Simulation Estimation of Dynamic Panel Tobit Models

Sheng-Kai Chang *

Department of Economics

Wayne State University

Abstract

In this paper I propose a computationally practical simulation estimator for large categories of the dynamic panel Tobit model with complicated dependence structures. I first apply the sequential decomposition methods introduced by Hendry and Richard (1992) to obtain the tractable simulated log-likelihood function of the dynamic panel Tobit model. I then maximize this log-likelihood function simulated through procedures based on a recursive algorithm formulated by Geweke-Hajivassiliou-Keane as well as a Gibbs sampling simulator. Monte Carlo experiments indicate that my simulation estimator performs strikingly well, even for a small simulation size.

JEL classification: C15; C23; C24.

Keywords: Panel Tobit model; GHK simulator; Gibbs sampling; Random effects.

*Corresponding author. Department of Economics, Wayne State University, Rm 2125, Faculty Administration Building, 656 W. Kirby, Detroit, MI 48202, USA. Tel.: +1-313-577-0603; fax: +1-313-577-0149. Email address: schang@wayne.edu (Sheng-Kai Chang).

1 Introduction

In the literature of panel data models, one critical issue is the estimation of limited dependent variable (LDV) models characterized by the presence of lagged dependent variables and serially correlated errors. A dynamic panel Tobit model is a leading example. The conventional techniques used in the estimation of linear panel data models are not applicable to the estimation of dynamic panel Tobit models due to the nature of the Tobit structure. Furthermore, the introduction of lagged dependent variables makes conventional estimation techniques even more difficult to apply.

One possible method for estimating the dynamic panel Tobit model is the fixed effects approach. Although the fixed effects model is valid under weak restrictions on the unobserved individual heterogeneity, there are some limitations of using this method in the context of a dynamic panel Tobit model. For example, Honoré (1993) estimates the panel Tobit model with lagged observed dependent variables through the fixed effects approach by creating the orthogonal conditions for method of moments estimators. A set of identification conditions for Honoré's model is provided by Honoré and Hu (2002). However, under such identification conditions, the model with time-dummies variables can not be estimated consistently.

Using the fixed effects approach creates even more problematic issues for the estimation of the dynamic panel Tobit model with lagged latent dependent variables. Hu (2003) estimates the censored panel data model with lagged latent dependent variables by applying the fixed effects method. Hu creates orthogonal conditions from the dynamic panel censored model for the Generalized Method of Moments (GMM) estimation. In addition to the same identification problem as Honoré's, the first observation of each individual is required to be uncensored with positive probability in order to construct the orthogonal conditions.

The other method proposed to handle the dynamic panel Tobit model is the random effects approach. By specifying the distribution of the error conditional on the regressors, the random effects estimators can be obtained through maximizing the corresponding likelihood function. A classic paper by Lillard and Willis (1978) as well as a recent paper by Geweke and Keane (2000) apply random effects techniques to panel earnings data without censoring. However, the likelihood function of the dynamic panel *censored* model is usually intractable since the dimension of an integral involved in its calculation is as large as the number of censoring periods in the model. Under such circumstances, simulation-based inference methods can be extremely useful.

The impact of simulation methods on the analysis of LDV models is profound, especially under recent advances in computing technologies. Various simulation estimation methods and procedures for drawing random variables have been proposed in the econometric literature. For instance, Lerman and Manski (1981) and Gourieroux and Monfort (1993) suggest the simulated maximum likelihood method (SML), McFadden (1989) performs the method of simulated moments (MSM), and Hajivassiliou and McFadden (1998) present the method of simulated scores (MSS).

Different simulators have also been proposed for simulating multinomial probabilities in LDV models. Among multivariate normal probability simulators, Hajivassiliou et al. (1996) suggest that the Geweke-Hajivassiliou-Keane (GHK) simulator is, in terms of root mean square errors, the most reliable simulator for approximating the multivariate normal distribution and its derivative among thirteen simulators they examined. In addition to the GHK simulator, Hajivassiliou and McFadden (1998) also indicate that the Gibbs sampling simulator has a satisfactory converging rate in terms of simulation bias.

This paper studies a practical, operational and versatile maximum likelihood procedure for general dynamic panel Tobit models using the GHK as well as the Gibbs sampling simulator. To do so, I obtain a tractable simulated log-likelihood function of the dynamic panel Tobit model by using the sequential factorization methods introduced by Hendry and Richard (1992).

The random effects approach is attractive in several aspects. The time-invariant, time-varying, and time-dummies variables can be incorporated in the model and they can be estimated consistently using my simulation estimator. In addition, identification is obtained in a straightforward manner, for example, under the assumption of normally distributed errors (Olsen, 1978). Most importantly, my estimation method allows for complicated dynamics.[1] The introduction of lagged latent dependent variables, lagged observed dependent variables, possibly with more than one lag, is straightforward. It is also easy to accommodate serial correlations in errors. Modifying the estimator to accommodate such specifications is done fairly easily, and in an intuitive manner. Furthermore, the predictions of the model can be generated by the estimation results through the random effects approach for dynamic panel Tobit models.

This simulation estimator is applied to study the earnings dynamics and the impact of Title VII of the 1964 Civil Rights Act on the convergence of the black-white earnings gap in Chang (2002). The rich dynamic structure of the earnings process can be identified from the top-coded data set by using my simulation estimation method.

An overview of the paper is as follows. Section 2 proposes a practical simulation estimator for dynamic panel Tobit models. I discuss the likelihood function simulation through the GHK and Gibbs sampling simulators and then show the consistency and asymptotic normality of the simulation estimator. Section 3 presents Monte Carlo experiments of the simulation estimator. Section 4 concludes the paper and proposes areas for future extensions.

[1] It has become substantial to allow rich dynamics in economic models since Heckman (1981) pointed out the importance of distinguishing true state dependence from spurious state dependence. For instance, Hyslop (1999) incorporates state dependence, serial correlation, and individual heterogeneity into the labor force participation model of married women.

2 Practical Simulation Estimators for Dynamic Panel Tobit Models

2.1 Basic Framework

The cross-sectional non-dynamic Tobit model introduced by Tobin (1958) can be represented as follows:

$$\begin{aligned} y_i^\star &= x_i\beta + \epsilon_i \\ y_i &= max\{y_i^\star, r\} \end{aligned} \tag{1}$$

where $y^\star$ is a latent dependent variable, x is a vector of exogenous variables, y represents an observed dependent variable, and r is a known constant. Without loss of generality, r can be rescaled to zero.

In a dynamic panel data framework, the Tobit model is described as:

$$\begin{aligned} y_{it}^\star &= x_{it}\beta + y_{i,t-1}^\star\lambda + \epsilon_{it} \\ y_{it} &= max\{y_{it}^\star, 0\} \\ \epsilon_{it} &= d_i + u_{it}, \quad t = 1, \ldots, T \quad i = 1, \ldots, N \end{aligned} \tag{2}$$

Note that model (2) is characterized by lagged *latent* dependent variables.[2] The component d_i is an unobserved individual specific random disturbance which is constant over time, and u_{it} is an idiosyncratic error which varies across time and individuals. Throughout this section, I assume that d_i and u_{it} are Gaussian conditional on $x_{i1}, \cdots, x_{iT}$.

[2]The dynamic panel Tobit model with lagged observed dependent variables can be obtained from (2) by replacing the first line of (2) with $y_{it}^\star = x_{it}\beta + y_{i,t-1}\lambda + \epsilon_{it}$. My simulation estimator is also applicable to the model with lagged observed dependent variables (see the results of Monte Carlo experiments section). Although I present the likelihood simulation based on the model (2), the likelihood simulation based on the model with the lagged observed dependent variables can be easily modified from the simulated likelihood function of (2); see Monte Carlo experiments section for details.

I consider two different specifications for the error terms in (2). The first is the *random effects* specification and is characterized by the following properties:

$$\begin{aligned} E[d_i|x_{i1},\cdots,x_{iT}] &= 0, \quad E[u_{it}|d_i,x_{i1},\cdots,x_{iT}]=0 \\ E[d_i^2|x_{i1},\cdots,x_{iT}] &= \sigma_d^2, \quad E[u_{it}^2|d_i,x_{i1},\cdots,x_{iT}]=\sigma_u^2 \\ E[d_i d_j|x_{i1},\cdots,x_{iT}] &= 0, \quad E[u_{it}u_{js}|d_i,x_{i1},\cdots,x_{iT}]=0 \end{aligned} \tag{3}$$

for all $i \neq j$ and $t \neq s$. The first line of (3) is the strict exogeneity assumption.[3] The homoskedasticity assumption is imposed on the second line of (3). The third line of (3) states that d_i and u_{it} are independent across all i and t.

