Center for Research on Economic and Social Theory

and

Department of Economics

Working Paper Series

Asymptotic Bias in Maximum Simulated Likelihood Estimation of Discrete Choice Models

Lung-fei Lee

July, 1992
Number 93-03

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

Asymptotic Bias in Maximum Simulated Likelihood Estimation of Discrete Choice Models

by

Lung-fei Lee*

Department of Economics
The University of Michigan
Ann Arbor, MI 48109-1220

Tel: 313-764-2363

July, 1992

* I appreciate having financial support from NSF under grant SES-9296071 for my research.

ABSTRACT

In this article, we have investigated a bias in an asymptotic expansion of the maximum simulated likelihood estimator introduced by Lerman and Manski for the estimation of discrete choice models. This bias occurs due to the nonlinearity of the derivatives of the log likelihood function and the statistically independent simulation errors of the choice probabilities across observations. This bias can be the dominated bias in an asymptotic expansion of the maximum simulated likelihood estimator when the number of simulated random variables per observation does not increase as fast as or faster than the sample size. The properly normalized maximum simulated likelihood estimator has even an asymptotic bias in its limiting distribution if the number of simulated random variables increase only as fast as the square root of the sample size. A bias-adjustment is introduced which can reduce the bias. Some Monte Carlo experiments have demonstrated the usefulness of the bias-adjustment procedure.

JEL classification number: 211
Key Words and Phrases:
Maximum Simulated Likelihood Estimation, Asymptotic Expansion, Asymptotic Bias, Rate of Stochastic Convergence, Stochastic Order, Bias-Adjustment.

Asymptotic Bias in Maximum Simulated Likelihood Estimation of Discrete Choice Models

by Lung-fei Lee

Department of Economics, The University of Michigan, Ann Arbor

July 1992

1. Introduction

In this article, we will study asymptotic biases of estimators which are derived from some maximum simulated likelihood methods for the discrete choice model. The estimation methods by simulation of moment equations or likelihood functions have been introduced by Lerman and Manski [8], Pakes [10], McFadden [9], and Pakes and Pollard [11]. The estimation method in Lerman and Manski [8] and Pakes [10] is a maximum simulated likelihood method. For any possible parameter value in the model, the response probabilities are approximated by simulated probabilities and the likelihood function is constructed with the response probabilities replaced by the simulated probabilities. The maximum simulated likelihood estimator is derived by maximizing the simulated likelihood function. Specifically, the simulated probabilities are generated independently for each individual decision maker in the sample. Subsequently, Lee [6] considered the case where the simulated probabilities need not to be statistically independent. The first case is termed the maximum simluated likelihood estimation with independently simulated moments to distinguish it from the latter case which is the maximum simulated likelihood estimation with dependently simulated moments. The asymptotic properties of consistency and asymptotic distributions of the method of simulated moments have been studied in McFadden [9] and Pakes and Pollard [11]. The asymptotic efficiency of the maximum simulated likelihood methods have been studied in Lee [6].

In this article, we will investigate some particular nonlinear features of the simulated likelihood estimation. We will investigate the possible asymptotic biases of the maximum simulated likelihood estimators. When the response probabilities are simulated independently, even though that the maximum simulated likelihood estimator can be consistent, asymptotic bias of the estimator may appear when the rate of convergence of the simulated probabilities is relatively low as compared with the sample size. Under some situations, the asymptotic distribution may be degenerated. Unless the number of simulated random variables increases as fast as the sample size, the maximum simulated likelihood estimator with independently simulated moments possesses a bias which has an order larger than the bias of the classical maximum likelihood estimator in an asymptotic expansion. The existence of such a bias is due to a particular feature of the estimator in that a second order term, consisting of sample average of squares of simulation errors, in the Taylor series expansion of the derivatives of the simulated log likelihood function with respect to the response probabilities can denominate a first order error term. We discuss bias adjustment procedures which can reduce or eliminate such a bias.

This article is organized as follows. In Section 2, the maximum simulated likelihood methods for the estimation of the discrete choice model are described. Regularity conditions that will be useful for our analysis are listed. In Section 3, we discuss the asymptotic biases in an asymptotic expansion of the simulated likelihood estimator with independently simulated moments. The order of the bias is derived and its practical implication in simulation estimation will be discussed. The features of bias will be compared across the classical maximum likelihood estimation and the simulated likelihood estimation with dependently simulated moments. Section 4 suggests some bias-adjustment procedures that will be useful to reduce or eliminate the leading bias due to simulation error. The practical implication of bias-adjustment will be discussed. Some Monte Carlo evidence on the performance of a bias-adjustment procedure is reported in Section 5. Our conclusions are in Section 6. All the proofs of the results are collected in the Appendix B. Some useful propositions that have applications in our analysis and are of interest in their own, are provided separately in the Appendix A.

2. Maximum Simulated Likelihood Estimation of The Discrete Choice Model

Consider a utility maximization model of discrete responses. Let $C = \{1, \cdots, L\}$ be a set of mutually exclusive and exhaustive alternatives in the discrete response model. For each alternative $l \in C$, the associated value of the alternative l is $U_l = x_l \alpha$, where α is a vector of individual weights distributed randomly in the population and x_l is a vector of measured attributes of alternative l. The 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 $x_1, \ldots, x_L$. Define

$$V_l = (U_1 - U_l, \cdots, U_{l-1} - U_l, U_{l+1} - U_l, \cdots, U_L - u_l), \qquad l = 1, \ldots, L \tag{2.1}$$

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 the alternative l is

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

Let d_{li} denote a response indicator for individual i, equal to one when the observed response is the alternative l, zero otherwise. With a sample of size n of independent observations, the log likelihood function for the discrete choice model is

$$\mathcal{L}_c(\theta) = \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \ln P(l|\theta, x_i). \tag{2.2}$$

A simulated likelihood approach will replace the hard-to-compute probabilities $P(l|\theta, x)$, $l = 1, \ldots, L$, by consistent simulators $f_{r,l}(\theta, x)$. Broad classes of simulators have been introduced in McFadden [9], Hajivassiliou and McFadden [4], Stern [13], Borsch-Supan and Hajivassiliou [2] and others. Consider, for example, the smooth simulators in McFadden [9] based on importance simulation technique. The smooth simulator based on recursive conditioning in Borsch-Supan and Hajivassiliou [2] is in fact also an importance simulator. Let $\gamma(v)$ be a density chosen for the simulation that has the negative orthant as its support. Let

$$h_l(v, x, \theta) = \frac{g_l(v|\theta, x)}{\gamma(v)}.$$

The response probability can be rewritten as

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

Averaging $h_l(v, x, \theta)$ over one or more Monte Carlo draws from $\gamma(v)$ gives a smooth unbiased estimator of $P(l|\theta, x)$. Suppose there are r Monte Carlo draws from $\gamma(v)$ for each sample observation. Let $v_j^{(i)}, j = 1, \ldots, r$ be the draws for observation i. Define

$$f_{r,l}(\theta, x_i) = \frac{1}{r} \sum_{j=1}^{r} h_l(v_j^{(i)}, x_i, \theta). \tag{2.3}$$

Conditional on x_i, $E(f_{r,l}(\theta, x_i)|x_i) = \int h_l(v, x_i, \theta)\gamma(v) dv = P(l|\theta, x_i)$ and hence $f_{r,l}(\theta, x_i)$ is a conditionally unbiased simulator (simulated moment). When r goes to infinity as n goes to infinity, $f_{r,l}(\theta, x_i)$ is also a consistent estimator of $P(l|x_i, \theta)$. By replacing the response probabilities with simulated moments, we are working with a simulated likelihood function, which is a pseudo-likelihood function. The log simulated likelihood function with independently simulated moments is

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

Let $\hat{\theta}_L$ denote the maximum simulated likelihood estimator derived from the maximization of $\mathcal{L}(\theta)$ in (2.4). For our analysis, the following regularity conditions are assumed for our model:

Assumption 1:

1. The sample observations (d_i, x_i) where $d_i = (d_{1i}, \ldots, d_{Li})$, $i = 1, \cdots, n$, are i.i.d.
2. The parameter space Θ is a compact convex subset of a K-dimensional Euclidean space and the true parameter vector θ_o is in the interior of Θ.
3. The support X of x is a compact set.
4. The choice probabilities $P(l|\theta, x)$ are continuous in $(\theta, x) \in \Theta \times X$ and are positive for each $(x, \theta) \in X \times \Theta$.
5. The choice probabilities $P(l|\theta, x)$, $l = 1, \cdots, L$, are continuously differentiable in θ up to the fourth order.
6. The θ_o is the unique minimizer of the function $E\{\sum_{l=1}^{L} P(l|\theta_o, x)\ln P(l|\theta, x)\}$.
7. 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 2:

1. The random vector v is simulated independently of d from a common conditional density function, conditional on x.
2. The simulated moment function $h(v, x, \theta)$ is continuously differentiable in θ up to the fourth order. The absolute values of each component of $h(v, x, \theta)$ and its first four order derivatives with respect to θ are dominated by a square integrable function $H(v)$.
3. $h(v, x, \theta)$ is a conditionally unbiased estimator of the vector of choice probabilities $P(\theta, x)$ conditional on x, for each $\theta \in \Theta$.
4. The conditional first four order moments of the function $h(v, x, \theta_o)$ and its first order derivative with respect to θ conditional on x exist.
5. The first ten order moments of $h(v, x, \theta_o)$ and its first three order derivatives with respect to θ at θ_o exist.

Assumption 3:

The number of random draws r for each individual goes to infinity as n goes to infinity.

These regularity conditions contain mostly familiar regularity conditions on the choice probabilities of the discrete choice model. Since our asymptotic expansions of the estimators involve higher order derivatives of the log simulated likelihood function and simulation errors up to certain high orders, the differentiability conditions and the moment conditions in Assumptions 1 and 2 are required to guarantee that those terms are well defined. Some of the domination by integrable function assumptions are used for the application of some uniform law of large number for the convergence of the simulated probabilities and the log likelihood function to their relevant limits.

The consistency of the simulated likelihood estimator $\hat{\theta}_L$ depend on asymptotic properties of the simulated moments. As r goes to infinity, the simulated moments are consistent estimates of the corresponding response probabilities. Since the support X of x and the parameter space Θ of θ are compact and $f_{r,l}(\theta, x_i)$ is dominated by an integrable function in Assumption 2.2, the convergence is also uniform in probability (see, e.g. Amemiya [1]), i.e.,

$$\sup_{\Theta \times X} \|f_{r,l}(\theta, x) - P(l|\theta, x)\| \xrightarrow{p} 0, \tag{2.5}$$

for all $l = 1, \ldots, L$. Assumptions 1.2-1.4 guarantee that the choice probabilities $P(l|\theta, x)$, $l = 1, \ldots, L$, are strictly bounded away from zero on $\Theta \times X$. Since $P(\theta, x)$ is bounded away from zero on $\Theta \times X$, it follows from (2.5) that

$$\sup_{\Theta} \left| \frac{1}{n}\mathcal{L}(\theta) - \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\ln P(l|\theta, x_i)\right| \xrightarrow{p} 0. \tag{2.6}$$

The uniform law of large numbers in Amemiya (1985) implies that

$$\sup_{\Theta} \left| \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\ln P(l|\theta, x_i) - E\left(\sum_{l=1}^{L} P(l|\theta_o, x)\ln P(l|\theta, x)\right)\right| \xrightarrow{p} 0. \tag{2.7}$$

Since $\frac{1}{n}\mathcal{L}(\theta)$ converges in probability uniformly to the limit function $E(\sum_{l=1}^{L} P(l|\theta_o, x)\ln P(\theta, x))$ on Θ and θ_o is the unique minimizer of this limit function by Assumption 1.6, $\hat{\theta}_L$ is consistent. For consistency, it is sufficient to have the random draws r go to infinity as n goes to infinity. Any particular growth rate of r related to n is not required. However, for asymptotic efficiency, r has to increase at a rate faster than $n^{1/2}$ (Lee [6]). In the subsequent sections, we will investigate the consequence that may occur when r increases at a rate equal to or lower than $n^{1/2}$. We will see that asymptotic bias may occur in such situations. We will show that when r increases slower than n there exists a dominated bias of order higher than the order of bias in an asymptotic expansion of the classical maximum likelihood estimator. Bias-adjustment procedures will be suggested to eliminate or reduce such a bias.

