Center for Research on Economic and Social Theory

and
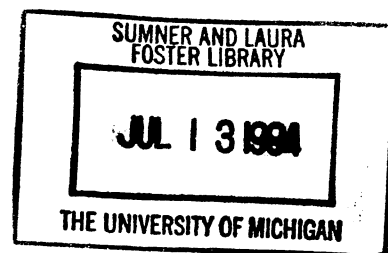
Department of Economics

Working Paper Series

Simulated Maximum Likelihood Estimation of Dynamic Discrete Choice Statistical Models---Some Monte Carlo Results

Lung-fei Lee

April 1994
Number 94-06

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, MI 48109-1220

Simulated Maximum Likelihood Estimation of Dynamic Discrete Choice Statistical Models
—— Some Monte Carlo Results

by Lung-Fei Lee

Department of Economics, The University of Michigan

Ann Arbor, MI 48109-1220

April 1994

Abstract

This article reports Monte Carlo results on the simulated maximum likelihood estimation of discrete panel statistical models. Among them are Markov, Generalized Poly, Renewal, and Habit Persistence Models with or without unobserved heterogeneity and serially correlated disturbances. We investigate statistical properties and computational performance of simulated maximum likelihood methods and a bias-correction procedure. With a moderate number of simulation draws for the construction of simulator and the bias adjusted procedure, most of these complex dynamic models can be adequately estimated for panels with length up to 30. The Polya model and the Renewal model can be accurately estimated for panels with 50 periods.

JEL subject codes: C13, C15, C23, C25

Key Words:

Discrete choice panel models, dynamic models, heterogeneity, serial correlation, simulated ML, bias-correction, Monte Carlo studies.

CORRESPONDENCE ADDRESS:

Lung-Fei Lee, Department of Economics, The University of Michigan, 611 Tapan Street, Ann Arbor, MI 48109-1220.

1. Introduction

In Heckman (1981a), a rich group of discrete time-discrete outcome stochastic processes has been introduced. It contains Bernoulli models, Markov models, renewal processes, Polya models, latent Markov models, and other familiar stochastic processes. The models are sufficiently flexible to accommodate time-varying explanatory variables, serial correlation for unobservable, and complex structural economic interrelationships among decisions over time. With rich specifications on both lagged dependent variables and correlated disturbances, these proposed models accommodate true and spurious state dependence and can be used to test competing theories on state dependence. Heckman (1981a) has provided discussions on various models in terms of data requirements, model identification, and implications for probability run patterns. Implementation of probability run patterns to test true state dependence against spurious state dependence can also be found in Heckman (1978) and Lee (1987).

For computation, Heckman (1981a) has focused on models with fixed effect, random components, or factor analytic schemes for unobserved heterogeneity but with serially independent distribution for the remaining disturbance. Such structures simplify computation as likelihood functions for such models involve single integral, which can be effectively implemented by the Gaussian Quadrature method. While the specification of random component for unobserved heterogeneity is a popular approach for panel data, the serially independent assumption for the remaining disturbance is restrictive. Likelihood functions for models with general serial correlation will in general involve multiple integrals. For discrete panel data models, the dimension of multiple integrals will be proportional to the number of time periods. The numerical implementation of such models is known to be computationally difficulty if not intractable. However, the recent development of simulation methods by Lerman and Manski (1981), McFadden (1989), Pakes and Pollard (1989), Borsch-Supan and Hajivassiliou (1993), and Keane (1994), among others, provides evidences that such a difficulty may be overcome with carefully designed Monte Carlo simulation estimation methods. A simulator due to Geweke (1989), Borsch-Supan and Hajivassiliou (1993), and Keane (1994) based on recursive simulation of truncated random variables is shown to provide a good unbiased simulator for multivariate

normal probabilities. Keane (1994) has emphasized its usefulness for the estimation of panel data model and has provided some Monte Carlo results for binary choice panel regression models with 8 periods. The Model Carlo results seem encouraging.

In this article, we are going to investigate the simulated maximum likelihood method for the estimation of various dynamic panel models introduced in Heckman (1981a). Special attention will be on the value and potential limitation for the estimation of long panels. We report Monte Carlo results for the simulation estimation of such models. An important aspect of the experiment is to investigate the discrimination ability of the true state dependence vs. spurious state dependence in various dynamic models. Other aspects of interest are on the simulated likelihood method itself. As the time length for a panel data increases, the total number of choice alternatives, which are composed of choice paths, increases exponentially and the dimension of integration increases linearly. The probability for any specific choice path will become smaller for longer panel data. One may expect that the simulator would become less accurate as time length increases. So one of the aspects is to investigate the quality of the simulated estimates for model parameters with long panel data. As the likelihood function is nonlinear, the simulated log likelihood function is a biased estimate of the exact log likelihood function. The simulated likelihood estimator can be consistent only if the number of simulated variables in the construction of simulator increases with sample size. For efficient estimation, the number of simulated variables has to increase with sample size at a certain rate. If the number of simulated variables does not increase fast enough or, in practice, the number is not large enough relative to sample size, the simulated likelihood estimator might have an asymptotic bias or a dominated finite sample bias (relative to standard error) due to simulation. The leading bias term due to simulation in the simulated likelihood estimator can be removed by a bias-correction procedure as suggested in Lee (1993). As a by-product, we will investigate value of the bias-correction procedure for the estimation of dynamic models with our Monte Carlo experiments.

The organization of this article is as follows. In Section 2, we provide an overview of various interesting models introduced in Heckman (1981a). It provides a guided tour for models studied in our Monte Carlo experiments. We review the formulation of simulators and simulated likelihood functions for such models and describe the bias-correction procedure. Monte Carlo results are reported for each model in subsequent sections. General conclusions are drawn in the last section.

2. Dynamic Models and Simulated Maximum Likelihood Estimation — an Overview

To analyze panel discrete responses in discrete time, Heckman (1981a) proposed a general framework in terms of latent dependent variables. Responses are observed from signs of latent dependent variables. The latent dependent variable at each period can be interpreted as utility difference across choice alternatives. Observed responses are results of utility maximization as in McFadden's discrete choice framework (1974). The latent dependent variable y_{it}^* for an individual i at period t is a function of vector of exogenous variables z_{it}, past occurrence of the event, and prior propensities to select a state:

$$y_{it}^* = z_{it}\beta + \sum_{l=1}^{\infty}\gamma_l y_{i,t-l} + \lambda\sum_{s=1}^{\infty}\prod_{l=1}^{s}y_{i,t-l} + \sum_{l=1}^{\infty}\delta_l y_{i,t-l}^* + v_{it}, \qquad (2.1)$$

and the observed dependent variable is

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0, \\ 0 & \text{if } y_{it}^* \le 0. \end{cases}$$

The initial conditions y_{it} and y_{it}^*, $t = 0, -1, \ldots$, are assumed to be fixed outside the model. The vector of disturbances $v_i = (v_{i1}, \cdots, v_{iT})'$ are independent of all z_{it} and are independent across individuals, but its components can be serially correlated over time.

The first term on the right-hand side of (2.1) represents the effect of exogenous variables on choice. The second term on the right-hand side of (2.1) represents the effect of the entire past history of the process on current choice. The third term represents the cumulative effect of the continuous experience in a state on current choice. This term could be generalized to allow for depreciation. The fourth term captures the notion of habit persistence. This general model can further be generalized to allow coefficients to depend on time as indicated in Heckman (1981a). This specification accommodates a wide variety of interesting stochastic models. The model that $y_{it}^* = z_{it}\beta + \sum_{l=1}^{\infty}\gamma_l y_{i,t-l} + v_{it}$ includes the Markov process and the Polya model. The specification $y_{it}^* = z_{it}\beta + \lambda\sum_{s=1}^{\infty}\prod_{l=1}^{s}y_{i,t-l} + v_{it}$ generates a renewal process. A latent Markov model emerges in $y_{it}^* = z_{it}\beta + \sum_{l=1}^{\infty}\delta_l y_{i,t-l}^* + v_{it}$. Our Monte Carlo experiment will study these various interesting models.

An important issue for panel models is to distinguish true state dependence from spurious state dependence. In true state dependence, past experience has a genuine behavioral effect in that an identical individual who has not experienced the event will behave differently than an individual who has experienced the event. The introduction of lagged variables of y or y^* captures such a dependence. In spurious state dependence, previous experience appears to influence the determinant of future event solely because it is

due to temporally correlated and persistent unobservable that determine choices. The unobservable heterogeneity for spurious state dependence can be introduced into the model with serially correlated disturbances and a random individual component in v_{it}; for example, $v_{it} = \xi_i + \epsilon_{it}$, where ϵ_{it} can be serially correlated and ξ_i is an individual component. With such a specification, the general framework (2.1) accommodates general sorts of heterogeneity and state dependence. To distinguish true state dependence from spurious state dependence, an alternative simple method based on lagged exogenous variables has been suggested in Chamberlain (1978). However, in contrary to the framework (2.1), that simple method can not make distinctions with regard to different types of state dependence, heterogeneity, and serial correlation. Hsiao (1986) provides a comprehensive survey on these models and related issues.

For the estimation of (2.1), Heckman (1981a) has proposed the method of maximum likelihood under the assumption that $v_{i1}, \ldots, v_{iT}$ are multivariate normally distributed. A general likelihood function can be found in his article.[1] For general correlated disturbances, the likelihood function involves multiple integrals. In a limited Monte Carlo experiment reported in Heckman (1981b) and an empirical application in Heckman (1981c), only fixed effect and random component models with $v_{it} = \xi_i + \epsilon_{it}$, where ϵ_{it} are serially uncorrelated, are estimated. For random component or one factor models, multivariate probability functions involve only single integrals, which can be effectively implemented by the Gaussian Quadrature method [Butler and Moffitt (1982)]. For more general correlated disturbances, we suggest estimation by simulated maximum likelihood methods. Simulation based estimation methods have recently been proposed by Lerman and Manski (1981), McFadden (1987), and Pakes and Pollard (1987), among others. For the implementation of simulated likelihood methods, we adopt the Geweke-Hajivassiliou-Keane (GHK) simulator. The GHK simulator is an unbiased simulator that can provide close approximation to some multivariate normal probabilities (Borsch-Supan and Hajivassiliou, 1993) and is of special interest for panel data models (Keane, 1994).

