Open access • Journal Article • DOI:10.1017/S0266466600013207

On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models — Source link

Lung-Fei Lee

Published on: 01 Dec 1992 - Econometric Theory (Center for Economic Research, Department of Economics, University of Minnesota)

Topics: Method of moments (statistics), Generalized method of moments, Method of simulated moments, Asymptotic theory (statistics) and Estimation theory

ON EFFICIENCY OF METHODS OF SIMULATED MOMENTS AND MAXIMUM SIMULATED LIKELIHOOD ESTIMATION OF DISCRETE RESPONSE MODELS

by

Lung-Fei Lee

Discussion Paper No. 260, September 1990

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, MN 55455

On Efficiency of Methods of Simulated Moments

and Maximum Simulated Likelihood Estimation of Discrete Response Models

by

Lung-Fei Lee

Department of Economics, University of Minnesota

Minneapolis, Minnesota 55455

First draft, dated June 1990

Revised draft, September 1990

ABSTRACT

This article has considered methods of simulated moments for estimation of discrete response models. We have introduced a modified method of simulated moments of McFadden [1989]. Using the same number of Monte Carlo draws as in McFadden's method of simulated moments, our estimator is asymptotically efficient relative to McFadden's estimator. In addition to the method of simulated moments, we have considered also maximum simulated likelihood estimation methods. The estimators are shown to be consistent and asymptotically normal without excessive number of Monte Carlo draws.

JEL classification number: 211

Keywords:

Simulated moments, method of moments, discrete choice models, simulated likelihood, parameter estimation, V-statistics, relative efficiency, asymptotic efficiency.

CORRESPONDING ADDRESS:

Lung-Fei Lee, 1035 Management and Economics, Department of Economics, University of Minnesota, 271 19th. Ave. S., Minneapolis, Mn 55455

On Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models

by

Lung-Fei Lee*

1. Introduction

McFadden [1989] has proposed a method of simulated moments for the estimation of discrete response models. The method of simulated moments uses Monte Carlo draws from some specific distributions to construct (asymptotically) unbiased estimates of conditional choice probabilities. The method of simulated moments avoids computation of complicated multivariate probability functions and is computationally tractable even for models with many choice alternatives. McFadden's method is different from the method of Monte Carlo integration in Lerman and Manski [1981]. Lerman and Manski [1981] has emphasized on the accuracy of estimating conditional response probabilities for each observation. Lerman and Manski [1981] has found out that a large number of Monte Carlo draws is required to provide accurate approximation to multivariate probit probabilities. McFadden's simulated method of moments has emphasized on the estimation of unknown parameters. The number of Monte Carlo draws to construct unbiased conditional choice probabilities for each individual observation does not need to be large. For each sample observation, only a fixed number of independent Monte Carlo draws is needed. McFadden [1989] has proposed several approaches to construct the (asymptotically) unbiased estimates of conditional choice probabilities. His estimators are all consistent and asymptotically normal. Since the number of draws for each observation is fixed and independent with sample size, the estimators derived

* I appreciate having financial support from NSF under grant no. SES-8809939 and SES-9010516 and computing grant support from the Minnesota Supercomputer Institute for my research. I would like to thank Hide Ichimuara, Michael Keane, Steven Stern and Scott Thompson for their useful comments. Amy J.L. Lee has provided valuable assistance on the simulation study in this article.

from the method of simulated moments are statistically less efficient than the classical method of moment estimator. Empirical applications of the method of simulated moments can be found, for examples, in Pakes [1986], Haijivassiliou and McFadden [1987] and Keane [1989].

In this article, we will consider the method of simulated moments based on smooth moment simulators. We point out that, given the same Monte Carlo draws, it is possible to improve asymptotic efficiency of estimators by slightly modifying the method of simulated moments of McFadden [1989]. Monte Carlo draws for each observation are exchangeable across observation and hence all of them can be used to construct unbiased estimates of conditional choice probabilities for each observation. These estimates are computationally more expensive as they involve many more terms for summation. But these simulated moments have smaller errors than the ones based on a smaller number of Monte Carlo draws. Thus if we used these unbiased estimates in the method of simulated moments, there might be statistical efficiency gain. However, potential efficiency gain might not be realized as the simulated moments are now dependent across observations. In this article, we will investigate the efficiency issue. Based on V-statistics theory, we show that the modified estimator is indeed asymptotically more efficient than the corresponding estimator in McFadden [1989] inspite of dependency. In addition to the method of simulated moments, we will consider also a simulated likelihood estimation method. This method provides an alternative estimation method without restricting ourself to the estimation of only moment equations but yet it will use the same number of Monte Carlo draws as in the method of simulated moments. Our modified method of simulated moments and simulated likelihood method can also be regarded as generalization of Pakes' simulation methods in Pakes [1986] to disaggregated data with smooth simulators [1].

[1] For aggregated data, since sample frequencies are sufficient statistics, our modified method of moment, McFadden's method and Pakes' method of moment will all coincide. In a time series content, simulated method of moment estimation similar to Pakes [1986] but with smooth simulated moments can be found in Lee and Ingram [1989].

This article is organized as follows. In Section 2, we present the modified method of simulated moments. Asymptotic properties and relative efficiency of the modified moment estimator will be investigated in Section 3. Section 4 presents a simulated likelihood estimation method. Its asymptotic properties are investigated. Section 5 will investigate an adding up property for moment simulators. This property has an interesting implication on the asymptotic distribution of the simulated likelihood estimation. In Section 6, we clarify the efficiency issue of dependently simulated moments versus independently simulated moments. Some Monte Carlo results on the finite sample performance of the proposed estimators are reported in Section 7. Conclusions are drawn in Section 8. Useful propositons for our anlysis are collected in an appendix.

2. Methods of Simulated Moments for Estimation of Discrete Response Models

McFadden [1989] has proposed a method of simulated moments for estimation of discrete response models. A typical discrete response model is a model with several choice alternatives. Let $C = \{1, \cdots, L\}$ be a set of mutually exclusive and exhaustive alternatives. For each alternative $l \in C$, the associated value is

$$u_l = x_l \alpha \tag{2.1}$$

where α is a vector of individual weights distributed randomly in the population and x_l is a vector of measured attributes of alterntive l. Response j is observed if $u_j \geq u_l$ for all $l \in C$. Assume that the distribution of α is known except for a vector of parameters θ of dimension k. Let x denote the vector consisting of all distinct explanatory variables in $x_1, \cdots, x_L$. Define

$$v_l = (u_1 - u_l, \cdots, u_{l-1} - u_l, u_{l+1} - u_l, \cdots, u_L - u_l)$$

and let $g_l(v|\theta, x)$ denote the density function of v_l conditional on x. The response probability $P(l|\theta, x)$ for alternative l is

$$P(l|\theta, x) = \int_{v \leq 0} g_l(v|\theta, x) dv. \tag{2.2}$$

Let d_{li} denote a response indicator for individual i, equal to one for the observed response, zero otherwise. The conventional method of moments estimation is

$$\min_{\theta} (D - P(\theta))' W' W (D - P(\theta)) \tag{2.3}$$

where $D - P(\theta)$ is the nL vector of residuals $d_{li} - P(l|\theta, x_i)$ stacked by observation and by alternative within observation and W is a matrix of instruments of rank $K \geq k$. The method of simulated moments of McFadden [1989] avoids the computation of $P(\theta)$ by replacing it with a simulator $f(\theta)$ that is (asymptotically) conditionally unbiased, given W and d, independent across observations and "well behaved" in θ. McFadden's method of simulated moments is

$$\min_{\theta} (D - f(\theta))' W' W (D - f(\theta)). \tag{2.4}$$

Both the articles of McFadden [1989] and Pakes and Pollard [1989] provide powerful asymptotic techniques which establish consistency and asymptotic distribution for a broad class of simulators which include both smooth and nonsmooth simulators. Consider, for example, the smooth simulators in McFadden [1989] based on importance simulation technique [2]. Let $\gamma(v)$ be a density chosen for the simulation that has the negative orthant as its support. Let

$$h_l(v,x,\theta) = \begin{cases} \frac{g_l(v|\theta,x)}{\gamma(v)}, & v < 0, \\ 0 \quad , & otherwise. \end{cases} \tag{2.5}$$

Then, (2.2) can be rewritten as

$$P(l|\theta,x) = \int h_l(v,x,\theta)\gamma(v)dv. \tag{2.6}$$

The density $\gamma(v)$ is usually chosen so that h_l is dominated by a function H independent of θ with $\int H\gamma dv$ finite. Averaging $h_l(v,x,\theta)$ for an observation, using one or more Monte Carlo draws from $\gamma(v)$ that are taken independently across observations and fixed for different θ gives a smooth unbiased estimator of $P(l|\theta,x)$. Suppose there are r Monte Carlo draws from $\gamma(v)$ for an observation. Let $v_j^{(i)}$, $j = 1,\cdots,r$ be the draws for observation i. Define

$$f_l(\theta,x_i) = \frac{1}{r}\sum_{j=1}^{r} h_l(v_j^{(i)},x_i,\theta). \tag{2.7}$$

Conditional on x_i,

$$\begin{aligned} E(f_l(\theta,x_i)|x_i) &= \int h_l(v,x_i,\theta)\gamma(v)dv \\ &= P(l|\theta,x_i) \end{aligned}$$

and hence $f_l(\theta,x_i)$ is an conditionally unbiased simulator. McFadden's estimator $\hat{\theta}_M$ of θ is derived from

$$\min_{\theta}(D - f(\theta))'W'W(D - f(\theta)) \tag{2.8}$$

[2] Importance simulation technique has useful applications in Bayesian inference in econometrics. For such applications, see, for example, Geweke [1989].

where $f(\theta) = (f'(\theta, x_1), \cdots, f'(\theta, x_n))'$ and $f'(\theta, x_i) = (f_1(\theta, x_i), \cdots, f_L(\theta, x_i))$ stacked with the simulated moments in (2.7).

