

Small Sample Properties of Bayesian Estimators of Labor Income Processes*

Taisuke Nakata[†]
Federal Reserve Board

Christopher Tonetti[‡]
Stanford GSB

March 31, 2014

Abstract

There exists an extensive literature estimating idiosyncratic labor income processes. While a wide variety of models are estimated, GMM estimators are almost always used. We examine the validity of using likelihood based estimation in this context by comparing the small sample properties of a Bayesian estimator to those of GMM. Our baseline studies estimators of a commonly used simple earnings process. We extend our analysis to more complex environments, allowing for real world phenomena such as time varying and heterogeneous parameters, missing data, unbalanced panels, and non-normal errors. The Bayesian estimators are demonstrated to have favorable bias and efficiency properties.

JEL: C32, C33, D12, D31, D91, E21

Keywords: Labor Income Process, Small Sample Properties, GMM, Bayesian Estimation, Error Component Models

*We thank Tim Cogley, Konrad Menzel, John Roberts, Tom Sargent, Matt Smith, and Gianluca Violante for useful comments and suggestions. The views expressed in this paper, and all errors and omissions, should be regarded as those solely of the authors, and are not necessarily those of the Federal Reserve Board of Governors or the Federal Reserve System.

[†]Corresponding author; Division of Research and Statistics, Federal Reserve Board, 20th Street and Constitution Avenue N.W. Washington, D.C. 20551; Email: taisuke.nakata@frb.gov; Phone: 202-452-2972.

[‡]Stanford GSB, 655 Knight Way, Stanford, CA 94305; Email: tonetti@stanford.edu.

1 Introduction

Measuring individual income risk is essential to answering a wide range of economic questions. For the vast majority of individuals, labor income is overwhelmingly the largest component of total income. Accordingly, accurate and precise estimation of labor earnings dynamics is important for analyzing people’s consumption and savings behavior, designing fiscal policy, and understanding the sources of and changes in inequality.

Dating back to Lillard and Willis (1978), Lillard and Weiss (1979), MaCurdy (1982), and Abowd and Card (1989), there is a history of fitting ARMA models to panel data to understand the labor income risk facing individuals. Error component models in which labor income is the sum of a transitory and a persistent shock are extremely common.¹ While no shortage of models have been estimated, a striking feature of the literature is that, almost always, the estimation routine is based on a minimum distance estimator, namely GMM.

Recent trends of increasing model complexity have led economists to depart from the simple GMM estimators that have dominated the literature. Sometimes these complex models are estimated by techniques similar to traditional GMM, like the Simulated Method of Moments.² However, sometimes a very different likelihood based approach, Bayesian estimation, is employed. An early example using Bayesian estimation is Geweke and Keane (2000), who jointly estimate earnings process parameters and marital status to analyze the transition probabilities between income quartiles over the life cycle. More recently, Jensen and Shore (2010, 2011) estimate complex labor income processes with a focus on heterogeneity in idiosyncratic risk and its evolution over time. Given ongoing advances in computational power and Markov chain Monte Carlo (MCMC) techniques, we are likely to see a growing use of these estimators.

While there exist decades of research that documents the properties of minimum distance estimators of labor income processes, there are no papers, to our knowledge, that systematically examine the small sample properties of likelihood based estimators on these types of error component models in panel settings.³ Although the benefits of Bayesian estimation are well known to theoretical econometricians, they are not widely understood across fields in the context of estimating labor income processes. Thus, as the profession adopts widespread use of Bayesian estimators for increasingly elaborate models, it is useful to document the properties of Bayesian estimators on simple income processes. This will provide a clear explanation of the Bayesian estimation routine, a better understanding of the source of estimation results, and a justification for its use in more complicated settings.

In this paper, we examine the validity of using likelihood based estimation by comparing the

¹Often the persistent component is assumed to follow a random walk as in MaCurdy (1982), Abowd and Card (1989), Gottschalk and Moffitt (1994), Meghir and Pistaferri (2004), and Blundell, Pistaferri, and Preston (2008).

²For example, Browning, Ejrnaes, and Alvarez (2010) estimate a model with many dimensions of heterogeneity using the Simulated Method of Moments.

³For example, a 1996 issue of the *Journal of Business & Economic Statistics* is dedicated to the small sample properties of GMM estimators, featuring notable papers on estimating covariance structures like Altonji and Segal (1996) and Clark (1996).

small sample properties of a Bayesian estimator to those of GMM. First, as a stepping stone, we derive the maximum likelihood estimator by translating the labor income process into linear Gaussian state space (LGSS) form and applying standard filtering procedures. With this machinery established in an easy-to-understand framework, we provide a concise, but self-contained, derivation of the Bayesian estimator. We then conduct a Monte Carlo analysis of a Bayesian estimator and a GMM estimator on a commonly used simple labor income process. Although the asymptotic dominance of properly specified likelihood based estimators is textbook, we provide the first systematic analysis of the small sample properties of likelihood based estimators of labor income processes. While the difference between estimators is typically modest, the Bayesian estimator is more efficient: parameter estimates are unbiased with smaller standard errors.

Although the initial findings demonstrate the better small sample properties of the Bayesian estimator, the exercise was performed on a simple model with ideal dataset conditions. In Section 4, we extend the analysis to more complex specifications that capture real world phenomena. In Section 4.1, we first look at an environment in which datasets are unbalanced panels and suffer from missing data. We modify the estimators to accommodate these realistic features of the dataset and study how the relative performance of the Bayesian estimator is affected. In Section 4.2, we modify the labor income process to allow for time-variation in the shock variances. Time variation in the variances of the labor income process has been documented in many papers and is essential in understanding the changes in labor market risks.⁴ In Section 4.3, we allow parameter heterogeneity across individuals. Many papers have found substantial heterogeneity in various aspects of the labor income process and emphasized its importance for understanding the risks people face. Finally, in Section 4.4, we extend the labor income process to allow for non-normal shocks. There is empirical evidence that labor income shocks are non-normal and have fat tails.⁵ To both learn about the shape of the error distributions and to address the concern of distribution misspecification, we let shocks be distributed according to a mixture of normals, which allows for a very flexible error distribution structure. These extensions establish that the beneficial small sample properties of Bayesian estimators are maintained in more complicated scenarios—thus demonstrating the usefulness of this estimator for applied economists.

We proceed in Section 2 to present the different estimators of the simple income process. Section 3 discusses their small sample properties and Section 4 discusses extensions to the simple income process, including time varying variances, missing data, unbalanced panels, heterogeneous parameters, and distributional assumptions on errors. Section 5 concludes.

⁴See for example Gottschalk and Moffitt (1994), Heathcote, Perri, and Violante (2010), and Blundell, Pistaferri, and Preston (2008) for GMM based estimations of the sequence of variances.

⁵See for example Hirano (1998) and Geweke and Keane (2000) for Bayesian estimations of income processes that allow for non-normal shocks.

2 Estimators of the Labor Income Process

It is well known from the labor economics literature (since Abowd and Card (1989)) that labor income is well described by an error components model, where residual labor earnings are the sum of a transitory and persistent shock. Often the transitory shock is i.i.d and the persistent shock follows an AR(1) process. Because the model fits well and is relatively simple, it has become very commonly used in the Labor and Macro literatures. Thus, we adopt this pervasive simple income process and present the GMM and Bayesian estimators. These estimators are developed to be used on panel data sets such as the PSID. Accordingly, our baseline labor income process is given by:

$$\epsilon_{i,t} = \rho\epsilon_{i,t-1} + \Sigma_{\eta}^{0.5}\eta_{i,t} \quad (1)$$

$$y_{i,t} = \epsilon_{i,t} + \Sigma_{\nu}^{0.5}\nu_{i,t} \quad (2)$$

where $y_{i,t}$ is the residual from a log income regression for an individual $i \in \{1, \dots, N\}$ at time $t \in \{1, \dots, T\}$.⁶ $\epsilon_{i,t}$ is the persistent component of income and is assumed to follow an AR(1) process. While $\{y_{i,t}\}_{i=1,N}^{t=1,T}$ is observed, $\{\epsilon_{i,t}\}_{i=1,N}^{t=1,T}$ is not. $\nu_{i,t}$ is the transitory component of income. $\eta_{i,t}$ is the shock to the persistent component of income. $\nu_{i,t}$ and $\eta_{i,t}$ are standard normal and Σ_{ν} , Σ_{η} are the variances of the transitory and persistent innovations, respectively. We also estimate properties of the initial state, $\epsilon_{i,0}$, which we assume is normally distributed with zero mean and variance Σ_{Z0} . $\nu_{i,t}$, $\eta_{i,t}$, and $\epsilon_{i,0}$ are independent of each other for all i and t . Thus the parameter vector to be estimated is $\theta := (\rho, \Sigma_{\nu}, \Sigma_{\eta}, \Sigma_{Z0})$. In the next subsections, we outline different techniques to estimate these parameters.

2.1 GMM

The standard strategy in the literature for estimating labor income processes is to use GMM. The goal of GMM estimation is to choose parameters that minimize the distance between empirical and theoretical moments. Identification and estimation rely on moments constructed from cross-sectional income autocovariances.⁷

We wish to estimate the system given by equations (1) and (2). There are $\frac{T(T-1)}{2}$ moments, where T is the total number of time periods. The total number of individuals, N , does not affect

⁶There are many ways to obtain residual labor income. To remove predictable components that are associated with the individual or the aggregate state of the economy, it is common to regress idiosyncratic labor income on a vector of observables such as individual demographics and variables that control for aggregates. There are different methods to remove aggregate components, such as using time dummy variables or a common correlated effect approach (Pesaran (2006)). While important for the macro implications of the idiosyncratic labor earnings process, our paper is concerned only with estimating the residual labor income process.

⁷This section develops the already well-understood GMM estimation routine. For other sources see Blundell, Pistaferri, and Preston (2008), Heathcote, Perri, and Violante (2010), or Guvenen (2009). We present the general case in which ρ is estimated.

the number of moments because the expectations are taken cross-sectionally over individuals.

$$M(\theta) = \text{vec} \begin{pmatrix} \mathbb{E}[y_{i,1} \cdot y_{i,1}] & \dots & \dots & \dots \\ \mathbb{E}[y_{i,2} \cdot y_{i,1}] & \mathbb{E}[y_{i,2} \cdot y_{i,2}] & \dots & \dots \\ \mathbb{E}[y_{i,3} \cdot y_{i,1}] & \mathbb{E}[y_{i,3} \cdot y_{i,2}] & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \mathbb{E}[y_{i,T} \cdot y_{i,1}] & \mathbb{E}[y_{i,T} \cdot y_{i,2}] & \dots & \mathbb{E}[y_{i,T} \cdot y_{i,T}] \end{pmatrix}$$

where the upper right triangle of the matrix is redundant by symmetry. Particular entries in the moment matrix $M(\theta)$ map into particular functions of the parameters $\theta = (\rho, \Sigma_\nu, \Sigma_\eta, \Sigma_{Z0})$. Define the cross-sectional moment $m_{t,j}(\theta)$ between agents at time t and $t + j$:

$$\begin{aligned} m_{t,j}(\theta) &= \mathbb{E}[y_{i,t} \cdot y_{i,t+j}] \\ &= \mathbb{E}[(\epsilon_{i,t} + \Sigma_\nu^{0.5} \nu_{i,t})(\epsilon_{i,t+j} + \Sigma_\nu^{0.5} \nu_{i,t+j})] \\ &= \begin{cases} \mathbb{E}[\epsilon_{i,t}^2] + \Sigma_\nu & \text{if } j = 0 \\ \rho^j \mathbb{E}[\epsilon_{i,t}^2] & \text{if } j > 0 \end{cases} \end{aligned}$$

where

$$\mathbb{E}[\epsilon_{i,t}^2] = \rho^{2t} \Sigma_{Z0} + \sum_{k=1}^t \rho^{2(t-k)} \Sigma_\eta$$

Simple algebraic manipulation of the above equations reveal (over) identification of the parameters.⁸

Define the estimator as

$$\hat{\theta}_{GMM} = \min_{\theta} (M(\theta) - \hat{M})' W (M(\theta) - \hat{M})$$

where \hat{M} stacks the sample covariance from the data.