To illustrate the construction of the GHK simulator and the simulated likelihood function for (2.1) that are relevant for our experiment, consider the specification $v_{it} = \sigma \xi_i + \epsilon_{it}$, where ϵ_{it} is a ARMA(p,q) process: $\epsilon_{it} = \sum_{l=1}^{p} \rho_l \epsilon_{i,t-l} + w_{it} + \sum_{j=1}^{q} \psi_j w_{i,t-j}$. The initial values for $\epsilon_{it}, w_{it}, y_{it}$ and y_{it}^* for $t \leq 0$ are known and are assumed to be zero. For normalization, w_{it} has a unit variance. Also ξ has a unit variance in the above specification. To simplify notation, the individual subscript i will be dropped in the derivation of the

[1] The likelihood formula there is correct only for models without lagged latent dependent variables. His likelihood formula needs to be revised for models with lagged latent dependent variables.

likelihood function for an individual. Let $a_t = x_t \beta + \sum_{l=1}^{\infty} \gamma_l y_{t-l} + \lambda \sum_{s=1}^{\infty} \prod_{l=1}^{s} y_{t-l} + \sum_{l=1}^{\infty} \delta_l y_{t-l}^* + \sigma \xi$. The joint probability for $y_1, \ldots, y_T$ conditional on exogenous variables and ξ is

$$P(y_1, \ldots, y_T | z_1, \ldots, z_T, \xi) = \int_{L_1}^{U_1} \cdots \int_{L_T}^{U_T} f(\epsilon_T | \epsilon_{T-1}, \cdots, \epsilon_1) f(\epsilon_{T-1} | \epsilon_{T-2}, \cdots, \epsilon_1) \cdots f(\epsilon_1) d\epsilon_T \ldots d\epsilon_1, \quad (2.2)$$

where $f(\epsilon_t | \epsilon_{t-1}, \cdots, \epsilon_1)$ is the conditional density of ϵ_t conditional on the past ϵs and the integral limits are

$$L_t = \begin{cases} -a_t & \text{if } y_t = 1, \\ -\infty & \text{if } y_t = 0, \end{cases} \qquad U_t = \begin{cases} \infty & \text{if } y_t = 1, \\ -a_t & \text{if } y_t = 0. \end{cases}$$

Given the initial conditions for ϵ and w, the ARMA process can be inverted recursively so that each w_t can be expressed as a function of ϵ_{t-s}, $s \geq 0$. Denote $b_t = a_t + \sum_{l=1}^{p} \rho_l \epsilon_{t-l} + \sum_{j=1}^{q} \psi_j w_{t-j}$. Let ϕ and Φ be, respectively, the standard normal density and distribution functions. Extend the limits L_t and U_t to

$$\bar{L}_t = \begin{cases} -b_t & \text{if } y_t = 1, \\ -\infty & \text{if } y_t = 0, \end{cases} \qquad \bar{U}_t = \begin{cases} \infty & \text{if } y_t = 1, \\ -b_t & \text{if } y_t = 0. \end{cases}$$

By a transformation of variables, (2.2) becomes

$$\begin{aligned}
P(y_1, &\ldots, y_T | z_1, \ldots, z_T, \xi) \\
&= \int_{L_1}^{U_1} \cdots \left(\int_{\bar{L}_{T-1}}^{\bar{U}_{T-1}} \left[\int_{\bar{L}_T}^{\bar{U}_T} \phi(w_T) dw_T \right] \phi(w_{T-1}) dw_{T-1} \right) \cdots \phi(w_1) dw_1 \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\Phi(\bar{U}_T) - \Phi(\bar{L}_T) \right) \prod_{s=1}^{T-1} \left(\Phi(\bar{U}_{T-s}) - \Phi(\bar{L}_{T-s}) \right) \phi_{[\bar{L}_{T-s}, \bar{U}_{T-s}]}(w_{T-s}) dw_{T-s} \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Phi((2y_T - 1)b_T) \prod_{s=1}^{T-1} \Phi((2y_{T-s} - 1)b_{T-s}) \phi_{[\bar{L}_{T-s}, \bar{U}_{T-s}]}(w_{T-s}) dw_{T-s},
\end{aligned} \quad (2.3)$$

because $\Phi(\bar{U}_t) - \Phi(\bar{L}_t) = \Phi((2y_t - 1)b_t)$, where $\phi_{[\bar{L}_t, \bar{U}_t]}$ is a truncated standard normal density function with support $[\bar{L}_t, \bar{U}_t]$. The probability of $y_1, \cdots, y_T$ conditional on $z_1, \cdots, z_T$ is $P(y_1, \cdots, y_T | z_1, \cdots, z_T) = \int_{-\infty}^{\infty} P(y_1, \cdots, y_T | z_1, \cdots, z_T, \xi) \phi(\xi) d\xi$. This function and (2.3) naturally suggest an unbiased estimator, namely, the GHK simulator.

The truncated standard normal variable w_t on $[\bar{L}_t, \bar{U}_t]$ can be easily simulated as transformation of a uniform random variable. Suppose u is a random draw from the uniform $[0, 1]$ distribution. Corresponding to $y_t = 0$, $\bar{L}_t = -\infty$ and $\bar{U}_t = -b_t$. A truncated standard normal variable w_t can be obtained from the relation $\Phi(w_t)/\Phi(-b_t) = u$. When $y_t = 1$, $\bar{L}_t = -b_t$ and $\bar{U}_t = \infty$. For the latter case, the truncated standard normal variable can be obtained from $[\Phi(w_t) - \Phi(-b_t)]/[1 - \Phi(-b_t)] = 1 - u$. Using the symmetric property of the standard normal density, $w_t = -(2y_t - 1)\Phi^{-1}[u\Phi((2y_t - 1)b_t)]$ is a random draw from the truncated standard normal distribution on $[\bar{L}_t, \bar{U}_t]$. The random variables for the GHK simulator can be recursively

generated. Initiate $\epsilon_0 = 0$ and other initial conditions. Generate $u_1, \cdots, u_{T-1}$ independent uniform $[0, 1]$ random variables. Generate ξ from a $N(0, 1)$ random variable generator. The random variables $w_1, \cdots, w_{T-1}$ can then be generated recursively from the following steps from $t = 1$ to $T - 1$:

1). Compute

$$w_t = -(2y_t - 1)\Phi^{-1}\left[u_t\Phi\left((2y_t - 1)\left(z_t\beta + \sum_{l=1}^{t}\gamma_l y_{t-l} + \lambda\sum_{s=1}^{t}\prod_{l=1}^{s} y_{t-l} + \sum_{l=1}^{t}\delta_l y_{t-l}^*\right.\right.\right.$$
$$\left.\left.\left. + \sigma\xi + \sum_{l=1}^{p}\rho_l\epsilon_{t-l} + \sum_{j=1}^{t}\psi_j w_{t-j}\right)\right)\right].$$

2). Update the disturbance process $\epsilon_t = \sum_{l=1}^{p}\rho_l\epsilon_{t-l} + w_t + \sum_{j=1}^{t}\psi_j w_{t-j}$.

3). Generate the latent dependent variable

$$y_t^* = z_t\beta + \sum_{l=1}^{t}\gamma_l y_{t-l} + \lambda\sum_{s=1}^{t}\prod_{l=1}^{s} y_{t-l} + \sum_{l=1}^{t}\delta_l y_{t-l}^* + \sigma\xi + \epsilon_t.$$

For special models, some of the above simulation steps may not be needed. For example, if the model does not involve lagged latent dependent variable y^*, simulation of the latent variables y_t^* will not be needed and the step 3) can be passed over. For each i, with m independently generated vector of random draws, the simulated log likelihood function is

$$\mathcal{L} = \sum_{i=1}^{n}\ln\left\{\frac{1}{m}\sum_{j=1}^{m}\Phi\left((2y_{it} - 1)(z_{it}\beta + \sum_{l=1}^{t}\gamma_l y_{i,t-l} + \lambda\sum_{s=1}^{t}\prod_{l=1}^{s} y_{i,t-l} + \sum_{l=1}^{t}\delta_l y_{i,t-l}^{*(j)}\right.\right.$$
$$\left.\left. + \sigma\xi_i^{(j)} + \sum_{l=1}^{p}\rho_l\epsilon_{i,t-l}^{(j)} + \sum_{j=1}^{t}\psi_j w_{i,t-j}^{(j)}\right)\right)\right\}, \tag{2.4}$$

where the superscript (j) denotes an independent simulation draw.

Asymptotic properties of the simulated maximum likelihood estimator (SMLE) have been studied in Hajivassiliou and McFadden (1990), Lee (1992, 1993) and Gourieroux and Monfort (1993), among others. As the choice probability simulators are independent across i, the simulated MLE can be asymptotically efficient when $\lim_{n\to\infty} n^{1/2}/m = 0$. However, when m increases at a rate slower than $n^{1/2}$, the limiting distribution of the SMLE becomes nasty. As shown in Lee (1993), if m increases at the rate of $n^{1/2}$, an asymptotic bias exists in the limiting distribution. The situation becomes worse when m increases at a rate slower than $n^{1/2}$ because the asymptotic bias will then dominate the variance. To reduce the dominated bias due to simulation, Lee (1993) has suggested a simple bias-correction procedure. For the discrete choice model with L alternatives, suppose the simulated log likelihood is written in a familiar format as $\mathcal{L}(\theta) = \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\ln f_{m,l,i}(z_i, \theta)$

where θ is the vector of parameters in the model, d_l is the dichotomous choice indicator for the alternative l and $f_{m,l,i} = \frac{1}{m}\sum_{j=1}^{m} h(w_i^{(j)}, z_i, \theta)$ is the simulated probability for the alternative l for the individual i. Let $\hat{\theta}$ be the SMLE. A bias-adjusted estimator $\hat{\theta}_A$ can be derived by correcting a bias term from $\hat{\theta}$:

$$\hat{\theta}_A = \hat{\theta} - \left\{\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial\ln f_{m,l,i}(z_i, \hat{\theta})}{\partial\theta}\frac{\partial\ln f_{m,l,i}(z_i, \hat{\theta})}{\partial\theta'}\right\}^{-1}\sum_{i=1}^{n}\omega_{m,i}(z_i, \hat{\theta})/m, \tag{2.5}$$

where

$$\omega_{m,i}(z_i, \theta) = \sum_{l=1}^{L}\left\{\frac{d_{li}}{f_{m,l,i}^2(z_i, \theta)}\left[\frac{\partial\ln f_{m,l,i}(z_i, \theta)}{\partial\theta}(S_{m,l,i}(z_i, \theta) - f_{m,l,i}^2(z_i, \theta))\right.\right.$$
$$\left.\left. - \left(C_{m,l,i}(z_i, \theta) - f_{m,l,i}(z_i, \theta)\frac{\partial f_{m,l,i}(z_i, \theta)}{\partial\theta}\right)\right]\right\},$$

with $S_{m,l,i}(z_i, \theta) = \frac{1}{m}\sum_{j=1}^{m} h_l^2(w_i^{(j)}, z_i, \theta)$ and $C_{m,l,i}(z_i, \theta) = \frac{1}{m}\sum_{j=1}^{m} h_l(w_i^{(j)}, z_i, \theta)\frac{\partial h_l(w_i^{(j)}, z_i, \theta)}{\partial\theta}$. The bias-corrected estimator $\hat{\theta}_A$ improves upon $\hat{\theta}$ by eliminating the leading bias term in $\hat{\theta}$ due to simulation. Another valuable property of $\hat{\theta}_A$ is that its asymptotic efficiency requires only that m goes to infinity at a rate faster than $n^{1/4}$ instead of $n^{1/2}$.

The following sections provide Monte Carlo results for several interesting panel regression and dynamic choice models. We are interested on the performance of simulated maximum likelihood methods; namely, finite sample statistical properties of the SMLE and bias-corrected estimators, the value of long panel data, and the time cost for estimation of various models. Some limited Monte Carlo results on the GHK simulator can be found in Borsch-Supan and Hajivassiliou (1993) for multiple choice models, and Keane (1994) for panel choice regression models up to eight time periods. Bias-corrected procedures have not been considered in their articles but in Lee (1993) for panel choice regression models and Markov models with four periods. Our present Monte Carlo experiments provide much more evidences on the value of simulated likelihood estimation. Our experiments investigate both the simulated maximum likelihood method and the bias-correction procedure. Panel models with periods much longer than eight are estimated. We are interested in the estimation of various dynamic models in addition to regression models. Practical values of simulation estimation are the ultimate goal of simulation based inference. The following Section 3 investigates only discrete choice regression models. Section 4 investigates Markov models. Polya and Renewal dynamic models are studied, respectively, in Sections 5 and 6. Section 7 considers Habit Persistent models. For all these models, first order autocorrelated disturbances are maintained. For models with unobserved heterogeneity, additional random components are included.

3. Panel Discrete Choice Regression Models

3.1 Models with Serially Correlated Disturbances

The discrete choice regression models are models without true state dependence. Throughout all the experiments, the stochastic process for the disturbance ϵ_{it} is assumed to be a AR(1) process. The model for generating the sample for our experimental study is a binary choice panel model:

$$y_{it}^* = \lambda + \beta x_{it} + \epsilon_{it} \tag{3.1}$$

and

$$\epsilon_{it} = \rho \epsilon_{i,t-1} + w_{it}, \tag{3.2}$$

where w_{it} are i.i.d. normal $N(0,1)$. The vector of exogenous variables z_{it} is $(1, x_{it})$.