The total number of Monte Carlo draws from $\gamma(v)$ is nr. Instead of dividing the nr random draws into n independent groups with r draws in each group, it seems possible to use all the nr draws to construct simulated moments for each observation. Define

$$f_{n,l}(\theta, x_i) = \frac{1}{nr} \sum_{j=1}^{nr} h_l(v_j, x_i, \theta). \tag{2.9}$$

The simulator $f_{n,l}(\theta, x_i)$ is apparently unbiased and has smaller variance than $f_l(\theta, x_i)$ in (2.7) for the estimation of the reponse probability $P(l|\theta, x_i)$. With these simulated moments, a possible modification of McFadden's simulated method of moments can be

$$\min_{\theta}(D - f_n(\theta))' W' W (D - f_n(\theta)) \tag{2.10}$$

where $f_n'(\theta) = (f_n'(\theta, x_1), \cdots, f_n'(\theta, x_n))$ and $f_n'(\theta, x_i) = (f_{n,1}(\theta, x_i), \cdots, f_{n,L}(\theta, x_i))$ from (2.9). Since the number of random draws used in (2.9) is proportional to the sample size n, $f_{n,l}(\theta, x_i)$ is a consistent estimate of the response probability $P(l|\theta, x_i)$. The simulated moment formulation in (2.9) captures the idea of simulated frequency in Lerman and Manski [1981]. However, the number of draws in Lerman and Manski [1981] has no specific relation with sample size and the draws may not even be fixed as θ varies. The estimation method in (2.10) differs from McFadden's method only in the formulation of simulated moment. The Monte Carlo draws and sample observations provide the same data base for estimation. The computation of $f_{n,l}$ in (2.9) will involve summation of much more terms than the computation of f_l in (2.7). However, $f_{n,l}$ is a more efficient estimate than f_l. The reduction of variance in $f_{n,l}$ might provide more efficient estimate for θ. On the other hand, this modified method of simulated moments violates an "independence" condition in McFadden's formulation in that $f_{n,l}(\theta, x_i)$ are statistically dependent across observations as the simulated moments $f_{n,l}(\theta, x_i)$ for each i contain the same set of random numbers. Such dependence might render the estimator to be less efficient. The relative efficiency of these estimators will be investigated in subsequent sections.

3. Asymptotic Properties and Relative Efficiency

To simplify notation, let us consider the case that $r = 1$, i.e., one Monte Carlo draw for each observation. For observation i, let $d_i = (d_{i1}, \cdots, d_{iL})'$ be the vector of choice indicators and w_i a $K \times L$ matrix of instrumental variables. Furthermore, let $h(v, x, \theta) = (h_1(v, x, \theta), \cdots, h_L(v, x, \theta))'$. To justify our asymptotic analysis, the following regularity conditions are assumed for our model:

ASSUMPTION 1.

1. The sample observations (d_i, x_i, w_i), $i = 1, \cdots, n$ are i.i.d.
2. The parameter space Θ is a compact covex subset of a k dimensional Euclidean space and the true parameter vector θ_o is in the interior of Θ.
3. The choice probability vector $P(\theta, x)$ is twice continuously differentiable in θ.
4. $P(\theta, x)$ and its first and second order derivatives in θ are dominated by a vector of integrable functions $G(x)$ which is independent with θ such that $E(|wG(x)|) < \infty$.
5. The second moments of the instrumental variables in w exist and are finite.

ASSUMPTION 2.

1. The matrix of instruments W has rank $K \geq k$.
2. $E(w\frac{\partial P(\theta_o, x)}{\partial \theta'})$ has full rank.
3. $E[w(P(\theta_o, x) - P(\theta, x))] = 0$ only at $\theta = \theta_o$.

ASSUMPTION 3.

1. The random vector v is simulated independently with x and w.
2. The simulated moment function $h(v, x, \theta)$ is continuously twice differentiable in θ.
3. $h(v, x, \theta)$ is a conditionally unbiased estimator of the choice probability $P(\theta, x)$ conditional on x and w, for each θ in Θ.
4. $h(v, x, \theta)$ and its first and second order derivatives in θ are dominated by a vector of integrable functions $H(v, x)$ which is independent with θ such that $E(|wH(v, x)|) < \infty$.

5. The second moment of $wh(v, x, \theta_o)$ exists and is finite.

The regularty conditions in Assumption 1 are basic regularity conditons for the discrete choice model. Assumption 2 are rank conditions and identification condition for θ_o. These conditions are standard conditons for estimation with the conventional method of moments. Assumption 3 are regularity conditions for the simulation design. Detailed discussion on such design can be found in McFadden [1989]. The domination conditions in Assumptions 1 and 3 guarantee that Lebesgue dominated convergence (LDC) theorem can be applied to interchange limiting, differentiation and integration operations. The domination conditions are also needed for uniform laws of large numbers to apply. The existence of second moments in Assumptions 1 and 3 are needed for central limit theorems to apply. Unbiased moment simulators are assumed in Assumption 3(3). This assumption can be relaxed to include appropriate asymptotically unbiased moment simulators with proper modification of our subsequent analysis [3].

Since $W = [w_1, \cdots, w_n]$,

$$\begin{aligned} \frac{1}{n} W(D - f_n(\theta)) &= \frac{1}{n} \sum_{i=1}^{n} w_i(d_i - f_n(\theta, x_i)) \\ &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i(d_i - h(v_j, x_i, \theta)) \end{aligned} \tag{3.1}$$

is a V-statistic (Serfling [1980]). Under our regularity conditions, the uniform law of large numbers for V-statistics in Proposition 1 of Appendix implies that

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i(d_i - h(v_j, x_i, \theta)) \xrightarrow{p} E[w(P(\theta_o, x) - P(\theta, x))] \tag{3.2}$$

uniformly in $\theta \in \Theta$, where $P(\theta, x) = (P(1|\theta, x), \cdots, P(L|\theta, x))'$. Hence

$$\frac{1}{n^2}(D - f_n(\theta))' W' W (D - f_n(\theta)) \xrightarrow{p} \| E[w(P(\theta_o, x) - P(\theta, x))] \|^2 \tag{3.3}$$

[3] Asymptotic analysis with smoothing parameters can follow the analysis of semiparametric estimation in Ichimura and Lee [1988] and Lee [1989].

uniformly in $\theta \in \Theta$. Since $P(\theta, x)$ is continuous in θ by Assumption 1, the LDC theorem implies that $E(wP(\theta, x))$ is continuous in θ. Under the identification condition that $E[w(P(\theta_o, x) - P(\theta, x))] = 0$ only at $\theta = \theta_o$ in Assumption 2, the estimator $\hat{\theta}_d$ derived from (2.10) is consistent.

By a Taylor series expansion,

$$\begin{aligned} 0 &= \frac{\partial f_n'(\hat{\theta}_d)}{\partial \theta} W'W(D - f_n(\hat{\theta}_d)) \\ &= \frac{\partial f_n'(\theta_o)}{\partial \theta} W'W(D - f_n(\theta_o)) - \left\{ \frac{\partial f_n'(\bar{\theta})}{\partial \theta} W'W \frac{\partial f_n(\bar{\theta})}{\partial \theta'} \right. \\ &\quad \left. - \left[\frac{\partial^2 f_n'(\bar{\theta})}{\partial \theta \partial \theta_1} W'W(D - f_n(\bar{\theta})), \cdots, \frac{\partial^2 f_n'(\bar{\theta})}{\partial \theta \partial \theta_k} W'W(D - f_n(\bar{\theta})) \right] \right\} (\hat{\theta}_d - \theta_o) \end{aligned} \tag{3.4}$$

where $\bar{\theta}$ lies between θ_o and $\hat{\theta}_d$. Since

$$\begin{aligned} \frac{1}{n} W \frac{\partial f_n(\theta)}{\partial \theta'} &= \frac{1}{n} \sum_{i=1}^{n} w_i \frac{\partial f_n(\theta, x_i)}{\partial \theta'} \\ &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \frac{\partial h(v_j, x_i, \theta)}{\partial \theta'} \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} E\left(\frac{\partial h(v, x, \theta)}{\partial \theta'} | x, w \right) &= \frac{\partial E(h(v, x, \theta)|x)}{\partial \theta'} \\ &= \frac{\partial P(\theta, x)}{\partial \theta'}, \end{aligned} \tag{3.6}$$

the uniform law of large numbers for V-statistics implies that

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \frac{\partial h(v_j, x_i, \theta)}{\partial \theta'} \xrightarrow{p} E\left(w \frac{\partial P(\theta, x)}{\partial \theta'} \right) \tag{3.7}$$

uniformly in $\theta \in \Theta$. Since $E\left(w \frac{\partial P(\theta, x)}{\partial \theta'}\right)$ is continuous at θ_o by the LDC theorem, and $\bar{\theta}$ is a consistent estimate of θ_o,

$$\frac{1}{n} W \frac{\partial f_n(\bar{\theta})}{\partial \theta'} \xrightarrow{p} E\left(w \frac{\partial P(\theta_o, x)}{\partial \theta'} \right). \tag{3.8}$$

Similarly, since

$$\frac{1}{n} W \frac{\partial^2 f_n(\theta)}{\partial \theta' \partial \theta_l} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \frac{\partial^2 h(v_j, x_i, \theta)}{\partial \theta' \partial \theta_l} \tag{3.9}$$

and

$$E\left(\frac{\partial^2 h(v, x, \theta)}{\partial \theta' \partial \theta_l} | x, w \right) = \frac{\partial^2 P(\theta, x)}{\partial \theta' \partial \theta_l}, \tag{3.10}$$

we have

$$\frac{1}{n} W \frac{\partial^2 f_n(\bar{\theta})}{\partial\theta' \partial\theta_l} \xrightarrow{p} E\left(w \frac{\partial^2 P(\theta_o, x)}{\partial\theta' \partial\theta_l} \right) \tag{3.11}$$

which is finite. On the other hand, the asymptotic distribution of $\frac{1}{\sqrt{n}} W(D - f_n(\theta_o))$ can be derived from the central limit theorem for V-statistics in Serfling [1980] (see Proposition 2 in Appendix). Since

$$\frac{1}{\sqrt{n}} W(D - f_n(\theta_o)) = \frac{1}{n\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_i(d_i - h(v_j, x_i, \theta)), \tag{3.12}$$

the kernel associated with this V-statistic is

$$\psi(z_i, z_j) = w_i(d_i - h(v_j, x_i, \theta_o)) \tag{3.13}$$

where $z = (w, x, d, v)$. The kernel ψ can be rewritten as

$$\psi(z_i, z_j) = w_i(d_i - P(\theta_o, x_i)) - w_i(h(v_j, x_i, \theta_o) - P(\theta_o, x_i)). \tag{3.14}$$

Since

$$E(\psi(z_i, z_j) | z_i) = w_i(d_i - P(\theta_o, x_i)) \tag{3.15}$$

and

$$E(\psi(z_j, z_i) | z_i) = -E[w(h(v_i, x, \theta_o) - P(\theta_o, x)) | v_i], \tag{3.16}$$

the central limit theorem for V-statistics implies that

$$\frac{1}{\sqrt{n}} W(D - f_n(\theta_o)) \xrightarrow{D} N(0, \Omega_d) \tag{3.17}$$

where

$$\begin{aligned} \Omega_d = {} & E\Big\{ \big[w(d - P(\theta_o, x)) - E(w[h(v, x, \theta_o) - P(\theta_o, x)] | v) \big] \\ & \cdot \big[w(d - P(\theta_o, x)) - E(w[h(v, x, \theta_o) - P(\theta_o, x)] | v) \big]' \Big\} \\ = {} & E\{ w(d - P(\theta_o, x))(d - P(\theta_o, x))' w' \} \\ & + E\Big\{ E(w[h(v, x, \theta_o) - P(\theta_o, x)] | v) E(w[h(v, x, \theta_o) - P(\theta_o, x)] | v)' \Big\}. \end{aligned} \tag{3.18}$$

