

# Semiparametric Bayesian Inference for Dynamic Tobit Panel Data Models with Unobserved Heterogeneity

Tong Li\* and Xiaoyong Zheng<sup>†</sup>

Indiana University

September 2004

## Abstract

This paper develops semiparametric Bayesian methods for inference of dynamic Tobit panel data models. Our approach only requires to specify the conditional mean dependence of the unobserved heterogeneity on the initial conditions and the strictly exogenous variables; no further distributional assumption on the unobserved heterogeneity is needed. Important quantities of economic interest such as the average partial effect and average transition probabilities can be readily obtained as a by-product of the Markov chain Monte Carlo (MCMC) run. We apply our method to study female labor supply using a panel data set from the National Longitudinal Survey of Youth 1979 (NLSY79).

**Keywords:** Average Partial Effect; Average Transition Probability; Censored Regression; Dirichlet Process Prior; Female Labor Supply; Infinite Mixture of Normals; Initial Conditions; Markov chain Monte Carlo; Nonlinear Panel Data Models

---

\*Contact Address: Department of Economics, Wylie Hall, Indiana University, Bloomington, IN 47405. Phone: (812) 855-2363, Fax: (812) 855-3736, Email: [toli@indiana.edu](mailto:toli@indiana.edu).

<sup>†</sup>Contact Address: Department of Economics, Wylie Hall, Indiana University, Bloomington, IN 47405. Phone: (812) 855-5308, Fax: (812) 855-3736, Email: [xzheng@indiana.edu](mailto:xzheng@indiana.edu).

# 1 Introduction

This paper develops semiparametric Bayesian methods for inference of dynamic Tobit panel data models with unobserved individual heterogeneity, and applies them to study female labor supply using the National Longitudinal Survey of Youth 1979 (NLSY79). Our approach only requires to specify the conditional mean dependence of the unobserved individual heterogeneity on the initial conditions and the strictly exogenous variables; no further distributional assumption on the unobserved heterogeneity is needed.

Because of their feasibility in modeling the unobserved individual heterogeneity and the state dependence at the same time, dynamic panel data models provide a general framework to study more complex economic relationships. For example, Chiappori and Salanie (2000) argue that they can be used to distinguish between moral hazard and adverse selection in auto insurance markets. Dynamic nonlinear panel data models, however, have presented challenges because of the difficulty arising from dealing with unobserved heterogeneity in general and initial conditions in particular. For Tobit models, for example, while they have been used widely in cross-section studies in labor economics and other applied microeconomics areas, they have been rarely applied in the dynamic panel data framework.<sup>1</sup>

As is well known, for dynamic panel data models with unobserved effects, an important issue is the treatment of the initial observations. For linear models with an additive unobserved effect, appropriate transformations such as differencing have been used to eliminate the unobserved effect, and GMM type estimators have been proposed to estimate the transformed model. For example, see Anderson and Hsiao (1982), Arellano and Bover (1995), Ahn and Schmidt (1995), Blundell and Bond (1998) and Hahn (1999), among others surveyed in Arellano and Honoré (2001) and Hsiao (2003). For nonlinear models, however, the treatment becomes more complicated because the unobserved effect in general cannot be eliminated through some transformations with only a few exceptions. For instance, Honoré (1993) derives the orthogonality conditions for the dynamic Tobit panel data models with

---

<sup>1</sup>To the best of our knowledge, this paper provides the first empirical application of dynamic Tobit panel data models with unobserved heterogeneity.

unobserved individual heterogeneity, and Honoré and Hu (2001) provide a set of sufficient conditions for the consistency and asymptotic normality of the estimator for these models. While their methods have an advantage in that they leave the distribution of the unobserved individual heterogeneity unspecified and allow arbitrary correlation between the unobserved heterogeneity and the explanatory variables, some restrictions on the data need to be met. For example, in Honoré and Hu (2001), time dummies are not allowed in the strictly exogenous covariates. Moreover, some quantities of economic interest such as the average partial effects cannot be estimated using their approach.

As summarized in Hsiao (2003), there have been mainly three different ways of treating initial observations in parametric inference of dynamic nonlinear panel data models. The first approach is to assume the initial conditions for each cross-section unit as nonrandom. The second (and more reasonable) approach is to allow the initial condition to be random, and to specify a joint distribution of all outcomes on the response including that in the initial period conditional on the unobserved heterogeneity term and observed strictly exogenous covariates. The third approach is to approximate the conditional distribution of the initial condition, as suggested by Heckman (1981a). Wooldridge (2003) discusses the advantages and disadvantages of these three approaches. He also suggests a simple alternative approach that is to model the distribution of the unobserved effect conditional on the initial observations and exogenous variables.<sup>2</sup> One of the advantages of Wooldridge's approach is that by specifying the (auxiliary) distribution of the unobserved heterogeneity conditional on the initial conditions to be normal, estimation of probit, Tobit and Poisson regression can be conducted using standard software. Average partial effects can also be estimated in a straightforward manner. This approach, on the other hand, can be subject to misspecification of the distribution of the unobserved effect conditional on the initial value and other exogenous covariates.

In this paper, we adopt an approach that is in a similar spirit to Wooldridge (2003) in modeling the relationship between the unobserved heterogeneity and the initial conditions,

---

<sup>2</sup>As acknowledged in Wooldridge (2003), this approach has been previously suggested for particular models including AR(1) without covariates by Chamberlain (1980), Blundell and Smith (1991) and Blundell and Bond (1998), and the binary response model with a lagged dependent variable by Arellano and Carrasco (2003) who take the distribution of the unobserved effect given the initial condition to be discrete.

but provide a more robust way of handling this relationship. Our approach is to only specify the conditional mean dependence of the unobserved effect on the initial conditions and strictly exogenous covariates, and leave the distribution of the remaining random error term unspecified. We adopt the recently developed Bayesian nonparametric method to estimate this unknown distribution. Specifically, we use an infinite mixture of normals. This is justified because Ferguson (1983) notes that any probability density function can be approximated arbitrarily closely in the  $L_1$  norm by a countable mixture of normal densities. Also, to implement this method we do not need to specify the number of components in approximating mixture of normals. During the Bayesian updating, after recovering the unobserved individual heterogeneity terms, we can use the Bayesian density estimation method to estimate their densities with a Dirichlet process prior. This approach is introduced by Lo (1984) and Ferguson (1983), with the later work by Escobar (1994), Escobar and West (1995), and West, Müller, and Escobar (1994) discussing its computational issues. It is worth noting that the advancement of the Markov chain Monte Carlo (MCMC) method makes the posterior analysis using the Bayesian nonparametric method computationally feasible, and stimulates its recent development.<sup>3</sup> Recently, this method has been adopted by econometricians in addressing various issues. See, e.g., Hirano (2002) for estimation of (linear) autoregressive panel data models, Chib and Hamilton (2002) for analysis of longitudinal data treatment models, Hasegawa and Kozumi (2003) for estimation of Lorenz curves, and Griffin and Steel (2004) for inference of stochastic frontier models.

We propose a semiparametric Bayesian method to estimate the dynamic Tobit panel data model with lagged censored dependent variables.<sup>4</sup> The semiparametric Bayesian approach offers several main advantages for dynamic Tobit panel data models. First, it is robust to misspecification of distributional assumptions about the unobserved individual heterogeneity.

---

<sup>3</sup>For surveys on the use of the MCMC techniques, see, e.g., Chib and Greenberg (1996) and Chib (2001).

<sup>4</sup>The model we consider here is referred to as the censored regression model with corner solution outcomes by Wooldridge (2002), as the zero value for the dependent variable can be viewed as a corner solution outcome from an optimization problem faced by an economic agent. This model is also considered by Honoré (1993) and Honoré and Hu (2001). Another type of the dynamic censored regression model, as considered in Hu (2002), is the dynamic Tobit model with lagged latent dependent variables, which is referred to as the censored regression model with data censoring or coding by Wooldridge (2002).

Second, some important quantities of economic interest such as the average partial effects and average transition probabilities can be readily obtained as a by-product of the MCMC run. Third, the Bayesian estimators are consistent and efficient. It is also worth noting that our approach can be readily extended to other dynamic nonlinear panel data models including binary choice models, ordered response models, censored regression models with lagged latent dependent variables, and Poisson regression models.

As an application of our approach, we study female labor supply using NLSY79. We estimate a (reduced-form) labor supply model that is of the Tobit form because a significant proportion of women in the data did not work from time to time. We also control for the state dependence and unobserved heterogeneity, and estimate the average partial effects of the key explanatory variables, and the average transition probabilities between working and not working. It is worth noting that while panel data provide a unique opportunity for empirical researchers to address important economic questions that cannot be addressed using cross-sectional or time series data sets, dynamic Tobit panel data models have been rarely used in applications due to the complications mentioned previously. The application in this paper, while demonstrating the usefulness of our approach, also has some findings that can be of interest to labor economists.

This paper is organized as follows. Section 2 lays out the model of interest. Section 3 is devoted to the estimation method, which is computationally efficient and utilizes modern MCMC techniques. In Section 4, we show how the average partial effects of the covariates on the dependent variable and the average transition probabilities of different states are calculated. Section 5 presents Monte Carlo results that demonstrate the usefulness and feasibility of our approach. In Section 6 we apply our approach to study the intertemporal labor supply of a panel of young women. Section 7 concludes.

## 2 The Model

Let  $y_{it}$  be the censored response variable of interest, where the indices  $i$  and  $t$  ( $i = 1, \dots, n$ ,  $t = 1, \dots, T$ ) refer to individual  $i$  and time period  $t$ , respectively.<sup>5</sup> We consider a dynamic Tobit panel data model where  $y_{it}$  depends parametrically on the covariate vector  $\mathbf{z}_{it}$ , the vector of lags of the dependent variable  $\mathbf{y}_{it-1}$  and the unobserved individual heterogeneity  $c_i$  in the form

$$y_{it} = \max\{0, \mathbf{z}_{it}\gamma + \mathbf{g}(\mathbf{y}_{it-1})\rho + c_i + u_{it}\}, \quad (1)$$

where  $\gamma$  is a vector of coefficients for the explanatory variables and  $\rho$  is a vector of lag coefficients, and  $u_{it}$  is a sequence of i.i.d. random variables distributed as Normal  $(0, \sigma_u^2)$ . As in Wooldridge (2003), the function  $\mathbf{g}(\cdot)$  allows the lagged censored dependent variable to appear in various ways. In this model,  $\mathbf{z}_{it}$  is strictly exogenous in the sense that conditional on the current  $\mathbf{z}_{it}$ ,  $\mathbf{y}_{it-1}$ , and  $c_i$ , the past and future  $\mathbf{z}_i$ 's do not affect the distribution of  $y_{it}$ . This rules out the dynamic feedback from past and future realizations of  $\mathbf{z}$  to the current realizations of the dependent variable. Therefore, the model is dynamic only because of the lagged dependent variable, not through the serial correlation of the error terms  $u_{it}$ .<sup>6</sup> The assumption that  $u_{it}$  is normal produces the dynamic Tobit panel data model with unobserved individual heterogeneity. Our approach can also be readily extended to the case where the distribution of  $u_{it}$  is a  $t$  distribution (Albert and Chib (1993)) or some mixture of normals (Geweke and Keane (2001)).

To complete the specification of the model, we need to make some assumptions regarding the relationship between the unobserved individual heterogeneity and the initial conditions. We make the following conditional mean dependence assumption of the unobserved hetero-

---

<sup>5</sup>The model can be easily extended to the unbalanced panel model. For ease of exposition, we focus on the balanced panel data model.