The exogenous variables x_{it} are generated as $x_{it} = (1/\sqrt{2})r_{it} + \sqrt{6}s_i$ where r_{it} are i.i.d. truncated normal $N(0,1)$ variables with support $[-2,2]$ and s_i are independent uniform variables with support on $[-1/2, 1/2]$. The variance of x is about 1 and its correlation coefficient over time is about 0.5. This process of generating exogenous variables is used for all the models in this article. The initial condition for (3.2) is $\epsilon_{i0} = 0$ for all i. Sample data are generated with $\beta = 1$, $\lambda = 0.2$ and $\rho = 0.4$. This ρ represents a moderate amount of autocorrelation for the disturbance. The sample size is 200. We have experimented with a small number of random draws and a moderate number of draws, namely, $m = 15$ and $m = 50$, for the construction of the GHK simulator. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^n \ln \left\{ \frac{1}{m} \sum_{j=1}^m \prod_{t=1}^T \Phi[D_{it}(\lambda + \beta x_{it} + \rho \epsilon_{i,t-1}^{(j)})] \right\}.$$

The number of periods for the panel data varies from $T = 8$ to the maximum of 100. For each case, the number of replications is 300. For each replication, the set of exogenous variables is redrawn. We report the empirical mean and standard deviation of estimates for each parameter. For the sake of easy comparison, we will also report root mean square errors for estimates and the CPU time in seconds per replication. All the computation is performed with IBM RS/6000 machines. The CPU time refers to the computation speed of a IBM RS/6000 Model 580 workstation.[2] The maximization algorithm used is a conjugate gradient method described in Press et al. (1986), Chapter 10. For all the cases and replications reported here, the algorithm

---

[2] Comparing the speed of Model 580 with the model 320H machine, the latter is slow by a time factor of 2.6.

converges without running into numerical problems.[3]

Table 1.A reports empirical means and standard deviations for both the SMLE (Bias unadjusted SML) and the bias-adjusted estimates (Bias adjusted SML). Root mean squared errors of estimates are reported in Table 1.B. For all the panels with periods from 8 to 100, the SMLE of the intercept λ and the regression slope parameter β have small biases. The SMLE of the autocorrelation coefficient ρ has shown some biases. As the number of time periods increases, the choice pattern becomes numerous and the choice probabilities become small. As smaller probabilities are more difficult to simulate than larger probabilities (Borsch-Supan and Hajivassiliou (1993)), biases of the SMLE for long panels become slightly larger. On the other hand, since longer panel data provide more sample information about the stochastic process, standard errors of estimates decrease. Root mean square errors also decrease for estimates with longer panels if time periods are not too long. However, for long panels with $T = 50$ or $T = 100$, root mean square errors for ρ are larger. For $T = 100$, root mean square errors for β are also larger. This is due to the circumstance that increasing bias may dominate the reduction of standard error. Biases of estimates are all reduced when the number of simulated random variables increases from $m = 15$ to 50. The improvement of the estimate of ρ is more apparent. The increased CPU time cost is close to be linear in m and T. The bias-corrected procedure eliminates some of the biases. The bias correction is valuable especially for small m with $m = 15$. Although standard errors of bias-corrected estimates are slightly larger, the reduction in bias dominates the slight increase of standard error and results in smaller root mean squares errors for estimates of ρ. Root mean square errors of bias-corrected estimates are in general smaller than root mean square errors of bias-unadjusted estimates for long panels. For short panels with $T = 8$ and 15, bias-corrected estimates with $m = 15$ can be at least as good as unadjusted SMLE with $m = 50$ in terms of similar bias. The additional CPU cost for bias-correction is negligible in the overall cost. In sum, the performance of the simulated ML methods seems satisfactory ever for long panels in this experiment.

Tables 2.A and 2.B report results for another data set. The sample is generated with ρ set to 0.85, a rather high correlation for the disturbance. The SMLE have much larger biases when they are compared with corresponding estimates in the previous sample with $\rho = 0.4$. Estimates with shorter periods $T = 8$ and 15 seem satisfactory even with $m = 15$. But for panels with longer periods, biases of estimates seem

---

[3] For models without unobserved heterogeneity, initial estimates are arbitrarily set to zero. For models with unobservable heterogeneity, the initial estimate of σ is set to one and the initial estimates of the other parameters are set to zero. We have also tried some other starting values such as $(1, -1, 0)$ and $(-1, 1, 0)$ for (β, λ, ρ) with which the algorithm converges to similar solutions.

severely downward. The results are better for the biased-corrected procedure. But for better improvement, larger m is desirable. With $m = 50$, the SMLE are reasonably good even for $T = 50$. For strong serial correlation, estimates for ρ have smaller standard errors but estimates of the intercept and the regression coefficient have larger errors than corresponding cases with a moderate serial correlation. With strong serial correlation, one may expect that there would be less variation and less changes on choices over time. Overall, simulation estimates perform better for models with moderate serial correlation than models with rather high serial correlation. The Monte Carlo results in Keane (1994) for a similar model with additional unobserved heterogeneity pointed out the same problem.[4]

3.2 Models with Unobserved Heterogeneous and Serially Correlated Disturbances

The model with unobserved heterogeneity generalizes the previous model (3.1)-(3.2) by introducing a random component as an additional disturbance:

$$y_{it}^* = \lambda + \beta x_{it} + \sigma \xi_i + \epsilon_{it}, \tag{3.3}$$

where ξ_i are i.i.d. $N(0, 1)$ and are independent of ϵ_{it}. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi[D_{it}(\lambda + \beta x_{it} + \sigma \xi_i^{(j)} + \rho \epsilon_{i,t-1}^{(j)})] \right\}.$$

Sample data are generated with parameters $\lambda = 0.2$, $\beta = 1.0$, $\sigma^2 = 0.5$, and $\rho = 0.4$. With the additional $\sigma \xi$, the serial correlation of two adjacent periods has a correlation coefficient about 0.6 and the fraction of variance due to the individual effect is about 0.3.

Tables 3.A and 3.B report simulated likelihood estimates and their root mean squares errors. All the estimates for β have small biases. Biases for the intercept λ are also small for T up to 50. There are some biases for the SMLE of σ and ρ. The SMLE of σ tend to increase as T increases and are biased upward for long panels with $T = 50$ or 100. On the other hand, biases of ρ are downward. So for long panel data, the correlation due to unobserved heterogeneity tends to be over estimated and the serial correlation of the remaining disturbance tends to be underestimated.[5] With m increases to 50, biases are reduced and the SMLE are reasonably good for T up to 30. The bias correction procedure reduces some of the biases and root mean square errors in β and ρ. Standard errors for estimates of β and ρ are decreasing as periods in

panels increase. For long panels with $T = 50$ and 100, biases can dominate standard errors that results in larger root mean square errors. Overall, those SMLE and bias-corrected estimates with $m = 50$ are good for panels with T up to 30 for this model. Comparing these estimates with estimates of panel regression models in (3.1) without unobserved heterogeneity, these estimates are slightly worse than the corresponding estimates in Table 1 with $\rho = 0.4$ but are much better than the estimates in Table 2 with $\rho = 0.85$. This may be due to the fact that the first lagged autocorrelation of this model is stronger than the autocorrelation of the model (3.1) with $\rho = 0.4$ but is weaker than the case with $\rho = 0.85$.

---

[4] As Keane (1994) has experimented with a very small m, namely 10 draws, and has neither used any bias correction procedure nor increased the number of simulation draws, his conclusion is too pessimistic.

[5] With higher autocorrelation but smaller fraction of variance due to the individual effect, the Monte Carlo result in Keane (1994) has also shown this pattern of bias for a panel of eight periods based on $m = 10$.

Table 1.A
SML—Model: Serial correlation only; no state dependence
True parameters: $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.4$

T	n	m	β	λ	ρ	time
			Bias unadjusted SML			
8	200	15	0.9939 (.0542)	.1999 (.0451)	.3757 (.0504)	21.83
15	200	15	0.9895 (.0406)	.2007 (.0337)	.3724 (.0322)	43.99
30	200	15	0.9854 (.0285)	.1984 (.0237)	.3633 (.0206)	97.27
50	200	15	0.9803 (.0227)	.1953 (.0164)	.3540 (.0170)	170.54
100	200	15	0.9719 (.0144)	.1948 (.0123)	.3369 (.0109)	367.64
8	200	50	0.9999 (.0549)	.2010 (.0453)	.3905 (.0503)	74.09
15	200	50	0.9955 (.0412)	.2019 (.0339)	.3876 (.0328)	151.01
30	200	50	0.9943 (.0288)	.2002 (.0240)	.3842 (.0216)	337.24
50	200	50	0.9904 (.0232)	.1975 (.0164)	.3782 (.0176)	570.69
100	200	50	0.9824 (.0146)	.1971 (.0123)	.3635 (.0108)	1242.82
			Bias adjusted SML			
8	200	15	1.0006 (.0551)	.2014 (.0453)	.3929 (.0517)	22.04
15	200	15	0.9969 (.0414)	.2023 (.0341)	.3905 (.0332)	44.40
30	200	15	0.9930 (.0291)	.1998 (.0239)	.3812 (.0212)	98.16
50	200	15	0.9871 (.0231)	.1966 (.0166)	.3702 (.0175)	172.05
100	200	15	0.9749 (.0145)	.1954 (.0124)	.3452 (.0110)	370.56
8	200	50	1.0024 (.0552)	.2015 (.0454)	.3968 (.0509)	74.80
15	200	50	0.9986 (.0416)	.2025 (.0341)	.3948 (.0335)	152.39
30	200	50	0.9985 (.0291)	.2010 (.0241)	.3935 (.0223)	340.09
50	200	50	0.9951 (.0235)	.1984 (.0165)	.3888 (.0183)	575.67
100	200	50	0.9850 (.0147)	.1976 (.0123)	.3705 (.0109)	1252.41

Table 1.B
SML—Model: Serial correlation only; no state dependence
Root mean squared errors

T	n	m	β	λ	ρ	β	λ	ρ
			Bias unadjusted SML			Bias adjusted SML		
8	200	15	.0545	.0450	.0559	.0550	.0453	.0521
15	200	15	.0419	.0336	.0423	.0415	.0341	.0345
30	200	15	.0320	.0237	.0421	.0299	.0239	.0283
50	200	15	.0300	.0170	.0490	.0264	.0169	.0346
100	200	15	.0316	.0134	.0641	.0290	.0132	.0559
8	200	50	.0548	.0453	.0511	.0552	.0454	.0509
15	200	50	.0414	.0339	.0351	.0416	.0341	.0338
30	200	50	.0293	.0239	.0268	.0291	.0241	.0232
50	200	50	.0251	.0166	.0280	.0240	.0166	.0214
100	200	50	.0229	.0126	.0380	.0210	.0125	.0314

<div style="display:flex">

Table 2.A
SML—Model: Serial correlation only; no state dependence
True parameters: $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.85$