The latter equality in (3.18) follows because v is simulated independently with x and w. It follows from (3.4) and the above asymptotic results that

$$\sqrt{n}(\hat{\theta}_d - \theta_\circ) \xrightarrow{D} N(0, \Sigma_d) \tag{3.19}$$

where

$$\begin{aligned}\Sigma_d = &\left\{E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})'E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})\right\}^{-1}\\ &\cdot E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})'\Omega_d E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})\left\{E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})'E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})\right\}^{-1}.\end{aligned} \tag{3.20}$$

The asymptotic covariance matrix of the estimator $\hat{\theta}_d$ can be compared with the asymptotic covariance matrix of McFadden's estimator $\hat{\theta}_M$. From McFadden [1989], we know that

$$\sqrt{n}(\hat{\theta}_M - \theta_\circ) \xrightarrow{D} N(0, \Sigma_M) \tag{3.21}$$

where

$$\begin{aligned}\Sigma_M = &\left\{E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})'E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})\right\}^{-1}\\ &\cdot E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})'\Omega_M E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})\left\{E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})'E(w\frac{\partial P(\theta_\circ,x)}{\partial\theta'})\right\}^{-1}\end{aligned} \tag{3.22}$$

and

$$\begin{aligned}\Omega_M = &E\left\{w(d - P(\theta_\circ,x))(d - P(\theta_\circ,x))'w'\right\}\\ &+ E\left\{w[h(v,x,\theta_\circ) - P(\theta_\circ,x)][h(v,x,\theta_\circ) - P(\theta_\circ,x)]'w'\right\}.\end{aligned} \tag{3.23}$$

By projection,

$$\begin{aligned}&E\left\{E[w(h(v,x,\theta_\circ) - P(\theta_\circ,x))|v]E[w(h(v,x,\theta_\circ) - P(\theta_\circ,x))|v]'\right\}\\ \le\ &E\left\{w(h(v,x,\theta_\circ) - P(\theta_\circ,x))(h(v,x,\theta_\circ) - P(\theta_\circ,x))'w'\right\}.\end{aligned} \tag{3.24}$$

Hence, the estimator $\hat{\theta}_d$ from (2.10) is asymptotically efficient relative to McFadden's estimator $\hat{\theta}_M$ form (2.8).

The above analysis can be generalized to the case with r Monte Carlo draws for each observation. For each individual i, let $v_1^i, \cdots, v_r^i$ be r draws. The results will follow by replacing the function $h(v_j, x_i, \theta)$ with $\frac{1}{r}\sum_{s=1}^{r} h(v_s^i, x_i, \theta)$ in the above analysis. For this general case, the estimator $\hat{\theta}_d$ will be asymptotically $N(\theta_o, \frac{1}{n}\Sigma_{d,r})$ where $\Sigma_{d,r}$ is the matrix in (3.20) with Ω_d replaced by the matrix $\Omega_{d,r}$:

$$\begin{aligned}\Omega_{d,r} = & E\{w(d - P(\theta_o, x))(d - P(\theta_o, x))'w'\} \\ & + \frac{1}{r}E\left\{E[w(h(v, x, \theta_o) - P(\theta_o, x))|v] \cdot E[w(h(v, x, \theta_o) - P(\theta_o, x))|v]'\right\}.\end{aligned} \tag{3.25}$$

The asymptotic covariance matrix of the corresponding McFadden's estimator will be $\frac{1}{n}\Sigma_{M,r}$ where $\Sigma_{M,r}$ is the matrix in (3.22) with Ω_M replaced by the matrix

$$\begin{aligned}\Omega_{M,r} = & E\{w(d - P(\theta_o, x))(d - P(\theta_o, x))'w'\} \\ & + \frac{1}{r}E\left\{w(h(v, x, \theta_o) - P(\theta_o, x))(h(v, x, \theta_o) - P(\theta_o, x))'w'\right\}.\end{aligned} \tag{3.26}$$

Comparing (3.25) with (3.26), the estimator $\hat{\theta}_d$ is asymptotically efficient relative to the estimator $\hat{\theta}_M$.

4. Maximum Simulated Likelihood Estimation

In the method of simulated moments of McFadden [1989], the instrumental variables w are independent with the Monte Carlo draws for the simulated moments. As implied by the method of maximum likelihood for discrete reponse models, the optimal instrument vector is $w = \frac{\partial \ln P'(\theta_o, x)}{\partial \theta}$. In McFadden's method, to approximate the optimal instrument with simulated moments, the random numbers used to compute $f(\theta)$ in (2.8) are independent of any simulation used in the construction of the instruments. Such design is needed in McFaddens' method to guarantee that the moment equations have zero mean. For our modified method, the simulated moments $f_n(\theta, x)$ are consistent for each observation. Therefore it might be possible to use the same Monte Carlo draws to compute both the simulated moments and the instrumental array. As the optimal moment equations are derived as the first order conditions for maximizing the likelihood function, it will be convenient to consider the likelihood method directly. In this section, we will analyze asymptotic properties of a maximum simulated likelihood method for the discrete choice model.

The log likelihood function for our discrete choice model is

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \ln P(l|\theta, x_i). \tag{4.1}$$

By replacing the response probabilities with simulated moments, we are working with a pseudo likelihood function. Without loss of generality, consider the simulated moments in (2.9) with $r = 1$. The log pseudo likelihood function is

$$L(\theta) = \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \ln f_{n,l}(\theta, x_i). \tag{4.2}$$

This estimation method can be regarded as a generalization of the simulated maximum likelihood method in Pakes [1986] (see also Pakes and Pollard [1989]) to disaggregated data with smooth simulator. For aggregated data, because the exogenous variables vector x is constant across observation, $f_{n,l}$ will simply be a simulated smooth probability estimator and $\frac{1}{n}\sum_{i=1}^{n} d_{li}$ will be the observed sample frequency for the alternative l[4]. This simulated likelihood approach may have computational advantage over the method of simulated moments in (2.10) when the number of choice alternatives L is large. For the method of simulated moments, for each observation, the simulated moments $f_{n,l}(\theta, x_i)$ need to be computed for all the alternatives l, $l = 1, \cdots, L$. For the simulated likelihood method, only the simulated moment corresponding to the chosen alternative needs to be computed for each observation. This estimation method will be of particular interest for estimation of panel data models where the total number of choice patterns over time can be quite large [5]. To justify our subsequent analysis, in addition to our assumptions 1 and 3, we assume that

ASSUMPTION 4.

1. The support X of x is a compact set.
2. The choice probability vector $P(\theta, x)$ is continuous in $(\theta, x) \in \Theta \times X$.
3. The conditional second moments of $h(v, x, \theta_o)$ and $\frac{\partial h(v,x,\theta_o)}{\partial \theta'}$ conditional on x exist and are uniformly bounded on X.

ASSUMPTION 5.

1. θ_o is the unique minimizer of the function $E(\sum_{l=1}^{L} P(l|\theta_o, x) \ln P(l|\theta, x))$.
2. The matrix $E\left[\sum_{l=1}^{L} P(l|\theta_o, x) \frac{\partial \ln P(l|\theta_o,x)}{\partial \theta} \frac{\partial \ln P(l|\theta_o,x)}{\partial \theta'}\right]$ is nonsingular.

Assumption 4 guarantees that the choice probabilities $P(l|\theta, x)$, $l = 1, \cdots, L$ are strictly bounded away from zero on $\Theta \times X$ which is needed to establish uniform convergence properties of the derivatives of $\ln f_{n,l}(\theta, x)$ with θ. The moment conditions are also needed for similar purpose.

[4] In Pakes [1986], the proposed simulated moment estimator is a simple frequency estimator.

[5] The difficulty of applying the method of simulated moments to panel data model has been pointed out in Keane [1989]. An possible alternative method of simulated moments have been suggested in Keane's article. McFadden [1989] has pointed out that for such models the use of simple frequency simulator instead of smooth simulator for the method of simulated moments may be more desirable. However, such method will involve nonsmooth function for minimization.

Assumption 5 are identification and rank conditions for the classical maximum likelihood estimation of discrete choice model. The identification condition can usually be justified with some other regularity conditions on choice probabilities via Jensen's inequality.

Let $\hat{\theta}_L$ denote the maximum simulated likelihood estimator. The asymptotic properties of consistency and asymptotic normality depend crucially on asymptotic properties of the simulated moments. Since the support X of x and the parameter space Θ of θ are compact sets, the uniform law of large numbers in Amemiya [1985] implies that

$$\sup_{\Theta \times X} \|f_n(\theta, x) - P(\theta, x)\| \xrightarrow{p} 0. \tag{4.3}$$

Since $P(\theta, x)$ is bounded away from zero on $\Theta \times X$ by Assumption 4, it follows from (4.3) that

$$\sup_{\Theta \times X} |\frac{1}{n} L(\theta) - \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \ln P(l|\theta, x_i)| \xrightarrow{p} 0. \tag{4.4}$$

The uniform law of large numbers in Amemiya [1985] implies also that

$$\sup_{\Theta} |\frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \ln P(l|\theta, x_i) - E(\sum_{l=1}^{L} P(l|\theta_o, x) \ln P(l|\theta, x))| \xrightarrow{p} 0. \tag{4.5}$$

Since $\frac{1}{n} L(\theta)$ converges in probability uniformly to the limit function $E(\sum_{l=1}^{L} P(l|\theta_o, x) \ln P(l|\theta, x))$ on Θ and θ_o is the unique minimizer of the limit function by Assumption 5, $\hat{\theta}_L$ is consistent.

