Simulation Estimation of Polychotomous-Choice Sample Selection Models

by Lung-fei Lee

Department of Economics

The Hong Kong University of Science and Technology

Kowloon, Hong Kong

June 1998

Abstract

This paper considers simulation estimation of sample selection models. Simulation estimation techniques are useful when sample selection criteria are complicated with polychotomous choice alternatives and correlated disturbances. Simulated likelihood and two-stage estimation methods are considered. For two-stage estimation, possible simulation estimation of outcome regression equations based on Gibbs sampler and variance-reduction techniques are considered. Various generalizations of the GHK simulator for likelihood simulation are possible. Monte Carlo results are provided for finite sample comparison on performance of various simulation methods. Issues on sensitivity of parameter estimates for distributional assumptions are investigated. Semiparametric estimation methods with simulation are also studied in our Monte Carlo experiments.

Key words: Selection bias, discrete choice, simulated moment, Gibbs sampler, variance reduction, simulated likelihood, misspecification, simulated likelihood ratio test, semiparametric estimation, Monte Carlo experiments.

JEL Classification: C15, C34, C35

CORRESPONDENCE ADDRESS:

Lung-fei Lee, Department of Economics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, HONG KONG. Tel: (852) 2358-7600; Fax: (852) 2358-2084; E-mail: lflee@ust.hk

Simulation Estimation of Polychotomous-Choice Sample Selection Models

by Lung-fei Lee*

1. Introduction

This paper considers the estimation of sample selection (SEL) models with polychotomous choices or multiple selection criteria. The focus is on the estimation of models which require estimation via simulation techniques. An example of such models is a model with multivariate normal disturbances. SEL models with polychotomous choices have been considered in Dubin and McFadden (1984), Lee (1983), and others, under extreme value distributions. Schmertmann (1994) discussed limitations of these model specifications. In some empirical studies, in order to achieve simplicity, researchers imposed restrictive and unrealistic assumptions such as independence across disturbances of equations. Such a practice resulted in controversies and debates (e.g., Duan et al. 1984, and Hay and Olsen 1984). In many circumstances with complex selection criteria, multivariate normal disturbances may be the preferred specification (Amemiya 1974, Hausman and Wise 1978,). With the development of simulation estimation methods, complex models can, in principle, be estimated (McFadden 1989, and Pakes and Pollard 1989). Methods of simulated moments (MSM), simulated pseudo-likelihood (SPL), and simulated maximum likelihood (SML) have been proposed in the literature (see McFadden 1989, Pakes and Pollard 1989, Lee 1992, Gouriéroux and Monfort 1993, and Borsch-Supan and Hajivassiliou 1993). Parameter estimates may, however, be sensitive to simulators used (McFadden and Ruud 1994). The GHK simulator introduced by Geweke (1991), Borsch-Supan and Hajivassiliou (1993), and Keane (1994) provides an adequate probability simulator for the multinomial probit model. Monte Carlo results on comparing various simulation methods and simulators have mostly focused on discrete choice (DC) models (Hajivassiliou et al. 1996, Borsch-Supan and Hajivassiliou 1993, and Geweke et al. 1994 and 1997).

This paper considers simulation estimation of SEL models. Two-stage estimation methods can be generalized for the estimation of outcome equations. Both biased and unbiased simulators for observed outcome equations are proposed. The same methods can be used for simulating other conditional outcome equations, such as opportunity costs of unchosen alternatives (Heckman 1990, and Lee 1995). A two-stage simulation estimation method (TSSM) is essentially MSM. SML methods can be more efficient. Various ways to generalize the GHK simulator to SEL models are possible. We investigate the performance of the various SML and TSSM methods in their finite sample performance by Monte Carlo studies. The studies provide

* Financial support from the Research Grant Council of Hong Kong under grant no. HKUST595/96H for my research is gratefully acknowledged.

numerical evidence on the feasibility and accuracy of various simulation approaches for the estimation of SEL models. Estimates of parameters in limited dependent variable (LDV) models may be sensitive to distributional specifications (see, e.g., surveys by Maddala 1983 and Amemiya 1984). A polychotomous-choice sample selection (P-SEL) model will be subject to the same criticism. We provide Monte Carlo results to demonstrate how sensitive the SML estimators (SMLEs) will be when the disturbances are skew-distributed or have bimodal distributions. Finally, we investigate a possible semiparametric MLE method originated by Gallant and Nychka (1987). The likelihood function of the semiparametric P-SEL model can be simulated. The simulated semiparametric likelihood (SSL) function can be used for estimation and the construction of specification error test of normality. Monte Carlo experiments are designed to investigate the finite sample properties of the semiparametric estimates. The finite sample levels of significance and power of SML tests are also investigated.

2. Sample Selection Models, Selected Outcomes, and Opportunity Costs

There are several broadly applicable SEL models with different choice processes or selection criteria.

A general P-SEL model with m alternatives and a continuous outcome equation is

$$y = x\beta + \epsilon, \quad (2.1)$$

and

$$U_j = z_j\gamma + u_j, \quad j = 1, \dots, m, \quad (2.2)$$

where the subscript j refers to the alternative j for an individual. A subscript i for an individual has been suppressed. The U_j represents the utility of the alternative j , and the outcome y can be observed only if the first alternative is chosen by the individual. In a utility maximization framework, the alternative j will be chosen if $U_j > U_l$ for all $l \neq j$, $l = 1, \dots, m$. As a normalization on the utility function, $U_m = 0$ for the alternative m for all individuals. Let $I_j = 1$ if the alternative j is chosen by the individual; $I_j = 0$, otherwise. A popular but theoretically restrictive specification of the choice component assumes that u s in (2.2) are i.i.d. Gumbel distributed. This specification results in logistic choice probabilities. A more general specification may assume normally distributed and correlated disturbances. For the latter specification, simulation estimation techniques are required, because implied choice probabilities involve multiple integrals which do not have a closed form expression.

Another SEL model allows multiple selection criteria. In such a model, an outcome can be observed only if several selection criteria are satisfied. Let $U_j = z_j\gamma + u_j$, $j = 1, \dots, m$, be m selection criteria, and

let $y = x\beta + \epsilon$ be an outcome equation. The outcome y for an individual can be observed only if $U_j > 0$ for all $j = 1, \dots, m$. This model can also be a sequential decision model with selectivity if the decision process is sequential. A simple but restrictive specification imposes an independence assumption on distributions of u s. More general specifications with correlated disturbances, such as those with a multivariate normal distribution, will require simulation estimation.