To implement the estimator, we need to choose the weighting matrix, W . Altonji and Segal (1996) show that the Optimal Minimum Distance (OMD) estimator, where W is the optimal weighting matrix, introduces significant small sample bias. They study the small sample properties of the GMM estimator with several alternative weighting matrices and recommend using the Equally

⁸The moments are only slightly modified when variances are time varying as in Section 4.2:

$$\begin{aligned} \mathbb{E}[y_{i,t} \cdot y_{i,t+j}] &= \begin{cases} \mathbb{E}[\epsilon_{i,t}^2] + \Sigma_{\nu,t} & \text{if } j = 0 \\ \rho^j \mathbb{E}[\epsilon_{i,t}^2] & \text{if } j > 0 \end{cases} \\ \mathbb{E}[\epsilon_{i,t}^2] &= \rho^{2t} \Sigma_{Z0} + \sum_{k=1}^t \rho^{2(t-k)} \Sigma_{\eta,k} \end{aligned}$$

Weighted Minimum Distance (EWMD) estimator, where W is the identity matrix. In light of their result, many papers in the literature use the EWMD estimator.⁹ Thus, we use the GMM estimator with identity weighting matrix throughout the paper. Extensions to different weighting matrices are straightforward.¹⁰

With all objects of the optimization problem defined, the estimates can be calculated computationally using an optimizer to choose parameters θ that minimize the distance between model and sample moments.

2.2 Likelihood Based Estimation

The maximum likelihood estimator is presented before deriving the Bayesian estimator. This establishes the linear Gaussian state space structure, which is used in both estimators. For the simple income process, it is feasible to compute the maximum likelihood estimator. However, the MLE estimator fails to scale with model complexity, as for more complicated specifications the likelihood is difficult or impossible to write in closed form. Bayesian estimation becomes useful in exactly these cases.

2.2.1 MLE

As its name suggests, the maximum likelihood estimate is defined as the parameter values that maximize the likelihood function.

$$\hat{\theta}_{MLE} = \max_{\theta} L(\{y_{i,t}\}_{i=1,N}^{t=1,T} | \theta)$$

For the simple labor income process, we can analytically compute the likelihood for any given θ , and thus it is feasible to solve this maximization problem. Analytical derivation of the likelihood function is straightforward once we notice that the labor income process forms a linear Gaussian state space (LGSS) system. We can draw from the time series econometrics literature to construct the likelihood using the Kalman filter.

Define the canonical form of a LGSS as

$$S_t = A_t S_{t-1} + B_t v_t \tag{3}$$

$$Y_t = C_t z_t + D_t S_t + w_t \tag{4}$$

where

⁹An alternative used by Blundell, Pistaferri, and Preston (2008) is the Diagonally Weighted Minimum Distance (DWMD) estimator, where W is the optimal weighting matrix with off-diagonal elements set to zero.

¹⁰We examined the performance of the OMD and DWMD estimators in unreported exercises and found that the EWMD estimator tends to outperform the OMD and DWMD estimators across various data generating processes. The OMD and DWMD estimators are sometimes slightly more efficient for the persistence parameter than the EWMD estimator, but such improved efficiency in the estimate of the persistence parameter always comes with substantially deteriorated efficiency in the estimates of other parameters.

$$\begin{aligned}
v_t &\sim N(0, Q_t) \\
w_t &\sim N(0, R_t) \\
S_0 &\sim N(S_{0|0}, P_{0|0})
\end{aligned}$$

Equation (3) is the state equation (or transition equation) and equation (4) is the observation equation (or measurement equation). S_t is a vector of latent state variables, and Y_t is the vector of observables. v_t is a vector of innovations to latent state variables, and w_t is a vector of measurement error. z_t is a vector of observed exogenous variables. To map this model into our labor income process of an individual Y_t is residual log income and S_t is the persistent component of income.¹¹ For notational convenience, let θ_t be a vector containing all parameters of the model ($A_t, B_t, C_t, D_t, Q_t, R_t, S_{0|0}, P_{0|0}$). In many applications, parameters are time-invariant ($\theta_t = \theta$ for all t), but this is not a necessary requirement. All processes discussed in this paper will hold A_t, B_t, C_t, D_t constant over time; later, when we discuss the time varying volatility case, Q_t and R_t will vary. Thus, our simple labor income process can be mapped into canonical form as

$$\begin{aligned}
A_t &= \rho, & B_t &= 1, & C_t &= 0, \\
D_t &= 1, & Q_t &= \Sigma_\eta, & R_t &= \Sigma_\nu, \\
z_t &= 0, & S_{0|0} &= 0, & P_{0|0} &= \Sigma_{Z0}
\end{aligned}$$

for all t, with parameter vector $\theta = (\rho, \Sigma_\nu, \Sigma_\eta, \Sigma_{Z0})$ as before.

We can now use the Kalman filter to derive the log likelihood of a LGSS, which is provided in Appendix A.1.¹² As idiosyncratic shocks are defined to be uncorrelated across individuals, one can obtain the log likelihood of the income data for all individuals by summing the log likelihood of each individual (i.e., $\log L(\{y_{i,t}\}_{i=1,N}^{t=1,T} | \theta) = \sum_{i=1}^N \log L(\{y_{i,t}\}^{t=1,T} | \theta)$). One obvious advantage of this methodology is the relative ease of estimating more complicated income processes, such as ARMA(p,q) processes, since they allow a LGSS representation. Given that the likelihood is defined in terms of canonical LGSS matrices, to estimate a different parameterization of the labor income process, all that remains is to map the process into state space form. With the analytical likelihood derived in Appendix A.1, we can now use an optimizer to numerically maximize the log likelihood to produce the MLE estimates.

¹¹For clarity of exposition, we set z_t to be zero and assume Y_t is residual income. However, z_t could include all of the traditional first stage conditioning variables like education, age, race, etc., to remove the predictable components of Y_t . This would allow estimation to be run in one step, instead of running a first-stage regression to recover idiosyncratic residual income. The general case is presented in the Appendix.

¹²For a detailed analysis, see chapter 13.4 in Hamilton (1994).

2.2.2 Bayesian Estimation: Gibbs Sampling

In computing the maximum likelihood estimator, two conditions are crucial. One is that we can compute the likelihood function analytically and the other is that the likelihood function is well behaved.¹³ For the simple labor income process, these two conditions are satisfied and the MLE estimates can be easily computed. However, they are not satisfied for more complex labor income processes. Bayesian methods are useful in those cases either because they don't require an analytical likelihood function or because the prior distribution smooths local maxima. For transparency of exposition, in this section we provide an overview of the Bayesian estimator for the standard income process, without the complications that it is well equipped to handle.

The goal of the Bayesian estimation is to characterize the posterior distribution of parameters, which is defined as the distribution of parameters conditional on data, $p\left(\theta|\{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$. The posterior distribution is related to two objects, the prior distribution and the likelihood function, by Bayes Theorem:

$$p\left(\theta|\{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \propto L(\{y_{i,t}\}_{i=1,N}^{t=1,T}|\theta)p(\theta)$$

where $p(\theta)$ is the prior distribution of parameters, which the researcher specifies. When the prior is uniform, the posterior is the same as the likelihood (as a function of parameters). The tighter the prior, the less the posterior reflects information from the data. Comparing the posterior to the prior provides a sense of how much the data informed the parameter estimates.¹⁴

To characterize the posterior distribution, the Bayesian estimator draws a large sample from $p\left(\theta|\{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$, which can be used to compute statistics about the parameters of interest. Once we characterize the posterior distribution, we can compute a point estimate according to some loss function. Following other applied work in the Bayesian literature, we use the median of the posterior distribution as the point estimate.¹⁵

To obtain this sample from the posterior, we use the Gibbs sampling algorithm. The Gibbs sampling algorithm sequentially draws from the distribution of each parameter and latent variable conditional on the previous draws for all other parameters. Since each draw is conditional on the previous draws, the samples are not independent. However, the stationary distribution of the draws thus generated can be shown to equal the joint posterior distribution of interest.¹⁶ This algorithm is useful whenever the joint density of the parameters and data is unknown, but the conditional distributions of each parameter is known. See Appendix A.2 for details about the Gibbs sampling procedure.

To sequentially draw from the conditional distributions of interest, there are three techniques

¹³If there exist multiple local maxima, the optimization routine may struggle to find the global maximum, i.e., the MLE estimates. This is especially true if the likelihood function is high dimensional.

¹⁴See Geweke (2005) or Gelman, Carlin, Stern, and Rubin (2009) for a Bayesian statistics reference.

¹⁵There is no a priori "correct" statistic to use as the point estimate. For example, one could use the mode, mean, or median of the posterior distribution. While the mode corresponds closely to the principle of maximum likelihood, the median is more robust computationally when sample sizes are limited.

¹⁶See Robert and Casella (2004) for a comprehensive treatment of MCMC theory and practice.

employed. The first is estimating the variance of a normal distribution with known mean. The second is estimating a linear regression model. The third is sampling the latent states of a LGSS model. To estimate Σ_ν , realize $y_{i,t} - \epsilon_{i,t} = \Sigma_\nu^{0.5} \nu_{i,t}$. Thus, conditioning on the epsilon sequence, we have a sample distributed $N(0, \Sigma_\nu)$. Given a conjugate prior, the posterior can be calculated analytically, allowing this iteration's Σ_ν to be set equal to a draw from the conditional posterior. To estimate ρ , realize from equation 1 that ρ is the coefficient in a linear regression model with known independent and dependent regressors with error distributed $N(0, \Sigma_\eta)$. Again, the conditional posterior can be calculated analytically, allowing this iteration's ρ to be set equal to a draw from the posterior.

The most complicated Gibbs block of this estimation is drawing the sequence of unobserved states in a LGSS. Given linearity and normality, the Kalman smoother provides the distribution of each state at each time conditional on all available data. However, because there is persistence in the states, knowing the particular draw of $\epsilon_{i,t}$ alters the conditional density of $\epsilon_{i,t-1}$.¹⁷ Carter and Kohn (1994) develop an algorithm that recursively updates, backwards through time, the Kalman smoothed conditional densities of each state at time t given the draw of the state at time $t + 1$.¹⁸ With these three tools, we are able to sequentially draw from the conditional distributions of the parameters and latent variables of interest.

3 Small Sample Properties

Now that the different estimators have been presented, we proceed to analyze the small sample properties of the estimators using Monte Carlo analysis. Our Monte Carlo simulation exercise is designed as follows. Simulate 100 panel datasets using the baseline labor income processes given by equations 1 and 2.¹⁹ In the benchmark simulation, each data set contains 500 people and 10 time periods (i.e., $N=500$ and $T=10$). However, to mimic the features of available panel datasets while also providing a more general depiction of the estimators' performance, we consider alternative values of $N \in \{100, 2000\}$ and $T \in \{5, 20\}$. Motivated by Blundell, Pistaferri, and Preston (2008), as a benchmark calibration we let the true parameter values be $(\rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}) = (1.00, 0.02, 0.05, 0.15)$.²⁰ For each of the 100 datasets generated, we compute the point estimates of the model parameters based on the Bayesian method and the GMM. When required by optimization or sampling, the same initial values are used across estimators.