It is worthwhile to compare the differences between the simulated likelihood estimation with independently simulated moments with the estimation with dependently simulated moments. For the dependent moments case, a total r number of random variables v_j, $j = 1, \cdots, r$ will be drawn independently of d and x for the construction of the simulated probabilities:

$$\bar{f}_{r,l}(\theta, x_i) = \frac{1}{r}\sum_{j=1}^{r} h_l(v_j, x_i, \theta), \qquad l = 1, \ldots, L, \tag{2.8}$$

for all i, $i = 1, \ldots, n$. The log simulated likelihood function with dependently simulated moments is

$$\bar{\mathcal{L}}(\theta) = \sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\ln \bar{f}_{r,l}(\theta, x_i). \tag{2.9}$$

The maximum simulated likelihood estimator of θ from (2.9) will be denoted by $\hat{\theta}_D$.

3. Biases in Maximum Simulated Likelihood Estimators

The maximum simulated likelihood estimator $\hat{\theta}_L$ with independently simulated moments from (2.4) satisfies the first order condition:

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

The asymptotic distribution of this estimator can be derived by a proper Taylor series expansion. By the mean-value theorem,

$$\sqrt{n}(\hat{\theta}_L - \theta_o) = -\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln f_{r,l}(\theta_o, x_i)}{\partial\theta}, \quad (3.1)$$

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

$$\sup_{\Theta\times X}\left\|\frac{\partial f_{r,l}(\theta, x)}{\partial\theta} - \frac{\partial P(l|\theta, x)}{\partial\theta}\right\| \xrightarrow{p} 0,$$

and

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

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

$$\frac{\partial \ln f_{r,l}(\theta, x)}{\partial\theta} = \frac{1}{f_{r,l}(\theta, x)}\frac{\partial f_{r,l}(\theta, x)}{\partial\theta} \xrightarrow{p} \frac{\partial \ln P(l|\theta, x)}{\partial\theta}$$

and

$$\frac{\partial^2 \ln f_{r,l}(\theta, x)}{\partial\theta\partial\theta'} = \frac{1}{f_{r,l}(\theta, x)}\frac{\partial^2 f_{r,l}(\theta, x)}{\partial\theta\partial\theta'} - \frac{1}{f_{r,l}^2(\theta, x)}\frac{\partial f_{r,l}(\theta, x)}{\partial\theta}\frac{\partial f_{r,l}(\theta, x)}{\partial\theta'} \xrightarrow{p} \frac{\partial^2 \ln P(l|\theta, x)}{\partial\theta\partial\theta'}$$

uniformly in $(\theta, x) \in \Theta \times X$. Since $\bar{\theta}$ is a consistent estimate of θ_o, it follows that

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

Therefore,

$$-\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right\}^{-1} = E\left[\sum_{l=1}^{L} P(l|\theta_o, x)\frac{\partial \ln P(l|\theta_o, x)}{\partial\theta}\frac{\partial P(l|\theta_o, x)}{\partial\theta'}\right]^{-1} + o_P(1). \quad (3.2)$$

It remains to consider the asymptotic distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln f_{r,l}(\theta_o, x_i)}{\partial\theta}$. It is useful to expand this function of $f_{r,l}(\theta_o, x_i)$ and $\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial\theta}$ at $P(l|\theta_o, x_i)$ and $\frac{\partial P(l|\theta_o, x_i)}{\partial\theta}$. By Lemma A.1 of the Appendix A with $m = 2$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln f_{r,l}(\theta_o, x_i)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + L_n + Q_n + C_n^{(R)}, \quad (3.3)$$

where

$$L_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{1}{P(l|\theta_o, x_i)}\left\{\left[\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial\theta}\right] - \frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta}[f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]\right\}$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{1}{P(l|\theta_o, x_i)}\left[\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial\theta} - \frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta}f_{r,l}(\theta_o, x_i)\right], \quad (3.4)$$

$$Q_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\left\{-\frac{1}{P^2(l|\theta_o, x_i)}\left[\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial\theta}\right][f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]\right.$$
$$\left. + \frac{1}{P^2(l|\theta_o, x_i)}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta}[f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^2\right\}, \quad (3.5)$$

and

$$C_n^{(R)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{1}{f_{r,l}(\theta_o, x_i)P^2(l|\theta_o, x_i)}\left\{\left[\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial\theta}\right][f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^2\right.$$
$$\left. - \frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta}[f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^3\right\}. \quad (3.6)$$

The L_n is a first order term involving the errors of the simulated response probabilities, namely, $f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)$, and the errors of their derivatives. The Q_n is a second order term involving the squared errors of the simulated probabilities and their derivatives. The $C_n^{(R)}$ is a remainder term of errors of high orders. In the following paragraphs, we will investigate the stochastic orders of L_n, Q_n and $C_n^{(R)}$. All the proofs can be found in the Appendix B.

Lemma 1: *Under our assumptions, L_n has zero mean but*

$$E(Q_n) = n^{1/2}r^{-1}\bar{\mu}, \quad (3.7)$$

where

$$\bar{\mu} = \sum_{l=1}^{L} E\left\{\frac{1}{P(l|\theta_o, x)}\left[\frac{\partial \ln P(l|\theta_o, x)}{\partial\theta}\text{var}(h_l(v, x)|x) - \text{cov}(h_l(v, x), \frac{\partial h_l(v, x)}{\partial\theta}|x)\right]\right\}. \quad (3.8)$$

Furthermore, $L_n = O_P(r^{-1/2})$, $Q_n - E(Q_n) = O_P(r^{-1})$, and $C_n^{(R)} = O_P(n^{1/2}r^{-3/2})$.

We see from this lemma that the second order term Q_n has a stochastic order $O_P(n^{1/2}r^{-1})$, which can be larger than the stochastic order $O_P(r^{-1/2})$ of the first order error term L_n if r does not increase as fast as the sample size n. This irregularity occurs because of the nonlinearity of the log simulated likelihood function and the simulated probabilities are independent across individuals. The errors of simulated probabilities do not appear additively in the first order condition of maximization as in the method of simulated moments of McFadden [9]. The nonlinearlity of the log likelihood function causes the presence of the quadratic term Q_n. The independence of the simulated probabilities across individuals causes the sample average in L_n to go to zero at the fast rate $r^{-1/2}$ instead of the rate $n^{1/2}r^{-1}$. This irregularity can cause that the limiting distribution of $\sqrt{n}(\hat{\theta}_L - \theta_o)$ does not center properly or its limiting distribution diverges when r does not increase faster than the square root of the sample size n. Indeed, these happen in the following theorem.

Theorem 2: *Under our assumptions,*

(i) if $\lim_{n\to\infty} n^{1/2}r^{-1} = 0$,

$$\sqrt{n}(\hat{\theta}_L - \theta_o) \xrightarrow{D} N(0, \Omega), \quad (3.9)$$

(ii) if $\lim_{n\to\infty} n^{1/2}r^{-1} = \lambda$ a finite positive constant,

$$\sqrt{n}(\hat{\theta}_L - \theta_o) \xrightarrow{D} N(\lambda\Omega\bar{\mu}, \Omega), \quad (3.10)$$

and

(iii) if $\lim_{n\to\infty} n^{1/2}r^{-1} = \infty$,

$$r(\hat{\theta}_L - \theta_o) \xrightarrow{p} \Omega\bar{\mu}, \tag{3.11}$$

where $\bar{\mu}$ *is defined in Lemma 1 and*

$$\Omega = \left\{ E\left[\sum_{l=1}^{L} P(l|\theta_o, x) \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta'} \right] \right\}^{-1}. \tag{3.12}$$

When r increases at a rate as fast as $n^{1/2}$, the limiting distribution of $\sqrt{n}(\hat{\theta}_L - \theta_o)$ is asymptotically normal but its does not properly locate at zero. An asymptotic bias exists which is proportional to $\Omega\bar{\mu}$. From (3.8), we see that $\bar{\mu}$ is in general nonzero. It can be a small constant vector if the conditional variance of $h_l(v,x)$ and the conditional covariance of $h_l(v,x)$ and $\frac{\partial h_l(v,x)}{\partial \theta}$ conditional on x are small for all x. The Monte Carlo experiments in Borsch-Supan and Haijivassiliou [2] provide some evidence that their smooth importance simulator based on recursive conditioning may have small variance. However, even though $\Omega\bar{\mu}$ can be small, the limiting distribution $\sqrt{n}(\hat{\theta}_L - \theta_o)$ will diverge when the increase of r is slower than $n^{1/2}$. When the increase of r is slower than $n^{1/2}$, the limiting distribution $\sqrt{n}(\hat{\theta}_L - \theta_o)$ diverges and the maximum simulated likelihood estimator $\hat{\theta}_L$ converges to θ_o at the slower rate r. The slow increasing r causes not only the simulated likelihood estimator to converge in probability to θ_o at a slow rate than $O_P(n^{-1/2})$ but also a degenerate distribution after proper normalization. Only when r increases faster than $n^{1/2}$, the limiting distribution of $\sqrt{n}(\hat{\theta}_L - \theta_o)$ is properly behaved and the maximum simulated likelihood estimator is asymptotically efficient.

However, even though the maximum simulated likelihood estimator $\hat{\theta}_L$ is asymptotically efficient when r increases faster than $n^{1/2}$, the bias in Q_n due to simulation may cause a bias in the asymptotic expansion of $\hat{\theta}_L - \theta_o$. Such a bias may dominate other biases in the estimator when r does not increase fast enough than the sample size. It is known that in a higher order asymptotic expansion, the bias of the classical maximum likelihood estimator may exist with an order of $O(n^{-1})$ for a sample of size n (see, e.g., Cox and Hinkley [3], p.310). Thus for the maximum likelihood estimation of the discrete response model with the exact likelihood function $\mathcal{L}_c$, the largest bias of the classical likelihood estimator is expected to have the order $O(n^{-1})$. For the simulated likelihood estimator, we expect such a bias order may exist in addition to the bias due to the simulation. This is shown in the theorem 4.