<sup>6</sup>Again, the approach in this paper can be easily extended to accommodate the case where  $u_{it}$  is serially correlated as in Chib and Jeliazkov (2003) for the analysis of dynamic binary choice panel data models. To focus on the main issue, we choose the i.i.d. specification of  $u_{it}$ .

geneity  $c_i$  on the initial conditions and observed strictly exogenous variables

$$E[c_i | \mathbf{y}_{i0}, \mathbf{z}_i] = c + \mathbf{g}_0(\mathbf{y}_{i0})\delta_1 + h(\mathbf{z}_i)\delta_2, \quad (2)$$

where  $c$  is a constant,  $\mathbf{y}_{i0}$  is the vector of initial values of the dependent variable  $y_i$ ,  $\mathbf{g}_0(\cdot)$  is a function that allows  $\mathbf{y}_{i0}$  to appear in a variety of ways, and  $h(\mathbf{z}_i)$  is a function of a set of explanatory variables that only vary over different individuals but are time-invariant. It can be the row vector of all (nonredundant) explanatory variables in all time periods. That is,  $h(\mathbf{z}_i) = (z_{i1}, z_{i2}, \dots, z_{iT})$  and each  $z_{it}$  can be multidimensional as in Wooldridge (2003). Alternatively it can be  $h(\mathbf{z}_i) = \bar{\mathbf{z}}_i$ , where  $\bar{\mathbf{z}}_i$  is the average of  $\mathbf{z}_{it}$  over all the time periods as in Chib and Jeliazkov (2003).<sup>7</sup>

From (2), we have

$$c_i = \mathbf{g}_0(\mathbf{y}_{i0})\delta_1 + h(\mathbf{z}_i)\delta_2 + \alpha_i, \quad (3)$$

where  $\alpha_i$  is assumed to be independent of  $\mathbf{y}_{i0}$  and  $\mathbf{z}_i$ , and contains  $c$  in (2). This specification of the unobserved individual heterogeneity captures its correlation with the initial observations of the dependent variable and the set of exogenous covariates. Therefore, it is in the same spirit as in Chamberlain (1980) and Wooldridge (2003). On the other hand and more importantly, we leave the distribution of  $\alpha_i$  unspecified. Instead, we will approximate its distribution using an infinite mixture of normals. This is justified because Ferguson (1983) notes that any probability density function can be approximately arbitrarily closely in the  $L_1$  norm by a countable mixture of normal densities

$$f(\cdot) = \sum_{j=1}^{\infty} p_j \phi(\cdot | \mu_j, \sigma_j^2), \quad (4)$$

where  $p_j \geq 0$ ,  $\sum_{j=1}^{\infty} p_j = 1$  and  $\phi(\cdot | \mu_j, \sigma_j^2)$  denotes the probability density function for a normal distribution with mean  $\mu_j$  and variance  $\sigma_j^2$ . Note that this is a different specification from

---

<sup>7</sup>For the identification purpose, those time-constant variables like race and gender cannot be in both  $\mathbf{z}_{it}$  and  $\bar{\mathbf{z}}_i$ .

the finite mixture of normals as used in Geweke and Keane (2001) since we do not have a prior number of components in the mixture of normals. Instead, we will use the Dirichlet process prior to carry out the Bayesian nonparametric density estimation and update the number of components in the infinite mixture of normals and the mean and variance for each component.

### 3 Estimation

#### 3.1 Fully Parametric Dynamic Tobit Panel Data Models

We first present the Bayesian estimation algorithm for the fully parametric dynamic Tobit panel data model used in Wooldridge (2003). This provides a benchmark for comparison with the semiparametric estimation method. The estimation algorithm is computationally fast and statistically efficient. As in Wooldridge (2003), this model is a slight modification of the above semiparametric model

$$\begin{aligned}
 y_{it} &= \max\{0, \mathbf{z}_{it}\gamma + \mathbf{g}(\mathbf{y}_{it-1})\rho + c_i + u_{it}\} \\
 c_i &= \mathbf{g}_0(\mathbf{y}_{i0})\delta_1 + h(\mathbf{z}_i)\delta_2 + \alpha_i \\
 \alpha_i &\sim \text{Normal}(\mu, \sigma^2).
 \end{aligned} \tag{5}$$

As can be seen from the above model, a special feature of the fully parametric model is that the unobserved individual heterogeneity term  $c_i$  is now assumed to be from a normal distribution.

For each individual  $i$ , we have the following conditional density for the dependent variables:

$$\begin{aligned}
 &f(y_{i1}, y_{i2}, \dots, y_{iT} | \mathbf{y}_{i0}, \mathbf{z}_{it}, c_i, \gamma, \rho) \\
 &= \prod_{t=1}^T \left\{ \left[ 1 - \Phi\left(\frac{\mathbf{z}_{it}\gamma + \mathbf{g}(\mathbf{y}_{it-1})\rho + c_i}{\sigma_u}\right) \right]^{1(y_{it}=0)} \left[ \frac{1}{\sigma_u} \phi\left(\frac{y_{it} - \mathbf{z}_{it}\gamma - \mathbf{g}(\mathbf{y}_{it-1})\rho - c_i}{\sigma_u}\right) \right]^{1(y_{it}>0)} \right\}. \tag{6}
 \end{aligned}$$



To implement a Gibbs sampler, we introduce the latent variable  $y_{it}^*$  for the dependent variable, and rewrite model (5) in the following form

$$\begin{aligned}
y_{it}^* &= \mathbf{z}_{it}\gamma + \mathbf{g}(\mathbf{y}_{it-1})\rho + c_i + u_{it} \\
y_{it} &= \mathbb{I}(y_{it}^* > 0)y_{it}^* \\
c_i &= \mathbf{g}_0(\mathbf{y}_{i0})\delta_1 + h(\mathbf{z}_i)\delta_2 + \alpha_i \\
\alpha_i &\sim \text{Normal}(\mu, \sigma^2).
\end{aligned} \tag{7}$$

As a result, likelihood (6) can be modified as follows to be conditioning on the latent variables  $y_{it}^*$  in addition to other conditioning variables included in (6)

$$\begin{aligned}
&f(y_{i1}, y_{i2}, \dots, y_{iT} | y_{i1}^*, y_{i2}^*, \dots, y_{iT}^*, \mathbf{y}_{i0}, \mathbf{z}_{it}, c_i, \gamma, \rho) = \\
&\prod_{t=1}^T \{1(y_{it} > 0)1(y_{it} = y_{it}^*) + 1(y_{it} = 0)1(y_{it}^* \leq 0)\} \\
&\times \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{1}{2\sigma_u^2} (y_{it}^* - \mathbf{z}_{it}\gamma - \mathbf{g}(\mathbf{y}_{it-1})\rho - c_i)^2\right)
\end{aligned} \tag{8}$$

Again, since we do not observe the latent variables  $y_{it}^*$  and  $c_i$  and integration over these variables will produce an analytically intractable likelihood, direct implementation of maximum likelihood estimation method or Bayesian MCMC would be difficult. Instead, we adopt the data augmentation approach suggested by Albert and Chib (1993), where the latent variables  $y_{it}^*$  and  $c_i$  are explicitly included in the MCMC iterations and are updated at each step. Another advantage of the data augmentation technique is that with the presence of  $y_{it}^*$  and  $c_i$ , updating the main parameters of interest,  $\gamma$  and  $\rho$ , becomes similar to the standard posterior updating for simple linear panel data models and therefore straightforward to implement.

Denote  $w_{it} = (\mathbf{z}_{it}, \mathbf{g}(\mathbf{y}_{it-1}))$ ,  $\beta = (\gamma', \rho)'$ ,  $x_i = (\mathbf{g}_0(\mathbf{y}_{i0}), h(\mathbf{z}_i), 1)$  and  $\delta = (\delta_1', \delta_2', \mu)'$ , we have the following algorithm.

**Algorithm 1** *MCMC for fully parametric dynamic Tobit panel data models*

1. Conditional on  $y_{it}$ ,  $w_{it}$ ,  $x_i$ ,  $\beta$ ,  $\delta$ ,  $\sigma^2$  and  $\sigma_u^2$  but marginalized over  $c_i$ ,  $y_{it}^*$  is updated from a normal distribution with mean  $w_{it}\beta + x_i\delta$  and variance  $\sigma^2 + \sigma_u^2$  with truncation at 0

from the left if the corresponding  $y_{it} = 0$ . If  $y_{it} > 0$ ,  $y_{it}^* = y_{it}$ .

2. Conditional on  $y_{it}^*$ ,  $c_i$  and  $w_{it}$ , update  $\sigma_u^2$  and  $\beta$  in one block. We use the improper flat prior for  $\beta$  and the independent gamma  $(\frac{N_1}{2}, \frac{R_1}{2})$  prior for  $1/\sigma_u^2$ , that is:  $1/\sigma_u^2 \propto (1/\sigma_u^2)^{\frac{N_1}{2}-1} e^{-R_1(1/\sigma_u^2)}$ . Therefore, the joint posterior distribution of  $1/\sigma_u^2$  and  $\beta$  conditional on data and other parameters is

$$\begin{aligned} & \text{posterior}(1/\sigma_u^2, \beta | y_{it}^*, c_i, w_{it}) \\ \propto & (1/\sigma_u^2)^{\frac{N_1}{2}-1} e^{-\frac{R_1}{2}(1/\sigma_u^2)} \prod_{i=1}^n \prod_{t=1}^T \frac{1}{\sqrt{\sigma_u^2}} \exp\left(-\frac{1}{2\sigma_u^2} (y_{it}^* - w_{it}\beta - c_i)^2\right). \end{aligned} \quad (9)$$

- a** To draw from this posterior, we draw  $1/\sigma_u^2$  marginalized over  $\beta$  first and then draw  $\beta | \sigma_u^2$ . Define

$$\hat{\beta} = \text{inv} \left( \sum_{i=1}^n \sum_{t=1}^T w'_{it} w_{it} \right) \times \left( \sum_{i=1}^n \sum_{t=1}^T w'_{it} (y_{it}^* - c_i) \right), \quad (10)$$

the posterior density of  $1/\sigma_u^2$  marginalized over  $\beta$  is

$$\begin{aligned} & \text{posterior}(1/\sigma_u^2 | y_{it}^*, c_i, w_{it}) \\ \propto & (1/\sigma_u^2)^{\frac{N_1+nT}{2}-1} \exp \left\{ -\frac{1}{\sigma_u^2} \times \frac{\left[ R_1 + \sum_{i=1}^n \sum_{t=1}^T (y_{it}^* - w_{it}\hat{\beta} - c_i)^2 \right]}{2} \right\}. \end{aligned} \quad (11)$$

That is, we draw  $1/\sigma_u^2$  from  $\text{gamma}(\frac{N_1+nT}{2}, \frac{R_1 + \sum_{i=1}^n \sum_{t=1}^T (y_{it}^* - w_{it}\hat{\beta} - c_i)^2}{2})$ .

- b** Second, we update  $\beta$  from  $\text{posterior}(\beta | 1/\sigma_u^2, y_{it}^*, c_i, w_{it})$ , which is

$$\text{Normal} \left( \hat{\beta}, \text{inv} \left( \frac{1}{\sigma_u^2} \sum_{i=1}^n \sum_{t=1}^T w'_{it} w_{it} \right) \right). \quad (12)$$

3. Conditional on  $y_{it}^*$ ,  $w_{it}$ ,  $x_i$ ,  $\beta$ ,  $\delta$ ,  $\sigma^2$ ,  $\sigma_u^2$ , update  $c_i$ . The posterior is

$$\prod_{t=1}^T \exp\left(-\frac{1}{2\sigma_u^2} (y_{it}^* - w_{it}\beta - c_i)^2\right) \exp\left(-\frac{1}{2\sigma^2} (c_i - x_i\delta)^2\right). \quad (13)$$

Therefore, we update  $c_i$  by drawing from  $\text{Normal}\left(c_i^*, \text{inv}\left(\frac{T}{\sigma_u^2} + \frac{1}{\sigma^2}\right)\right)$  where  $c_i^* = \text{inv}\left(\frac{T}{\sigma_u^2} + \frac{1}{\sigma^2}\right) \times \left(\frac{1}{\sigma_u^2} \sum_{t=1}^T (y_{it}^* - w_{it}\beta) + \frac{1}{\sigma^2} x_i\delta\right)$ .