For the Bayesian estimation, we need to specify prior distributions of parameters. We choose

¹⁷We could sample each $\epsilon_{i,t}$ as a separate block. This is the approach taken by Jacquier, Polson, and Rossi (1994) when estimating a stochastic volatility model. However, this would lead to a highly auto-correlated, slowly converging Markov chain.

¹⁸For some extensions, it is easy to replace the Carter and Kohn (1994) algorithm with the more computationally efficient algorithm developed by Durbin and Koopman (2002). See the online technical appendix for detailed description of the Durbin and Koopman (2002) algorithm.

¹⁹Monte Carlo analysis is often based on a large number of datasets. However, due to the computational intensity of the Bayesian estimator, we present results with a smaller than usual number of simulations. A robustness check on the baseline specification with a larger number of datasets (500) replicated the results of the 100 sample exercise.

²⁰These parameter values are standard and similar to the estimates found throughout the traditional literature.

prior parameters such that the variance of the prior distribution is very large and thus the posterior distribution is mainly determined by the likelihood function. Specifically, we set the prior for ρ to be Normal with mean 0 and an arbitrarily large variance, truncated at support $[-1,1]$. For Σ_{Z0} , Σ_η , and Σ_ν we use the inverse-gamma distribution with degree of freedom 2 and location parameter equal 0.01.

For each estimator, we then compute the mean, standard deviation, and the root mean square error (RMSE) of the point estimates across the 100 datasets. The difference between the mean and true parameter values measures the bias of the estimator, while the standard deviation measures the efficiency. The root mean square error measures the average squared deviation of the estimator from the true parameter value, combining information on the bias and standard deviation of the estimator. Accordingly, we emphasize root mean square error as the overall measure of the estimators' efficiency.

Table 1 shows the small sample properties of the two estimators. For each parameter and estimator, the first number is the mean of the point estimates across 100 simulated datasets, the second number in the parentheses is the standard deviation of the point estimates, and the last number in the square brackets is the RMSE.

The Bayesian estimator performs better than the GMM estimator in terms of RMSEs. For all parameter estimates, the RMSEs of the Bayesian estimator are smaller than those of the GMM estimator. While the differences tend to be small as both estimators have very low RMSEs, they are sometimes non-trivial. For example, the RMSEs of the Bayesian estimator are about 30 percent and 50 percent smaller than the GMM estimator for Σ_η and Σ_ν . The outperformance of the Bayesian estimator comes from both smaller bias and smaller standard deviation. For all parameters, the mean of the Bayesian estimator is closer to the true parameter values than that of the GMM estimator and the standard deviation of the Bayesian estimator is smaller.

Improved efficiency could result from using better moment conditions or from using informative priors. To demonstrate that the improved performance of the Bayesian estimator is not driven by the priors, we first re-estimate the model, setting the prior mean of ρ to 0.5 and scaling up the variances' prior means by a factor of 10. This very different prior specification did not significantly affect the posterior distribution. This means that the posterior distribution is mainly determined by the likelihood function. Second, we know that the GMM estimator with optimal orthogonality conditions is asymptotically and numerically equivalent to MLE, since the asymptotically optimal GMM estimator uses the score for moment conditions. For the baseline specification, we present the MLE estimates in Table 1, which have very similar RMSEs to the Bayesian estimates.²¹ Thus, efficiency improvements come from using a likelihood based estimator, not from imposing informative priors.

Outperformance of the Bayesian estimator is robust to different specifications of the data generating process. Table 2 reports the RMSEs for alternative values of N (=100,2000), T (=5,20),

²¹In comparing the MLE and Bayesian estimates, it should be noted that we use the median of the posterior as the Bayesian point estimate, where the mode of the posterior more closely corresponds to the MLE estimate. This difference becomes especially important when comparing truncated parameters, like ρ .

and ρ ($=0.8$). In most cases, the RMSEs of the Bayesian estimators are smaller or roughly equal for all parameter values. For ρ , the estimators give roughly equal RMSEs with alternative T and N . However, with $\rho = 0.8$, the Bayesian estimator delivers smaller RMSEs. For Σ_{Z0} , one estimator does not consistently perform better than the other across different cases. For Σ_η and Σ_ν , the RMSEs of the Bayesian estimator are consistently smaller, sometimes by a large amount.

3.1 Robustness to Model Misspecification

While the Bayesian estimator performs well in small samples on a properly specified model, how do its small sample properties compare to the GMM estimator when the model is misspecified? We address this question with two exercises. First, we estimate the model given by equations (1) and (2) assuming that $\rho = 1$, when the true data generating process has $\rho = 0.95$. As displayed in Table 3, the Bayesian estimator has significantly lower RMSEs for both Σ_η (0.005 vs. 0.014) and Σ_ν (0.004 vs. 0.023) and a modestly lower RMSE for Σ_{Z0} (0.031 vs. 0.033).

Second, we estimate the same model assuming errors are normally distributed, when they are in fact distributed according to a mixture of normals, using the data generating process detailed in Section 4.4. As shown in Table 4, the Bayesian and GMM estimators perform similarly, with the Bayesian estimator resulting in moderately smaller RMSEs on Σ_η (0.0021 vs. 0.0026), Σ_ν (0.0026 vs. 0.0042), and Σ_{Z0} (0.0127 vs. 0.0135), and an identical RMSE for ρ (0.0048).

Thus, even though the Bayesian estimator heavily relies on parametric assumptions, it can still have as good or better small sample properties when the model is misspecified.

4 Robustness to Extended Labor Income Processes

The previous section showed that the Bayesian estimator outperforms the GMM estimator in finite samples on a simple labor income process. However, when using real data, researchers often need to modify the simple process to address unique features of the dataset or to incorporate additional features into the model. Since some of these modifications are quite common or even necessary, it would be useful to know how robust are the findings based on the simple labor income process. Accordingly, this section considers four such modifications to the baseline earnings model and documents the performance of the Bayesian estimator in each setting. The modifications we consider are (1) unbalanced panel with missing data, (2) time-varying volatilities, (3) heterogeneous parameters, (4) non-normal shocks. Adjusting for an unbalanced panel with missing data is necessary whenever one estimates the model using real data. Time-varying volatilities are particularly common in the literature, regardless of the estimation methods used. Recent literature has documented evidence suggesting substantial heterogeneity in income process parameters across individuals. While non-normality is often ignored in GMM estimation, previous studies using Bayesian methods often estimated the income process allowing for non-normality and found significant improvements in model fit.

Of course, there are many other interesting extensions to the labor income process. These

include a richer modeling of volatility evolution (such as GARCH or stochastic volatility), time or age variation in other parameters (such as the autoregressive coefficient), and a richer decomposition of labor income (such as ARMA(p,q)).²² There has also been a growing interest in allowing for different dimensions of heterogeneity, in income profiles, initial levels, autoregressive coefficients, variances of idiosyncratic shocks, and more.²³ Most of these extensions can be handled within the LGSS structure and can thus be estimated given the methods previously discussed. However, our purpose in this section is not to be comprehensive, but to investigate how some common complications of the income processes may affect the relative performance of the Bayesian estimator. Thus, we focus on the aforementioned four extensions.

4.1 Data Structure

Most panel datasets used for estimating labor income processes are unbalanced. That is, the dataset contains different cohorts and does not necessarily measure the initial labor market experiences of all cohorts. They are also known to contain many missing observations. This can be due to a variety of reasons, such as misreporting, attrition of the participants, and top-coding.

To examine how the Bayesian estimator performs on an unbalanced panel dataset with missing observations, we modify the baseline data generating process in two ways. First, we assign a 5 percent probability of being missing to each of the $T \times N$ observations. Second, we let a fraction (30 percent) of the population enter the dataset when they are age 4 and leave at age 10 to make the data unbalanced. As in the benchmark calibration of the previous simple labor income process, we set $N=500$ and $T=10$, and let the true parameter values be $(\rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}) = (1.00, 0.02, 0.05, 0.15)$. The Gibbs sampling algorithm needs to be modified to take missing data into account, but the estimation of a LGSS model with missing observations has been well studied and the modification is straightforward. The details on how to modify the Gibbs sampling procedure are given in Appendix A.3. Prior distributions are specified in the same way as in the baseline labor income process.

Table 5 compares the Bayesian estimator and the GMM estimator with this data structure. The Bayesian estimator continues to perform better than the GMM estimator. While the RMSEs for Σ_{Z0} are roughly equal across the two estimators (0.0147 for the Bayesian estimator and 0.0157 for the GMM estimator), the RMSEs associated with the Bayesian estimator are substantially smaller than those from the GMM estimator for ρ , Σ_η , and Σ_ν . As in the baseline exercise, the smaller RMSEs of the Bayesian estimator comes from its smaller bias and smaller standard deviation. For all parameters except ρ , the mean of the Bayesian estimator is closer to the true parameter values than that of the GMM estimator. For all parameters, the standard deviation of the Bayesian estimator is smaller.

The overall outperformance of the Bayesian estimator remains robust to different probabilities of missing observations and different fractions of the second cohort, as well as to alternative values

²²It is important to properly model measurement error, as documented by French (2004).

²³See Browning, Hansen, and Heckman (1999) and Jensen and Shore (2010) for a discussion.

of N , T , and ρ . These robustness results are available from the authors upon request.

4.2 Time Varying Variances

There is ample evidence that the variances of labor income shocks are time-varying. Many authors have documented a significant low frequency rise in the variances of income shocks over time. For example, Blundell, Pistaferri, and Preston (2008) and Gottschalk and Moffitt (1994) analyze the PSID and find increasing variances of shocks during the 1980s. Heathcote, Perri, and Violante (2010) conduct extensive analysis on wage inequality and show the continuous rise in income volatility over the past 40 years. Others have found that there is a fluctuation in shock volatility at the business cycle frequency. For example, using the PSID, Storesletten, Telmer, and Yaron (2004) found that the variance of income shocks is larger during recessions than during booms. Understanding the evolution of volatility in the labor income process is important in itself, as it allows us to understand the change in the risk environment facing workers. It is also important in analyzing many macroeconomic topics, such as the welfare cost of business cycles or the degree of self-insurance and market incompleteness.

While there are many different ways to model time variation in the variances, the majority of the literature models heteroskedasticity by treating the variance at each time as a different parameter.²⁴ We follow the literature and modify the baseline labor income process so that each variance parameter is specific to the time period. That is, our new data generating process is described by

$$\epsilon_{i,t} = \rho\epsilon_{i,t-1} + \Sigma_{\eta,t}^{0.5}\eta_{i,t} \quad (5)$$

$$y_{i,t} = \epsilon_{i,t} + \Sigma_{\nu,t}^{0.5}\nu_{i,t} \quad (6)$$

where $\Sigma_{\nu,t}$, $\Sigma_{\eta,t}$ are the variances of the transitory and persistent innovations at time t , respectively. We focus on the case with $T=10$ and calibrate the variances according to the estimates of Blundell, Pistaferri, and Preston (2008).²⁵ As before, we set $N=500$ and let the true parameter values for the rest of the parameters be $(\rho, \Sigma_{Z0}) = (1.00, 0.15)$.

The linear Gaussian state space structure is designed to allow parameters to vary over time. The only critical modification in the estimation routine is to feed the correct parameter values into the Kalman filter at the right times. To map our new process into a LGSS, we need to define A_t , B_t , C_t , D_t , Q_t , R_t for all t . We set $A_t = \rho$, $B_t = 1$, $C_t = 0$, $D_t = 1 \forall t$, keeping these matrices constant over time as in the benchmark simple labor income process. We set $Q_1 = \Sigma_{\eta,1}, \dots, Q_T = \Sigma_{\eta,T}$ and $R_1 = \Sigma_{\nu,1}, \dots, R_T = \Sigma_{\nu,T}$. Estimation then proceeds as in the stationary variances case. Prior distributions for ρ and Σ_{Z0} are specified in the same way as in the baseline labor income process.

