Center for Research on Economic and Social Theory
and
Department of Economics
Working Paper Series

Simulated Maximum Likelihood Estimation of Discrete Models with Group Data

Lung-fei Lee

November, 1993
Number 93-29

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

Simulated Maximum Likelihood Estimation of Discrete Response Models

with Group Data

by

Lung-Fei Lee

Department of Economics

The University of Michigan

Ann Arbor, MI 48109-1220

Tel: 313-764-2363

November, 1993

Abstract

This article has compared the performance of two methods of simulated maximum likelihood for the estimation of discrete choice models with group data. One method of simulated likelihood uses simulators which are statistically independent across individuals in the sample. The alternative method allows simulators to be correlated across individuals. The comparisons are based on the criteria of statistical efficiency and computation time cost. As the simulated maximum likelihood method with dependent simulators can take into account the presence of sufficient statistics in group data, it can have advantages over the simulated likelihood method with independent simulators in term of computation cost saving and statistical efficiency. The computation time cost of the simulated likelihood method with dependent simulators can be inexpensive as the method of simulated moments of McFadden (1989). This is so especially for group data with either large sample sizes or small number of groups. Besides theoretical analysis, Monte Carlo results provide some evidence.

Correspondence Address:

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

Simulated Maximum Likelihood Estimation of Discrete Response Models

with Group Data

Lung-Fei Lee*

University of Michigan

Ann Arbor, Michigan

Date: November 1993

1. Introduction

Estimation methods by simulation have been introduced by Lerman and Manski (1981), McFadden (1989), and Pakes and Pollard (1989), among others. There are method of simulated moments (MSM) and simulated maximum likelihood method (SML)(or simulated M-estimation method). McFadden (1989) has concentrated on the MSM that use additive structure of moment equations. Efficiency of the MSM depends on chosen moment equations. The SML method is conceptually more straightforward. Earlier work by Lerman and Manski (1981) and Pakes (1986) have shown the need of a large amount of simulations for the SML. The MSM of McFadden is introduced to overcome such a burden. However, recent work by Borsch-Supan and Hajivassiliou (1993) has displayed that only a moderate amount of simulations is needed for the SML method if the probability simulator is a good one. The simulated likelihood function is constructed by replacing the response probabilities by simulated probabilities. The simulated maximum likelihood estimator (SMLE) is derived by maximizing the simulated likelihood function. Following McFadden (1989), the simulated probabilities in Borsch-Supan and Hajivassiliou (1993) are generated independently for each individual decision unit in the sample. Incidentally, the simulated likelihood method used in the empirical study of patents in Pakes involves statistically dependent simulators, i.e., simulators that can be correlated across different individual units.

The asymptotic properties of the MSM have been studied in McFadden (1989) and Pakes and Pollard (1989). Lee (1992a,1992b) has pointed out some differences on asymptotic properties of the SML method with dependently simulated probabilities (SML-DS) and the SML with independently simulated probabilities

* I appreciate having financial support from NSF under grants SES-9208620 and SBR-9223325 for my research. This article was motivated by comments on my previous work on simulated maximum likelihood estimation by Professors Ariel Pakes and Steven Stern.

(SML-IS). If choice probabilities involve some continuous explanatory variables, the SML-IS will be preferred to the SML-DS in terms of statistical efficiency when computations costs are equalizing. The dependency of simulators can reduce statistical efficiency due to correlation of simulated probabilities. Some Monte Carlo results are reported to confirm such performances in finite samples in Lee (1992a).

In contrary to the previous studies, this article considers the circumstance with aggregated (or grouped) data for the estimation of discrete choice models. For models with aggregated data, as the sample observations can be divided into groups according to values of explanatory variables and choice patterns, there are sufficient statistics summarizing the sample information. For such models, it seems natural to approach estimation with dependent simulators than with independent simulators. In this article, we will investigate the issues of practical values of dependent simulators vs. independent simulators. Monte Carlo comparisons of statistical and computational efficiencies are reported. The general conclusion is that when the number of aggregated groups are not large, the SML-DS may be preferred to the SML-IS in terms of both statistical and computational efficiencies.

This article is organized as follows. In section 2, we review the asymptotic properties of SML methods with dependent and independent simulators derived in our previous work. In section 3 we will focus on the efficiency comparison when the computation costs are controlled for. Monte Carlo results are provided in section 4 to report the performance of the two methods in finite samples. Section 5 summarize the findings.

2. Asymptotic Properties of Simulated Likelihood Estimation Methods: Review

For a discrete choice model with L alternatives, let $P(l|x,\theta)$ be the response probability of alternative l, $l = 1,\ldots,L$, where x denotes a vector of explanatory variables and θ is a finite dimensional parameter vector. Let d_l be the dichotomous indicator of alternative l, which takes the value 1 when the alternative l is chosen and is zero, otherwise. With a cross-section independent sample of size n, the log likelihood function is

$$\mathcal{L}_c(\theta) = \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li} \ln P(l|x_i,\theta). \tag{2.1}$$

The simulated likelihood method replaces the computational difficult response probabilities by some statistical simulated probabilities. Various simulators have been introduced in McFadden (1989), Hajivassiliou and McFadden (1990), Borsch-Supan and Hajivassiliou (1993), and Stern (1992), among others. So far, Monte Carlo evidence in Borsch-Supan and Hajivassiliou suggests that the importance sampling simulator suggested in Borsch-Supan and Hajivassiliou (1993), which is known as the Geweke-Hajivassiliou-Keane simulator [see Geweke (1989) and Keane (1990)], provides the most accurate stochastic simulator to the choice probabilities. In this article, our concern is not on the performance of any particular simulator, but on the statistical and computational efficiencies of using simulators that are statistically independent or not for SML estimation.

Let $h_l(v,x,\theta)$, $l = 1,\ldots,L$, be smooth kernel functions (a terminology from U-statistics, not nonparametric kernel estimation) used in the construction of simulators such that $\int h_l(v,x,\theta)\gamma(v)dv = P(l|x,\theta)$, where $\gamma(v)$ is the density function of simulated random variable. Averaging over Monte Carlo draws from $\gamma(v)$ gives a smooth unbiased estimator of $P_l(x,\theta)$. Let i denote an individual in the sample with explanatory variable x_i. Simulators that are independent across observations can be constructed as

$$f^I_{r,l,i}(x_i,\theta) = \frac{1}{r}\sum_{j=1}^{r} h_l(v_i^{(j)}, x_i, \theta), \tag{2.2}$$

where the $v_i^{(j)}$, $j = 1,\ldots,r$, are r independent random draws from $\gamma(.)$ for each individual i, which are also independent across different i, $i = 1,\ldots,n$. Dependent simulators can be constructed by drawing m random variables, say, $v^{(1)},\ldots,v^{(m)}$, and using them to construct the simulators

$$f^D_{m,l}(x_i,\theta) = \frac{1}{m}\sum_{j=1}^{m} h_l(v^{(j)}, x_i, \theta) \tag{2.3}$$

for all i, $i = 1, \ldots, n$. Based on these simulators, the simulated log likelihood function with independent simulators is

$$\mathcal{L}_I(\theta) = \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li} \ln f^I_{r,l,i}(x_i, \theta), \tag{2.4}$$

and the one with dependent simulators is

$$\mathcal{L}_D(\theta) = \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li} \ln f^D_{m,l}(x_i, \theta). \tag{2.5}$$