4. Conditional on  $c_i$  and  $x_i$ , update  $\sigma^2$  and  $\delta$  in one block. We use the improper flat prior for  $\delta$  and independent gamma  $(\frac{N_2}{2}, \frac{R_2}{2})$  prior for  $1/\sigma^2$ , that is:  $1/\sigma^2 \propto (1/\sigma^2)^{\frac{N_2}{2}-1} e^{-R_2(1/\sigma^2)}$ . Therefore, the joint posterior distribution of  $1/\sigma^2$  and  $\delta$  is

$$\begin{aligned} & \text{posterior}(1/\sigma^2, \delta | x_i, c_i) \\ \propto & (1/\sigma^2)^{\frac{N_2}{2}-1} e^{-\frac{R_2}{2}(1/\sigma^2)} \prod_{i=1}^n \frac{1}{\sqrt{\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (c_i - x_i\delta)^2\right). \end{aligned} \quad (14)$$

**a** To draw from this posterior, we draw  $\text{posterior}(1/\sigma^2 | c_i, x_i)$  first and then draw  $\text{posterior}(\delta | 1/\sigma^2, c_i, x_i)$ . Define

$$\widehat{\delta} = \text{inv}\left(\sum_{i=1}^n x_i' x_i\right) \times \left(\sum_{i=1}^n x_i' c_i\right), \quad (15)$$

the posterior density of  $1/\sigma^2$  marginalized over  $\delta$  is

$$\begin{aligned} & \text{posterior}(1/\sigma^2 | c_i, x_i) \\ \propto & (1/\sigma^2)^{\frac{N_2+n}{2}-1} \exp\left[-\frac{1}{\sigma^2} \times \frac{R_2 + \sum_{i=1}^n (c_i - x_i\widehat{\delta})^2}{2}\right]. \end{aligned} \quad (16)$$

That is, we draw  $1/\sigma^2$  from  $\text{gamma}\left(\frac{N_2+n}{2}, \frac{R_2 + \sum_{i=1}^n (c_i - x_i\widehat{\delta})^2}{2}\right)$ .

**b** Second, we update  $\delta$  from

$$posterior(\delta|1/\sigma^2, c_i, x_i) = Normal\left(\widehat{\delta}, inv\left(\frac{1}{\sigma^2}\sum_{i=1}^n x_i'x_i\right)\right). \quad (17)$$

## 3.2 Semiparametric Dynamic Tobit Panel Data Models

This subsection considers the case where the normal distribution assumption for the error term  $\alpha_i$  and hence for the unobserved individual heterogeneity term  $c_i$  is relaxed. Instead of imposing any distributional assumptions for  $\alpha_i$ , we approximate it using an infinite mixture of normals. As stated in Section 2, with this approach, each  $\alpha_i$  is assumed to be distributed as Normal  $(\mu_i, \sigma_i^2)$ . Our semiparametric Bayesian estimation consists of two main parts. At each iteration, in the first part, using the augmented latent variables  $c_i$  and the current value of parameters, we can recover the error term  $\alpha_i$  in the unobserved individual heterogeneity through the relationship  $\alpha_i = c_i - \tilde{\mathbf{x}}_i\tilde{\delta}$ , where  $\tilde{\mathbf{x}}_i = (\mathbf{g}_0(\mathbf{y}_{i0}), h(\mathbf{z}_i))$  and  $\tilde{\delta} = (\delta'_1, \delta'_2)'$ . After recovering the error terms, we can use a Bayesian approach to estimate their densities with a Dirichlet process prior for the unknown densities. We update the number of components (denoted by  $m_c$ , say) in approximating mixture of normals and the mean and variance of the normal denoted by  $\mu_i$  and  $\sigma_i^2$  for each  $i$ . In the second part of each iteration, we update the model parameters and values for the latent variables.

### 3.2.1 Bayesian Nonparametric Estimation

Our objective here for the nonparametric Bayesian estimation is to update the number of components in approximating mixture of normals and the mean and variance for each component given a set of observations  $\alpha_i$ . Following Escobar and West (1995), Hirano (2002) and Hasegawa and Kozumi (2003), we denote

$$q(\alpha_i|\theta_i) = q(\alpha_i|\mu_i, \sigma_i^2) = \phi(\alpha_i|\mu_i, \sigma_i^2) \quad (18)$$

where  $\theta_i = (\mu_i, \sigma_i^2)$  is a sample from some unknown distribution  $P$ . Conditional on  $P$ , the density for  $\alpha_i$  is

$$s(\alpha_i|P) = \int q(\alpha_i|\theta_i)dP. \quad (19)$$

This gives us a nonparametric class of distributions.

In order to carry out a Bayesian analysis, we assume that the distribution  $P$  has the Dirichlet process prior introduced by Ferguson (1973, 1974). The Dirichlet process is a probability measure on the space of all distributions and can be used as a prior distribution on the space of probability distributions. Let  $\Theta$  be a space and  $\Gamma$  a  $\sigma$ -field of subsets of  $\Theta$ , and let  $\tau P_0$  be a finite non-null measure on  $(\Theta, \Gamma)$  where  $P_0$  is a proper base probability distribution and  $\tau$  is a precision parameter. Then, a stochastic process  $P$  is a Dirichlet process if, for any given partition,  $A_1, \dots, A_q$  of the parameter space  $\Theta$ , the random vector  $(P(A_1), \dots, P(A_q))$  has a Dirichlet distribution with a parameter vector  $(\tau P_0(A_1), \dots, \tau P_0(A_q))$ . We denote  $DP(\tau P_0)$  for a Dirichlet process with base measure  $\tau P_0$  for the rest of the paper.

The Dirichlet process is a probability distribution on the space of probability distributions on  $(\Theta, \Gamma)$  and selects a discrete distribution with probability one. The discreteness of the probability distribution selected by the Dirichlet process seems to be unsuitable for modelling smooth densities. However, it can be related to the infinite normal mixture model as noted by Ferguson (1983), Lo (1984) and further explained by Ghosal, Ghosh and Ramamoorthi (1999).

To get the posterior distributions, we need to integrate over the parameter space. Escobar (1994) shows that with the  $\theta_i$ 's as the latent variables, integrating  $P$  over its prior distribution gives the sequence of  $\theta_i$ ,  $i = 1, \dots, n$  as follows

$$\theta_i|\theta_1, \theta_2, \dots, \theta_{i-1} = \begin{cases} = \theta_j & \text{with probability } \frac{1}{\alpha+i-1} \\ \sim P_0 & \text{with probability } \frac{\alpha}{\alpha+i-1} \end{cases} \quad (20)$$

where  $\theta_1 \sim P_0$ . As a result,  $\alpha_i$ 's are partitioned into  $m_c$  groups such that all  $\alpha_i$  in the same group have the same  $\theta_i$  while those in different groups differ. These  $m_c$  distinct values  $\theta_i$  are a sample from  $P_0$ .

To complete the model specification, we choose the base prior distribution  $P_0$  to be the conjugate normal/inverse-gamma distribution

$$dP_0(\mu_i, \sigma_i^2) \propto \text{Normal}(\mu_0, \tau_0 \sigma_i^2) IG\left(\frac{N_0}{2}, \frac{R_0}{2}\right) \quad (21)$$

where  $IG(a, b)$  denotes the inverse-gamma distribution with parameters  $a$  and  $b$ . Finally, as in Escobar and West (1995),  $\tau$  is assumed to follow a gamma distribution

$$\tau \sim \text{gamma}(d_1, d_2). \quad (22)$$

Let  $\xi_j$ ,  $j = 1, 2, \dots, m_c$  denote the  $m_c$  distinct values of  $\theta_i$  and

$$\theta^{(i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n) \quad (23)$$

be the set of values of  $\theta$  for units other than  $i$ . Furthermore, let superscript  $(i)$  refer to variables defined on all units other than  $i$  and hence  $\xi^{(i)} = (\xi_1^{(i)}, \dots, \xi_{m_c^{(i)}}^{(i)})$  are the distinct values among  $(\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n)$ . Escobar (1994) shows that using the Dirichlet process prior leads to a useful set of conditional distributions. Specifically,

$$P|\theta^{(i)} \sim DP\left(\tau P_0 + \sum_{j=1}^{m_c^{(i)}} n_j^{(i)} \delta(\xi_j^{(i)})\right) \quad (24)$$

where  $n_j^{(i)}$  is the number of  $\theta_i$  taking the value of  $\xi_j^{(i)}$  and  $\delta(\xi_j^{(i)})$  represents unit point mass at  $\theta_i = \xi_j^{(i)}$ . Therefore,

$$\theta_i|\theta^{(i)} \sim E(P|\theta^{(i)}) \sim \frac{\alpha}{\alpha + n - 1} P_0 + \frac{1}{\alpha + n - 1} \sum_{j=1}^{m_c^{(i)}} n_j^{(i)} \delta(\xi_j^{(i)}). \quad (25)$$

One way of sampling  $\theta_i$  and determines the number of components in approximating mixture of normals  $m_c$  is to use the density given in equation (25). But sampling  $\theta_i$  in this model can be eased by introducing the configuration vector  $S = \{S_1, \dots, S_n\}$ . That is,  $S_i = j$

if and only if  $\theta_i = \xi_j$ . With this setting sampling  $\theta_i$  is equivalent to sampling  $S$  and  $\xi_j$ ,  $j = 1, 2, \dots, m_c$ . West, Müller and Escobar (1994) propose an efficient algorithm as follows:

1. Sampling  $S_i$ ,  $i = 1, 2, \dots, n$  from the conditional distribution

$$q(S_i = j | \xi^{(i)}, P_0) \propto \begin{cases} \tau q_t(\alpha_i | \mu_0, (1 + \tau_0)R_0/n_0, n_0) & \text{if } j = 0 \\ n_j^{(i)} \phi(\alpha_i | \xi_j^{(i)}) & \text{if } j > 0 \end{cases} \quad (26)$$

where  $q_t(\alpha | \mu, \sigma^2, \nu)$  is the density of the  $t$ -distribution with mean  $\mu$ , scale factor  $\sigma^2$  and  $\nu$  degrees of freedom and  $\phi(\alpha | \xi_j^{(i)})$  denotes the normal probability density characterized by  $\xi_j^{(i)}$ .

2. Sampling  $\xi_j$ ,  $j = 1, 2, \dots, m_c$  from the conditional distribution

$$\begin{aligned} q(\xi_j | \xi^{(j)}, S, P_0) &\propto \prod_{\{\alpha_i: S_i=j\}} \phi(\alpha_i | \xi_j) dP_0 \\ &\propto \text{Normal}(\mu_1, \tau_1 \sigma_j^2) \text{IG}\left(\frac{N_3}{2}, \frac{R_3}{2}\right) \end{aligned} \quad (27)$$

where

$$\begin{aligned} \mu_1 &= \frac{\tau_0 \sum_{\{\alpha_i: S_i=j\}} \alpha_i + \mu_0}{\tau_0 n_j + 1}, \tau_1 = \frac{\tau_0}{\tau_0 n_j + 1}, N_3 = N_0 + n_j, \\ R_3 &= R_0 + \frac{n_j \left( \frac{1}{n_j} \sum_{\{\alpha_i: S_i=j\}} \alpha_i - \mu_0 \right)^2}{\tau_0 n_j + 1} + \sum_{\{\alpha_i: S_i=j\}} \left( \alpha_i - \frac{1}{n_j} \sum_{\{\alpha_i: S_i=j\}} \alpha_i \right)^2 \end{aligned}$$

and  $n_j$  is the number of observations such that  $S_i = j$ .

Finally, Escobar and West (1995) show that with a beta distributed variable  $\eta \sim \text{Beta}(\tau + 1, n)$ , the full conditional distribution of  $\tau$  is given by

$$\tau \sim z \times \text{gamma}(d_1 + m_c, d_2 - \log \eta) + (1 - z) \times \text{gamma}(d_1 + m_c - 1, d_2 - \log \eta) \quad (28)$$

where  $\frac{z}{1-z} = \frac{d_1+m_c-1}{n(d_2-\log \eta)}$ .