²⁴As a notable alternative, Meghir and Pistaferri (2004) model the evolution of the variance as a GARCH process.

²⁵Since the timing convention in our model is slightly different from that of Blundell, Pistaferri, and Preston (2008), $\Sigma_{\eta,1}$ and $\Sigma_{\eta,10}$ are not separately identified from Σ_{Z0} and $\Sigma_{\nu,10}$, respectively. Thus, we set $\Sigma_{\eta,1}$ and $\Sigma_{\eta,10}$ to their true values in the estimation.

For the prior distributions of $\Sigma_{\eta,t}$ and $\Sigma_{\nu,t}$, we use the inverse-gamma distribution with degree of freedom 2 and location parameter equal to 0.01 for all t .

Table 6 compares the performance of the two estimators. As in the previous cases, the Bayesian estimator tends to deliver smaller RMSEs than the GMM estimator. For the time specific shock variances, the Bayesian estimator delivers substantially—about 30-50 percent—smaller RMSEs than the GMM estimator, except for $\Sigma_{\eta,2}$ and $\Sigma_{\nu,1}$ where the differences are small between two estimators. The outperformance of the Bayesian estimator comes from both smaller bias and standard deviation. For the $\Sigma_{\eta,2}$ and $\Sigma_{\nu,1}$ as well as ρ and Σ_{Z0} , the RMSEs of the two estimators are roughly the same. As before, these results are robust to alternative parameterizations of the data generating process, and the robustness results are available from the authors upon request.

For both estimators, the estimates of each variance parameter are not as precise as in the benchmark case. As an example, consider the RMSEs of Σ_{ν} in the benchmark DGP and $\Sigma_{\nu,10}$ in the DGP with time-varying variance. The true parameter values (0.05 and 0.0506) are roughly the same. However, for both the Bayesian and GMM estimators, the RMSEs of $\Sigma_{\nu,10}$ are about three times larger than those of Σ_{ν} . The reduction in the precision of the estimators is not surprising because we are estimating a larger number of parameters using the same number of observations. Nevertheless, both estimators remain sufficiently precise to closely track the rise and fall of the variances over time.

4.3 Heterogeneity

Understanding the heterogeneity in income process parameters is essential for answering questions with distributional aspects—across individuals and over the life cycle. Thus, many authors have estimated labor income processes allowing for heterogeneity in parameters across individuals. Lillard and Weiss (1979), Baker (1997), and Guvenen (2009) studied labor income processes in which the deterministic growth rate of income is heterogeneous across people. Jensen and Shore (2010, 2011) estimated models with heterogeneity in the variances of shocks to income processes. Browning, Ejrnaes, and Alvarez (2010) went further and estimated labor income processes where all parameters are allowed to be heterogeneous across individuals.

While many authors have stayed within the framework of the method-of-moment estimator to handle heterogeneity, some have employed Bayesian methods. To obtain some idea about the performance of the Bayesian estimator in the presence of heterogeneity, we modify our benchmark labor income process to introduce heterogeneity in one parameter—the deterministic growth rate of labor income. This modified process is often referred to as the labor income process with heterogeneous income profiles (HIP), and is specified as follows:

$$\epsilon_{i,t} = \rho\epsilon_{i,t-1} + \Sigma_{\eta}^{0.5}\eta_{i,t} \quad (7)$$

$$y_{i,t} = \beta i t + \epsilon_{i,t} + \Sigma_{\nu}^{0.5}\nu_{i,t} \quad (8)$$

where β_i is normally distributed with mean zero and variance Σ_β . Our benchmark labor income process is a special case of this process with $\Sigma_\beta = 0$. To estimate the parameters of this model, the Gibbs sampler needs to be modified to include two additional blocks—one for drawing Σ_β and the other for drawing $\{\beta_i\}_{i=1}^N$. Appendix A.4 shows the details. We calibrate Σ_β based on the estimate from Guvenen (2009). As before, we set $N=500$ and $T=10$, and calibrate the other parameters to be $(\rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}) = (1.00, 0.02, 0.05, 0.15)$. For the prior distribution of Σ_β , we use the inverse-gamma distribution with degree of freedom 2 and location parameter equal 0.01. Prior distributions of the other parameters are the same as in the baseline labor income process.

Table 7 shows the results of the Monte Carlo experiment. The Bayesian estimator performs favorably compared to the GMM estimator in the model with heterogeneous income profiles. While both estimates are somewhat larger (0.0006) than the actual value (0.0004) for Σ_β , the Bayesian estimator is more efficient and has a lower RMSE. Estimating this additional parameter does not alter the relative performance of the two estimators for other parameters. In particular, even though the Bayesian estimator’s RMSE for ρ becomes somewhat larger than the GMM counterpart, the Bayesian estimator continues to dominate the GMM estimator for all of the variance parameters.

4.4 Non-Normal Shocks

There is some evidence that shocks to the labor income process are not normally distributed. Hirano (1998) and Geweke and Keane (2002) both find that labor income shocks are fat-tailed. Capturing such non-normality is important in itself, but is also important for understanding earnings mobility and the persistence of poverty.

Even if one is not interested in the higher moments of the shock distribution, modeling non-normality is important as misspecification of the error distribution can reduce the efficiency of the estimator. The main efficiency gains in MLE arise because the estimator takes advantage of the parametric form of the shock distributions. The choice between GMM and likelihood-based estimation can then be viewed as a trade-off between smaller standard errors vs. misspecified distributions. To take advantage of increased efficiency, but to limit the possible damage from assuming the wrong distribution, the aforementioned authors allow for shocks to be distributed according to a mixture of normals. Mixtures of normal distributions are very flexible, allowing for a more robust estimation procedure. Models with few mixture components have been shown to perform well in small samples and can be viewed as a flexible parametric model that is an alternative to non-parametric density estimation without the associated curse of dimensionality.²⁶

In the following Monte Carlo analysis, we follow the previous papers and modify the baseline DGP so that the shocks are distributed according to a mixture of normal distributions. For simplicity of exposition, we use a mixture of two normal distributions. We also abstract from skewness by fixing the means of the normal distributions to be zero. Thus, our new data generating process

²⁶Norets (2010); Norets and Pelenis (forthcoming, 2012) document the theory and practice of approximating continuous distributions using a finite number of mixtures of normal distributions. They show “that large classes of conditional densities can be approximated in the Kullback-Leibler distance by different specifications of finite smooth mixtures of normal densities or regressions.”

is described as

$$\epsilon_{i,t} = \rho\epsilon_{i,t-1} + \eta_{i,t} \tag{9}$$

$$y_{i,t} = \epsilon_{i,t} + \nu_{i,t} \tag{10}$$

With probability $\pi_{\eta,1}$, $\eta_{i,t}$ is drawn from a normal distribution with variance $h_{\eta,1}$. With probability $\pi_{\eta,2} = (1 - \pi_{\eta,1})$, $\eta_{i,t}$ is drawn from a normal distribution with variance $h_{\eta,2}$. Similarly, $\nu_{i,t}$ is drawn from a normal distribution with variance $h_{\nu,1}$ with probability $\pi_{\nu,1}$, and $\nu_{i,t}$ is drawn from a normal distribution with variance $h_{\nu,2}$ with probability $\pi_{\nu,2} = (1 - \pi_{\nu,1})$. Notice that we have increased the number of parameters to describe the distribution of η and ν . In the benchmark income process, we only needed to estimate four parameters, $(\rho, \Sigma_{Z0}, \Sigma_{\eta}, \Sigma_{\nu})$. Here, we need to estimate eight parameters, $(\rho, \Sigma_{Z0}, \pi_{\eta,1}, h_{\eta,1}, h_{\eta,2}, \pi_{\nu,1}, h_{\nu,1}, h_{\nu,2})$. The added distributional flexibility comes at the cost of additional parameters to estimate. With sufficient sample size, the gains should outweigh the losses.

We calibrate the parameters of the mixture distribution so that both transitory and persistent shocks are fat-tailed (positive excess kurtosis). We set $N=500$ and $T=10$, and calibrate the rest of the parameters to be $(\rho, \Sigma_{Z0}) = (1.00, 0.15)$. The details on how to modify the Gibbs sampling algorithm to handle non-normality is given in Appendix A.5. Prior distributions for ρ and Σ_{Z0} are specified in the same way as in the baseline labor income process. Prior distributions of parameters specific to the mixture normal model is discussed in Appendix A.5. Since it is not common to estimate kurtosis in the GMM framework, we simply apply the unmodified GMM estimator.

Table 8 compares the finite sample properties of the Bayesian estimator and the GMM estimator. For the Bayesian estimator, in addition to the estimates for all parameters, we report the second and fourth moments of the shock distribution implied by the point estimates of the parameters governing the mixture distribution. To concisely convey the results, we report the average of the point estimates across 100 datasets for the parameters governing the mixture.

Even though the Bayesian estimator is now estimating a larger number of parameters, its performance remains almost as good as in the benchmark case. The RMSEs of the Bayesian estimator are roughly the same as those in the benchmark labor income process, except for Σ_{η} . The RMSEs of the Bayesian estimators remain smaller than those of the GMM estimator for the estimated second moments of η , ν , and the initial state $\epsilon_{i,0}$. The Bayesian estimator also performs well in estimating the parameters governing the mixture distribution. For parameters governing the distribution of ν , the mean of the estimator comes very close to the true values, and thus, the implied kurtosis estimate is very close to the true kurtosis. For parameters governing the distribution of η , the means of the estimators are not as close to the true values as in the case for ν . For example, the estimator overstates the probability that ν is drawn from the distribution with larger variance by 4.4 percentage points. As a result, the implied kurtosis estimate is slightly downward biased, but its average estimate (8.29) is reasonably close to the true value of 8.97. For both η and ν ,

the standard deviation and thus RMSEs for the kurtosis estimates are larger compared to other parameters. This is not surprising given that they are identified from the realizations of very large infrequent shocks.

Estimating the model with mixtures of normal distributions is substantially more computationally intensive than estimating the models considered earlier. Thus, we are limited in examining how robust these results are to alternative specifications of the parameters and priors. Nevertheless, our Monte Carlo exercise suggests that the Bayesian estimator is useful in estimating the labor income processes even when the errors are not normally distributed.

5 Conclusion

In this paper, we have examined the validity of performing Bayesian estimation on widely used error component models of labor income processes. Although the differences are not large, the Monte Carlo analysis reveals that Bayesian parameter estimates have favorable efficiency and bias properties relative to GMM in small samples. After accounting for documented features of the data—such as missing values, heteroskedasticity, parameter heterogeneity, and non-normal errors—the beneficial small sample properties of the Bayesian estimator remain. The favorable small sample properties provide a justification for employing Bayesian techniques in future research. The detailed derivation of the estimators provides a clear exposition of the estimation routines and a better understanding of the source of the estimation results. It is also designed to help applied economists develop Bayesian estimation routines for related labor income processes. Our analysis of the extensive Monte Carlo experiments suggests that Bayesian methods are appropriate for the relevant types of processes and data typically used by applied labor economists.

