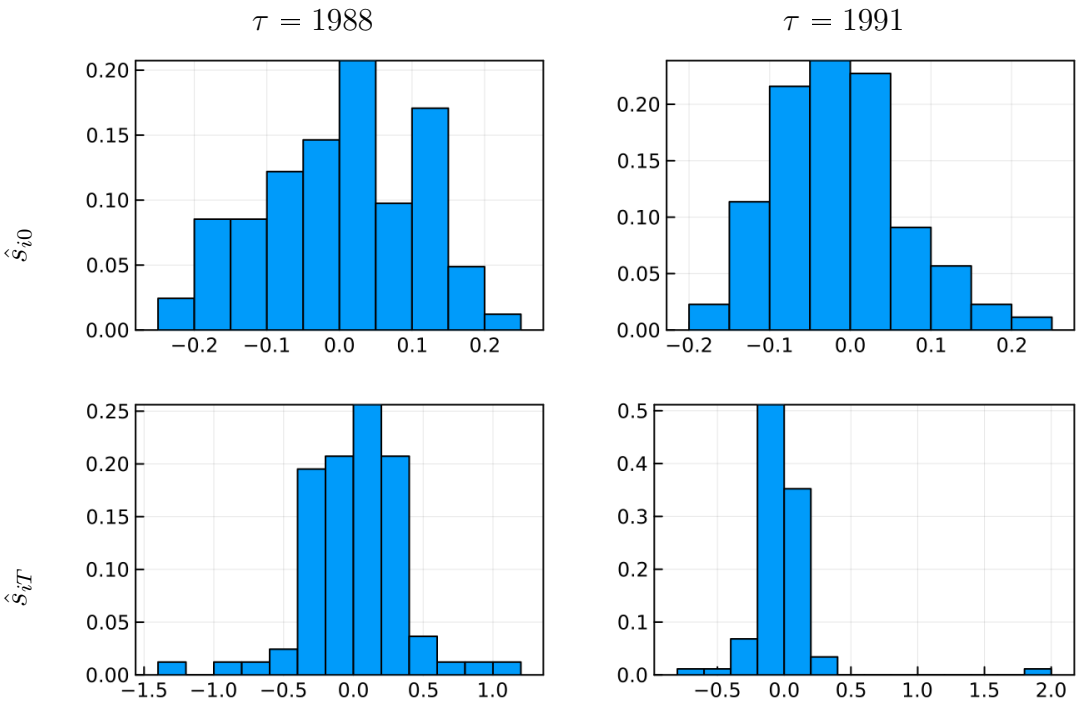
[13] This is slightly different from the posterior mode of the core-group size shown in Figure 3.

Figure A-5: Posterior Probability of Unit i Belonging to Core Group, $T = 20$



In Figure A-5 we plot for each unit the posterior probability of belonging to the core group, defined as $\mathbb{P}\{z_i^l = 0 | Y_{1:N,1:T}\}$.

Figure A-6: Histograms of Posterior Means $\hat{s}_{i0}$ and $\hat{s}_{iT}$, $T = 20$



In Figure A-6 we plot the cross-sectional distribution of the posterior mean estimates $\hat{s}_{i0}$ and $\hat{s}_{iT}$.

