JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **23**: 699–728 (2008)
Published online 2 September 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.1017

SEMIPARAMETRIC BAYESIAN INFERENCE FOR DYNAMIC TOBIT PANEL DATA MODELS WITH UNOBSERVED HETEROGENEITY

TONG LI[a]* AND XIAOYONG ZHENG[b]

[a] *Department of Economics, Vanderbilt University, Nashville, TN, USA*
[b] *Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC, USA*

SUMMARY

This paper develops semiparametric Bayesian methods for inference of dynamic Tobit panel data models. Our approach requires that the conditional mean dependence of the unobserved heterogeneity on the initial conditions and the strictly exogenous variables be specified. 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 run. We apply our method to study female labor supply using a panel data set from the National Longitudinal Survey of Youth 1979. Copyright © 2008 John Wiley & Sons, Ltd.

Received 26 July 2005; Revised 15 November 2007

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 requires that the conditional mean dependence of the unobserved individual heterogeneity on the initial conditions and the strictly exogenous variables be specified.

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.[1]

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,

* Correspondence to: Tong Li, Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA. E-mail: tong.li@vanderbilt.edu

[1] The only exception, to the best of our knowledge, is that Hu (2002) estimates a dynamic Tobit panel data model with the lagged uncensored variable as one of the explanatory variables.

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 unobserved individual heterogeneity, and Honoré and Hu (2004) 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 (2004), time dummies are not allowed in the strictly exogenous covariates.

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 (2005) 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.[2] 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.[3] 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 (2005) in modeling the relationship between the unobserved heterogeneity and the initial conditions, 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. Thus we propose a semiparametric Bayesian method to estimate the dynamic Tobit panel data model with lagged censored dependent variables.[4] The Bayesian nonparametric approach is introduced by Lo (1984) and Ferguson (1983), with the later work by Escobar (1994), Escobar and West (1995), and West

[2] As acknowledged in Wooldridge (2005), 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.

[3] Loudermilk (2006) applies the same idea as Wooldridge (2005) to dynamic panel data model with fractional dependent variables.

[4] 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 (2004). Another type of dynamic censored regression model, as considered in Hu (2002), is the dynamic Tobit model with lagged latent

DOI: 10.1002/jae

et al. (1994) discussing its computational issues and by Ghosal *et al*. (1999) for the posterior consistency of the Bayesian nonparametric estimation. Recently, this method has been adopted by econometricians in addressing various issues. See, for example, 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.

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. Second, some important quantities of economic interest such as the average partial effects and the average transition probabilities (ATP) can be readily obtained as a by-product of the Markov chain Monte Carlo (MCMC) run. 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 transition probabilities between working and not working. As such, our paper contributes to the literature by estimating the dynamic Tobit panel data model with corner solution outcomes with an empirical application. 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 APEs 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. The related algorithms are included in the Appendix.

2. THE MODEL

Let y_{it} be the censored response variable of interest, where the indices i and t ($i = 1, \ldots, n$, $t = 1, \ldots, T$) refer to individual i and time period t, respectively.[5] 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

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

dependent variables, which is referred to as the censored regression model with data censoring or coding by Wooldridge (2002).

[5] The model can be easily extended to the unbalanced panel model. For ease of exposition, we focus on the balanced panel data model.

DOI: 10.1002/jae

where γ is a vector of coefficients for the explanatory variables and ρ is a vector of lag coefficients, and u_{it} is a sequence of i.i.d. random variables distributed as normal (0, σ_u^2).[6] As in Wooldridge (2005), 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}.[7] 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).[8]

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 heterogeneity c_i on the initial conditions and observed strictly exogenous variables

$$E[c_i|\mathbf{y}_{i0}, \mathbf{z}_i] = c + h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta \tag{2}$$

where c is a constant, $\mathbf{y}_{i0}$ is the vector of initial values of the dependent variable y_i, and $\mathbf{z}_i$ is a set of explanatory variables that only vary over different individuals but are time-invariant.[9] It can be the row vector of all (nonredundant) explanatory variables in all time periods. That is, $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots z_{iT})$ and each z_{it} can be multidimensional as in Wooldridge (2005). Alternatively it can be $\mathbf{z}_i = \bar{\mathbf{z}}_i$, where $\bar{\mathbf{z}}_i$ is the average of $\mathbf{z}_{it}$ over all the time periods as in Chib and Jeliazkov (2006).[10] $h(\cdot)$ is a function that allows $\mathbf{y}_{i0}$ and $\mathbf{z}_i$ to appear in a variety of ways, including nonlinearities and interactions between $\mathbf{y}_{i0}$ and $\mathbf{z}_i$.

We can rewrite (2) as

$$c_i = h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \alpha_i \tag{3}$$

where α_i is an error term. It 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 (2005). On the other hand, and more importantly, we leave the distribution of α_i unspecified.[11] Instead, we will approximate its

[6] Note that for ease of exposition u_{it} is assumed to be normally distributed and homoskedastic. In Appendix III, we give an algorithm for the more general case where the distribution of u_{it} is unspecified and can be heteroskedastic.

[7] 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 (2006) 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}.

[8] For surveys on the use of the MCMC techniques in econometrics, see, for example, Chib and Greenberg (1996) and Chib (2001).

[9] Implicitly, this specification rules out the possibility that $\mathbf{y}_{i0}$ and all past values of $\mathbf{y}_{it}$ in periods prior to period 0 depend on the unobserved individual heterogeneity c_i. Otherwise, as pointed out by Honoré (2002), the distribution of c_i, conditional on the strictly exogenous variables and $\mathbf{y}_{i0}$ can depend on the time series properties of $\mathbf{z}_{it}$ in a very complicated way. In view of this, following one of the referees' suggestion, in the Monte Carlo studies and empirical application that follow, we investigate how robust our approach is to the specification (2).

[10] For the identification purpose, those time-constant variables like race and gender cannot be in both $\mathbf{z}_{it}$ and $\bar{\mathbf{z}}_i$.

[11] We maintain the homoskedasticity assumption on α_i as is commonly assumed in the Tobit models. In addition, we need to assume that α_i is independent of $\mathbf{y}_{i0}$ and $\mathbf{z}_i$, which is a strong assumption as well.

DOI: 10.1002/jae

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

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

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 μ_j and variance σ_j^2. Note that this is a different specification from 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

With the specifications in Section 2, for each individual i we have the following conditional density for the dependent variables:

$$\begin{aligned} & f(y_{i1}, y_{i2}, \ldots, y_{iT}|\mathbf{y}_{i0}, \mathbf{z}_i, c_i, \gamma, \rho) \\ & = \prod_{t=1}^{T} \left\{ \left[1 - \Phi\left(\frac{\mathbf{z}_{it}\gamma + \mathbf{g}(\mathbf{y}_{it-1})\rho + c_i}{\sigma_u} \right) \right]^{1(y_{it}=0)} \left[\frac{1}{\sigma_u} \phi\left(\frac{y_{it} - \mathbf{z}_{it}\gamma - \mathbf{g}(\mathbf{y}_{it-1})\rho - c_i}{\sigma_u} \right) \right]^{1(y_{it}>0)} \right\} \end{aligned} \tag{5}$$

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

$$\begin{aligned} y_{it}^* &= \mathbf{z}_{it}\gamma + \mathbf{g}(\mathbf{y}_{it-1})\rho + c_i + u_{it} \\ y_{it} &= \mathbf{1}(y_{it}^* > 0) y_{it}^* \\ c_i &= h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \alpha_i \\ \alpha_i|\mathbf{y}_{i0}, \mathbf{z}_i &\sim \text{i.i.d. normal}(\mu_i, \sigma_i^2) \end{aligned} \tag{6}$$

As a result, likelihood (5) can be modified as follows to be conditioning on the latent variables y_{it}^* in addition to other conditioning variables included in (5):

$$\begin{aligned} & f(y_{i1}, y_{i2}, \ldots, y_{iT}|y_{i1}^*, y_{i2}^*, \ldots, y_{iT}^*, \mathbf{y}_{i0}, \mathbf{z}_i, c_i, \gamma, \rho) = \\ & \prod_{t=1}^{T} \{1(y_{it} > 0) 1(y_{it} = y_{it}^*) + 1(y_{it} = 0) 1(y_{it}^* \leq 0)\} \\ & * \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{1}{2\sigma_u^2} (y_{it}^* - \mathbf{z}_{it}\gamma - \mathbf{g}(\mathbf{y}_{it-1})\rho - c_i)^2 \right) \end{aligned} \tag{7}$$

DOI: 10.1002/jae

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, γ and ρ, becomes similar to the standard posterior updating for simple linear panel data models and therefore straightforward to implement.