The simulated maximum likelihood estimator (SMLE) can be derived by maximizing the above (pseudo-likelihood) functions with respect to θ. Let $\hat{\theta}_D$ and $\hat{\theta}_I$ denote, respectively, the SMLE of θ based on dependent simulators and independent simulators. The independent simulator approach is the one emphasized in the simulation estimation literature [McFadden (1989), Hajivassiliou and McFadden (1990), Borsch-Supan and Hajivassiliou (1993), and many more]. Some asymptotic properties for both the SML-IS and SML-DS are derived in Lee (1992a, 1992b) among others under some mild regularity conditions [see also Gourieroux and Monfort (1993)]. Due to the nonlinearity of log simulated likelihood functions in terms of simulation errors $f^I_{r,l,i}(x_i, \theta) - P(l|x_i, \theta)$ and $f^D_{m,l}(x_i, \theta) - P(l|x_i, \theta)$. Consistency of the SMLEs of θ requires that r and m go to infinity.

The rate of convergence and the limiting distribution of $\hat{\theta}_D - \theta_o$ where θ_o denotes the true parameter vector can be derived by the theory of generalized U-statistics [Lee (1990)]. The rate of convergence depends on the sample size n relative to the size m of total random draws. If $\lim_{n\to\infty}(n/m) = \lambda$ is finite, the rate of convergence is of order $1/\sqrt{n}$. On the other hand, if $\lim_{n\to\infty}(n/m) = \infty$, the rate of convergence can only be of order $1/\sqrt{m}$. When $\lim_{n\to\infty}(n/m) = \lambda$ is finite,

$$\sqrt{n}(\hat{\theta}_D - \theta_o) \xrightarrow{D} N(0, \Sigma_D), \tag{2.6}$$

where

$$\Sigma_D = \Sigma + \lambda\Delta, \tag{2.7}$$

with

$$\Sigma = \left\{ E\left[\sum_{l=1}^{L} P(l|x, \theta_o) \frac{\partial \ln P(l|x, \theta_o)}{\partial\theta} \frac{\partial \ln P(l|x, \theta_o)}{\partial\theta'}\right]\right\}^{-1}, \tag{2.8}$$

and

$$\begin{aligned}\Delta = &\Sigma \cdot E\left\{\sum_{l=1}^{L} E\left[\frac{\partial h_l(v, x, \theta_o)}{\partial\theta} - \frac{\partial \ln P(l|x, \theta_o)}{\partial\theta} h_l(v, x, \theta_o)\bigg| v\right]\right. \\ &\left.\times \sum_{l=1}^{L} E\left[\frac{\partial h_l(v, x, \theta_o)}{\partial\theta} - \frac{\partial \ln P(l|x, \theta_o)}{\partial\theta} h_l(v, x, \theta_o)\bigg| v\right]'\right\} \cdot \Sigma.\end{aligned} \tag{2.9}$$

The $\hat{\theta}_D$ is not asymptotically efficient unless m converges to infinite faster than n, that is, unless $\lambda = 0$. The second term in (2.7) reflects the error introduced in the simulated likelihood due to simulation of the response probabilities. For the case that m goes to infinite at a rate slower than n, $\lambda = \infty$ and

$$\sqrt{m}(\hat{\theta}_D - \theta_o) \xrightarrow{D} N(0, \Delta), \tag{2.10}$$

as noises in the simulators dominate the model noise.

For the approach with independent simulators, when $\lim_{n\to\infty}(n^{1/2}/r) = 0$, the asymptotic distribution of $\hat{\theta}_I$ will be asymptotically efficient, i.e.,

$$\sqrt{n}(\hat{\theta}_I - \theta_o) \xrightarrow{D} N(0, \Sigma). \tag{2.11}$$

This includes the case that r increases proportionally to n. Under such designs, $\hat{\theta}_I$ is efficient relative to $\hat{\theta}_D$. However, when r increases slower than $\sqrt{n}$, the limiting distribution of $\hat{\theta}_I$ becomes nasty. When $\lim_{n\to\infty}(n^{1/2}r^{-1}) = \mu$ is a finite positive constant,

$$\sqrt{n}(\hat{\theta}_I - \theta_o) \xrightarrow{D} N(\mu\Sigma\omega, \Sigma), \tag{2.12}$$

where

$$\omega = \sum_{l=1}^{L} E\left\{\frac{1}{P(l|x, \theta_o)}\left[\frac{\partial \ln P(l|x, \theta_o)}{\partial\theta} \mathrm{var}(h_l(v, x, \theta_o)|x) - \mathrm{cov}\left(h_l(v, x, \theta_o), \frac{\partial h_l(v, x, \theta_o)}{\partial\theta}\bigg| x\right)\right]\right\}. \tag{2.13}$$

Thus when r increases at the rate $\sqrt{n}$, an asymptotic bias exists in the limiting distribution. The situation becomes worse when r increases at a rate slower than $\sqrt{n}$ because the asymptotic bias will dominate the variance. Indeed, when $\lim_{n\to\infty}(n^{1/2}r^{-1}) = \infty$,

$$r(\hat{\theta}_I - \theta_o) \xrightarrow{P} \Sigma\omega, \tag{2.14}$$

[see Lee (1992b), Theorem 2]. In conclusion, while the θ_I may be efficient relative to θ_D when m and r are approximately the same, there might be biases in the limiting distribution of $\hat{\theta}_I$ when r is not large compared to n. It is interesting that biases due to simulation in $\hat{\theta}_I$ can be reduced by a bias-corrected procedure. A bias-adjusted estimator $\hat{\theta}_A$ can be derived by correcting a bias term directly from $\hat{\theta}_I$:

$$\hat{\theta}_A = \hat{\theta}_I - \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li} \frac{\partial \ln f^I_{r,l,i}(x_i, \hat{\theta}_I)}{\partial\theta} \frac{\partial \ln f^I_{r,l,i}(x_i, \hat{\theta}_I)}{\partial\theta'}\right\}^{-1} \frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{n} \nu_r(x_i, \hat{\theta}_I), \tag{2.15}$$

where $\nu_r(x_i,\theta) = \omega_{r,i}(x_i,\theta)/r$ with

$$\begin{aligned}\omega_{r,i}(x_i,\theta) = \sum_{l=1}^{L}\Bigg\{\frac{d_{li}}{[f^I_{r,l,i}(x_i,\theta)]^2}\Bigg[&\frac{\partial \ln f^I_{r,l,i}(x_i,\theta)}{\partial\theta}(S_{r,l}(x_i,\theta) - [f^I_{r,l,i}(x_i,\theta)]^2) \\ &- \left(C_{r,l}(x_i,\theta) - f^I_{r,l,i}(x_i,\theta)\frac{\partial f^I_{r,l,i}(x_i,\theta)}{\partial\theta}\right)\Bigg]\Bigg\},\end{aligned} \tag{2.16}$$

$$S_{r,l}(x_i,\theta) = \frac{1}{r}\sum_{j=1}^{r} h_l^2(v_i^{(j)}, x_i, \theta),$$

and

$$C_{r,l,i}(x_i,\theta) = \frac{1}{r}\sum_{j=1}^{r} h_l(v_i^{(j)}, x_i, \theta)\frac{\partial h_l(v_i^{(j)}, x_i, \theta)}{\partial\theta}.$$