Simulation techniques can be useful for estimating unknown parameters in these models. In addition, one may be interested in the computation of observed or potential outcome equations conditional on choice status. Heckman (1990) and Lee (1995) discuss opportunity costs for unchosen alternatives in SEL models. In a polychotomous choice model, the expected outcome equation is $E(y|x, I_1 = 1) = x\beta + E(\epsilon|x, U_1 > U_l, l = 2, \dots, m)$. In a multiple selection criteria model, the conditional outcome equation is $E(y|x, U_1 > 0, \dots, U_m > 0) = x\beta + E(\epsilon|x, U_1 > 0, \dots, U_m > 0)$. For an individual who has chosen the alternative l , the expected opportunity cost for the potential outcome of the alternative 1 is $E(y|x, I_l = 1) = x\beta + E(\epsilon|x, U_l > U_r, r = 1, \dots, l-1, l+1, \dots, m)$. These equations under a multivariate distribution, such as the multivariate normal distribution, for disturbances may involve multiple integrals which need to be evaluated by simulation techniques.

3. Two-stage Estimation with Simulation

Unknown parameters of the SEL model can be estimated by two-stage methods. With simulation, a TSSM is similar to the MSM. Unknown parameters and regressors of a model are suppressed for simplicity in the following presentations if that does not lead to confusion. The choice model that will be explicitly analyzed is the polychotomous choice model. Generalizations of simulation methods to models with multiple selection criteria are similar. In the first stage, the DC equations in (2.2) are estimated. The likelihood function of the DC model is

$$L(\bar{I}) = \prod_{j=1}^m \left[\int g(\bar{U}_{m-1}, I_j = 1) d\bar{U}_{m-1} \right]^{I_j}, \quad (3.1)$$

where $\bar{U}_{m-1} = (U_1, \dots, U_{m-1})$ and $\bar{I} = (I_1, \dots, I_m)$. This likelihood function can be simulated with the GHK probability simulator. The probabilities of $U_j - U_l > 0$ for $l = 1, \dots, m$ but $l \neq j$, are determined by the multivariate normal distribution of $v_j = (u_1 - u_j, \dots, u_{j-1} - u_j, u_{j+1} - u_j, \dots, u_m - u_j)$. Denote $w_j = (z_j - z_1, \dots, z_j - z_{j-1}, z_j - z_{j+1}, \dots, z_j - z_m)$. The variance matrix of v_j can be decomposed by the Cholesky decomposition $H_j H_j'$ where H_j is a lower triangular matrix. It follows that $v_{jl} = \sum_{k=1}^l h_{jlk} \eta_{jk}$, where v_{jl} is the l th component of v_j , h_{jlk} is the (l, k) entry of H_j , and $\eta_j = (\eta_{j1}, \dots, \eta_{j,m-1})$ is a standard

normal vector (with zero mean and an identity variance matrix), i.e., $v_j = H_j \eta_j$. Define $L_{j1} = w_{j1} \gamma / h_{j11}$ and $L_{jl} = [w_{jl} \gamma - \sum_{k=1}^{l-1} h_{jlk} \eta_k] / h_{jll}$ for $l = 2, \dots, m-1$. It follows that

$$\begin{aligned} P(I_j = 1) &= \int_{-\infty}^{L_{j,m-1}} \cdots \int_{-\infty}^{L_{j1}} \prod_{l=1}^{m-1} \phi(\eta_{jl}) d\eta_{jl} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{l=1}^{m-1} \Phi(L_{jl}) \phi_{(-\infty, L_{jl}]}(\eta_{jl}) d\eta_{jl}, \end{aligned} \quad (3.2)$$

where ϕ and Φ are, respectively, the standard normal density and distribution, and $\phi_{(a,b]}$ is a truncated standard normal density with support on $(a, b]$. The GHK sampler is to generate sequentially truncated standard normal variables from $\prod_{l=1}^{m-1} \phi_{(-\infty, L_{jl}]}(\eta_{jl})$. With S simulation runs, the GHK likelihood simulator of the DC model is

$$\hat{L}_S^C(\bar{I}) = \prod_{j=1}^m \left\{ \frac{1}{S} \sum_{s=1}^S \prod_{l=1}^{m-1} \Phi(L_{jl}^{(s)}) \right\}^{I_j}. \quad (3.3)$$

The parameter γ of (2.2) can be estimated by the maximum likelihood (ML) method applied to the simulated likelihood (SL) function (3.3).

With the first-stage estimate $\hat{\gamma}$, the outcome equation can be estimated in the second stage by MSM. Simulated moment equations can be derived from observed outcome equations. TSSM may be of special interest for the case where $E(\epsilon | \bar{U}_{m-1}) = \delta \psi(\bar{U}_{m-1}, \gamma)$ for some function $\psi(\bar{U}_{m-1}, \gamma)$ involving only parameters of the DC equations. Unknown parameters of the observed outcome equation,

$$E(y | I_1 = 1) = x\beta + \delta E(\psi(\bar{U}_{m-1}, \gamma) | I_1 = 1), \quad (3.4)$$

can be estimated by an instrumental variable (IV) procedure after $E(\psi(\bar{U}_{m-1}, \gamma) | I_1 = 1)$ is replaced by a simulated function $\hat{E}_S(\psi(\bar{U}_{m-1}, \gamma))$ evaluated at $\hat{\gamma}$. Simulators of $E(\psi(\bar{U}_{m-1}, \gamma) | I_1 = 1)$ can be derived in various ways. We consider first the recursive simulation technique underlying the GHK simulator. This recursive simulation technique does not provide an unbiased simulator. Unbiased simulators are then derived. Variance reduction based on conditioning may also be useful.