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 α_i in the unobserved individual heterogeneity through the relationship $\alpha_i = c_i - h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta$. After recovering the error terms, we can use a Bayesian approach as detailed in Appendix II 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 μ_i and σ_i^2 for each i. In the second part of each iteration, we update the model parameters and values for the latent variables. Incorporating other parts of the model, we have the following algorithm for the semiparametric dynamic Tobit data models.[12]

Denoting $w_{it} = (\mathbf{z}_{it}, \mathbf{g}(\mathbf{y}_{it-1}))$, $\beta = (\gamma', \rho')'$, we have the following algorithm.[13]

Algorithm 1 *MCMC for semiparametric dynamic Tobit panel data models*

1. Conditional on y_{it}, w_{it}, $h(\mathbf{y}_{i0}, \mathbf{z}_i)$, β, δ, μ_i, σ_i^2 and σ_u^2 but marginalized over c_i, y_{it}^* is updated from a normal distribution with mean $w_{it}\beta + h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \mu_i$ 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 σ_u^2 and β in one block. Using the improper flat prior for β and the independent gamma $\left(\frac{N_1}{2}, \frac{R_1}{2}\right)$ prior for $1/\sigma_u^2$, that is: $1/\sigma_u^2 \propto (1/\sigma_u^2)^{\frac{N_1}{2}-1} e^{-R_1(1/\sigma_u^2)}$,
   (a) draw $1/\sigma_u^2$ from gamma $\left(\dfrac{N_1 + nT}{2}, \dfrac{R_1 + \sum_{i=1}^n \sum_{t=1}^T (y_{it}^* - w_{it}\widehat{\beta} - c_i)^2}{2}\right)$ where $\widehat{\beta} = \text{inv}$ $\left(\sum_{i=1}^n \sum_{t=1}^T w_{it}' w_{it}\right) \times \left(\sum_{i=1}^n \sum_{t=1}^T w_{it}'(y_{it}^* - c_i)\right)$ and
   (b) update β from a normal distribution with mean $\widehat{\beta}$ and variance $\text{inv}\left(\dfrac{1}{\sigma_u^2} \sum_{i=1}^n \sum_{t=1}^T w_{it}' w_{it}\right)$.
3. Conditional on y_{it}^*, w_{it}, $h(\mathbf{y}_{i0}, \mathbf{z}_i)$, β, δ, μ_i, σ_i^2, σ_u^2, update c_i by drawing from a normal distribution with mean c_i^* and variance $\text{inv}\left(\dfrac{T}{\sigma_u^2} + \dfrac{1}{\sigma_i^2}\right)$, where $c_i^* = \text{inv}\left(\dfrac{T}{\sigma_u^2} + \dfrac{1}{\sigma_i^2}\right) \times$ $\left(\dfrac{1}{\sigma_u^2} \sum_{t=1}^T (y_{it}^* - w_{it}\beta) + \dfrac{1}{\sigma_i^2}(h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \mu_i)\right)$.

[12] The Bayesian estimation algorithm for the Wooldridge (2005) model is presented in Appendix III. This model is used as the benchmark for comparison. Two other closely related models, which are the dynamic tobit model with lagged uncensored variables and the dynamic model with unknown distribution of u_{it}, and their algorithms in the semiparametric Bayesian framework, are also included in Appendix III.
[13] See Appendix I for details of this algorithm.

DOI: 10.1002/jae

4. Conditional on c_i, $h(\mathbf{y}_{i0}, \mathbf{z}_i)$, μ_i and σ_i^2 update δ. Using the improper flat prior for δ, we update δ by drawing from a normal distribution with mean $\widehat{\delta}$ and variance $\mathrm{inv}\left(\sum_{i=1}^{n} \frac{h(\mathbf{y}_{i0}, \mathbf{z}_i)' h(\mathbf{y}_{i0}, \mathbf{z}_i)}{\sigma_i^2}\right)$, where

$$\widehat{\delta} = \mathrm{inv}\left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} h(\mathbf{y}_{i0}, \mathbf{z}_i)' h(\mathbf{y}_{i0}, \mathbf{z}_i)\right) \times \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} h(\mathbf{y}_{i0}, \mathbf{z}_i)'(c_i - \mu_i)\right)$$

5. Lastly, recover $\alpha_i = c_i - h(\mathbf{y}_{i0}, \mathbf{z}_i)'\delta$ and update the number of components m_c in approximating mixture of normals and the mean and variance of the normal denoted by μ_i and σ_i^2 for each i using the Bayesian nonparametric density estimation method described in Appendix II.

4. AVERAGE PARTIAL EFFECTS AND TRANSITION PROBABILITIES

4.1. Average Partial Effects

For nonlinear models, besides the estimation of parameters, obtaining the APEs 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. The method proposed by Wooldridge (2005) 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 APEs will also need the implementation of simulation 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 APEs 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 jth covariate w_{it}^j. For the dynamic Tobit panel data model, if w_{it}^j is continuous and there are no interaction terms involved, then the partial effect of the covariate on the expected value of the dependent variable is

$$m_{it}^j = \frac{\partial E(y_{it}|w_{it}, c_i, \beta, \sigma_u)}{\partial w_{it}^j} = \Phi\left(\frac{w_{it}\beta + c_i}{\sigma_u}\right)\beta^j \tag{8}$$

where $\Phi(\cdot)$ denotes the cumulative standard normal distribution 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

$$E(y_{it}|w_{it}, c_i, \beta, \sigma_u) = \Phi\left(\frac{w_{it}\beta + c_i}{\sigma_u}\right)(w_{it}\beta + c_i) + \sigma_u \phi\left(\frac{w_{it}\beta + c_i}{\sigma_u}\right) \tag{9}$$

If the covariate w_{it}^j is involved 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 c_i usually does not have a clear measurement unit and is usually not observed. Therefore, it is usually more interesting 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 additional simulation 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 APEs conditional on the observed data, but marginalized over all unknowns including the model parameters and the unobserved individual heterogeneity.[14] 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

$$\pi(m_{it}^j|\text{data}) = \int \pi(m_{it}^j|\text{data}, c_i, \beta, \sigma_u)\text{d}\pi(c_i, \beta, \sigma_u|\text{data}) \tag{10}$$

A sample of m_{it}^j can be produced by the method of composition using the draws of c_i, β, σ_u from steps 2 and 3 in the algorithm described above. 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 (ith observation in t period) mean partial effect, when w_{it}^j is continuous and there is no interaction terms for this covariate, can be estimated as

$$\overline{m_{it}^j} \approx G^{-1}\sum_{g=1}^{G} m_{it}^{j(g)} \text{ where } m_{it}^{j(g)} = \Phi\left(\frac{w_{it}\beta^{(g)} + c_i^{(g)}}{\sigma_u^{(g)}}\right)\beta^{j(g)} \tag{11}$$

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

$$m^j = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} m_{it}^j}{nT} \tag{12}$$

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 dataset 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

[14] Chib and Hamilton (2002) use this approach to obtain the average treatment effects.

DOI: 10.1002/jae

$$p_{it}^{01} = pr(y_{it} > 0 | y_{it-1} = 0, \mathbf{z}_{it}, c_i) \tag{13}$$
$$= \Phi\left(\frac{\mathbf{z}_{it}\gamma + c_i}{\sigma_u}\right)$$

and $p_{it}^{00} = 1 - p_{it}^{01}$. The probability for individual i to quit working in period t conditional on working $\overline{y}_{it-1}$ hours in period $t-1$ is

$$p_{it}^{10} = pr(y_{it} = 0 | y_{it-1} = \overline{y}_{it-1}, \mathbf{z}_{it}, c_i) = 1 - \Phi\left(\frac{\overline{y}_{it-1}\rho + \mathbf{z}_{it}\gamma + c_i}{\sigma_u}\right) \tag{14}$$

Again, these probabilities depend on the unobserved heterogeneity. Thus, the ATPs can be computed in the same way as for the APEs during the MCMC model estimations.

5. SIMULATION STUDIES

We perform several 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 almost 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 implement the estimation method proposed by Wooldridge (2005) to the same simulated dataset and compare the results from that method to the results from the Bayesian estimation. The standard errors for the APEs and the ATPs using the Wooldridge method are obtained by bootstrap with 1000 iterations. In general, the results from Wooldridge's method is close to the result from the fully parametric Bayesian algorithm since both methods have the same underlying assumptions.[15]

The first experiment is designed as the following:

$$y_{it}^* = z_{it}\gamma + \rho y_{it-1} + c_i + u_{it}$$
$$c_i = y_{i0}\delta_1 + \mathbf{z}_i\delta_2 + \alpha_i \tag{15}$$

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. $\mathbf{z}_i$ is the simple average of z_{it} across the time periods. α_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: $\gamma = 1$, $\rho = 0.6$, $\delta_1 = 0.3$, $\delta_2 = 0.2$.[16]

[15] As pointed out by a referee, at issue is how to compare APEs from classical and Bayesian methods. In what follows, we compare the mean of the posterior distribution from the Bayesian method with the point estimate from the classical method, while comparing the standard deviation of the posterior distribution from the Bayesian method with the standard error calculated using the Delta method from the classical approach.

[16] Since in simulation experiments we observe both the observed heterogeneity and the unobserved heterogeneity, true values of the average partial effects and the average transition probabilities can be calculated using their formulas in the previous section. Also, since different experiments assume different error distributions, these quantities are different for different experiments.

DOI: 10.1002/jae

Columns 3 and 4 in Table I summarize the results using the fully parametric approach. Columns 5 and 6 in Table I report the results using the semiparametric Bayesian estimation method and the last two columns of Table I collect the results using the Wooldridge method. Overall, from this table, we can see that when the error term in the unobserved individual heterogeneity is normal, both the two estimation procedures which assume normality for the error term (the fully parametric Bayesian algorithm and the Wooldridge method) and the semiparametric Bayesian algorithm perform well.