The bias-corrected estimator $\hat\theta_A$ improves upon $\hat\theta_I$ by eliminating the leading bias term in $\hat\theta_I$ due to simulation. Another valuable property of $\hat\theta_A$ is that the asymptotic efficiency of $\hat\theta_A$ requires only that r goes to infinity faster than the rate $n^{1/4}$ instead of $n^{1/2}$ [Lee (1992b)].

The computation of the likelihood functions $\mathcal{L}_D$ and $\mathcal{L}_I$ involve double summations – one over the sample observations n and one over the number of simulated random variables m (or r) in the construction of simulated probabilities. When there are continuous regressors in the model, the choice probabilities will, in general, different across individuals. For the continuous regressors' case, the computation costs of these two simulated likelihood functions might be expected to be similar when m and r are about the same. The total number of simulated random variables for the dependent simulators case will be m and the total number of simulated random variables for the independent simulators case is nr, which will be n times large. However, the cost for generating the random variables will, in general, be a small component of the total cost for the derivation of the SMLE in an iteration algorithm. As the SML-IS can be more efficient relative to the SML-DS with the computation cost controlled for, the SML-IS is the preferred simulated likelihood approach. However, when the explanatory variables are all discrete or the individuals in the sample observations can be grouped together according to their common value of x and choice patterns, the SML-DS might have its value if the number of groups is not really large. This is the issue that we like to address in the subsequent sections.

3. Simulated Likelihood and Group Data

For the model with group data, it is quite natural to use dependent simulators since they can take into account the simplicity of sufficient statistics. Suppose that x takes on only K values, say, $c_1, \ldots, c_K$, where K is a finite integer. The individuals in the sample who have the same value of x and are observed to choose the same alternative can be put into a group. The sufficient statistics are $n_{l,k} = \sum_{\{i|x_i=c_k\}} d_{l,i}$, for $l = 1, \cdots, L$, and $k = 1, \ldots, K$. The log likelihood function in (2.1) becomes

$$\mathcal{L}_c(\theta) = \sum_{k=1}^{K}\sum_{l=1}^{L} n_{l,k}\ln P(l|c_k,\theta). \tag{3.1}$$

The simulated log likelihood function (2.5) becomes

$$\mathcal{L}_D(\theta) = \sum_{k=1}^{K}\sum_{l=1}^{L} n_{l,k}\ln f^D_{m,l}(c_k,\theta). \tag{3.2}$$

However, the simulated log likelihood function $\mathcal{L}_I(\theta)$ in (2.4) can not be simplified because the simulator $f_{r,l,i}(c_k,\theta)$ where $i \in \{i|x_i = c_k\}$ depends on i through the simulated random variables $v_i^{(1)}, \ldots, v_i^{(r)}$. The independently simulated probabilities can be numerically different for individuals in a group. On the other hand, the simulated probabilities with dependent simulators are the same for all the individuals in a group. Besides the above simulated likelihood approaches, it is possible to introduce a hybrid approach that allows the simulated probabilities to be the same for all individuals in a group but independent across different groups. Let $v_k^{(j)}$, $j = 1, \cdots, m$; $k = 1, \cdots, K$, be mK independent draws from $\gamma(.)$. Define the simulators

$$f^{D,k}_{m,l}(c_k,\theta) = \frac{1}{m}\sum_{j=1}^{m} h_l(v_k^{(j)}, c_k, \theta), \tag{3.3}$$

for alternative j and group k. The simulated log likelihood function with (3.3) is

$$\mathcal{L}_H(\theta) = \sum_{k=1}^{K}\sum_{l=1}^{L} n_{l,k}\ln f^{D,k}_{m,l}(c_k,\theta). \tag{3.4}$$

The simulated maximum likelihood estimate derived from (3.4) will be termed the SML estimator with hybrid simulators (SML-HS). As shown below, the simulated likelihood in (3.4) can be put into the framework of simulated likelihood with dependent simulators. Therefore, the asymptotic analysis in Lee (1992a) is directly applicable to derive the consistency and limiting distribution for the hybrid SML estimator. Let $I_{c_k}(x)$ be a dichotomous indicator such that $I_{c_k}(x) = 1$ if $x = c_k$; 0, otherwise. The $\mathcal{L}_H(\theta)$ can be rewritten as

$$\begin{aligned}\mathcal{L}_H(\theta) &= \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\left\{\sum_{k=1}^{K} I_{c_k}(x_i)\ln f^{D,k}_{m,l}(c_k,\theta)\right\} \\ &= \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\ln f^H_{m,l}(x_i,\theta),\end{aligned}$$

where $f^H_{m,l}(x_i,\theta) = \frac{1}{m}\sum_{j=1}^m h^H_l(v^{(j)}, x_i, \theta)$ with

$$h^H_l(v^{(j)}, x_i, \theta) = \sum_{k=1}^K I_{c_k}(x_i) h_l(v_k^{(j)}, c_k, \theta), \tag{3.5}$$

and $v^{(j)} = (v_1^{(j)}, \cdots, v_K^{(j)})$. So the simulated likelihood in (3.4) is essentially a simulated likelihood with dependent simulators. It uses simulators $f^H_{m,l}(x,\theta)$ instead of $f^D_{m,l}(x,\theta)$.

With group data, the SML-DS or SML-HS may have computational advantage over the SML-IS. The outer summations in (3.2) and (3.4) are taken over K instead of n. In contrary, the outer summation for $\mathcal{L}_I(\theta)$ in (2.4) is taking over n. For the computation of (3.2) and (2.4), the proper comparison is, however, not to compare K with n, but rather to compare the product KL with n. The reason is that for the SML-IS only response probability corresponding to the chosen alternative for an individual needs to be simulated. So the total number of simulated probabilities will be n for the independent simulators' case, no matter how large is L. On the other hand, for the SML-DS or SML-HS, members within a group identified by the value of x (an exogenous group) may have chosen different alternatives, so the total number of simulated probabilities can be KL. If n is not too large, it is possible that for each exogenous group some alternatives might not be chosen. So there may be less than KL simulated probabilities. However, one may expect that the total number of simulated probabilities for the dependency case will be close to KL for large n and moderate L. For each simulated probability, it takes m summations over the kernel function for the dependency case and its takes r summations for the independence case. To compare the two simulated likelihood approaches, one should consider both the statistical efficiency and computational costs. It is convenient to consider one conditional on the other. For our theoretical discussion, we presume that $KLm = nr$ might represent the same computation time cost for both the simulated likelihood approaches. In practice, this of course might not be exact. Conditional on $KLm = nr$, it may be sensible to discuss statistical relative efficiency as if computation cost has been controlled for.