By virtue of (3), the covariance structure of ϵ_{it} can be written as:

$$E[\epsilon_{it}\epsilon_{i,t-j}|x_{i1},\cdots,x_{iT}] = \begin{cases} \sigma_\epsilon^2 & for \quad j=0 \\ \rho\sigma_\epsilon^2 & for \quad j\neq 0 \end{cases} \tag{4}$$

where ρ is the fraction of the variance of ϵ due to the individual random effects, and $\sigma_\epsilon^2 = \sigma_d^2 + \sigma_u^2$. The correlation between the errors term ϵ_{it} for any two different time periods is always ρ. Under this specification, the covariance structure of the random effects model can also be represented as:

$$\Sigma_{RE} = \begin{bmatrix} \sigma_d^2+\sigma_u^2 & \sigma_d^2 & \cdots & \sigma_d^2 \\ \sigma_d^2 & \sigma_d^2+\sigma_u^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_d^2 \\ \sigma_d^2 & \cdots & \sigma_d^2 & \sigma_d^2+\sigma_u^2 \end{bmatrix} = \begin{bmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon^2 & \cdots & \rho\sigma_\epsilon^2 \\ \rho\sigma_\epsilon^2 & \sigma_\epsilon^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho\sigma_\epsilon^2 \\ \rho\sigma_\epsilon^2 & \cdots & \rho\sigma_\epsilon^2 & \sigma_\epsilon^2 \end{bmatrix} \tag{5}$$

where Σ_{RE} is a $T \times T$ matrix.

The second specification I consider is the *random effects plus AR(1)* specification. The only difference in this specification and the random effects model is that the error

[3] The assumption $E[d_i|x_{i1},\cdots,x_{iT}]=0$ can be replaced by the assumption $E[d_i|x_{i1},\cdots,x_{iT}]=\omega_0+\omega_1 x_{i1}+\cdots+\omega_T x_{iT}$ with little change to my simulation method. This case will be discussed in details in section 2.7. See also Chamberlain (1984) and Jacubson (1988).

term ϵ_{it} in (2) is

$$\begin{aligned} \epsilon_{it} &= d_i + v_{it} \\ v_{it} &= \zeta v_{i,t-1} + u_{it}. \end{aligned} \tag{6}$$

where u_{it} and d_i continue to satisfy all of the conditions in (3).

The covariance structure of (6) can be described as:

$$\Sigma_{RE+AR(1)} = \frac{\sigma_u^2}{1-\zeta^2} \begin{bmatrix} 1 & \zeta & \cdots & \zeta^{T-1} \\ \zeta & 1 & \ddots & \zeta^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \zeta^{T-1} & \zeta^{T-2} & \cdots & 1 \end{bmatrix} + \sigma_d^2 J_T \tag{7}$$

where $\Sigma_{RE+AR(1)}$ is a $T \times T$ matrix and J_T is a $T \times T$ matrix of ones.[4]

I assume that the stationary assumption

$$|\zeta| < 1 \tag{8}$$

is also satisfied for the random effects plus AR(1) errors model.[5] Both covariance structures (5) and (7) will be used for deriving the simulation estimator.

2.2 Simulation Estimation

Let I_{it} be a censored indicator function, such that

$$I_{it} = \begin{cases} 1 & for \quad y_{it}^{\star} > 0 \\ 0 & for \quad y_{it}^{\star} \leq 0 \end{cases}$$

Therefore, for individual i, if $I_{it} = 1$ then $y_{it}^{\star}$ is observed and $y_{it} = y_{it}^{\star}$. On the other hand, $y_{it}^{\star}$ is censored and its value is not observed (i.e. $y_{it} = 0$) if $I_{it} = 0$.

[4]Since the variance parameters σ_d^2 and σ_v^2 cannot be identified for Probit model, the normalization condition $\sigma_d^2 + \sigma_v^2 = 1$ is used. However, for the Tobit model, both σ_d^2 and σ_v^2 can be identified without the normalization condition $\sigma_d^2 + \sigma_v^2 = 1$.

[5]The unit root of the AR(1) process is excluded from this stationary assumption.

In order to obtain the tractable simulated likelihood function, individual $i's$ likelihood function L_i is decomposed into products of sequential conditional density functions. Thus, the likelihood function for individual i is represented as:

$$L_i = \int_{-\infty}^{0} \cdots \int_{-\infty}^{0} \prod_{t=1}^{T} g(y_{it}, y_{it}^{\star} | y_{i,t-1}, y_{i,t-1}^{\star}) dy_i^{\star} \tag{9}$$

where the number of dimensions for $y_i^{\star}$ is the same as the number of censoring periods for individual i, and $y_{i,0}$ and $y_{i,0}^{\star}$ are assumed to be known values.[6]

Following the suggestions by Hendry and Richard (1992) and Lee (1999), the joint density function of y_{it} and $y_{it}^{\star}$, $g(y_{it}, y_{it}^{\star} | y_{i,t-1}, y_{i,t-1}^{\star})$ in (9), can be further decomposed into the products of different forms of conditional densities:

$$\begin{aligned} & g(y_{it}, y_{it}^{\star} | y_{i,t-1}, y_{i,t-1}^{\star}) \\ & = q_1(y_{it} | y_{it}^{\star}, y_{i,t-1}, y_{i,t-1}^{\star}) \times h_1(y_{it}^{\star} | y_{i,t-1}, y_{i,t-1}^{\star}) \qquad (10) \\ & = q_2(y_{it} | y_{i,t-1}, y_{i,t-1}^{\star}) \times h_2(y_{it}^{\star} | y_{it}, y_{i,t-1}, y_{i,t-1}^{\star}) \qquad (11) \end{aligned}$$

where q_1 and q_2 are conventional Tobit likelihood functions, [7] and h_1 and h_2 are importance sampling densities from which the latent variables $y_{it}^{\star}$ are drawn. In addition, Hendry and Richard (1992) claim that the simulation estimator based on (11) is a better approach than (10) in terms of simulation efficiency.

Thus, the likelihood function for the dynamic panel Tobit model can be written as:

$$L = \prod_{i=1}^{N} L_i$$

The simulated likelihood function[8] based on the decomposition of (10) can be obtained by using an indicator function. Let $I_{(-\infty,0)}$ be the indicator function of the set $(-\infty, 0)$.

[6] y_{i0} and $y_{i0}^{\star}$ are assumed to be zero in my likelihood simulation.

[7] $q_1 = [f(y_{it} | y_{i,t-1}, y_{i,t-1}^{\star})]^{I_{it}} [P(I_{it} = 0 | y_{i,t-1}, y_{i,t}^{\star})]^{1-I_{it}}$ in (10), and $q_2 = [f(y_{it} | y_{i,t-1}, y_{i,t-1}^{\star})]^{I_{it}} [P(I_{it} = 0 | y_{i,t-1}, y_{i,t-1}^{\star})]^{1-I_{it}}$ in (11).

[8] The initial conditions $y_{i0}^{\star(r)}$ and y_{i0} are assumed to be zero throughout the likelihood simulation.

Then with a finite number of simulation runs R, the unbiased likelihood simulator for individual i is

$$\widetilde{L}_i = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}[f(y_{it}|y_{i,t-1}, y^{\star(r)}_{i,t-1})]^{I_{it}}[I_{(-\infty,0)}(y^{\star(r)}_{it})]^{1-I_{it}} \tag{12}$$

where $y^{\star(r)}_{it}$ in (12) is drawn from the conditional density function $h_1(y^{\star}_{it}|y_{i,t-1}\ , y^{\star(r)}_{i,t-1})$. Let $\widetilde{l}_i = ln(\widetilde{L}_i)$, then the simulated log-likelihood function with R simulation runs can be represented as:

$$\widetilde{l}_R = ln(\prod_{i=1}^{N}\widetilde{L}_i) = \sum_{i=1}^{N}\widetilde{l}_i \tag{13}$$

The likelihood simulation (13) is based on the frequency simulator. Due to the indicator function in (12), (13) is not a smooth function in the parameter space. Furthermore, Hajivassiliou et al. (1996) and Lee (1999) show that the simulation estimator based on (13) is inefficient in terms of the simulation variance compared with the GHK-based estimators.

The GHK simulator can do a much better job than the frequency simulator for dynamic panel Tobit models. The method behind the GHK simulator is to allow current information to feedback to the simulation procedure. That is, $y^{\star}_{it}$ is recursively drawn from the univariate conditional probability $h_2(y^{\star}_{it}|I_{it} = 0, y_{i,t-1}, y^{\star(r)}_{i,t-1})$ which is conditional on not only the past, but also the current sample information.