The second experiment is designed in the same way as the first experiment, except that now we generate α_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 α_i highly non-normal. In this case, as we see from the results of Monte Carlo simulations in Table II, the fully parametric Bayesian updating algorithm and the Wooldridge method, which assume normality, significantly bias the parameters of interest downwards by 10–15%. On the other hand, as seen from Table II, the semiparametric

Table I. Results from the experiment where the unobserved heterogeneity term is normally distributed

	True value	Para. Bayesian		Semipara. Bayesian		Wooldridge	
		Mean	SD	Mean	SD	Estimate	SE
γ	1	1.0157	0.0197	1.0170	0.0198	1.0163	0.0197
ρ	0.6	0.6112	0.0165	0.6127	0.0169	0.6115	0.0166
δ_1	0.3	0.2523	0.0586	0.2492	0.0577	0.2482	0.0563
δ_2	0.2	0.0588	0.0810	0.0594	0.0808	0.0604	0.0801
σ_u^2	1	1.1032	0.0310	1.0107	0.0302	0.9880	0.0300
μ	0	-0.0208	0.0462	n.a.	n.a.	-0.0215	0.0465
σ^2	1	0.9772	0.0718	n.a.	n.a.	0.9781	0.0702
APE_z	0.6087	0.6136	0.0113	0.6142	0.0114	0.6183	0.0190
$\text{APE}_{y_{-1}}$	0.3652	0.3692	0.0099	0.3701	0.0102	0.3721	0.0144
p^{00}	0.4791	0.4865	0.0056	0.4869	0.0057	0.4887	0.0158
p^{01}	0.5209	0.5135	0.0056	0.5131	0.0057	0.5113	0.0158
p^{10}	0.3137	0.3184	0.0051	0.3183	0.0052	0.3549	0.0148
$\overline{y}_{it-1}$	1.2652						

Table II. Results from the experiment where the unobserved heterogeneity term is non-normal

	True value	Para. Bayesian		Semipara. Bayesian		Wooldridge	
		Mean	SD	Mean	SD	Estimate	SE
γ	1	0.9480	0.0230	1.0138	0.0243	0.9379	0.0197
ρ	0.6	0.4888	0.0221	0.5953	0.0220	0.4827	0.0166
δ_1	0.3	0.1825	0.0905	0.2482	0.0349	0.1259	0.0563
δ_2	0.2	0.0618	0.1223	0.1677	0.0554	0.0305	0.0801
σ_u^2	1	1.0726	0.0382	1.0040	0.0342	0.9368	0.0292
APE_z	0.4041	0.3858	0.0090	0.4070	0.0089	0.4113	0.0198
$\text{APE}_{y_{-1}}$	0.2424	0.1989	0.0089	0.2390	0.0085	0.2117	0.0140
p^{00}	0.6630	0.6447	0.0055	0.6651	0.0050	0.6041	0.0181
p^{01}	0.3370	0.3553	0.0055	0.3349	0.0050	0.3959	0.0181
p^{10}	0.5971	0.5846	0.0046	0.6016	0.0034	0.5559	0.0187
$\overline{y}_{it-1}$	0.6798						

DOI: 10.1002/jae

Bayesian algorithm is robust to the specification of the distribution of the error terms. For the estimates of the APEs and the ATPs, using the fully parametric Bayesian updating algorithm or the Wooldridge method leads to smaller biases compared with that of the estimates. However, the semiparametric Bayesian algorithm still performs significantly better.

The robustness gained by using the semiparametric Bayesian algorithm instead of the more user-friendly Wooldridge method comes with a computational cost. Using a Pentium® 4 2.4 GB processor, the computational time for the fully parametric algorithm is 1.05 s per 10 draws and 5.18 s per 10 draws for the semiparametric algorithm. On the other hand, the Wooldridge method can be implemented within just a few seconds. However, as mentioned above, a second main advantage of the Bayesian method is that the APEs and the ATPs can be obtained easily as a by-product of the estimation algorithm. Therefore, after the estimation using the Bayesian approach, we can easily get the posterior distributions for the APEs and the ATPs. This is not the case when we use the Wooldridge approach.

Additional simulation experiments are conducted to see whether the semiparametric Bayesian algorithm is robust to slight complications and/or misspecifications in the model. In these additional experiments, data are generated under slightly different assumptions but the estimation algorithm remains the same as in the second experiment. The results from these additional experiments are collected in Table III. When u_{it} is non-normal ($u_{it} \sim$ gamma(1,1/2)) or heteroskedastic ($u_{it} \sim$ normal($0,\sigma_{it}^2$) and $\sigma_{it}^2 \sim$ uniform[0,2]) or z_{it} follows an AR(1) process with a lag coefficient 0.5, comparing the corresponding columns of Table III with columns 5 and 6 of Table II, we can see these slight misspecifications lead to slightly worse parameter estimates, with even less worse estimates for the APEs and the ATPs. On the other hand, when u_{it} follows an AR(1) process with a lag coefficient 0.5, the resulting estimates become significantly worse. This indicates the semiparametric Bayesian algorithm is not very robust to this additional source of dynamics. Finally, we carry out an experiment where we generate the data for five periods as usual and then use only the last three periods of data for estimation. In more detail, we treat the data in period 2 as the initial period and $\mathbf{z}_i$ is calculated as the simple average of $\mathbf{z}$ in periods 3, 4 and 5. The results are reported in the last three columns of Table III. As one can see from the table, the estimates are reasonably close to the true values. This indicates that the current specification of the initial condition provides a good approximation when it is actually slightly misspecified. From these experiments, we can conclude that the semiparametric Bayesian algorithm is robust to slight complications and/or misspecifications in the model.

6. APPLICATION: INTERTEMPORAL FEMALE LABOR SUPPLY

In this section, we illustrate the proposed method by estimating a reduced-form intertemporal female labor supply model. This dataset consists of a panel of 1115 young women over 7 years (1987–1993). The dataset is from the NLSY79, which can be obtained from the US Department of Labor. The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14–22 years of age when first surveyed in 1979. The survey runs every year from 1979 to 1994 and every 2 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 (1995) for returns of schooling, and Olsen (1994) for fertility, among others. We use the number of average working hours per week, which is

Table III. Results from additional experiments where the unobserved heterogeneity term is non-normal

	u_{it} non-normal			u_{it} heteroskedasticity			u_{it} AR(1)		
	True	Mean	SD	True	Mean	SD	True	Mean	SD
γ	1	0.9888	0.0122	1	0.9841	0.0220	1	1.0503	0.0240
ρ	0.6	0.5834	0.0104	0.6	0.6110	0.0182	0.6	0.7221	0.0202
δ_1	0.3	0.3303	0.0202	0.3	0.2582	0.0379	0.3	0.1447	0.0402
δ_2	0.2	0.2871	0.0349	0.2	0.1117	0.0519	0.2	0.0490	0.0807
σ_u^2	1	3.4650	0.1031	1	0.9837	0.0298	1	1.0998	0.0408
APE_z	0.4529	0.4645	0.0064	0.4426	0.4396	0.0092	0.4065	0.4121	0.0088
$\text{APE}_{y_{-1}}$	0.2718	0.2741	0.0047	0.2655	0.2713	0.0080	0.2439	0.2833	0.0080
p^{00}	0.6480	0.5791	0.0049	0.6356	0.6351	0.0044	0.6630	0.6822	0.0046
p^{01}	0.3520	0.4209	0.0049	0.3664	0.3649	0.0044	0.3370	0.3187	0.0046
p^{10}	0.5457	0.5251	0.0051	0.5528	0.5570	0.0033	0.5970	0.5996	0.0036
$\overline{y}_{it-1}$	1.0170			0.7384			0.7482		

	z_{it} AR(1)			Last 3 years		
	True	Mean	SD	True	Mean	SD
γ	1	0.9922	0.0249	1	0.9571	0.0320
ρ	0.6	0.5688	0.0205	0.6	0.6129	0.0376
δ_1	0.3	0.2608	0.0350	0.3	0.1695	0.0490
δ_2	0.2	0.2215	0.0415	0.2	0.1437	0.0606
σ_u^2	1	1.0231	0.0376	1	0.9910	0.0509
APE_z	0.3840	0.3812	0.0091	0.4187	0.4013	0.0139
$\text{APE}_{y_{-1}}$	0.2304	0.2185	0.0080	0.2512	0.2570	0.0168
p^{00}	0.6703	0.6668	0.0042	0.6600	0.6595	0.0071
p^{01}	0.3297	0.3332	0.0042	0.3400	0.3405	0.0071
p^{10}	0.6030	0.6045	0.0034	0.5895	0.5795	0.0051
$\overline{y}_{it-1}$	0.6798			0.7371		

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 IV. Among these variables, only the race variables, Black and Hispanic, are time-invariant. Covariates similar to those used in Table IV are common in the empirical labor supply literature. Examples include Chib and Jeliazkov (2006) for study of married women labor force participation, Shaw (1994) for research on the persistence of female labor supply, and 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

$$
\begin{aligned}
y_{it} &= \max\{0, \mathbf{z}_{it}\gamma + y_{it-1}\rho + c_i + u_{it}\} \\
c_i &= y_{i0}\delta_1 + \mathbf{z}_i\delta_2 + \alpha_i \\
u_{it} &\sim \text{normal}(0, \sigma_u^2)
\end{aligned}
\tag{16}
$$

where $t = 0$ corresponds to the initial year 1987, $y_{it} = \text{Hours}_{it}$ and $\mathbf{z}_{it} = (\text{Age}_{it}, \text{Educ}_{it}, \text{Income}_{it}, \text{Married}_{it}, \text{Northcentral}_{it}, \text{Northeast}_{it}, \text{South}_{it}, \text{Old}_{it}, \text{Young}_{it}, \text{School}_{it}, \text{SMSA}_{it}, \text{Unemp}_{it}, \text{Urban}_{it}, (\text{Age}_{it})^2, \text{Age}_{it}{}^*\text{Educ}_{it}, (\text{Age}_{it})^{2*}\text{Educ}_{it}, \text{Hours}_{it-1})$. Specification of $\mathbf{z}_i$ requires some discussion.