T	n	m	β	λ	ρ	time
			Bias unadjusted SML			
8	200	15	0.9671 (.0650)	.1922 (.0734)	.8217 (.0339)	24.69
15	200	15	0.9439 (.0478)	.1908 (.0712)	.8176 (.0203)	49.23
30	200	15	0.9040 (.0338)	.1857 (.0614)	.7984 (.0148)	103.21
50	200	15	0.8643 (.0288)	.1718 (.0451)	.7779 (.0116)	170.20
100	200	15	0.8228 (.0179)	.1596 (.0355)	.7496 (.0089)	353.02
8	200	50	0.9867 (.0656)	.1966 (.0739)	.8379 (.0322)	84.54
15	200	50	0.9763 (.0498)	.1965 (.0724)	.8384 (.0186)	165.56
30	200	50	0.9486 (.0360)	.1944 (.0639)	.8241 (.0140)	353.36
50	200	50	0.9081 (.0291)	.1811 (.0464)	.8056 (.0112)	561.33
100	200	50	0.8572 (.0186)	.1685 (.0377)	.7743 (.0084)	1164.10
			Bias adjusted SML			
8	200	15	0.9862 (.0669)	.1956 (.0748)	.8364 (.0334)	24.90
15	200	15	0.9611 (.0492)	.1937 (.0731)	.8273 (.0200)	49.64
30	200	15	0.9159 (.0344)	.1878 (.0629)	.8039 (.0146)	104.11
50	200	15	0.8718 (.0290)	.1729 (.0460)	.7813 (.0114)	171.71
100	200	15	0.8233 (.0180)	.1595 (.0359)	.7499 (.0090)	355.94
8	200	50	0.9955 (.0666)	.1983 (.0745)	.8444 (.0322)	85.25
15	200	50	0.9882 (.0508)	.1988 (.0737)	.8448 (.0185)	166.94
30	200	50	0.9607 (.0367)	.1968 (.0654)	.8293 (.0138)	356.19
50	200	50	0.9166 (.0293)	.1826 (.0475)	.8092 (.0112)	566.29
100	200	50	0.8593 (.0189)	.1685 (.0382)	.7752 (.0085)	1173.70

Table 2.B
SML—Model: Serial correlation only; no state dependence
Root mean squared errors

T	n	m	β	λ	ρ	β	λ	ρ
			Bias unadjusted SML			Bias adjusted SML		
8	200	15	.0728	.0737	.0441	.0682	.0748	.0360
15	200	15	.0737	.0717	.0382	.0626	.0733	.0302
30	200	15	.1018	.0629	.0537	.0909	.0639	.0483
50	200	15	.1387	.0531	.0731	.1315	.0533	.0697
100	200	15	.1781	.0537	.1008	.1776	.0541	.1005
8	200	50	.0669	.0739	.0343	.0666	.0744	.0326
15	200	50	.0551	.0724	.0219	.0521	.0736	.0192
30	200	50	.0627	.0640	.0294	.0537	.0653	.0248
50	200	50	.0964	.0500	.0458	.0883	.0505	.0423
100	200	50	.1440	.0491	.0762	.1420	.0495	.0753

</div>

<table>
<tr><td colspan="2" align="center">Table 3.A</td></tr>
</table>

Table 3.A
SML–Serially correlated disturbances; random component; no state dependence
True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5} = 0.7071..$, and $\rho = 0.4$

T	n	m	β	λ	σ	ρ	time
					Bias unadjusted SML		
8	200	15	0.9694 (.0650)	.1992 (.0724)	.6408 (.1362)	.4018 (.0867)	23.28
15	200	15	0.9724 (.0456)	.1933 (.0661)	.6942 (.0741)	.3755 (.0452)	39.34
30	200	15	0.9579 (.0344)	.1902 (.0696)	.7181 (.0628)	.3472 (.0293)	89.87
50	200	15	0.9559 (.0277)	.1956 (.0745)	.7404 (.0681)	.3279 (.0183)	177.12
100	200	15	0.9486 (.0202)	.1879 (.1001)	.7903 (.0839)	.2983 (.0133)	410.54
8	200	50	0.9912 (.0658)	.2063 (.0757)	.6948 (.0974)	.3864 (.0800)	73.80
15	200	50	0.9918 (.0453)	.1959 (.0633)	.7102 (.0662)	.3865 (.0434)	123.30
30	200	50	0.9802 (.0350)	.2033 (.0654)	.7124 (.0552)	.3739 (.0272)	300.90
50	200	50	0.9798 (.0273)	.2029 (.0652)	.7253 (.0525)	.3613 (.0179)	572.90
100	200	50	0.9715 (.0178)	.1869 (.0827)	.7521 (.0692)	.3342 (.0128)	1394.93
					Bias adjusted SML		
8	200	15	0.9884 (.0672)	.2030 (.0743)	.6906 (.1302)	.3938 (.0906)	23.40
15	200	15	0.9864 (.0468)	.1958 (.0682)	.7143 (.0739)	.3843 (.0466)	39.76
30	200	15	0.9682 (.0348)	.1920 (.0718)	.7297 (.0643)	.3598 (.0297)	90.73
50	200	15	0.9637 (.0282)	.1968 (.0765)	.7482 (.0701)	.3387 (.0184)	178.60
100	200	15	0.9512 (.0208)	.1880 (.1009)	.7926 (.0849)	.3031 (.0133)	413.55
8	200	50	0.9989 (.0666)	.2082 (.0764)	.7092 (.0942)	.3844 (.0811)	74.54
15	200	50	0.9989 (.0460)	.1975 (.0641)	.7151 (.0659)	.3938 (.0443)	124.73
30	200	50	0.9878 (.0356)	.2054 (.0666)	.7158 (.0560)	.3860 (.0280)	303.82
50	200	50	0.9871 (.0277)	.2042 (.0675)	.7297 (.0539)	.3741 (.0181)	577.72
100	200	50	0.9743 (.0179)	.1867 (.0840)	.7545 (.0703)	.3414 (.0130)	1404.82

Table 3.B
SML–Serially correlated disturbances; random component; no state dependence
Root mean squared error

T	n	m	β	λ	σ	ρ	β	λ	σ	ρ
			Bias unadjusted SML				Bias adjusted SML			
8	200	15	.0717	.0722	.1513	.0866	.0681	.0742	.1311	.0906
15	200	15	.0532	.0663	.0751	.0514	.0487	.0682	.0741	.0490
30	200	15	.0544	.0702	.0637	.0604	.0471	.0722	.0680	.0500
50	200	15	.0520	.0745	.0757	.0744	.0459	.0765	.0812	.0640
100	200	15	.0552	.1006	.1180	.1025	.0531	.1014	.1204	.0978
8	200	50	.0663	.0758	.0980	.0810	.0665	.0767	.0941	.0825
15	200	50	.0460	.0633	.0662	.0454	.0459	.0641	.0663	.0446
30	200	50	.0402	.0654	.0553	.0377	.0376	.0668	.0566	.0312
50	200	50	.0339	.0652	.0554	.0426	.0305	.0676	.0583	.0316
100	200	50	.0336	.0836	.0825	.0670	.0313	.0849	.0846	.0601

4. Markov Models

4.1 Markov Model with Serially Correlated Disturbances

The Markov dynamic choice model considered in the experiment is

$$y_{it}^* = \beta x_{it} + \lambda d_{i,t-1} + \epsilon_{it}, \tag{4.1}$$

where

$$\epsilon_{it} = \rho \epsilon_{i,t-1} + w_{it},$$

and w_{it} are i.i.d. $N(0,1)$. The stochastic process starts at the initial period with $\epsilon_{i0} = 0$ and $d_{i0} = 0$ for all i. This model is a first order Markov model as choices made last period are the only prior choices relevant to current choice. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi[D_{it}(\beta x_{it} + \lambda d_{i,t-1} + \rho \epsilon_{i,t-1}^{(j)})] \right\}.$$

This model differs from the model in (3.1) in that the λ represents the effect of true state dependence. The x variable is generated by the same stochastic process before.

Tables 4.A and 4.B report simulated likelihood estimates and their root mean square errors for the case where true parameters are set to $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.4$. With a sample of size $n = 200$ and $m = 15$, biases of the SMLE of β are small. But there are upward biases in the SMLE of λ. The SMLE of ρ are biased downward. So the dynamic effect can be over stated but the serial correlation of disturbances can be underestimated. Standard errors of all the SMLE decrease as panels become longer. For the longest panel with $T = 100$, biases dominate standard errors and their root mean square errors become larger than the root mean square errors for panels with shorter periods. With $m = 50$, biases of the SMLE of λ and ρ have been substantially reduced. The bias correction procedure can also substantially reduce biases of the SMLE. By solely increasing the sample size n (from 200 to 500) but without increasing the number of simulated random variables m, biases can not be reduced as predicted by theory. Overall, estimates of this model are reasonably good for panels with T up to 50. With $m = 50$, bias-corrected estimates have small root mean square errors among all the estimates. Comparing these estimates with estimates of the regression model without state dependence in Tables 1.A and 1.B, standard errors and root mean square errors of estimates of the Markov dynamic model are larger. When small m ($m = 15$), is utilized, the estimate of the dynamic effect of the Markov model is more biased.

Tables 5.A and 5.B provide estimates for the case with a high autocorrelation $\rho = 0.85$. Biases and standard errors of the SMLE for both β and λ become larger, even though standard errors for ρ are smaller. With the bias-correction procedure and $m = 50$, estimates can be quite accurate for shorter panels with T up to 15. Biases and root mean square errors tend to be larger for panels with T equal to 30 or more.

4.2 Markov Model with Unobserved Heterogeneity and Serially Correlated Disturbances

The Markov dynamic choice model considered for Monte Carlo study in this section is

$$y_{it}^* = \beta x_{it} + \lambda d_{i,t-1} + \sigma \xi_i + \epsilon_{it}, \tag{4.2}$$

where

$$\epsilon_{it} = \rho \epsilon_{i,t-1} + w_{it},$$

and ξ_i and w_{it} are i.i.d. $N(0,1)$. The stochastic process starts at the initial period with $\epsilon_{i0} = 0$ and $d_{i0} = 0$ for all i. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi[D_{it}(\beta x_{it} + \lambda d_{i,t-1} + \sigma \xi_i^{(j)} + \rho \epsilon_{i,t-1}^{(j)})] \right\}.$$

Sample data are generated with $\beta = 1$, $\lambda = 0.2$, $\sigma^2 = 0.5$, and $\rho = 0.4$.

The SMLE and their root mean square errors are reported on Tables 6.A and 6.B. The SMLE of β are biased downward, biases of the dynamic effect λ are upward, and estimates of ρ are biased downward. The magnitude of bias increases as the length of the panel becomes longer. The SMLE of σ with small $m = 15$ are biased downward for short panels but biases become upward for long panels. Directions of bias of the SMLE for β, λ and ρ are similar to the Markov model (4.1) without unobserved heterogeneity. Biases for β for the Markov model with additional unobserved heterogeneity are slightly larger. Biases for ρ are much more severe when T are 30 or longer. Biases of the SMLE of the dynamic effect λ are also quite severe for T as long as 30 or longer. By increasing m to 50, biases of estimates are reduced. Bias-corrected estimates have smaller biases. With $m = 50$, biases of bias-corrected estimates for all the parameters seem quite small for T up to 15. In terms of root mean squares errors, estimates are improved with T up to 30. When T becomes long as 50 or 100, root mean squares errors of estimates can be larger than the ones with shorter T. By increasing m to a large 100, biases and root mean squares errors can further be reduced but the time cost is double. Bias-corrected estimates have slightly larger standard errors but their root mean square errors are, in general, smaller. In conclusion, for this model, bias-corrected estimates with $m = 50$ can be reasonably good for T up to 30.