Thus, the simulated log-likelihood function based on the GHK simulator for individual i can be described as:

$$\widehat{L}_i = \frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T}[f(y_{it}|y_{i,t-1}, y^{\star(r)}_{i,t-1})]^{I_{it}}[P(I_{it} = 0|y_{i,t-1}, y^{\star(r)}_{i,t-1})]^{1-I_{it}} \tag{14}$$

(14) can be viewed as the simulated likelihood function corresponding to the decomposition of (11). By letting $\widehat{l}_i = ln(\widehat{L}_i)$, the simulated log-likelihood function with R simulation draws can be represented as:

$$\widehat{l}_R = ln(\prod_{i=1}^{N}\widehat{L}_i) = \sum_{i=1}^{N}\widehat{l}_i \tag{15}$$

Now, let $l_i = ln(L_i)$, where L_i is defined in (9). With this notation, the log-likelihood function simulated by (13) and (15) is

$$l = \sum_{i=1}^{N} l_i \tag{16}$$

I assume that the error terms in (2) have a normal distribution, that is, $\epsilon_i \sim N(0, \Sigma_{RE})$, where $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})'$ is a $T \times 1$ column vector, and so $E[\epsilon_i \epsilon_i' | x_{i1}, \cdots, x_{iT}] = \Sigma_{RE}$. Under this assumption, the simulation estimator of (16) based on the GHK simulator is expected to be useful for the dynamic panel Tobit models since it is well recognized that the GHK simulator is very accurate for the simulation of a multivariate normal distribution (Hajivassiliou et al., 1996).

In addition to the GHK simulator, the Gibbs sampling simulator is also known to have a theoretically excellent convergence rate in terms of simulation bias when it is applied to LDV models (Hajivassiliou and McFadden, 1998). The Gibbs sampling simulator is also easily implemented for the multivariate normal probability. Therefore, both the GHK and Gibbs sampling simulators will be implemented for simulating the log-likelihood function of the dynamic panel Tobit model.

2.3 The GHK Simulator

Let A be the Cholesky decomposition of Σ_{RE}, that is, $\Sigma_{RE} = AA'$, where A is a lower-triangular matrix:

$$A = \begin{bmatrix} A_{11} & 0 & \cdots & 0 \\ A_{21} & A_{22} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ A_{T1} & A_{T2} & \cdots & A_{TT} \end{bmatrix} \tag{17}$$

Given this structure, we can write $\epsilon_i = A\eta_i$, where $\eta_i \sim N(0, I)$ and $\eta_i = (\eta_{i1}, \ldots, \eta_{iT})'$.

Let θ be the vector of parameters of interest in the random effects model, i.e., $\theta = (\beta, \lambda, \rho, \sigma_\epsilon^2)$. Given the sample, suppose that for individual i the total number of instances of censoring is m_i and that censoring occurs at time $t_1, \ldots, t_{m_i}$. I then draw random variables $\xi_{it}^{(r)}$ from the uniform random number generator on [0,1], where $t = t_1, \ldots, t_{m_i}$, $i = 1, \ldots, N$ and $r = 1, \ldots R$. In order to guarantee the validity of the stochastic equicontinuity condition[9] for my simulation estimator, the random numbers are drawn once and kept fixed when θ varies.[10]

To simplify my exposition, I assume that $y_{i0}^{\star} = 0$ for all i. This assumption can be relaxed in empirical applications, see section 2.7 for details. Then, for the r^{th} simulation run, the latent variable $y^{\star}$ at censoring period t for individual i can be simulated by:

$$y_{it}^{\star(r)} = x_{it}\beta + y_{i,t-1}^{\star(r)}\lambda + A_t(\theta)\overline{\eta}_{it}^{(r)}(\theta) \tag{18}$$

where $\overline{\eta}_{it}^{(r)}(\theta) = (\eta_{i1}^{(r)}(\theta), \ldots, \eta_{it}^{(r)}(\theta), 0, \ldots, 0)'$ and A_t is the t^{th} row of (17). Also, $y_{i,t-1}^{\star(r)} = y_{i,t-1}$ if censoring does not occur at $t-1$.

In order to generate $y_{it}^{\star}$ from the conditional density $h_2(y_{it}^{\star}|I_{it} = 0, y_{i,t-1}, y_{i,t-1}^{\star(r)})$, the truncated normal random variables are drawn recursively for censored observations. Let Φ be the cumulative standard normal function. Then, the truncated normal random variables $\eta_{it}^{(r)}$ can be simulated or calculated by

$$\eta_{it}^{(r)} = \begin{cases} \Phi^{-1}(\xi_{it}^{(r)}\Phi(\frac{-x_{it}\beta - y_{i,t-1}^{\star(r)}\lambda - \sum_{k=1}^{t-1} A_{tk}\eta_{ik}^{(r)}}{A_{tt}})) & for \quad t \in \{t_1, \ldots, t_{m_i}\} \\ \frac{y_{it}^{\star(r)} - x_{it}\beta - y_{i,t-1}^{\star(r)}\lambda - \sum_{k=1}^{t-1} A_{tk}\eta_{ik}^{(r)}}{A_{tt}} & for \quad t \notin \{t_1, \ldots, t_{m_i}\} \end{cases} \tag{19}$$

where $y_{it}^{\star(r)} = y_{it}$ in the second row of (19) since $y_{it}^{\star(r)}$ is observable when $t \notin \{t_1, \ldots, t_{m_i}\}$. Therefore, $y_{it}^{\star(r)}$ can be calculated recursively through (18) for each time period and each person.

The simulated likelihood function (14) can be obtained through (18) and (19). Specifically, $f(y_{it}|y_{i,t-1}, y_{i,t-1}^{\star(r)})$ and $P(I_{it} = 0|y_{i,t-1}, y_{i,t-1}^{\star(r)})$ in (14) can be calculated by

$$f(y_{it}|y_{i,t-1}, y_{i,t-1}^{\star(r)}) = \frac{1}{A_{tt}}\phi(\frac{y_{it} - x_{it}\beta - y_{i,t-1}^{\star(r)}\lambda - \sum_{k=1}^{t-1} A_{tk}\eta_{ik}^{(r)}}{A_{tt}}) \tag{20}$$

[9]See Hajivassiliou and McFadden (1998).

[10]See McFadden (1989) for reference.

for $t \notin \{t_1, \ldots, t_{m_i}\}$, where ϕ is the standard normal density function, and

$$P(I_{it} = 0 | y_{i,t-1}, y^{\star(r)}_{i,t-1}) = \Phi(\frac{-x_{it}\beta - y^{\star(r)}_{i,t-1}\lambda - \sum_{k=1}^{t-1} A_{tk}\eta^{(r)}_{ik}}{A_{tt}}) \tag{21}$$

for $t \in \{t_1, \ldots, t_{m_i}\}$. Thus, by combining (20) and (21) with (14), a practical simulation estimator of θ can be obtained for maximizing the simulated log-likelihood function (15).

For the random effects plus AR(1) errors model, I can simply recalculate A as $\Sigma_{RE+AR(1)} = AA'$, where A is still the Cholesky decomposition of $\Sigma_{RE+AR(1)}$. Then, the simulation estimators of the random effects plus AR(1) errors model will be obtained through maximizing the simulated log-likelihood function by plugging A into (18) to (21).

2.4 Gibbs Sampling Simulator

Since Hajivassiliou and McFadden (1998) show that the Gibbs sampling simulator provides a theoretically superior convergence rate in terms of simulation bias, I present the likelihood simulation of (14) based on the Gibbs sampling simulator in this section.