The estimator $\hat{\theta}_L$ satisfies the first order condition:

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln f_{n,l}(\hat{\theta}_L, x_i)}{\partial \theta} = 0. \tag{4.6}$$

By a Taylor expansion,

$$\sqrt{n}(\hat{\theta}_L - \theta_o) = -\left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial^2 \ln f_{n,l}(\bar{\theta}, x_i)}{\partial \theta \partial \theta'} \right\}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln f_{n,l}(\theta_o, x_i)}{\partial \theta} \tag{4.7}$$

where $\bar{\theta}$ lies between $\hat{\theta}_L$ and θ_o. By the uniform law of large numbers in Amemiya [1985],

$$\sup_{\Theta \times X} \|\frac{\partial f_{n,l}(\theta, x)}{\partial \theta} - \frac{\partial P(l|\theta, x)}{\partial \theta}\| \xrightarrow{p} 0. \tag{4.8}$$

and

$$\sup_{\Theta\times X} \| \frac{\partial^2 f_{n,l}(\theta,x)}{\partial\theta\partial\theta'} - \frac{\partial^2 P(l|\theta,x)}{\partial\theta\partial\theta'} \| \xrightarrow{p} 0 \tag{4.9}$$

for all $l = 1, \cdots, L$. It follows that

$$\begin{aligned} \frac{\partial \ln f_{n,l}(\theta,x)}{\partial\theta} &= \frac{1}{f_{n,l}(\theta,x)} \frac{\partial f_{n,l}(\theta,x)}{\partial\theta} \\ &\xrightarrow{p} \frac{\partial \ln P(l|\theta,x)}{\partial\theta} \end{aligned} \tag{4.10}$$

and

$$\begin{aligned} \frac{\partial^2 \ln f_{n,l}(\theta,x)}{\partial\theta\partial\theta'} &= \frac{1}{f_{n,l}(\theta,x)} \frac{\partial^2 f_{n,l}(\theta,x)}{\partial\theta\partial\theta'} - \frac{1}{f^2_{n,l}(\theta,x)} \frac{\partial f_{n,l}(\theta,x)}{\partial\theta} \frac{\partial f_{n,l}(\theta,x)}{\partial\theta'} \\ &\xrightarrow{p} \frac{\partial^2 \ln P(l|\theta,x)}{\partial\theta\partial\theta'} \end{aligned} \tag{4.11}$$

uniformly in $(\theta,x) \in \Theta \times X$. Since $\frac{1}{n}\sum_{i=1}^n \sum_{l=1}^L d_{li} \frac{\partial^2}{\partial\theta\partial\theta'} \ln P(l|\theta_o,x)$ converges in probability to $E(d_l \frac{\partial^2}{\partial\theta\partial\theta'} \ln P(l|\theta_o,x))$ and $\bar{\theta}$ is a consistent estimate of θ_o, (4.11) implies that

$$\frac{1}{n}\sum_{i=1}^n \sum_{l=1}^L d_{li} \frac{\partial^2 \ln f_{n,l}(\bar{\theta},x_i)}{\partial\theta\partial\theta'} \xrightarrow{p} E\{\sum_{l=1}^L d_l \frac{\partial^2 \ln P(l|\theta_o,x)}{\partial\theta\partial\theta'}\}. \tag{4.12}$$

It is well known that

$$E\{\sum_{l=1}^L d_l \frac{\partial^2 \ln P(l|\theta_o,x)}{\partial\theta\partial\theta'}\} = -E[\sum_{l=1}^L P(l|\theta_o,x) \frac{\partial \ln P(l|\theta_o,x)}{\partial\theta} \frac{\partial \ln P(l|\theta_o,x)}{\partial\theta'}]. \tag{4.13}$$

The first order term in (4.7) can be analyzed with Taylor series expansion. By a Taylor expansion up to the second order,

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{l=1}^L d_{li} \frac{\partial}{\partial\theta} \ln f_{n,l}(\theta_o,x_i) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{l=1}^L d_{li} \frac{\partial \ln P(l|\theta_o,x_i)}{\partial\theta} + L_n + R_n \end{aligned} \tag{4.14}$$

where

$$L_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{l=1}^L d_{li} \frac{1}{P(l|\theta_o,x_i)} \left[\frac{\partial f_{n,l}(\theta_o,x_i)}{\partial\theta} - \frac{\partial \ln P(l|\theta_o,x_i)}{\partial\theta} f_{n,l}(\theta_o,x_i)\right] \tag{4.15}$$

and

$$\begin{aligned} R_n =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \sum_{l=1}^L d_{li}\Big[- \frac{1}{\tilde{f}^2_{n,l}(i)} (\frac{\partial f_{n,l}(\theta_o,x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o,x_i)}{\partial\theta})(f_{n,l}(\theta_o,x_i) - P(l|\theta_o,x_i)) \\ &+ \frac{1}{\tilde{f}^3_{n,l}(i)} \frac{\partial \tilde{f}_{n,l}(i)}{\partial\theta} (f_{n,l}(\theta_o,x_i) - P(l|\theta_o,x_i))^2\Big] \end{aligned} \tag{4.16}$$

where $\tilde{f}_{n,l}(i)$ lies between $f_{n,l}(\theta_o, x_i)$ and $P(l|\theta_o, x_i)$, and $\frac{\partial \tilde{f}_{n,l}(i)}{\partial \theta}$ lies between $\frac{\partial f_{n,l}(\theta_o, x_i)}{\partial \theta}$ and $\frac{\partial P(l|\theta_o, x_i)}{\partial \theta}$. Since $P(\theta_o, x)$ is bounded away from zero on X, (4.3) implies that $\tilde{f}_{n,l}(i)$ is bounded away from zero on X in probability. Hence

$$\begin{aligned} \|R_n\| =& O_p(1) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \| \frac{\partial f_{n,l}(\theta_o, x_i)}{\partial \theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial \theta} \| \cdot |f_{n,l}(\theta_o, x_i) - P(l|\theta_o, x_i)| \\ &+ O_p(1) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} |f_{n,l}(\theta_o, x_i) - P(l|\theta_o, x_i)|^2 . \end{aligned} \tag{4.17}$$

Since X is compact, and $var(h_l(v, x, \theta_o)|x)$ and $var(\frac{\partial}{\partial \theta} h_l(v, x, \theta_o)|x)$, $l = 1, \cdots, L$ are uniformly bounded on X by Assumption 4,

$$\begin{aligned} \sup_X E((f_{n,l}(\theta_o, x) - P(l|\theta_o, x))^2 |x) &= \frac{1}{n} \sup_X var(h_l(v, x, \theta_o)|x) \\ &= O(\frac{1}{n}) \end{aligned} \tag{4.18}$$

and

$$\begin{aligned} \sup_X E((\frac{\partial f_{n,l}(\theta_o, x)}{\partial \theta_j} - \frac{\partial P(l|\theta_o, x)}{\partial \theta_j})^2 |x) &= \frac{1}{n} \sup_X var(\frac{\partial h_l(v, x, \theta_o)}{\partial \theta_j} |x) \\ &= O(\frac{1}{n}) \end{aligned} \tag{4.19}$$

for each component θ_j of θ. By Markov inequality and (4.18),

$$\begin{aligned} & P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} |f_{n,l}(\theta_o, x_i) - P(l|\theta_o, x_i)|^2 > \epsilon \right) \\ & \leq \frac{\sqrt{n}}{\epsilon} \sum_{l=1}^{L} E(f_{n,l}(\theta_o, x) - P(l|\theta_o, x))^2 \\ & = \frac{\sqrt{n}}{\epsilon} \sum_{l=1}^{L} E \left\{ E[(f_{n,l}(\theta_o, x) - P(l|\theta_o, x))^2 |x] \right\} \\ & = O(\frac{1}{\sqrt{n}}). \end{aligned} \tag{4.20}$$

Similarly, for each component θ_j of θ, by Markov and Cauchy inequalities

$$\begin{aligned} & P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} | \frac{\partial f_{n,l}(\theta_o, x_i)}{\partial \theta_j} - \frac{\partial P(l|\theta_o, x_i)}{\partial \theta_j} | \cdot |f_{n,l}(\theta_o, x_i) - P(l|\theta_o, x_i)| > \epsilon \right) \\ & \leq \frac{\sqrt{n}}{\epsilon} \sum_{l=1}^{L} E(| \frac{\partial f_{n,l}(\theta_o, x)}{\partial \theta_j} - \frac{\partial P(l|\theta_o, x)}{\partial \theta_j} | \cdot |f_{n,l}(\theta_o, x) - P(l|\theta_o, x)|) \\ & \leq \frac{\sqrt{n}}{\epsilon} \sum_{l=1}^{L} \left\{ E(\frac{\partial f_{n,l}(\theta_o, x)}{\partial \theta_j} - \frac{\partial P(l|\theta_o, x)}{\partial \theta_j})^2 \cdot E(f_{n,l}(\theta_o, x) - P(l|\theta_o, x))^2 \right\}^{\frac{1}{2}} \\ & = O(\frac{1}{\sqrt{n}}). \end{aligned} \tag{4.21}$$

Equations (4.17), (4.20) and (4.21) imply that $R_n \xrightarrow{p} 0$.

It remains to analyze the first two terms in (4.14). Define a kernel

$$\phi(z_i, z_j) = \sum_{l=1}^{L} d_{li}\{\frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} + \frac{1}{P(l|\theta_o, x_i)}[\frac{\partial h_l(v_j, x_i, \theta_o)}{\partial \theta} - \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} h_l(v_j, x_i, \theta_o)]\} \tag{4.22}$$

where $z = (x, d, v)$. We have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} + L_n = \frac{1}{n\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi(z_i, z_j). \tag{4.23}$$

The asymptotic distribution of (4.23) can be derived from the central limit theorem for V-statistics. As $h_l(v_j, x_i, \theta_o)$ is a conditional unbiased estimate of $P(l|\theta_o, x_i)$,

$$E(\phi(z_i, z_j)|z_i) = \sum_{l=1}^{L} d_{li} \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} \tag{4.24}$$

and

$$E(\phi(z_j, z_i)|z_i) = \sum_{l=1}^{L} E\left\{\frac{\partial h_l(v_i, x, \theta_o)}{\partial \theta} - \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} h_l(v_i, x, \theta_o)|v_i\right\}. \tag{4.25}$$

The central limit theorem for V-statistics implies that

$$\frac{1}{n\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi(z_i, z_j) \xrightarrow{D} N(0, \Omega_L) \tag{4.26}$$

where

$$\begin{aligned} \Omega_L =& E[\sum_{l=1}^{L} P(l|\theta_o, x) \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta'}] \\ &+ E\{\sum_{l=1}^{L} E[\frac{\partial h_l(v, x, \theta_o)}{\partial \theta} - \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} h_l(v, x, \theta_o)|v] \\ &\cdot \sum_{l=1}^{L} E[\frac{\partial h_l(v, x, \theta_o)}{\partial \theta} - \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} h_l(v, x, \theta_o)|v]'\}. \end{aligned} \tag{4.27}$$

Finally, (4.7), (4.12), (4.14), (4.23) and (4.26) imply that

$$\sqrt{n}(\hat{\theta}_L - \theta_o) \xrightarrow{D} N(0, \Sigma_L) \tag{4.28}$$

where

$$\Sigma_L = \Sigma_m \Omega_L \Sigma_m \tag{4.29}$$

and

$$\Sigma_m = \left\{ E\left[\sum_{l=1}^{L} P(l|\theta_o, x) \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta'} \right] \right\}^{-1}. \tag{4.30}$$