DOI: 10.1002/jae

Table IV. Summary statistics[a]

Variable	Explanation	Mean	SD
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 members' income/1000	27.8697	77.5495
Married	1 if married, 0 otherwise	0.6408	0.4798
North-central	1 if in the north-central 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 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

[a] The sample consists of a cross-section of 1115 individuals over 6 years.

As mentioned earlier, $\mathbf{z}_i$ can be the row vector of all (nonredundant) explanatory variables in all time periods. That is, $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots z_{iT})$ and each z_{it} can be multidimensional as in Wooldridge (2005). Alternatively it can be $\mathbf{z}_i = \overline{\mathbf{z}}_i$, where $\overline{\mathbf{z}}_i$ is the average of $\mathbf{z}_{it}$ over all the time periods as in Chib and Jeliazkov (2006). As pointed by a referee, conditioning on $\mathbf{y}_{i0}$ changes the partial correlation between c_i and different elements of $\mathbf{z}_i$ as the correlation between $\mathbf{y}_{i0}$ and $\mathbf{z}_{it}$ is higher for smaller t. To balance between flexibility and computational convenience, we use a combination of both the Wooldridge (2005) and the Chib and Jeliazkov (2006) methods by specifying the unobserved heterogeneity to have different correlations with the Income and Unemployed variables in different years. For other covariates, the average of $\mathbf{z}_{it}$ over different years is used. Finally, we leave the distribution of α_i unspecified, and will use the semiparametric Bayesian estimation procedure we propose to estimate model (16). The resulting parameter estimates are reported in Tables V–VII. The chain is of length 5000 draws following burn-ins of 1000 draws. For comparison purpose, we also implement the estimation method proposed by Wooldridge (2005) and the results are presented side by side with the results from the semiparametric Bayesian method in the tables.

6.1. Evidence of Non-normality

Figure 1 shows the estimated predictive density and mean adjusted estimated density using the Wooldridge (2005) method for error term α_i, respectively, 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 is highly non-normal. Interestingly, the estimated density of the dependent variable also demonstrates two modes at 0 and 40, which correspond to the two biggest modes of the estimated predictive density for error term α_i.[17] This

[17] It is quite intuitive that 0 and 40 are two modes for the working hours. While all people in the dataset either worked or did not work, meaning that 0 serves also as a truncation point, there are 1679 out of 7805 observations who worked

DOI: 10.1002/jae

Table V. Results from the main equation

Variable	Semipara. Bayesian		Wooldridge	
	Mean	SD	Estimate	SE
Age	−17.1609	5.8806	20.9112	62.0059
Educ	2.7382	1.1587	6.9962	6.8640
Income	−0.0014	0.0021	−0.0013	0.0022
Married	−1.6441	0.5789	−1.5378	0.6264
North-central	4.5166	2.1442	4.4158	2.0973
Northeast	7.1109	2.8630	7.4688	2.8310
South	2.4227	1.6907	1.8795	2.0352
Old	−1.6349	0.4848	−1.8653	0.4904
Young	−3.5942	0.3908	−3.6680	0.3998
School	−3.5532	0.6851	−3.4932	0.7170
SMSA	1.2385	0.6346	1.1809	0.6237
Unemp	−0.0994	0.3902	−0.0895	0.3896
Urban	2.2375	1.0176	1.7895	1.0475
$(\text{Age})^2$	2.7675	1.7282	−3.4182	10.4039
(Age)*(Educ.)	−1.6470	0.8536	−4.4614	4.5905
$(\text{Age})^2$*(Educ.)	0.2890	0.1722	0.7485	0.7691
Hours_{-1}	0.6604	0.0180	0.6623	0.0177

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. On the other hand, as one can see from Table VI, the estimated mean of α_i from using the Wooldridge (2005) method is -137.5178 with a standard error 223.4611, meaning that it is insignificant. However, the mean of estimated predictive density for α_i is around 30 from Figure 1. Such a difference could be attributed to the fact that while most of the parameter estimates from the two methods are close, there are some estimates that are quite far from each other. For example, in the auxiliary equation, the mean estimate for the coefficient of 'mean of age' is -16.6546 from the semiparametric Bayesian method, while the estimate is 60.8567 from the Wooldridge method, thus driving the constant term in the auxiliary equation, which is the mean of α_i in the Wooldridge model, to the left. To compare the shapes of the density estimates for α_i from both methods, we readjust the mean of the estimated density using the Wooldridge method to be at the mode of the predictive density. As illustrated in Figure 1, these two density estimates have similar shapes.

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 0.6604 with a standard deviation of 0.0180. This indicates that the number of average working

more than 40 hours. As pointed by a referee, having 40 hours as another mode in addition to 0 may raise some issues in appropriateness of the Tobit model, as the Tobit model allows 0 to be a focal point but no other focal point. On the other hand, the Tobit model has been widely used in studying female labor supply using similar datasets to ours. See, for example, Smith and Blundell (1986), Newey (1987), and Blundell and Smith (1994). Thus we use the Tobit model in the empirical illustration; a more general model is left for future research.

DOI: 10.1002/jae

Table VI. Results from the auxiliary equation

Variable	Semipara. Bayesian		Wooldridge	
	Mean	SD	Estimate	SE
Mean of Age	−16.6546	8.1162	60.8657	164.0453
Mean of Educ.	3.8939	2.8354	10.7897	17.9147
Income88	$-1.0177e^{-5}$	$1.5219e^{-5}$	$-1.2608e^{-5}$	$1.5806e^{-5}$
Income89	$1.3971e^{-7}$	$1.6995e^{-6}$	$4.7255e^{-5}$	$1.7988e^{-6}$
Income90	$-8.0455e^{-6}$	$1.4549e^{-7}$	$-9.3265e^{-6}$	$1.4932e^{-5}$
Income91	$-3.5600e^{-5}$	$1.3892e^{-5}$	$-3.8300e^{-5}$	$1.4522e^{-5}$
Income92	$2.8021e^{-6}$	$2.1973e^{-6}$	$2.9475e^{-6}$	$2.2203e^{-6}$
Income93	$-1.8894e^{-5}$	$1.0158e^{-5}$	$-2.1936e^{-5}$	$1.0683e^{-5}$
Mean of Married	3.3093	0.7854	3.5075	0.8871
Mean of North-central	−3.9694	2.2472	−4.0261	2.1978
Mean of Northeast	−7.4293	2.9827	−8.0558	2.9367
Mean of South	−1.2816	1.7215	−1.0045	2.1319
Mean of Old	1.5426	0.5357	1.7372	0.5647
Mean of # of Young	−0.5010	0.5002	−0.4549	0.5246
Mean of School	3.8450	1.2181	2.8492	1.4067
Mean of SMSA	−1.8822	0.8101	−1.7956	0.7992
Unemp88	−0.2519	0.5167	−0.4378	0.5658
Unemp89	−0.6222	0.6903	−0.7996	0.7626
Unemp90	1.1868	0.6470	1.4897	0.7038
Unemp91	1.9344	0.5954	1.3464	0.5824
Unemp92	−0.3055	0.6361	−0.0088	0.6732
Unemp93	−1.3410	0.5691	−1.1023	0.5879
Mean of Urban	−2.4430	1.1143	−2.0211	1.1856
Mean of $(\text{Age})^2$	−1.5916	0.6605	−1.6110	0.6942
Mean of (Age)*(Educ.)	−0.0672	0.8666	−0.5652	0.8738
Mean of $(\text{Age})^2$*(Educ)	5.0340	2.4630	−8.6898	27.5972
Black	−1.0037	2.0046	−6.1194	12.1161
Hispanic	−0.0314	0.3669	0.8672	2.0363
Hours_0	−0.1479	0.0149	0.1558	0.0167
Cons.	n.a.	n.a.	−137.5178	223.4611
σ_α^2	n.a	n.a	21.8033	2.7503

hours per week a woman supplies this year will be highly influenced by the number she supplied in the previous year. Specifically, Table VII 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 0.59 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 *et al*. (1988). Empirically, 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 have roughly the same age composition) is 0.216 using the fixed-effects approach and 0.427 using the OLS. The difference may arise for 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 start from 1987. As documented by Blau (1998), over the past 25 years woman's labor supply not only has increased on the extensive

DOI: 10.1002/jae

Table VII. Average partial effects and average transition probabilities

Variable	Semipara. Bayesian		Wooldridge	
	Mean	SD	Estimate	SE
Age	0.2642	0.9053	0.2092	1.7452
Educ.	0.3390	0.3689	0.3570	0.5040
Income	−0.0012	0.0020	−0.0012	0.0020
Married	−0.7523	0.3134	−1.0985	0.5771
North-central	1.0623	0.4029	1.4097	1.9388
Northeast	0.7546	0.2465	1.0305	2.6820
South	0.6596	0.5449	0.7926	1.8738
Old	−1.8055	0.4431	−1.6420	0.4511
Young	−3.3919	0.3696	−3.2290	0.3735
School	−0.2049	0.0423	−0.2689	0.6460
SMSA	0.3371	0.1663	0.4584	0.5736
Unemp.	−0.0387	0.1441	−0.0440	0.3572
Urban	0.9933	0.5694	1.4712	0.9558
Hours_{-1}	0.5881	0.0146	0.5830	0.0200
p^{01}	0.8609	0.0074	0.6957	0.0332
p^{00}	0.1391	0.0074	0.3043	0.0332
p^{10}	0.0005	0.0005	0.0488	0.0067