Gibbs sampling is a Markov process that proceeds as follows. First, let A be the Cholesky decomposition of Σ_{RE} or $\Sigma_{RE+AR(1)}$. I then write $\epsilon_i = A\eta_i$, where $\eta_i \sim N(0, I)$ and $\eta_i = (\eta_{i1}, \ldots, \eta_{iT})'$. Thus, the Gibbs sampling recursive procedure for individuals $i = 1, \cdots, N$ in the Gibbs iterations $r = 1, \cdots, R$ for time $t = 1, \cdots, T$ of each simulation run can be defined as:

$$\eta^{(r)}_{it} = \begin{cases} \kappa^{(r)}_{it} + \sigma_t \Phi^{-1}(\xi^{(r)}_{it} \Phi(\frac{-\kappa^{(r)}_{it}}{\sigma_t})) \\ \frac{y^{\star(r)}_{it} - x_{it}\beta - y^{\star(r)}_{i,t-1}\lambda - \sum_{k=1}^{t-1} A_{Tk}\eta^{(r)}_{ik} - \sum_{k=t+1}^{T} A_{Tk}\eta^{(r-1)}_{ik}}{A_{Tt}} \end{cases} \tag{22}$$

The first line of (22) is for $t \in \{t_1, \ldots, t_{m_i}\}$, and the second line of (22) is for $t \notin \{t_1, \ldots, t_{m_i}\}$. Recall that the latent variable $y^\star_{it}$ is observed if $t \notin \{t_1, \ldots, t_{m_i}\}$. In addition, ξ is drawn from identically independent distribution uniform on $[0, 1]$. In (22), if the distribution of η_i is assumed to be $\eta_i \sim N(\mu, \Lambda)$, then

$$\kappa^{(r)}_{it} = \mu_t + \Lambda_{t,-t}\Lambda^{-1}_{-t,-t} \begin{bmatrix} \upsilon^{(r)}_{<t} \\ \upsilon^{(r-1)}_{>t} \end{bmatrix} \tag{23}$$

and

$$\sigma_t = [\Lambda_{tt} - \Lambda_{t,-t}\Lambda^{-1}_{-t,-t}\Lambda_{-t,t}]^{1/2} \tag{24}$$

where $\mu = [\mu_1, \cdots \mu_T]$, Λ is a $T \times T$ covariance matrix, $v_{<t} = [A_{T1}\eta^{(r)}_{i1} \quad A_{T2}\eta^{(r)}_{i2} \cdots A_{T,t-1}\eta^{(r)}_{i,t-1}]'$, and $v_{>t} = [A_{T,t+1}\eta^{(r-1)}_{i,t+1} A_{T,t+2}\eta^{(r-1)}_{i,t+2} \cdots A_{T,T}\eta^{(r-1)}_{iT}]'$. Also, $'-t'$ represents that the matrix is without its original t^{th} row or t^{th} column. For example, $\Lambda_{-r,-t}$ is a $(T-1)\times(T-1)$ matrix which represents the Λ matrix but without its r^{th} row and t^{th} column.

Since $\mu_t = 0$ for all t, and $\Lambda = I_T$ in (23) and (24),[11] (22) can be further simplified as:

$$\eta^{(r)}_{it} = \begin{cases} \Phi^{-1}(\xi^{(r)}_{it}\Phi(\frac{-x_{it}\beta - y^{\star(r)}_{i,t-1}\lambda - \sum_{k=1}^{t-1} A_{Tk}\eta^{(r)}_{ik} - \sum_{k=t+1}^{T} A_{Tk}\eta^{(r-1)}_{ik}}{A_{Tt}})) \\ \frac{y^{\star(r)}_{it} - x_{it}\beta - y^{\star(r)}_{i,t-1}\lambda - \sum_{k=1}^{t-1} A_{Tk}\eta^{(r)}_{ik} - \sum_{k=t+1}^{T} A_{Tk}\eta^{(r-1)}_{ik}}{A_{Tt}} \end{cases} \tag{25}$$

Again, the first line of (25) is for $t \in \{t_1, \ldots, t_{m_i}\}$, and the second line of (25) is for $t \notin \{t_1, \ldots, t_{m_i}\}$.[12]

R Gibbs iterations of $\eta^{(r)}_{it}$ can be generated through (25). The limiting distribution of η_{it} can be approximated by discarding the first h_R of R Gibbs iterations, and averaging the remaining $(R - h_R)$ iterations. For example, if $R = 1,000$ and $h_R = 200$, then the target distribution of η_{it} can be approximated by averaging the last 800 Gibbs iterations of $\eta^{(r)}_{it}$.

A starting value of $\eta^{(r)}_{it}$ is needed for the Gibbs sampling procedure. Since any sequence of $\eta^{(r)}_{it}$ with positive probability in the Markov Chain is qualified to be a starting value, I simply use one GHK draw of $\eta^{(r)}_{it}$ as the initial value.

The simulated log-likelihood function (15) based on the Gibbs sampling simulator is constructed by entering in the approximate distribution of η_{it} into (20) and (21), and the corresponding simulation estimator can then be obtained through maximizing this simulated log-likelihood function.

[11] I_T is the $T \times T$ identity matrix.

[12] The Gibbs sampling simulator for panel Probit Models is also discussed by Keane (1993).

2.5 Asymptotic Properties of Simulation Estimators

The consistency and asymptotic normality results established in this section are based on (15). First, I define the score and the simulated score function for dynamic panel Tobit models. Let

$$s_i = \frac{dl_i(\theta)}{d\theta} = \frac{L_{i\theta}(\theta)}{L_i(\theta)}$$

be the associated score function for individual i, where L_i is defined in (9), and $L_{i\theta}$ is the first derivative of the likelihood function L_i with respect to θ. The simulated score function, $\widehat{s}_i$, is defined in the same way as the score function:

$$\widehat{s}_i = \frac{d\widehat{l}_i}{d\theta} = \frac{\widehat{L}_{i\theta}(\theta)}{\widehat{L}_i}$$

where $\widehat{L}_i$ is defined in (14), and $\widehat{L}_{i\theta}$ is the first derivative of the simulated likelihood function $\widehat{L}_i$ with respect to θ.

The necessary conditions and the proof for consistency and asymptotic normality of general simulated score estimators in LDV models are provided by Hajivassiliou and McFadden (1998). Since the dynamic panel Tobit model is a special case of a LDV model, the consistency and asymptotic normality of my simulation estimator can be obtained directly from Hajivassiliou and McFadden's results. Some of the necessary conditions presented in Hajivassiliou and McFadden (1998) are even automatically satisfied for dynamic panel Tobit models.

I refer to either the GHK based or the Gibbs sampling based version at my simulated maximum likelihood estimator as the *SML-GHK* or the *SML-Gibbs* estimator. The following theorem develops the consistency and asymptotic normality of the maximum simulated likelihood estimator for the dynamic panel Tobit model.

Theorem 1 *Assume that the observations and simulators are independently and identically distributed. Assume also that the moment existence assumption in Hajivassiliou* *and McFadden (1998) is satisfied. Let* $\widehat{\theta}_N \equiv argmax_\theta \Sigma_{i=1}^N \widehat{l}_i(\theta)$*, where* $\widehat{l}_i(\theta)$ *is simulated by the GHK or Gibbs sampling simulator with an appropriate choice of the number of simulations specified in the proof and is obtained by requiring that the Monte Carlo random numbers are not redrawn when* θ *is changed. Then,* $\widehat{\theta}_N \longrightarrow \theta^\star$ *in probability, and* $\sqrt{N}(\widehat{\theta}_N - \theta^\star) \longrightarrow N(0, J^{-1}(J+Q)J^{-1})$ *in distribution, where* $\theta^\star$ *is the true value of* θ *in the parameter space* Θ*,* $J = -E_i[s_{i\theta}(\theta^\star)]$*,* $Q = E_i[(\widehat{s}_i(\theta^\star) - E_r[\widehat{s}_i(\theta^\star)])(\widehat{s}_i(\theta^\star) - E_r[\widehat{s}_i(\theta^\star)])']$*,* E_i *is an expectation over the distribution of observations, and* E_r *is an expectation over simulation draws.*

Proof : See Appendix A.

2.6 Estimating the Asymptotic Variance

For the asymptotic distribution to be useful in practice, J and $J+Q$ in Theorem 1 should be consistently estimated.

First, let $\Omega = J + Q$. An estimator of Ω is then

$$\widehat{\Omega} = \frac{1}{N}\Sigma_{i=1}^N[\widehat{s}_i(\widehat{\theta}_N)\widehat{s}_i(\widehat{\theta}_N)']$$

where $\widehat{\theta}_N$, obtained through maximizing the simulated likelihood with N observations, is a consistent estimator for $\theta^\star$. $\widehat{\Omega}$ is therefore also a consistent estimator of Ω. J can be estimated by:

$$\widehat{J} = \frac{1}{N}\Sigma_{i=1}^N \widehat{s}_{i\theta}(\widehat{\theta}_N)$$

$\widehat{J}$ is simply the Hessian matrix of the simulated log-likelihood function of the dynamic panel Tobit model. Therefore, the consistent estimator of the asymptotic variance of my simulation estimator can be written as $\widehat{J}^{-1}\widehat{\Omega}\widehat{J}^{-1}$.