The classical maximum likelihood estimator from (4.1) is asymptotically efficient with asymptotic covariance matrix Σ_m. Since

$$\begin{aligned} \Sigma_L = \Sigma_m + \Sigma_m E\{ &\sum_{l=1}^{L} E[\frac{\partial h_l(v,x,\theta_o)}{\partial \theta} - \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} h_l(v,x,\theta_o)|v] \\ &\cdot \sum_{l=1}^{L} E[\frac{\partial h_l(v,x,\theta_o)}{\partial \theta} - \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} h_l(v,x,\theta_o)|v]' \} \Sigma_m \end{aligned} \tag{4.31}$$

the maximum simulated likelihood estimator $\hat{\theta}_L$ is asymptotically inefficient as the second term in (4.31) does not vanish in general as sample size increases. This term reflects the error introduced in the simulated likelihood function. Comparing the asymptotic distribution of $\hat{\theta}_L$ in (4.28) with the asymptotic distribution of $\hat{\theta}_d$ with the optimal instrument $w_i = \frac{\partial \ln P'(\theta_o, x_i)}{\partial \theta}$ in (3.19) (as if it were completely observable), the difference appears only in the matrices Ω_L and Ω_d. With the optimal instrument, Ω_d in (3.18) becomes

$$\begin{aligned} \Omega_d =& E[\sum_{l=1}^{L} P(l|\theta_o, x) \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta'}] \\ &+ E\{ \sum_{l=1}^{L} E[\frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} (h_l(v,x,\theta_o) - P(l|\theta_o, x))|v] \\ &\cdot \sum_{l=1}^{L} E[\frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} (h_l(v,x,\theta_o) - P(l|\theta_o, x))|v]' \}. \end{aligned} \tag{4.32}$$

Since

$$\begin{aligned} &\sum_{l=1}^{L} E[\frac{\partial h_l(v,x,\theta_o)}{\partial \theta} - \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} h_l(v,x,\theta_o)|v] \\ =& \sum_{l=1}^{L} E[(\frac{\partial h_l(v,x,\theta_o)}{\partial \theta} - \frac{\partial P(l|\theta_o, x)}{\partial \theta}) - \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} (h_l(v,x,\theta_o) - P(l|\theta_o, x))|v], \end{aligned} \tag{4.33}$$

the component $\sum_{l=1}^{L} E(\frac{\partial h_l(v,x,\theta_o)}{\partial \theta} - \frac{\partial P(l|\theta_o,x)}{\partial \theta}|v)$ may be interpreted as the additional error introduced in the simulation of instruments.

Generalization of the above analysis to the case with r Monte Carlo draws for each observation is straightforward. The results will follow by replacing the function $h(v_j, x_i, \theta)$ with $\frac{1}{r}\sum_{s=1}^{r} h(v_s^i, x_i, \theta)$ in the analysis. With r Monte Carlo draws for each observation, the estimator $\hat{\theta}_L$ will be asymptotically $N(\theta_o, \frac{1}{n}\Sigma_{L,r})$ where

$$\begin{aligned}\Sigma_{L,r} = \Sigma_m + \frac{1}{r}\Sigma_m E\{ &\sum_{l=1}^{L} E[\frac{\partial h_l(v,x,\theta_o)}{\partial \theta} - \frac{\partial \ln P(l|\theta_o,x)}{\partial \theta} h_l(v,x,\theta_o)|v] \\ &\cdot \sum_{l=1}^{L} E[\frac{\partial h_l(v,x,\theta_o)}{\partial \theta} - \frac{\partial \ln P(l|\theta_o,x)}{\partial \theta} h_l(v,x,\theta_o)|v]' \}\Sigma_m\end{aligned} \tag{4.34}$$

5. Simulators with Adding Up Property

As we have pointed out in the previous section that the asymptotic covariance of the maximum simulated likelihood estimator $\hat{\theta}_L$ from (4.1) differs from the asymptotic covariance of the modified simulated moment estimator $\hat{\theta}_d$ with instruments $w_i = \frac{\partial \ln P'(\theta_o, x_i)}{\partial \theta}$ in that an additional error component $\sum_{l=1}^{L} E(\frac{\partial h_l(v,x,\theta_o)}{\partial \theta} - \frac{\partial P(l|\theta_o,x)}{\partial \theta}|v)$ has been introduced. However, this additional error can vanish if $\sum_{l=1}^{L} h_l(v,x,\theta_o) = 1$. Functions h_l with this property imply that the simulators $f_{n,l}(\theta,x)$ from (2.9) have an adding up property that $\sum_{l=1}^{L} f_{n,l}(\theta,x) = 1$. This adding up property is intuitively desirable as $f_{n,l}(\theta,x)$ are estimates of probabilities. Eventhought this adding up property does not seem to play an explicit role in McFadden's method of simulated moments, it has an important implication for our method of simulated likelihood estimation. Maximum simulated likelihood estimators derived from simulators with adding up property have the same limiting distribution of the estimator derived from the modified method of simulated moments with instruments $w_i = \frac{\partial \ln P'(\theta_o, x_i)}{\partial \theta}$ as if they were completely observable. Simulated moment estimators with adding up property have been introduced in McFadden [1989] and Stern [1987]. Stern's simulated moment estimator is appealing for models with error components.

Simulators derived from the importance sampling technique, in general, may not have the adding up property since h_l is simply ratios of densities. Adding up property can always be satisfied by normalizing the orignial simulators. Adding up property is satisfied for the normalized moment simulators $f_{n,l}(\theta,x)/\sum_{l=1}^{L} f_{n,l}(\theta,x)$. This suggests the maximum simulated likelihood estimation with the following function :

$$L(\theta) = \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li} \ln \frac{f_{n,l}(\theta,x_i)}{\sum_{j=1}^{L} f_{n,j}(\theta,x_i)}. \tag{5.1}$$

Let $\tilde{\theta}_L$ be the maximum simulated likelihood estimator derived from (5.1). Consistency of $\tilde{\theta}_L$ is apparent from arguments in the previous section. Asymptotic distribution of $\tilde{\theta}_L$ remains to be investigated. By a Taylor expansion,

$$\sqrt{n}(\tilde{\theta}_L - \theta_o) = -\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{n,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'} - \frac{\partial^2 \ln \sum_{j=1}^{L} f_{n,j}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right)\right\}^{-1}$$
$$\cdot \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\sum_{l=1}^{L} d_{li}\frac{\partial \ln f_{n,l}(\theta_o, x_i)}{\partial\theta} - \frac{\partial \ln \sum_{j=1}^{L} f_{n,j}(\theta_o, x_i)}{\partial\theta}\right). \tag{5.2}$$

It follows from (4.11) that

$$\frac{\partial^2 \ln \sum_{l=1}^{L} f_{n,l}(\theta, x)}{\partial\theta\partial\theta'} \xrightarrow{p} \frac{\partial^2 \ln \sum_{l=1}^{L} P(l|\theta, x)}{\partial\theta\partial\theta'}$$
$$= 0 \tag{5.3}$$

uniformly in $(\theta, x) \in \Theta \times X$. Hence

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{n,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'} - \frac{\partial^2 \ln \sum_{j=1}^{L} f_{n,j}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right) \xrightarrow{p} E\left\{\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln P(l|\theta_o, x)}{\partial\theta\partial\theta'}\right\}. \tag{5.4}$$

The remaining term in (5.2) can be analyzed with a similar expansion in (4.14) and (4.23):

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\sum_{l=1}^{L} d_{li}\frac{\partial \ln f_{n,l}(\theta_o, x_i)}{\partial\theta} - \frac{\partial \ln \sum_{j=1}^{L} f_{n,j}(\theta_o, x_i)}{\partial\theta}\right)$$
$$= \frac{1}{n\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi(z_i, z_j) + o_p(1) \tag{5.5}$$

where

$$\phi(z_i, z_j) = \sum_{l=1}^{L} d_{li}\left\{\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + \frac{1}{P(l|\theta_o, x_i)}\left[\frac{\partial h_l(v_j, x_i, \theta_o)}{\partial\theta} - \frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta}h_l(v_j, x_i, \theta_o)\right]\right\}$$
$$- \sum_{l=1}^{L}\frac{\partial h_l(v_j, x_i, \theta_o)}{\partial\theta}. \tag{5.6}$$

It is apparent that

$$E(\phi(z_i, z_j)|z_i) = \sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta}$$

and

$$E(\phi(z_j, z_i)|z_i) = -\sum_{l=1}^{L} E\left\{\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta}h_l(v_i, x, \theta_o)|v_i\right\}.$$

Therefore

$$\sqrt{n}(\tilde{\theta}_L - \theta_o) \xrightarrow{D} N(0, \Sigma_m\Omega_d\Sigma_m) \tag{5.7}$$

where Σ_m is the matrix in (4.30) and Ω_d is the matrix in (4.32). This estimator has the same limiting distribution as the estimator derived from the modified method of simulated momemts with instruments $w_i = \frac{\partial \ln P'(\theta_o, x_i)}{\partial \theta}$ as if they were completely observable. Adding up property has indeed an interesting implication on the asymptotic property of our simulated likelihood estimators.

6. Dependent and Independent Simulated Moments

The estimator $\hat{\theta}_d$ derived from the method of simulated moments with $f_{n,l}(\theta, x_i)$ in (2.9) is not asymptotically efficient relative to the conventional method of moment estimator $\hat{\theta}_c$ derived from (2.3). The conventional moment estimator is asymptotically normal and

$$\sqrt{n}(\hat{\theta}_c - \theta_o) \xrightarrow{D} N(0,\ \Sigma_c) \tag{6.1}$$

where

$$\begin{aligned}\Sigma_c = &\left\{E\left(w\frac{\partial P(\theta_o, x)}{\partial \theta'}\right)' E\left(w\frac{\partial P(\theta_o, x)}{\partial \theta'}\right)\right\}^{-1} \\ &\cdot E\left(w\frac{\partial P(\theta_o, x)}{\partial \theta'}\right)' \Omega_c E\left(w\frac{\partial P(\theta_o, x)}{\partial \theta'}\right)\left\{E\left(w\frac{\partial P(\theta_o, x)}{\partial \theta'}\right)' E\left(w\frac{\partial P(\theta_o, x)}{\partial \theta'}\right)\right\}^{-1}\end{aligned} \tag{6.2}$$

and

$$\Omega_c = E\{w(d - P(\theta_o, x))(d - P(\theta_o, x)'w'\}$$

(see McFadden [1989]). The estimator $\hat{\theta}_d$ is not asymptotically efficient because the second term of Ω_d in (3.18) (or in (3.25) with r draws) does not vanish asymptotically. The loss of efficiency is due to dependency of $f_n(\theta, x_i)$ across observation.