Thus, with a set of  $\alpha_i$ , we updated  $\theta_i$  for each  $i$ , the number of components in the countable infinite mixture of normals  $m_c$  and  $\tau$ , the precision parameter in the base measure.

### 3.2.2 The Algorithm

Incorporating other parts of the model, we have the following algorithm for the semiparametric dynamic Tobit data models.

**Algorithm 2** *MCMC for semiparametric dynamic Tobit panel data models*

1. Conditional on  $y_{it}$ ,  $w_{it}$ ,  $\tilde{\mathbf{x}}_i$ ,  $\beta$ ,  $\tilde{\delta}$ ,  $\mu_i$ ,  $\sigma_i^2$  and  $\sigma_u^2$  but marginalized over  $c_i$ ,  $y_{it}^*$  is updated from a normal distribution with mean  $w_{it}\beta + \tilde{\mathbf{x}}_i\tilde{\delta} + \mu_i$  and variance  $\sigma_i^2 + \sigma_u^2$  with truncation at 0 from the left if the corresponding  $y_{it} = 0$ . If  $y_{it} > 0$ ,  $y_{it}^* = y_{it}$ .
2. Conditional on  $y_{it}^*$ ,  $c_i$  and  $w_{it}$ , update  $\sigma_u^2$  and  $\beta$  in one block. We use the improper flat prior for  $\beta$  and the independent gamma  $(\frac{N_1}{2}, \frac{R_1}{2})$  prior for  $1/\sigma_u^2$ , that is:  $1/\sigma_u^2 \propto (1/\sigma_u^2)^{\frac{N_1}{2}-1} e^{-R_1(1/\sigma_u^2)}$ . Therefore, the joint posterior distribution of  $1/\sigma_u^2$  and  $\beta$  conditional on data and other parameters is

$$\begin{aligned} & \text{posterior}(1/\sigma_u^2, \beta | y_{it}^*, c_i, w_{it}) \\ & \propto (1/\sigma_u^2)^{\frac{N_1}{2}-1} e^{-\frac{R_1}{2}(1/\sigma_u^2)} \prod_{i=1}^n \prod_{t=1}^T \frac{1}{\sqrt{\sigma_u^2}} \exp\left(-\frac{1}{2\sigma_u^2} (y_{it}^* - w_{it}\beta - c_i)^2\right). \end{aligned} \quad (29)$$

- a** To draw from this posterior, we draw  $1/\sigma_u^2$  marginalized over  $\beta$  first and then draw  $\beta | \sigma_u^2$ . Define

$$\hat{\beta} = inv \left( \sum_{i=1}^n \sum_{t=1}^T w'_{it} w_{it} \right) \times \left( \sum_{i=1}^n \sum_{t=1}^T w'_{it} (y_{it}^* - c_i) \right), \quad (30)$$



the posterior density of  $1/\sigma_u^2$  marginalized over  $\beta$  is

$$\begin{aligned} & \text{posterior}(1/\sigma_u^2 | y_{it}^*, c_i, w_{it}) \\ & \propto (1/\sigma_u^2)^{\frac{N_1+nT}{2}-1} \exp \left\{ -\frac{1}{\sigma_u^2} \times \frac{\left[ R_1 + \sum_{i=1}^n \sum_{t=1}^T (y_{it}^* - w_{it}\hat{\beta} - c_i)^2 \right]}{2} \right\}. \end{aligned} \quad (31)$$

That is, we draw  $1/\sigma_u^2$  from  $\text{gamma}(\frac{N_1+nT}{2}, \frac{R_1 + \sum_{i=1}^n \sum_{t=1}^T (y_{it}^* - w_{it}\hat{\beta} - c_i)^2}{2})$ .

**b** Second, we update  $\beta$  from  $\text{posterior}(\beta | 1/\sigma_u^2, y_{it}^*, c_i, w_{it})$ , which is

$$\text{Normal} \left( \hat{\beta}, \text{inv} \left( \frac{1}{\sigma_u^2} \sum_{i=1}^n \sum_{t=1}^T w_{it}' w_{it} \right) \right). \quad (32)$$

3. Conditional on  $y_{it}^*, w_{it}, \tilde{\mathbf{x}}_i, \beta, \tilde{\delta}, \mu_i, \sigma_i^2, \sigma_u^2$ , update  $c_i$ . The posterior is

$$\prod_{t=1}^T \exp\left(-\frac{1}{2\sigma_u^2} (y_{it}^* - w_{it}\beta - c_i)^2\right) \exp\left(-\frac{1}{2\sigma_i^2} (c_i - \tilde{\mathbf{x}}_i\tilde{\delta} - \mu_i)^2\right). \quad (33)$$

Therefore, we update  $c_i$  by drawing from  $\text{Normal}(c_i^*, \text{inv}(\frac{T}{\sigma_u^2} + \frac{1}{\sigma_i^2}))$  where  $c_i^* = \text{inv}(\frac{T}{\sigma_u^2} + \frac{1}{\sigma_i^2}) \times (\frac{1}{\sigma_u^2} \sum_{t=1}^T (y_{it}^* - w_{it}\beta) + \frac{1}{\sigma_i^2} (x_i\delta + \mu_i))$ .

4. Conditional on  $c_i, \tilde{\mathbf{x}}_i, \mu_i$  and  $\sigma_i^2$  update  $\tilde{\delta}$ . We use the improper flat prior for  $\tilde{\delta}$ .

Therefore, the posterior distribution of  $\tilde{\delta}$  is

$$\begin{aligned} & \text{posterior}(\tilde{\delta} | \tilde{\mathbf{x}}_i, c_i, \mu_i, \sigma_i^2) \\ & \propto \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma_i^2} (c_i - \tilde{\mathbf{x}}_i\tilde{\delta} - \mu_i)^2\right). \end{aligned} \quad (34)$$

Define

$$\hat{\delta} = \text{inv} \left( \sum_{i=1}^n \frac{1}{\sigma_i^2} \tilde{\mathbf{x}}_i' \tilde{\mathbf{x}}_i \right) \times \left( \sum_{i=1}^n \frac{1}{\sigma_i^2} \tilde{\mathbf{x}}_i' (c_i - \mu_i) \right) \quad (35)$$

we update  $\tilde{\delta}$  by drawing from  $\text{Normal}(\hat{\delta}, \text{inv}(\sum_{i=1}^n \frac{\tilde{\mathbf{x}}_i' \tilde{\mathbf{x}}_i}{\sigma_i^2}))$ .

5. Finally, recover  $\alpha_i = c_i - \tilde{\mathbf{x}}_i \tilde{\delta}$  and update the number of components  $m_c$  in approximating mixture of normals and the mean and variance of the normal denoted by  $\mu_i$  and  $\sigma_i^2$  for each  $i$  using the Bayesian nonparametric density estimation method described in the previous subsection.

### 3.2.3 Predictive Densities

In the Bayesian framework, a useful and informative way to draw implications from the unknown densities of the error term  $\alpha_i$  is to study its predictive distribution. Parallel to equation (25), conditional on  $(\theta_1, \dots, \theta_n)$ , for the new unit  $i = n + 1$ , we have

$$\theta_{i+1} | (\theta_1, \dots, \theta_n) \sim \frac{\alpha}{\alpha + n} P_0 + \frac{1}{\alpha + n} \sum_{j=1}^{m_c} n_j \delta(\xi_j), \quad (36)$$

where  $n_j$  is the number of  $\theta_i$ 's taking the value  $\xi_j$ . Thus, the distribution of  $\alpha_{i+1}$  conditional on the data can be rewritten as

$$\begin{aligned} q(\alpha_{i+1} | \theta_1, \dots, \theta_n) &= \frac{\alpha}{\alpha + n} q_t(\alpha_{i+1} | \mu_0, (1 + \tau_0)R_0/n_0, n_0) \\ &+ \frac{1}{\alpha + n} \sum_{j=1}^{m_c} n_j \phi(\alpha_{i+1} | \xi_j). \end{aligned} \quad (37)$$

As a result, the predictive distribution for the error term  $\alpha_i$  can be obtained as

$$q(\alpha_{i+1} | data) = \int q(\alpha_{i+1} | \theta_1, \dots, \theta_n) \pi(\theta_1, \dots, \theta_n | data) d(\theta_1, \dots, \theta_n). \quad (38)$$

Since the Gibbs sampler provides draws for  $\theta_i$ 's, we can use the Monte Carlo method to integrate out  $\theta_i$ 's to estimate  $q(\alpha_{i+1} | data)$  as

$$\hat{q}(\alpha_{i+1} | data) = \frac{1}{M} \sum_{i=1}^M q(\alpha_{i+1} | \theta_1^{(i)}, \dots, \theta_n^{(i)}), \quad (39)$$

where  $(\theta_1^{(i)}, \dots, \theta_n^{(i)})$  is a simulated sample of  $(\theta_1, \dots, \theta_n)$ .

## 4 Average Partial Effects and Transition Probabilities

### 4.1 Average Partial Effects

For nonlinear models, besides the estimation of parameters, obtaining the average partial effects is necessary to assess the effects of any change in the covariates on the dependent variable. This is important since it can be used to evaluate policies. Easy calculation of the average partial effect is also one main advantage of the semiparametric Bayesian method we propose in this paper. In the classical framework, though the methods adopted by Honoré (1993) do not require any distributional assumption for the unobserved heterogeneity, the partial effects are not identified using his approach. On the other hand, the method proposed by Wooldridge (2003) depends on the assumption of some special distributions for the unobserved individual heterogeneity. If the distribution is assumed to be something other than the normal, parameter estimation and calculation of the average partial effects will also need the implementation of simulation-based methods because of the integration involved. The semiparametric Bayesian method proposed in this paper relaxes the distributional assumption for the unobserved individual heterogeneity term, but at the same time, the calculation of the average partial effects becomes a by-product of the MCMC estimation procedure and hence does not add in any additional computation burden.

Let  $m_{it}^j$  denote some partial effects of the  $j$ th covariate  $w_{it}^j$ . For the dynamic Tobit panel data model, if  $w_{it}^j$  is continuous and there is no interaction terms involved, then the partial effect of the covariate on the expected value of the dependent variable is:

$$m_{it}^j = \frac{\partial E(y_{it}|w_{it}, c_i, \beta, \sigma_u)}{\partial w_{it}^j} = \Phi\left(\frac{w_{it}\beta + c_i}{\sigma_u}\right)\beta^j. \quad (40)$$

where  $\Phi(\cdot)$  denotes the cumulative standard normal density and  $\phi(\cdot)$  denotes the standard normal density. If  $w_{it}^j$  is discrete,  $m_{it}^j$  equals the difference between the values  $E(y_{it}|w_{it}, c_i, \beta, \sigma_u)$

takes when  $w_{it}^j = 1$  and when  $w_{it}^j = 0$ , respectively, where

$$E(y_{it}|w_{it}, c_i, \beta, \sigma_u) = \Phi\left(\frac{w_{it}\beta + c_i}{\sigma_u}\right)(w_{it}\beta + c_i) + \sigma_u\phi\left(\frac{w_{it}\beta + c_i}{\sigma_u}\right). \quad (41)$$

If the covariate  $w_{it}^j$  involves in some interaction terms, then the partial effects become more complex. However, they share a similar feature to these partial effects shown above in that they also depend on the unobserved heterogeneity.