To get a better approximation, it is useful to increase the expansion in (3.3) with additional terms:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln f_{r,l}(\theta_o, x_i)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} + L_n + Q_n + C_n + D_n + R_n, \tag{3.13}$$

where L_n and Q_n are the same as in (3.4) and (3.5); $C_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} c_{ri}$; and $D_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_{ri}$ where

$$\begin{aligned} c_{ri} = \sum_{l=1}^{L} d_{li} \frac{1}{P^3(l|\theta_o, x_i)} \Big\{ & [\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial \theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial \theta}][f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^2 \\ & - \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} [f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^3 \Big\}, \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} d_{ri} = \sum_{l=1}^{L} d_{li} \frac{1}{P^4(l|\theta_o, x_i)} \Big\{ & \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} [f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^4 \\ & - [\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial \theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial \theta}][f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^3 \Big\}; \end{aligned} \tag{3.15}$$

and

$$\begin{aligned} R_n = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{1}{f_{r,l}(\theta_o, x_i) P^4(l|\theta_o, x_i)} \Big\{ & \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} [f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^5 \\ & - [\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial \theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial \theta}][f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^4 \Big\} \end{aligned} \tag{3.16}$$

is the remainder term. The stochastic orders of these higher order terms are reported in Lemma 3.

Lemma 3: *Under our assumptions,* $E(C_n) = n^{1/2}r^{-2}\bar{c}$ *and* $E(D_n) = n^{1/2}r^{-2}\bar{d}(r)$ *where*

$$\begin{aligned} \bar{c} = E\Big(\sum_{l=1}^{L} d_{li} \frac{1}{P^3(l|\theta_o, x_i)} \Big\{ & E([\frac{\partial h_l(v, x, \theta_o)}{\partial \theta} - \frac{\partial P(l|\theta_o, x)}{\partial \theta}][h_l(v, x) - P(l|\theta_o, x)]^2 | x) \\ & - \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} E([h_l(v, x, \theta_o) - P(l|\theta_o, x)]^3 | x) \Big\} \Big). \end{aligned} \tag{3.17}$$

and $\lim_{r\to\infty} \bar{d}(r) = \bar{d}$ *with*

$$\bar{d} = 3E\Big(\sum_{l=1}^{L} \frac{d_l}{P^4(l|\theta_o, x)} \{ \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta} \mathrm{var}^2(h_l(v, x)|x) - \mathrm{var}(h_l(v, x)|x) \mathrm{cov}(\frac{\partial h_l(v, x)}{\partial \theta}, h_l(v, x)|x) \} \Big). \tag{3.18}$$

Furthermore, $C_n - E(C_n) = O_P(r^{-3/2})$, $D_n - E(D_n) = O_P(r^{-2})$ *and* $R_n = O_P(n^{1/2}r^{-5/2})$.

The terms C_n and D_n have the order $O_P(n^{1/2}r^{-2})$, which are smaller than the order $O_P(n^{1/2}r^{-1})$ of Q_n. While the mean of Q_n is $O(n^{1/2}r^{-1})$, both the means of C_n and D_n are $O(n^{1/2}r^{-2})$. The means of C_n and D_n are also due to simulation but they have smaller order than the mean of Q_n. The interesting thing to note is that the mean of the higher order term C_n decreases by a factor of r^{-1} instead of $r^{-1/2}$ even though the stochastic order $O_P(r^{-3/2})$ of $C_n - E(C_n)$ decreases by a factor of $r^{-1/2}$ as compared with the stochastic order of $Q_n - E(Q_n)$, which is $O_P(r^{-1})$ in Lemma 1.

The following theorem provides all the leading biases in the asymptotic expansion of the maximum simulated likelihood estimator with independently simulated moments.

Theorem 4: *Under our assumptions,*

$$\begin{aligned} \sqrt{n}(\hat{\theta}_L - \theta_o) = \Omega \Big\{ & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P_l(\theta_o, x_i)}{\partial \theta} + L_n + Q_n + B_{1,n} + B_{2,n} \Big\} \\ & + O_P(\max[n^{-1/2}r^{-1/2}, n^{-1}, r^{-1}, n^{1/2}r^{-2}]), \end{aligned} \tag{3.19}$$

where L_n *and* Q_n *have been defined in* (3.4) *and* (3.5),

$$B_{1,n} = n^{-1/2} \Big\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[\sum_{l=1}^{L} d_{li} \frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial \theta \partial \theta'} - E(\sum_{l=1}^{L} d_l \frac{\partial^2 \ln P(l|\theta_o, x)}{\partial \theta \partial \theta'}) \Big] \Big\} \Omega \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta}, \tag{3.20}$$

and

$$B_{2,n} = \frac{n^{-1/2}}{2} \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta'} \Omega E(\sum_{l=1}^{L} d_l \frac{\partial^3 \ln P(l|\theta_o, x)}{\partial \theta_1 \partial \theta \partial \theta'}) \Omega \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} \\ \vdots \\ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta'} \Omega E(\sum_{l=1}^{L} d_l \frac{\partial^3 \ln P(l|\theta_o, x)}{\partial \theta_K \partial \theta \partial \theta'}) \Omega \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{l=1}^{L} d_{li} \frac{\partial \ln P(l|\theta_o, x_i)}{\partial \theta} \end{pmatrix}. \tag{3.21}$$

The terms L_n, Q_n, $B_{1,n}$, *and* $B_{2,n}$ *have the following properties:*

(1) L_n *has zero mean and* $L_n = O_P(r^{-1/2})$,
(2) $Q_n - E(Q_n) = O_P(r^{-1})$ *and* $E(Q_n) = n^{1/2}r^{-1}\bar{\mu}$ *with* $\bar{\mu}$ *defined in* (3.8) *of Lemma 1.*
(3) $B_{1,n} - E(B_{1,n}) = O_P(n^{-1/2})$ *and* $E(B_{1,n}) = n^{-1/2}\Delta_1$, *where*

$$\Delta_1 = E\left\{\left[\left(\sum_{l=1}^{L} d_l \frac{\partial \ln P(l|\theta_o, x)}{\partial \theta'}\right) \otimes \left(\sum_{l=1}^{L} d_l \frac{\partial^2 \ln P(l|\theta_o, x)}{\partial\theta\partial\theta'} - E\left(\sum_{l=1}^{L} d_l \frac{\partial \ln P(l|\theta_o, x)}{\partial\theta\partial\theta'}\right)\right)\right] \operatorname{vec}(\Omega)\right\}, \quad (3.22)$$

and

(4) $B_{2,n} - E(B_{2,n}) = O_P(n^{-1/2})$ *and* $E(B_{2,n}) = n^{-1/2}\Delta_2$, *where*

$$\Delta_2 = \frac{1}{2}\begin{pmatrix} \operatorname{tr}(\Omega E(\sum_{l=1}^{L} d_l \frac{\partial^3 \ln P(l|\theta_o, x)}{\partial\theta_1\partial\theta\partial\theta'})) \\ \vdots \\ \operatorname{tr}(\Omega E(\sum_{l=1}^{L} d_l \frac{\partial^3 \ln P(l|\theta_o, x)}{\partial\theta_K\partial\theta\partial\theta'})) \end{pmatrix}. \quad (3.23)$$

From Theorem 4, we see that the leading components of biases of the simulated likelihood estimator in an asymptotic expansion are generated by the terms Q_n, $B_{1,n}$ and $B_{2,n}$. The order of biases of $B_{1,n}$ and $B_{2,n}$ is $O(n^{-1/2})$. While the bias in Q_n is due to the effect of simulating the likelihood function, the biases in $B_{1,n}$ and $B_{2,n}$ are due to the nonlinearity of the likelihood estimation. As a corollary of the above asymptotic expansion, it provides an asymptotic expansion of the classical maximum likelihood estimator derived from maximizing the exact likelihood function (2.2). For the classical maximum likelihood estimator $\hat{\theta}_c$ of the discrete choice model from (2.2). The corresponding asymptotic expansion of $\hat{\theta}_c$ will be

$$\sqrt{n}(\hat{\theta}_c - \theta_o) = \Omega\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P_l(\theta_o, x_i)}{\partial\theta} + B_{1,n} + B_{2,n}\right\} + O_P(n^{-1}). \quad (3.24)$$

This expansion can be seen either by inspection of the proof of Theorem 4 or by letting r being infinity in the formulas of Theorem 4. The terms $B_{1,n}$ and $B_{2,n}$ appear because the second order derivatives of the log likelihood function with respect to θ are functions of both exogenous variables and endogenous variables. For the normal regression model, the second order derivatives of the log normal likelihood function are functions of the regressors only and the maximum likelihood estimator, which is the least square estimator, is known to be unbiased.

The biases in $B_{1,n}$ and $B_{2,n}$ in the approximation of $\sqrt{n}(\hat{\theta}_L - \theta_o)$ can dominate the bias in Q_n only if r increases faster than the increase of the sample size n. If r increases faster than or propositional to n, the leading biases will be of order $O(n^{-1/2})$, which is the same bias order in the classical likelihood estimator in (3.24). If r increases slower than n, the order $O(n^{1/2}r^{-1})$ of bias in Q_n will be larger than $O(n^{-1/2})$ and it will be the dominated bias. When r increases slower than n, the bias of the simulation likelihood estimator is severe than the bias of the classical likelihood estimator. The practical implication of this analysis is that in order to achieve asymptotic efficiency and to keep the bias of the simulated likelihood estimator with independently simulated moment to the usual order $O(n^{-1/2})$, it is necessary to have the number of simulation r for each individual to increase as fast as or faster than the sample size n. To suppress the bias due to simulation, it is necessary to increase r faster than n. These designs seem demanding in terms of the total number of simulated variables, which will be proportional to or larger than n^2, and the computational burden. In a subsequent section, we will discuss bias-adjustment procedures that may be useful to reduce the bias to some satisfactory degree.

So far, our analyses have concentrated on the simulated likelihood estimator with independently simulated moments. It will be of interest to compare and clarify the possible differences between the simulatd likelihood estimator with independently simulated moments and the likelihood estimator with dependently simulated moments. It is also useful to point out the differences of the maximum simulated likelihood approaches and the method of simulated scores introduced by Hajivassiliou and McFadden [4].

The maximum simulated likelihood estimator $\hat{\theta}_D$ with dependently simulated moments is derived from (2.9). By the mean-value theorem,

$$\sqrt{n}(\hat{\theta}_D - \theta_o) = -\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln \bar{f}_{r,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right\}^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln \bar{f}_{r,l}(\theta_o, x_i)}{\partial\theta}. \quad (3.25)$$

The consistency of the simulated likelihood estimator will require that r increases to infinity as n goes to infinity. The major difference between the two simulated likelihood estimators is due to the asymptotic distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln \bar{f}_{r,l}(\theta_o, x_i)}{\partial\theta}$. By an expansion similar to (3.3),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln \bar{f}_{r,l}(\theta_o, x_i)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + \bar{L}_n + \bar{Q}_n + \bar{C}_n^{(R)}, \quad (3.26)$$

where $\bar{L}_n$, $\bar{Q}_n$ and $\bar{C}_n^{(R)}$ have the similar expressions in (3.4), (3.5) and (3.6) with the dependently simulated moments $\bar{f}_{r,l}(\theta_o, x_i)$ replacing the independently simulated moments $f_{r,l}(\theta_o, x_i)$.

The following theorem for the maximum simulated likelihood estimator with dependently simulated moments will be useful to clarify some of the differences between the two cases.