References

- ABOWD, J. M., AND D. CARD (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–445.
- ALTONJI, J. G., AND L. SEGAL (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business and Economic Statistics*, 14(3), 353–366.
- BAKER, M. (1997): "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics*, 15(2), 338–375.
- BLUNDELL, R., L. PISTAFERRI, AND I. PRESTON (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98:5, 1887–1921.
- BROWNING, M., M. EJRNAES, AND J. ALVAREZ (2010): "Modelling Income Processes With Lots of Heterogeneity," *Review of Economic Studies*, 77(4), 1353–1381.
- BROWNING, M., L. P. HANSEN, AND J. J. HECKMAN (1999): *Micro data and general equilibrium models*. 1 of *Handbook of Macroeconomics*, chap. 8, pp. 543–633. Elsevier, 1 edn.
- CARTER, C. K., AND R. KOHN (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81(3), 541–553.
- CLARK, T. E. (1996): "Small-Sample Properties of Estimators of Nonlinear Models of Covariance Structure," *Journal of Business and Economic Statistics*, 14(3), 367–373.
- DURBIN, J., AND S. J. KOOPMAN (2002): "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika*, 89(3), 603–616.
- FRENCH, E. (2004): "The Labor Supply Response to (Mismeasured but) Predictable Wage Changes," *The Review of Economics and Statistics*, 86(2), 602–613.
- GELMAN, A., J. B. CARLIN, H. S. STERN, AND D. B. RUBIN (2009): *Bayesian Data Analysis*. Chapman & Hall.
- GEWEKE, J. (2005): *Contemporary Bayesian Econometrics and Statistics*. Wiley Series in Probability and Statistics.
- (2007): "Interpretation and Inference in Mixture Models: Simple MCMC Works," *Computational Statistics & Data Analysis*, 51(7), 3529–3550.
- GEWEKE, J., AND M. KEANE (2000): "An Empirical Analysis of Earnings Dynamics Among Men in the PSID: 1968-1989," *Journal of Econometrics*, 96, 293–356.
- GOTTSCHALK, P., AND R. A. MOFFITT (1994): "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity*, 25(2), 217–272.
- GUVENEN, F. (2009): "An empirical investigation of labor income processes," *Review of Economic Dynamics*, 12(1), 58–79.
- HAMILTON, J. D. (1994): *Time Series Analysis*. Princeton University Press, 1 edn.
- HEATHCOTE, J., F. PERRI, AND G. L. VIOLANTE (2010): "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States 1967-2006," *Review of Economic Dynamics*, 13(1), 15–51.
- HIRANO, K. (1998): "A Semiparametric Model for Labor Earnings Dynamics," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, ed. by D. D. Dey, P. Muller, and D. Sinha, chap. 20, pp. 355–367. Springer.

- JACQUIER, E., N. G. POLSON, AND P. E. ROSSI (1994): “Bayesian Analysis of Stochastic Volatility Models,” *Journal of Business and Economic Statistics*, 12(4), 371–389.
- JENSEN, S. T., AND S. H. SHORE (2010): “Changes in the Distribution of Income Volatility,” *mimeo*.
- (2011): “Semi-Parametric Bayesian Modelling of Income Volatility Heterogeneity,” *Journal of the American Statistical Association*, 106, 1280–1290.
- LILLARD, L. A., AND Y. WEISS (1979): “Components of Variation in Panel Earnings Data: American Scientists 1960-1970,” *Econometrica*, 47(2), 437–454.
- LILLARD, L. A., AND R. J. WILLIS (1978): “Dynamic Aspects of Earning Mobility,” *Econometrica*, 46(5), 985–1012.
- 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(1), 83–114.
- MEGHIR, C., AND L. PISTAFERRI (2004): “Income Variance Dynamics and Heterogeneity,” *Econometrica*, 72(1), 1–32.
- NORETS, A. (2010): “Approximation of Conditional Densities by Smooth Mixtures of Regressions,” *Annals of Statistics*, 38(3), 1733–1766.
- NORETS, A., AND J. PELENIS (2012): “Bayesian modeling of joint and conditional distributions,” *Journal of Econometrics*, 168(2), 332–346.
- (forthcoming): “Posterior Consistency in Conditional Density Estimation by Covariate Dependent Mixtures,” *Economic Theory*.
- PESARAN, H. (2006): “Estimation and Inference in Large Heterogeneous Panels with Multifactor Error Structure,” *Econometrica*, 74(4), 967–1012.
- ROBERT, C. P., AND G. CASELLA (2004): *Monte Carlo Statistical Methods*, Springer Texts in Statistics. Springer, 2nd edn.
- STORESLETTEN, K., C. I. TELMER, AND A. YARON (2004): “Cyclical Dynamics in Idiosyncratic Labor Market Risk,” *Journal of Political Economy*, 112(3), 695–717.

Tables

Table 1: Finite-Sample Properties of the Bayesian and GMM Estimators:
Benchmark Labor Income Process (N=500, T=10)

Parameter	True Value	Bayesian	GMM	MLE
ρ	1.0000	0.9953 (0.0032) [0.0057]	0.9963 (0.0047) [0.0060]	0.9995 (0.0037) [0.0047]
Σ_{Z0}	0.1500	0.1532 (0.0112) [0.0116]	0.1543 (0.0117) [0.0125]	0.1535 (0.0110) [0.0115]
Σ_{η}	0.0200	0.0204 (0.0016) [0.0016]	0.0210 (0.0020) [0.0023]	0.0201 (0.0015) [0.0015]
Σ_{ν}	0.0500	0.0494 (0.0015) [0.0016]	0.0481 (0.0030) [0.0035]	0.0499 (0.0015) [0.0015]

*For each parameter and each estimator, the top number is the mean of the estimates across samples, the number in parentheses is the standard deviation, and the number in square brackets is the root mean square error.

Table 2: RMSEs of the Bayesian and GMM Estimators for Alternative Calibrations

Parameter	N=500, T=5		N=500, T=20	
	Bayesian	GMM	Bayesian	GMM
ρ	0.0121	0.0122	0.0028	0.0029
Σ_{Z0}	0.0138	0.0133	0.0141	0.0152
Σ_{η}	0.0032	0.0050	0.0010	0.0017
Σ_{ν}	0.0028	0.0048	0.0011	0.0042
Parameter	N=100, T=10		N=2000, T=10	
	Bayesian	GMM	Bayesian	GMM
ρ	0.0116	0.0116	0.0025	0.0024
Σ_{Z0}	0.0305	0.0322	0.0063	0.0065
Σ_{η}	0.0035	0.0059	0.0007	0.0012
Σ_{ν}	0.0037	0.0092	0.0008	0.0019
Parameter	$\rho = 0.8$ (N=500, T=10)			
	Bayesian	GMM		
ρ	0.0140	0.0171		
Σ_{Z0}	0.0176	0.0165		
Σ_{η}	0.0021	0.0025		
Σ_{ν}	0.0019	0.0025		

Table 3: Finite-Sample Properties of the Bayesian and GMM Estimators:
 ρ -misspecified Model (N=500, T=10)

Parameter	True Value	Bayesian	GMM
Σ_{Z_0}	0.1500	0.1203 (0.0088) [0.0310]	0.1183 (0.0092) [0.0326]
Σ_{η}	0.0200	0.0155 (0.0012) [0.0047]	0.0055 (0.0014) [0.0144]
Σ_{ν}	0.0500	0.0537 (0.0017) [0.0041]	0.0729 (0.0029) [0.0230]

*The model defined by equations (1) and (2) is estimated assuming $\rho = 1$, but the true data generating process is given by equations (1) and (2) with $\rho = 0.95$. For each parameter and each estimator, the top number is the mean of the estimates across samples, the number in parentheses is the standard deviation, and the number in square brackets is the root mean square error.

Table 4: Finite-Sample Properties of the Bayesian and GMM Estimators:
Mixture-misspecified Model (N=500, T=10)

Parameter	True Value	Bayesian	GMM
ρ	1.0000	0.996 (0.002) [0.005]	0.9998 (0.0039) [0.0048]
Σ_{Z_0}	0.1500	0.149 (0.013) [0.013]	0.1506 (0.0135) [0.0135]
Σ_{η}	0.0200	0.021 (0.002) [0.002]	0.0213 (0.0023) [0.0026]
Σ_{ν}	0.0500	0.050 (0.003) [0.003]	0.0476 (0.0035) [0.0042]

*The model defined by equations (1) and (2) is estimated assuming that the error terms are distributed according to Normal distributions, but the true data generating process is given by equations (1) and (2) with the error terms distributed as mixtures of Normal distributions, as outline in Section 4.4. Thus, the “True” variances are the implied true variances of η and ν given the Mixture DGP. For each parameter and each estimator, the top number is the mean of the estimates across samples, the number in parentheses is the standard deviation, and the number in square brackets is the root mean square error.

Table 5: Finite-Sample Properties of the Bayesian and GMM Estimators:
 Unbalanced Panel Dataset with Missing Observations (N=500, T=10)

Parameter	True Value	Bayesian	GMM
ρ	1.0000	0.9952 (0.0025) [0.0055]	0.9955 (0.0056) [0.0072]
Σ_{Z0}	0.1500	0.1551 (0.0139) [0.0147]	0.1551 (0.0148) [0.0157]
Σ_{η}	0.0200	0.0205 (0.0016) [0.0017]	0.0213 (0.0033) [0.0036]
Σ_{ν}	0.0500	0.0493 (0.0017) [0.0018]	0.0478 (0.0036) [0.0043]

*For each parameter and each estimator, the top number is the mean of the estimates across samples, the number in parentheses is the standard deviation, and the number in square brackets is the root mean square error.

Table 6: Finite-Sample Properties of the Bayesian and GMM Estimators:
Time-Varying Variances (N=500, T=10)

Parameter	True Value	Mean		RMSE	
		Bayesian	GMM	Bayesian	GMM
ρ	1.0000	0.9954	0.9970	0.0055	0.0050
Σ_{Z_0}	0.1500	0.1501	0.1497	0.0120	0.0121
Σ_{η_2}	0.0207	0.0207	0.0213	0.0070	0.0068
Σ_{η_3}	0.0301	0.0317	0.0325	0.0052	0.0081
Σ_{η_4}	0.0274	0.0281	0.0282	0.0047	0.0063
Σ_{η_5}	0.0293	0.0292	0.0284	0.0050	0.0077
Σ_{η_6}	0.0222	0.0225	0.0240	0.0042	0.0074
Σ_{η_7}	0.0289	0.0290	0.0288	0.0046	0.0067
Σ_{η_8}	0.0157	0.0159	0.0181	0.0052	0.0080
Σ_{η_9}	0.0185	0.0186	0.0202	0.0052	0.0081
Σ_{ν_1}	0.0415	0.0412	0.0399	0.0066	0.0070
Σ_{ν_2}	0.0318	0.0312	0.0300	0.0038	0.0067
Σ_{ν_3}	0.0372	0.0360	0.0351	0.0046	0.0070
Σ_{ν_4}	0.0286	0.0281	0.0275	0.0036	0.0062
Σ_{ν_5}	0.0286	0.0283	0.0265	0.0037	0.0067
Σ_{ν_6}	0.0351	0.0348	0.0338	0.0040	0.0070
Σ_{ν_7}	0.0380	0.0375	0.0354	0.0039	0.0072
Σ_{ν_8}	0.0544	0.0535	0.0520	0.0054	0.0087
Σ_{ν_9}	0.0369	0.0363	0.0354	0.0045	0.0087
$\Sigma_{\nu_{10}}$	0.0506	0.0501	0.0509	0.0051	0.0096

Table 7: Finite-Sample Properties of the Bayesian and GMM Estimators:
Labor Income Process with HIP (N=500, T=10)

Parameter	True Value	Bayesian	GMM
ρ	1.0000	0.9936 (0.0037) [0.0074]	0.9960 (0.0040) [0.0062]
Σ_{Z0}	0.1500	0.1530 (0.0121) [0.0124]	0.1538 (0.0144) [0.0149]
Σ_{η}	0.0200	0.0197 (0.0021) [0.0021]	0.0195 (0.0045) [0.0045]
Σ_{ν}	0.0500	0.0498 (0.0019) [0.0019]	0.0493 (0.0041) [0.0041]
Σ_{β}	0.0004	0.0006 (0.0003) [0.0003]	0.0006 (0.0005) [0.0005]

*For each parameter and each estimator, the top number is the mean of the estimates across samples, the number in parentheses is the standard deviation, and the number in square brackets is the root mean square error.