The presence of the unobserved individual heterogeneity makes the calculation of the partial effects difficult since it usually does not have a clear measurement unit and is usually not observed. Therefore, it is usually more meaningful and useful to obtain the so-called average partial effect, which is the partial effect averaged across the population distribution of the unobserved heterogeneity. In other words, the unobserved heterogeneity needs to be integrated out. In the classical framework, if the integration does not have a closed-form, then we need to use simulation-based methods to calculate these effects. In the Bayesian framework, this calculation comes as a by-product during the estimation of the model parameters. More specifically, we can obtain summaries of the average partial effects conditional on the observed data, but marginalized over all unknowns including the model parameters and the unobserved individual heterogeneity.<sup>8</sup> To fix the ideas, by definition, the posterior density of  $m_{it}^j$  conditional on the observed data, but marginalized over all the unknowns, is

$$\pi(m_{it}^j|\text{data}) = \int \pi(m_{it}^j|\text{data}, c_i, \beta, \sigma_u) d\pi(c_i, \beta, \sigma_u|\text{data}). \quad (42)$$

A sample of  $m_{it}^j$  can be produced by the method of composition using the draws of  $c_i, \beta, \sigma_u$  from steps 2 and 3 in the algorithm described in Section 3.2.2. Given a posterior sample of  $m_{it}^j$  from  $\pi(m_{it}^j|\text{data})$ , which we denote by  $\{m_{it}^{j(g)}\}$ , the unit ( $i$ th observation in  $t$  period) mean partial effect, when  $w_{it}^j$  is continuous and there is no interaction terms for this covariate, can

---

<sup>8</sup>Chib and Hamilton (2002) use this approach to obtain the average treatment effects.

be estimated as

$$\overline{m_{it}^j} \approx G^{-1} \sum_{g=1}^G m_{it}^{j(g)} \text{ where } m_{it}^{j(g)} = \Phi\left(\frac{w_{it}\beta^{(g)} + c_i^{(g)}}{\sigma_u^{(g)}}\right)\beta^{j(g)} \quad (43)$$

At a more aggregate level, the average partial effect for a randomly selected observation from the population may be defined as

$$m^j = \frac{\sum_{i=1}^n \sum_{t=1}^T m_{it}^j}{nT} \quad (44)$$

whose posterior distribution is again available from the posterior sample on  $m_{it}^j$ .

## 4.2 Average Transition Probabilities

Another quantity of economic interest is the transition probability to and from different states. For example, to study the female labor supply, policy makers may want to know what is the probability of working in the next period conditional on not working in this period, or what is the probability of not working in the next period conditional on working in this period, as well as the effects of various covariates on these probabilities. It is worth noting that a panel data set can provide a researcher with a unique opportunity to assess the transition probabilities, as it contains sequential observations over time for the same individual.

Suppose we have a dynamic Tobit model with only one lag and denote the state of working as state 1 and not working as state 0. Then the probability for individual  $i$  to transfer from state 0 to state 1 in period  $t$  is

$$\begin{aligned} p_{it}^{01} &= pr(y_{it} > 0 | y_{it-1} = 0, \mathbf{z}_{it}, c_i) \\ &= \Phi\left(\frac{\mathbf{z}_{it}\gamma + c_i}{\sigma_u}\right) \end{aligned} \quad (45)$$

and the probability for individual  $i$  to stay at state 0 is  $p_{it}^{00} = 1 - p_{it}^{01}$ . The probability for

individual  $i$  to quit working in period  $t$  conditional on working  $\bar{y}_{it-1}$  hours in period  $t - 1$  is

$$p_{it}^{10} = pr(y_{it} = 0 | y_{it-1} = \bar{y}_{it-1}, \mathbf{z}_{it}, c_i) = 1 - \Phi\left(\frac{\bar{y}_{it-1}\rho + \mathbf{z}_{it}\gamma + c_i}{\sigma_u}\right). \quad (46)$$

Again, these probabilities depend on the unobserved heterogeneity. Thus, the average transition probabilities can be computed in the same way as for the average partial effects during the MCMC model estimations.

## 5 Simulation Studies

We perform two simulation studies. The first one is the case where the error term in the unobserved individual heterogeneity is normal. In this case, we find that the fully parametric and the semiparametric algorithms perform equally well. In the second experiment, we use an extremely non-normal distribution for the unobserved effect term. In this case, the fully parametric algorithm, which assumes normality of this term, becomes inconsistent, but the semiparametric algorithm is robust to this specification and still performs well.

We also demonstrate the good performance of our MCMC chains by plotting the autocorrelation functions and histograms for the sampled parameters. All the chains are of length of 5,000 draws following burn-ins of 1,000 draws.

The first experiment is designed as the following

$$\begin{aligned} y_{it}^* &= z_{it}\gamma + \rho y_{it-1} + c_i + u_{it} \\ c_i &= y_{i0}\delta_1 + \bar{\mathbf{z}}_i\delta_2 + \alpha_i \end{aligned} \quad (47)$$

Set  $N = 1000$ ,  $T = 5$  (this is roughly the sample size in our data).  $z_{it}$  (single dimension) and  $u_{it}$  are from independent standard normals.  $\bar{\mathbf{z}}_i$  is the simple average of  $z_{it}$  across the time periods.  $\alpha_i$  is from normal with mean 0 and variance 1.  $y_{i0}$  is also from standard normal but censored from left at 0. Parameters are set as the following:  $\gamma = 1$ ,  $\rho = 0.6$ ,  $\delta_1 = 0.3$ ,  $\delta_2 = 0.2$ .

Table 1 summarizes the results using the fully parametric approach. Figures 1(a) and 1(b) present the time series of the draws, histogram of the draws and the autocorrelation functions. Table 2 reports the results using the semiparametric Bayesian estimation method and Figures 2(a) and 2(b) present the MCMC convergence diagnostics. Overall, from these tables and graphs, we can see when the error term in the unobserved individual heterogeneity is normal, both the fully parametric Bayesian algorithm which assumes normality for the error term and the semiparametric Bayesian algorithm perform well. Also, the MCMC chains converge quickly and mix well.

The second experiment is designed in the same way as the first experiment except that now we generate  $\alpha_i$  according to a two-component mixture of gamma distributions. 30% of the error term comes from the gamma (1, 1/10) and the other 70% of the error term comes from  $-5 + \text{gamma}(1, 1/2)$ . This mixture makes the error term  $\alpha_i$  highly nonnormal. In this case, as we see from the results of Monte Carlo simulations in Table 3, the fully parametric Bayesian updating algorithm which assumes normality significantly biases the parameters of interest downwards by 10 to 15 percent. On the other hand, as seen from Table 4, the semiparametric Bayesian algorithm is robust to the specification of the distribution of the error terms. Again, from the output of Figures 4(a) and 4(b), we can see that this semiparametric Bayesian MCMC algorithm converges to the true value quickly and the chain mixes well. Using a Pentium® 4 2.4 GB processor, the computational time for the fully parametric algorithm is 1.05 seconds per 10 draws and 5.18 seconds per 10 draws for the semiparametric algorithm. Therefore, the semiparametric Bayesian method is computationally efficient.

## 6 Application: Intertemporal Female Labor Supply

In this section, we illustrate the proposed method by estimating a reduced-form intertemporal female labor supply model. The data set consists of a panel of 1,115 young women over 7 years (1987-1993). This data set is from the National Logitudinal Survey of Youth 1979 (NLSY79), which can be obtained from the U.S. Department of Labor. The NLSY79 is a

nationally representative sample of 12,686 young men and women who were 14 to 22 years of age when first surveyed in 1979. The survey runs every year from 1979 to 1994 and every two years thereafter and collects information on subjects' labor market performance, environmental variables, human capital and other socioeconomic variables. It has been widely used by social scientists, especially labor economists, during the past decade. Examples include Farber (1994) for worker mobility, Kane and Rouse (1996) for returns of schooling, Olsen (1994) for fertility, among others. We use the number of average working hours per week, which is defined as the total number of working hours in a given year divided by 52, as the measure of labor supply and the dependent variable, together with a set of 15 covariates, which are presented and explained in Table 5. Among these variables, only the race variables, Black and Hispanic, are time-invariant. Covariates similar to those used in Table 5 are common in the empirical labor supply literature. Examples include Chib and Jeliazkov (2003) for study of married women labor force participation, Shaw (1994) for research on the persistence of female labor supply, Nakamura and Nakamura (1994) for examination of the effects of children and recent work experience on female labor supply.

A simple (reduced-form) dynamic model of female labor supply is

$$\begin{aligned}
y_{it} &= \max\{0, \mathbf{z}_{it}\gamma + y_{it-1}\rho + c_i + u_{it}\} \\
c_i &= y_{i0}\delta_1 + \bar{\mathbf{z}}_i\delta_2 + \alpha_i \\
u_{it} &\sim \text{Normal}(0, \sigma_u^2),
\end{aligned} \tag{48}$$

where  $t = 0$  corresponds to the initial year 1987,  $y_{it} = \text{Hours}_{it}$ ,  $\mathbf{z}_{it} = (\text{Age}_{it}, \text{Educ}_{it}, \text{Income}_{it}, \text{Married}_{it}, \text{Northcentral}_{it}, \text{Northeast}_{it}, \text{South}_{it}, \text{Old}_{it}, \text{Young}_{it}, \text{School}_{it}, \text{SMSA}_{it}, \text{Unemp}_{it}, \text{Urban}_{it}, (\text{Age}_{it})^2, \text{Age}_{it} * \text{Educ}_{it}, (\text{Age}_{it})^2 * \text{Educ}_{it}, \text{Hours}_{it-1})$  and  $\bar{\mathbf{z}}_i$  is the average of  $\mathbf{z}_{it}$  over these years. Note that a Tobit model is appropriate because over seven years in our data, the proportions of non-working women are 15.5%, 15.1%, 14.3%, 15.8%, 15.8%, 17.1%, and 16.1%, respectively. Here we leave the distribution of  $\alpha_i$  unspecified, and will use the semi-parametric Bayesian estimation procedure we propose to estimate model (48). The resulting



parameter estimates are reported in Table 6, Table 7 and the estimates for the average partial effects are reported in Table 8. The chain is of length of 5,000 draws following burn-ins of 1,000 draws.

## 6.1 Evidence of Nonnormality

Figure 5 shows the estimated predictive density for the error term  $\alpha_i$ , together with the kernel estimate for the density of the dependent variable  $y_{it}$ . As we can see, the predictive density under the semiparametric model has multiple modes and hence highly non-normal. Interestingly, the estimated density of the dependent variable also demonstrates two modes, which correspond to the two biggest modes of the estimated predictive density for the error term  $\alpha_i$ . This indicates that the modes of density of the dependent variable are mainly driven by the modes of the density of the error term in the unobserved heterogeneity term; the role of unobserved heterogeneity is significant. Therefore, correctly modeling the distribution of the unobserved heterogeneity is crucial to get accurate results.

## 6.2 Empirical Findings

The semiparametric Bayesian estimation of our empirical model yields some interesting findings. First, even after controlling for the unobserved effect using the equation for  $c_i$  semiparametrically, the lagged number of working hours is very important. The point estimate for this variable is 0.6988 with a standard deviation of 0.0183. This indicates that the number of average working hours per week a woman supplies this year will be highly influenced by the number she supplied in the previous year. Specifically, Table 8 tells us that on average, if the average working hours per week in the previous year increases by 1 hour, the average number of working hours in this year will increase for about 0.64 hours. This finding is reasonable since there are a lot of potential sources of state dependence from the theoretical point of view. Examples include human capital accumulation as in Heckman (1981b), job search cost as in Hyslop (1999), child care needs as in Nakamura and Nakamura (1994) and intertemporal nonseparability of preference for leisure as in Hotz, Kydland and Sedlacek (1988). Empir-

ically, the persistence estimate is higher than that found in the previous literature. Using the Panel Study of Income Dynamics (PSID) from 1967 to 1987, Shaw (1994) found that the persistence estimate for the group of people aged between 25 and 34 (our data has roughly the same age composition) is 0.216 using the fixed effects approach and 0.427 using the OLS. The difference may come from several reasons. First, Shaw (1994) concentrates on a sample of women who all work, while our Tobit specification allows the possibility of non-working. Second, Shaw's data is from 1967 to 1987, while our data starts from 1987. As documented by Blau (1998), over the past 25 years, woman's labor supply not only has increased on the extensive (participation) margin, but also on the intensive (hours of work) margin. Third, Shaw's sample only consists of white women, while our data has white, black and Hispanic women.