Table 4.A

SML—serial correlated disturbances; Markov state dependence
True parameters: $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.4$

T	n	m	β	λ	ρ	time
			Bias unadjusted SML			
8	200	15	0.9928 (.0562)	.2217 (.0754)	.3682 (.0622)	20.04
15	200	15	0.9853 (.0425)	.2293 (.0520)	.3609 (.0373)	40.21
30	200	15	0.9799 (.0287)	.2339 (.0341)	.3505 (.0260)	81.65
50	200	15	0.9743 (.0235)	.2391 (.0270)	.3382 (.0191)	146.25
100	200	15	0.9648 (.0145)	.2598 (.0185)	.3153 (.0141)	318.58
8	200	50	1.0007 (.0562)	.2076 (.0768)	.3890 (.0626)	67.77
15	200	50	0.9939 (.0434)	.2138 (.0536)	.3828 (.0388)	137.84
30	200	50	0.9920 (.0293)	.2128 (.0350)	.3804 (.0277)	277.19
50	200	50	0.9877 (.0240)	.2161 (.0281)	.3713 (.0201)	483.74
100	200	50	0.9780 (.0144)	.2368 (.0193)	.3497 (.0142)	1062.70
15	500	15	0.9828 (.0233)	.2289 (.0300)	.3622 (.0235)	101.65
30	500	15	0.9795 (.0161)	.2340 (.0206)	.3515 (.0168)	239.84
			Bias adjusted SML			
8	200	15	1.0020 (.0576)	.2057 (.0766)	.3922 (.0640)	20.27
15	200	15	0.9953 (.0436)	.2123 (.0533)	.3862 (.0387)	40.62
30	200	15	0.9899 (.0295)	.2170 (.0352)	.3753 (.0270)	82.49
50	200	15	0.9833 (.0238)	.2240 (.0277)	.3605 (.0197)	147.71
100	200	15	0.9691 (.0148)	.2521 (.0187)	.3275 (.0143)	321.51
8	200	50	1.0042 (.0568)	.2014 (.0774)	.3979 (.0633)	68.48
15	200	50	0.9982 (.0440)	.2063 (.0544)	.3933 (.0397)	139.23
30	200	50	0.9977 (.0298)	.2030 (.0358)	.3939 (.0286)	280.02
50	200	50	0.9940 (.0244)	.2052 (.0288)	.3865 (.0211)	488.51
100	200	50	0.9819 (.0147)	.2297 (.0195)	.3603 (.0146)	1072.31
15	500	15	0.9927 (.0240)	.2121 (.0307)	.3870 (.0246)	102.74
30	500	15	0.9894 (.0164)	.2174 (.0212)	.3760 (.0174)	242.01

Table 4.B
SML—serial correlated disturbances; Markov state dependence
Root mean squared errors

T	n	m	β	λ	ρ	β	λ	ρ
			Bias unadjusted SML			Bias adjusted SML		
8	200	15	.0565	.0783	.0697	.0575	.0767	.0644
15	200	15	.0449	.0596	.0540	.0438	.0546	.0410
30	200	15	.0350	.0481	.0559	.0311	.0390	.0366
50	200	15	.0348	.0475	.0647	.0290	.0367	.0441
100	200	15	.0381	.0626	.0859	.0343	.0553	.0739
8	200	50	.0561	.0770	.0634	.0569	.0772	.0632
15	200	50	.0438	.0553	.0424	.0440	.0546	.0402
30	200	50	.0303	.0372	.0339	.0298	.0358	.0292
50	200	50	.0270	.0323	.0350	.0251	.0292	.0250
100	200	50	.0263	.0415	.0523	.0232	.0355	.0423

Table 5.A
SML–Serially correlated disturbances; Markov state dependence
True parameters: $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.85$

T	n	m	β	λ	ρ	time
			Bias unadjusted SML			
8	200	15	0.9744 (.0683)	.2231 (.0965)	.8179 (.0387)	24.42
15	200	15	0.9478 (.0501)	.2358 (.0695)	.8118 (.0232)	47.40
30	200	15	0.9077 (.0357)	.2626 (.0520)	.7892 (.0182)	104.58
50	200	15	0.8689 (.0282)	.2895 (.0413)	.7621 (.0157)	188.52
100	200	15	0.8277 (.0192)	.3455 (.0349)	.7222 (.0114)	386.83
8	200	50	0.9943 (.0684)	.2093 (.0948)	.8371 (.0350)	83.45
15	200	50	0.9794 (.0506)	.2135 (.0689)	.8366 (.0203)	159.61
30	200	50	0.9507 (.0372)	.2322 (.0476)	.8202 (.0159)	346.88
50	200	50	0.9116 (.0295)	.2520 (.0380)	.7975 (.0136)	627.23
100	200	50	0.8647 (.0201)	.3094 (.0340)	.7566 (.0104)	1352.99
			Bias adjusted SML			
8	200	15	0.9932 (.0706)	.2093 (.0973)	.8355 (.0377)	24.64
15	200	15	0.9642 (.0515)	.2248 (.0697)	.8231 (.0225)	47.84
30	200	15	0.9194 (.0365)	.2559 (.0525)	.7960 (.0177)	105.46
50	200	15	0.8765 (.0283)	.2851 (.0416)	.7665 (.0153)	189.94
100	200	15	0.8290 (.0195)	.3422 (.0350)	.7235 (.0114)	389.84
8	200	50	1.0028 (.0695)	.2034 (.0953)	.8446 (.0347)	84.16
15	200	50	0.9907 (.0516)	.2066 (.0692)	.8439 (.0200)	159.61
30	200	50	0.9622 (.0379)	.2259 (.0477)	.8263 (.0154)	346.88
50	200	50	0.9199 (.0297)	.2474 (.0384)	.8018 (.0134)	631.98
100	200	50	0.8674 (.0202)	.3059 (.0342)	.7584 (.0104)	1362.58

Table 5.B
SML–Serially correlated disturbances; Markov state dependence
Root mean squared errors

T	n	m	β	λ	ρ	β	λ	ρ
			Bias unadjusted SML			Bias adjusted SML		
8	200	15	.0729	.0991	.0502	.0708	.0976	.0403
15	200	15	.0723	.0781	.0447	.0626	.0738	.0351
30	200	15	.0989	.0813	.0634	.0885	.0766	.0568
50	200	15	.1341	.0985	.0893	.1267	.0947	.0849
100	200	15	.1734	.1496	.1283	.1721	.1464	.1270
8	200	50	.0686	.0951	.0372	.0694	.0952	.0351
15	200	50	.0545	.0701	.0243	.0523	.0694	.0209
30	200	50	.0617	.0574	.0338	.0535	.0542	.0283
50	200	50	.0932	.0644	.0542	.0855	.0609	.0500
100	200	50	.1368	.1146	.0940	.1342	.1113	.0922

Table 6.A

SML–Serially correlated disturbances; random component; Markov state dependence

True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5} = 0.7071..$, and $\rho = 0.4$

T	n	m	β	λ	σ	ρ	time
				Bias unadjusted SML			
8	200	15	0.9662 (.0637)	.2317 (.1081)	.6299 (.1358)	.3876 (.1124)	25.24
15	200	15	0.9654 (.0448)	.2667 (.0827)	.6766 (.0759)	.3427 (.0692)	44.65
30	200	15	0.9516 (.0347)	.3216 (.0652)	.6904 (.0622)	.2905 (.0464)	91.16
50	200	15	0.9510 (.0269)	.3775 (.0562)	.7107 (.0640)	.2525 (.0337)	157.31
100	200	15	0.9459 (.0212)	.4817 (.0526)	.7662 (.0799)	.1860 (.0266)	353.14
8	200	50	0.9884 (.0660)	.2224 (.1073)	.6851 (.0963)	.3787 (.1055)	81.35
15	200	50	0.9876 (.0458)	.2227 (.0768)	.6978 (.0712)	.3795 (.0620)	152.19
30	200	50	0.9763 (.0344)	.2598 (.0637)	.6946 (.0575)	.3442 (.0464)	334.48
50	200	50	0.9775 (.0268)	.3030 (.0541)	.6956 (.0532)	.3153 (.0341)	558.50
100	200	50	0.9713 (.0179)	.3848 (.0400)	.7168 (.0663)	.2610 (.0214)	1178.51
8	200	100	0.9927 (.0658)	.2167 (.1056)	.6942 (.0927)	.3819 (.1027)	161.69
15	200	100	0.9933 (.0467)	.2120 (.0783)	.7021 (.0714)	.3890 (.0620)	300.13
30	200	100	0.9839 (.0334)	.2337 (.0619)	.6975 (.0562)	.3671 (.0432)	667.17
50	200	100	0.9868 (.0269)	.2640 (.0513)	.6972 (.0510)	.3462 (.0334)	1151.58
100	200	100	0.9790 (.0180)	.3382 (.0387)	.7027 (.0542)	.2970 (.0201)	2378.40
				Bias adjusted SML			
8	200	15	0.9871 (.0661)	.2146 (.1111)	.6819 (.1314)	.3880 (.1171)	25.46
15	200	15	0.9812 (.0459)	.2435 (.0849)	.7000 (.0759)	.3610 (.0715)	45.07
30	200	15	0.9633 (.0353)	.2990 (.0673)	.7039 (.0636)	.3120 (.0479)	92.02
50	200	15	0.9602 (.0276)	.3604 (.0575)	.7197 (.0655)	.2708 (.0348)	158.78
100	200	15	0.9493 (.0214)	.4747 (.0533)	.7688 (.0810)	.1947 (.0274)	356.08
8	200	50	0.9969 (.0670)	.2147 (.1087)	.7007 (.0938)	.3808 (.1075)	82.09
15	200	50	0.9958 (.0465)	.2062 (.0781)	.7057 (.0714)	.3944 (.0633)	153.63
30	200	50	0.9847 (.0349)	.2346 (.0652)	.7020 (.0587)	.3668 (.0477)	337.42
50	200	50	0.9855 (.0273)	.2749 (.0555)	.7034 (.0550)	.3401 (.0347)	563.44
100	200	50	0.9747 (.0180)	.3667 (.0400)	.7203 (.0672)	.2768 (.0217)	1188.43
8	200	100	0.9973 (.0664)	.2123 (.1063)	.7017 (.0915)	.3836 (.1038)	163.14
15	200	100	0.9982 (.0471)	.2016 (.0791)	.7058 (.0714)	.3986 (.0629)	302.97
30	200	100	0.9901 (.0338)	.2144 (.0634)	.7023 (.0573)	.3843 (.0443)	683.03
50	200	100	0.9934 (.0276)	.2385 (.0525)	.7042 (.0525)	.3678 (.0342)	1161.44
100	200	100	0.9822 (.0182)	.3183 (.0386)	.7069 (.0552)	.3132 (.0202)	2398.21

Table 6.B

SML–Serially correlated disturbances; random component; Markov state dependence

Root mean squared error

T	n	m	β	λ	σ	ρ	β	λ	σ	ρ
				Bias unadjusted SML				Bias adjusted SML		
8	200	15	.0720	.1124	.1560	.1129	.0672	.1118	.1335	.1175
15	200	15	.0565	.1062	.0817	.0897	.0496	.0952	.0761	.0814
30	200	15	.0596	.1379	.0643	.1189	.0509	.1197	.0636	.1002
50	200	15	.0559	.1862	.0639	.1513	.0484	.1704	.0666	.1338
100	200	15	.0581	.2865	.0993	.2156	.0550	.2799	.1017	.2072
8	200	50	.0669	.1094	.0986	.1075	.0669	.1095	.0939	.1090
15	200	50	.0473	.0800	.0716	.0652	.0466	.0782	.0713	.0635
30	200	50	.0417	.0873	.0588	.0725	.0380	.0737	.0589	.0580
50	200	50	.0349	.1163	.0544	.0913	.0309	.0931	.0550	.0692
100	200	50	.0338	.1891	.0669	.1406	.0310	.1714	.0683	.1251
8	200	100	.0661	.1067	.0934	.1041	.0663	.1068	.0915	.1049
15	200	100	.0471	.0791	.0714	.0629	.0471	.0790	.0713	.0628
30	200	100	.0370	.0704	.0570	.0543	.0351	.0649	.0574	.0469
50	200	100	.0299	.0819	.0519	.0633	.0283	.0650	.0525	.0469
100	200	100	.0277	.1435	.0543	.1050	.0254	.1244	.0551	.0891

5. Generalized Polya Models

5.1 Polya Models with Serially Correlated Disturbances

The generalized Polya model with a depreciation factor δ is considered for our Monte Carlo experiment. The Polya model is specified as

$$y_{it}^{*} = z_{it}\beta + \lambda \sum_{j=1}^{t} \delta^{j-1} d_{i,t-j} + \epsilon_{it}, \tag{5.1}$$

and

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + w_{it},$$

where w_{it} are i.i.d. $N(0,1)$. The initial conditions are $\epsilon_{i0} = 0$ and $d_{i0} = 0$ for all i. In this model, the entire history of the process is relevant to current decision making. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi[D_{it}(\beta z_{it} + \lambda \sum_{j=1}^{t} \delta^{j} d_{i,t-1} + \rho\epsilon_{i,t-1}^{(j)})] \right\}.$$

For comparison purpose, the discount factor δ is assumed to be a known constant and is set to be 0.7. Sample data are generated with parameters $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.4$.