The contribution of simulation noise Q to Ω goes to zero as $R \longrightarrow \infty$ for the GHK simulator and the asymptotic variance goes to J^{-1} (or Ω^{-1}). However, in practice, it is more accurate for simulation estimators to always use $\widehat{J}^{-1}\widehat{\Omega}\widehat{J}^{-1}$ as estimators of asymptotic variance in the finite sample simulation estimation.

2.7 Extension of the Basic Model

The extension of the basic model (2) is considered in this section. First, the assumption of independence of the exogenous variable x and the unobserved individual heterogeneity is relaxed. Then, the initial condition problem of dynamic panel Tobit models is discussed.

Following Chamberlain (1984), the unobserved individual heterogeneity can be assumed to be correlated with the exogenous variable x_{it} for all t in a linear way. Thus, let the error term ϵ_{it} in (2) be

$$\epsilon_{it} = c_i + u_{it}$$

and

$$c_i = \omega_0 + \omega_1 x_{i1} + \cdots + \omega_T x_{iT} + d_i$$

where d_i and u_{it} satisfy all conditions in (3). Under such a specification, a constant should not be included in x_{it} since it can not be identified from ω_0.[13] Moreover, time dummies are also excluded from x_{it} in c_i because they do not vary across i.[14] Otherwise, ω_0, ω_1, ..., ω_T along with β and λ in model (2) can be consistently estimated using the simulation estimation methods proposed above. However, the dimension of the parameters to be estimated will increase at the rate T. When the time period T or the dimension of x is small, computation is not an issue for estimating ω_0, ω_1, ..., ω_T, β and λ. Otherwise, following Wooldridge (2002), c_i can be further simplified as:

$$c_i = \omega_0 + \omega \overline{x}_i + d_i, \qquad \overline{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$$

[13]One way to include a constant in x_{it} is to assume the independence of time constant x_{it} and the unobserved individual heterogeneity c_i. Under such a circumstance, the parameter of time constant x_{it} can be consistently estimated.

[14]The time dummies can be included in the exogenous variables x_{it} in (2).

and ω_0 and ω can be consistently estimated by the proposed simulation estimation method.[15]

One advantage of using this specification is no matter how large the time period T is in the data set, the number of parameters to be estimated will only be affected by the dimension of x_i. For a panel data set with a large T, this specification method for unobserved individual heterogeneity will not only allow for correlation between c_i and x_i, but also make computation easier for parameter estimation.

The other issue that should be dealt with carefully in dynamic panel Tobit models is the initial condition problem. One possible answer for the initial condition problem is to assume that the initial value y_{i0} is nonrandom. However, this assumption is too strong since it implies that y_{i0} is independent with the unobserved individual heterogeneity c_i. The more realistic assumption is to specify the steady state conditional distribution of y_{i0} given x_i and c_i. Since the conditional distribution of c_i given x_i can be specified in a linear way as above, the simulated log-likelihood function $f(y_{i0}, y_{i1}, \cdots, y_{iT}|x_i)$ for individual i can be obtained directly if the conditional distribution of y_{i0} given x_i and c_i is known. However, even if the conditional distribution of y_{i0} given x_i and c_i is assumed to be at the steady state, it is still difficult to find the conditional distribution of y_{i0} given x_i and c_i.[16]

There are two methods proposed to solve this difficulty. The first one is by Heckman (1981). Heckman's approach is to approximate the conditional distribution of y_{i0} given x_i and c_i, and then multiply it with the known conditional density function of period 1 and after. For example, in the dynamic panel Tobit structures, the conditional distribution of y_{i0} given x_i and c_i can also be assumed to be normally distributed. Then, the log-likelihood function of $f(y_{i0}, y_{i1}, \cdots, y_{iT}|x_i)$ is obtained and the parameters are estimated by using the simulation estimation methods in the previous section.

The other method is proposed by Wooldridge (2002). Wooldridge's approach is to

[15]The time dummies are also not included in $\overline{x}_i$.

[16]See Hsiao (1986), section 7.4 for a further discussion.

specify the distribution of c_i given y_{i0} and x_i. For example, the unobserved individual heterogeneity c_i can be specified as the following

$$c_i = \omega_0 + \omega_1 x_{i1} + \cdots + \omega_T x_{iT} + \psi y_{i0} + d_i$$

Under such a specification, the distribution of y_{i0} given x_i and c_i remains unknown, and the simulation estimation is performed by maximizing the conditional log-likelihood function of $f(y_{i1}, \cdots, y_{iT}|y_{i0}, x_i)$.

For the dynamic panel Tobit model, although the conditional distribution of y_{i0} given c_i and x_i does not need to be specified for Wooldridge's approach, one advantage to using Heckman's approach for the initial condition problem is that the first period dependent variables are allowed to be censored. Since the conditional distribution of y_{i0} given x_i and c_i is approximated for a certain distribution, the censored initial values can be simulated through the same procedure as in the previous sections. On the other hand, Wooldridge's approach does not work if some initial period dependent variables are censored in the data set. Therefore, if all initial values can be observed in the data set, Wooldridge's method may be preferable since the distributional assumption of y_{i0} given x_i and c_i is not necessary for an estimation. However, the approximated conditional distribution of y_{i0} given x_i and c_i should be specified if some initial values are censored in the data set.[17]

The initial condition method of Heckman or Wooldridge can be combined with Chamberlain's approach of unobserved individual heterogeneity for the estimation of dynamic panel Tobit models. In addition, when the time period T is large, $\overline{x}_i$ can be used instead of x_{it} in the conditional distribution specification of c_i, especially when the dimension of x_i is also large.

[17] A good application for Wooldridge's method is dynamic panel Tobit models with lagged observed dependent variables. On the other hand, Heckman's approach can be applied to dynamic panel Tobit models with lagged latent dependent variables.

3 Monte Carlo Experiments

In this section, Monte Carlo experiments are performed in order to illustrate the finite sample properties for the SML estimator. All Monte Carlo experiments are based on the GHK simulator.[18]

For each experiment, 1,000 Monte Carlo data sets are generated from true parameter values. For each Monte Carlo data set, the simulation estimator is obtained through maximizing the simulated log-likelihood function. The results of the mean of the estimated estimators, the sample standard deviation, the estimated standard deviation, and the empirical probability of the 1% significant level of t statistics for the biases that are based on the empirical standard errors are reported for all experiments. Two types of models are considered in the experiments: the random effects model and the random effects plus AR(1) errors model.[19]

For the random effects model, the Monte Carlo data set is generated from (2), (3), and (4). Let $d_i \sim N(0,3)$, $u_{it} \sim N(0,2)$, $\beta = 1.2$ and $\lambda = 0.2$, in which case $\theta^\star = (\beta, \lambda, \rho, \sigma_\epsilon^2) = (1.2, 0.2, 0.6, 5)$, where $\theta^\star$ represents the true parameter values. The exogenous variable x_t is drawn from $N(0,1)$ and starting values for the optimizations for the random effects model are $\theta_{start} = (0.5, 0.5, 0.5, 2)$.[20] Thus, under this type of Monte Carlo experiment design, the average censoring fraction of a Monte Carlo data set over 1,000 samples is around 0.5 except for the constant x case (see Table 3 below). The likelihood simulation assumes a zero initial value through experiments, that is, $y_{i0}^\star = 0$ for all i.

In all Monte Carlo experiments, the number of people, N, is set at 250, and each

[18] I also have conducted some preliminary simulations using a Gibbs sampler. The results, which are available upon request, are similar to the ones with GHK.

[19] The optimization subroutine used for the Monte Carlo experiments is the CO procedure from the Gauss software and Newton's algorithm is used for maximization.

[20] All of these numbers, including the true parameter values and starting values, were arbitrarily chosen in my experiments.

individual is observed in eight time periods. Ten draws are used to form the GHK simulator, that is, $T = 8, R = 10$ and $N = 250$. In each Monte Carlo experiment, the linear maximum likelihood estimation (MLE) estimator which ignores the censoring problem is provided in order to provide a comparison with my estimator.

Table 1 shows the results for the SML-GHK estimator. My simulation estimator clearly outperforms the linear MLE estimator, especially for the estimators of β and σ_ϵ^2. Furthermore, the mean of the standard errors underestimates the sample standard deviation of the estimated parameters for the linear MLE estimator, while for my simulation estimator the mean of the standard errors provides a good estimator for the sample standard deviation of the estimated parameters.

In order to investigate the case that λ is near the unit root, the results of $\lambda = 0.9$ are presented in Table 2.[21] My SML-GHK estimator still outperforms the linear MLE estimator even when the coefficient of the lagged dependent variable is near the unit root, especially for the estimators of β and σ_ϵ^2.