To clarify this issue, consider a different simulator constructed with n independent Monte Carlo draws for each observation. The total number of random draws from $\gamma(v)$ is n^2. Let $v_j^{(i)}$, $j = 1, \cdots, n$ be the n draws for observation i. Define

$$f^*_{n,l}(\theta, x_i) = \frac{1}{n}\sum_{j=1}^{n} h_l(v_j^{(i)}, x_i, \theta). \tag{6.3}$$

Let $\hat{\theta}_I$ denote the estimator derived from

$$\min_{\theta}(D - f^*_n(\theta))'W'W(D - f^*_n(\theta)). \tag{6.4}$$

Since

$$\frac{1}{n}W(D - f_n^*(\theta)) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i(d_i - h(v_j^{(i)}, x_i, \theta)), \tag{6.5}$$

the uniform law of large numbers in Proposition 3 of the appendix implies that

$$\frac{1}{n}W(D - f_n^*(\theta)) \xrightarrow{p} E[w(P(\theta_o, x) - P(\theta, x))] \tag{6.6}$$

uniformly in $\theta \in \Theta$. Under the identification condition of Assumption 2, $\hat{\theta}_I$ is consistent. Similarly, Proposition 3 implies that for any consistent estimate $\bar{\theta}$ of θ_o,

$$\frac{1}{n}W\frac{\partial f_n^*(\bar{\theta})}{\partial \theta'} \xrightarrow{p} E\left(w\frac{\partial P(\theta_o, x)}{\partial \theta'}\right) \tag{6.7}$$

and

$$\frac{1}{n}W\frac{\partial^2 f_n^*(\bar{\theta})}{\partial \theta' \partial \theta_l} \xrightarrow{p} E\left(w\frac{\partial^2 P(\theta_o, x)}{\partial \theta' \partial \theta_l}\right). \tag{6.8}$$

On the other hand, the central limit theorem in Proposition 4 of the appendix implies that

$$\frac{1}{\sqrt{n}}W(D - f_n^*(\theta_o)) \xrightarrow{D} N(0,\ E[w(d - P(\theta_o, x))(d - P(\theta_o, x))'w']). \tag{6.9}$$

Hence, it follows from a Taylor series expansion similar to (3.4) that

$$\sqrt{n}(\hat{\theta}_I - \theta_o) \xrightarrow{D} N(0,\ \Sigma_c). \tag{6.10}$$

The estimator $\hat{\theta}_I$ is asymptotically efficient as the conventional moment estimator $\hat{\theta}_c$. This estimator $\hat{\theta}_I$ is asymptotically efficient relative to $\hat{\theta}_d$. However, the efficiency gain is obtained at the expense of taking much larger number of Monte Carlo draws.

Similar conclusions hold for the maximum simulated likelihood estimation with the simulated moments in (6.3). Proposition 4 implies that the corresponding term similar to L_n in (4.15) will now converge in distribution to zero and hence converge in probability to zero. The corresponding simulated likelihood estimator is asymptotically efficient.

7. Some Monte Carlo Simulations

In this section, we report some limited Monte Carlo results on the finite sample performance of our estimators. We have experimented with two discrete choice models. The first model is a binary choice model and the second model is a dynamic discrete choice panel data model.

The binary choice model in the simulation is specified as

$$y^* = \beta_1 + \beta_2 x + \epsilon \tag{7.1}$$

where y^* is a latent dependent variable which sign determines the observed indicator d as $d = 1$ if $y^* > 0$ and $d = 0$, otherwise. The disturbance ϵ is generated by a standard normal variable $N(0,\ 1)$. The exogenous variable x is a truncated normal $N(0,\ 1)$ variable with support on $[-2,\ 2]$. The true parameters in the simulation are $\beta_1 = 0$ and $\beta_2 = 1$. This model is a simple probit model. As the model can easily be estimated by the classical maximum likelihood method, the probit MLE provides the ideal estimator in comparison with estimators derived from methods of simulated moments. Throught the experiment, simulated moments are derived from importance sampling technique. Random numbers for the construction of simulated moments are drawn from the standard exponential distribution. The exponential distribution is desirable for models with normal disturbances because it satisfies easily the domination condition in Assumption 3.

Table 1 reports the finite sample performance of the probit MLE, the maximum simulated likelihood estimates (MSL) and the maximum simulated likelihood estimates based on normalized simulated moments (Normalized MSL). Data of various sample sizes, namely 25, 50, 100 and 200, are considered. Conjugate gradient routine in Press et al [1986], pp.305-306, is used to implement the maximum simulated likelihood estimation. Table 1 reports summary statistics of mean (Mean), standard deviation (SD) and root mean square error (RMSE) of estimates based on 200 repetitions for each case. The first column block reports the probit MLE. The second and third column blocks report the MSL estimates with one Monte Carlo draw (r=1) and two Monte Carlo draws (r=2) for each observation respectively. The estimates of the Normalized MSL reported in the last column block are based on r=1. Comparing the estimates with the true parameters, all the estimates of the intercept term β_1 have rather small biases. The biases of the estimates of β_2 are more severe but are decreasing as sample size increases. The standard errors of all the estimates are also decreasing as sample size increases. For the estimation of the intercept term, there are not much substantial differences across the different estimators. However, except for the very small sample of size 25, there is a ranking across the different estimators of the regression coefficient β_2. In most cases, the probit MLE have the smallest biases and variances. The second best estimation procedure is the Normalized MSL and the worse one is the MSL with r=1. As expected, MSL estimates with more random draws (r=2) are better than MSL estimates with one random draw (r=1). The interesting observation from this simulation is that the Normalized SML performs even better than the MSL procedure with r=2. It performs also favourably in comparison with the probit MLE estimates. As a measure of relative efficiency (RE) of a simulation estimator in comparison with the probit MLE, we can consider the ratio of the RMSE of a probit estimator over the RMSE of the corresponding simulation estimator. For the sample size 100, the RE of the MSL (r=1) of β_2 is about 69%. It increases to 75% for the MSL with r=2. For the sample size 200, the efficiency of the MSL (r=1) is only 62% and the efficiency of the MSL (r=2) is 72%. With both sample sizes, the efficiency of the Normalized MSL is 97%. It is interesting to note that for very small sample of size 25 the several simulation estimators may perform well as the probit MLE.

Table 2 reports simulation results on estimates derived from various methods of simulated moments. For the simple binary choice model, natural instrumental variables are $w_i = (1, x_i)$. With these instruments, McFadden's estimator from (2.8) can be derived from the solution of the equations:

$$\sum_{i=1}^{n} w_i'(d_i - f_1(\theta, x_i)) = 0. \tag{7.2}$$

Estimators for the modified method of simulated moments can be derived with the proper simulated moment of $P(1|\theta, x)$ replacing f_1 in (7.2). Table 2 reports the estimates from McFadden's method of simulated moments (McFadden's SM), the modified method with dependent simulated moments (SM with dep. moments) as well as the method with independent simulated moments (SM with indep. moments). Probit MLE are again the ideal estimates for comparison. For McFadden's method, we have experimented with simulated moments based on different numbers of draws, namely r=1, 10 and 50. As expected, as sample size increases from 100 to 200, biases and variances of the various SM estimators decrease. As the number of draws r increases, the variances of McFadden's SM estimator as well as their corresponding RMSE decrease. With one random draw for each observation, i.e., r=1, the relative efficiency of McFadden's SM estimator for β_2 with sample size 100 is only 21% and it increases to 24% with sample size 200. The relative efficiency of the SM estimator of β_2 based on dependent moment is 72% with sample size 100 and it is 63% with sample size 200. Relative efficiency of McFadden's SM estimator increases as the number of draws r increases. The SM estimator based on dependent moments with $r = 1$ is still more efficient than McFadden's SM estimator with $r = 10$. On the other hand with $r = 50$, McFadden's SM estimator becomes much more efficient. The SM estimtor based on independent moments with $r = N$ is compared favourably with the probit MLE. Comparing the estimates across Tables 1 and 2, the MSL with r=1 and the SM with dependent moments are compatible. The Normalized MSL is compatible with the SM estimator with independent moments.

In Table 3, we report some Monte Carlo results on the estimation of a dynamic discrete choice panel data model. The model is specified as follows:

$$y_{it}^* = \beta x_{it} + \lambda d_{i,t-1} + u_{it} \tag{7.3}$$

and

$$u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad i = 1, \cdots, n; \ t = 1, \cdots, T \tag{7.4}$$

where $d_{i,t}$ is the observable dichotomous indicator of the latent variable y^*_{it}. The disturbances ϵ_{it} are i.i.d normal $N(0, \sigma^2)$. In order to normalize the variance of u to be unity, σ^2 is set to equal to $1 - \rho^2$. To capture possible correlation of the regressor x_{it} over time, x_{it} is specified to possess an error component structure:

$$x_{it} = \frac{1}{\sqrt{2}} z_{it} + \sqrt{6} w_i \tag{7.5}$$

where z_{it} are i.i.d. truncated normal $N(0,\ 1)$ variables with support $[-2,\ 2]$ and w_i are independent uniform variates with support on $[-\frac{1}{2},\ \frac{1}{2}]$. The variance of x is about 1 and its correlation coefficient over time is about 0.5. To start the dynamic process, the initial condition is specified as $d_{i,0} = 0$ for all i. The true parameters in the model are $\beta = 1.0$, $\lambda = 0.2$ and $\rho = 0.4$. With T time periods, the product of T independent univariate standard exponential densities is chosen as the T-dimensional multivariate importance sampling density for this model. Table 3 reports MSL estimates with dependent moments and MSL estimates with independent moments. The estimates tend to underestimate the regression coefficient β and the serial correlation coefficient ρ but tend to overestimate the dynamic coefficient λ of this model. As sample size increases, the biases of the MSL tend to decrease and their variances and RMSE decrease monotonically. With adding up property, the Normalized MSL estimates have smaller biases as well as smaller variances than the MSL estimates with the same sample size. The biases of the Normalized MSL for the regression coefficient β is remarkly small. The MSL with independent moments do not necessarily have smaller biases as compared with the MSL with dependent moments, however their variances are smaller. The MSL with independent moments have smaller RMSE than the MSL with dependent moments for the data with the same sample size. The RMSE of the MSL with independent moments are also smaller than the RMSE of the Normalized MSL except for the estimation of ρ. The above estimates are derived for panel data with 4 time periods. With longer panel T=6, the MSL estimates have smaller variances but larger biases than the corresponding MSL estimates with shorter panel T=4.

For panel data with T periods and N cross sectional units, the total number of sample points is NT. We might expect that for longer panels the variances of the estimates would decrease as the sample sizes have increased. However, the number of different choice patterns in longer panels is much larger than the number of choice patterns in shorter panels. For T=4, the total number of choice patterns for the binary choice dynamic model is 16 but the number increases to 64 for T=6. With many choice alternatives, the choice probabilities are in general more difficult to be estimated. These features might explain the above Monte Carlo findings. Anyhow, in terms of RMSE, longer panel data is still preferable to short panel data.