The SMLE and the bias-corrected estimates for this model are reported in Tables 7.A and 7.B. We have experimented with $T = 8, 15, 30,$ and 50. Biases of all the SMLE of β and λ are very small even for $m = 15$. There are some downward biases for the SMLE of ρ for large T with $m = 15$. These biases can be easily reduced by increasing m to 50 or by the bias correction procedure. Even with $m = 15$, the bias-corrected estimates are quit good. Estimates of the dynamic effect λ are accurate for all these T. Comparing estimates of this model with corresponding estimates of the Markov model in Table 4, the Polya model can be better estimated. Estimates of the dynamic effect λ not only have small biases but also have much smaller standard errors. The generalized Polya model has apparently much stronger state dependence property than the Markov property. That may explain the excellent statistical properties of these estimates. On the other hand, the computational time cost for this model is relatively more expensive. The additional time cost is due to the need of computing more tail probabilities in this model.

5.2 Polya Models with Unobserved Heterogeneity and Serially Correlated Disturbances

With a random component ξ to capture unobserved heterogeneity, the generalized Polya model for our experiment is specified as

$$y_{it}^{*} = z_{it}\beta + \lambda \sum_{j=1}^{t} \delta^{j-1} d_{i,t-j} + \sigma\xi_i + \epsilon_{it}, \tag{5.2}$$

and

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + w_{it},$$

where ξ_i and w_{it} are i.i.d. $N(0,1)$. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi \left[D_{it} \left(\beta z_{it} + \lambda \sum_{j=1}^{t} \delta^{j} d_{i,t-1} + \sigma\xi_i^{(j)} + \rho\epsilon_{i,t-1}^{(j)} \right) \right] \right\}.$$

Tables 8.A and 8.B report the SMLE and bias-corrected estimates and their root mean square errors. There are some downward biases in the SMLE of β and ρ and upward biases for estimating the dynamic effect λ. These biases are not large. Estimates of σ are biased downward for small T. There seems to be a trade-off between estimates of σ and ρ. As T becomes longer, the value of estimates of σ increases but the value of estimates of ρ decreases. For $T = 50$ and $m = 15$, the bias of ρ is 25 percent. The bias-correction procedure reduces biases. With $m = 50$ and the bias-correction procedure, estimates of this model are reasonably good for T up to 50. Comparing these estimates with the corresponding estimates of the Markov model with unobserved heterogeneity in Table 6, the dynamic effect λ and the serial correlation coefficient ρ in the Polya model can be much better estimated. They have small biases and smaller standard errors. This may be due to the stronger dynamic effect in this model. With the additional unobserved heterogeneity, estimates are less precise when compared with estimates of the Polya model without unobserved heterogeneity in (5.1). The time cost for estimation, however, turns out to be less due to the fact that choice probabilities are now less extreme as an additional disturbance is introduced.

Table 7.A
SML—serial correlated disturbances; the Polya model
True parameters: $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.4$

T	n	m	β	λ	ρ	time
			Bias unadjusted SML			
8	200	15	0.9949 (.0544)	.2054 (.0343)	.3821 (.0539)	31.32
15	200	15	0.9921 (.0450)	.2056 (.0209)	.3801 (.0348)	67.58
30	200	15	0.9856 (.0297)	.2047 (.0130)	.3709 (.0226)	152.63
50	200	15	0.9788 (.0232)	.2022 (.0082)	.3547 (.0197)	276.65
8	200	50	1.0026 (.0547)	.2018 (.0351)	.4007 (.0537)	107.75
15	200	50	1.0000 (.0459)	.2031 (.0215)	.3979 (.0361)	231.64
30	200	50	0.9970 (.0310)	.2028 (.0132)	.3953 (.0232)	505.92
50	200	50	0.9917 (.0235)	.2002 (.0084)	.3832 (.0208)	928.09
			Bias adjusted SML			
8	200	15	1.0036 (.0557)	.2016 (.0348)	.4023 (.0559)	31.55
15	200	15	1.0013 (.0460)	.2032 (.0213)	.4011 (.0363)	68.02
30	200	15	0.9950 (.0303)	.2029 (.0132)	.3914 (.0236)	153.48
50	200	15	0.9872 (.0238)	.2010 (.0084)	.3731 (.0204)	278.10
8	200	50	1.0059 (.0552)	.2003 (.0353)	.4082 (.0544)	108.48
15	200	50	1.0040 (.0465)	.2020 (.0217)	.4066 (.0370)	233.15
30	200	50	1.0022 (.0314)	.2016 (.0133)	.4062 (.0242)	508.90
50	200	50	0.9977 (.0241)	.1991 (.0086)	.3957 (.0217)	932.85

Table 7.B
SML–Serially correlated disturbances; the Polya model
Root mean squared errors

T	n	m	β	λ	ρ	β	λ	ρ
			Bias unadjusted SML			Bias adjusted SML		
8	200	15	.0546	.0347	.0567	.0557	.0348	.0558
15	200	15	.0456	.0216	.0400	.0459	.0215	.0362
30	200	15	.0330	.0138	.0368	.0306	.0135	.0250
50	200	15	.0314	.0084	.0493	.0270	.0084	.0338
8	200	50	.0547	.0351	.0536	.0554	.0352	.0550
15	200	50	.0459	.0217	.0361	.0465	.0218	.0375
30	200	50	.0311	.0134	.0237	.0315	.0134	.0249
50	200	50	.0249	.0084	.0267	.0241	.0086	.0221

Table 8.A
SML–Serially correlated disturbances; random component; the Polya model
True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5}$, and $\rho = 0.4$

T	n	m	β	λ	σ	ρ	time
			Bias unadjusted SML				
8	200	15	0.9648 (.0648)	.2201 (.0471)	.6317 (.1373)	.3903 (.0894)	25.79
15	200	15	0.9647 (.0476)	.2172 (.0296)	.6718 (.0787)	.3679 (.0516)	53.77
30	200	15	0.9503 (.0369)	.2197 (.0257)	.6946 (.0735)	.3370 (.0306)	136.88
50	200	15	0.9469 (.0286)	.2344 (.0221)	.7079 (.0700)	.3091 (.0230)	263.49
8	200	50	0.9891 (.0671)	.2111 (.0470)	.6869 (.1069)	.3843 (.0773)	85.07
15	200	50	0.9871 (.0482)	.2071 (.0306)	.6945 (.0721)	.3850 (.0523)	182.97
30	200	50	0.9773 (.0384)	.2056 (.0256)	.7066 (.0632)	.3727 (.0316)	423.29
50	200	50	0.9734 (.0289)	.2177 (.0222)	.7056 (.0605)	.3508 (.0236)	837.40
			Bias adjusted SML				
8	200	15	0.9884 (.0675)	.2089 (.0488)	.6917 (.1327)	.3876 (.0936)	26.01
15	200	15	0.9818 (.0490)	.2097 (.0306)	.6993 (.0790)	.3805 (.0537)	54.20
30	200	15	0.9628 (.0377)	.2143 (.0267)	.7102 (.0757)	.3535 (.0318)	137.75
50	200	15	0.9565 (.0293)	.2313 (.0228)	.7174 (.0720)	.3230 (.0237)	264.98
8	200	50	0.9987 (.0682)	.2064 (.0477)	.7058 (.1048)	.3845 (.0788)	85.80
15	200	50	0.9958 (.0491)	.2028 (.0312)	.7036 (.0723)	.3946 (.0536)	184.42
30	200	50	0.9864 (.0393)	.1996 (.0266)	.7150 (.0649)	.3885 (.0331)	426.18
50	200	50	0.9820 (.0296)	.2116 (.0234)	.7135 (.0629)	.3677 (.0247)	842.32

Table 8.B
SML–Serially correlated disturbances; random component; the Polya model
Root mean squared error

T	n	m	β	λ	σ	ρ	β	λ	σ	ρ
			Bias unadjusted SML				Bias adjusted SML			
8	200	15	.0736	.0511	.1565	.0898	.0684	.0495	.1334	.0942
15	200	15	.0592	.0342	.0861	.0607	.0522	.0321	.0793	.0570
30	200	15	.0618	.0323	.0744	.0700	.0529	.0302	.0756	.0563
50	200	15	.0603	.0409	.0699	.0938	.0525	.0387	.0726	.0805
8	200	50	.0679	.0482	.1086	.0787	.0681	.0480	.1046	.0802
15	200	50	.0498	.0313	.0730	.0543	.0492	.0313	.0723	.0538
30	200	50	.0445	.0262	.0631	.0417	.0415	.0266	.0653	.0350
50	200	50	.0392	.0283	.0604	.0546	.0346	.0261	.0631	.0407

6. Renewal Models

6.1 Renewal Model with Serially Correlated Disturbances

A renewal model with a discount factor δ is considered. This model for our study is specified as

$$y_{it}^* = z_{it}\beta + \lambda \sum_{j=1}^{t} \delta^{j-1} \prod_{l=1}^{j} d_{i,t-l} + \epsilon_{it}, \tag{6.1}$$

where

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + w_{it}$$

and w_{it} are i.i.d. $N(0,1)$. The process is assumed to start at the initial period with $\epsilon_{i0} = 0$. The renewal model captures the essence that the current continuous duration in one state is a determinant of the decision to remain in or exit the state. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi \left[D_{it} \left(\beta z_{it} + \lambda \sum_{j=1}^{t} \delta^{j-1} \prod_{l=1}^{j} d_{i,t-l} + \rho\epsilon_{i,t-1}^{(j)} \right) \right] \right\}.$$

The discount factor δ is assumed to be a known constant with the value of 0.7. The true parameters are set to $\beta = 1$, $\lambda = 0.2$ and $\rho = 0.4$ as for the other models.

The SMLE and bias-corrected estimates for this model and their root mean square errors are reported in Tables 9.A and 9.B. The SMLE for this model are quite good. There is some biases in estimates of ρ. These biases can be reduced by increasing m to 50 or by the bias correction procedure. These estimates are good for all the panels with T up to 50. Estimates of both the dynamic effect λ and the serial correlation coefficient ρ of this model are better in terms of smaller biases and standard errors than estimates of the Markov model in Table 4. Comparing them with estimates of the Polya model in Table 9, these estimates are only slightly worse. These results are compatible with model specification in that the true state dependence in the Renewal model is stronger than the state dependence in the Markov model but is weaker than the state dependence of the Polya model.