Nakamura and Nakamura (1994) state that "According to the dominant economic model, the labor supply of individuals is determined by the intersections of their reservation and offered wage functions: an individual's reservation wage being the compensation required for the individual to be willing to work one more unit time period, such as an hour, and the offered wage being what an employer would be willing to pay for this labor input. With this theoretical context, factors that act to raise the reservation wage or lower the offered wage of an individual will tend to decrease his or her labor supply. Economists have argued that family responsibilities, and especially children, can affect both the reservation and offered wage rates of women and that the impacts on their labor supply will be predominantly negative." See Nakamura and Nakamura (1992) and Browning (1992) for a literature survey. Consistent with previous findings on this issue as in Chib and Jeliazkov (2003), Shaw (1994), Nakamura and Nakamura (1994) and Gronau (1973), we find that on average, an additional pre-school child (who is no more than 6 years old) will reduce the woman's labor supply by 3.28 hours per week and an additional school-age child (who is at least 6 years old) will reduce the woman's labor supply by 1.79 hours per week. These results are intuitive since younger children require more attention from the mother and hence reduces the woman's labor supply more than that from a woman with older children.

Gronau (1973) finds that education has a considerable effect on the woman's value of time: the shadow price of time of college graduates exceeds that of elementary school graduates, other things being equal, by over 20 percent. Formal education is considered the prime source of changes in productivity in the market sector and hence is likely to have a positive effect on the woman's labor supply, which is confirmed by our findings that on average, an additional year of education will increase the woman's weekly labor supply for about 2.76 hours. Another important exogenous variable that affects the woman's labor supply decision is the woman's age, which is documented in detail in Chib and Jeliazkov (2003). Chib and Jeliazkov (2003) study the effect of age on female labor participation decision nonparametrically and find that there is a declining labor force participation rate for women who are aged between late 20's and early 30's (which is roughly the age composition for our data). We find similar results here. Marginally, on average, as the woman becomes one year older, she is going to reduce the weekly labor supply for about 4.86 hours. This is because for the group of young women who aged between late 20's and early 30's, this is the period that they get married and have young children. In other words, their family responsibilities increase a lot over this period and hence they choose to supply less labor. More interestingly, from Table 6, we find that the estimate on the interactive variable for age and years of education is negative with a small standard deviation, which means that other things being equal, age has a larger negative effect on female labor supply when they have more years of education. This phenomenon may be due to the special age composition of the data. Since we are concentrating on a sample of women who are aged between late 20's and early 30's, for women in this age group, if they have high years of education, they may get married and have children in recent years and the reduction in labor supply is substantial compared with those with fewer years of education and may get married early and have older children.

For other exogenous variables, as predicted by the theory, being married has a negative effect on the woman's labor supply. Women living in the northcentral and northeast areas work more than women living in the west. Enrolling in a school will on average reduce the woman's labor supply by 0.20 hours per week. Living in the urban area will also increase

a woman's labor supply, which may be attributed to the fact that there are more working opportunities in the urban area compared with that of the rural area. Other variables like other family member's income, living in a metropolitan area and the local unemployment do not have much effects on the woman's labor supply.

The parameter estimates for the auxiliary equation for the unobserved heterogeneity, which are reported in Table 7, also reveal several interesting aspects of the data. First, the initial status on the number of working hours is very important and implies that there is a substantial positive correlation between the unobserved heterogeneity and the initial condition. Second, age, other family members' income, marital status, geographical areas, number of pre-school children and whether the woman lives in the urban area turn out to be correlated with the unobserved heterogeneity.

One interesting empirical question is whether the labor supply stability, as reflected by the coefficient estimate for the lag dependent variable, differs by race. To investigate this question, we divide the sample into two groups, one for the white and the other for the non-white, and calculate the average partial effects for the two groups separately according to equation (44). The results are reported in Table 9. For a white woman, on average, the effect of an additional weekly working hours in the past year on the current year is 0.61 hour, while for a non-white woman, this effect is only 0.55. This reflects that white women enjoy a more stable labor supply than those non-white women. Also, the negative effect coming from greater family responsibilities like the marriage and the number of children has greater impact on the labor supply of white women than that of non-white women. This may be explained by that the living standard for non-white women is lower and they cannot afford much reduction in labor supply even facing high demand of their non-market activities from the family responsibilities. On the other hand, age and year of education have roughly the same effects on both white and non-white women.

To assess the strength of the state dependence, we estimate the average transition probability from non-working in the last period to working in the current period and the average transition probability from working at the average number of hours during the last period but

stop working during the current period. Calculations are done according to equations (45) and (46) and the results are reported in Table 10. For a young woman who does not work in the previous period, the estimated probability of working in the current period, averaged across the unobserved heterogeneity, is 0.8681. At the same time, the estimated average probability for her to remain out of the labor force is 0.1319. On the other hand, if the woman supplies the average number of hours in the previous period, the average probability of quitting the job this period is close to zero.

## 7 Conclusion

In this paper, we propose a semiparametric Bayesian method to estimate dynamic Tobit panel data models with unobserved heterogeneity. This method allows us to model the relationship between the unobserved heterogeneity and initial conditions in a more robust way in that only the conditional mean dependence of the unobserved heterogeneity on the initial conditions is needed. Moreover, this method offers considerable computational advantages thanks to the modern MCMC and data augmentation techniques. With this method, the average partial effect and average transition probabilities can be estimated readily along with the parameter estimates. Simulation studies demonstrate the good finite sample properties of our method and its robustness.

We apply our method to study female labor supply using a panel data set from NLSY79. With our method, we are able to assess the effects of the key regressors on female labor supply. Moreover, we are also able to evaluate the relationship between the unobserved heterogeneity and the key regressors, and measure the strength of the state dependence, and hence offer insight on the dynamics of female labor supply.

## REFERENCES

- Ahn, S. C. and P. Schmidt (1995): "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5-27.
- Albert, J. and S. Chib (1993): "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669-679.
- Anderson, T. W. and C. Hsiao (1982): "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 67-82.
- Arellano, M. and O. Bover (1995): "Another Look at the Instrumental Variables Estimation of Error-Component Models," *Journal of Econometrics*, 68, 29-51.
- Arellano, M. and R. Carrasco (2003): "Binary Choice Panel Data Models with Predetermined Variables," *Journal of Econometrics*, 115, 125-157.
- Arellano, M. and B. Honoré (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics*, Volume 5, ed. by J. Heckman and E. Leamer. Amsterdam: North-Holland, 3229-3296.
- Blau, F. (1998): "Trends in the Well-Being of American Woman, 1970-1995," *Journal of Economic Literature*, 36, 112-165.
- Blundell, R. and S. Bond (1998): "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, 87, 115-143.
- Blundell, R. and R.J. Smith (1991): "Initial Conditions and Efficient Estimation in Dynamic Panel Data Models," *Annals d'Economie et de Statistique*, 20/21, 109-123.
- Browning, M. (1992): "Children and Household Economic Behavior," *Journal of Economic Literature*, 30, 3, 1434-1475.
- Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225-238.

- Chiappori, P. A. and B. Salanie (2000): “Testing for Adverse Selection in Insurance Markets,” *Journal of Political Economy*, 108, 56-78.
- Chib, S. (2001): “Markov Chain Monte Carlo Methods: Computation and Inference,” in *Handbook of Econometrics*, Volume 5, ed. by J. Heckman and E. Leamer. Amsterdam: North Holland, 3569-3649.
- Chib, S. and E. Greenberg (1996): “Markov Chain Monte Carlo Simulation Methods in Econometrics,” *Econometric Theory*, 12, 409-431.
- Chib, S. and B. H. Hamilton (2002): “Semiparametric Bayes Analysis of Longitudinal Data Treatment Models,” *Journal of Econometrics*, 110, 67-89.
- Chib, S. and I. Jeliazkov (2003): “Semiparametric Hierarchical Bayes Analysis of Discrete Data with State Dependence and Serial Correlation,” Working paper, Washington University in St. Louis.
- Escobar, M. D. (1994): “Estimating Normal Means With a Dirichlet Process Prior,” *Journal of the American Statistical Association*, 89, 268-277.
- Escobar, M. D. and M. West (1995): “Bayesian Density Estimation and Inference Using Mixtures,” *Journal of the American Statistical Association*, 90, 577-588.
- Farber, H. S. (1994): “The Analysis of Interfirm Worker Mobility,” *Journal of Labor Economics* 12, 4, 554-593.
- Ferguson, T.S. (1973): “A Bayesian Analysis of Some Nonparametric Problems,” *The Annals of Statistics*, 1, 209-230.
- Ferguson, T.S. (1974): “Prior Distributions on Spaces of Probability Measures,” *The Annals of Statistics*, 2, 615-629.
- Ferguson, T. S. (1983): “Bayesian Density Estimation by Mixtures of Normal Distributions,” in *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His*

- Sixtieth Birthday*, ed. by H. Rizvi and J. Rustagi. New York: Academic Press, pp. 287-302.
- Geweke, J. and M. Keane (2001): "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics*, Volume 5, ed. by J. Heckman and E. Leamer. Amsterdam: North-Holland, 3462-3571.
- Griffin, J. E. and M. F. J. Steel (2004): "Semiparametric Bayesian Inference for Stochastic Frontier Models," *Journal of Econometrics*, forthcoming.
- Ghosal, S., J. K. Ghosh and R. V. Ramamoorthi (1999): "Posterior Consistency of Dirichlet Mixtures in Density Estimation," *Annals of Statistics*, 27, 143-158.
- Gronau, R. (1973): "The Effect of Children on the Housewife's Value of Time," *Journal of Political Economy*, 81, 2, S168-S199.
- Hahn, J. (1999): "How Informative is the Initial Condition in the Dynamic Panel Data Model with Fixed Effects?" *Journal of Econometrics*, 93, 309-326.
- Hasegawa and Kozumi (2003): "Estimation of Lorenz Curves: A Bayesian Nonparametric Approach," *Journal of Econometrics*, 115, 277-291.
- Heckman, J. (1981a): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in *Structural Analysis of Discrete Panel Data with Econometric Applications*, ed. by C. Manski and D. McFadden. Cambridge: MIT Press, 179-195.
- Heckman, J. (1981b): "Heterogeneity and State Dependence," in *Studies in Labor Markets*, ed. by S. Rosen. University of Chicago Press, 91-131.
- Hirano, K. (2002): "Semiparametric Bayesian Inference in Autoregressive Panel Data Models," *Econometrica*, 70, 781-799.



- Honoré, B. (1993): “Orthogonality Conditions for Tobit Models with Fixed Effects,” *Journal of Econometrics*, 59, 35-61.
- Honoré, B. and L. Hu (2001): “Estimation of Cross Sectional and Panel data Censored Regression Models with Endogeneity,” Unpublished Manuscript, Princeton University.
- Hotz, V., Kydland, F. and G. Sedlacek (1988): “Intertemporal Preferences and Labor Supply,” *Econometrica*, 56, 335-360.
- Hsiao, C. (2003): *Analysis of Panel Data*, Second Edition. Cambridge, U.K.: Cambridge University Press.
- Hu, L. (2002): “Estimation of a Censored Dynamic Panel Data Model,” *Econometrica*, 70, 2499-2517.
- Hyslop, D. (1999): “State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Woman,” *Econometrica*, 67, 1255-1294.
- Kane, T. J. and C. E. Rouse (1995): “Labor-Market Returns to Two-and Four-Year College,” *American Economic Review* 85, 3, 600-614.
- Lo, A. Y. (1984): “On a Class of Bayesian Nonparametric Estimates: I. Density Estimates,” *The Annals of Statistics*, 12, 351-357.
- Nakamura, A. and M. Nakamura (1992): “The Econometrics of Female Labor Supply and Children,” *Econometric Reviews*, 11, 1, 1-71.
- Nakamura, A. and M. Nakamura (1994): “Predicting Female Labor Supply: Effects of Children and Recent Work Experience,” *Journal of Human Resources*, 29, 2, 304-327.
- Olsen, R. J.(1994): “Fertility and the Size of the U.S. Labor Force,” *Journal of Economic Literature* 32, 1, 60-100.
- Shaw, K (1994): “The Persistence of Female Labor Supply: Empirical Evidence and Implications,” *Journal of Human Resources*, 29, 2, 348-378.