Consider the estimation of $E(y | I_1 = 1) = x\beta + E(\epsilon | I_1 = 1)$ given values of unknown parameters and x . Let $f(\epsilon | v_1)$ be the conditional density of ϵ conditional on v_1 , where v_1 has the Cholesky decomposition

$v_1 = H_1 \eta_1$. It follows that

$$\begin{aligned}
E(\epsilon|I_1 = 1) &= \frac{1}{P(I_1 = 1)} \int_{-\infty}^{\infty} \int_{-\infty}^{L_{1,m-1}} \cdots \int_{-\infty}^{L_{11}} \epsilon f(\epsilon|v_1 = H_1 \eta_1) \prod_{l=1}^{m-1} \phi(\eta_{1l}) d\eta_{1l} \cdot d\epsilon \\
&= \frac{1}{P(I_1 = 1)} \int_{-\infty}^{L_{1,m-1}} \cdots \int_{-\infty}^{L_{11}} E(\epsilon|v_1 = H_1 \eta_1) \prod_{l=1}^{m-1} \phi(\eta_{1l}) d\eta_{1l} \\
&= \frac{1}{P(I_1 = 1)} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} E(\epsilon|v_1 = H_1 \eta_1) \prod_{l=1}^{m-1} \Phi(L_{1l}) \phi_{[-\infty, L_{1l}]}(\eta_{1l}) d\eta_{1l}.
\end{aligned} \tag{3.5}$$

The GHK sampler $\prod_{l=1}^{m-1} \phi_{[-\infty, L_{1l}]}(\eta_{1l})$ can be used simultaneously to simulate both the numerator and denominator in (3.5). With S simulation runs from the GHK sampler, (3.5) can be simulated as

$$\begin{aligned}
\hat{E}_S(\epsilon|I_1 = 1) &= \frac{1}{S} \sum_{s=1}^S E(\epsilon|v_1 = H_1 \eta_1^{(s)}) \prod_{l=1}^{m-1} \Phi(L_{1l}^{(s)}) / \frac{1}{S} \sum_{s=1}^S \prod_{l=1}^{m-1} \Phi(L_{1l}^{(s)}) \\
&= \sum_{s=1}^S E(\epsilon|v_1 = H_1 \eta_1^{(s)}) \omega^{(s)},
\end{aligned} \tag{3.6}$$

where $\omega^{(s)} = \prod_{l=1}^{m-1} \Phi(L_{1l}^{(s)}) / \sum_{r=1}^S \prod_{l=1}^{m-1} \Phi(L_{1l}^{(r)})$. $\hat{E}_S(\epsilon|I_1 = 1)$ is a consistent estimator of $E(\epsilon|I_1 = 1)$ as S goes to infinity. But, as a ratio of simulated functions, this simulator is biased when S is fixed and finite.

If we are interested in unbiased simulators for $E(\epsilon|I_1 = 1)$, Gibbs sampler (GS) may be useful (Geman and Geman 1984, and Tanner and Wong 1987). Since $E(\epsilon|I_1 = 1) = \int \int \epsilon f(\epsilon, \bar{U}_{m-1}|I_1 = 1) d\epsilon d\bar{U}_{m-1}$, the density $f(\epsilon, \bar{U}_{m-1}|I_1 = 1)$ of ϵ and $\bar{U}_{m-1}$ conditional on $I_1 = 1$ is naturally a density for simulation draws. An attractive approach for drawing random variables from $f(\epsilon, \bar{U}_{m-1}|I_1 = 1)$ is GS. GS is based on random variables drawn from full conditional marginal densities $f(\epsilon|I_1 = 1, \bar{U}_{m-1})$ and $g(U_j|I_1 = 1, \epsilon, \bar{U}_{(m-1, -j)})$ where $\bar{U}_{(m-1, -j)}$ denotes the subvector of $\bar{U}_{(m-1)}$ with U_j deleted. GS is based on Markov chain simulation. It is a popular approach for drawing random variables from a posterior density in the Bayesian methodology (see, for examples, Albert and Chib 1993, Geweke, Keane and Runkle 1994 and 1997, and McCulloch and Rossi 1994). It can also be useful in the sampling approach (Hajivassiliou, McFadden and Ruud 1996). With $(\epsilon^{(s)}, \bar{U}_{m-1}^{(s)})$ drawn from $f(\epsilon, \bar{U}_{m-1}|I_1 = 1)$, a simulator for $E(\epsilon|I_1 = 1)$ is

$$\hat{E}_{1,S}(\epsilon|I_1 = 1) = \frac{1}{S} \sum_{s=1}^S \epsilon^{(s)}. \tag{3.7}$$

This simulator is unbiased as it is a sample average of random draws.

It is possible to have unbiased simulators with smaller variances than that in (3.7) under some circumstances. Because $E(\epsilon|I_1 = 1, \bar{U}_{m-1}) = E(\epsilon|\bar{U}_{m-1})$, $E(\epsilon|I_1 = 1) = \int E(\epsilon|\bar{U}_{m-1}) g(\bar{U}_{m-1}|I_1 = 1) d\bar{U}_{m-1}$. If $\bar{U}_{m-1}$ can be drawn from $g(\bar{U}_{m-1}|I_1 = 1)$ and $E(\epsilon|I_1 = 1)$ is computationally tractable, a simulator will be

$$\hat{E}_{2,S}(\epsilon|I_1 = 1) = \frac{1}{S} \sum_{s=1}^S E(\epsilon|\bar{U}_{m-1}^{(s)}). \tag{3.8}$$

GS is a feasible approach if random variables can be drawn from full conditional marginal densities of $g(\bar{U}_{m-1}|I_1 = 1)$. The simulator in (3.8) is unbiased. It has a smaller simulation or Monte Carlo variance (MCV) than that of (3.7). This can be shown as follows. For any two random vectors w and z , the identity $\text{var}(w) = E[\text{var}(w|z)] + \text{var}[E(w|z)]$ implies that $\text{var}[E(w|z)] \leq \text{var}(w)$. The MCV of $\hat{E}_{1,S}(\epsilon|I_1 = 1)$ is $\text{var}[\epsilon|I_1 = 1]$ and that of $\hat{E}_{2,S}(\epsilon|I_1 = 1)$ is $\text{var}[E(\epsilon|I_1 = 1, \bar{U}_{m-1})]$. Because $\text{var}[E(\epsilon|I_1 = 1, \bar{U}_{m-1})] \leq \text{var}(\epsilon|I_1 = 1)$, $\hat{E}_{2,S}(\epsilon|I_1 = 1)$ has a smaller MCV than that of $\hat{E}_{1,S}(\epsilon|I_1 = 1)$.