6.2 Renewal Models with Unobserved Heterogeneity and Serially Correlated Disturbances

Incorporating a random component ξ, the model with a discount factor δ is

$$y_{it}^* = z_{it}\beta + \lambda \sum_{j=1}^{t} \delta^{j-1} \prod_{l=1}^{j} d_{i,t-l} + \sigma\xi_i + \epsilon_{it}, \tag{6.2}$$

where

$$\epsilon_{it} = \rho\epsilon_{i,t-1} + w_{it}$$

and ξ_i and w_{it} are i.i.d. $N(0,1)$. The process is assumed to start at the initial period with $\epsilon_{i0} = 0$. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi \left[D_{it} \left(\beta z_{it} + \lambda \sum_{j=1}^{t} \delta^{j-1} \prod_{l=1}^{j} d_{i,t-l} + \sigma\xi_i^{(j)} + \rho\epsilon_{i,t-1}^{(j)} \right) \right] \right\}.$$

The discount factor is known to be 0.7. The sample data are generated with $\beta = 1$, $\lambda = 0.2$, $\sigma^2 = 0.5$, and $\rho = 0.4$.

Results for the estimation of this model are reported on Tables 10.A and 10.B. There are upward biases in the SMLE of λ and downward biases in estimates of ρ. The magnitude of bias is increasing as T becomes longer. Biases in estimates of β and ρ are relatively small. With large $m = 50$, all the biases are reduced. The bias-correction procedure has also reduced the bias. The bias-corrected estimates with $m = 50$ are pretty good. They have small biases and small root mean square errors. The model can be estimated adequately for panels with T up to 50. As the state dependence in this model is relatively weaker than a corresponding Polya model, estimates of this model are slightly less accurate in terms of biases and standard errors than estimates of the Polya model in Table 8. On the other hand, these estimates are much better than estimates of the Markov model with unknown heterogeneity in Table 6. The computing time cost of this model with unobserved heterogeneity is slightly less than the time cost of the renewal model without unobserved heterogeneity in (6.1). It is about the same computing cost for a Polya model with unobserved heterogeneity.

Table 9.A
SML—serial correlated disturbances; the Renewal model
True parameters: $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.4$

T	n	m	β	λ	ρ	time
			Bias unadjusted SML			
8	200	15	0.9953 (.0559)	.2077 (.0392)	.3757 (.0611)	27.90
15	200	15	0.9871 (.0432)	.2104 (.0259)	.3716 (.0363)	65.01
30	200	15	0.9806 (.0316)	.2117 (.0161)	.3613 (.0252)	142.37
50	200	15	0.9755 (.0240)	.2131 (.0106)	.3476 (.0188)	262.83
8	200	50	1.0040 (.0561)	.2004 (.0402)	.3969 (.0628)	95.30
15	200	50	0.9970 (.0442)	.2037 (.0272)	.3936 (.0381)	221.65
30	200	50	0.9946 (.0328)	.2039 (.0171)	.3905 (.0270)	473.70
50	200	50	0.9909 (.0244)	.2048 (.0113)	.3799 (.0207)	868.00
			Bias adjusted SML			
8	200	15	1.0054 (.0575)	.1996 (.0402)	.3997 (.0636)	28.13
15	200	15	0.9982 (.0444)	.2032 (.0269)	.3968 (.0382)	65.42
30	200	15	0.9919 (.0326)	.2051 (.0168)	.3860 (.0267)	143.21
50	200	15	0.9856 (.0245)	.2075 (.0109)	.3696 (.0196)	264.27
8	200	50	1.0079 (.0567)	.1972 (.0407)	.4059 (.0638)	96.02
15	200	50	1.0018 (.0449)	.2004 (.0277)	.4043 (.0393)	223.03
30	200	50	1.0011 (.0336)	.1999 (.0177)	.4041 (.0283)	476.66
50	200	50	0.9981 (.0249)	.2007 (.0117)	.3950 (.0219)	872.76

Table 9.B
SML–Serially correlated disturbances; the Renewal model
Root mean squared errors

T	n	m	Bias unadjusted SML			Bias adjusted SML		
			β	λ	ρ	β	λ	ρ
8	200	15	.0560	.0398	.0656	.0577	.0401	.0635
15	200	15	.0450	.0279	.0461	.0443	.0270	.0383
30	200	15	.0371	.0199	.0461	.0335	.0175	.0301
50	200	15	.0343	.0168	.0557	.0284	.0132	.0362
8	200	50	.0561	.0402	.0628	.0571	.0407	.0640
15	200	50	.0442	.0274	.0386	.0448	.0277	.0394
30	200	50	.0332	.0175	.0286	.0336	.0177	.0286
50	200	50	.0260	.0123	.0288	.0249	.0117	.0224

Table 10.A
SML–Serially correlated disturbances; random component; the Renewal model
True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5}$, and $\rho = 0.4$

T	n	m	β	λ	σ	ρ	time
			Bias unadjusted SML				
8	200	15	0.9645 (.0646)	.2249 (.0558)	.6342 (.1251)	.3815 (.0943)	27.79
15	200	15	0.9610 (.0464)	.2318 (.0343)	.6617 (.0783)	.3458 (.0587)	55.96
30	200	15	0.9507 (.0355)	.2367 (.0269)	.6796 (.0635)	.3220 (.0376)	131.49
50	200	15	0.9465 (.0288)	.2562 (.0223)	.7008 (.0699)	.2876 (.0259)	260.57
8	200	50	0.9894 (.0670)	.2146 (.0554)	.6873 (.1056)	.3806 (.0884)	93.14
15	200	50	0.9862 (.0464)	.2135 (.0366)	.6933 (.0755)	.3746 (.0583)	202.58
30	200	50	0.9736 (.0354)	.2191 (.0265)	.6746 (.0555)	.3607 (.0370)	440.09
50	200	50	0.9757 (.0292)	.2285 (.0222)	.6991 (.0488)	.3444 (.0265)	820.62
			Bias adjusted SML				
8	200	15	0.9882 (.0671)	.2120 (.0577)	.6925 (.1211)	.3832 (.0981)	28.00
15	200	15	0.9788 (.0476)	.2201 (.0357)	.6898 (.0788)	.3635 (.0613)	56.39
30	200	15	0.9641 (.0362)	.2261 (.0279)	.6968 (.0650)	.3437 (.0390)	132.36
50	200	15	0.9569 (.0296)	.2489 (.0231)	.7120 (.0717)	.3059 (.0268)	262.02
8	200	50	0.9989 (.0680)	.2091 (.0562)	.7054 (.1035)	.3833 (.0904)	93.88
15	200	50	0.9954 (.0473)	.2061 (.0377)	.7036 (.0760)	.3880 (.0601)	204.02
30	200	50	0.9833 (.0359)	.2089 (.0277)	.6839 (.0568)	.3813 (.0387)	443.03
50	200	50	0.9850 (.0299)	.2169 (.0234)	.7098 (.0500)	.3679 (.0278)	825.54

Table 10.B
SML–Serially correlated disturbances; random component; the Renewal model
Root mean squared error

T	n	m	Bias unadjusted SML				Bias adjusted SML			
			β	λ	σ	ρ	β	λ	σ	ρ
8	200	15	.0736	.0610	.1446	.0959	.0681	.0588	.1217	.0994
15	200	15	.0605	.0467	.0904	.0799	.0520	.0409	.0806	.0713
30	200	15	.0607	.0454	.0691	.0866	.0509	.0382	.0657	.0685
50	200	15	.0607	.0605	.0701	.1153	.0523	.0540	.0717	.0978
8	200	50	.0677	.0572	.1073	.0904	.0679	.0569	.1033	.0918
15	200	50	.0483	.0389	.0766	.0635	.0474	.0381	.0759	.0612
30	200	50	.0441	.0326	.0642	.0539	.0396	.0291	.0612	.0429
50	200	50	.0380	.0361	.0494	.0616	.0335	.0289	.0500	.0424

7. Habit Persistent Models

7.1 Models with Serially Correlated Disturbances

The habit persistent model is specified as follows:

$$y_{it}^* = z_{it}\beta + \lambda y_{i,t-1}^* + \epsilon_{it}, \qquad (7.1)$$

where

$$\epsilon_{it} = \rho \epsilon_{i,t-1} + w_{it}$$

and w_{it} are i.i.d. $N(0,1)$. The process starts at the initial period with $y_{i0}^* = 0$ and $\epsilon_{i0} = 0$. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi \left[D_{it} \left(\beta z_{it} + \lambda y_{i,t-1}^{*(j)} + \rho \epsilon_{i,t-1}^{(j)} \right) \right] \right\}.$$

The true parameters are $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.4$.

The SMLE and bias-corrected estimates and their root mean square errors are reported in Tables 11.A and 11.B. There are downward biases in the SMLE for all the parameters β, λ and ρ. Biases for β and λ are relatively smaller than biases for ρ. With m increased to 50, biases are reduced and those biases for β and λ are small. The bias-correction procedure can further reduce bias and biases for ρ can become small. Bias-corrected estimates have also smaller mean squared errors than those of the unadjusted SMLE. The parameters in this model can be accurately estimated with T up to 30 or even 50. Comparing these estimates with estimates of the previous dynamic models, the main difference is on the direction of bias of estimates of the dynamic effect λ. While estimates of λ tend to be upward in Markov, Polya and Renewal models, the bias in the habit persistent model tends to be downward. The dynamic coefficient λ can be better estimated in terms of smaller bias and standard error than the corresponding estimates of the Markov model in Table 4. However, they are not better estimated than the ones in Polya and Renewal models. The habit persistent model above is a first order latent Markov model. One may suspect that the dynamic effect for this model is stronger than the dynamic effect of the corresponding Markov model but is weaker than the accumulated effects of Polya and Renewal models. On the other hand, estimates of β and ρ in this model have larger root mean square errors than the ones in the Markov model. The β and ρ in Polya and Renewal models are also better estimated. In terms of time cost, it takes more time to estimate the habit persistence model than the Markov model but it costs less than the computation of Polya and Renewal models.

7.2 Models with Unobserved Heterogeneity and Serially Correlated Disturbances

The habit persistent model with a random component is specified as follows:

$$y_{it}^* = z_{it}\beta + \lambda y_{i,t-1}^* + \sigma \xi_i + \epsilon_{it}, \qquad (7.2)$$

where

$$\epsilon_{it} = \rho \epsilon_{i,t-1} + w_{it}$$

and ξ_i and w_{it} are i.i.d. $N(0,1)$. The process is assumed to start at the initial period with $y_{i0}^* = 0$ and $\epsilon_{i0} = 0$. The simulated log likelihood function for this model is

$$\mathcal{L} = \sum_{i=1}^{n} \ln \left\{ \frac{1}{m} \sum_{j=1}^{m} \prod_{t=1}^{T} \Phi[D_{it}(\beta z_{it} + \lambda y_{i,t-1}^{*(j)} + \sigma \xi_i^{(j)} + \rho \epsilon_{i,t-1}^{(j)})] \right\}.$$

As for the other models with unknown heterogeneity, the true parameters are $\beta = 1$, $\lambda = 0.2$, $\sigma^2 = 0.5$, and $\rho = 0.4$ to generate sample observations.

Tables 12.A and 12.B report results for this model. With the additional heterogeneity, estimates of β, λ and ρ have larger downward biases than the ones for the habit persistent model without unobserved heterogeneity. The magnitude of bias becomes larger as T becomes longer. For σ, there is some bias in its estimate for panels with small T. As m increases, biases decrease. The bias-corrected procedure can further reduce bias and root mean squares errors are in general also smaller. With $m = 50$ and the bias-corrected procedure, estimates are reasonably good for panels with T up to 15 or 30. Estimates of λ and ρ are more accurate than corresponding estimates of the Markov model with unknown heterogeneity in Table 6 in terms of bias and root mean square error. On the other hand, estimates of β and ρ in this model are less precise in terms of bias and root mean square error than estimates of the Markov model with unobserved heterogeneity. Estimates of this habit persistent model are less accurate than estimates of the corresponding Polya and Renewal models.