West, M., P. Müller and M. D. Escobar (1994): “Hierarchical Priors and Mixture Models, With Application in Regression and Density Estimation,” in *Aspects of Uncertainty. A Tribute to D. V. Lindley*, ed. by A. F. M. Smith and P. R. Freeman. London: Wiley, pp. 363-386.

Wooldridge, J.M. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Wooldridge, J.M. (2003): “Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity,” *Journal of Applied Econometrics*, forthcoming.

**Table 1** Parameter Estimates Using the Fully Parametric Bayesian Algorithm for the Experiment where the Unobserved Heterogeneity Term is Normally Distributed

	$\gamma$	$\rho$	$\delta_1$	$\delta_2$	$\sigma_u^2$	$\mu$	$\sigma^2$
true value	1	0.6	0.3	0.2	1	0	1
mean	1.0153	0.6107	0.2537	0.0592	1.0142	-0.0198	0.9763
standard deviation	0.0200	0.0170	0.0593	0.0807	0.0306	0.0471	0.0710

**Table 2** Parameter Estimates Using the Semiparametric Bayesian Algorithm for the Experiment where the Unobserved Heterogeneity Term is Normally Distributed

	$\gamma$	$\rho$	$\delta_1$	$\delta_2$	$\sigma_u^2$
true value	1	0.6	0.3	0.2	1
mean	1.0154	0.6111	0.2495	0.0638	1.0133
standard deviation	0.0199	0.0619	0.0590	0.0836	0.0314

**Table 3** Parameter Estimates Using the Fully Parametric Bayesian Algorithm for the Experiment where the Unobserved Heterogeneity Term is Non-normal

	$\gamma$	$\rho$	$\delta_1$	$\delta_2$	$\sigma_u^2$
true value	1	0.6	0.3	0.2	1
mean	0.9188	0.4869	0.2913	0.1958	1.0476
standard deviation	0.0231	0.0226	0.0871	0.1189	0.0389

**Table 4** Parameter Estimates Using the Semiparametric Bayesian Algorithm for the Experiment where the Unobserved Heterogeneity Term is Non-normal

	$\gamma$	$\rho$	$\delta_1$	$\delta_2$	$\sigma_u^2$
true value	1	0.6	0.3	0.2	1
mean	0.9813	0.5891	0.2927	0.1689	0.9978
standard deviation	0.0248	0.0200	0.0337	0.0618	0.0345

**Table 5** Summary Statistics<sup>9</sup>

Variable	Explanation	Mean	Standard Deviation
Hours	Average working hours per week	28.1407	17.5463
Age	Age/10	2.9590	0.2743
Educ	Yeas of education	13.2749	2.2652
Income	Other family member's income/1000	27.8697	77.5495
Married	1 if Married, 0 otherwise	0.6408	0.4798
Northcentral	1 if in the Northcentral area, 0 otherwise	0.3087	0.4620
Northeast	1 if in the Northeast area, 0 otherwise	0.1254	0.3312
South	1 if in the South area, 0 otherwise	0.3859	0.4869
Old	# of Children aged 6 or above	0.6842	0.9644
Young	# of Children aged under 6	0.6365	0.7829
School	1 if in the School, 0 otherwise	0.0691	0.2536
SMSA	1 if in a Metropolitan area, 0 otherwise	0.3520	0.4776
Unemp	Local unemployment rate greater than 6%	0.4495	0.4995
Urban	Urban area	0.7484	0.4339
Black	1 if Black, 0 otherwise	0.1211	0.3264
Hispanic	1 if Hispanic, 0 otherwise	0.0610	0.2394

<sup>9</sup>The sample consists of a cross section of 1115 individuals over 6 years.

**Table 6** Results from the Main Equation Using the Semiparametric Bayesian Algorithm

Variable	Mean	Stan. Dev.	Variable	Mean	Stan. Dev.
Age	-13.7280	6.3795	School	-3.5640	0.7775
Educ	3.0097	1.1593	SMSA	0.9588	0.7446
Income	-0.0015	0.0021	Unemp	-0.1399	0.3967
Married	-1.8655	0.6184	Urban	1.9300	0.9003
Northcentral	3.3389	1.6993	(Age) <sup>2</sup>	2.2431	1.8486
Northeast	8.2803	2.9172	(Age)*(Educ)	-2.0172	0.8799
South	2.3966	1.6434	(Age) <sup>2</sup> *(Educ)	0.3425	0.1801
Old	-1.9595	0.4281	Hours <sub>-1</sub>	0.6988	0.0183
Young	-3.6062	0.4131			

**Table 7** Results from the Auxiliary Equation Using the Semiparametric Bayesian

Algorithm		
Variable	Mean	Standard Deviation
Mean of Age	-13.5491	8.2735
Mean of Educ	1.8529	2.5484
Mean of Income	-0.0117	0.0040
Mean of Married	2.1281	0.7538
Mean of Northcentral	-3.7333	1.6673
Mean of Northeast	-9.3105	2.9051
Mean of South	1.8837	0.4267
Mean of Old	0.0154	0.5108
Mean of # of Young	2.8209	1.3460
Mean of School	-1.5690	0.9442
Mean of SMSA	-1.6270	1.6718
Mean of Unemp	0.5313	0.6431
Mean of Urban	-2.0950	0.9638
Mean of (Age) <sup>2</sup>	3.5318	0.5484
Mean of (Age)*(Educ)	-0.0986	0.7098
Mean of (Age) <sup>2</sup> *(Educ)	-0.1096	2.5242
Black	-1.0819	1.8804
Hispanic	-0.1441	0.3548
Hours <sub>0</sub>	0.1130	0.0151

**Table 8** Average Partial Effects

Variable	Mean	Stan. Dev.	Variable	Mean	Stan. Dev.
Age	-4.8580	1.9339	Young	-1.7856	0.3835
Educ	2.7642	1.3540	Old	-3.2830	0.3731
Income	-0.0013	0.0020	School	-0.1964	0.0443
Married	-0.8594	0.3009	SMSA	0.2167	0.1693
Northcentral	0.5816	0.2831	Unemp	-0.0480	0.1342
Northeast	0.6705	0.2593	Urban	0.9726	0.4259
South	0.5759	0.3996	Hours <sub>-1</sub>	0.6369	0.0179

**Table 9** Average Partial Effects by Race

Variable	White		Non-white	
	Mean	Standard Deviation	Mean	Standard Deviation
Age	-4.8491	1.8846	-4.3152	1.6945
Educ	2.8020	1.3429	2.5183	1.2080
Income	-0.0014	0.0020	-0.0013	0.0018
Married	-0.7274	0.3097	-0.4056	0.1847
Northcentral	1.0011	0.3813	0.1571	0.1129
Northeast	0.7698	0.2095	0.2762	0.1245
South	0.4835	0.4065	-0.1580	0.0475
Young	-3.3378	0.3177	-2.9992	0.2883
Old	-1.6215	0.3564	-1.4572	0.3220
School	-0.2050	0.0430	-0.1580	0.0475
SMSA	0.2961	0.1714	0.0880	0.0760
Unemp	-0.0353	0.1310	-0.0251	0.1003
Urban	0.8138	0.5708	0.6336	0.4564
Hours <sub>-1</sub>	0.6134	0.0163	0.5512	0.0168

**Table 10** Average Transition Probabilities

Variable	Mean	Standard Deviation
$p^{01}$	0.8681	0.0076
$p^{00}$	0.1319	0.0076
$p^{10}$	0.0072	0.0040



Figure 1: Time Series, Histograms and Autocorrelations of the MCMC Draws Using the Fully Parametric Bayesian Algorithm when the Unobserved Heterogeneity Term is Normal.

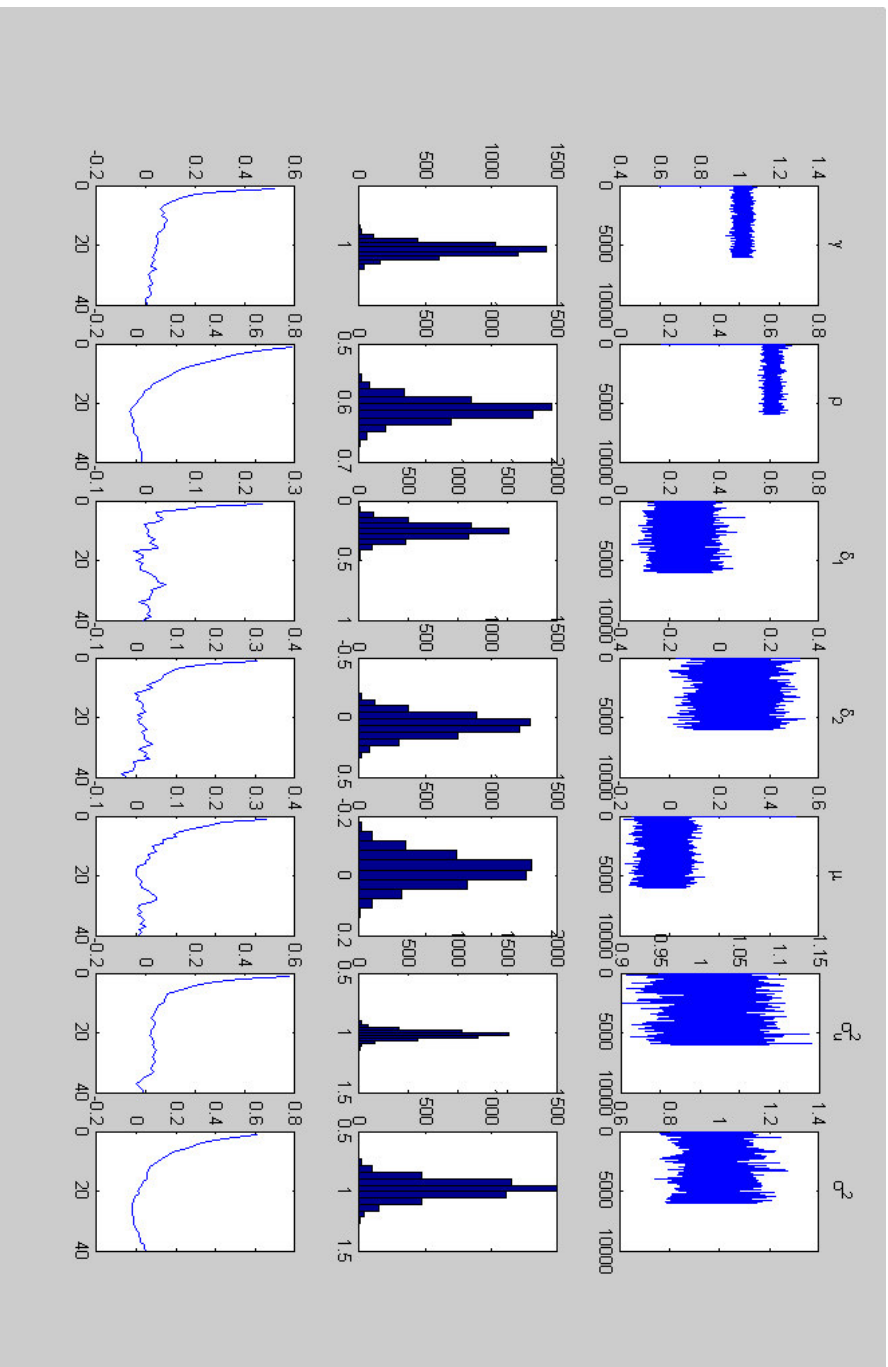


Figure 2: Time Series, Histograms and Autocorregrams of the MCMC Draws Using the Semiparametric Bayesian Algorithm when the Unobserved Heterogeneity is Normal.