Table 8: Finite-Sample Properties of the Bayesian and GMM Estimators:
Non-Normal Error Distributions (N=500, T=10)

Parameter	True Value	Bayesian	GMM
ρ	1.0000	0.9960 (0.0023) [0.0047]	0.9987 (0.0036) [0.0045]
Σ_{Z0}	0.1500	0.1490 (0.0115) [0.0115]	0.1492 (0.0135) [0.0135]
Σ_{η}	0.0200	0.0208 (0.0014) [0.0016]	0.0214 (0.0024) [0.0028]
Kurtosis of η	8.9743	8.2942 (1.1551) [1.3354]	na (na) [na]
$h_{\eta,1}$	0.0059	0.0054	na
$h_{\eta,2}$	0.0766	0.0721	na
$\pi_{\eta,1}$	0.8000	0.7560	na
$\pi_{\eta,2}$	0.2000	0.2440	na
Σ_{ν}	0.0500	0.0492 (0.0023) [0.0025]	0.0479 (0.0035) [0.0042]
Kurtosis of ν	9.0112	9.1119 (0.5790) [0.5848]	na (na) [na]
$h_{\nu,1}$	0.0146	0.0141	na
$h_{\nu,2}$	0.1914	0.1902	na
$\pi_{\nu,1}$	0.8000	0.7989	na
$\pi_{\nu,2}$	0.2000	0.2011	na

*For each parameter and each estimator, the top number is the mean of the estimates across samples, the number in parentheses is the standard deviation, and the number in square brackets is the root mean square error.

A Online Technical Appendix

A.1 Log Likelihood of LGSS

Define the canonical linear Gaussian state space (LGSS) form.

$$\begin{aligned} S_t &= A_t S_{t-1} + B_t v_t \\ Y_t &= C_t z_t + D_t S_t + w_t \end{aligned}$$

Then the log likelihood of parameter vector $\theta = (A_t, B_t, C_t, D_t, Q_t, R_t, S_{0|0}, P_{0|0})$ is given by:

$$\begin{aligned} &\ln p(Y^T | \theta) \\ &= \ln \prod_{t=1}^T p(Y_t | Y^{t-1}, \theta) \\ &= \sum_{t=1}^T \ln p(Y_t | Y^{t-1}, \theta) \\ &= \sum_{t=1}^T \ln[(2\pi)^{-0.5k} \det[P_{t|t-1}^{-0.5}] \exp[-0.5(Y_t - C_t z_t - D_t S_{t|t-1})' P_{t|t-1}^{-1} (Y_t - C_t z_t - D_t S_{t|t-1})]] \\ &= \sum_{t=1}^T \ln[(2\pi)^{-0.5k} \det P_{t|t-1}^{-0.5}] + \sum_{t=1}^T \ln[\exp[-0.5(Y_t - C_t z_t - D_t S_{t|t-1})' P_{t|t-1}^{-1} (Y_t - C_t z_t - D_t S_{t|t-1})]] \\ &= -0.5Tk \ln(2\pi) - 0.5 \sum_{t=1}^T \ln[\det P_{t|t-1}] + \sum_{t=1}^T [-0.5(Y_t - C_t z_t - D_t S_{t|t-1})' P_{t|t-1}^{-1} (Y_t - C_t z_t - D_t S_{t|t-1})] \end{aligned}$$

A.2 Bayesian Estimation Procedure

The labor income process we want to estimate is:

$$\begin{aligned} y_{i,t} &= X_{i,t} \beta + \epsilon_{i,t} + \Sigma_\nu^{0.5} \nu_{i,t} \\ \epsilon_{i,t} &= \rho \epsilon_{i,t-1} + \Sigma_\eta^{0.5} \eta_{i,t} \end{aligned}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. $\eta_{i,t}$ and $\nu_{i,t}$ are independent across i and t , and are distributed as standard normal:

$$\eta_{i,t} \sim N(0, 1)$$

and

$$\nu_{i,t} \sim N(0, 1)$$

$\epsilon_{i,0}$ is also distributed as normal:

$$\epsilon_{i,0} \sim N(0, \Sigma_{Z0})$$

$X_{i,t}$ consists of a polynomial of age, race, education, urban residence controls, etc.

Our goal is to draw from $p(\beta, \rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T} | \{y_{i,t}\}_{i=1,N}^{t=1,T})$.

A.2.1 Prior Distribution

We specify the prior distributions for β and ρ to be normal with arbitrarily large variances, truncating $\rho \in [0, 1]$ to ensure stationarity.

$$\begin{aligned}\beta &\sim N(\beta_0, \Gamma_{0,\beta}) \\ \rho &\sim N(\rho_0, \Gamma_{0,\rho}) \mathbb{I}[-1 \leq \rho \leq 1]\end{aligned}$$

We specify the prior distributions for Σ_η , Σ_ν , and Σ_{Z0} to be inverse-gamma.

$$\begin{aligned}\Sigma_\eta &\sim IG(T_{0,\eta}, S_{0,\eta}^{-1}) \\ \Sigma_\nu &\sim IG(T_{0,\nu}, S_{0,\nu}^{-1}) \\ \Sigma_{Z0} &\sim IG(T_{0,Z0}, S_{0,Z0}^{-1})\end{aligned}$$

A.2.2 Gibbs Sampling

We use Gibbs sampling with the following 6 blocks.

$$\begin{aligned}\text{Block 1} &: p\left(\beta \mid \rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 2} &: p\left(\rho \mid \beta, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 3} &: p\left(\Sigma_\eta \mid \beta, \rho, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 4} &: p\left(\Sigma_\nu \mid \beta, \rho, \Sigma_\eta, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 5} &: p\left(\Sigma_{Z0} \mid \beta, \rho, \Sigma_\eta, \Sigma_\nu, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 6} &: p\left(\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T} \mid \beta, \rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right)\end{aligned}$$

Throughout Block 1 to Block 4, the formulae are the same as in the standard case with one person, except that we have T*N observations, instead of T observations.

Block 1: β

It is identical to the problem of drawing from the posterior of coefficients in a linear regression model. The posterior distribution is normal, and its mean and variance are analytically computed (see Chapter 2 of Geweke (2005)).

$$p\left(\beta \mid \rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \sim N(\mu_{1,\beta}, \Gamma_{1,\beta})$$

where²⁷

²⁷In the case of an informative prior, the formula is given by

$$\begin{aligned}\Gamma_{1,\beta} &= \Sigma_\nu \left[\sum_{i=1}^N \sum_{t=1}^T X_{i,t} X_{i,t}' \right]^{-1} \\ \mu_{1,\beta} &= \Gamma_{1,\beta} \left[\frac{1}{\Sigma_\nu} \sum_{i=1}^N \sum_{t=1}^T X_{i,t} (y_{i,t} - \epsilon_{i,t})' \right]\end{aligned}$$

Block 2: ρ

The same as in Block 1, except for truncation (see p.169 of Geweke (2005)).

$$p\left(\rho|\beta, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \sim N(\mu_{1,\rho}, \Gamma_{1,\rho}) [-1 \leq \rho \leq 1]$$

where

$$\begin{aligned}\Gamma_{1,\rho} &= \Sigma_\eta \left[\sum_{i=1}^N \sum_{t=1}^T \epsilon_{i,t-1}^2 \right]^{-1} \\ \mu_{1,\rho} &= \Gamma_{1,\rho} \left[\frac{1}{\Sigma_\eta} \sum_{i=1}^N \sum_{t=1}^T \epsilon_{i,t-1} \epsilon_{i,t} \right]\end{aligned}$$

Block 3: Σ_ν

$$p\left(\Sigma_\nu|\beta, \rho, \Sigma_\eta, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \sim IG(T_{1,\nu}, S_{1,\nu}^{-1})$$

where

$$\begin{aligned}T_{1,\nu} &= T_{0,\nu} + TN \\ S_{1,\nu} &= S_{0,\nu} + \sum_{i=1}^N \sum_{t=1}^T (y_{i,t} - X_{i,t}\beta - \epsilon_{i,t})^2\end{aligned}$$

$$\begin{aligned}\Gamma_1 &= \left[\Gamma_0^{-1} + \frac{1}{\sigma^2} X'X \right]^{-1} \\ \mu_1 &= \Gamma_1 \left[\Gamma_0^{-1} \mu_0 + \frac{1}{\sigma^2} X'y \right]\end{aligned}$$

As Γ_0 goes to infinity,

$$\begin{aligned}\Gamma_1 &\rightarrow \sigma^2 [X'X]^{-1} \\ \mu_1 &\rightarrow [X'X]^{-1} [X'y]\end{aligned}$$

which coincides with β_{ols} and $var(\beta_{ols})$.

Block 4: Σ_η

$$p\left(\Sigma_\eta|\beta, \rho, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \sim IG(T_{1,\eta}, S_{1,\eta}^{-1})$$

where

$$\begin{aligned} T_{1,\eta} &= T_{0,\eta} + TN \\ S_{1,\eta} &= S_{0,\eta} + \sum_{i=1}^N \sum_{t=1}^T (\epsilon_{i,t} - \rho\epsilon_{i,t-1})^2 \end{aligned}$$

Block 5: Σ_{Z0}

$$p\left(\Sigma_{Z0}|\beta, \rho, \Sigma_\eta, \Sigma_\nu, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \sim IG(T_{1,Z0}, S_{1,Z0}^{-1})$$

where

$$\begin{aligned} T_{1,Z0} &= T_{1,Z0} + N \\ S_{1,Z0} &= S_{0,Z0} + \sum_{i=1}^N \epsilon_{i,0}^2 \end{aligned}$$

Block 6: $\{\epsilon_t\}$

For each individual i , we apply the Carter-Kohn algorithm.²⁸

Step 1: Specify $S_{0|0}$ and $P_{0|0}$

Step 2: Kalman-Filter

Starting from $S_{0|0}$ and $P_{0|0}$, compute $\{P_{t|t-1}, P_{t|t}\}_{t=1}^T$ and $\{S_{t|t}\}_{t=1}^T$ using the Kalman Filter.

$$\begin{aligned} S_{t|t-1} &= A_t S_{t-1|t-1} \\ P_{t|t-1} &= A_t P_{t-1|t-1} A_t' + B Q_t B' \\ G_t &= D_t P_{t|t-1} D_t' + R_t \\ K_t &= P_{t|t-1} D_t' G_t^{-1} \\ S_{t|t} &= S_{t|t-1} + K_t (Y_t - C_t z_t - D_t S_{t|t-1}) \\ P_{t|t} &= (I_{n_s} - K_t D_t) P_{t|t-1} \end{aligned}$$

where $S_{s|t} \equiv E[S_s|Y^t]$ and $P_{s|t} \equiv Var[S_s|Y^t]$.

Step 3: Carter-Kohn's backward recursion:

At each $t=T$ through 1 going backward, compute

$$\begin{aligned} S_{t-1|t} &= S_{t-1|t-1} + P_{t-1|t-1} A_t' P_{t|t-1}^{-1} (S_t - A_t S_{t-1|t-1}) \\ P_{t-1|t} &= P_{t-1|t-1} - P_{t-1|t-1} A_t' P_{t|t-1}^{-1} A_t P_{t-1|t-1} \end{aligned}$$

²⁸For details see Carter and Kohn (1994).

based on S_t , and draw S_{t-1} from $N(S_{t-1|t}, P_{t-1|t})$.