The above Monte Carlo simulations are performed by a Cray-XMP machine in Minnesota Supercomputer Institute. While different machines have different functions and computing speed, it may still be worthy to report the CPU times in computing some of the estimates. For the estimation of the panel data model with N=100 and T=4, it took 204.44 seconds of CPU to compute the 200 MSL estimates with dependent moments reported in Table 3; 398.83 seconds to compute the MSL estimates with independent moments and 2441.73 seconds to compute the Normalized MSL estimates [6]. These computing times do not include time cost to draw random numbers for the construction of simulated moments. The computational cost of the MSL with independent moments is about 2 times more in this case because it involves the computation of N^2 different numbers of importance sampling density functions instead of just N different importance densities for the case with dependent moments. The CPU time cost of the 200 Normalized MSL estimates is almost 12 times more expensive than the computation of the MSL estimates with dependent moments. For the MSL estimates with dependent moments, only a single choice probability needs to be computed for each cross sectional unit while for the Normalized MSL, all 2^T choice probabilities need to be computed for normalization purpose as adding up property does not hold for the

[6] Starting initial estimates for iteration are set at (0, 0, 0). We have experimented with initial estimates set at the true parameter vector. Convergence with the latter estimates are only sightly faster.

original simulators. With 200 cross sectional units, the CPU time for the computation of 200 MSL estimates with dependent moments increases to 646.01 seconds which is about 3 times more than the computation of the MSL with 100 units. McFadden's methods of simulated moments use independent moments and therefore are similar in the MSL estimation with independent moments in this aspect. The number of random draws r in McFadden's method does not need to tie with sample size. With M=50, the computation time of 200 MSL estimates with independent moments for the model with N=100 and T=4 is 220.21 seconds and it reduces to 148.30 seconds when M=30. In this example, the computation cost with M=50 is larger than the cost of the MSL with dependent moments but its cost is 27.5% less when M=30 [7].

[7] We note that the computational cost of McFadden's method of simulated moments depends on the number of moments used in the estimation. For the above comparison, it corresponds essentially to the use of a single moment equation.

Table 1.

Binary Choice Model: Maximum Simulated Likelihood and Probit Estimates

True parameters: $\beta_1 = 0, \quad \beta_2 = 1$

N		Probit MLE			MSL (r=1)				MSL (r=2)				Normalized MSL			
		Mean	SD	RMSE	Mean	SD	RMSE	RE	Mean	SD	RMSE	RE	Mean	SD	RMSE	RE
25	β_1	0.0134	.3098	.3038	0.0072	.2970	.2911	1.04	0.0171	.3281	.3219	.94	0.0138	.3188	.3127	.97
	β_2	1.2449	.6956	.7242	1.1809	.7342	.7418	.98	1.2230	.7211	.7409	.98	1.2519	.6780	.7105	1.02
50	β_1	-0.0020	.2276	.2253	-0.0089	.2402	.2380	.95	-0.0003	.2313	.2290	.98	0.0039	.2263	.2241	1.01
	β_2	1.0928	.3889	.3960	1.1341	.5609	.5712	.69	1.0844	.4525	.4558	.87	1.0992	.3967	.4050	.98
100	β_1	-0.0150	.1400	.1401	-0.0144	.1400	.1400	1.00	-0.0136	.1419	.1418	.99	-0.0142	.1401	.1401	1.00
	β_2	1.0338	.2134	.2150	1.0443	.3099	.3115	.69	1.0470	.2846	.2870	.75	1.0371	.2189	.2209	.97
200	β_1	-0.0001	.0993	.0991	-0.0010	.1021	.1018	.97	0.0004	.1020	.1017	.97	0.0006	.0999	.0997	.99
	β_2	1.0069	.1469	.1467	1.0254	.2356	.2364	.62	1.0186	.2035	.2038	.72	1.0080	.1509	.1507	.97

Table 2.

Binary Choice Model: Methods of Simulated Moments and Probit MLE

True parameters: $\beta_1 = 0, \quad \beta_2 = 1$

Method			Sample Size N=100				Sample Size N=200			
			Mean	SD	RMSE	RE	Mean	SD	RMSE	RE
Probit ML		β_1	-0.0150	.1400	.1401	1.00	-0.0001	.0993	.0991	1.00
		β_2	1.0338	.2134	.2150	1.00	1.0069	.1469	.1467	1.00
McFadden's SM	r=1	β_1	-0.0218	.2019	.2021	.69	0.0014	.1582	.1578	.63
		β_2	1.1565	1.0319	1.0386	.21	1.0917	.6177	.6229	.24
	r=10	β_1	-0.0116	.1831	.1826	.77	0.0120	.1352	.1354	.73
		β_2	1.0635	.4104	.4133	.52	1.0575	.3172	.3216	.46
	r=50	β_1	-0.0128	.1451	.1449	.97	0.0003	.1023	.1020	.97
		β_2	1.0377	.2367	.2385	.90	1.0078	.1586	.1584	.93
SM with indep. moments	r=N	β_1	-0.0163	.1393	.1396	1.00	-0.0010	.0990	.0988	1.00
		β_2	1.0345	.2207	.2223	.97	1.0061	.1487	.1485	.99
SM with dep. moments	r=1	β_1	0.0098	.1779	.1773	.79	0.0088	.1203	.1203	.82
		β_2	1.0339	.2980	.2984	.72	1.0248	.2319	.2326	.63

Table 3.

Panel Data Model and Maximum Simulated Likelihood Estimation

True parameters: $\beta = 1$, $\lambda = 0.2$, $\rho = 0.4$

Method	T	N	r	β			λ			ρ		
				Mean	SD	RMSE	Mean	SD	RMSE	Mean	SD	RMSE
MSL(dep. moments)	4	50	1	.9147	.1948	.2109	.2715	.2680	.2748	.2757	.2478	.2750
	4	100	1	.9436	.1521	.1615	.2460	.1955	.1999	.3138	.1853	.2035
	4	200	1	.9687	.1158	.1197	.2496	.1333	.1419	.3243	.1263	.1470
Normalized MSL	4	100	1	1.0006	.1185	.1179	.2229	.1645	.1653	.3549	.1469	.1530
MSL(indep. moments)	4	100	N	.9481	.0875	.1014	.2548	.1404	.1501	.2751	.1182	.1716
MSL(dep. moments)	6	100	1	.9009	.1124	.1494	.2463	.1435	.1501	.2591	.1307	.1917

8. Conclusion

In this article, we have considered a modified method of simulated moments for the estimation of discrete choice models. Similar to McFadden's method of simulated moments, excessive number of Monte Carlo draws are unnecessary for the simulation estimator of the parameters in the model to be consistent asymptotically normal. As in McFadden's simulation design, for each observation, only one or a few random numbers from a specific distribution will be drawn. It differs from McFadden's method only in the way of constructing simulated moments with the generated random numbers. As the random draws are exchangeable across observation, all of them can be used in the construction of simulated moments for each observation. Such simulated moments will involve summation of much more terms than McFadden's simulated moments. The computational cost will be more expensive but the simulation estimator of the parameters is asymptotically efficient relative to McFadden's estimator. This design can easily be adapted to other estimation methods without restricting our attention to the selection of moment equations. One of such other methods that have been considered in this article is a maximum simulated likelihood method. The maximum simulated likelihood estimator is shown to be consistent asymptotically normal. Adding up property of probability simulator plays an interesting role in the maximum simulated likelihood estimation. With adding up property, the maximum simulated likelihood estimator is asymptotically equivalent to the estimator derived from the modified method of simulated moments with optimum instruments suggested from the classical maximum likelihood method as if they were completely observable. The maximum simulated likelihood method may be of particular interest for panel data models where the total number of choice patterns over time can be large. These proposed estimation methods can also be regarded as generalizations of the simulation estimation methods in Pakes [1986] (Pakes and Pollard [1989]) to disaggregated data. It differs from the Monte Carlo integration approach in Lerman and Manski [1981] in two important aspects. As in McFadden's approach, our estimation method has emphasized on estimation of unknown parameters instead of choice probabilities and the total number of Monte Carlo draws does not need to be excessive large. As the number of random draws for each observation is fixed, the derived estimators are in general not statistically efficient. Statistical efficient estimators can be derived only at the expense of much more Monte Carlo draws. Our estimation methods will be useful if excessive number of Monte Carlo draws is impractical or undesirable.

In the article, we have considered only smooth simulators. Statistical analysis with such simulators is much simpler than nonsmooth simulators. However, with the recent development of empirical process theory in Pakes and Pollard [1989] for simulators and U-statistics empirical process theory in Nolan and Pollard [1987, 1988], we expect that our method could be generalized to cover nonsmooth estimators. Such possible generalization will be investigated in separate articles. We expect also that our methods can be generalized to the estimation of limited dependent variable models.

Appendix

Proposition 1. (A Uniform Law of Large Numbers for V-statistics)

Let $\{y_i\}$ be a random sample. Let $g(y_{i_1}, y_{i_2}, \theta)$ be a measurable function of (y_{i_1}, y_{i_2}) for each $\theta \in \Theta$, which is a compact subset of a Euclidean space, and is a continuous function of $\theta \in \Theta$ for each (y_{i_1}, y_{i_2}). Suppose that $E \sup_{\theta \in \Theta} |g(y_{i_1}, y_{i_2}, \theta)| < \infty$ for all $1 \leq i_1, i_2 \leq 2$. Then

$$\frac{1}{n^2} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} g(y_{i_1}, y_{i_2}, \theta) \xrightarrow{p} E(g(y_1, y_2, \theta))$$

uniformly in $\theta \in \Theta$.

PROOF: A law of large number for U-statistics is available in Serfling ([1980], p.206). It states that for any kernel h, if $E|h(y_1, y_2)| < \infty$, the associated U-statistic U_n with kernel h will converge in probability to $E(h(y_1, y_2))$. The V-statistic V_n with kernel h is closedly related to the U-statistic U_n. If $E|h(y_{i_1}, y_{i_2})| < \infty$ for $1 \leq i_1, i_2 \leq 2$, then $E|U_n - V_n| = O(\frac{1}{n})$ (see Serfling [1980], p.206). Hence V_n will converge in probability to $E(h(y_1, y_2))$ also. With this law of large numbers for V-statistics, the uniform convergence property in the proposition can be proved with exactly the same argument in Amemiya ([1985], Theorem 4.2.1).

Q.E.D.