Table A-2: Descriptive Statistics for Experience in the Starting Period

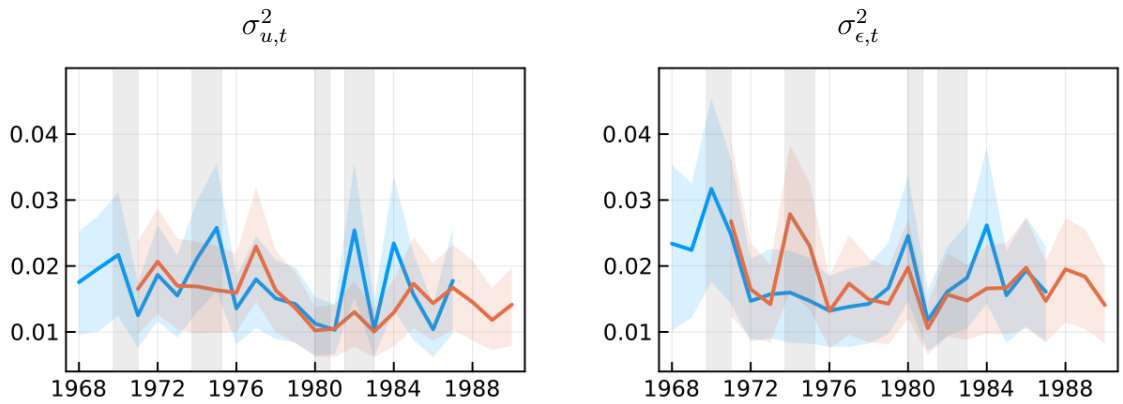
τ	Mean	Std	Min	25%	Median	75%	Max
1988	10.27	4.22	2	7	10	13	20
1991	9.39	3.92	2	7	9	11	20

Table A-2 provides descriptive statistics of the distribution of years of experience in the starting period of the two samples considered in the main text.

Time Path of Average Idiosyncratic Income Risk (Balanced Panels). Figure A-7 depicts posterior means and credible intervals for $\sigma^2_{u,t}$ and $\sigma^2_{\epsilon,t}$ as a function of time. We overlay the posteriors for the two samples: $\tau = 1988$ and $\tau = 1991$. Years with quarters that were classified by the NBER as recessions are indicated by the grey shaded areas. It is important to note that the cross sections for the two panels in any particular year are not identical. This explains the variation in the estimates of the year-specific standard deviations.

While the volatility spikes do not line up perfectly with the NBER recessions, there is a positive correlation. For the 1988 sample $\sigma^2_{u,t}$ spikes in the 1970, 1974, and 1982 recessions. On the other hand, it does not spike in the 1980 recession, but does have a pronounced peak during the recovery in 1984. The innovation variance $\sigma^2_{\epsilon,t}$ peaks in 1970 and 1980, but not in the other recessions. Note that individuals in our sample are always employed. Thus, we are not capturing unemployment risk, which contributes to the cyclical variation of income risk.

Figure A-7: Evolution of Average Idiosyncratic Volatility



Notes: Posterior means and 90% credible intervals. Blue: $\tau = 1988$; Red: $\tau = 1991$. Grey shading denotes NBER-dated recessions.

Estimates of Initial States (Balanced Panels). Posteriors for the mean and variance of the distribution of s_{i0} are reported in Table A-3. The estimated mean of the initial state s_{i0} is zero. Using the posterior median for v_{s_0} from the 1988 sample and ignoring the estimation uncertainty, the standard deviation of s_{i0} is 0.17. A back-of-the-envelope calculation using posterior median estimates of ρ, $v_{\delta\sigma_\epsilon}$, and the average level of $\sigma^2_{\epsilon,t}$, leads to an unconditional standard deviation of s_{it} of 1.25. Thus, focusing on a particular cohort, over time the income inequality due to the persistent stochastic component increases as the variance of s_{it} rises

Table A-3: Posterior Estimates for the Initial-State Distribution, $T = 20$

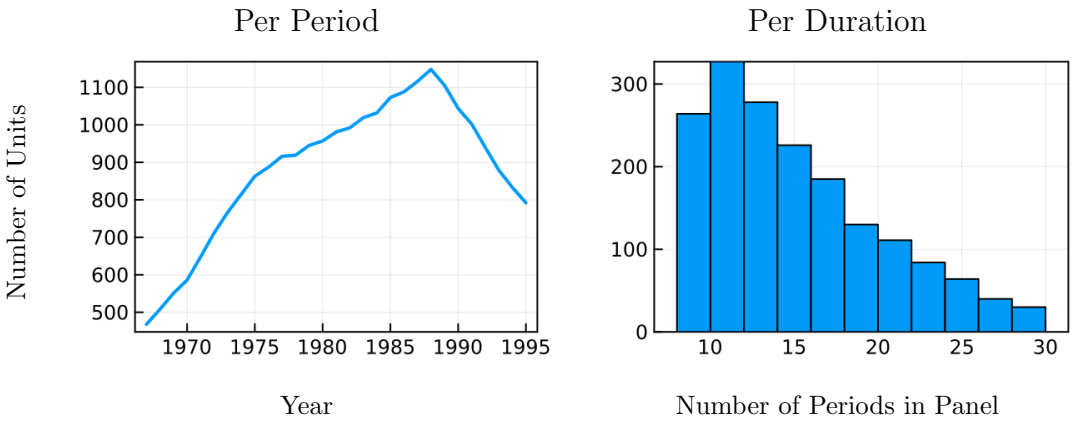
	$\tau = 1988$		$\tau = 1991$	
Para	Median	HPD	Median	HPD
μ_{s_0}	0.00	[-0.06, 0.06]	-0.01	[-0.06, 0.04]
v_{s_0}	0.03	[0.01, 0.06]	0.02	[0.01, 0.04]