Table 11.A

Table 11.A
SML—serial correlation; the Habit Persistence model
True parameters: $\beta = 1$, $\lambda = 0.2$, and $\rho = 0.4$

T	n	m	β	λ	ρ	time
			Bias unadjusted SML			
8	200	15	0.9849 (.0577)	.1932 (.0521)	.3666 (.0660)	26.04
15	200	15	0.9737 (.0450)	.1960 (.0358)	.3512 (.0453)	56.81
30	200	15	0.9644 (.0303)	.1887 (.0267)	.3365 (.0339)	133.45
50	200	15	0.9546 (.0255)	.1863 (.0224)	.3172 (.0287)	247.35
8	200	50	0.9952 (.0591)	.1977 (.0511)	.3852 (.0646)	87.70
15	200	50	0.9886 (.0463)	.2013 (.0342)	.3776 (.0422)	194.82
30	200	50	0.9823 (.0315)	.1942 (.0249)	.3699 (.0313)	441.81
50	200	50	0.9734 (.0258)	.1902 (.0209)	.3536 (.0258)	811.17
			Bias adjusted SML			
8	200	15	0.9973 (.0593)	.1976 (.0515)	.3892 (.0647)	26.25
15	200	15	0.9866 (.0464)	.1992 (.0349)	.3753 (.0444)	57.22
30	200	15	0.9754 (.0310)	.1908 (.0257)	.3576 (.0326)	134.35
50	200	15	0.9636 (.0257)	.1874 (.0222)	.3343 (.0277)	248.78
8	200	50	1.0003 (.0599)	.1998 (.0509)	.3939 (.0642)	88.42
15	200	50	.9955 (.0472)	.2036 (.0338)	.3892 (.0420)	196.21
30	200	50	.9905 (.0321)	.1967 (.0244)	.3843 (.0307)	444.69
50	200	50	.9812 (.0261)	.1920 (.0207)	.3676 (.0253)	816.01

Table 11.B
SML–Serially correlation; the Habit Persistence model
Root mean squared errors

T	n	m	β	λ	ρ	β	λ	ρ
			Bias unadjusted SML			**Bias adjusted SML**		
8	200	15	.0595	.0525	.0738	.0593	.0515	.0655
15	200	15	.0521	.0359	.0665	.0482	.0349	.0507
30	200	15	.0467	.0289	.0720	.0395	.0273	.0534
50	200	15	.0520	.0262	.0876	.0445	.0255	.0713
8	200	50	.0592	.0511	.0661	.0598	.0508	.0644
15	200	50	.0476	.0342	.0477	.0473	.0340	.0433
30	200	50	.0361	.0256	.0434	.0335	.0246	.0344
50	200	50	.0370	.0231	.0531	.0321	.0222	.0411

Table 12.A
SML-Serially correlation; random component; the Habit Persistence model
True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5}$, and $\rho = 0.4$

T	n	m	β	λ	σ	ρ	time
			Bias unadjusted SML				
8	200	15	0.9560 (.0692)	.1833 (.0747)	.6292 (.1511)	.4027 (.1366)	33.62
15	200	15	0.9446 (.0467)	.1816 (.0602)	.6836 (.0899)	.3560 (.0928)	73.32
30	200	15	0.9134 (.0356)	.1772 (.0450)	.6978 (.0818)	.3127 (.0569)	164.78
50	200	15	0.9035 (.0294)	.1683 (.0395)	.7090 (.0790)	.2922 (.0482)	289.60
8	200	50	0.9872 (.0711)	.1992 (.0651)	.6893 (.1104)	.3842 (.1086)	109.89
15	200	50	0.9771 (.0497)	.1977 (.0500)	.7013 (.0809)	.3716 (.0740)	241.21
30	200	50	0.9489 (.0367)	.1929 (.0398)	.6962 (.0788)	.3458 (.0486)	560.23
50	200	50	0.9402 (.0294)	.1818 (.0301)	.6958 (.0671)	.3298 (.0356)	1013.38
			Bias adjusted SML				
8	200	15	0.9789 (.0727)	.1918 (.0700)	.6740 (.1465)	.4004 (.1333)	33.85
15	200	15	0.9623 (.0482)	.1872 (.0563)	.7030 (.0903)	.3680 (.0880)	73.76
30	200	15	0.9251 (.0366)	.1791 (.0427)	.7087 (.0830)	.3265 (.0542)	165.70
50	200	15	0.9114 (.0303)	.1688 (.0385)	.7157 (.0800)	.3028 (.0466)	291.07
8	200	50	0.9984 (.0727)	.2035 (.0635)	.7021 (.1073)	.3857 (.1061)	110.64
15	200	50	0.9894 (.0508)	.2024 (.0483)	.7068 (.0812)	.3828 (.0720)	242.70
30	200	50	0.9610 (.0378)	.1967 (.0383)	.7019 (.0801)	.3603 (.0469)	563.22
50	200	50	0.9497 (.0300)	.1839 (.0294)	.7016 (.0686)	.3430 (.0345)	1018.62

Table 12.B
SML-Serially correlation; random component; the Habit Persistence model
Root mean squared error

T	n	m	β	λ	σ	ρ	β	λ	σ	ρ
			Bias unadjusted SML				**Bias adjusted SML**			
8	200	15	.0819	.0764	.1698	.1364	.0755	.0703	.1500	.1330
15	200	15	.0725	.0629	.0928	.1026	.0611	.0577	.0902	.0935
30	200	15	.0936	.0504	.0822	.1042	.0834	.0475	.0829	.0913
50	200	15	.1009	.0505	.0789	.1181	.0937	.0495	.0803	.1078
8	200	50	.0721	.0650	.1117	.1096	.0726	.0634	.1072	.1069
15	200	50	.0547	.0499	.0809	.0791	.0518	.0483	.0810	.0739
30	200	50	.0628	.0404	.0794	.0728	.0543	.0384	.0801	.0614
50	200	50	.0666	.0352	.0679	.0787	.0585	.0335	.0687	.0666

8. Conclusions

This article has reported Monte Carlo results of simulated maximum likelihood methods for the estimation of discrete choice regression and dynamic models. Choice probabilities are simulated with the GHK recursive estimator. We investigate finite sample performance of the simulated likelihood estimator and a bias-correction procedure which can eliminate the leading bias due to simulation. The discrete choice models are panel choice models introduced in Heckman (1981a), that include discrete choice regression models with true state dependence, spurious correlation, and unobserved heterogeneity. Even with a binary choice in each time period, the pattern or path of choices can be very large as time periods become longer. Our intention is two folds. We would like to investigate finite sample performance of the simulated maximum likelihood estimator for panel data dynamic models, limitation of simulated likelihood methods, and the value of long panels compared with relatively short panels. The second purpose is to study the econometric estimation of various interesting dynamic discrete choice models, the ability of estimating dynamic effects, and the distinction of true state dependence from spurious state dependence and unobserved heterogeneity.

As an estimation method, the simulated maximum likelihood approach is computationally tractable and numerically stable. We have experimented with panel data with time length up to 100. There is not a single case of failure of convergence of the iterative maximization algorithm. Models with moderate serial correlation can be more accurately estimated than models with high serial correlation in terms of smaller bias and standard error. Regression models are better estimated than Markov state dependence models. Dynamic models with strong dynamic effects such as the generalized Polya model and the renewal model can be better estimated than the Markov model and the latent Markov model. In terms of time cost, the estimation of generalized Polya and renewal models are more expensive than the estimation of the Markov model and the latent Markov model. With a small number of simulation draws for the construction of simulator and relatively long panels, dynamic effects of Markov, Polya and renewal models can be overestimated and the spurious correlation can be underestimated. Only for the latent Markov model, both the dynamic and spurious correlation effects are underestimated. While standard errors of estimates tend to be smaller for panels with longer periods, the bias for panel data with long periods such as 100 can dominate the standard error and results in a larger root mean square error in some of the models. Biases of the SMLE can usually be reduced with a larger number of simulated variables for the construction of simulator and by the bias correction procedure. With the bias correction procedure and a moderate number of simulated draws for the

simulator, say 50, most of the dynamic models with moderate autocorrelation can be adequately estimated for panels up to 30 periods. Due to strong state dependence, Polya and Renewal models can be adequately estimated up to 50 periods in our Monte Carlo experiments.

While our Monte Carlo experiments have studied many important aspects of simulation estimation methods and efficient estimation of dynamic panel statistical models, many interesting issues have not been addressed. We have not investigated statistical inference procedures on model discrimination or model selection based on simulation estimation methods. For the estimation of dynamic panel models, we have not investigated the initial value problem, which can be quite difficult and is an open challenging question [Heckman (1981b)]. Furthermore, we have not investigated intertemporal optimization dynamic choice models where extra complexity will be encountered in backward evaluation of value functions in the presence of serially correlated disturbances [Pakes (1987), Rust (1987), and Wolpin (1984)]. All these issues remain for further investigation.

References

Borsch-Supan, A. and V.A. Hajivassiliou (1993). "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models." *Journal of Econometrics* 58: 347-368.

Butler, J.S. and R. Moffitt (1982). "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model." *Econometrica* 50: 761-764.

Chamberlain, G. (1982). "Multivariate Regression Models for Panel Data." *Journal of Econometrics* 18: 5-46.

Geweke, J. (1989). "Efficient Simulation from the Multivariate Normal Distribution Subject to Linear Inequality Constraints and the Evaluation of Constraint Probabilities." memo. Duke University, Durham, NC.

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 (1990). "The Method of Simulated Scores for the Estimation of LDV models: With an Application to the Problem of External Debt Crises." Cowles Foundation Discussion Paper No. 967, Yale University, New Haven, CT.

Heckman, J.J. (1978). "Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence Against the Hypothesis of Spurious State Dependence." *Annales de l'INSEE* 30-1: 227-269.

—— (1981a). "Statistical Models for Discrete Panel Data," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C.F. Manski and D. McFadden, pp.114-178. Cambridge: MIT Press.

—— (1981b). "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process." In *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C.F. Manski and D. McFadden, pp. 179-195. Cambridge: MIT Press.

—— (1981c). "Heterogeneity and State Dependence." In *Studies in Labor Markets*, ed. by S. Rosen, pp. 91-139. Chicago: University of Chicago Press.

Hsiao, C. (1986). *Analysis of Panel Data.* Cambridge: Cambridge University Press.

Keane, M.P. (1994). "A Computationally Practical Simulation Estimator for Panel Data." *Econometrica* 62: 95-116.

Lee, L.F. (1987). "Non-Parametric Testing of Discrete Panel Data Models." *Journal of Econometrics* 34: 147-177.

Lee, L.F. (1992). "On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models." *Econometric Theory* 8: 518-552.

Lee, L.F. (1993). "Asymptotic Bias in Maximum Simulated Likelihood Estimation of Discrete Choice Models." CREST and Department of Economics Working Paper Series, Number 93-03. The University of Michigan, Ann Arbor, Michigan.

Lerman, S. and C. Manski (1981). "On the Use of Simulated Frequencies to Approximate Choice Probabilities." In *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden. Cambridge: MIT Press.

McFadden, D. (1973). "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, ed. P. Zarembka. New York: Academic Press.

McFadden, D. (1989). "A Method of Simulated Moments of Estimation of Discrete Choice Models without Numerical Integration." *Econometrica* 57: 995-1026.

Pakes, A. (1987). "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica* 54: 755-784.

Pakes, A. and D. Pollard (1989). "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 54: 755-785.

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling (1986). *Numerical Recipes.* New York: Cambridge University Press.

Rust, J. (1987). "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica* 55: 999-1035.

Wolpin, K.I. (1984). "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality." *Journal of Political Economy* 92: 852-874.

DATE DUE

MAR 31 1998	