Theorem 5: *Under our assumptions,*

$$\bar{L}_n = O_P(n^{1/2}r^{-1/2}) \quad \text{and} \quad \bar{Q}_n = O_P(n^{1/2}r^{-1}). \quad (3.27)$$

When $\lim_{n\to\infty} n/r = \lambda_1$ *exists and is finite,*

$$\sqrt{n}(\hat{\theta}_D - \theta_o) \xrightarrow{D} N(0, \Omega[\Omega^{-1} + \lambda_1 V]\Omega). \quad (3.28)$$

On the other hand, when $\lim_{n\to\infty} n/r = \infty$,

$$\sqrt{r}(\hat{\theta}_D - \theta_o) \xrightarrow{D} N(0, \Omega V \Omega), \quad (3.29)$$

where

$$V = E\left\{\sum_{l=1}^{L} E\left[\frac{\partial h_l(v, x, \theta_o)}{\partial\theta} - \frac{\partial \ln P(l|\theta_o, x)}{\partial\theta}h_l(v, x, \theta_o)\Big|v\right] \times \sum_{l=1}^{L} E\left[\frac{\partial h_l(v, x, \theta_o)}{\partial\theta} - \frac{\partial \ln P(l|\theta_o, x)}{\partial\theta}h_l(v, x, \theta_o)\Big|v\right]'\right\}. \quad (3.30)$$

The major difference between the dependent moments case and the independent moments case is on the stochastic order of the first order term $\bar{L}_n$. As contrary to the term L_n in (3.4) which has order $O_P(r^{-1/2})$, the term $\bar{L}_n$ in the above theorem has order $O_P(n^{1/2}r^{-1/2})$, which is apparently much larger. Both L_n and $\bar{L}_n$ have zero mean. It is the variance of $\bar{L}_n$ that is larger. This is so due to the dependence of the simulated moments across individuals. The sample average across individuals in $\bar{L}_n$ does not reduce the variation of the simulation errors due to dependence. The quadratic error term $\bar{Q}_n$ has the same mean as Q_n and hence have the same order as Q_n. Sample averaging cross products of independent moments have no effect on reducing the mean of the products of simulated moments. The first error term $\bar{L}_n$ has an stochastic order larger than the second error term $\bar{Q}_n$. The second order error term $\bar{Q}_n$ does not dominate the leading first order error term $\bar{L}_n$. When r increases as fast as or faster than n, the simulated likelihood estimator $\hat{\theta}_D$ is $\sqrt{n}$-consistency. It is asymptotically efficient when r increases faster than n since in that case λ_1 in (3.28) is zero. When r increases in a rate slow than n, the estimator $\hat{\theta}_D$ converges to θ_o at a slow rate $\sqrt{r}$ rather than $\sqrt{n}$. However, since the $\bar{L}_n$ dominates $\bar{Q}_n$ in stochastic order, the limiting distribution of the properly normalized statistic $\sqrt{r}(\hat{\theta}_D - \theta_o)$ is located at the center zero. The normalization of $\hat{\theta}_D - \theta_o$ with $\sqrt{r}$ instead of $\sqrt{n}$ when r increases slower than n rescales the bias $E(\bar{Q}_n)$ in the expansion (3.26) by a factor $\sqrt{r/n}$. The order of $\sqrt{r/n}\bar{Q}_n$ becomes $O_P(r^{-1/2})$. The leading bias term in the asymptotic expansion of $\sqrt{r}(\hat{\theta}_D - \theta_o)$ has order $O(r^{-1/2})$ when r increases slower than n. When r increases as fast as or faster than n, the bias order $O(n^{1/2}r^{-1})$ of $\bar{Q}_n$ will apparently be equal to or smaller than $O(n^{-1/2})$. In summary, for the simulated likelihood estimator with dependent moments, it has in general larger asymptotic variance than the estimator $\hat{\theta}_L$, however, the properly normalized statistic of $\hat{\theta}_D - \theta_o$ has asymptotically normal distribution located at the center zero.

Hajivassiliou and McFadden [4] introduced a method of simulated scores which overcomes the bias problem in the maximum simulated likelihood approach. Their method is based on the method of simulated moments (see, McFadden [9]). The maximum simulated likelihood estimator $\hat{\theta}_L$ with independently simulated moments in (2.4) satisfies the first order equation:

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

For the case that $f_{r,l}(\theta, x_i)$ has the adding-up property, i.e., $\sum_{l=1}^{L} f_{r,l}(\theta, x_i) = 1$ for all x_i, the first order condition will be equivalent to

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L}(d_{li} - f_{r,l}(\hat{\theta}_L, x_i))\frac{\partial \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial \theta} = 0. \tag{3.31}$$

Instead of (3.31), the Hajivassiliou and McFadden estimator $\hat{\theta}_{HM}$ is solved from the following equation:

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L}(d_{li} - \hat{f}_{r,l}(\hat{\theta}_{HM}, x_i))\frac{\partial \ln \tilde{f}_{r,l}(\hat{\theta}_{HM}, x_i)}{\partial \theta} = 0, \tag{3.32}$$

where $\hat{f}_{r,l}(\theta, x_i)$ and $\tilde{f}_{r,l}(\theta, x_i)$ are two independently simulated probabilities. As $\hat{f}_{r,l}(\theta_o, x)$ is an unbiased estimator of $P(l|\theta_o, x)$, the expectation of $\sum_{l=1}^{L}(d_{li} - \hat{f}_{r,l}(\theta_o, x_i))\frac{\partial \ln \tilde{f}_{r,l}(\theta_o, x_i)}{\partial \theta}$ is zero. So the bias problem in a asymptotic expansion of (3.32) does not exist. Because of the unbiasedness design in their simulation method, the Hajivassiliou and McFadden estimator can be asymptotically efficient if r goes to infinity at any rate as n goes to infinity. The independence of the simulated probabilities $\hat{f}_{r,l}(\theta, x_i)$ and the simulated instrumental variables $\frac{\partial \ln \tilde{f}_{r,l}(\theta, x_i)}{\partial \theta}$ is the crucial setting in their method.

4. Bias Adjustment

The slow increasing r case can cause not only the simulated likelihood estimator $\hat{\theta}_L$ to converge in probability to θ_o at a slow rate but also converge to a degenerate distribution after normalization. The latter phenomenon has some useful implication for correcting the asymptotic bias in the case (*ii*) and the potential bias in the case (*i*). Let r_1 be the number of simulated random variables such that $\lim_{n\to\infty} r_1 = \infty$ and $\lim_{n\to\infty} n^{-1/2}r_1 > 0$ and let $\hat{\theta}_{L,1}$ be the corresponding maximum simulated likelihood estimator of θ_o. Select a slow increasing r_2 such that $\lim_{n\to\infty} r_2 = \infty$ but $\lim_{n\to\infty} r_2 n^{-1/2} = 0$. Let $\hat{\theta}_{L,2}$ be the maximum simulated likelihood estimator derived with r_2 number of simulated random variables. Define an estimator $\hat{\theta}_L^*$ where

$$\hat{\theta}_L^* = \left(\hat{\theta}_{L,1} - \frac{r_2}{r_1}\hat{\theta}_{L,2}\right) \Big/ \left(1 - \frac{r_2}{r_1}\right).$$

We note that $\hat{\theta}_L^* - \theta_o = (1 - r_2/r_1)^{-1}\left[(\hat{\theta}_{L,1} - \theta_o) - (r_2/r_1)(\hat{\theta}_{L,2} - \theta_o)\right]$ and

$$\sqrt{n}(\hat{\theta}_L^* - \theta_o) = (1 - r_2/r_1)^{-1}\left[\sqrt{n}(\hat{\theta}_{L,1} - \theta_o) - (\sqrt{n}/r_1)r_2(\hat{\theta}_{L,2} - \theta_o)\right].$$

The asymptotic distribution of $\sqrt{n}(\hat{\theta}_{L,1} - \theta_o)$ satisfies either (3.9) of the case (*i*) or (3.10) of the case (*ii*) in Theorem 2. The probability limit of $r_2(\hat{\theta}_{L,2} - \theta_o)$ is the one in (3.11) of the case (*iii*). As $\lim_{n\to\infty} \frac{\sqrt{n}}{r_1} = \lambda$, the asymptotic bias in $\sqrt{n}(\hat{\theta}_{L,1} - \theta_o)$ is eliminated by substracting the limit of $(\sqrt{n}/r_1)r_2(\hat{\theta}_{L,2} - \theta_o)$. Since r_2/r_1 converges to zero,

$$\sqrt{n}(\hat{\theta}_L^* - \theta_o) \xrightarrow{D} N(0, \Omega).$$

This bias correction involves the computation of two estimates. As it will be shown in subsequent sections, the bias can be corrected directly without the computation of different estimates. The above correction is illustrative.

A bias-adjusted estimator $\hat{\theta}_A$ can be derived by correcting a bias term directly from the simulated likelihood estimator $\hat{\theta}_L$:

$$\hat{\theta}_A = \hat{\theta}_L + \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial\theta\partial\theta'}\right\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\nu_r(\hat{\theta}_L, x_i), \tag{4.1}$$

where $\nu_r(\theta, x_i) = \frac{1}{r}\mu_r(\theta, x_i)$ with

$$\mu_r(\theta, x_i) = \sum_{l=1}^{L}\left\{\frac{d_{li}}{f_{r,l}^2(\theta, x_i)}\left[\frac{\partial \ln f_{r,l}(\theta, x_i)}{\partial \theta}(S_{r,l}(\theta, x_i) - f_{r,l}^2(\theta, x_i)) - \left(C_{r,l}(\theta, x_i) - f_{r,l}(\theta, x_i)\frac{\partial f_{r,l}(\theta, x_i)}{\partial \theta}\right)\right]\right\},$$

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

and

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

As an alternative of (4.1), a bias-adjusted estimator can also be defined as

$$\hat{\theta}_A = \hat{\theta}_L - \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial \theta}\frac{\partial \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial \theta'}\right\}^{-1}\frac{1}{n}\sum_{i=1}^{n}\nu_r(\hat{\theta}_L, x_i), \tag{4.1'}$$

Conditional on x_i, $S_{r,l}(\hat{\theta}_L, x_i) - f_{r,l}^2(\hat{\theta}_L, x_i)$ and $C_{r,l}(\hat{\theta}_L, x_i) - f_{r,l}(\hat{\theta}_L, x_i)\frac{\partial f_{r,l}(\hat{\theta}_L, x_i)}{\partial \theta}$ are respectively consistent estimates of $\mathrm{var}(h_l(v, x_i)|x_i)$ and $\mathrm{cov}(h_l(v, x_i), \frac{\partial h_l(v, x_i)}{\partial \theta}|x_i)$. The sample average $\frac{1}{n}\sum_{i=1}^{n}\mu_r(\hat{\theta}_L, x_i)$ is a consistent estimate of $\bar{\mu}$ in Lemma 1. The second term on the right hand side of (4.1) adjusts approximately the bias $\frac{1}{r}\Omega\bar{\mu}$ from $\hat{\theta}_L$. We note that this bias-adjustment procedure is applicable to any simulated likelihood estimator from (2.4) without imposing any rate of growth on r except that r goes to infinity when n goes to infinity. Of course if r has already be chosen to increase faster than n, the above bias-adjustment is unnecessary and redundant. The following theorem demonstrates that the bias of $O(n^{1/2}r^{-1})$ has been eliminated.