Denote $G = KL$. The number of alternatives L in a choice model is predetermined and is a constant. The number of exogenous groups K may be finite but for some circumstances it may increase with sample size n. Let us consider first the case that K is a fixed finite constant. From section 2, the asymptotic distribution of $\hat\theta_D$ and its asymptotic efficiency depend on n/m. On the other hand, the asymptotic distribution of $\hat\theta_I$ depends on $n^{1/2}/r$. The $\hat\theta_A$ depends on $n^{1/4}/r$. As $Gm = nr$ and G is a fixed finite constant, $n/m = O(1/r)$. It follows that if r is fixed, $\hat\theta_I$ and $\hat\theta_A$ will be inconsistent but $\hat\theta_D$ is consistent and asymptotically normal.[1]

[1] The computation advantage of the MSM of McFadden (1989) is that r can be a fixed finite integer. The above analysis suggests that the SML-DS may have the same computation advantage as the MSM. However with r fixed, both the SML-DS and the method of simulated moments are consistent but not asymptotic efficient.

When r goes to infinity, n/m must go to zero. Therefore while $\hat\theta_I$ and $\hat\theta_A$ become consistent, the $\hat\theta_D$ must be asymptotic efficient. This provides very strong justification from the statistical point of view in favor of the SML-DS or SML-HS for the estimation of discrete choice models with group data.

It is of interest to consider the possibility that K increases (and therefore G) as sample size increases. This may be the case, as n increases the x may take more distinct values. Depending on the rate of increase for G, there are many different patterns. The time cost condition $Gm = nr$ implies that $m = nr/G$ and $n/m = G/r$. For fixed finite r, m goes to infinity when $G = o(n)$. That is, as long as the number of groups increases at a rate slower than n, the $\hat\theta_D$ will be consistent even $\hat\theta_I$ and $\hat\theta_A$ are not. $\hat\theta_D$ is asymptotic normal but inefficient under such a circumstance. When $\hat\theta_I$ and $\hat\theta_A$ become consistent (i.e., r goes to infinity), the $\hat\theta_D$ can become asymptotically efficient as long as G increases at a rate slower than r. This is so, because $\frac{n}{m}$ goes to zero when $G = o(r)$. When G increases at the rate of n, it would be similar to the continuous regressor case and the $\hat\theta_I$ and $\hat\theta_A$ will be favorable. When G increases at a rate not slower than r, the detailed comparison will depend on how fast r goes to infinity. Table 1 provides a list of results. The comparisons can be generalized to the bias-adjusted SMLE. The necessary modifications in Table 1 are simply to replace the power factor $\frac{1}{2}$ of n by $\frac{1}{4}$ in $\Delta = \frac{1}{2}(1-\epsilon)$, $\frac{1}{2}$, and $\frac{1}{2}+\delta$ for the relevant columns. In summary, when G is not large as compared with n, the $\hat\theta_D$ can be a favorable estimator.

REMARKS: McFadden (1989) introduced the MSM. For the MSM, there are no fundamental differences in the use of independent simulators or dependent simulators for group data. Let $d_i = (d_{1,i}, \cdots, d_{L,i})'$ be the vector of discrete choice indicators. For independent simulators, let $f^I_{r,i}(x,\theta) = (f^I_{r,1,i}(x,\theta), \cdots, f^I_{r,L,i}(x,\theta))'$ and for the dependent (hybrid) simulators, let $f^H_m(x,\theta) = (f^H_{m,1}(x,\theta), \cdots, f^H_{m,L}(x,\theta))'$. With instrumental variables matrix w_i, the simulated moments with independent simulators are

$$\sum_{i=1}^n w_i(d_i - f^I_{r,i}(x_i,\theta)), \tag{3.6}$$

and the simulated moments with dependent simulators will be

$$\sum_{i=1}^n w_i(d_i - f^H_m(x_i,\theta)). \tag{3.7}$$

With group data, it is natural that w_i will be the same for all i in an exogenous group. Let w_{c_k} denote the common value of w_i in the group with $x = c_k$. Let $n_k = (n_{1,k}, \cdots, n_{L,k})'$ and let $n_{(c_k)}$ be the number of i in the exogenous group with $x = c_k$. It follows that

$$\sum_{i=1}^{n} w_i(d_i - f^I_{r,i}(x_i, \theta)) = \sum_{k=1}^{K} w_{c_k}(n_k - \sum_{\{i|x_i=c_k\}} f^I_{r,i}(c_k, \theta)), \tag{3.6$'$}$$

and

$$\sum_{i=1}^{n} w_i(d_i - f^D_m(x_i, \theta)) = \sum_{k=1}^{K} w_{c_k}(n_k - n_{(c_k)} f^H_m(c_k, \theta)). \tag{3.7$'$}$$

From (2.2), we have $\frac{1}{n_{(c_k)}} \sum_{\{i|x_i=c_k\}} f^I_{r,l,i}(c_k, \theta) = \frac{1}{rn_{(c_k)}} \sum_{i=1}^{n_{(c_k)}} \sum_{j=1}^{r} h_l(v_i^{(j)}, c_k, \theta)$, which is a probability simulator that is common for all individuals in an exogenous group. The latter simulator and the hybrid simulator $f^H_{m,l}(c_k, \theta)$ differ only in the number of simulated random variables used in their constructions.[2] When $K = 1$, these simulators are the same.

[2] This difference does raise issues on sampling design that may be an interesting research topic for future investigation.

Table 1: Asymptotic Properties of SMLE

Criterion: $nr = Gm$
Keys: C– consistency, $\overline{C}$– inconsistent estimator, n^s– stochastic convergence at the rate $O_p(1/n^s)$;
N– asymptotic normal, B–asymptotic bias, D–degenerate limiting distribution;
E– asymptotic efficient, $\overline{E}$– not asymptotic efficient.