A.3 Bayesian Estimation Procedure with Missing Data

This section presents how to modify Gibbs Block 6 to handle missing observations. In terms of estimation, unbalanced panel datasets can be seen as containing systematic missing observations and does not introduce extra complication.

The LGSS with missing observations can be written as

$$\begin{aligned} S_t &= A_t S_{t-1} + B v_t \\ Y_t &= C_t z_t + D_t S_t + w_t \end{aligned}$$

where

$$\begin{aligned} v_t &\sim N(0, Q_t) \\ w_t &\sim N(0, R_t) \\ S_0 &\sim N(S_{0|0}, P_{0|0}) \end{aligned}$$

When working with missing data, it is convenient to let Y_t denote a vector containing the incomes of multiple individuals at time t .²⁹ Note that some elements of Y_t may be missing, i.e., $Y_{i,t} = 'NaN'$. Let $n_{m,t}$ be the number of missing observations at time t . Let W_t be a $(n_y - n_{m,t}) - by - n_y$ matrix that selects the elements of Y_t that are not missing. For example, suppose Y_t is 6-by-1, and the second and fourth elements are missing.

$$Y_t = \begin{pmatrix} 3.5 \\ 'NaN' \\ 2.1 \\ 'NaN' \\ 3.2 \\ 5.9 \end{pmatrix}$$

Then, W_t is a 4-by-6 matrix with the following structure.

$$W_t = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

so that

$$\bar{Y}_t \equiv W_t Y_t = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3.5 \\ 'NaN' \\ 2.1 \\ 'NaN' \\ 3.2 \\ 5.9 \end{pmatrix} = \begin{pmatrix} 3.5 \\ 2.1 \\ 3.2 \\ 5.9 \end{pmatrix}$$

When no value is missing, W_t is a $n_y - by - n_y$ identity matrix, and $\bar{Y}_t = Y_t$. Using this variable W_t , we transform the observation equation for each t to obtain

²⁹In all other sections, Y_t refers to the income of a single individual.

$$\begin{aligned}
S_t &= A_t S_{t-1} + B v_t \\
W_t Y_t &= W_t C_t z_t + W_t D_t S_t + W_t w_t
\end{aligned}$$

Finally, create an indicator variable ϕ_t which takes the value 0 when all values of Y_t are missing, and 1 otherwise. If $\phi_t = 0$, set $W_t = I_{n_y}$.

A.3.1 Estimation Algorithm

To modify Block 6: Gibbs Sampling, we must only modify Step 2. Other steps are unmodified.

Step 2: Kalman-Filter

Starting from $S_{0|0}$ and $P_{0|0}$, compute $\{P_{t|t-1}, P_{t|t}\}_{t=1}^T$ and $\{S_{t|t}\}_{t=1}^T$ using Kalman Filter.

$$\begin{aligned}
S_{t|t-1} &= A_t S_{t-1|t-1} \\
P_{t|t-1} &= A_t P_{t-1|t-1} A_t' + B Q_t B' \\
G_t &= W_t D_t P_{t|t-1} D_t' W_t' + W_t R_t W_t' \\
K_t &= P_{t|t-1} D_t' W_t' G_t^{-1} \\
S_{t|t} &= S_{t|t-1} + \phi_t K_t (W_t Y_t - W_t C_t z_t - W_t D_t S_{t|t-1}) \\
P_{t|t} &= (I_{n_s} - K_t W_t D_t) P_{t|t-1}
\end{aligned}$$

where $S_{s|t} \equiv E[S_s|Y^t]$ and $P_{s|t} \equiv Var[S_s|Y^t]$. In case where all values of Y_t are missing, simply set

$$\begin{aligned}
S_{t|t-1} &= A_t S_{t-1|t-1} \\
P_{t|t-1} &= A_t P_{t-1|t-1} A_t' + B Q_t B' \\
S_{t|t} &= S_{t|t-1} \\
P_{t|t} &= P_{t|t-1}
\end{aligned}$$

A.4 Bayesian Estimation Procedure with Heterogeneous Income Profiles

The labor income process we want to estimate is:

$$\begin{aligned}
y_{i,t} &= X_{i,t} \beta_i + \epsilon_{i,t} + \Sigma_\nu^{0.5} \nu_{i,t} \\
\epsilon_{i,t} &= \rho \epsilon_{i,t-1} + \Sigma_\eta^{0.5} \eta_{i,t}
\end{aligned}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$. $\eta_{i,t}$ and $\nu_{i,t}$ are independent across i and t , and are distributed as standard normal. $\epsilon_{i,0}$ is distributed as normal with variance Σ_{Z0} . For each i , β_i is distributed as mean zero and variance Σ_β :

$$\beta_i \sim N(0, \Sigma_\beta)$$

$X_{i,t}$ consists of a polynomial of age, race, education, urban residence controls, etc. We use β^N to denote $\{\beta_i\}_{i=1}^N$. Our goal is to draw from $p\left(\beta^N, \rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T} | \{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$.

A.4.1 Prior Distribution

Select the prior distribution for ρ to be normal and the distributions for Σ_η , Σ_ν , and Σ_{Z0} to be inverse-gamma. We specify the prior distributions for Σ_β to be inverse-gamma.

$$\Sigma_\beta \sim IG(T_{0,\beta}, S_{0,\beta}^{-1})$$

A.4.2 Gibbs Sampling

We use Gibbs sampling with the following 7 blocks.

$$\begin{aligned} \text{Block 1} & : p\left(\beta^N | \rho, \Sigma_\beta, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 2} & : p\left(\Sigma_\beta | \rho, \beta^N, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 3} & : p\left(\rho | \beta^N, \Sigma_\beta, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 4} & : p\left(\Sigma_\eta | \beta^N, \rho, \Sigma_\beta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 5} & : p\left(\Sigma_\nu | \beta^N, \rho, \Sigma_\beta, \Sigma_\eta, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 6} & : p\left(\Sigma_{Z0} | \beta^N, \rho, \Sigma_\beta, \Sigma_\eta, \Sigma_\nu, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \text{Block 7} & : p\left(\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T} | \beta^N, \Sigma_\beta, \rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \end{aligned}$$

Modification of Blocks 3–7 are straightforward. We focus on block 1 and 2.

Block 1: β^N

As in the previous case, the posterior is normally distributed. However, we now need to compute the posterior mean and variance for each individual.

$$p\left(\beta_i | \rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \Sigma_\beta, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \sim N(\mu_{i,1,\beta}, \Gamma_{i,1,\beta})$$

where

$$\begin{aligned} \Gamma_{i,1,\beta} & = \left[\Sigma_\beta^{-1} + \frac{1}{\Sigma_\nu} X_i' X_i \right]^{-1} \\ \mu_{i,1,\beta} & = \Gamma_{i,1,\beta} \left[\frac{1}{\Sigma_\nu} X_i' (y_i - \epsilon_i) \right] \end{aligned}$$

Block 2: Σ_β

$$p\left(\Sigma_\beta | \beta^N, \rho, \Sigma_\eta, \Sigma_\nu, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=0,N}^{t=1,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \sim IG(T_{1,\beta}, S_{1,\beta}^{-1})$$

where

$$\begin{aligned}
T_{1,\beta} &= T_{0,\beta} + N \\
S_{1,\beta} &= S_{0,\beta} + \sum_{i=1}^N (\beta_i)^2
\end{aligned}$$

A.5 Bayesian Estimation Procedure with a Mixture of Normal Distributions

The labor income process we want to estimate is:

$$\begin{aligned}
y_{i,t} &= X_{i,t}\beta + \epsilon_{i,t} + \nu_{i,t} \\
\epsilon_{i,t} &= \rho\epsilon_{i,t-1} + \eta_{i,t}
\end{aligned}$$

for $i = 1, \dots, N$ and $t = 1, \dots, T$.

$\eta_{i,t}$ and $\nu_{i,t}$ are independent across i and t . They are distributed according to a mixture of m_η and m_ν number of normals, respectively.

$$\begin{aligned}
\eta_{i,t} | s_{\eta,i,t}=j &\sim N(0, h_\eta h_\eta^j) \\
p(s_{\eta,i,t} = j) &= \pi_\eta^j \quad \forall j \in 1, 2, \dots, m_\eta \\
\nu_{i,t} | s_{\nu,i,t}=j &\sim N(0, h_\nu h_\nu^j) \\
p(s_{\nu,i,t} = j) &= \pi_\nu^j \quad \forall j \in 1, 2, \dots, m_\nu
\end{aligned}$$

For the shock $k \in \{\eta, \nu\}$, $h_k h_k^j$ is the variance of distribution j , and π_k^j is the probability weight for distribution j .^{30,31} $s_{k,i,t}$ indicates from which distribution of the mixture the innovation for individual i in time t came.

As before, $\epsilon_{i,0}$ is distributed as normal:

$$\epsilon_{i,0} \sim N(0, \Sigma_{Z0})$$

Our goal is to draw from

$$p\left(\beta, \rho, h_\nu, \{h_\nu^j\}_{j=1}^{m_\nu}, \{\pi_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\eta}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T} | \{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

A.5.1 Prior

We specify the prior for β and ρ to be normal with arbitrarily large variance.

³⁰Note: h and h_j are not identified separately. Since we are only interested in h times h^j the lack of identification is inconsequential. h is added for numerical/computational reasons. See Gelman, Carlin, Stern, and Rubin (2009).

³¹The likelihood function of the mixture model is invariant to permutation of the components of the mixture. In our application, we simply label the component with a smaller variance as the first component and the component with a larger variance as the second component. For more detailed discussions, see Geweke (2007).

$$\begin{aligned}\beta &\sim N(\beta_0, \Gamma_{0,\beta}) \\ \rho &\sim N(\rho_0, \Gamma_{0,\rho}) \mathbb{I}[-1 \leq \rho \leq 1]\end{aligned}$$

We specify the prior for Σ_{Z_0} , h_ν , $\{h_\nu^j\}_{j=1}^{m_\nu}$, h_η , and $\{h_\eta^j\}_{j=1}^{m_\eta}$ to be inverse-gamma.

$$\begin{aligned}\Sigma_{Z_0} &\sim IG\left(T_{0,Z_0}, S_{0,Z_0}^{-1}\right) \\ h_\nu &\sim IG\left(T_{h_\nu,0}, S_{h_\nu,0}^{-1}\right) \\ h_\nu^j &\sim IG\left(T_{h_\nu^j,0}, S_{h_\nu^j,0}^{-1}\right) \\ h_\eta &\sim IG\left(T_{h_\eta,0}, S_{h_\eta,0}^{-1}\right) \\ h_\eta^j &\sim IG\left(T_{h_\eta^j,0}, S_{h_\eta^j,0}^{-1}\right)\end{aligned}$$

We specify the prior for $\{\pi_\nu^j\}_{j=1}^{m_\nu}$ and $\{\pi_\eta^j\}_{j=1}^{m_\eta}$ to be Dirichlet.

$$\begin{aligned}p\left(\{\pi_\eta^j\}_{j=1}^{m_\eta}\right) &\sim D(r, r, \dots, r) \\ p\left(\{\pi_\nu^j\}_{j=1}^{m_\nu}\right) &\sim D(r, r, \dots, r)\end{aligned}$$

In the estimation of the model with two normal mixtures in Section 4.4, the true value for h_η and h_ν are set to one. For prior distributions, we set $T_{h_\nu,0} = T_{h_\nu^1,0} = T_{h_\nu^2,0} = 2$, $S_{h_\nu,0} = 1$, $S_{h_\nu^1,0} = S_{h_\nu^2,0} = 0.01$, $T_{h_\eta,0} = T_{h_\eta^1,0} = T_{h_\eta^2,0} = 2$, $S_{h_\eta,0} = 1$, $S_{h_\eta^1,0} = S_{h_\eta^2,0} = 0.01$. For the parameter of the Dirichlet distribution, we set $r = 0.5$ for both π_η and π_ν .