Theorem 6: *Under our assumptions,*

$$\sqrt{n}(\hat{\theta}_A - \theta_o) = \Omega\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + L_n + [Q_n - E(Q_n)] + B_{1,n} + B_{2,n}\Big\} + O_P(\max[n^{-1/2}r^{-1/2}, n^{-1}, r^{-1}, n^{1/2}r^{-2}]). \quad (4.2)$$

The bias $E(Q_n)$ has been removed from $\hat{\theta}_L$. Since $Q - E(Q_n)$ has zero mean in the asymptotic expansion of $\sqrt{n}(\hat{\theta}_A - \theta_o)$, the biases in $\hat{\theta}_A$ do not contain a bias term of $O(n^{1/2}r^{-1})$. The biases in $B_{1,n}$ and $B_{2,n}$ of order $O(n^{-1/2})$ remain but they are due to the nonlinearity of the log likelihood function and not due to simulation. The bias order $O_P(n^{-1/2})$ remains because our bias adjustment is designed to eliminate the leading bias term due to simulation. It is possible to remove these biases but it is out of scope of this article and will not be discussed here. The biases due to simulation error have not been completely eliminated, but the remaining biases due to simulation are of much smaller orders. From the asymptotic expansion in (4.2), the largest remaining biases can have orders no larger than the orders $O(n^{-1/2})$, $O(n^{1/2}r^{-2})$ or $O(r^{-1})$. With a little bit of analysis, we see that the dominated bias orders can only be $O(n^{-1/2})$ or $O(n^{1/2}r^{-2})$. This is so, because when r increases faster than $n^{1/2}$, the order $O(n^{-1/2})$ is larger than $O(r^{-1})$. When r increases proportionally to $n^{1/2}$, $O_P(r^{-1})$ is the same order as $O(n^{-1/2})$ or $O(n^{1/2}r^{-2})$. Finally when r increases slower than $n^{1/2}$, the order $O(n^{1/2}r^{-2})$ is larger than $O(r^{-1})$. In all cases, the order of the largest remaining bias in $\sqrt{n}(\hat{\theta}_A - \theta_o)$ can not be larger than $O(\max[n^{-1/2}, n^{1/2}r^{-2}])$. If r increases at a rate proportional to $n^{1/4}$, an asymptotic bias will be there in $\sqrt{n}(\hat{\theta}_A - \theta_o)$. If r increases slower than $n^{1/4}$, the asymptotic distribution will diverge. The bias-adjusted estimator $\hat{\theta}_A$ will be asymptotically efficient if the number of simulated random variables r increases faster than $n^{1/4}$. The bias-adjusted estimator $\hat{\theta}_A$ will be asymptotically efficient and has biases with orders no larger than $O(n^{-1})$ as in the classical maximum likelihood estimator if r increases as fast as or faster than $n^{1/2}$. Comparing this requirement with the requirement that r increases as fast as or faster than n for the bias-unadjusted estimator $\hat{\theta}_L$, it is much less demanding on the total number of simulation for the whole sample and the computational burden on the evaluation of the simulated probabilities. The practical implication is that it is possible to have an asymptotically efficient estimator of θ with a bias of order no larger than $O(n^{-1})$ by the method of maximum simulated likelihood with independent moments with a simulated design by letting the number of simulated variables r to increase as fast as or faster than $n^{1/2}$.

5. Some Monte Carlo Evidence

In this section, we report some Monte Carlo experiments on the maximum simulated likelihood estimation with independently simulated moments and the bias-adjustment procedure (4.1)'. The model is a dynamic discrete choice panel data model used in Lee [6]. The model is specified as follows:

$$y_{it}^* = \beta x_{it} + \lambda d_{i,t-1} + u_{it} \quad (5.1)$$

and

$$u_{it} = \rho u_{i,t-1} + \epsilon_{it}, \quad i = 1, \cdots, n;\ t = 1, \cdots, T, \quad (5.2)$$

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 be 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:

$$x_{it} = \frac{1}{\sqrt{2}} z_{it} + \sqrt{6} s_i, \quad (5.3)$$

where z_{it} are i.i.d. truncated normal $N(0,\ 1)$ variables with support $[-2,\ 2]$ and s_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. A four time period model is used for our experiment, i.e., $T = 4$. The sample size n for each replication is either 100 or 200. The number of simulated random variables is ranged from 10 to 100 per observation. The number of replications for each case is about 600.

Table 1 reports the maximum simulated likelihood estimates and their bias-adjusted estimates. The maximum simulated likelihood estimation for this model tends to underestimate the regression coefficient β and the serial correlation coefficient ρ but tends to overestimate the dynamic coefficient λ of this model. The bias-adjustment procedure has apparently reduced the biases of the estimates of regression coefficient β and the serial correlation coefficient ρ. The standard deviations (SD) of all the bias-adjusted estimates are slightly larger. However, the gains in the bias reduction of the estimates of β and ρ are substantial and their root mean squared errors (RMSE) are reduced. The biases in the estimates of the dynamic coefficient λ do not change much and their RMSE increase slightly. Overall, the bias-adjustment procedure does reduce the biases of the maximum simulated likelihood estimates.

TABLE 1.
Maximum Simulated Likelihood Estimation with Independently Simulated Moments and Bias-Adjustment
True parameters: $\beta = 1$, $\lambda = 0.2$, $\rho = 0.4$

T	n	r		Mean	SD	RMSE	Median	LQ	UQ
				Bias-Unadjusted Simulated Likelihood Estimates					
4	100	10	β	.7900	.0694	.2212	.7851	.7427	.8306
			λ	.2435	.1206	.1281	.2493	.1688	.3249
			ρ	.1932	.1027	.2309	.1977	.1298	.2609
4	100	30	β	.8791	.0790	.1444	.8768	.8256	.9335
			λ	.2454	.1298	.1374	.2437	.1564	.3352
			ρ	.2508	.1147	.1881	.2606	.1733	.3260
4	100	50	β	.9099	.0784	.1194	.9069	.8544	.9630
			λ	.2467	.1335	.1413	.2462	.1508	.3422
			ρ	.2662	.1074	.1715	.2745	.1989	.3431
4	100	100	β	.9414	.0895	.1069	.9374	.8729	.9985
			λ	.2558	.1350	.1460	.2476	.1560	.3460
			ρ	.2761	.1165	.1700	.2867	.1969	.3573
4	200	15	β	.8242	.0537	.1838	.8250	.7880	.8569
			λ	.2449	.0901	.1006	.2430	.1841	.3069
			ρ	.2187	.0830	.1993	.2231	.1664	.2738
4	200	50	β	.9083	.0541	.1064	.9069	.8718	.9438
			λ	.2506	.0942	.1069	.2591	.1859	.3093
			ρ	.2644	.0820	.1584	.2710	.2122	.3230
T	n	r		**Bias-Adjusted Estimates**					
4	100	10	β	.8493	.0740	.1679	.8417	.8000	.8962
			λ	.2458	.1324	.1400	.2557	.1604	.3398
			ρ	.2301	.1195	.2077	.2335	.1557	.3070
4	100	30	β	.9330	.0876	.1103	.9303	.8707	.9928
			λ	.2475	.1421	.1498	.2442	.1467	.3480
			ρ	.2843	.1305	.1743	.2943	.1985	.3752
4	100	50	β	.9567	.0871	.0972	.9545	.8958	1.0154
			λ	.2492	.1438	.1519	.2486	.1502	.3483
			ρ	.2940	.1180	.1585	.3049	.2181	.3795
4	100	100	β	.9780	.0987	.1011	.9753	.9057	1.0419
			λ	.2592	.1430	.1547	.2517	.1555	.3485
			ρ	.2958	.1250	.1627	.3122	.2131	.3821
4	200	15	β	.8811	.0575	.1321	.8822	.8408	.9210
			λ	.2479	.0987	.1096	.2452	.1771	.3146
			ρ	.2535	.0934	.1737	.2622	.1960	.3178
4	200	50	β	.9533	.0598	.0758	.9503	.9123	.9927
			λ	.2538	.1007	.1141	.2571	.1831	.3157
			ρ	.2903	.0908	.1424	.3003	.2315	.3533

6. Conclusions

In this article, we have pointed out a bias in the asymptotic expansion of the maximum simulated likelihood estimator $\hat{\theta}_L$ with independently simulated moments. This bias in the asymptotic expansion of $\sqrt{n}(\hat{\theta}_L - \theta_o)$ has an order $O(n^{1/2}r^{-1})$ where n is the sample size and r is the number of simulated random variables for each sample observation. This bias occurs due to the nonlinearity of the derivatives of the log likelihood function and the statistically independent simulation errors. The classical maximum likelihood estimator is known to have biases of order no larger than $O(n^{-1/2})$ in an asymptotic expansion of $\sqrt{n}(\hat{\theta}_c - \theta_o)$ where $\hat{\theta}_c$ is the classical maximum likelihood estimator. The maximum simulated likelihood estimator $\hat{\theta}_L$ is asymptotically efficient when r goes to infinity with a rate faster than $n^{1/2}$. However, the bias due to simulation will dominate the biases of $O(n^{-1/2})$ unless r goes to infinity with a rate as fast as or faster than n. This is quite a demanding requirement for the simulation estimation with independently simulated moments. If r is increasing only as fast as $n^{1/2}$, an asymptotic bias will be present in the limiting distribution of $\sqrt{n}(\hat{\theta}_L - \theta_o)$. If r increases slower than $n^{1/2}$, $\sqrt{n}(\hat{\theta}_L - \theta_o)$ will diverge. Under such a circumstance, $\hat{\theta}_L$ can only be r-consistent and the properly normalized asymptotic distribution of $\hat{\theta}_L$ is a degenerated distribution.

In this article, we have also suggested some bias-adjustment procedures to correct for the dominated bias due to simulation. The bias can be reduced to an order $O(n^{1/2}r^{-2})$ in an asymptotic expansion of $\sqrt{n}(\hat{\theta}_A - \theta_o)$ where $\hat{\theta}_A$ is a bias-adjusted estimator constructed from $\hat{\theta}_L$. The bias-adjusted estimator is asymptotically efficient if r increases faster than $n^{1/4}$. The biases in $\sqrt{n}(\hat{\theta}_A - \theta_o)$ will have orders no larger than $O(n^{-1/2})$ if r increases as fast as or faster than $n^{1/2}$. This requirement is much less demanding for the implementation of the simulated likelihood estimation with independently simulated moments. Some Monte Carlo experiments have demonstrated that the bias-adjustment procedure is valuable.

This article has concentrated on the analysis of the asymptotic expansion of biases of the simulated likelihood estimator. The analysis has useful implications on point estimation as well as on the construction of Wald test statistics and Scoring test statistics (Lagrange Multiplier test statistics). The implementation of such statistics and the likelihood ratio tests statistics will be left for future investigation.

Appendix A

Lemma A.1: *If b and $\hat{b}$ are not zero, then*

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \sum_{k=1}^{m}(-1)^{k-1}\frac{1}{b^k}[(\hat{a}-a)(\hat{b}-b)^{k-1} - \frac{a}{b}(\hat{b}-b)^k] + (-1)^m\frac{1}{\hat{b}b^m}[(\hat{a}-a)(\hat{b}-b)^m - \frac{a}{b}(\hat{b}-b)^{m+1}],$$

for all $m \geq 0$, where $\sum_{k=1}^{0}$ is understood to be zero as a convention.