	r							
	$O(n^{\Delta})$							
	$\Delta = 0$		$\Delta = \epsilon/2$		$\Delta = 1/2$		$\Delta = 1/2 + \delta$	
K	$\hat{\theta}_I$	$\hat{\theta}_D$	$\hat{\theta}_I$	$\hat{\theta}_D$	$\hat{\theta}_I$	$\hat{\theta}_D$	$\hat{\theta}_I$	$\hat{\theta}_D$
$O(1)$	$\overline{C}$	$C, n^{1/2}$	$C, n^{\epsilon/2}$	$C, n^{1/2}$	$C, n^{1/2}$	$C, n^{1/2}$	$C, n^{1/2}$	$C, n^{1/2}$
	–	N	B, D	N	B, N	N	N	N
	–	$\overline{E}$	$\overline{E}$	E	$\overline{E}$	E	E	E
$O(n^{\Delta\eta})$	$\overline{C}$	$C, n^{1/2}$	$C, n^{\epsilon/2}$	$C, n^{1/2}$	$C, n^{1/2}$	$C, n^{1/2}$	$C, n^{1/2}$	$C, n^{1/2}$
	–	N	B, D	N	B, N	N	N	N
	–	$\overline{E}$	$\overline{E}$	E	$\overline{E}$	E	E	E
$O(n^{\Delta})$	$\overline{C}$	$C, n^{1/2}$	$C, n^{\epsilon/2}$	$C, n^{1/2}$	$C, n^{1/2}$	$C, n^{1/2}$	$C, n^{1/2}$	$C, n^{1/2}$
	–	N	B, D	N	B, N	N	N	N
	–	$\overline{E}$	$\overline{E}$	$\overline{E}$	$\overline{E}$	$\overline{E}$	E	$\overline{E}$
$O(n^{\Delta+\eta})$	$\overline{C}$	$C, n^{(1-\eta)/2}$	$C, n^{\epsilon/2}$	$C, n^{(1-\eta)/2}$	$C, n^{1/2}$	$C, n^{(1-\eta)/2}$	$C, n^{1/2}$	$C, n^{(1-\eta)/2}$
	–	N	B, D	N	B, N	N	N	N
	–	$\overline{E}$	$\overline{E}$	$\overline{E}$	$\overline{E}$	$\overline{E}$	E	$\overline{E}$

Notes: $\Delta > 0$, $\delta > 0$, $0 < \eta < 1$, and $0 < \epsilon < 1$ in the above Table.

4. Some Monte Carlo Results

In this section, we report some Monte Carlo experiment results for the simulated likelihood estimation methods in terms of statistical efficiency and computing time cost. Two discrete choice panel data models are investigated.

The first model (Model 1) is a dynamic discrete choice model with individual random component and serial disturbance:

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

and

$$\epsilon_{it} = \rho \epsilon_{i,t-1} + (1-\rho^2)^{1/2} w_{it}, \tag{4.2}$$

where y_{it}^{*} is the latent dependent variable which sign determines the choice d_{it} as $d_{it} = 1$ if $y_{it}^{*} > 0$ and $d_{it} = 0$, otherwise; ξ_i is an individual random $N(0,1)$; and ϵ_{it} is a stationary $AR(1)$ disturbance with w_{it} being a $N(0,1)$ random variable. The process is assumed to start at $t = 1$ with $d_{i,0} = 0$. The disturbances ξ_i and ϵ_{it} are independent. The ϵ_{it} in (4.2) has a unit variance, which reflects the normalization used for this model. The value of the true parameter vector $(\beta, \lambda, \sigma, \rho)$ is $(1, 0.2, \sqrt{0.5}, 0.4)$.

Since the Geweke-Hajivassiliou-Keane recursive simulator is a powerful simulator for discrete choice probabilities, we use a similar simulator for our study. Their recursive simulation procedure can be easily adopted to our model (4.1)-(4.2). Define a sign indicator D_t such that $D_t = 1$ if $d_t = 1$ and $D_t = -1$ if $d_t = 0$. Let Φ denote the standard normal distribution function. The recursive simulators are generated by the following steps:

1.) Generate ϵ_0 from $N(0,1)$, generate independently ξ from $N(0,1)$, and generate $u_1, \ldots, u_T$ independent uniform $[0,1]$ random variables.

2.) Compute w_t and ϵ_t recursively from $t = 1$ to $t = T$ as follows:

2.1) Compute $w_t \equiv \Phi^{-1}\left[u_t \Phi\left(D_t(\beta + \lambda d_{t-1} + \sigma\xi + \rho\epsilon_{t-1})/\sqrt{1-\rho^2}\right)\right]$.

2.2) Update the AR disturbance process as $\epsilon_t = \rho\epsilon_{t-1} + (1-\rho^2)^{\frac{1}{2}} w_t$.

The simulated log likelihood function with independent simulators is

$$\mathcal{L}_I = \sum_{i=1}^{n} \ln\left\{\frac{1}{r}\sum_{j=1}^{r}\prod_{t=1}^{T}\Phi\left[D_{it}\left((\beta + \lambda d_{i,t-1} + \sigma\xi_i^{(j)} + \rho\epsilon_{i,t-1}^{(j)})/\sqrt{1-\rho^2}\right)\right]\right\}. \tag{4.3}$$

The simulated log likelihood function with dependent simulators is

$$\mathcal{L}_D = \sum_{(D_1,\ldots,D_T)} n_{D_1\cdots D_T} \ln\left\{\frac{1}{m}\sum_{j=1}^{m}\prod_{t=1}^{T}\Phi\left[D_t\left((\beta + \lambda d_{t-1} + \sigma\xi^{(j)} + \rho\epsilon_{t-1}^{(j)})/\sqrt{1-\rho^2}\right)\right]\right\}, \tag{4.4}$$

where $n_{D_1\cdots D_T}$ is the number of sample observations that have $(D_1, \cdots, D_T)$ as chosen choice path over the sample periods.

The Monte Carlo results for the SML-DS based on (4.4) are reported in Table 2. The results for the SML-IS based on (4.3) are reported in Table 3. The number of periods is set to four. So the model corresponds to a discrete choice model with 16 alternatives. Samples with various sizes n and numbers of simulated variables m or r are tried. For each design, the means and standard errors of estimates based on 600 replications are reported. In addition, the average CPU time (in seconds) and average number of iterations (for successful convergence of a maximization algorithm) for a replication are reported.[3] The latter statistics provide information about computing time cost.

From Table 2, we see that the SML-DS provides accurate estimates of β and σ for all the sample sizes from 100 to 400 unless m is quite small. With $m = 60$ or more, the biases for β and σ are small. For the estimates of the dynamic effect λ and serial correlation coefficient ρ, there are biases. The bias for λ is upward. The bias of ρ is downward. Such biases are not due to errors in the use of simulated probabilities. They are finite sample biases. As sample sizes increase from 100 to 400, these biases for λ and ρ are reduced. As n and/or m increase, the standard errors of the estimates are decreased as expected. The time cost for the computation of the SML-DS is approximately linear in m. With m fixed, the computing cost has increased slightly as n increases from 100 to 400. The computing cost for the dependent simulator approach depends on the number of groups but not on sample sizes. For small n, some discrete choice patterns (alternatives) may not be observed in the sample. When the alternatives are not chosen, they do not contribute to the likelihood function and these choice probabilities need not be computed. So for small n, the effective groups can be less than 16. When n becomes larger, it is likely each of all the 16 alternatives may be chosen by some individuals in the sample and their probabilities need be computed.