Further variance reduction of simulators may be feasible. The conditional expectation can be rewritten as $E(\epsilon|I_1 = 1) = \int E[\epsilon|I_1 = 1, \bar{U}_{(m-1,-j)}]g(\bar{U}_{(m-1,-j)}|I_1 = 1)d\bar{U}_{(m-1,-j)}$ for $j = 1, \dots, m-1$. With $j = 1$,

$$\begin{aligned} E[\epsilon|I_1 = 1, \bar{U}_{(m-1,-1)}] &= E[\epsilon|\bar{U}_{(m-1,-1)}, U_1 > \bar{U}_{(m-1,-1)}] \\ &= \int E(\epsilon|\bar{U}_{m-1})g(U_1|\bar{U}_{(m-1,-1)}, U_1 > \bar{U}_{(m-1,-1)})dU_1. \end{aligned}$$

This function can be computationally tractable for some models as it involves only a single integration of $E(\epsilon|\bar{U}_{m-1})$ with respect to a univariate truncated conditional density of U_1 . Similarly, for $j = 2, \dots, m-1$,

$$E[\epsilon|I_1 = 1, \bar{U}_{(m-1,-j)}] = \int E(\epsilon|\bar{U}_{m-1})g(U_j|\bar{U}_{(m-1,-j)}, U_1 > U_j)dU_j.$$

These suggest that any of the following simulators can be an unbiased simulator of $E(\epsilon|I_1 = 1)$:

$$\hat{E}_{3,S}^{(j)}(\epsilon|I_1 = 1) = \frac{1}{S} \sum_{s=1}^S E[\epsilon|I_1 = 1, \bar{U}_{(m-1,-j)}^{(s)}], \quad j = 1, \dots, m-1, \quad (3.9)$$

where $\bar{U}_{m-1}^{(s)}$, $s = 1, \dots, S$ represent random draws from $g(\bar{U}_{m-1}|I_1 = 1)$. Each of these simulators can have smaller MCV than that of $\hat{E}_{2,S}(\epsilon|I_1 = 1)$. This is so, because

$$E(\epsilon|I_1 = 1, \bar{U}_{(m-1,-j)}) = E[E(\epsilon|\bar{U}_{m-1})|I_1 = 1, \bar{U}_{(m-1,-j)}],$$

therefore $\text{var}[E(\epsilon|I_1 = 1, \bar{U}_{(m-1,-j)})] \leq \text{var}[E(\epsilon|\bar{U}_{m-1})]$.

REMARKS: The $m-1$ simulators can be combined into a weighted average. Let $\omega = (\omega_1, \dots, \omega_{m-1})$, where $\omega_j \geq 0$ for all $j = 1, \dots, m-1$ and $\sum_{j=1}^{m-1} \omega_j = 1$, be a vector of weights. A weighted simulator with ω is

$$\hat{E}_{\omega,S}(\epsilon|I_1 = 1) = \sum_{j=1}^{m-1} \omega_j \hat{E}_{3,S}^{(j)}(\epsilon|I_1 = 1).$$

This weighted simulator with arbitrary weights can have smaller MCV than that of $\hat{E}_{2,S}(\epsilon|I_1 = 1)$. For any random variable z with finite second moment, $E^2(z) \leq E(z^2)$. This inequality implies that

$$\left[\sum_{j=1}^{m-1} \omega_j \hat{E}_{3,S}^{(j)}(\epsilon|I_1 = 1) - E(\epsilon|I_1 = 1) \right]^2 \leq \sum_{j=1}^{m-1} \omega_j [\hat{E}_{3,S}^{(j)}(\epsilon|I_1 = 1) - E(\epsilon|I_1 = 1)]^2,$$

hence,

$$E \left[\sum_{j=1}^{m-1} \omega_j \hat{E}_{3,S}^{(j)}(\epsilon|I_1=1) - E(\epsilon|I_1=1) \right]^2 \leq \sum_{j=1}^{m-1} \omega_j E[\hat{E}_{3,S}^{(j)}(\epsilon|I_1=1) - E(\epsilon|I_1=1)]^2 \\ \leq \text{var}[\hat{E}_{2,S}(\epsilon|I_1=1)],$$

that is, $\text{var}[\hat{E}_{\omega,S}(\epsilon|I_1=1)] \leq \text{var}[\hat{E}_{2,S}(\epsilon|I_1=1)]$. From these inequalities, it can be seen that the weighted simulator can have smaller MCV than some $\hat{E}_{3,S}^{(j)}(\epsilon|I_1=1)$, but it may not have smaller MCV than all of them. The latter issue is similar to the estimation of a common mean when samples are not i.i.d. A simulator that takes into account possible heteroskedastic variances and autocorrelations among $\hat{E}_{3,S}^{(j)}(\epsilon|I_1=1)$ for $j = 1, \dots, m-1$ by a proper generalized least squares formulation can be relatively efficient. The latter will be computationally involved as heteroskedastic variances and autocovariances need to be simulated independently of $\hat{E}_{3,S}^{(j)}(\epsilon|I_1=1)$ to maintain the unbiasedness property of the generalized least squares simulator.

With a simulated bias-correction term $\hat{E}_S(\psi(\bar{U}_{m-1}, \gamma))$ evaluated at $\hat{\gamma}$, the TSSM can estimate the outcome equation (3.4). Let $\hat{z}_S = (x, \hat{E}_S(\psi(\bar{U}_{m-1}, \hat{\gamma})))$ and w be a vector of IVs. Then $\alpha = (\beta', \delta')'$ can be estimated as

$$\hat{\alpha} = \left[\sum_{i=1}^n I_{1i} w_i \hat{z}_{S,i}' \right]^{-1} \sum_{i=1}^n I_{1i} w_i y_i, \quad (3.10)$$

where the subscript i refers to an individual and n is the sample size. IVs for the bias-correction term can be derived by using another independent simulator of this term. This TSSM is a simulated IV (SIV) procedure. The TSSM estimates (TSSMEs) with any simulators in (3.6)-(3.9) are consistent as both n and S go to infinity. The TSSMEs with any unbiased simulators in (3.7)-(3.9) are consistent for a fixed finite S . The TSSME with the recursive simulator (3.6) is consistent only when S goes to infinity. The recursive simulation approach with a finite fixed S can still be useful if the following moment equation implied by the model is used for estimation:

$$I_1 y = I_1 x \beta + \delta E(I_1 \psi(\bar{U}_{m-1}, \gamma)) + \epsilon_1, \quad (3.11)$$

where ϵ_1 is a disturbance term with zero mean. The GHK sampler can provide an unbiased simulator for $E(I_1 \psi(\bar{U}_{m-1}, \gamma))$ and x can be used as an IV for $I_1 x$. Asymptotic distribution and variance of these TSSMEs can be derived as in the MSM approach (McFadden 1989, and Pakes and Pollard 1989) and by a Taylor expansion taking into account the influence of the asymptotic distribution of $\hat{\gamma}$ (Amemiya 1978).