Proof: This lemma can be proved by induction. Since $\frac{1}{\hat{b}} = \frac{1}{b}[1 - \frac{1}{\hat{b}}(\hat{b}-b)]$, we have

$$\frac{\hat{a}}{\hat{b}} - \frac{a}{b} = \frac{1}{\hat{b}}(\hat{a} - \frac{a}{b}\hat{b}) = \frac{1}{\hat{b}}[(\hat{a}-a) - \frac{a}{b}(\hat{b}-b)].$$

Suppose the result holds for $m-1$. It follows that

$$\begin{aligned}\frac{\hat{a}}{\hat{b}} - \frac{a}{b} &= \sum_{k=1}^{m-1}(-1)^{k-1}\frac{1}{b^k}[(\hat{a}-a)(\hat{b}-b)^{k-1} - \frac{a}{b}(\hat{b}-b)^k] \\ &\quad + (-1)^{m-1}\frac{1}{\hat{b}b^{m-1}}[(\hat{a}-a)(\hat{b}-b)^{m-1} - \frac{a}{b}(\hat{b}-b)^m] \\ &= \sum_{k=1}^{m-1}(-1)^{k-1}\frac{1}{b^k}[(\hat{a}-a)(\hat{b}-b)^{k-1} - \frac{a}{b}(\hat{b}-b)^k] \\ &\quad + (-1)^{m-1}\frac{1}{b}[1 - \frac{1}{\hat{b}}(\hat{b}-b)]\frac{1}{b^{m-1}}[(\hat{a}-a)(\hat{b}-b)^{m-1} - \frac{a}{b}(\hat{b}-b)^m] \\ &= \sum_{k=1}^{m}(-1)^{k-1}\frac{1}{b^k}[(\hat{a}-a)(\hat{b}-b)^{k-1} - \frac{a}{b}(\hat{b}-b)^k] \\ &\quad + (-1)^m\frac{1}{\hat{b}}\frac{1}{b^m}[(\hat{a}-a)(\hat{b}-b)^m - \frac{a}{b}(\hat{b}-b)^{m+1}].\end{aligned}$$

Q.E.D.

Lemma A.2: *Let z_i, $i = 1, \cdots, r$, be independent random variables with mean 0. Let ν be an even integer. Then*

$$E|\sum_{i=1}^{r} z_i|^\nu \leq A_\nu r^{\frac{\nu}{2}-1}\sum_{i=1}^{r} E|z_i|^\nu,$$

where A_ν is a universal constant depending only on ν.

Proof: This is Lemma A on page 304 in Serfling [12].

Proposition A.3: *Suppose that $(v_1^{(i)}, \ldots, v_r^{(i)}, x_i)$, $i = 1, \ldots, n$, are identically distributed random vectors. Let*

$$q_{ri} = s(x_i) \cdot \left[\frac{1}{r}\sum_{j=1}^{r} p(v_j^{(i)}, x_i)\right]^{m_1} \cdot \left[\frac{1}{r}\sum_{j=1}^{r} t(v_j^{(i)}, x_i)\right]^{m_2},$$

where m_1 and m_2 are nonnegative intergers and s, p and t are measurable functions such that, conditional on x, $E(p(v,x)|x) = E(t(v,x)|x) = 0$. If $m_1 > 0$ but $m_2 = 0$, assume that $E(s^2(x)p^{2m_1}(v,x))$ is finite. Otherwise, assume that $E(|s(x)|^a p^{2m_1 a}(v,x))$ and $E(|s(x)|^b t^{2m_2 b}(v,x))$ are finite, where $\frac{1}{a} + \frac{1}{b} = 1$, some $a > 1$. Then

$$\frac{1}{n}\sum_{i=1}^{n}|q_{ri}| = O_P(r^{-\frac{m_1+m_2}{2}}).$$

Furthermore, if $(v_1^{(i)}, \ldots, v_r^{(i)}, x_i)$, $i = 1, \ldots, n$, are i.i.d. random vector, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(q_{ri} - \mu_{q_r}) = O_P(r^{-\frac{m_1+m_2}{2}}),$$

where $\mu_{q_r} = E(q_{ri})$ is the mean of q_{ri}.

Proof: By the Holder inequality,

$$E(q_{ri}^2) \leq E^{1/a}\{|s(x)|^a[\frac{1}{r}\sum_{j=1}^{r}p(v_j,x)]^{2m_1 a}\}E^{1/b}\{|s(x)|^b[\frac{1}{r}\sum_{j=1}^{r}t(v_j,x)]^{2m_2 b}\}.$$

The Lemma A.2 implies that

$$E([\frac{1}{r}\sum_{j=1}^{r}p(v_j,x)]^{2m_1 a}|x) \leq \frac{c}{r^{m_1 a}}E(p^{2m_1 a}(v,x)|x)$$

and $E([\frac{1}{r}\sum_{j=1}^{r}t(v_j,x)]^{2m_2 b}|x) \leq \frac{c}{r^{m_2 b}}E(t^{2m_2 b}(v,x)|x)$ where c is a constant. Therefore, $E(q_{ri}^2) \leq \frac{c}{r^{m_1+m_2}}d$, where $d = E^{1/a}\{|s(x)|^a p^{2m_1 a}(v,x)\}E^{1/b}\{|s(x)|^b t^{2m_2 b}(v,x)\}$. By the Markov inequality and the inequality of absolute moments,

$$P(n^{-1}r^{\frac{m_1+m_2}{2}}\sum_{i=1}^{n}|q_{ri}| \geq \epsilon) \leq \epsilon^{-1}r^{\frac{m_1+m_2}{2}}E(|q_{ri}|) \leq \epsilon^{-1}r^{\frac{m_1+m_2}{2}}E^{1/2}(|q_{ri}|^2) \leq \epsilon^{-1}c^{1/2}d^{1/2}.$$

Since ϵ is arbitary, $n^{-1}r^{\frac{m_1+m_2}{2}}\sum_{i=1}^{n}|q_{ri}| = O_P(1)$. This proves the first part of the results.

When the random vectors $(v_1^{(i)}, \ldots, v_r^{(i)}, x_i)$ are i.i.d. for all i, the Chebyshev inequality implies

$$P(|n^{-1/2}r^{\frac{m_1+m_2}{2}}\sum_{i=1}^{n}(q_{ri} - \mu_{q_r})| \geq \epsilon) \leq \epsilon^{-2}r^{m_1+m_2}E(q_{ri} - \mu_{q_r})^2 \leq \epsilon^{-2}cd.$$

Since ϵ is arbitrary, $n^{-1/2}r^{\frac{m_1+m_2}{2}}\sum_{i=1}^{n}(q_{ri} - \mu_{q_r}) = O_P(1)$. Hence $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(r_{ri} - \mu_{q_r}) = O_P(r^{-\frac{m_1+m_2}{2}})$. This proves the second part of the results. The case with $m_2 = 0$ can be proved without using the Holder inequality. Q.E.D.

Proposition A.4: *Let $\{v_j\}$ be i.i.d. random variables. Suppose that the measurable functions $p_l(v)$, $l = 1, \ldots, 4$, have mean zero. Then*

(i)

$$E\{\prod_{l=1}^{k+1}[\frac{1}{r}\sum_{j=1}^{r}p_l(v_j)]\} = \frac{1}{r^k}E\{\prod_{l=1}^{k+1}p_l(v)\}, \qquad k = 1, 2;$$

(ii)

$$E\{\prod_{l=1}^{4}[\frac{1}{r}\sum_{j=1}^{r}p_l(v_j)]\} = \frac{1}{r^2}\Big\{\frac{1}{r}E[\prod_{l=1}^{4}p_l(v)] + \frac{r-1}{r}[E(p_1(v)p_2(v))E(p_3(v)p_4(v)) + E(p_1(v)p_3(v))E(p_2(v)p_4(v)) + E(p_1(v)p_4(v))E(p_2(v)p_3(v))]\Big\}.$$

Proof: The proof is straightforward by using the independence property of the vs. Q.E.D.

Proposition A.5: *Let $(v_1^{(i)}, \ldots, v_r^{(i)}, x_i)$, $i = 1, \ldots, n$, be i.i.d. random vectors, and $p(v,x)$ is a measurable function with finite fourth order moment. Denote $p_r(x_i) = \frac{1}{r}\sum_{j=1}^{r}p(v_j^{(i)}, x_i)$. Suppose that h is a* *twice differentiable measurable function such that its second order derivative is uniformly continuous and bounded. Furthermore suppose that $p_r(x)$ converges in probability uniformly in x to $E(p(v,x)|x)$. Then*

$$\frac{1}{n}\sum_{i=1}^{n} h(p_r(x_i)) = \frac{1}{n}\sum_{i=1}^{n} h(E[p(v,x)|x]) + O_P(\max[n^{-1/2}r^{-1/2}, r^{-1}]).$$

Proof: By a Taylor expansion up to the second order,

$$\frac{1}{n}\sum_{i=1}^{n} h(p_r(x_i)) = \frac{1}{n}\sum_{i=1}^{n} h(E[p(v,x_i)|x_i]) + F_n + S_n,$$

where

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \nabla' h(E[p(v,x_i)|x_i])(p_r(x_i) - E[p(v,x_i)|x_i]),$$

and

$$S_n = \frac{1}{2n}\sum_{i=1}^{n}(p_r(x_i) - E[p_r(v,x_i)|x_i])'\nabla^2 h(\bar{p}_r(x_i))(p_r(x_i) - E[p_r(v,x_i)|x_i]),$$

with ∇ and ∇^2 being respectively the gradient operator and the matrix of second order derivatives. The second result of Proposition A.3 implies $F_n = O_P(n^{-1/2}r^{-1/2})$. Since $p_r(x)$ converges in probability to $E(p(v,x)|x)$ uniformly in x and $\nabla^2 h(.)$ is uniformly continuous and bounded, $\sup_x \|\nabla^2 h(\bar{p}_r(x_i))\|$ is bounded in probability. As

$$|S_n| \le \sup_x \|\nabla^2 h(\bar{p}_r(x))\| \cdot \frac{1}{2n}\sum_{i=1}^{n}\|p_r(x_i) - E(p(v,x_i)|x_i)\|^2,$$

the first part of Proposition A.3 implies that $S_n = O_P(r^{-1})$. Q.E.D.

Proposition A.6: *Suppose $\{v_j\}$ is a sequence of i.i.d. random variables, and $\{x_i\}$ is a sequence of i.i.d. random variables which are independent of v_j, for all j. Let $\psi(v,x)$ be a measurable function of (v,x) with zero mean and finite variance.*

(i) Suppose that r is an integer-valued function of n such that $\lim_{n\to\infty}\frac{n}{r} = \lambda_1$ exists and is finite, then

$$\frac{1}{r\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r}\psi(v_j,x_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E(\psi(v,x_i)|x_i) + \frac{\sqrt{n}}{r}\sum_{j=1}^{r}E(\psi(v_j,x)|v_j) + o_P(1)$$
$$\xrightarrow{D} N\big(0, \mathrm{var}\{E[\psi(v,x)|x]\} + \lambda_1 \mathrm{var}\{E[\psi(v,x)|v]\}\big).$$

(ii) Suppose that that $\lim_{n\to\infty}\frac{r}{n} = \lambda_2$ exists and is finite, then

$$\frac{1}{n\sqrt{r}}\sum_{i=1}^{n}\sum_{j=1}^{r}\psi(v_j,x_i) = \frac{\sqrt{r}}{n}\sum_{i=1}^{n}E(\psi(v,x_i)|x_i) + \frac{1}{\sqrt{r}}\sum_{j=1}^{r}E(\psi(v_j,x)|v_j) + o_P(1)$$
$$\xrightarrow{D} N\big(0, \lambda_2\mathrm{var}\{E[\psi(v,x)|x]\} + \mathrm{var}\{E[\psi(v,x)|v]\}\big).$$

Proof: These are central limit theorems for generalized U-statistics in Lehmann [7] (see, also A.J. Lee [5]). Q.E.D.