The SML-IS estimates are reported in Table 3. There are biases in the SML-IS estimates. Except the estimates of ρ, these biases are reduced as r increases. The biases for all the estimates tend to be smaller when both n and r increase. For the estimate of β, if only n increases but r does not, the biases do not necessarily become smaller. The latter phenomenon is not surprising theoretically. The bias adjusted SML estimates have effectively smaller biases than the bias unadjusted SML except the estimates of ρ. Their variances are,

[3] The maximization algorithm is a conjugate gradient method described in Press et at. [1986], Chapter 10. The Fortran programs ran in an IBM RS/6000 Model 580 workstation. Comparing it with the speed of the IBM RS/6000 Model 320H workstation, the 320H machine is slow by a time factor of 2.6.

however, slightly larger. For the estimates of ρ, as r increases but n fixed, their biases may even be worsened. The bias behavior of the estimates of ρ can better be understood when they are compared with the SML-DS estimates in Table 2. There are both finite sample bias and simulation bias in these estimates. For the SML-IS, the biases due to simulation happen to be biased upward but the finite sample biases are biased downward. So under certain combinations, the combined biases can be smaller. As r becomes large, the biases due to simulation are reduced and the remaining biases are mainly due to finite sample biases. The computation costs of the SML-IS are also linear in n or r. The computation cost depends on the sample size because the number of simulated probabilities is propositional to n. Comparing the time costs from Table 2 and Table 3, the SML-IS are much more expensive. With time cost being equal, a moderate r corresponds to a much larger m. As an example with $n = 400$, the time cost for $r = 20$ for the SML-IS can be used for the computation of the SML-DS with m greater than 500. The formula $m = nr/G$ has slightly overestimated the time cost for the SML-DS. With equal time cost, the actual m can be larger than nr/G. The extra time saving for the SML-DS is due to the slightly smaller number of iterations for successful convergence of the numerical algorithm. As large m may be used with moderate time cost, the biases due to simulation in the SML-DS are smaller than the corresponding biases in the SML-IS. The variances of the SML-IS for β are slightly smaller than the variances of the corresponding SML-DS estimates. On the other hand, the variances of the SML-DS for σ and ρ are smaller except small m. As the variances of the SML-DS are on average compatible with the variances of the SML-IS estimates, the performance of the SML-DS seems better.

One might expect that the advantage of the dependent simulator approach would disappear or its performance might be worsened when there are many more groups in a sample data. To investigate this possibility, we simulate another model (Model 2):

$$y_{it}^{*} = \beta x_i + \lambda d_{i,t-1} + \sigma \xi_i + \epsilon_{it}, \tag{4.5}$$

where ϵ_{it} is a $AR(1)$ process as in (4.2). The x_i takes with equal probabilities on ten different values from -0.8 to 1.0 with increment 0.2. With 16 alternatives in choices, the (maximum) number of groups is 160. Tables 4A-4C report, respectively, the results of the SML-IS with or without bias adjustment, the SML-DS estimates, and the SML-HS estimates. The results reported in Table 4A and Table 4B are, respectively, for the sample sizes $n = 100$ and $n = 200$. Since the number of groups is no larger than 160, the dependent simulators approach and the hybrid approach have some time saving advantage. The m can be 1.8 to 2.5 times larger than r for a similar time cost. Overall, the SML-DS provides better estimates of β and σ in

terms of smaller biases, and even smaller variances for $n = 200$. However, both the SML-IS approaches with or without bias adjustment provide better estimates for λ and ρ. It is interested to see that the SML-HS is a mixture of the independent simulator and the dependent simulator approaches. There are no obvious winners here for $n = 100$ or 200. Table 4C provides results for $n = 400$. As n becomes larger, the m can now be more than 3.5 times larger than r with a similar time cost. The SML approach with dependent simulators provides better estimates for β and σ. The λ and ρ are still better estimated by the SML with independent simulators and their bias-adjusted estimates. Their differences are, however, smaller than the cases with smaller sample sizes. The hybrid SML estimator becomes more interesting. It provides better estimates of λ than the SML-IS and SML-DS approaches. On average, the hybrid simulator seems the preferred one for this case with $n = 400$.

Table 2
Model 1: SML—Dependent Simulators (SML-DS)
True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5} \doteq 0.7071$, and $\rho = 0.4$

n	m	β	λ	σ	ρ	time	#iteration
100	30	1.0122 (.3275)	.2730 (.2562)	.7508 (.4218)	.3026 (.2522)	2.4998	8.27
100	60	0.9894 (.2624)	.2625 (.2569)	.6912 (.3396)	.3180 (.2472)	4.9221	8.12
100	120	0.9912 (.2424)	.2568 (.2470)	.6914 (.3281)	.3275 (.2435)	9.8890	8.18
100	180	0.9898 (.2420)	.2560 (.2495)	.6877 (.3185)	.3266 (.2424)	14.9338	8.14
100	375	0.9881 (.2366)	.2566 (.2455)	.6829 (.3096)	.3267 (.2419)	30.9701	8.14
200	30	1.0149 (.2561)	.2647 (.1917)	.7737 (.3232)	.3190 (.2006)	2.7735	7.78
200	60	0.9947 (.2076)	.2500 (.1866)	.7120 (.2749)	.3413 (.1957)	5.6663	7.91
200	120	0.9963 (.1739)	.2456 (.1824)	.7123 (.2449)	.3469 (.1893)	11.2064	7.84
200	180	0.9892 (.1647)	.2424 (.1801)	.6974 (.2443)	.3520 (.1907)	16.9629	7.91
200	375	0.9872 (.1547)	.2408 (.1788)	.6916 (.2277)	.3556 (.1877)	35.7265	7.98
200	750	0.9872 (.1515)	.2419 (.1804)	.6915 (.2212)	.3535 (.1899)	70.9905	7.91
400	30	1.0143 (.2151)	.2571 (.1478)	.7802 (.2561)	.3270 (.1571)	2.9658	7.73
400	60	0.9955 (.1604)	.2416 (.1353)	.7243 (.2068)	.3486 (.1434)	5.9923	7.79
400	120	0.9933 (.1352)	.2345 (.1293)	.7159 (.1901)	.3621 (.1392)	12.0979	7.86
400	180	0.9908 (.1243)	.2309 (.1304)	.7054 (.1754)	.3662 (.1366)	18.2331	7.89
400	500	0.9866 (.1151)	.2272 (.1247)	.6941 (.1644)	.3715 (.1332)	50.5710	7.86
400	750	0.9877 (.1097)	.2287 (.1245)	.6944 (.1634)	.3693 (.1356)	75.6673	7.80