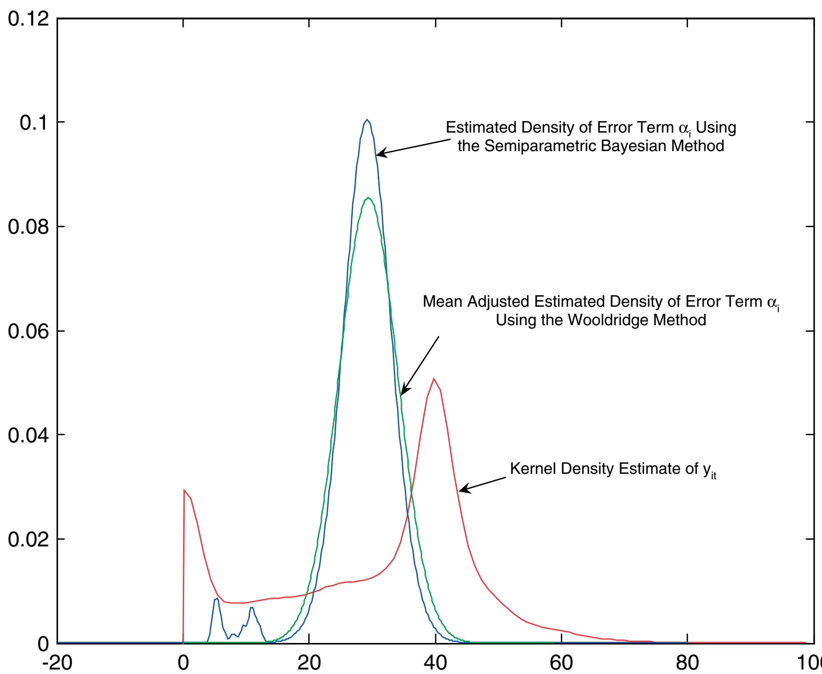

Figure 1. Estimated predictive density for α_i. This figure is available in color online at www.interscience.wiley.com/journal/jae

DOI: 10.1002/jae

(participation) margin, but also on the intensive (hours of work) margin. Third, Shaw's sample only consists of white women, while our data have 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 (2006), 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 3.40 hours per week and an additional school-age child (who is at least 6 years old) will reduce the woman's labor supply by 1.81 hours per week. These results are intuitive since younger children require more attention from the mother and hence reduce 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%. 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 0.34 hours. More interestingly, from Table V, 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 twenties and early thirties, 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 north-central 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 or in a metropolitan 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 and metropolitan area compared with that of the rural area. Other variables like other family members' income and local unemployment do not have much effect on the woman's labor supply.

The parameter estimates for the auxiliary equation for the unobserved heterogeneity, which are reported in Table VI, 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 negative correlation between the unobserved heterogeneity and the initial condition. Second, age, other

DOI: 10.1002/jae

family members' income in the year of 1991, marital status, geographical areas, number of school-age children and unemployment rates in years 1991 and 1993 turn out to be correlated with the unobserved heterogeneity.

To assess the strength of state dependence, we also 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 (13) and (14) and the results are reported in Table VII. 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 0.8609. At the same time, the estimated average probability for her to remain out of the labor force is 0.1391. 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.

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 APEs for the two groups separately according to equation (12). The results from the semiparametric Bayesian estimation are reported in Table VIII. For a white woman, on average, the effect of an additional weekly working hours in the past year on the current year is 0.60 hour, while for a non-white woman this effect is only 0.54. This reflects that white women enjoy a more stable labor supply than that of non-white women. Also, the negative effect coming from greater family responsibilities like 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 the fact 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, years of education and other family members' income have roughly the same effects on both white and non-white women.

Since the Wooldridge approach only requires a small fraction of the time required by the more computationally intensive semiparametric Bayesian approach to obtain the point estimates for

Table VIII. Average partial effects by race

Variable	White		Non-white	
	Mean	SD	Mean	SD
Age	0.2932	0.9237	0.1336	0.8372
Educ.	0.3454	0.3757	0.3105	0.3386
Income	−0.0012	0.0020	−0.0011	0.0018
Married	−0.8190	0.3413	−0.4526	0.1922
North-central	1.2383	0.4689	0.2717	0.1178
Northeast	0.8357	0.2707	0.3902	0.1429
South	0.5951	0.4910	0.9491	0.7900
Old	−1.8390	0.4511	−1.6550	0.4075
Young	−3.4549	0.3765	−3.1088	0.3411
School	−0.2108	0.0436	−0.1788	0.0439
SMSA	0.3791	0.1868	0.1482	0.0803
Unemp.	−0.0399	0.1487	−0.0334	0.1250
Urban	1.0055	0.5758	0.9383	0.5458
Hours_{-1}	0.5990	0.0148	0.5390	0.0153

parameters and other quantities of economic interest like APEs and ATPs, we provide here a comparison between the results from the two methods. This kind of cost–benefit analysis can help empirical researchers make informed decisions on which method to use. We focus on APEs and ATPs as they are the most interesting economic quantities. The standard errors in the Wooldridge approach are obtained by the Delta method. From Table VII, we can see that all of the estimates for APEs and ATPs from both approaches have the same signs. Moreover, it is interesting to see that APEs are quite close for most of the variables. For ATPs, on the other hand, the results from both methods are a bit different. Thus we can conclude that the APEs from both methods are very

Table IX. Results using the last 4 years of data

Variable	Main equation		APE and ATP	
	Mean	SD	Mean	SD
Age	−15.1304	9.6171	3.1915	1.3843
Educ.	4.4729	2.4333	0.6398	0.4512
Income	0.0033	0.0032	0.0031	0.0029
Married	−1.5678	1.0585	−0.6745	0.4184
North-central	5.2496	2.9615	1.6521	0.5838
Northeast	12.2063	3.7661	1.2333	0.3468
South	3.3369	2.3340	0.7281	0.6779
Old	−2.8490	0.6212	−2.3198	0.4726
Young	−3.6199	0.5139	−2.8623	0.4720
School	−2.2384	0.9809	−0.1075	0.0514
SMSA	0.1566	0.9268	0.1521	0.2221
Unemp.	−0.3862	0.5447	−0.0620	0.1442
Urban	3.1351	1.1274	2.0844	0.7569
$(\text{Age})^2$	2.3489	2.9104	n.a.	n.a.
(Age)*(Educ.)	−3.3312	1.6875	n.a.	n.a.
$(\text{Age})^2$*(Educ.)	0.5961	0.3216	n.a.	n.a.
Hours_{-1}	0.6714	0.0357	0.5894	0.0266
p^{01}	n.a.	n.a.	0.8520	0.0108
p^{00}	n.a.	n.a.	0.1480	0.0108
p^{10}	n.a.	n.a.	0.0104	0.0059

Auxiliary equation

Variable	Mean	SD	Variable	Mean	SD
Mean of Age	−15.2915	11.0559	Unemp90	1.0614	0.5619
Mean of Educ.	−4.2236	3.9961	Unemp91	1.5603	0.6312
Income90	$7.1584e^{-6}$	$1.4188e^{-5}$	Unemp92	−0.6232	0.7553
Income91	$-3.3385e^{-5}$	$1.7020e^{-5}$	Unemp93	−0.2732	0.6260
Income92	$5.6635e^{-7}$	$2.3216e^{-6}$	Mean of Urban	−3.4665	1.1574
Income93	$-5.5409e^{-5}$	$1.1950e^{-5}$	Mean of $(\text{Age})^2$	−1.0468	0.7475
Mean of Married	3.9371	1.2346	Mean of (Age)*(Educ.)	0.3054	0.9765
Mean of North-central	−4.9470	3.0165	Mean of $(\text{Age})^2$*(Educ.)	3.7591	3.4085
Mean of Northeast	−11.4319	3.8582	Black	4.2035	2.6559
Mean of South	−2.8452	2.3874	Hispanic	−0.8291	0.4719
Mean of Old	2.5985	0.6359	Hours_0	0.1599	0.0336
Mean of # of Young	0.4297	0.7039			
Mean of School	−0.1815	1.5518			
Mean of SMSA	−0.9156	1.0790			

DOI: 10.1002/jae

similar though the underlying heterogeneity distribution can be different, and such a difference in the underlying heterogeneity distribution may contribute to the different ATPs.[18]

Finally, we repeat the empirical analysis using only the last 4 years of data (1990–1993 with 1989 as period 0). This practice will provide evidence on whether the current specification of the initial condition provides a good approximation. The results using the semiparametric approach are collected in Table IX. Again we focus on the discussion of APEs and ATPs. Comparing the corresponding columns in Tables VII and IX, we find that for all but one insignificant variable (Income), the signs of the APEs and ATPs remain the same. Quantitatively, the results are very close to each other for some variables like the APE of Hours_{-1}, which is also one of the quantities of main economic interest. However, for other variables like Age, the differences are larger. Thus we conclude that the two analyses yield qualitatively similar but quantitatively slightly different results and the current specification of the initial condition provides a good approximation.

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 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 dataset 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. The findings from our approach offer insight on the dynamics of female labor supply, and contribute to the empirical labor literature.

APPENDIX I: DETAILS OF MCMC ALGORITHMS

1. Draw σ_u^2 and β. The joint posterior distribution of $1/\sigma_u^2$ and β conditional on data and other parameters is

$$posterior(1/\sigma_u^2, \beta/y_{it}^*, c_i w_{it}) = (1/\sigma_u^2)^{\frac{N_1}{2}-1} e^{-R_1(1/\sigma_u^2)} \prod_{i=1}^{n}\prod_{t=1}^{T}\left[\frac{1}{\sqrt{\sigma_u^2}}\exp\left(-\frac{1}{2\sigma_u^2}(y_{it}^* - w_{it}\beta - c_i)^2\right)\right]$$

[18] Of course, this does not necessarily mean that the estimates for ATPs from the semiparametric Bayesian method are better than those from the Wooldridge approach, as the Tobit model we specify here could also be subject to some misspecification, as discussed in footnote 17.