Appendix B

Proof of Lemma 1:

It is apparent from (3.4) that the mean of L_n is zero. Proposition A.3 of Appendix A implies that $L_n = O_P(r^{-1/2})$. The Q_n can be rewritten as $Q_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q_{ri}$, where

$$q_{ri} = \sum_{l=1}^{L}\frac{d_{li}}{P^2(l|\theta_o,x_i)}\Big\{\frac{\partial \ln P(l|\theta_o,x_i)}{\theta}[f_{r,l}(\theta_o,x_i) - P(l|\theta_o,x_i)]^2$$
$$- [f_{r,l}(\theta_o,x_i) - P(l|\theta_o,x_i)]\left[\frac{\partial f_{r,l}(\theta_o,x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o,x_i)}{\partial\theta}\right]\Big\}.$$

Since $E(q_{ri}|d_i,x_i) = \sum_{l=1}^{L}\frac{d_{li}}{P^2(l|\theta_o,x_i)}\Big(\frac{\partial \ln P(l|\theta_o,x_i)}{\partial\theta}\cdot\frac{1}{r}\mathrm{var}(h_l(v,x_i)|x_i) - \frac{1}{r}\mathrm{cov}\big(h_l(v,x_i), \frac{\partial h_l(v,x_i)}{\partial\theta}|x_i\big)\Big)$, the mean μ_q of q_{ri} is $\mu_q = E(q_{ri}) = E\{E(q_{ri}|d_i,x_i)\} = \frac{1}{r}\bar{\mu}$, where $\bar{\mu}$ is defined in the statement of Lemma 1. From Proposition A.3 of the Appendix A, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(q_{ri} - \mu_q) = O_P(r^{-1})$. The remainder term $C_n^{(R)}$ has a smaller order than Q_n since

$$\|C_n^{(R)}\| \le O_P(1)\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}|f_{r,l}(\theta_o,x_i) - P(l|\theta_o,x_i)|^2\cdot\Big\|\frac{\partial f_{r,l}(\theta_o,x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o,x_i)}{\partial\theta}\Big\|$$
$$+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}\|f_{r,l}(\theta_o,x_i) - P(l|\theta_o,x_i)\|^3\Big\}$$
$$\le O_P(n^{1/2}r^{-3/2}),$$

by using the first part of the results in Proposition A.3. Q.E.D.

Proof of Theorem 2:

From Lemma 1,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}d_{li}\frac{\partial \ln f_{r,l}(\theta_o,x_i)}{\partial\theta}$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}d_{li}\frac{\partial \ln P_l(\theta_o,x_i)}{\partial\theta} + L_n + Q_n + C_n^{(R)}$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}d_{li}\frac{\partial \ln P_l(\theta_o,x_i)}{\partial\theta} + O_P(r^{-1/2}) + [O_P(r^{-1}) + n^{1/2}r^{-1}\bar{\mu}] + O_P(n^{1/2}r^{-3/2}).$$

Hence,

$$\sqrt{n}(\hat{\theta}_L - \theta_o) = -\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L}d_{li}\frac{\partial^2 \ln f_{r,l}(\bar{\theta},x_i)}{\partial\theta\partial\theta'}\right\}^{-1}$$
$$\times\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}d_{li}\frac{\partial \ln P_l(\theta_o,x_i)}{\partial\theta} + n^{1/2}r^{-1}\bar{\mu} + O_P(\max[r^{-1/2}, n^{1/2}r^{-3/2}])\Big\}.$$

For the case (i), as $n^{1/2}r^{-1}$ goes to zero, $\sqrt{n}(\hat{\theta}_L - \theta_o) = \Omega\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}d_{li}\frac{\partial \ln P_l(\theta_o,x_i)}{\partial\theta} + o_P(1)$. The result of (i) follows from the classical central limit theorem. Under the case (ii),

$$\sqrt{n}(\hat{\theta}_L - \theta_o) = \left\{E\left[\sum_{l=1}^{L}P(l|\theta_o,x)\frac{\partial \ln P(l|\theta_o,x)}{\partial\theta}\frac{\partial \ln P(l|\theta_o,x)}{\partial\theta'}\right]\right\}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{L}d_{li}\frac{\partial \ln P(l|\theta_o,x_i)}{\partial\theta} + \lambda\bar{\mu}\right\} + o_P(1).$$

which converges in distribution to $N(\lambda\Omega\bar{\mu}, \Omega)$. The case *(iii)* occurs when r increases at a rate slower than $\sqrt{n}$. Let $k_n = r/\sqrt{n}$. It follows that

$$\begin{aligned} r(\hat{\theta}_L - \theta_o) &= k_n\sqrt{n}(\hat{\theta}_L - \theta_o) \\ &= -\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right\}^{-1} \\ &\quad \times \left\{\frac{k_n}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + k_n n^{1/2} r^{-1}\bar{\mu} + k_n O_P(\max[r^{-1/2}, n^{1/2}r^{-3/2}])\right\} \\ &= -\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right\}^{-1} \{O_P(k_n) + \bar{\mu} + O_P(\max[k_n r^{-1/2}, r^{-1/2}])\} \\ &= \Omega\bar{\mu} + o_P(1), \end{aligned}$$

because $r \to \infty$ and $k_n \to 0$. Q.E.D.

Proof of Lemma 3:

Proposition A.3 and Proposition A.4 imply that

$$C_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(c_{ri} - E(c_{ri})) + \sqrt{n}E(c_{ri}) = O_P(r^{-3/2}) + \sqrt{n}r^{-2}\bar{c},$$

where $\bar{c}$ is defined in the statement of Lemma 3, and

$$D_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(d_{ri} - E(d_{ri})) + \sqrt{n}r^{-2}\bar{d}(r) = O_P(r^{-2}) + \sqrt{n}r^{-2}\bar{d}(r),$$

where

$$\begin{aligned} \bar{d}(r) = E\Big(\sum_{l=1}^{L}\frac{1}{P^3(l|\theta_o, x_i)}\Big\{&\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta}E([f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^4|x_i) \\ &- E([\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial\theta}][f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^3|x_i)\Big\}\Big). \end{aligned}$$

Proposition A.4 shows that

$$E([f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)]^4|x_i) = \frac{1}{r^2}\Big\{3\frac{r-1}{r}\mathrm{var}^2(h_l(v, x_i)|x_i) + \frac{1}{r}E([h_l(v, x_i) - P(l|\theta_o, x_i)]^4|x_i)\Big\},$$

and

$$\begin{aligned} &E([\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial\theta}][f_{r,l}(\theta_o, x_i) - P(L|\theta_o, x_i)]^3|x_i) \\ &= \frac{1}{r^2}\Big\{3\frac{r-1}{r}E([\frac{\partial h_l(v, x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial\theta}][h_l(v, x_i) - P(l|\theta_o, x_i)]|x_i)\mathrm{var}(h_l(v, x_i)|x_i) \\ &\quad + \frac{1}{r}E([\frac{\partial h_l(v, x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial\theta}][h_l(v, x_i) - P(l|\theta_o, x_i)]^3|x_i)\Big\}. \end{aligned}$$

For the remainder term R_n,

$$\begin{aligned} \|R_n\| \le O_P(1)\Big\{&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}|f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)|^5 \\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L}\|\frac{\partial f_{r,l}(\theta_o, x_i)}{\partial\theta} - \frac{\partial P(l|\theta_o, x_i)}{\partial\theta}\| \cdot |f_{r,l}(\theta_o, x_i) - P(l|\theta_o, x_i)|^4\Big\} \\ = O_P(&n^{1/2}r^{-5/2}), \end{aligned}$$

from Proposition A.3. Q.E.D.

Proof of Theorem 4:

For each component θ_k of θ,

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln f_{r,l}(\theta_o, x_i)}{\partial\theta_k} + \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\theta_o, x_i)}{\partial\theta_k\partial\theta'}(\hat{\theta}_L - \theta_o) \\ &+ \frac{1}{2n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}(\hat{\theta}_L - \theta_o)'\frac{\partial^3 \ln f_{r,l}(\theta_o, x_i)}{\partial\theta_k\partial\theta\partial\theta'}(\hat{\theta}_L - \theta_o) \\ &+ \frac{1}{3!n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\sum_{k_1=1}^{K}\sum_{k_2=2}^{K}\sum_{k_3=1}^{K}\frac{\partial^4 \ln f_{r,l}(\bar{\theta}, x_i)}{\partial\theta_k\partial\theta_{k_1}\partial\theta_{k_2}\partial\theta_{k_3}}(\hat{\theta}_{L,k_1} - \theta_{ok_1})(\hat{\theta}_{L,k_2} - \theta_{ok_2})(\hat{\theta}_{L,k_3} - \theta_{ok_3}). \end{aligned} \tag{B.1}$$

From (3.13) and Lemma 3,

$$\begin{aligned} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln f_{r,l}(\theta_o, x_i)}{\partial\theta} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + L_n + Q_n + O_P(\max[r^{-3/2}, n^{1/2}r^{-2}]) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + O_P(\max[r^{-1/2}, n^{1/2}r^{-1}]). \end{aligned} \tag{B.2}$$

Proposition A.5 implies that

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\theta_o, x_i)}{\partial\theta\partial\theta'} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta} + O_P(\max[n^{-1/2}r^{-1/2}, r^{-1}]), \tag{B.3}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^3 \ln f_{r,l}(\theta_o, x_i)}{\partial\theta_k\partial\theta\partial\theta'} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^3 \ln P(l|\theta_o, x_i)}{\partial\theta_k\partial\theta\partial\theta'} + O_P(\max[n^{-1/2}r^{-1/2}, r^{-1}]). \tag{B.4}$$

It follows that

$$\begin{aligned} \sqrt{n}(\hat{\theta}_L - \theta_o) &= \Omega\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln f_{r,l}(\theta_o, x_i)}{\partial\theta} \\ &+ \Omega\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta'} - E(\sum_{l=1}^{L} d_l\frac{\partial^2 \ln P(l|\theta_o, x)}{\partial\theta\partial\theta'})\}\sqrt{n}(\hat{\theta}_L - \theta_o) \\ &+ O_P(\max[n^{-1/2}r^{-1/2}, r^{-1}] \cdot \sqrt{n}\|\hat{\theta}_L - \theta_o\|) \\ &+ \frac{1}{2}\Omega\begin{pmatrix} \sqrt{n}(\hat{\theta}_L - \theta_o)'\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^3 \ln P(l|\theta_o, x_i)}{\partial\theta_1\partial\theta\partial\theta'}(\hat{\theta}_L - \theta_o) \\ \vdots \\ \sqrt{n}(\hat{\theta}_L - \theta_o)'\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^3 \ln P(l|\theta_o, x_i)}{\partial\theta_K\partial\theta\partial\theta'}(\hat{\theta}_L - \theta_o) \end{pmatrix} \\ &+ O_P(\max[n^{-1/2}r^{-1/2}, r^{-1}] \cdot \sqrt{n}\|\hat{\theta}_L - \theta_o\|^2) \\ &+ O_P(\sqrt{n}\|\hat{\theta}_L - \theta_o\|^3). \end{aligned} \tag{B.5}$$

From (3.1), (3.2) and $(B.2)$,

$$\begin{aligned} \sqrt{n}(\hat{\theta}_L - \theta_o) &= \Omega\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + O_P(\max[r^{-1/2}, n^{1/2}r^{-1}]) \\ &= O_P(\max[1, n^{1/2}r^{-1}]), \end{aligned} \tag{B.6}$$

which implies that

$$\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta'} - E(\sum_{l=1}^{L} d_l \frac{\partial^2 \ln P(l|\theta_o, x)}{\partial\theta\partial\theta'})\}\sqrt{n}(\hat{\theta}_L - \theta_o)$$
$$= B_{1,n} + O_P(n^{-1/2})O_P(\max[r^{-1/2}, n^{1/2}r^{-1}]), \qquad (B.7)$$

where $B_{1,n}$ is defined in the statement of Theorem 4, and

$$\sqrt{n}(\hat{\theta}_L - \theta_o)'\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^3 \ln P(l|\dot{\theta}_o, x_i)}{\partial\theta_k\partial\theta\partial\theta'}(\hat{\theta}_L - \theta_o)$$
$$= n^{-1/2}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta'})\Omega(\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^3 \ln P(l|\theta_o, x_i)}{\partial\theta_k\partial\theta\partial\theta'})\Omega(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta})$$
$$+ n^{-1/2}O_P(\max[r^{-1/2}, n^{1/2}r^{-1}])$$
$$= n^{-1/2}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta'})\Omega E(\sum_{l=1}^{L} d_l\frac{\partial^3 \ln P(l|\theta_o, x)}{\partial\theta_k\partial\theta\partial\theta'})\Omega(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta})$$
$$+ O_P(n^{-1}) + O_P(\max[n^{-1/2}r^{-1/2}, r^{-1}]). \qquad (B.8)$$

Substituting $(B.2), (B.7), (B.8)$, and $\hat{\theta}_L - \theta_o = O_P(\max[n^{-1/2}, r^{-1}])$ into $(B.5)$, we have the asymptotic expansion in (3.19). Obviously, $B_{1,n}$ and $B_{2,n}$ have the order $O_P(n^{-1/2})$. The computation of the means of $B_{1,n}$ and $B_{2,n}$ are trivial. Q.E.D.