One of the advantages to using random effects panel data models is that the constant parameter in the model can be identified. Table 3 reports the Monte Carlo evidence of a random effects model with a constant.[22] β is set to be -1. The average censoring fraction of the Monte Carlo data set over 1,000 samples is 0.68498 which is higher than Table 1 and Table 2. However, the estimation results of my simulation estimator in Table 3 are still attractive, especially for the estimators of β and σ_ϵ^2. As evidenced in Table 3, the time invariant constant in dynamic panel Tobit models can be successfully identified through my simulation estimators.

I next turn to the random effects plus AR(1) errors model. The data generating process is $\beta = 1.2$, $\lambda = 0.2$ and with AR(1) errors process $\zeta = 0.2$ for Table 4, and $\zeta = 0.9$ for Table 5. For simplicity, the normalization $\sigma_d^2 + \sigma_v^2 = 1$ is used, and σ_d^2 is set

[21] All other variables are left unchanged.

[22] In other words, $x_{it} = 1$ for all i and t in (2).

to be 0.8 in both Table 4 and Table 5.[23] Therefore, $d_i \sim N(0, 0.8)$, $u_{it} \sim N(0, 0.192)$, and $\theta^\star = (\beta, \lambda, \zeta, \sigma_d^2) = (1.2, 0.2, 0.2, 0.8)$ in Table 4. And $d_i \sim N(0, 0.8)$, $u_{it} \sim N(0, 0.038)$, and $\theta^\star = (\beta, \lambda, \zeta, \sigma_d^2) = (1.2, 0.2, 0.9, 0.8)$ in Table 5. The exogenous variable x_{it} is again generated from $N(0, 1)$, and starting values of the optimizations used in these experiments are set to be $\theta_{start} = (0.5, 0.5, 0.5, 0.5)$ in both Table 4 and Table 5.

The Monte Carlo evidence of the random effects plus AR(1) errors model with $\zeta = 0.2$ is presented in Table 4. The results of Table 4 are impressive for the mean of the estimated parameters. My SML-GHK estimator outperforms the linear MLE estimator, especially for the estimators of β and ζ. As before, the mean of the estimated standard errors underestimate the sample standard deviation of the estimated parameters for the linear MLE estimator. However, for my simulation estimator, the mean of the estimated standard errors provides a good estimator for the sample standard deviation of the estimated parameters. My simulation estimators can therefore identify the complicated dependence structures including λ, ρ, and ζ when the value of ζ is small.

The Monte Carlo evidence of the random effects plus AR(1) errors model with $\zeta = 0.9$ is presented in Table 5. Although the results of β and λ for my simulation estimator are still attractive, the estimation of ζ is much worse than the case with $\zeta = 0.2$ when the AR(1) errors are near the unit root process. Furthermore, the results of σ_d^2 are also less attractive. The estimate of ζ is downward-biased, while the estimate of σ_d^2 is upward-biased.[24] Therefore in this case the simulated maximum likelihood estimators do not provide a good estimator of ζ and σ_d^2. They are, however, still better estimates than those arising out of a linear MLE procedure.

[23] The normalization assumption is necessary for panel Probit models for identifying both σ_d^2 and σ_v^2. However, σ_d^2 and σ_v^2 can both be identified in dynamic panel Tobit models without this assumption. The $\sigma_d^2 + \sigma_v^2 = 1$ used here is just for simplicity.

[24] The direction of the bias for ζ and σ_d^2 found in the current context of a dynamic panel Tobit model is the same as the finding in Keane (1994) for a panel Probit model.

The model with lagged observed dependent variables is also examined with Monte Carlo experiments. This model can be characterized as:

$$\begin{aligned} y^{\star}_{it} &= x_{it}\beta + y_{i,t-1}\lambda + \epsilon_{it} \\ y_{it} &= max\{y^{\star}_{it}, 0\} \\ \epsilon_{it} &= d_i + u_{it}, \quad t = 1, \ldots, T \quad i = 1, \ldots, N \end{aligned} \tag{26}$$

The simulated likelihood function of model (26) can be obtained in the same way as model (2) except that $y_{i,t-1}$ is used in (18) to (21), instead of using $y^{\star(r)}_{i,t-1}$.[25]

The Monte Carlo evidence for model (26) is presented in Table 6. The Monte Carlo evidence shows that the SML-GHK estimators also performs as well as in Table 1 when the dynamic panel Tobit model is one with lagged observed dependent variables. Thus, as suggested by the Monte Carlo experiments, my simulation estimator can be applied to the dynamic panel Tobit models with either lagged latent dependent variables or lagged observed dependent variables.

Furthermore, in order to show that my simulation estimator is applicable to estimate the model with a flexible functional form of lagged dependent variables, the following AR(2) model is also considered:

$$\begin{aligned} y^{\star}_{it} &= y^{\star}_{i,t-1}\beta + y^{\star}_{i,t-2}\lambda + \epsilon_{it} \\ y_{it} &= max\{y^{\star}_{it}, 0\} \\ \epsilon_{it} &= d_i + v_{it} \quad t = 1, \ldots, T \quad i = 1, \ldots, N \end{aligned}$$

The results of Monte Carlo experiments for this model are presented in Table 7.[26]

[25] That is, $y^{\star(r)}_{it} = x_{it}\beta + y_{i,t-1}\lambda + A_t(\theta)\overline{\eta}^{(r)}_{it}(\theta)$ for (18), $\eta^{(r)}_{it} = \begin{cases} \Phi^{-1}(\xi^{(r)}_{it}\Phi(\frac{-x_{it}\beta - y_{i,t-1}\lambda - \sum_{k=1}^{t-1} A_{tk}\eta^{(r)}_{ik}}{A_{tt}})) & for \quad t \in \{t_1, \ldots, t_{m_i}\} \\ \frac{y^{\star(r)}_{it} - x_{it}\beta - y_{i,t-1}\lambda - \sum_{k=1}^{t-1} A_{tk}\eta^{(r)}_{ik}}{A_{tt}} & for \quad t \notin \{t_1, \ldots, t_{m_i}\} \end{cases}$ for (19), $f(y_{it}|y_{i,t-1}, y^{\star(r)}_{i,t-1}) = \frac{1}{A_{tt}}\phi(\frac{y_{it} - x_{it}\beta - y_{i,t-1}\lambda - \sum_{k=1}^{t-1} A_{tk}\eta^{(r)}_{ik}}{A_{tt}})$ for (20), and $P(I_{it} = 0|y_{i,t-1}, y^{\star(r)}_{i,t-1}) = \Phi(\frac{-x_{it}\beta - y_{i,t-1}\lambda - \sum_{k=1}^{t-1} A_{tk}\eta^{(r)}_{ik}}{A_{tt}})$ for (21).

[26] Here, let $y^{\star(r)}_{it} = y^{\star}_{i,t-1}\beta + y^{\star}_{i,t-2}\lambda + A_t(\theta)\overline{\eta}^{(r)}_{it}(\theta)$ for (18), $\eta^{(r)}_{it} =$

For a model with these complicated dependence structures, the estimation through the fixed effects approach is almost impossible. However, the proposed simulation estimator in Table 7 does perform as well as in Tables 1 through 6. Therefore, my simulation estimator is potentially able to estimate large categories of dynamic panel Tobit models with very complicated dependence structures, a task which is sometimes cumbersome or even impossible for the fixed effects approach.

4 Conclusion

This paper proposed a computationally practical simulation estimator for dynamic panel Tobit models based on maximizing simulated log-likelihood functions. This estimator allows for the estimation of dynamic panel Tobit models with complicated error structures. Monte Carlo results indicated that the simulation estimator performs well, even for a small simulation size.