DOI: 10.1002/jae

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

$$\widehat{\beta} = \text{inv}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}' w_{it}\right) \times \left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}'(y_{it}^* - c_i)\right) \tag{A.1}$$

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

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

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

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

$$\text{normal}\left(\widehat{\beta}, \text{inv}\left(\frac{1}{\sigma_u^2}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}' w_{it}\right)\right) \tag{A.3}$$

2. Draw c_i. The posterior is

$$\prod_{t=1}^{T} \exp\left(-\frac{1}{2\sigma_u^2}(y_{it}^* - w_{it}\beta - c_i)^2\right) \exp\left(-\frac{1}{2\sigma_i^2}(c_i - h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta - \mu_i)^2\right) \tag{A.4}$$

Therefore, we update c_i by drawing from normal $\left(c_i^*, \text{inv}\left(\dfrac{T}{\sigma_u^2} + \dfrac{1}{\sigma_i^2}\right)\right)$ where $c_i^* = \text{inv}\left(\dfrac{T}{\sigma_u^2} + \dfrac{1}{\sigma_i^2}\right) \times \left(\dfrac{1}{\sigma_u^2}\sum_{t=1}^{T}(y_{it}^* - w_{it}\beta) + \dfrac{1}{\sigma_i^2}(h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \mu_i)\right)$.

3. Draw δ. The posterior distribution of δ is

$$\begin{aligned}&\text{posterior}(\delta|h(\mathbf{y}_{i0}, \mathbf{z}_i), c_i, \mu_i, \sigma_i^2)\\ &\propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma_i^2}(c_i - h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta - \mu_i)^2\right)\end{aligned} \tag{A.5}$$

Define

$$\widehat{\delta} = \text{inv}\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}h(\mathbf{y}_{i0}, \mathbf{z}_i)'h(\mathbf{y}_{i0}, \mathbf{z}_i)\right) \times \left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}h(\mathbf{y}_{i0}, \mathbf{z}_i)'(c_i - \mu_i)\right) \tag{A.6}$$

DOI: 10.1002/jae

We update δ by drawing from normal $\left(\widehat{\delta}, \text{inv}\left(\sum_{i=1}^{n} \frac{h(\mathbf{y}_{i0}, \mathbf{z}_i)' h(\mathbf{y}_{i0}, \mathbf{z}_i)}{\sigma_i^2}\right)\right)$.

APPENDIX II. BAYESIAN NONPARAMETRIC ESTIMATION

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

$$q(\alpha_i|\theta_i) = q(\alpha_i|\mu_i, \sigma_i^2) = \phi(\alpha_i|\mu_i, \sigma_i^2) \tag{A.7}$$

where $\theta_i = (\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$ is a sample from some unknown distribution P on the space (Θ, Γ) and Γ is a σ-field of subsets of Θ. Conditional on P, the density for α_i is

$$s(\alpha_i|P) = \int q(\alpha_i|\theta_i) \mathrm{d}P \tag{A.8}$$

This gives us a nonparametric class of distributions. How rich this class is depends on the choice of q and P. For Bayesian analysis, we need to specify a prior for P, which in turn implies a prior for $s(\alpha_i|P)$. If the prior for P has unit mass at some point θ^*, then $s(\alpha_i|P)$ is just equal to $q(\alpha_i|\theta^*)$, which is the same as a parametric specification of s. If P is assumed to have a Dirichlet process prior as introduced by Ferguson (1973, 1974), then it gives a prior structure for any random smooth density.

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 τP_0 be a finite non-null measure on (Θ, Γ), where P_0 is a proper base probability distribution and τ is a precision parameter. Then, a stochastic process P is a Dirichlet process if, for any given partition, $A_1, \ldots, A_q$ of the parameter space Θ, the random vector $(P(A_1), \ldots, P(A_q))$ has a Dirichlet distribution with a parameter vector $(\tau P_0(A_1), \ldots, \tau P_0(A_q))$. We denote $\mathrm{DP}(\tau P_0)$ for a Dirichlet process with base measure τP_0 for the rest of the paper.

One important feature of the Dirichlet process is that it selects a discrete distribution with probability one. The discreteness of the probability distribution selected by the Dirichlet process seems to be unsuitable for modeling 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 *et al.* (1999). This can be seen most clearly in the representation derived by Sethuraman (1994). Let $w_1, w_2, \ldots$ be i.i.d. Beta$(1, \tau)$ and let $p_1 = w_1$, $p_j = w_j \prod_{i=1}^{j-1}(1 - w_i)$, $j > 1$. Finally, let $\theta_1, \theta_2, \ldots$ be i.i.d. from P_0. Then if $P(d\theta) = \sum_{j=1}^{\infty} p_j \delta_{\theta_j}(d\theta)$, then $P \sim \mathrm{DP}(\tau P_0)$, where $\delta_{\theta_j}(A) = 1$ if $\theta_j \in A$ and $\delta_{\theta_j}(A) = 0$ otherwise. With (A.7) and (A.8), this implies

$$s(\alpha_i|P) = \sum_{j=1}^{\infty} p_j q(\alpha_i|\theta_j) = \sum_{j=1}^{\infty} p_j \phi(\alpha_i|\mu_j, \sigma_j^2) = f(\cdot) \tag{A.9}$$

This demonstrates that the infinite countable mixture of normals is just an alternative representation of the density for α_i under the Dirichlet process prior. Furthermore, Escobar (1994) shows

DOI: 10.1002/jae

that with the θ_is as the latent variables, integrating P over its prior distribution gives the sequence of $\theta_i, i = 1, \ldots, n$ as follows:

$$\theta_i|\theta_1, \theta_2, \ldots, \theta_{i-1} = \begin{cases} =\theta_j \text{ with probability} \dfrac{1}{\tau + i - 1} \\ \sim P_0 \text{ with probability} \dfrac{\tau}{\tau + i - 1} \end{cases} \tag{A.10}$$

where $\theta_1 \sim P_0$. As a result, α_is are partitioned into m_c groups such that all α_is in the same group have the same θ_i while those in different groups differ. These m_c distinct values θ_i are a sample from P_0. The degree of clustering is determined by τ. Small values of τ will tend to generate a sample with all the data points located in a few clusters, whereas larger values of τ will give a more diverse sample.

Also, using the Dirichlet process prior leads to a useful set of conditional distributions. Let ξ_j, $j = 1, 2, \ldots, m_c$ denote the m_c distinct values of θ_i and

$$\theta^{(i)} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n) \tag{A.11}$$

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

$$P|\theta^{(i)} \sim DP\left(\tau P_0 + \sum_{j=1}^{m_c^{(i)}} n_j^{(i)} \delta(\xi_j^{(i)})\right) \tag{A.12}$$

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

$$\theta_i|\theta^{(i)} \sim E(P|\theta^{(i)}) \sim \frac{\tau}{\tau + n - 1} P_0 + \frac{1}{\tau + n - 1} \sum_{j=1}^{m_c^{(i)}} n_j^{(i)} \delta(\xi_j^{(i)}) \tag{A.13}$$

is the prior distribution of θ_i conditional on $\theta^{(i)}$ and P_0.

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

$$dP_0(\mu_i, \sigma_i^2) \propto \text{normal}(\mu_0, \tau_0 \sigma_i^2) IG\left(\frac{N_0}{2}, \frac{R_0}{2}\right) \tag{A.14}$$

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

$$\tau \sim \text{gamma}(d_1, d_2). \tag{A.15}$$

One way of sampling θ_i from its posterior distribution and determining the number of components in approximating a mixture of normals m_c is to apply Bayes' theorem to the density

DOI: 10.1002/jae

given in equation (A.13) and draw from the resulting posterior density directly. But sampling θ_i from its posterior distribution in this model can be eased by introducing the configuration vector $S = \{S_1, \ldots, S_n\}$. That is, $S_i = j$ if and only if $\theta_i = \xi_j$. With this setting sampling θ_i is equivalent to sampling S and ξ_j, $j = 1, 2, \ldots, m_c$. West *et al.* (1994) propose an efficient algorithm as follows:

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

$$q(S_i = j|\xi^{(i)}, P_0) \propto \begin{cases} \tau q_t(\alpha_i|\mu_0, (1+\tau_0)R_0/n_0, n_0) & \text{if } j = 0 \\ n_j^{(i)}\phi(\alpha_i|\xi_j^{(i)}) & \text{if } j > 0 \end{cases} \tag{A.16}$$

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

2. Sampling ξ_j, $j = 1, 2, \ldots, m_c$ from the conditional distribution

$$\begin{aligned} q(\xi_j|\xi^{(i)}, S, P_0) &\propto \prod_{\{\alpha_i:S_i=j\}} \phi(\alpha_i|\xi_j)dP_0 \\ &\propto \text{normal}(\mu_1, \tau_1\sigma_j^2)IG\left(\frac{N_3}{2}, \frac{R_3}{2}\right) \end{aligned} \tag{A.17}$$

where

$$\mu_1 = \frac{\tau_0 \sum_{\{\alpha_i:S_i=j\}} \alpha_i + \mu_0}{\tau_0 n_j + 1}, \tau_1 = \frac{\tau_0}{\tau_0 n_j + 1}, N_3 = N_0 + n_j$$

$$R_3 = R_0 + \frac{n_j\left(\frac{1}{n_j}\sum_{\{\alpha_i:S_i=j\}} \alpha_i - \mu_0\right)^2}{\tau_0 n_j + 1} + \sum_{\{\alpha_i:S_i=j\}}\left(\alpha_i - \frac{1}{n_j}\sum_{\{\alpha_i:S_i=j\}} \alpha_i\right)^2$$

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 τ is given by

$$\tau \sim z \times \text{gamma}(d_1 + m_c, d_2 - \log\eta) + (1 - z) \times \text{gamma}(d_1 + m_c - 1, d_2 - \log\eta) \tag{A.18}$$

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

Thus, with a set of α_i, we updated θ_i for each i, the number of components in the countable infinite mixture of normals m_c, and τ, the precision parameter in the base measure.