Notes: We report 90% posterior HPD intervals.

from v_{s_0} to the unconditional variance of the s_t process. This is reflected in the cross-sectional dispersion of the posterior mean estimates $\hat{s}_{it}$, when comparing period $t = 1$ to $t = T = 20$: the estimates range from -0.2 to 0.2 at the beginning of the sample, and from -1 to 1 at the end of the sample.

Further Results for the Unbalanced Panel. The unbalanced sample is similar to the full sample described above, except that we require that individuals have at least nine *consecutive* observations. Figure A-8 provides some information about the composition of $\mathcal{U}$.

Figure A-8: Number of Units in Unbalanced Panel



Parameter estimates from the baseline model and the HIP and RIP versions are reported in Table A-4. Consistent with Figure 6, the estimates of q^α and q^ρ are very close to one, implying full heterogeneity. The posterior median estimates of α_0 and the return to experience $\alpha_1/10$ are, at -0.21 and .017 respectively, very close to the estimates for the balanced samples $\mathcal{B}_{20,1988}$ and $\mathcal{B}_{20,1991}$ reported in Table 5. Due to the larger sample size, the HPD intervals for those parameters are much narrower for the unbalanced sample.

Table A-4: Posterior Distribution of θ, Unbalanced Panel

	S&S		HIP		RIP	
Para	Median	HPD	Median	HPD	Median	HPD
α_0	-0.21	[-0.23, -0.20]	-0.21	[-0.23, -0.20]	-0.37	[-0.39, -0.35]
$\alpha_1/10$	.017	[.016, .018]	.017	[.016, .018]	.020	[.019, .021]
ρ	0.63	[0.60, 0.65]	0.63	[0.61, 0.65]	0.97	[0.97, 0.98]
μ_{s_0}	-0.23	[-0.26, -0.20]	-0.23	[-0.26, -0.20]	-0.02	[-0.04, -0.00]
v_{s_0}	0.26	[0.23, 0.29]	0.27	[0.24, 0.30]	0.14	[0.13, 0.15]
q^α	0.99	[0.97, 1.00]	-	-	-	-
q^ρ	0.89	[0.77, 1.00]	-	-	-	-
q^{σ_u}	1.00	[0.99, 1.00]	1.00	[0.99, 1.00]	1.00	[0.99, 1.00]
q^{σ_ϵ}	1.00	[0.99, 1.00]	1.00	[0.99, 1.00]	1.00	[0.99, 1.00]
$v_{\delta^{\alpha_0}}$	0.09	[0.08, 0.10]	0.08	[0.07, 0.10]	-	-
$v_{\delta^{\alpha_1}}/100$	.0002	[.0002, .0002]	.0002	[.0002, .0002]	-	-
v_{δ^ρ}	0.05	[0.04, 0.05]	0.04	[0.04, 0.05]	-	-
$v_{\delta^{\sigma_u}}$	42.56	[17.75, 78.24]	43.03	[18.72, 79.81]	45.37	[19.04, 82.14]
$v_{\delta^{\sigma_\epsilon}}$	35.85	[15.23, 67.73]	35.56	[16.31, 68.49]	30.07	[12.90, 53.81]
LPS Diff.		-		-.0003		.0072
MSE Diff. (%)		-		0.18		-1.44

Notes: For the parameters we report 90% posterior HPD intervals. α_1 and $v_{\delta^{\alpha_1}}$ are rescaled to undo the division of experience by 10 in the definition of x_{it}, see (26). LPS and MSE differentials for the out-of-sample predictions are computed with respect to the baseline S&S model.