Table 3
Model 1: SML—Independent Simulators (SML-IS)
True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5} \doteq 0.7071$, and $\rho = 0.4$

n	r	β	λ	σ	ρ	time	#iteration
		Bias unadjusted SML-IS					
100	10	0.8600 (.2064)	.3002 (.2451)	.4418 (.3410)	.3743 (.2161)	6.5768	8.93
100	20	0.9180 (.2214)	.2689 (.2489)	.5271 (.3823)	.3657 (.2494)	13.3913	8.93
100	30	0.9430 (.2294)	.2610 (.2485)	.5764 (.3667)	.3561 (.2472)	19.8165	8.79
100	50	0.9536 (.2253)	.2622 (.2481)	.6028 (.3540)	.3442 (.2527)	32.4477	8.51
100	75	0.9678 (.2310)	.2592 (.2451)	.6309 (.3401)	.3370 (.2497)	48.9048	8.46
100	100	0.9722 (.2285)	.2593 (.2457)	.6431 (.3351)	.3344 (.2462)	63.7048	8.28
100	200	0.9798 (.2311)	.2563 (.2488)	.6593 (.3123)	.3321 (.2459)	127.4555	—
200	10	0.8385 (.1238)	.2863 (.1693)	.3870 (.2607)	.4168 (.1609)	13.4586	8.89
200	15	0.8820 (.1343)	.2646 (.1752)	.4764 (.2815)	.4073 (.1780)	20.2135	8.76
200	30	0.9253 (.1403)	.2477 (.1805)	.5538 (.2893)	.3914 (.2048)	40.2684	8.63
200	50	0.9492 (.1448)	.2442 (.1795)	.6086 (.2746)	.3777 (.1977)	66.3789	8.44
200	75	0.9614 (.1488)	.2439 (.1820)	.6344 (.2653)	.3696 (.1975)	98.5324	8.25
400	10	0.8239 (.0815)	.2786 (.1210)	.3515 (.2063)	.4377 (.1203)	28.3107	9.02
400	20	0.8963 (.1004)	.2506 (.1250)	.5173 (.2263)	.4121 (.1434)	54.1355	8.38
400	30	0.9249 (.1056)	.2373 (.1272)	.5726 (.2117)	.4068 (.1449)	88.8999	8.26
400	50	0.9465 (.1064)	.2336 (.1277)	.6123 (.2093)	.3934 (.1521)	134.3421	8.22
		Bias Adjusted SML-IS					
100	10	0.9313 (.2335)	.2348 (.2628)	.5458 (.3900)	.3907 (.2629)	6.6214	—
100	20	0.9720 (.2448)	.2317 (.2634)	.6018 (.4100)	.3638 (.2889)	13.4800	—
100	30	0.9860 (.2463)	.2359 (.2594)	.6365 (.3792)	.3492 (.2764)	19.9489	—
100	50	0.9829 (.2361)	.2483 (.2559)	.6459 (.3577)	.3350 (.2731)	32.6678	—
100	75	0.9892 (.2387)	.2503 (.2506)	.6632 (.3392)	.3286 (.2639)	49.2385	—
100	100	0.9888 (.2347)	.2529 (.2500)	.6686 (.3331)	.3274 (.2568)	63.7048	—
100	200	0.9887 (.2340)	.2533 (.2510)	.6736 (.3094)	.3277 (.2511)	128.3383	—
200	10	0.9056 (.1404)	.2173 (.1817)	.4851 (.3056)	.4428 (.1973)	13.5470	—
200	15	0.9431 (.1507)	.2154 (.1866)	.5656 (.3125)	.4159 (.2110)	20.3467	—
200	30	0.9675 (.1509)	.2222 (.1894)	.6145 (.3023)	.3868 (.2286)	40.5315	—
200	50	0.9792 (.1519)	.2300 (.1856)	.6523 (.2794)	.3698 (.2135)	66.8177	—
200	75	0.9831 (.1537)	.2351 (.1864)	.6666 (.2654)	.3622 (.2080)	99.1950	—
400	10	0.8874 (.0919)	.2083 (.1304)	.4468 (.2463)	.4695 (.1483)	28.4887	—
400	20	0.9511 (.1120)	.2127 (.1340)	.6006 (.2432)	.4138 (.1668)	54.4894	—
400	30	0.9694 (.1144)	.2122 (.1346)	.6397 (.2201)	.4017 (.1627)	89.4788	—
400	50	0.9775 (.1119)	.2196 (.1325)	.6587 (.2120)	.3856 (.1634)	135.2229	—

Table 4A
Model 2: SML
True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5} \doteq 0.7071$, and $\rho = 0.4$

		β	λ	σ	ρ	time	#iteration
n	r	**SML-IS**					
100	10	0.8942 (.2180)	.2624 (.2031)	.4390 (.3201)	.3945 (.2009)	5.7118	8.56
100	30	0.9523 (.2372)	.2368 (.2124)	.5682 (.3323)	.3757 (.2262)	16.1881	7.93
100	60	0.9681 (.2414)	.2344 (.2129)	.6017 (.3142)	.3660 (.2291)	32.0720	7.71
100	100	0.9760 (.2432)	.2347 (.2134)	.6132 (.3169)	.3627 (.2253)	52.3226	7.66
n	r	**Bias adjusted SML-IS**					
100	10	0.9417 (.2422)	.2144 (.2184)	.5330 (.3654)	.4005 (.2450)	5.7118	—
100	30	0.9816 (.2504)	.2196 (.2208)	.6182 (.3405)	.3692 (.2479)	16.3189	—
100	60	0.9855 (.2483)	.2262 (.2179)	.6310 (.3144)	.3603 (.2407)	32.3391	—
100	100	0.9872 (.2475)	.2299 (.2163)	.6323 (.3158)	.3587 (.2322)	52.7601	—
n	m	**SML-DS**					
100	10	0.9492 (.2442)	.3029 (.2934)	.6555 (.4340)	.3174 (.2653)	2.9911	7.49
100	30	0.9871 (.2537)	.2620 (.2430)	.6628 (.3345)	.3380 (.2299)	8.4911	7.27
100	60	0.9856 (.2480)	.2503 (.2306)	.6387 (.3169)	.3481 (.2327)	17.1010	7.42
100	100	0.9968 (.2515)	.2426 (.2193)	.6574 (.3098)	.3537 (.2193)	28.3476	7.34
100	170	0.9960 (.2468)	.2344 (.2062)	.6569 (.2940)	.3576 (.2191)	47.7947	7.29
n	m	**SML-HS**					
100	10	0.9147 (.2751)	.2646 (.2141)	.4747 (.3620)	.3865 (.2158)	3.3444	8.60
100	30	0.9726 (.2604)	.2381 (.2074)	.5856 (.3448)	.3711 (.2267)	9.2843	7.88
100	60	0.9856 (.2574)	.2268 (.2058)	.6114 (.3230)	.3770 (.2180)	17.9849	7.79
100	100	0.9894 (.2583)	.2332 (.2072)	.6388 (.3072)	.3623 (.2191)	29.0862	7.51
100	170	0.9907 (.2480)	.2247 (.2060)	.6296 (.2989)	.3759 (.2205)	50.1789	7.58

Table 4B
Model 2: SML
True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5} \doteq 0.7071$, and $\rho = 0.4$