Predictive Densities

In the Bayesian framework, a useful and informative way to draw implications from the unknown densities of the error term α_i is to study its predictive distribution. Parallel to equation (A.13), conditional on $(\theta_1, \ldots, \theta_n)$, for the new unit $i = n + 1$, we have

$$\theta_{n+1} | (\theta_1, \ldots, \theta_n) \sim \frac{\tau}{\tau + n} P_0 + \frac{1}{\tau + n} \sum_{j=1}^{m_c} n_j \delta(\xi_j) \tag{A.19}$$

where n_j is the number of θ_is taking the value ξ_j. Thus, the distribution of α_{n+1} conditional on the data can be rewritten as

$$\begin{aligned} q(\alpha_{n+1} | \theta_1, \ldots, \theta_n) = & \frac{\tau}{\tau + n} q_t(\alpha_{n+1} | \mu_0, (1 + \tau_0) R_0 / n_0, n_0) \\ & + \frac{1}{\tau + n} \sum_{j=1}^{m_c} n_j \phi(\alpha_{n+1} | \xi_j) \end{aligned} \tag{A.20}$$

Thus the predictive distribution for the error term α_i $(i = n + 1)$ can be obtained as

$$q(\alpha_{n+1} | \text{data}) = \int q(\alpha_{n+1} | \theta_1, \ldots, \theta_n) \pi(\theta_1, \ldots, \theta_n | \text{data}) \mathrm{d}(\theta_1, \ldots, \theta_n) \tag{A.21}$$

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

$$\widehat{q}(\alpha_{n+1} | \text{data}) = \frac{1}{M} \sum_{i=1}^{M} q(\alpha_{n+1} | \theta_1^{(i)}, \ldots, \theta_n^{(i)}) \tag{A.22}$$

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

APPENDIX III: MCMC ALGORITHMS FOR RELATED MODELS

Wooldridge (2005) Model

As in Wooldridge (2005), this model is a slight modification of the above semiparametric model:

$$\begin{aligned} y_{it} &= \max\{0, \mathbf{z}_{it}\gamma + \mathbf{g}(\mathbf{y}_{it-1})\rho + c_i + u_{it}\} \\ c_i &= h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \alpha_i \\ \alpha_i &\sim \text{normal}(\mu, \sigma^2) \end{aligned} \tag{A.23}$$

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

1. Conditional on y_{it}, w_{it}, x_i, β, Δ, σ^2 and σ_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 $y_{it} > 0$, $y_{it}^* = y_{it}$.

DOI: 10.1002/jae

2. Conditional on y_{it}^*, c_i and w_{it}, update σ_u^2 and β in one block. Using the improper flat prior for β and the independent gamma $\left(\frac{N_1}{2}, \frac{R_1}{2}\right)$ prior for $1/\sigma_u^2$, that is: $1/\sigma_u^2 \propto (1/\sigma_u^2)^{\frac{N_1}{2}-1} e^{-R_1(1/\sigma_u^2)}$,
(a) draw $1/\sigma_u^2$ from gamma $\left(\dfrac{N_1+nT}{2}, \dfrac{R_1+\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}^*-w_{it}\widehat{\beta}-c_i)^2}{2}\right)$, where $\widehat{\beta} = \text{inv}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}' w_{it}\right) \times \left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}'(y_{it}^*-c_i)\right)$ and
(b) update β from a normal distribution with mean $\widehat{\beta}$ and variance $\text{inv}\left(\frac{1}{\sigma_u^2}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}' w_{it}\right)$.
3. Conditional on y_{it}^*, w_{it}, x_i, β, δ, σ^2, σ_u^2, update c_i by drawing from a normal distribution with mean c_i^* and variance $\text{inv}\left(\frac{T}{\sigma_u^2}+\frac{1}{\sigma^2}\right)$, 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 σ^2 and Δ in one block. Using the improper flat prior for Δ and independent gamma $\left(\frac{N_2}{2}, \frac{R_2}{2}\right)$ prior for $1/\sigma^2$, that is: $1/\sigma^2 \propto (1/\sigma^2)^{\frac{N_2}{2}-1} e^{-R_2(1/\sigma^2)}$,
(a) draw $1/\sigma^2$ from gamma $\left(\dfrac{N_2+n}{2}, \dfrac{R_2+\sum_{i=1}^{n}(c_i-x_i\widehat{\Delta})^2}{2}\right)$, where $\widehat{\Delta} = \text{inv}\left(\sum_{i=1}^{n} x_i' x_i\right) \times \left(\sum_{i=1}^{n} x_i' c_i\right)$, and
(b) update Δ from a normal distribution with mean $\widehat{\Delta}$ and variance $\text{inv}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i' x_i\right)$.

Dynamic Tobit Model with the Lagged Uncensored Variable

The model with lagged uncensored variable as studied by Hu (2002) is the following:

$$\begin{aligned} y_{it}^* &= \mathbf{z}_{it}\gamma + \mathbf{g}(\mathbf{y}_{it-1}^*)\rho + c_i + u_{it} \\ &= w_{it}^*\beta + c_i + u_{it} \\ c_i &= h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \alpha_i \end{aligned} \tag{A.24}$$

where $w_{it}^* = (\mathbf{z}_{it}, \mathbf{g}(\mathbf{y}_{it-1}^*))$ and u_{it} is assumed to be normal variable with mean 0 and variance σ_u^2. In order to estimate the model, we have to make the additional assumption that $y_{i0}^* = y_{i0}$, that is, all observations of the dependent variable in the initial period are uncensored, which is also made in Hu (2002). Instead of imposing any distributional assumptions for α_i, we approximate it using an infinite mixture of normals. We have the following algorithm:

1. Conditional on y_{it}, w_{it}^*, $h(\mathbf{y}_{i0}, \mathbf{z}_i)$, β, δ, μ_i, σ_i^2 and σ_u^2 but marginalized over c_i, y_{it}^* (sequentially from y_{i1}^* to y_{iT}^*) is updated from a normal distribution with mean $w_{it}^*\beta + h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \mu_i$ 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 σ_u^2 and β in one block. Using the improper flat prior for β and the independent gamma $\left(\frac{N_1}{2}, \frac{R_1}{2}\right)$ prior for $1/\sigma_u^2$, that is: $1/\sigma_u^2 \propto (1/\sigma_u^2)^{\frac{N_1}{2}-1} e^{-R_1(1/\sigma_u^2)}$,

DOI: 10.1002/jae

(a) draw $1/\sigma_u^2$ from gamma $\left(\frac{N_1 + nT}{2}, \frac{R_1 + \sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it}^* - w_{it}^*\widehat{\beta} - c_i)^2}{2}\right)$, where $\widehat{\beta} = \text{inv}\left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}^{*'} w_{it}^*\right) \times \left(\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}^{*'}(y_{it}^* - c_i)\right)$ and

(b) update β from a normal distribution with mean $\widehat{\beta}$ and variance inv $\left(\frac{1}{\sigma_u^2}\sum_{i=1}^{n}\sum_{t=1}^{T} w_{it}^{*'} w_{it}^*\right)$.

3. Conditional on y_{it}^*, w_{it}^*, $h(\mathbf{y}_{i0}, \mathbf{z}_i)$, β, δ, μ_i, σ_i^2, σ_u^2, update c_i by drawing from a normal distribution with mean c_i^* and variance inv $\left(\frac{T}{\sigma_u^2} + \frac{1}{\sigma_i^2}\right)$, where $c_i^* = \text{inv}\left(\frac{T}{\sigma_u^2} + \frac{1}{\sigma_i^2}\right) \times \left(\frac{1}{\sigma_u^2}\sum_{t=1}^{T}(y_{it}^* - w_{it}^*\beta) + \frac{1}{\sigma_i^2}(h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \mu_i)\right)$.
4. Conditional on c_i, $h(\mathbf{y}_{i0}, \mathbf{z}_i)$, μ_i and σ_i^2 update δ. Using the improper flat prior for δ, we update δ by drawing from a normal distribution with mean $\widehat{\delta}$ and variance inv $\left(\sum_{i=1}^{n}\frac{h(\mathbf{y}_{i0}, \mathbf{z}_i)'h(\mathbf{y}_{i0}, \mathbf{z}_i)}{\sigma_i^2}\right)$, where

$$\widehat{\delta} = \text{inv}\left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}h(\mathbf{y}_{i0}, \mathbf{z}_i)'h(\mathbf{y}_{i0}, \mathbf{z}_i)\right) \times \left(\sum_{i=1}^{n}\frac{1}{\sigma_i^2}h(\mathbf{y}_{i0}, \mathbf{z}_i)'(c_i - \mu_i)\right) \tag{A.25}$$

5. Finally, recover $\alpha_i = c_i - h(\mathbf{y}_{i0}, \mathbf{z}_i)'\delta$ and update the number of components m_c in approximating a mixture of normals and the mean and variance of the normal denoted by μ_i and σ_i^2 for each i using the Bayesian nonparametric density estimation method described in Appendix II.