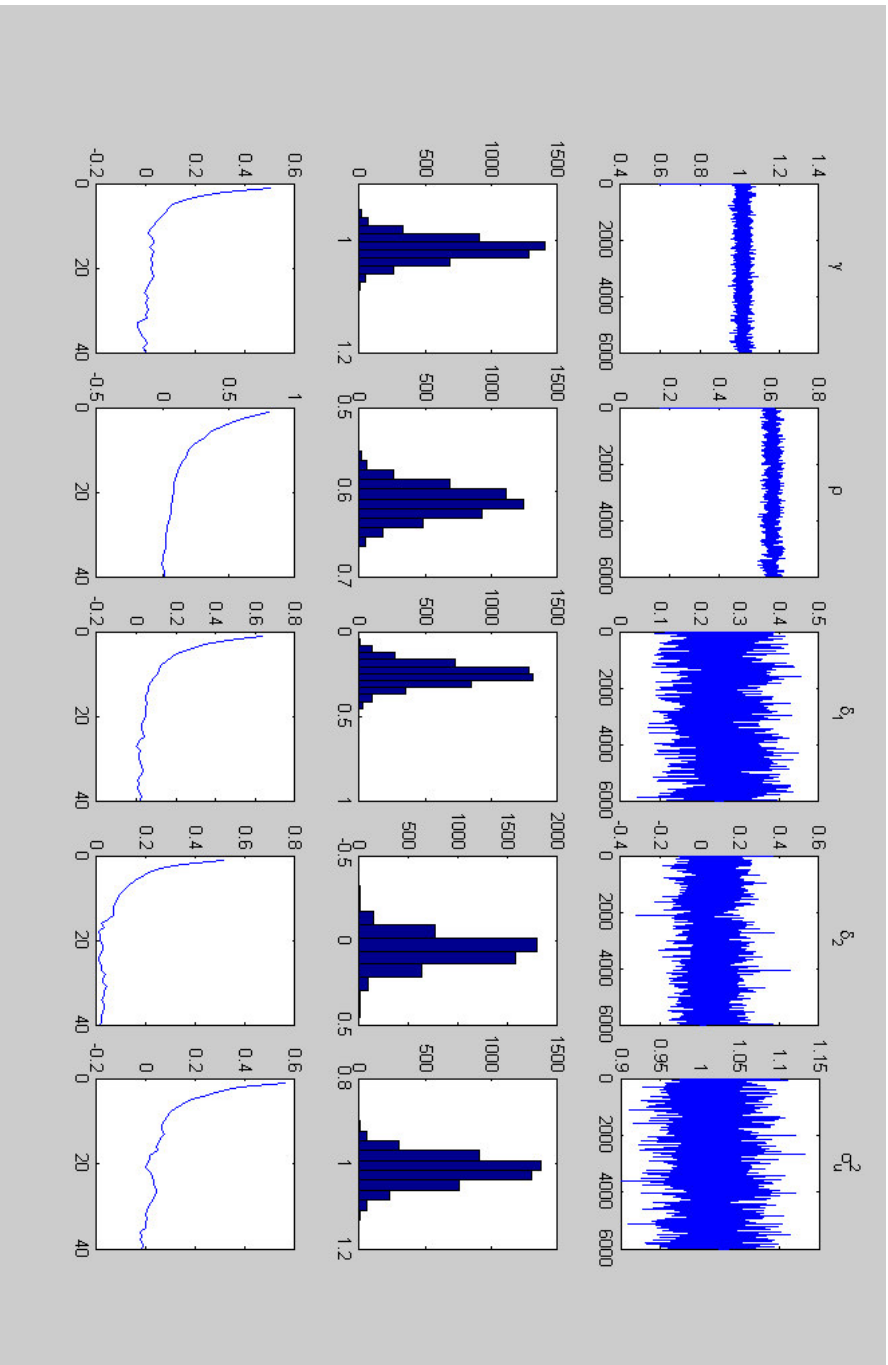


Figure 3: Time Series, Histograms and Autocorrelations of the MCMC Draws Using Fully Parametric Bayesian Algorithm when Unobserved Heterogeneity is Non-normal.

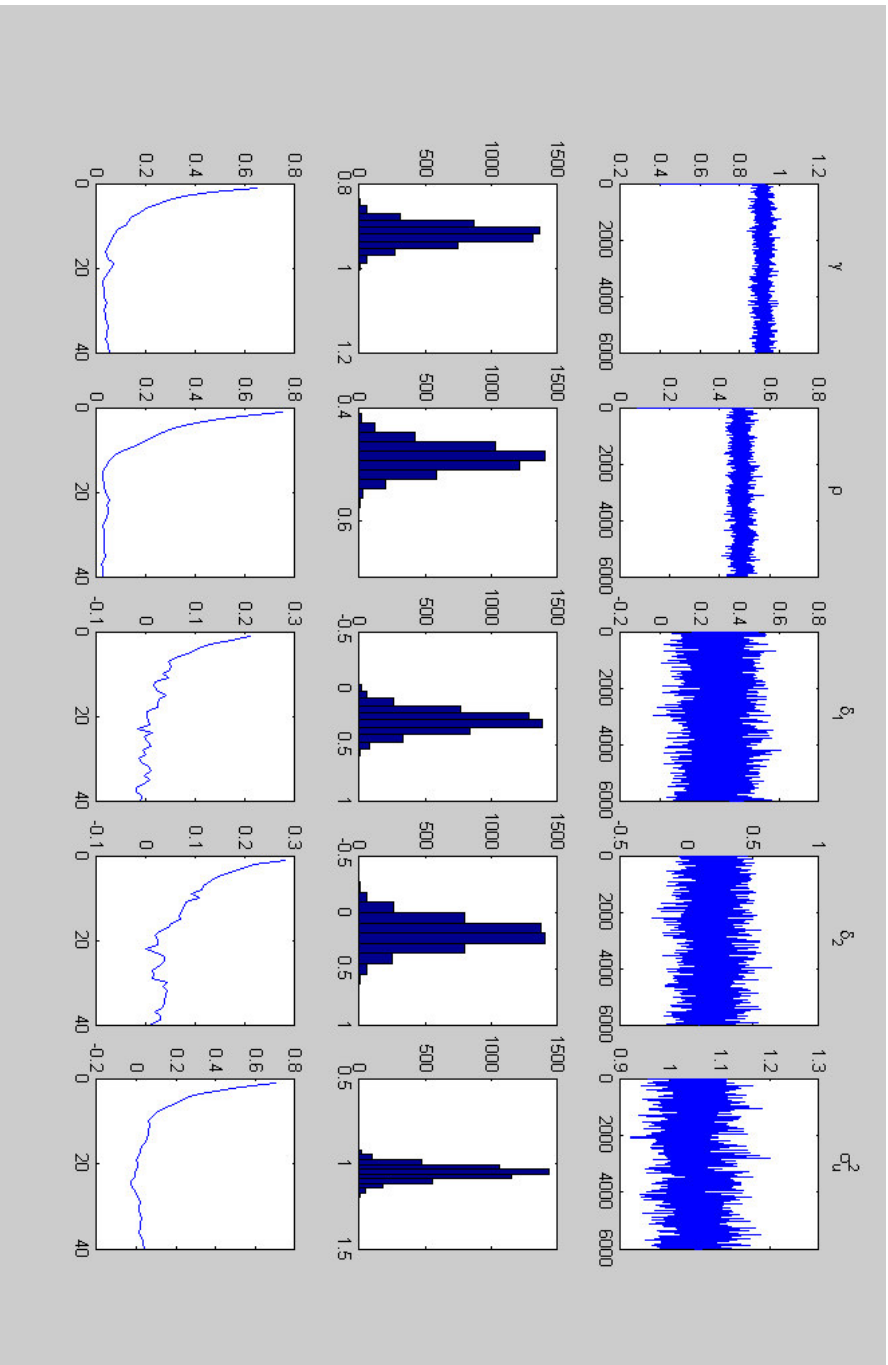


Figure 4: Time Series, Histograms, Autocorrelations of the MCMC Draws Using Semiparametric Bayesian Algorithm when Unobserved Heterogeneity is Non-normal.

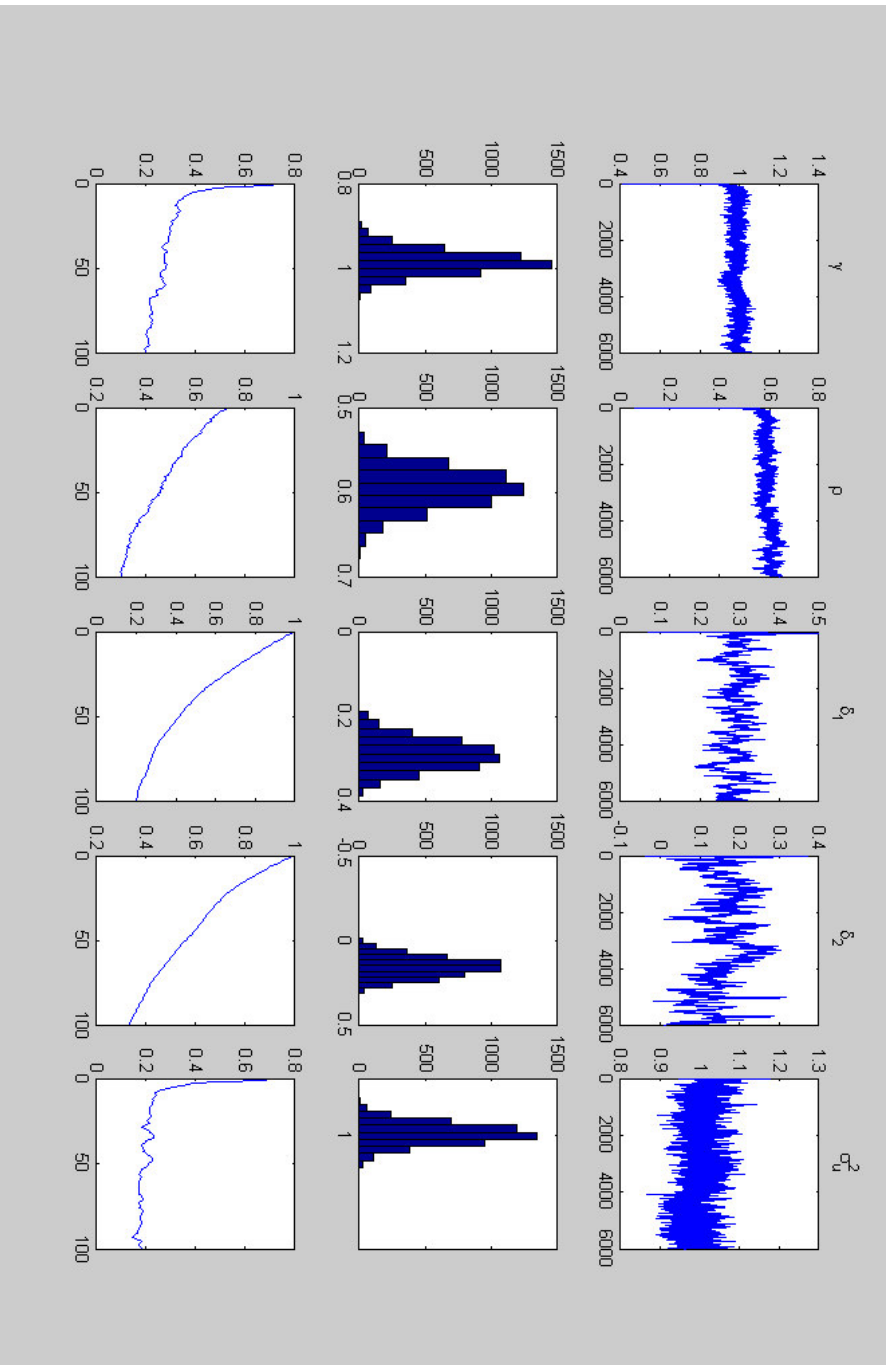


Figure 5: Estimated Predictive Density Estimate for  $\alpha_i^2$