Proof of Theorem 5:

By the expansion (3.26),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln \tilde{f}_{r,l}(\theta_o, x_i)}{\partial\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + \bar{L}_n + \bar{Q}_n + \bar{C}_n^{(R)}.$$

The term $\bar{L}_n$ can be rewritten as $\bar{L}_n = \frac{1}{r\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r}\tilde{\phi}(v_j^{(i)}, z_i)$, where $z_i = (d_i, x_i)$ and

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

Since $E(\tilde{\phi}(v, z)|z) = 0$, the central limit theorem of the generalized U-statistic in Proposition A.6 implies that when $\lim_{n\to\infty} n/r = \lambda_1$ exists and is finite, $\bar{L}_n \xrightarrow{D} N(0, \lambda_1 V)$, but when $\lim_{n\to\infty} n/r = \infty$, $\sqrt{r/n}\bar{L}_n \xrightarrow{D} N(0, V)$ where V is defined in (3.30). In either cases, we have $\bar{L}_n = O_P(n^{1/2}r^{-1/2})$. The first part of the results of Proposition A.3 implies that $\bar{Q}_n = O_P(n^{1/2}r^{-1})$ and $\bar{C}_n = O_P(n^{1/2}r^{-3/2})$. Therefore,

$$\sqrt{n}(\hat{\theta}_D - \theta_o) = -\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln \tilde{f}_{r,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right\}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + \bar{L}_n + O_P(n^{1/2}r^{-1})\right\}.$$

When $\lim_{n\to\infty} n/r = \lambda_1$ exists and is finite, Proposition A.6 implies

$$\sqrt{n}(\hat{\theta}_D - \theta_o) \xrightarrow{D} N(0, \Omega[\Omega^{-1} + \lambda_1 V]\Omega).$$

On the other hand, when $\lim_{n\to\infty} n/r = \infty$,

$$\sqrt{r}(\hat{\theta}_D - \theta_o)$$
$$= \sqrt{r/n}\cdot\sqrt{n}(\hat{\theta}_D - \theta_o)$$
$$= -\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln \tilde{f}_{r,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right\}^{-1}\left\{\sqrt{r/n}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta} + \sqrt{r/n}\bar{L}_n + O_P(r^{-1/2})\right\}$$
$$= -\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln \tilde{f}_{r,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'}\right\}^{-1}\left\{\sqrt{r/n}\bar{L}_n + o_P(1)\right\}$$
$$\xrightarrow{D} N(0, \Omega V \Omega),$$

where the third equality follows because $\lim_{n\to\infty} r/n = 0$ and $\lim_{n\to\infty} r = \infty$. Q.E.D.

Proof of Theorem 6:

Define $\nu(\theta, x_i) = \frac{1}{r}\mu(\theta, x_i)$ where

$$\mu(\theta, x_i) = \sum_{l=1}^{L}\frac{d_{li}}{P^2(l|\theta_o, x_i)}\left[\frac{\partial \ln P(l|\theta_o, x_i)}{\partial\theta}\mathrm{var}(h_l(v, x_i)|x_i) - \mathrm{cov}(h_l(v, x_i), \frac{\partial h_l(v, x_i)}{\partial\theta}|x_i)\right].$$

Consider each terms separately in the decomposition:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_r(\hat{\theta}_L, x_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu(\theta_o, x_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\nu_r(\theta_o, x_i) - \nu(\theta_o, x_i)] + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\nu_r(\hat{\theta}_L, x_i) - \nu_r(\theta_o, x_i)].$$

Since $E(\nu(\theta_o, x)) = r^{-1}\bar{\mu}$, it follows that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu(\theta_o, x_i) = \frac{1}{r\sqrt{n}}\sum_{i=1}^{n}(\mu_i - \bar{\mu}) + n^{1/2}r^{-1}\bar{\mu} = n^{1/2}r^{-1}\bar{\mu} + O(r^{-1}).$$

Proposition A.5 implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\nu_r(\theta_o, x_i) - \nu(\theta_o, x_i)] = r^{-1}n^{1/2}\cdot\frac{1}{n}\sum_{i=1}^{n}[\mu_r(\theta_o, x_i) - \mu(\theta_o, x_i)] = O_P(\max[r^{-3/2}, n^{1/2}r^{-2}]).$$

By a Taylor expansion,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\nu_r(\hat{\theta}_L, x_i) - \nu_r(\theta_o, x_i)] = r^{-1}n^{1/2}(\hat{\theta}_L - \theta_o)\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mu_r(\bar{\theta}, x_i)}{\partial\theta'} = O_P(\max[r^{-1}, n^{1/2}r^{-2}]),$$

because $\sqrt{n}(\hat{\theta}_L - \theta_o) = O_P(\max[1, n^{1/2}r^{-1}])$ in $(B.6)$. Therefore,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_r(\hat{\theta}_L, x_i) = n^{1/2}r^{-1}\bar{\mu} + O_P(\max[r^{-1}, n^{1/2}r^{-2}]) = O_P(n^{1/2}r^{-1}). \qquad (B.9)$$

Denote

$$T_{1,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta'} - E(\sum_{l=1}^{L} d_l\frac{\partial^2 \ln P(l|\theta_o, x)}{\partial\theta\partial\theta'}),$$

$$T_{2,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\theta_o, x_i)}{\partial\theta\partial\theta'} - \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta'},$$

and

$$T_{3,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial\theta\partial\theta'} - \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\theta_o, x_i)}{\partial\theta\partial\theta'}.$$

It is obvious that $T_{1,n} = O_P(n^{-1/2})$. Proposition A.5 implies that $T_{2,n} = O_P(\max[n^{-1/2}r^{-1/2}, r^{-1}])$. By the mean value theorem and $(B.6)$,

$$T_{3,n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\sum_{k=1}^{K}\frac{\partial^3 \ln f_{r,l}(\bar{\theta}, x_i)}{\partial\theta\partial\theta'\theta_k}(\hat{\theta}_{L,k} - \theta_{o,k}) = O_P(\max[n^{-1/2}, r^{-1}]).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial\theta\partial\theta'} - E(\sum_{l=1}^{L} d_l\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta'}) = T_{1,n} + T_{2,n} + T_{3,n} = O_P(\max[n^{-1/2}, r^{-1}]),$$

and

$$\begin{aligned}
&\Big\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial\theta\partial\theta'}\Big\}^{-1} - \Big\{E(\sum_{l=1}^{L} d_l\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta'})\Big\}^{-1} \\
&= -\Big\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial\theta\partial\theta'}\Big\}^{-1} \\
&\times \Big[\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial\theta\partial\theta'} - E(\sum_{l=1}^{L} d_l\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta'})\Big]\Big\{E(\sum_{l=1}^{L} d_l\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta'})\Big\}^{-1} \\
&= O_P(\max[n^{-1/2}, r^{-1}]).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\Big(\Big\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial\theta\partial\theta'}\Big\}^{-1} - \Big\{E(\sum_{l=1}^{L} d_l\frac{\partial^2 \ln P(l|\theta_o, x_i)}{\partial\theta\partial\theta'})\Big\}^{-1}\Big)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_r(\hat{\theta}_L, x_i) \qquad (B.10) \\
&= O_P(\max[r^{-1}, n^{1/2}r^{-2}]).
\end{aligned}$$

Hence it follows from (B.10) and (B.9) that

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_A - \theta_o) &= \sqrt{n}(\hat{\theta}_L - \theta_o) + \Big\{\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L} d_{li}\frac{\partial^2 \ln f_{r,l}(\hat{\theta}_L, x_i)}{\partial\theta\partial\theta'}\Big\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_r(\hat{\theta}_L, x_i) \\
&= \sqrt{(}\hat{\theta}_L - \theta_o) - \Omega\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nu_r(\hat{\theta}_L, x_i) + O_P(\max[r^{-1}, n^{1/2}r^{-2}]) \\
&= \sqrt{(}\hat{\theta}_L - \theta_o) - \Omega\cdot n^{1/2}r^{-1}\bar{\mu} + O_P(\max[r^{-1}, n^{1/2}r^{-2}]).
\end{aligned}$$

The final result follows from this equation and (3.19) in Theorem 4. The proof of the estimator defined in (4.1)' is similar. Q.E.D.

References

1. Amemiya, T. *Advanced Econometrics*. Harvard University Press, Cambridge, Mass. 1985.
2. Borsch-Supan, A. and V. Haijivassiliou, Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. Cowles Foundation Discussion Paper No. 960, Yale University, 1990.
3. Cox, D.R. and D.V. Hinkley, *Theoretical Statistics*. Princeton: Princeton University Press, 1974.
4. Hajivassiliou, V. and D. McFadden, The debt repayment crises LDC's: estimation by the method of simulated moments. Yale University, Working paper, 1987.
5. Lee, A.J. *U-statistics: theory and practice*. New York: Marcel Dekker Inc., 1990.
6. Lee, L.F. On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. Forthcoming in *Econometric Theory*, 1992.
7. Lehmann, E.L. *Nonparametrics: statistical methods based on ranks*. San Francisco: Holden Day, 1966.
8. Lerman, S., and C. Manski. On the use of simulated frequencies to approximate choice probabilities. In C. Manski and D. McFadden (ed.), *Structural Analysis of Discrete Data with Econometric Applications*, Chapter 7 and pp. 305-319. Cambridge: MIT Press, 1981.
9. McFadden, D. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometirca* 57 (1989): 995-1026.
10. Pakes, A. Patents as options: some estimates of the value of holding european patent stocks. *Econometrica* 54 (1986): 755-785.
11. Pakes, A. and D. Pollard. Simulation and the asymptotics of optimization estimators. *Econometrica* 57 (1989): 1027-1057.
12. Serfling, R.J. *Approximation theorems of mathematical statistics*. New York: Wiley, 1980.
13. Stern, S. A method for smoothing simulated moments of discrete probilities in multivariate probit models. *Econometrica* 60, pp. 943-952, 1992.