Dynamic Model with Unknown Distribution of u_{it}

The model with unknown distribution of u_{it} is the following:

$$\begin{aligned} y_{it}^* &= \mathbf{z}_{it}\gamma + \mathbf{g}(\mathbf{y}_{it-1})\rho + h(\mathbf{y}_{i0}, \mathbf{z}_i)\delta + \alpha_i + u_{it} \\ &= \widetilde{w}_{it}\widetilde{\beta} + v_{it} \\ y_{it} &= \mathbf{I}(y_{it}^* > 0)y_{it}^* \end{aligned} \tag{A.26}$$

where $\widetilde{w}_{it} = (\mathbf{z}_{it}, \mathbf{g}(\mathbf{y}_{it-1}), h(\mathbf{y}_{i0}, \mathbf{z}_i))$, $\widetilde{\beta} = (\gamma', \rho', \delta')'$, $v_{it} = \alpha_i + u_{it}$ and the density of v_{it} is unknown. We approximate its density using an infinite mixture of normals. As stated in Section 2, with this approach each v_{it} is assumed to be distributed as normal $(\mu_{it}, \sigma_{it}^2)$. Our semiparametric Bayesian estimation consists of two main parts. At each iteration, in the first part, using the current value of parameters, we can recover the error term v_{it} through the relationship $v_{it} = y_{it}^* - \widetilde{w}_{it}\widetilde{\beta}$. After recovering the error terms, we can use a Bayesian approach as detailed in Appendix II 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 a mixture of normals and the mean and variance of the normal denoted by μ_{it} and σ_{it}^2 for each it. In the second part of each iteration, we update the model parameters. Incorporating other parts of the model, we have the following algorithm:

1. Conditional on y_{it}, $\widetilde{w}_{it}$, $\widetilde{\beta}$, μ_{it} and σ_{it}^2, y_{it}^* is updated from a normal distribution with mean $\widetilde{w}_{it}\widetilde{\beta} + \mu_{it}$ and variance σ_{it}^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}^*, $\widetilde{w}_{it}$, μ_{it} and σ_{it}^2, update $\widetilde{\beta}$. Using the improper flat prior for $\widetilde{\beta}$, update $\widetilde{\beta}$ from a normal distribution with mean inv $\left(\sum_{i=1}^n \sum_{t=1}^T \frac{1}{\sigma_{it}^2}\widetilde{w}_{it}'\widetilde{w}_{it}\right) \times \left(\sum_{i=1}^n \sum_{t=1}^T \frac{1}{\sigma_{it}^2} w_{it}'(y_{it}^* - \mu_{it})\right)$ and variance inv $\left(\sum_{i=1}^n \sum_{t=1}^T \frac{1}{\sigma_{it}^2} w_{it}' w_{it}\right)$.
3. Finally, recover $v_{it} = y_{it}^* - \widetilde{w}_{it}\widetilde{\beta}$ and update the number of components m_c in approximating a mixture of normals and the mean and variance of the normal denoted by μ_{it} and σ_{it}^2 for each it using the Bayesian nonparametric density estimation method described in Appendix II.

ACKNOWLEDGEMENTS

We are grateful to Editor, Hashem Pesaran, and two referees for their constructive comments that greatly improve the paper. All remaining errors are our own.

REFERENCES

Ahn SC, Schmidt P. 1995. Efficient estimation of models for dynamic panel data. *Journal of Econometrics* **68**: 5–27.

Albert J, Chib S. 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**: 669–679.

Anderson TW, Hsiao C. 1982. Formulation and estimation of dynamic models using panel data. *Journal of Econometrics* **18**: 67–82.

Arellano M, Bover O. 1995. Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics* **68**: 29–51.

Arellano M, Carrasco R. 2003. Binary choice panel data models with predetermined variables. *Journal of Econometrics* **115**: 125–157.

Arellano M, Honoré B. 2001. Panel data models: some recent developments. In *Handbook of Econometrics*, Vol. 5, Heckman J, Leamer E (eds). North-Holland: Amsterdam; 3229–3296.

Blau F. 1998. Trends in the well-being of american woman, 1970–1995. *Journal of Economic Literature* **36**: 112–165.

Blundell R, Bond S. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* **87**: 115–143.

Blundell R, Smith RJ. 1991. Initial conditions and efficient estimation in dynamic panel data models. *Annals d'Economie et de Statistique* **20/21**: 109–123.

Blundell R, Smith RJ. 1994. Coherency and estimation in simultaneous models with censored or qualitative dependent variables. *Journal of Econometrics* **64**: 355–373.

Browning M. 1992. Children and household economic behavior. *Journal of Economic Literature* **30**: 1434–1475.

Chamberlain G. 1980. Analysis of covariance with qualitative data. *Review of Economic Studies* **47**: 225–238.

Chiappori PA, Salanie B. 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*, Vol. 5, Heckman J, Leamer E (eds). North-Holland: Amsterdam; 3569–3652.

Chib S, Greenberg E. 1996. Markov chain Monte Carlo simulation methods in econometrics. *Econometric Theory* **12**: 409–431.

Chib S, Hamilton BH. 2002. Semiparametric Bayes analysis of longitudinal data treatment models. *Journal of Econometrics* **110**: 67–89.

Chib S, Jeliazkov I. 2006. Inference in semiparametric dynamic models for binary longitudinal data. *Journal of the American Statistical Association* **101**: 685–700.
Escobar MD. 1994. Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association* **89**: 268–277.
Escobar MD, West M. 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**: 577–588.
Farber HS. 1994. The analysis of interfirm worker mobility. *Journal of Labor Economics* **12**: 554–593.
Ferguson TS. 1973. A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**: 209–230.
Ferguson TS. 1974. Prior distributions on spaces of probability measures. *Annals of Statistics* **2**: 615–629.
Ferguson TS. 1983. Bayesian density estimation by mixtures of normal distributions. In *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His Sixtieth Birthday*, Rizvi H, Rustagi J (eds). Academic Press: New York; 287–302.
Geweke J, Keane M. 2001. Computationally intensive methods for integration in econometrics. In *Handbook of Econometrics*, Vol. 5. Heckman J, Leamer E (eds). North-Holland: Amsterdam; 3463–3568.
Ghosal S, Ghosh JK, Ramamoorthi RV. 1999. Posterior consistency of Dirichlet mixtures in density estimation. *Annals of Statistics* **27**: 143–158.
Griffin JE, Steel MFJ. 2004. Semiparametric Bayesian inference for stochastic frontier models. *Journal of Econometrics* **123**: 121–152.
Gronau R. 1973. The effect of children on the housewife's value of time. *Journal of Political Economy* **81**: 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 H, Kozumi H. 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*, Manski C, McFadden D (eds). MIT Press: Cambridge, MA; 179–195.
Heckman J. 1981b. Heterogeneity and state dependence. In *Studies in Labor Markets*, Rosen S (ed.). University of Chicago Press: Chicago, IL; 91–140.
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. 2002. Nonlinear models with panel data. *Portuguese Economic Journal* **1**: 163–179.
Honoré B, Hu L. 2004. Estimation of cross sectional and panel data censored regression models with endogeneity. *Journal of Econometrics* **122**: 293–316.
Hotz V, Kydland F, Sedlacek G. 1988. Intertemporal preferences and labor supply. *Econometrica* **56**: 335–360.
Hsiao C. 2003. *Analysis of Panel Data* (2nd edn). Cambridge University Press: Cambridge, UK.
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 TJ, Rouse CE. 1995. Labor-market returns to two-and four-year college. *American Economic Review* **85**: 600–614.
Lo AY. 1984. On a class of Bayesian nonparametric estimates: I. Density estimates. *Annals of Statistics* **12**: 351–357.
Loudermilk MS. 2007. Estimation of fractional dependent variables in dynamic panel data models with an application to firm dividend policy. *Journal of Business and Economic Statistics* **25**: 462–472.
Nakamura A, Nakamura M. 1992. The econometrics of female labor supply and children. *Econometric Reviews* **11**: 1–71.
Nakamura A, Nakamura M. 1994. Predicting female labor supply: effects of children and recent work experience. *Journal of Human Resources* **29**: 304–327.
Newey WK. 1987. Specification tests for distributional assumptions in the Tobit model. *Journal of Econometrics* **34**: 125–145.
Olsen RJ. 1994. Fertility and the size of the US labor force. *Journal of Economic Literature* **32**: 60–100.

DOI: 10.1002/jae

Sethuraman J. 1994. A constructive definition of dirichlet priors. *Statistica Sinica* **4**: 639–650.

Shaw K. 1994. The persistence of female labor supply: empirical evidence and implications. *Journal of Human Resources* **29**: 348–378.

Smith RJ, Blundell RW. 1986. An exogeneity test for a simultaneous equation Tobit model with an application to labor supply. *Econometrica* **54**: 679–685.

West M, Müller P, Escobar MD. 1994. Hierarchical priors and mixture models, with application in regression and density estimation. In *Aspects of Uncertainty: A Tribute to D. V. Lindley*, Smith AFM, Freeman PR (eds). Wiley: Chichester; 363–386.

Wooldridge JM. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

Wooldridge JM. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* **20**: 39–54.

DOI: 10.1002/jae