A.5.2 Estimation Algorithm

Gibbs sampling consists of the following 6 main blocks:

Block 1:

$$p\left(\rho | \beta, h_\nu, \{h_\nu^j\}_{j=1}^{m_\nu}, \{\pi_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\eta}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z_0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 2:

$$p\left(\rho | \beta, h_\nu, \{h_\nu^j\}_{j=1}^{m_\nu}, \{\pi_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\eta}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z_0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 3.1:

$$p\left(h_\nu | \beta, \rho, \{h_\nu^j\}_{j=1}^{m_\nu}, \{\pi_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\eta}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z_0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 3.2:

$$p\left(\{h_\nu^j\}_{j=1}^{m_\nu} | \beta, \rho, h_\nu, \{\pi_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\eta}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z_0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 3.3:

$$p\left(\{\pi_\nu^j\}_{j=1}^{m_\nu} | \beta, \rho, h_\nu, \{h_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\eta}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z_0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 3.4:

$$p\left(\{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}|\beta,\rho,h_{\nu},\{h_{\nu}^j\}_{j=1}^{m_{\nu}},\{\pi_{\nu}^j\}_{j=1}^{m_{\nu}},h_{\eta},\{h_{\eta}^j\}_{j=1}^{m_{\eta}},\{\pi_{\eta}^j\}_{j=1}^{m_{\eta}},\{s_{\eta,i,t}\}_{i=1,N}^{t=1,T},\Sigma_{Z0},\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T},\{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 4.1:

$$p\left(h_{\eta}|\beta,\rho,h_{\nu},\{h_{\nu}^j\}_{j=1}^{m_{\nu}},\{\pi_{\nu}^j\}_{j=1}^{m_{\nu}},\{s_{\nu,i,t}\}_{i=1,N}^{t=1,T},\{h_{\eta}^j\}_{j=1}^{m_{\eta}},\{\pi_{\eta}^j\}_{j=1}^{m_{\eta}},\{s_{\eta,i,t}\}_{i=1,N}^{t=1,T},\Sigma_{Z0},\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T},\{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 4.2:

$$p\left(\{h_{\eta}^j\}_{j=1}^{m_{\eta}}|\beta,\rho,h_{\nu},\{h_{\nu}^j\}_{j=1}^{m_{\nu}},\{\pi_{\nu}^j\}_{j=1}^{m_{\nu}},\{s_{\nu,i,t}\}_{i=1,N}^{t=1,T},\{\pi_{\eta}^j\}_{j=1}^{m_{\eta}},\{s_{\eta,i,t}\}_{i=1,N}^{t=1,T},\Sigma_{Z0},\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T},\{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 4.3:

$$p\left(\{\pi_{\eta}^j\}_{j=1}^{m_{\eta}}|\beta,\rho,h_{\nu},\{h_{\nu}^j\}_{j=1}^{m_{\nu}},\{\pi_{\nu}^j\}_{j=1}^{m_{\nu}},\{s_{\nu,i,t}\}_{i=1,N}^{t=1,T},h_{\eta},\{h_{\eta}^j\}_{j=1}^{m_{\eta}},\{s_{\eta,i,t}\}_{i=1,N}^{t=1,T},\Sigma_{Z0},\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T},\{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 4.4:

$$p\left(\{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}|\beta,\rho,h_{\nu},\{h_{\nu}^j\}_{j=1}^{m_{\nu}},\{\pi_{\nu}^j\}_{j=1}^{m_{\nu}},\{s_{\nu,i,t}\}_{i=1,N}^{t=1,T},h_{\eta},\{h_{\eta}^j\}_{j=1}^{m_{\eta}},\{\pi_{\eta}^j\}_{j=1}^{m_{\eta}},\Sigma_{Z0},\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T},\{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 5:

$$p\left(\Sigma_{Z0}|\beta,\rho,h_{\nu},\{h_{\nu}^j\}_{j=1}^{m_{\nu}},\{\pi_{\nu}^j\}_{j=1}^{m_{\nu}},\{s_{\nu,i,t}\}_{i=1,N}^{t=1,T},h_{\eta},\{h_{\eta}^j\}_{j=1}^{m_{\eta}},\{\pi_{\eta}^j\}_{j=1}^{m_{\eta}},\{s_{\eta,i,t}\}_{i=1,N}^{t=1,T},\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T},\{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Block 6:

$$p\left(\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}|\beta,\rho,h_{\nu},\{h_{\nu}^j\}_{j=1}^{m_{\nu}},\{\pi_{\nu}^j\}_{j=1}^{m_{\nu}},\{s_{\nu,i,t}\}_{i=1,N}^{t=1,T},h_{\eta},\{h_{\eta}^j\}_{j=1}^{m_{\eta}},\{\pi_{\eta}^j\}_{j=1}^{m_{\eta}},\{s_{\eta,i,t}\}_{i=1,N}^{t=1,T},\Sigma_{Z0},\{y_{i,t}\}_{i=1,N}^{t=1,T}\right)$$

Notation

Let $(NT)_{\nu}^j$ be the number of i - t pairs that have a ν innovation drawn from the j^{th} distribution in the ν mixture. That is,

$$(NT)_{\nu}^j = \sum_{i=1}^N \sum_{t=1}^T \delta(s_{\nu,i,t}, j)$$

where, δ is the Kronecker delta function: $\delta(a, b) = 1$ if $a = b$ and $\delta(a, b) = 0$ if $a \neq b$.

Block 1

It is identical to the problem of drawing from the posterior of coefficients in a linear regression model. The posterior distribution is normal, and its mean and variance are analytically computed.

$$p\left(\beta|\rho,h_{\nu},\{h_{\nu}^j\}_{j=1}^{m_{\nu}},\{\pi_{\nu}^j\}_{j=1}^{m_{\nu}},\{s_{\nu,i,t}\}_{i=1,N}^{t=1,T},h_{\eta},\{h_{\eta}^j\}_{j=1}^{m_{\eta}},\{\pi_{\eta}^j\}_{j=1}^{m_{\eta}},\{s_{\eta,i,t}\}_{i=1,N}^{t=1,T},\Sigma_{Z0},\{\epsilon_{i,t}\}_{i=1,N}^{t=0,T},\{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \sim N(\mu_{1,\beta}, \Gamma_{1,\beta})$$

where

$$\Gamma_{1,\beta} = \left[\sum_{i=1}^N \sum_{t=1}^T \frac{1}{h_{\nu} h_{\nu}^{s_{\nu,i,t}}} X_{i,t} X_{i,t}' \right]^{-1}$$

$$\mu_{1,\beta} = \Gamma_{1,\beta} \left[\sum_{i=1}^N \sum_{t=1}^T \frac{1}{h_{\nu} h_{\nu}^{s_{\nu,i,t}}} X_{i,t} (y_{i,t} - \epsilon_{i,t})' \right]$$

Block 2

$$p\left(\rho|\beta, h_\nu, \{h_\nu^j\}_{j=1}^{m_\nu}, \{\pi_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\eta}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \sim N(\mu_{1,\rho}, \Gamma_{1,\rho}) [-1 \leq \rho \leq 1]$$

where

$$\Gamma_{1,\rho} = \left[\sum_{i=1}^N \sum_{t=1}^T \frac{1}{h_\eta h_\nu^{s_{\eta,i,t}}} \epsilon_{i,t-1}^2 \right]^{-1} \\ \mu_{1,\rho} = \Gamma_{1,\rho} \left[\sum_{i=1}^N \sum_{t=1}^T \frac{1}{h_\nu h_\nu^{s_{\nu,i,t}}} \epsilon_{i,t-1} \epsilon_{i,t} \right]$$

Block 3.1

$$p\left(h_\nu|\beta, \rho, \{h_\nu^j\}_{j=1}^{m_\nu}, \{\pi_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\eta}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \sim IG\left(S_{h_\nu,1}^{-1}, T_{h_\nu,1}\right)$$

where

$$T_{h_\nu,1} = T_{h_\nu,0} + m_\nu + NT \\ S_{h_\nu,1} = S_{h_\nu,0} + \sum_{i=1}^N \sum_{t=1}^T \frac{1}{h_\nu^{s_{\nu,i,t}}} (y_{i,t} - X_{i,t}\beta - \epsilon_{i,t})^2$$

Block 3.2

$$p\left(\{h_\nu^j\}_{j=1}^{m_\nu}|\beta, \rho, h_\nu, \{\pi_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\eta}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \sim IG\left(S_{h_\nu^j,1}^{-1}, T_{h_\nu^j,1}\right)$$

where

$$T_{h_\nu^j,1} = T_{\nu,0}^j + (NT)_\nu^j \\ S_{h_\nu^j,1} = S_{h_\nu^j,0} + \frac{1}{h_\nu} \sum_{i=1}^N \sum_{t=1}^T \delta(s_{\nu,i,t}, j) (y_{i,t} - X_{i,t}\beta - \epsilon_{i,t})^2$$

for all $1 \leq j \leq m_\nu$.

Block 3.3

$$p\left(\{\pi_\nu^j\}_{j=1}^{m_\nu}|\beta, \rho, h_\nu, \{h_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\nu}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \sim D(b_{\nu,1}, b_{\nu,2}, \dots, b_{\nu,m_\nu})$$

where

$$b_{\nu,j} = r + (NT)_\nu^j$$

for all $1 \leq j \leq m_\nu$.

Block 3.4

$$p\left(s_{\nu,i,t} = j|\beta, \rho, h_\nu, \{h_\nu^j\}_{j=1}^{m_\nu}, \{\pi_\nu^j\}_{j=1}^{m_\nu}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\nu}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \Sigma_{Z0}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \sim \pi_\nu^j (h_\nu^j)^{-0.5} \exp\left[-\frac{(y_{i,t} - X_{i,t}\beta - \epsilon_{i,t})^2}{2h_\nu h_\nu^j}\right]$$

for all $1 \leq t \leq T, 1 \leq i \leq N$.

Block 4

This block is analogous to Block 3. Replace each ν subscript with η and $y_{i,t} - X_{i,t}\beta - \epsilon_{i,t}$ with $\epsilon_{i,t} - \rho\epsilon_{i,t-1}$.

Block 5

$$p\left(\Sigma_{Z0}|\beta, \rho, h_\nu, \{h_\nu^j\}_{j=1}^{m_\nu}, \{\pi_\nu^j\}_{j=1}^{m_\nu}, \{s_{\nu,i,t}\}_{i=1,N}^{t=1,T}, h_\eta, \{h_\eta^j\}_{j=1}^{m_\eta}, \{\pi_\eta^j\}_{j=1}^{m_\nu}, \{s_{\eta,i,t}\}_{i=1,N}^{t=1,T}, \{\epsilon_{i,t}\}_{i=1,N}^{t=0,T}, \{y_{i,t}\}_{i=1,N}^{t=1,T}\right) \\ \sim IG\left(T_{1,Z0}, S_{1,Z0}^{-1}\right)$$

where

$$T_{1,Z0} = T_{1,Z0} + N \\ S_{1,Z0} = S_{0,Z0} + \sum_{i=1}^N \epsilon_{i,0}^2$$

Block 6

For each individual i , we apply the Carter-Kohn algorithm. To cast our labor income processes into our canonical form,

$$A_t = \rho, \quad B_t = 1, \quad C_t = \beta, \\ D_t = 1, \quad Q_t = h_\eta h_\eta^{s_{\eta,i,t}}, \quad R_t = h_\nu h_\nu^{s_{\nu,i,t}}, \\ z_t = X_{i,t}, \quad S_{0|0} = 0, \quad P_{0|0} = \Sigma_{Z0}$$