Proposition 2. (A Central Limit Theorem for V-statistics)

Let $\{y_i\}$ be a random sample and $g(y_{i_1}, y_{i_2})$ be a measurable vector value function of (y_{i_1}, y_{i_2}) with zero mean. Suppose that $E|g(y_{i_1}, y_{i_2})|^2 < \infty$ for all $1 \leq i_1, i_2 \leq 2$. Then

$$\frac{1}{n\sqrt{n}} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} g(y_{i_1}, y_{i_2}) \xrightarrow{D} N(0, \Sigma)$$

where

$$\Sigma = E\{[E(g(y_1, y_2)|y_1) + E(g(y_2, y_1)|y_1)][E(g(y_1, y_2)|y_1) + E(g(y_2, y_1)|y_1)]'\}.$$

PROOF: This central limit theorem can be found in Serfling [1980], p.192 and p.206.

Q.E.D.

Proposition 3. *Suppose $\{v_j^{(i)}\}$ are independently and identically distributed sequences of random variables. $\{x_i\}$ is a sequence of i.i.d. random variables which are independent with $v_j^{(i)}$, for all i, j. Let $g(v, x, \theta)$ be a measurable function of (v, x) for each $\theta \in \Theta$, which is a compact subset of a Euclidean space, and is a continuous function of $\theta \in \Theta$ for each (v, x). Suppose that $E \sup_{\theta \in \Theta} |g(v, x, \theta)|^2 < \infty$, then*

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} g(v_j^{(i)}, x_i, \theta) \xrightarrow{p} E(g(v, x, \theta))$$

uniformly in $\theta \in \Theta$.

PROOF: For any square integrable function $h(v, x)$, by independence,

$$var(\frac{1}{n} \sum_{i=1}^{n} [\frac{1}{n} \sum_{j=1}^{n} h(v_j^{(i)}, x_i)]) = \frac{1}{n} var(\frac{1}{n} \sum_{j=1}^{n} h(v_j, x)).$$

Since $h(v, x)$ is square integrable, it follows that $var(E[h(v, x)|x])$ and $E\{var(h(v, x)|x)\}$ are all finite. Let $z_n = \frac{1}{n} \sum_{j=1}^{n} h(v_j, x)$. The variance of z_n is

$$\begin{aligned} var(z_n) &= E(\frac{1}{n} \sum_{j=1}^{n} h(v_j, x) - E(h(v, x)|x))^2 + E\left(E(h(v, x)|x) - E(h(v, x))\right)^2 \\ &= \frac{1}{n} E\{var(h(v, x)|x)\} + var(E[h(v, x)|x]). \end{aligned} \tag{A.1}$$

It follows that by Chebyshev's ineqality

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} h(v_j^{(i)}, x_i) \xrightarrow{p} E(h(v, x)).$$

With this law of large numbers, the uniform convergence property in this proposition follows from the argument in Amemiya ([1985], Theorem 4.2.1).

Q.E.D.

Proposition 4. *Suppose $\{v_j^{(i)}\}$ are independently and identically distributed sequences of random variables. $\{x_i\}$ is a sequence of i.i.d. random variables which are independent with $v_j^{(i)}$, for all i, j. Let $g(v, x)$ be a measurable function. Suppose that there exists a square integrable* *function $h(x)$ such that $|g(v,x)| \le h(x)$ for all v, then*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n} g(v_j^{(i)}, x_i) - E(g(v,x)) \xrightarrow{D} N(0, var(E[g(v,x)|x])).$$

PROOF: Denote $z_{ni} = \frac{1}{n}\sum_{j=1}^{n} g(v_j^{(i)}, x_i)$. Under our independency assumptions, $z_{n1}, \cdots, z_{nn}$ are mutually independent. $E(z_{ni}) = E(g(v,x)) = \mu$ is a constant. The variance of z_{ni}

$$var(z_{ni}) = \frac{1}{n}E\{var(g(v,x)|x)\} + var(E[g(v,x)|x])$$

is a constant independent of i, which converges to $var(E[g(v,x)|x])$. Since the sequence

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\frac{1}{n}\sum_{j=1}^{n} g(v_j^{(i)}, x_i) - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(z_{ni} - \mu)$$

is a double array sequence, a central limit theorem of double array will be needed. The central limit theorem of double array in Chung [1974] can be applied if Lindeberg condition is satisfied, i.e., for any $\epsilon > 0$,

$$E[I((z_{ni} - \mu)^2 > \epsilon n\bar{\sigma}_n^2)(z_{ni} - \mu)^2] \to 0$$

where $\bar{\sigma}_n^2$ is the variance of z_{ni}. Let $(\Omega \times X, \nu \times \xi)$ be the common probability space that (v,x) lives on. Denote $f_n(\omega, x) = (\frac{1}{n}\sum_{j=1}^{n} g(v_j(\omega), x) - \mu)^2$. It follows that

$$\begin{aligned} &E[I((z_{ni} - \mu)^2 > \epsilon n\bar{\sigma}_n^2)(z_{ni} - \mu)^2] \\ &= \int_{A_n} f_n(\omega, x)d\nu(\omega)d\xi(x) \end{aligned} \tag{A.2}$$

where $A_n = \{(\omega, x)|(\frac{1}{n}\sum_{j=1}^{n} g(v_j(\omega), x) - \mu)^2 > \epsilon n\bar{\sigma}_n^2\}$. Lindeberg condition will be satisfied if $(A.2)$ converges to zero. By Chebyshev's inequality

$$\begin{aligned} (\nu \times \xi)(A_n) &= P(|\frac{1}{n}\sum_{j=1}^{n} g(v_j, x) - \mu| > \sqrt{\epsilon n}\bar{\sigma}_n) \\ &\le \frac{1}{\epsilon n\bar{\sigma}_n^2}E(\frac{1}{n}\sum_{j=1}^{n} g(v_j, x) - \mu)^2 \\ &= \frac{1}{\epsilon n} \end{aligned}$$

which converges to zero. For each x, the strong law of large numbers for i.i.d. variables implies that $\frac{1}{n}\sum_{j=1}^{n} g(v_j, x) \stackrel{a.s.}{\rightarrow} E(g(v,x)|x)$. Hence $f_n(\omega, x) \rightarrow f(x) = (E(g(v,x)|x) - \mu)^2$, $a.e.[\nu \times \xi]$ on $\Omega \times X$. The following lemma completes the proof.

Q.E.D.

The following lemma generalizes the familiar result for an integrable function (see, e.g. Royden [1963], Chapter 4, Proposition 13) to a sequence of integrable functions.

Lemma 5. *Suppose that $f_n(x)$ is a sequence of nonnegative integrable functions with measure ν. If $\lim_{n\to\infty} f_n(x) = f(x)$ a.e.$[\nu]$, and there exists an integrable function $h(x)$ such that $|f_n(x)| \leq h(x)$ for all x, then for any $\epsilon > 0$, there exist $\delta > 0$ and n_0 such that for all measurable subset A with $\nu(A) < \delta$, $\int_A f_n(x)d\nu(x) < \epsilon$ for all $n \geq n_0$.*

PROOF: Suppose not, there exists an $\epsilon > 0$ such that for any $\delta > 0$ and n, one can find A and m_n with $\nu(A) < \delta$, but $\int_A f_{m_n}(x)d\nu(x) \geq \epsilon$. In particular, there exist A_n with $\nu(A_n) < \frac{1}{2^n}$ and $\int_{A_n} f_{m_n}(x)d\nu(x) \geq \epsilon$. Let $g_n(x) = f_{m_n}(x)I_{A_n}(x)$. The functions $g_n(x)$ converge weakly to 0 except x in the set $\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i$. Since

$$\nu(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i) \leq \nu(\bigcup_{i=n}^{\infty} A_i) \leq \sum_{i=n}^{\infty}\nu(A_i) \leq \sum_{i=n}^{\infty}\frac{1}{2^i}$$

converges to zero as n goes to ∞, $g_n(x) \rightarrow 0$, $a.e.$ Let $h_n(x) = f_{m_n}(x) - g_n(x) = f_{m_n}(x)[1 - I_{A_n}(x)]$. $\{h_n\}$ is a sequence of nonnegative functions. Furthermore, $h_n(x) \rightarrow f(x)$ $a.e.$ By Fatous's lemma and LDC theorem,

$$\begin{aligned}\int f(x)d\nu(x) &\leq \liminf_{n\to\infty}\int h_n(x)d\nu(x)\\ &\leq \lim_{n\to\infty}\int f_{m_n}(x)d\nu(x) - \liminf_{n\to\infty}\int g_n(x)d\nu(x)\\ &= \int f(x)d\nu(x) - \liminf_{n\to\infty}\int_{A_n} f_{m_n}(x)d\nu(x)\\ &\leq \int f(x)d\nu(x) - \epsilon\end{aligned}$$

which is a contradiction.

Q.E.D.

References

1. Amemiya, T. [1985], *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.

2. Chung, K.L. [1974], *A Course in Probability Theory*, 2nd ed. New York: Academic Press.

3. Geweke, J. [1989], "Bayesian Inference in Econometric Models with Monte Carlo integration", *Econometrica* 57: 1317-1339.

4. Hajivassiliou, V., and D. McFadden [1987], "The Debt Repayment Crises LDC's: Estimation by the Method of Simulated Moments", Yale University, Working Paper.

5. Ichimura, H. and L.F. Lee [1988], "Semiparametric Estimation of Multiple Index Models: Single Equation Estimation", manuscript, Department of Economics, University of Minnesota, forthcoming in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, edited by W.A. Barnett, J. Powell and G. Tauchen.

6. Keane, M.P. [1989], "A Computationally Practical Simulation Estimator for Panel Data, With Applications to Labor Supply and Real Wage Movement Over the Business Cycle", Discussion Paper 16, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis.

7. Lee, L.F. [1989], "Semiparametric Maximum Profile Likelihood Estimation of Polytomous and Sequential Choice Models", Discussion Paper No. 253, Center for Economic Research, Department of Economics, University of Minnesota, December 1989.

8. Lee, B.S. and B.F. Ingram [1989], "Simulation Estimation of Time Series Models", manuscript, Carlson School of Management, University of Minnesota.

9. 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, 305-319.

10. McFadden, D. [1989], "A Method of Simulated Moments for Estimation of Discrete Response

Models Without Numerical Integration", *Econometrica* 57: 995-1026.

11. Nolan, D. and D. Pollard [1987], "U-processes: Rates of Convergence", *The Annals of Statistics* 15: 780-799.

12. Nolan, D. and D. Pollard [1988], "Functional Limit Theorems for U-processes", *The Annals of Probability* 16: 1291-1298.

13. Pakes, A. [1986], "Patents as Options: 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. Press, W.H., B.P. Flannery, S.A. Teukolsky and W.T. Vetterling [1986], *Numerical Recipes*, Cambridge University Press, Cambridge, New York.

16. Royden, H.L. [1963], *Real Analysis*, MacMillan Company, New York.

17. Serfling, R.J. [1980], *Approximation Theroems of Mathematical Statistics*, Wiley, New York.

18. Stern, S. [1987], "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models", Manuscript, Department of Economics, University of Virginia.