One thing that should be improved in future work for my simulation estimators is when AR(1) errors are near a unit root process. Under such circumstances, the maximum simulated likelihood estimator is not a wise choice, as highlighted in Table 5. One way to solve this problem is to increase the number of observations. An alternate method is to use the method of simulated moments (MSM). For example, Keane (1994) successfully corrected the bias in the simulated maximum likelihood method by using the MSM estimator for the panel Probit model when AR(1) errors are near the unit root process. Thus, it is possible to fix the bias of my simulation estimators by using the MSM through

---

$\begin{cases} \Phi^{-1}(\xi_{it}^{(r)}\Phi(\frac{-y^{\star}_{i,t-1}\beta-y^{\star}_{i,t-2}\lambda-\sum_{k=1}^{t-1}A_{tk}\eta_{ik}^{(r)}}{A_{tt}})) & for \quad t\in\{t_1,\dots,t_{m_i}\} \\ \frac{y_{it}^{\star(r)}-y^{\star}_{i,t-1}\beta-y^{\star}_{i,t-2}\lambda-\sum_{k=1}^{t-1}A_{tk}\eta_{ik}^{(r)}}{A_{tt}} & for \quad t\notin\{t_1,\dots,t_{m_i}\} \end{cases}$ for (19), $f(y_{it}|y_{i,t-1},y^{\star(r)}_{i,t-1}) = \frac{1}{A_{tt}}\phi(\frac{y_{it}-y^{\star}_{i,t-1}\beta-y^{\star}_{i,t-2}\lambda-\sum_{k=1}^{t-1}A_{tk}\eta_{ik}^{(r)}}{A_{tt}})$ for (20), and $P(I_{it} = 0|y_{i,t-1},y^{\star(r)}_{i,t-1}) = \Phi(\frac{-y^{\star}_{i,t-1}\beta-y^{\star}_{i,t-2}\lambda-\sum_{k=1}^{t-1}A_{tk}\eta_{ik}^{(r)}}{A_{tt}})$ for (21). In addition, the normalization $\sigma_d^2+\sigma_v^2 = 1$ is used here for simplicity.

a suitable moment condition for dynamic panel Tobit models.

In addition, the normality assumption should be relaxed for my model specification. If this assumption is false, the SML estimator will no longer be consistent due to model misspecification. How to drop the normality assumption or how to replace it with a more flexible one is a promising extension of this paper.

Appendix A

Proof of Theorem 1. Olsen (1978) has already proved that the Tobit log-likelihood function is globally concave and has a unique maximum. The panel Tobit log-likelihood function (16) is the summation of Tobit log-likelihood functions and so is a strictly concave function. Therefore, the identification and concavity assumptions in Hajivassiliou and McFadden (1998) are fulfilled for dynamic panel Tobit models. Further, with this concave log-likelihood function, the consistency results of my simulation estimator can be obtained without assuming compactness of the parameter space Θ (Newey and McFadden, 1994).

Amemiya (1973) shows that the log-likelihood functions as well as their score functions for Tobit models are differentiable on Θ, implying that they are also continuous on Θ. These results are also valid for dynamic panel Tobit Models.

In addition, McFadden (1989) shows that the simulation residual process will be stochastic equicontinuous if the Monte Carlo random numbers used to simulate the log-likelihood function are not redrawn when θ is varied.

According to Corollary 1 in Hajivassiliou and McFadden (1998), the simulation bias will converge uniformly to zero, in probability, if the simulation process is unbiased, or if the bias of simulation is dominated by a positive $O(\frac{1}{\sqrt{N}})$ function which is independent of θ. The GHK and Gibbs sampling simulation estimators are not unbiased estimators. However, when the GHK estimator is based on R simulations with $\frac{R}{\sqrt{N}} \longrightarrow \infty$ of the sample size N, as $N \longrightarrow \infty$ and $R \longrightarrow \infty$, the simulation bias will converge to zero uniformly in θ. The simulation bias will also converge to zero uniformly in θ if the Gibbs sampling estimator is based on n_R Gibbs sampling runs for each R simulations with $\frac{n_R}{\ln N} \longrightarrow \infty$ as $N \longrightarrow \infty$ and $n_R \longrightarrow \infty$. Since $\widehat{l}_i(\theta)$ is simulated by using the GHK and Gibbs sampling simulators with big enough simulation runs, the simulation bias will converge to zero in probability uniformly.

Thus, if $\widehat{\theta}_N \equiv argmax_\theta \Sigma_{i=1}^N \widehat{l}_i(\theta)$, then $\widehat{\theta}_N$ solves $\Sigma_{i=1}^N \widehat{s}_i(\theta) = 0$. The consistency and asymptotic normality of my simulation estimator for dynamic panel Tobit models follows immediately from Theorem 1 in Hajivassiliou and McFadden (1998). *Q.E.D.*

Acknowledgements

I thank Arthur Goldberger, Bruce Hansen, Gautam Tripathi for their helpful discussions and suggestions. I am especially indebted to Yuichi Kitamura for his encouragement and invaluable guidance. All errors which remain are mine.

Reference

Amemiya, T., 1973, Regression analysis when the dependent variable is truncated normal, Econometrica 41, 997-1016.

Chamberlain, G., 1984, Panel data, in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics, Vol. 2 (Elsevier Science Publishers BV, Amsterdam) 1247-1318.

Chang, S., 2002, Simulation estimation of dynamic panel Tobit models with an application to the convergence of the black-white earnings gap and income dynamics, unpublished manuscript, Wayne State University.

Geweke, J. and M. Keane, 2000, An empirical analysis of earnings dynamics among men in the PSID: 1968-1989, Journal of Econometrics 96, 293-356.

Gourieroux, C. and A. Monfort, 1993, Simulation-based inference: a survey with special reference to panel data models, Journal of Econometrics 59, 5-33.

Hajivassiliou, V. and D. McFadden, 1998, The method of simulated scores for estimating limited dependent variable models, Econometrica 66, 863-896.

Hajivassiliou, V., McFadden, D. and P. Ruud, 1996, Simulation of multivariate normal rectangle probabilities and their derivatives: theoretical and computational results, Journal of Econometrics 72, 85-134.

Heckman, J., 1981, Statistical models for discrete panel data, In : C.F. Manski and D. McFadden, eds., Structural analysis of discrete panel data with econometric applications (MIT press) 114-178.

Hendry, D.F. and J.F. Richard, 1992, Likelihood evaluation for dynamic latent variables models, In : H.M. Amman, D.A. Belsley, and L.F. Pau, eds., Computational economics and econometrics (Kluwer, Amsterdam).

Honoré, B., 1993, Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables, Journal of Econometrics 59, 35-61.

Honoré, B. and L. Hu, 2002, Estimation of cross sectional and panel data censored regression models with endogeneity, unpublished manuscript, Princeton University.

Hsiao, C., 1986, Analysis of panel data (Cambridge, Cambridge University Press).

Hu, L., 2003, Estimation of a censored dynamic panel data model, Econometrica, forthcoming.

Hyslop, D.R., 1999, State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married woman, Econometrica 67, 1255-1294.

Jacubson, G., 1988, The sensitivity of labor supply parameter estimates to unobserved individual effects: fixed and random effects estimates in a nonlinear model using panel data, Journal of Labor Economics 6, 302-329.

Keane, M., 1993, Simulation estimation for panel data models with limited dependent variables, In: G.S. Maddala, C.R. Rao and H.D. Vinod, Eds., Handbook of Statistics Vol. 11 (Elsevier Science Publishers BV, Amsterdam) 545-571.

Keane, M., 1994, A computationally practical simulation estimator for panel data, Econometrica 62, 95-116.

Lee, L., 1999, Estimation of dynamic and ARCH Tobit models, Journal of Econometrics 92, 355-390.

Lerman, S. and C. Manski, 1981, On the use of simulated frequencies to approximate choice probabilities, In : C.F. Manski and D. McFadden, eds., Structural analysis of discrete panel data with econometric applications (MIT press), 305-319.

Lillard, L. and R. Willis, 1978, Dynamic aspects of earning mobility, Econometrica 46, 985-1012.

McFadden, D., 1989, A method of simulated moments for estimation of discrete choice models without numerical integration, Econometrica 57, 995-1026.

Newey, W.K. and D. McFadden, 1994, Large sample estimation and hypothesis testing, In: R. Engel and D. McFadden, eds., Handbook of Econometrics Vol. IV (Elsevier Science Publishers BV, Amsterdam), 2111-2245.

Olsen, R., 1978, Note on the uniqueness of the maximum likelihood estimator for the Tobit model, Econometrica 46, 1211-1215.

Tobin,J., 1958, Estimation of relationships for limited dependent variables, Econometrica 26, 24-36.

Wooldridge, J., 2002, Econometric analysis of cross section and panel data (MIT Press).

Table 1. Random Effects Model with $\lambda = 0.2$

Linear MLE Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\|t_{bias}\| > z^*_{0.01})$
β	1.200	0.59731	0.60600	0.03447	1.000
λ	0.200	0.17442	0.03692	0.02645	0.057
ρ	0.600	0.64256	0.05223	0.02998	0.129
σ^2_ϵ	5.000	2.85870	2.16951	0.29246	1.000

SML-GHK Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\|t_{bias}\| > z^*_{0.01})$
β	1.200	1.19731	0.04470	0.04450	0.015
λ	0.200	0.19859	0.02469	0.02460	0.012
ρ	0.600	0.59293	0.03421	0.03329	0.009
σ^2_ϵ	5.000	4.96961	0.39384	0.39121	0.007

Note: 1000 Monte Carlo data sets were created using true parameters $\theta^\star$ shown. $Mean(\widehat{\theta})$ refers to the mean of the estimated estimators over all 1000 data sets. $Std(\widehat{\theta})$ refers to the sample standard deviation of the estimated parameters. $Mean(\widehat{Std}(\widehat{\theta}))$ refers to the mean of the estimated parameter standard errors over all 1000 data sets. t_{bias} stands for the t-statistics of the biases which is based on the empirical standard errors, and $z^*_{0.01} = 2.57$. $P(|t_{bias}| > z^*_{0.01})$ is calculated based on the empirical probability.