		β	λ	σ	ρ	time	#iteration
n	r	**SML-IS**					
200	15	0.9024 (.1613)	.2383 (.1485)	.4716 (.2952)	.4178 (.1767)	17.3225	8.35
200	30	0.9451 (.1762)	.2316 (.1482)	.5755 (.2743)	.3949 (.1766)	32.5617	7.75
200	60	0.9645 (.1803)	.2246 (.1482)	.6092 (.2616)	.3902 (.1778)	65.7558	7.58
n	r	**Bias adjusted SML-IS**					
200	15	0.9425 (.1770)	.2029 (.1594)	.5486 (.3246)	.4196 (.2091)	17.4534	—
200	30	0.9752 (.1861)	.2139 (.1552)	.6279 (.2829)	.3889 (.1944)	32.8239	—
200	60	0.9828 (.1860)	.2164 (.1522)	.6402 (.2625)	.3845 (.1870)	66.2960	—
n	m	**SML-DS**					
200	20	0.9666 (.1731)	.2736 (.2265)	.6748 (.3128)	.3331 (.2134)	8.6362	7.04
200	40	0.9745 (.1739)	.2532 (.1980)	.6610 (.2753)	.3513 (.2012)	17.2600	7.15
200	80	0.9823 (.1749)	.2348 (.1849)	.6618 (.2540)	.3665 (.1894)	34.5802	7.19
200	150	0.9914 (.1719)	.2353 (.1686)	.6920 (.2355)	.3598 (.1791)	64.2499	7.06
n	m	**SML-HS**					
200	20	0.9413 (.1988)	.2417 (.1719)	.5652 (.2999)	.3867 (.1919)	9.3663	7.74
200	40	0.9582 (.1789)	.2250 (.1604)	.6072 (.2814)	.3878 (.1855)	18.4547	7.69
200	80	0.9752 (.1758)	.2276 (.1550)	.6544 (.2547)	.3724 (.1774)	35.3625	7.37
200	150	0.9893 (.1793)	.2261 (.1568)	.6716 (.2409)	.3683 (.1772)	65.2645	7.22

Table 4C
Model 2: SML
True parameters: $\beta = 1$, $\lambda = 0.2$, $\sigma = \sqrt{0.5} \doteq 0.7071$, and $\rho = 0.4$

		β	λ	σ	ρ	time	#iteration
n	r			SML-IS			
400	20	0.9252 (.1219)	.2304 (.1081)	.5388 (.2300)	.4172 (.1379)	45.1274	7.81
400	30	0.9494 (.1204)	.2277 (.1100)	.5974 (.2123)	.4014 (.1393)	71.9591	7.44
400	60	0.9663 (.1222)	.2202 (.1047)	.6322 (.2013)	.3972 (.1341)	134.6610	7.42
n	r			Bias adjusted SML-IS			
400	20	0.9646 (.1335)	.2043 (.1164)	.6124 (.2464)	.4126 (.1592)	45.4756	—
400	30	0.9815 (.1280)	.2109 (.1160)	.6534 (.2191)	.3938 (.1535)	72.5335	—
400	60	0.9854 (.1262)	.2122 (.1077)	.6648 (.2009)	.3910 (.1407)	135.7403	—
n	m			SML-DS			
400	50	0.9882 (.1278)	.2420 (.1594)	.6884 (.2083)	.3599 (.1557)	29.7029	6.93
400	100	0.9963 (.1240)	.2346 (.1340)	.6907 (.1832)	.3690 (.1395)	59.5965	6.96
400	150	0.9981 (.1218)	.2297 (.1275)	.7009 (.1707)	.3713 (.1315)	88.1383	6.92
400	200	1.0001 (.1231)	.2297 (.1242)	.6965 (.1655)	.3705 (.1314)	118.6600	6.94
n	m			SML-HS			
400	50	0.9711 (.1360)	.2171 (.1104)	.6336 (.2144)	.3995 (.1348)	31.7462	7.44
400	100	0.9857 (.1289)	.2189 (.1097)	.6644 (.1863)	.3883 (.1283)	61.6710	7.28
400	150	0.9952 (.1297)	.2185 (.1052)	.6765 (.1772)	.3850 (.1266)	91.0383	7.17
400	200	0.9961 (.1260)	.2176 (.1071)	.6822 (.1705)	.3838 (.1244)	121.2005	7.15

5. Conclusion

This article has considered simulation methods for the estimation of discrete choice models with group data. Sufficient statistics in terms of observed frequencies are available in group data. The simulated maximum likelihood estimation method can take into account such a characteristic when the simulated probabilities are allowed to be correlated across individual decision units in the sample. The maximum simulated likelihood approach with dependent simulators can be compared with the approach with independent simulators in terms of computation time cost and statistical efficiency. When the number of aggregated groups is not too large, the simulated maximum likelihood method with dependent simulators may be attractive in terms of computational cost saving and statistical efficiency. Conditional on time costs being equal, the maximum simulated likelihood approach with dependent simulators can be statistically efficient relative to the approach with independent simulators. As the computation cost of the simulated likelihood approach with dependent simulators does not, in general, increase in proportion to sample size, the dependent simulator approach is especially valuable for the estimation of models with large sample sizes. Monte Carlo results are provided to show our claims.

This article has addressed only simulated likelihood methods. In addition to the simulated maximum likelihood approach, our discussions are also relevant to other M-estimation methods such as the Pearson or Neyman minimum chi-square estimation methods, or quasi-likelihood methods [Laroque and Salanie (1990), Gourieroux and Monfort (1993)]. For the simulated method of moments of McFadden, the comparisons with dependent or independent simulators are irrelevant as they are essentially indistinguishable. The simulated likelihood approaches are interesting because their implementations are straightforward. The computation cost for the simulated maximum likelihood estimator with dependent simulators can be as low as the method of simulated moments when the number of aggregate groups is not large. Besides estimation, for the approaches with dependent simulators, one may also construct goodness-of-fit statistics from the simulated likelihood function. The statistical testing issues, however, are subjects for future research.

References

1. 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.

2. 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.

3. Gourieroux C. and A. Monfort (1993), "Simulation-based Inference: A Survey with Special Reference to Panel Data Models." *Journal of Econometrics* 59: 5-33.

4. 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.

5. Keane, M.P. (1990), "A Computationally Practical Simulation Estimator for Panel Data, with Applications to Estimating Temporal Dependence in Employment and Wages." Discussion Paper, University of Minnesota, Minneapolis, MN.

6. Laroque, G. and B. Salanie (1989), "Estimation of Multi-market Fix-Price Models: An Application of Pseudo Maximum Likelihood Methods." *Econometrica* 57: 831-860.

7. Lee, A.J. (1990), *U-Statistics: Theory and Practice.* New York: Marcel Dekker.

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

9. Lee, L.F. (1992b), "Asymptotic Bias in Maximum Simulated Likelihood Estimation of Discrete Choice Models." CREST and Department of Economics Working Paper Series, Number 93-03.

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

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

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

13. Pakes, A. (1986), "Some Estimates of the Value of Holding European Patent Stocks." *Econometrica* 54: 755-785.

14. Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators." *Econometrica* 57: 1027-1057.

15. Stern, S. (1992), "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models." *Econometrica* 60: 943-952.