An important difference between the estimates from the $\mathcal{U}$ and $\mathcal{B}_{T,\tau}$ samples manifests itself in regard to the autocorrelation of the persistent income component ρ and the prevalence of return-to-experience heterogeneity captured by q^α. For the unbalanced sample $\hat{\rho} = 0.63$, which is considerably lower than the estimates obtained for the balanced samples and the numbers generally reported in the literature. In regard to q^α, recall that the estimates for the two balanced panels reported in Table 5 were 0.34 and 0.81. In the unbalanced panel almost all individuals deviate from the core group: $\hat{q}^\alpha = 0.99$. The variance of the return-to-experience discrepancies, on the other hand, is of the same magnitude as in the balanced panels: $\hat{v}_{\delta^{\alpha_1}} = .0002$, compared with .0002 and .0001. Qualitatively, the low estimate of ρ is consistent with the more prevalent return-to-experience heterogeneity, as the two have been found to be negatively related in the existing literature.

The low estimate of ρ and the high estimate of q^α imply that return-to-experience heterogeneity and transitory shocks play a more important role for income inequality within cohorts. For the $\mathcal{B}_{20,1988}$ sample in Figure 5 we found that with five years of experience, α

heterogeneity generates about 8% of the income inequality and the transitory income shock about 4%. With eight years of experience the contribution is essentially zero. Using the estimates from the unbalanced panel instead, α heterogeneity and transitory shocks explain 40% and 10% of the income inequality within the cohort after five years of work experience and 25% and 5% after ten years.

Comparing the $v_{\delta^{\sigma_u}}$ and $v_{\delta^{\sigma_\epsilon}}$ estimates from the unbalanced panel to those from the balanced panels in Table 5, one also notices a substantial increase in the shock-variance heterogeneity. For instance, for $\mathcal{B}_{20,1988}$ the posterior median estimates are $v_{\delta^{\sigma_u}} = 6.97$ and $v_{\delta^{\sigma_\epsilon}} = 4.58$, whereas they are 42.56 and 35.85 for the unbalanced panel. To interpret the magnitude of $\delta_i^{\sigma_\epsilon}$ (and its variance), recall that it has to be multiplied by $\sigma^2_{\epsilon,t}$, which we plotted in Figure A-7 for the balanced panels. The values for the unbalanced panel are slightly smaller. Overall, the cross-sectional averages of $\sigma^2_{\epsilon,t}\delta_i^{\sigma_\epsilon}$ are quite similar across samples.

The remaining columns of Table A-4 contain estimates for HIP and RIP. The HIP estimates, generated conditional on $q^\alpha = q^\rho = 1$, are essentially equal to the baseline estimates, which feature $\hat{q}^\alpha = 0.99$ and $\hat{q}^\rho = 0.89$. The RIP estimates feature a substantially larger $\hat{\rho} = 0.97$. Moreover, the estimated return to experience is about 0.3 percentage points larger than for S&S and HIP. In the bottom rows of Table A-4 we report results for the one-step-ahead out-of-sample predictive performance, where the last observation of each unit is held out. As previously in Figure 7, we compute LPS differentials and relative MSE differentials (in percent) with respect to the baseline S&S model. For units whose last observation falls before 1996, the predictive density uses the posterior draws of the variances of that year, which are estimated from the other units observed in that year; for units whose last observation is in 1996 we use the 1995 values. Negative values indicate an improvement. The differentials between S&S and HIP are essentially zero. The RIP specification with the more persistent income process improves the point forecasts slightly: its MSE is about 1.4% lower than that of S&S. Its density forecasts, however, are less accurate. Summing the differences in log predictive scores over the 1,739 units, we obtain log odds of about 12.5 in favor of S&S. Thus, the parsimony of the RIP specification improves the point forecasts on average across units, but not the predictive densities.