Table 2. Random Effects Model with $\lambda = 0.9$

Linear MLE Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	1.200	0.59921	0.60412	0.04195	1.000
λ	0.900	0.89129	0.01817	0.01585	0.021
ρ	0.600	0.56370	0.05487	0.04101	0.046
σ^2_ϵ	5.000	2.86319	2.16380	0.29840	1.000
SML-GHK Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	1.200	1.19107	0.04975	0.04883	0.013
λ	0.900	0.89949	0.01454	0.01448	0.009
ρ	0.600	0.57525	0.05090	0.04438	0.025
σ^2_ϵ	5.000	4.86279	0.46854	0.44718	0.020

Note: 1000 Monte Carlo data sets were created using true parameters $\theta^\star$ shown. $Mean(\widehat{\theta})$ refers to the mean of the estimated estimators over all 1000 data sets. $Std(\widehat{\theta})$ refers to the sample standard deviation of the estimated parameters. $Mean(\widehat{Std}(\widehat{\theta}))$ refers to the mean of the estimated parameter standard errors over all 1000 data sets. t_{bias} stands for the t-statistics of the biases which is based on the empirical standard errors, and $z^*_{0.01} = 2.57$. $P(|t_{bias}| > z^*_{0.01})$ is calculated based on the empirical probability.

Table 3. Random Effects Model with Constant x

Linear MLE Estimator

Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	-1.000	0.44266	1.44849	0.04835	1.000
λ	0.200	0.17215	0.05197	0.04364	0.027
ρ	0.600	0.48404	0.12812	0.05352	0.338
σ^2_ϵ	5.000	0.94716	4.06912	0.12725	1.000

SML-GHK Estimator

Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	-1.000	-0.96955	0.13748	0.13344	0.012
λ	0.200	0.20020	0.04205	0.04184	0.009
ρ	0.600	0.58205	0.04673	0.04294	0.017
σ^2_ϵ	5.000	4.86966	0.51239	0.49173	0.012

Note: 1000 Monte Carlo data sets were created using true parameters $\theta^\star$ shown. $Mean(\widehat{\theta})$ refers to the mean of the estimated estimators over all 1000 data sets. $Std(\widehat{\theta})$ refers to the sample standard deviation of the estimated parameters. $Mean(\widehat{Std}(\widehat{\theta}))$ refers to the mean of the estimated parameter standard errors over all 1000 data sets. t_{bias} stands for the t-statistics of the biases which is based on the empirical standard errors, and $z^*_{0.01} = 2.57$. $P(\lvert t_{bias}\rvert > z^*_{0.01})$ is calculated based on the empirical probability.

Table 4. Random Effects plus AR(1) Errors Model with $\zeta = 0.2$

Linear MLE Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	1.200	0.59679	0.60505	0.02767	1.000
λ	0.200	0.16890	0.03693	0.01978	0.153
ζ	0.200	0.00851	0.19496	0.03452	0.997
σ_d^2	0.800	0.72841	0.07271	0.01186	1.000
SML-GHK Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	1.200	1.19662	0.01816	0.01781	0.008
λ	0.200	0.20007	0.01200	0.01198	0.013
ζ	0.200	0.18015	0.05135	0.04727	0.015
σ_d^2	0.800	0.79545	0.01201	0.01109	0.016

Note: 1000 Monte Carlo data sets were created using true parameters $\theta^\star$ shown. $Mean(\widehat{\theta})$ refers to the mean of the estimated estimators over all 1000 data sets. $Std(\widehat{\theta})$ refers to the sample standard deviation of the estimated parameters. $Mean(\widehat{Std}(\widehat{\theta}))$ refers to the mean of the estimated parameter standard errors over all 1000 data sets. t_{bias} stands for the t-statistics of the biases which is based on the empirical standard errors, and $z^*_{0.01} = 2.57$. $P(|t_{bias}| > z^*_{0.01})$ is calculated based on the empirical probability.

Table 5. Random Effects plus AR(1) Errors Model with $\zeta = 0.9$

Linear MLE Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	1.200	0.59712	0.60478	0.02872	1.000
λ	0.200	0.17854	0.02851	0.01869	0.085
ζ	0.900	0.00111	0.90137	0.03473	1.000
σ_d^2	0.800	0.77724	0.02472	0.00951	0.433

SML-GHK Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	1.200	1.19739	0.00872	0.00830	0.011
λ	0.200	0.19918	0.00602	0.0595	0.012
ζ	0.900	0.77327	0.13961	0.05792	0.356
σ_d^2	0.800	0.87993	0.10082	0.06115	0.008

Note: 1000 Monte Carlo data sets were created using true parameters $\theta^\star$ shown. $Mean(\widehat{\theta})$ refers to the mean of the estimated estimators over all 1000 data sets. $Std(\widehat{\theta})$ refers to the sample standard deviation of the estimated parameters. $Mean(\widehat{Std}(\widehat{\theta}))$ refers to the mean of the estimated parameter standard errors over all 1000 data sets. t_{bias} stands for the t-statistics of the biases which is based on the empirical standard errors, and $z^*_{0.01} = 2.57$. $P(\lvert t_{bias}\rvert > z^*_{0.01})$ is calculated based on the empirical probability.

Table 6. Random Effects Model with Lagged Observed Variables

Linear MLE Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	1.200	0.63051	0.57166	0.03418	1.000
λ	0.200	0.15773	0.04959	0.02576	0.174
ρ	0.600	0.65083	0.05800	0.02771	0.236
σ^2_ϵ	5.000	3.00092	2.02438	0.29236	1.000
SML-GHK Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	1.200	1.19696	0.04340	0.04323	0.012
λ	0.200	0.20457	0.02883	0.02843	0.011
ρ	0.600	0.59132	0.03244	0.03121	0.016
σ^2_ϵ	5.000	4.94647	0.38644	0.38213	0.010

Note: 1000 Monte Carlo data sets were created using true parameters $\theta^\star$ shown. $Mean(\widehat{\theta})$ refers to the mean of the estimated estimators over all 1000 data sets. $Std(\widehat{\theta})$ refers to the sample standard deviation of the estimated parameters. $Mean(\widehat{Std}(\widehat{\theta}))$ refers to the mean of the estimated parameter standard errors over all 1000 data sets. t_{bias} stands for the t-statistics of the biases which is based on the empirical standard errors, and $z^*_{0.01} = 2.57$. $P(|t_{bias}| > z^*_{0.01})$ is calculated based on the empirical probability.

Table 7. Random effects with AR(2) Model

Linear MLE Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	0.300	0.30307	0.02564	0.02542	0.011
λ	0.800	0.79191	0.02781	0.02656	0.019
ρ	0.600	0.61903	0.04365	0.03920	0.018
σ^2_ϵ	5.000	2.47612	2.54595	0.29329	1.000

SML-GHK Estimator					
Parameter	$\theta^\star$	$Mean(\widehat{\theta})$	$Std(\widehat{\theta})$	$Mean(\widehat{Std}(\widehat{\theta}))$	$P(\lvert t_{bias}\rvert > z^*_{0.01})$
β	0.300	0.30789	0.02657	0.02533	0.018
λ	0.800	0.79392	0.02665	0.02590	0.016
ρ	0.600	0.57628	0.04938	0.04322	0.028
σ^2_ϵ	5.000	4.82409	0.48419	0.45029	0.009

Note: 1000 Monte Carlo data sets were created using true parameters $\theta^\star$ shown. $Mean(\widehat{\theta})$ refers to the mean of the estimated estimators over all 1000 data sets. $Std(\widehat{\theta})$ refers to the sample standard deviation of the estimated parameters. $Mean(\widehat{Std}(\widehat{\theta}))$ refers to the mean of the estimated parameter standard errors over all 1000 data sets. t_{bias} stands for the t-statistics of the biases which is based on the empirical standard errors, and $z^*_{0.01} = 2.57$. $P(|t_{bias}| > z^*_{0.01})$ is calculated based on the empirical probability.