4. Simulated Maximum Likelihood Estimation

Instead of the TSSM method, SML methods can be constructed to estimate the whole model. The likelihood function for an observation in the SEL model is

$$L(x, \bar{I}, y) = \left[\int f(\epsilon, \bar{U}_{m-1}, I_1 = 1) d\bar{U}_{m-1} \right]^{I_1} \prod_{j=2}^m \left[\int g(\bar{U}_{m-1}, I_j = 1) d\bar{U}_{m-1} \right]^{I_j}, \quad (4.1)$$

where $\epsilon = y - x\beta$. It can be simulated without bias by simulating each likelihood components. The GHK simulator of choice probability can be generalized for the simulation. There are two possible ways to simulate the likelihood function in (4.1).

Based on the factorization $f(\epsilon, \bar{U}_{m-1}, I_1 = 1) = f(\epsilon | \bar{U}_{m-1}) g(\bar{U}_{m-1}, I_1 = 1)$,

$$\begin{aligned} \int f(\epsilon, \bar{U}_{m-1}, I_1 = 1) d\bar{U}_{m-1} &= \int_{-\infty}^{L_{1,m-1}} \cdots \int_{-\infty}^{L_{11}} f(\epsilon | v_1 = H_1 \eta_1) \prod_{l=1}^{m-1} \phi(\eta_{1l}) d\eta_{1l} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\epsilon | v_1 = H_1 \eta_1) \prod_{l=1}^{m-1} \Phi(L_{1l}) \phi_{(-\infty, L_{1l}]}(\eta_{1l}) d\eta_{1l}. \end{aligned} \quad (4.2)$$

The likelihood function in (4.1) can thus be simulated as

$$\tilde{L}_S(\bar{I}, y) = \left\{ \frac{1}{S} \sum_{s=1}^S f(\epsilon | v_1 = H_1 \eta_1^{(s)}) \prod_{l=1}^{m-1} \Phi(L_{1l}^{(s)}) \right\}^{I_1} \prod_{j=2}^m \left\{ \frac{1}{S} \sum_{s=1}^S \prod_{l=1}^{m-1} \Phi(L_{jl}^{(s)}) \right\}^{I_j}, \quad (4.3)$$

where the random variables are drawn from $\prod_{l=1}^{m-1} \phi_{(-\infty, L_{jl}]}(\eta_{jl})$ for each $j = 1, \dots, m$.

The second approach depends on the factorization $f(\epsilon, \bar{U}_{m-1}, I_1) = g(\bar{U}_{m-1}, I_1 | \epsilon) f(\epsilon)$. Let CC' be the Cholesky decomposition of the variance matrix of (ϵ, v'_1) . It follows that $(\epsilon, v'_1)' = C\zeta$ where ζ is an m -dimensional standard normal vector. With this Cholesky representation, $\epsilon = C_{11}\zeta_1$ and $v_{1k} = \sum_{l=1}^{k+1} C_{k+1,l}\zeta_l$ for each $k = 1, \dots, m-1$. Define $K_k = [w_{1k}\gamma - \sum_{l=1}^k C_{k+1,l}\zeta_l] / C_{k+1,k+1}$ for $k = 1, \dots, m-1$. It follows that

$$\begin{aligned} \int f(\epsilon, \bar{U}_{m-1}, I_1) d\bar{U}_{m-1} &= \int g(\bar{U}_{m-1}, I_1 | \epsilon) d\bar{U}_{m-1} \cdot f(\epsilon) \\ &= \int_{-\infty}^{K_{m-1}} \cdots \int_{-\infty}^{K_1} \prod_{l=2}^m \phi(\zeta_l) d\zeta_l \cdot f(\epsilon) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{l=1}^{m-1} \Phi(K_l) \phi_{(-\infty, K_l]}(\zeta_{l+1}) d\zeta_{l+1} \cdot f(\epsilon), \end{aligned} \quad (4.4)$$

where $\zeta_1 = \epsilon / C_{11}$. With S simulation runs from $\prod_{l=1}^{m-1} \phi_{(-\infty, K_l]}(\zeta_{l+1})$ for $j = 1, \dots, m$, (4.4) can be simulated and the SL is

$$\hat{L}_S(\bar{I}, y) = \left\{ \frac{1}{S} \sum_{s=1}^S \prod_{l=1}^{m-1} \Phi(K_l^{(s)}) f(\epsilon) \right\}^{I_1} \prod_{j=2}^m \left\{ \frac{1}{S} \sum_{s=1}^S \prod_{l=1}^{m-1} \Phi(L_{jl}^{(s)}) \right\}^{I_j}. \quad (4.5)$$

Asymptotic properties of the SMLE can be found in Lee (1992) and Gouriéroux and Monfort (1993), among others. The difference of the two SLs in (4.3) and (4.5) is on the likelihood components of observed

outcomes. Accuracy of these SLs in a finite sample will be compared by Monte Carlo studies. The variance of a log SL is due to two sources of variation – simulation and sample. An SL is a function of sample observation z and simulated draws ξ , say, $\hat{L}^S(z, \xi)$. The variance of $\hat{L}^S$ is a sum of two terms

$$\text{var}(\hat{L}^S(z, \xi)) = E[\text{var}(\hat{L}^S(z, \xi)|z)] + \text{var}[E(\hat{L}^S(z, \xi)|z)]. \quad (4.6)$$

The $\text{var}(\hat{L}^S(y, \xi)|z)$ is the MCV conditional on a sample. The first term of (4.6) is the mean of MCVs over sample. For an unbiased SL, the second term captures the sample variation of the likelihood function. As the second term is the same for all unbiased SLs, the comparison of unbiased SLs can be based on the average of MCVs.

5. Semiparametric Maximum Likelihood Estimation

For the estimation of LDV models, because a parametric distribution will impose specific probability and regression functions of choices and outcomes, estimates of parameters may not be robust to a distributional assumption. A large body of literature on semiparametric methods exists. In this section, we consider the semiparametric ML (SEML) method by Gallant and Nychka (1987). Related two-stage methods for the estimation of outcome equations can be found in Lee (1982) and Newley et al.(1990). The SEML method is based on a modified Hermite expansion of an unknown multivariate density around a multivariate normal density. Consider a scalar random variable w first. Let $f(w)$ be the (unknown) density of w . Gallant and Nychka (1987) have shown that, under some regularity conditions, f can be expanded with respect to a normal density of $N(\mu, \sigma^2)$ as

$$f(w) = \left[\sum_{j=0}^{\infty} c_j H_j \left(\frac{w - \mu}{\sigma} \right) \right]^2 \frac{1}{\sigma} \phi \left(\frac{w - \mu}{\sigma} \right) / \sum_{j=0}^{\infty} \frac{c_j^2}{j!}, \quad (5.1)$$

where $c_0 = 1$ and H_j s are the univariate Hermite polynomials where

$$H_j(w) = w^j - \frac{j^{[2]}}{2 \cdot 1!} w^{j-2} + \frac{j^{[4]}}{2^2 \cdot 2!} w^{j-4} - \frac{j^{[6]}}{2^3 \cdot 3!} w^{j-6} + \dots, \quad (5.2)$$

and $j^{[r]} = j(j-1) \cdots (j-(r-1))$. In practice, a finite order expansion will be used for estimation. An r th order approximation is

$$f_r(w) = \left[\sum_{j=0}^r c_j H_j \left(\frac{w - \mu}{\sigma} \right) \right]^2 \frac{1}{\sigma} \phi \left(\frac{w - \mu}{\sigma} \right) / \sum_{j=0}^r \frac{c_j^2}{j!}. \quad (5.3)$$

The $f_r(w)$ is a proper density function as its integration with w is a unity. $f_r(w)$ will reduce to a normal density with mean μ and variance σ^2 when $c_j = 0$ for all $j = 1, \dots, r$. If w is an m -dimensional random

vector, its density can be expanded with respect to a multivariate normal density of $N(\mu, \Sigma)$ as

$$f(w) = \frac{\{\sum_{j_1=0}^{\infty} \cdots \sum_{j_m=0}^{\infty} c_{j_1 \cdots j_m} \prod_{l=1}^m H_{j_l}([P^{-1}(w-\mu)]_l)\}^2}{\sum_{i_1=0}^{\infty} \cdots \sum_{i_m=0}^{\infty} c_{i_1 \cdots i_m}^2 \prod_{l=1}^m i_l!} \cdot \frac{|\Sigma|^{-1/2}}{(2\pi)^{m/2}} \exp\{-\frac{1}{2}(w-\mu)'\Sigma^{-1}(w-\mu)\}, \quad (5.4)$$

where $c_{0 \cdots 0} = 1$, $PP' = \Sigma$ and $[b]_l$ denotes the l th component of a vector b . Equivalently, write $w = \mu + P\bar{s}_m$.

The density of $f(\bar{s}_m)$ of $\bar{s}_m$ where $\bar{s}_m = (s_1, \dots, s_m)$ has the expansion

$$f(\bar{s}_m) = S_m(\bar{s}_m) \prod_{l=1}^m \phi(s_l), \quad (5.5)$$

where

$$S_m(\bar{s}_m) = \frac{\{\sum_{j_1=0}^{\infty} \cdots \sum_{j_m=0}^{\infty} c_{j_1 \cdots j_m} \prod_{l=1}^m H_{j_l}(s_l)\}^2}{\sum_{i_1=0}^{\infty} \cdots \sum_{i_m=0}^{\infty} c_{i_1 \cdots i_m}^2 \prod_{l=1}^m i_l!}. \quad (5.6)$$

The marginal density of $\bar{s}_{m-1}$ where $\bar{s}_{m-1} = (s_1, \dots, s_{m-1})$ implied by (5.5) is

$$f(\bar{s}_{m-1}) = S_{m-1}(\bar{s}_{m-1}) \prod_{l=1}^{m-1} \phi(s_l), \quad (5.7)$$

and

$$\begin{aligned} S_{m-1}(\bar{s}_{m-1}) &= \frac{\sum_{j_m=0}^{\infty} j_m! \{\sum_{j_1=0}^{\infty} \cdots \sum_{j_{m-1}=0}^{\infty} c_{j_1 \cdots j_m} \prod_{l=1}^{m-1} H_{j_l}(s_l)\}^2}{\sum_{i_m=0}^{\infty} i_m! \sum_{i_1=0}^{\infty} \cdots \sum_{i_{m-1}=0}^{\infty} c_{i_1 \cdots i_m}^2 \prod_{l=1}^{m-1} i_l!} \\ &= \sum_{j_m=0}^{\infty} \omega_{j_m} \frac{\{\sum_{j_1=0}^{\infty} \cdots \sum_{j_{m-1}=0}^{\infty} c_{j_1 \cdots j_m} \prod_{l=1}^{m-1} H_{j_l}(s_l)\}^2}{\sum_{i_1=0}^{\infty} \cdots \sum_{i_{m-1}=0}^{\infty} c_{i_1 \cdots i_m}^2 \prod_{l=1}^{m-1} i_l!}, \end{aligned} \quad (5.8)$$

and

$$\omega_{j_m} = \frac{j_m! \sum_{j_1=0}^{\infty} \cdots \sum_{j_{m-1}=0}^{\infty} c_{j_1 \cdots j_m} \prod_{l=1}^{m-1} i_l!}{\sum_{i_m=0}^{\infty} i_m! \sum_{i_1=0}^{\infty} \cdots \sum_{i_{m-1}=0}^{\infty} c_{i_1 \cdots i_m}^2 \prod_{l=1}^{m-1} i_l!}, \quad j_m = 0, 1, \dots \quad (5.9)$$

The implied marginal density is a weighted mixture of Hermite expanded densities. It can be a single Hermite expanded density if independence of s_m and $\bar{s}_{m-1}$ is imposed on (5.5).

To expand the densities of disturbances v_j s and (ϵ, v'_1) with Hermite expansion, one may proceed as follows. Let

$$(\bar{u}'_{m-1}, \epsilon)' = B_m \bar{\xi}_m, \quad (5.10)$$

where B_m is a lower diagonal $m \times m$ matrix, and the density of $\bar{\xi}_m$ has a Hermite expansion density in (5.5)-(5.6), i.e., $f(\bar{\xi}_m) = S_m(\bar{\xi}_m) \prod_{l=1}^m \phi(\xi_l)$. Let B_{m-1} be the matrix B_m with its m th row and column deleted. It follows from (5.10) that $\bar{u}_{m-1} = B_{m-1} \bar{\xi}_{m-1}$. The implied marginal density of $\bar{\xi}_{m-1}$ is the one in (5.7)-(5.9), i.e., $f(\bar{\xi}_{m-1}) = S_{m-1}(\bar{\xi}_{m-1}) \prod_{l=1}^{m-1} \phi(\xi_l)$. Let R_j be the transformation matrix such that $R_j \bar{u}_{m-1} = v_j$. Let H_j be the lower diagonal matrix in a Cholesky decomposition $H_j H_j' = R_j B_{m-1} B_{m-1}' R_j'$. Thus $v_j = H_j \eta_j$. The density of η_j can be derived from that of $\bar{\xi}_{m-1}$. The identity $H_j H_j' = R_j B_{m-1} B_{m-1}' R_j'$ implies that

$B_{m-1}^{-1}R_j^{-1}H_j$ is an orthonormal matrix. Therefore, $H_j'R_j'^{-1}B_{m-1}'B_{m-1}^{-1}R_j^{-1}H_j = I$ and $|B_{m-1}^{-1}R_j^{-1}H_j| = 1$.

Since $\eta_j = H_j^{-1}R_jB_{m-1}\bar{\xi}_{m-1}$, it follows that

$$\begin{aligned} f(\eta_j) &= |H_jR_j^{-1}B_{m-1}^{-1}| \cdot S_{m-1}(B_{m-1}^{-1}R_j^{-1}H_j\eta_j) \frac{1}{(2\pi)^{(m-1)/2}} \exp\left\{-\frac{1}{2}\eta_j'H_j'R_j'^{-1}B_{m-1}'B_{m-1}^{-1}R_j^{-1}H_j\eta_j\right\} \\ &= S_{m-1}(B_{m-1}^{-1}R_j^{-1}H_j\eta_j) \prod_{l=1}^{m-1} \phi(\eta_{jl}). \end{aligned} \quad (5.11)$$

This implies that

$$\begin{aligned} P(I_j = 1) &= \int_{-\infty}^{L_{j,m-1}} \cdots \int_{-\infty}^{L_{j1}} S_{m-1}(B_{m-1}^{-1}R_j^{-1}H_j\eta_j) \prod_{l=1}^{m-1} \phi(\eta_{jl}) d\eta_{jl} \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} S_{m-1}(B_{m-1}^{-1}R_j^{-1}H_j\eta_j) \prod_{l=1}^{m-1} \Phi(L_{jl}) \phi_{(-\infty, L_{jl}]}(\eta_{jl}) d\eta_{jl}. \end{aligned} \quad (5.12)$$

It remains that $\int f(\epsilon, \bar{U}_{m-1}, I_1 = 1) d\bar{U}_{m-1}$ should be considered with the Hermite expansion. Let Q be the transformation matrix such that $Q(\bar{u}'_{m-1}, \epsilon)' = (\epsilon, v'_1)'$. Let C be the lower diagonal matrix of the Cholesky decomposition $CC' = QB_mB'_mQ'$. Let $(\epsilon, v'_1)' = C\zeta$. It follows that $\zeta = C^{-1}QB_m\bar{\xi}_m$ and

$$f(\zeta) = S_m(B_m^{-1}Q^{-1}C\zeta) \prod_{l=1}^m \phi(\zeta_l). \quad (5.13)$$

This implies that

$$\begin{aligned} \int f(\epsilon, \bar{U}_{m-1}, I_1 = 1) d\bar{U}_{m-1} &= \int_{-\infty}^{K_{m-1}} \cdots \int_{-\infty}^{K_1} S_m(B_m^{-1}Q^{-1}C\zeta) \prod_{l=2}^m \phi(\zeta_l) d\zeta_l \cdot \frac{1}{C_{11}} \phi\left(\frac{\epsilon}{C_{11}}\right) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} S_m(B_m^{-1}Q^{-1}C\zeta) \prod_{l=1}^{m-1} \Phi(K_l) \phi_{(-\infty, K_l]}(\zeta_{l+1}) d\zeta_{l+1} \cdot \frac{1}{C_{11}} \phi\left(\frac{\epsilon}{C_{11}}\right), \end{aligned} \quad (5.14)$$

where $\zeta_1 = \epsilon/C_{11}$.

The GHK samplers $\prod_{l=1}^{m-1} \phi_{(-\infty, L_{jl}]}(\eta_{jl})$ and $\prod_{l=1}^{m-1} \phi_{(-\infty, K_l]}(\zeta_{l+1})$ can be used for simulating the likelihood in (4.1) as

$$\begin{aligned} \tilde{L}_S(\bar{I}, y) &= \left\{ \frac{1}{S} \sum_{s=1}^S S_m(B_m^{-1}Q^{-1}C\zeta^{(s)}) \prod_{l=1}^{m-1} \Phi(K_l^{(s)}) \cdot \frac{1}{C_{11}} \phi\left(\frac{\epsilon}{C_{11}}\right) \right\}^{I_1} \\ &\quad \cdot \prod_{j=2}^m \left\{ \frac{1}{S} \sum_{s=1}^S S_{m-1}(B_{m-1}^{-1}R_j^{-1}H_j\eta_j^{(s)}) \prod_{l=1}^{m-1} \Phi(L_{jl}^{(s)}) \right\}^{I_j}. \end{aligned} \quad (5.15)$$

The simulated SEML estimate (SSMLE) can be derived by maximizing the semiparametric SL (SSL) in (5.15) with respect to all the parameters including those of cs in (5.6).

6. Monte Carlo Experiments

This section will report Monte Carlo results on finite sample performance of various simulators.

6.1 Experimental Designs

In the Monte Carlo experiment, a trichotomous-choice SEL model is studied. The utility equations are

$$U_j = z_j\gamma + u_j, \quad j = 1, 2, \quad (6.1)$$

and $U_3 = 0$ as a normalization. The outcome for alternative 1 can be observed when it is chosen, which is

$$y = x\beta + \epsilon. \quad (6.2)$$

u_1 , u_2 and ϵ in (6.1)-(6.2) are jointly normally distributed. To impose the positive-definiteness of the covariance matrix of disturbances, a Cholesky's decomposition can be used: $u_1 = b_{11}\xi_1$, $u_2 = b_{21}\xi_1 + b_{22}\xi_2$, and $\epsilon = \delta_1\xi_1 + \delta_2\xi_2 + \delta_3\xi_3$, where ξ_1 , ξ_2 and ξ_3 are independent standard normal variables. To impose a scale normalization, $b_{11} = 1$ is assumed. In the experiment, the true parameters are $\beta = 1$, $\gamma = 1$, $(b_{21}, b_{22}) = (0.6, 0.8)$, and $(\delta_1, \delta_2, \delta_3) = (1, 0, 1)$. u_1 , u_2 and ϵ are correlated with each other. The implied marginal distributions of u_1 , u_2 and ϵ are, respectively, $N(0, 1)$, $N(0, 1)$ and $N(0, 2)$, and the covariances of u_1 and u_2 , u_1 and ϵ , and u_2 and ϵ are, respectively, 0.6, 1 and 0.6. x is generated from an $N(0, 4/3)$ distribution. This specification implies that the regression equation (6.2) has an $R^2 = 0.4$ for the underlying population before selection. For the utility functions in (6.1), $z_1 = w_1 + w_3$ and $z_2 = w_2 + w_3$ where w s are independently $N(0, 1/2)$ distributed. Thus, z_1 and z_2 are both $N(0, 1)$ but are correlated with a covariance 0.5. In this model (Model-1), x in (6.2) is independent of z_1 and z_2 in (6.1).

To investigate whether possible correlations between x and z_1 and z_2 have important effects on parameter estimates, two additional models will be estimated. In Model-2, x is generated as $x = (4/3)^{1/2}[w_3 + w_4]$ where w_4 is an $N(0, 1/2)$ independent of w_1 , w_2 and w_3 . The resulting x in Model-2 is an $N(0, 4/3)$, as in Model-1, but is correlated with z_1 and z_2 with a 0.5 correlation coefficient for each. To impose stronger correlation, a Model-3 is specified with $x = (4/3)^{1/2}z_1$, which implies that x and z_1 are perfectly correlated and x is correlated with z_2 with a correlation coefficient of 0.5.

The TSSMEs of the outcome equation (6.2) based on the recursive simulator in (3.6) and unbiased GS simulators in (3.8) and (3.9) will be compared. They will also be compared with SMLs based on (4.3) or (4.5). The optimization subroutine used for SML estimation is the DFP algorithm from the package GQOPT version 6.08. The programs are written in fortran and were run in both SUN Sparc and IBM RISC/6000 workstations. Estimates with various sample size n and simulation runs S will be reported. For each case,

the number of replications is 400. There are a few cases with a slightly smaller number of replications because of nonconvergence. For these cases, we report the number of successful convergent replications (con#). Summary statistics on the empirical mean (Mean) and empirical standard deviation (SD) based on the replications are reported. For SMLEs, estimated standard errors (Est.SD) will also be reported. The Est.SD of SMLE is derived by an inverse of the negative numerical second-order derivatives of a log SL from GQOPT. For some estimates, average CPU time in seconds (T) per replication is reported for reference. The reported CPU seconds are based on an IBM RISC/6000 Model 580 single processor workstation.

6.2 Monte Carlo Results on TSSM Estimation

The TSSMEs are reported in Table 2. The first-stage estimates of the DC model (Table 1) are SMLEs derived with the GHK simulator for choice probabilities. The second column of Table 1 reports estimates when unit variance restrictions are imposed on disturbances, i.e., $b(2, 1)^2 + b(2, 2)^2 = 1$. Estimates for the model without this variance restriction are reported in the first column of Table 1. The SMLEs of γ and all parameter estimates in the unrestricted model have small biases. A slightly larger bias in the restricted estimate is observed for the estimate of $b(2, 1)$. There are some slight improvements in the maximized log SL value (F) when S increases from 30 to 100, but the SMLEs are not sensitive to S . As the restriction is correctly imposed, the SDs of restricted estimates are smaller than those of unrestricted estimates. As n increases from 200 to 500, the SDs decrease with a rate on average compatible to the theoretical square-root n rate. These results provide additional Monte Carlo evidence on the good performance of the GHK simulator for the estimation of DC multinomial probit models (Hajivassiliou et al. 1996, Borsch-Supan and Hajivassiliou 1993, Keane 1994, and Geweke et al. 1994).

Table 2 reports the TSSMEs of the outcome equation. Eq.(6.2) implies that $y = x\beta + \delta_1\xi_1 + \delta_2\xi_2 + \delta_3\xi_3$. Conditional on samples with $I_1 = 1$, ξ_1 and ξ_2 are simulated. The SIV method, SIV-1, is based on the recursive simulator (3.6), the SIV-Gibbs-1, is based on $\hat{E}_{2,S}$ of (3.8) and the SIV-Gibbs-2 method is based on $\hat{E}_{3,S}^{(1)}$ of (3.9). IVs for the bias-correction term are independently simulated through a similar procedure in each approach. The IV procedure based on (3.11) by simulating $E(I_1\xi_1)$ and $E(I_1\xi_2)$ by the recursive simulator and using x as IVs for I_1x is reported as SIV-2 in Table 2. Finally, we report an ordinary least squares (SOLS) estimator of the bias-corrected simulated outcome equation conditional on $I_1 = 1$ with the recursive simulator. This SOLS estimator is consistent if S goes to infinity, but would be inconsistent if a constant S were used. The inconsistency comes from the correlation of the simulators' errors with the

simulators themselves in the simulated bias-corrected equation. All the samples before selection are based on $n = 200$. The subsamples for individuals who choose the first alternative are, on average, about one third. Draws S vary from 30 to 100. The first-stage estimate for each case corresponds to the SMLE of DC with the same n and S . For GS approach, the deterministic components, $z_1\hat{\gamma}$ and $z_2\hat{\gamma}$, are used as starting points for the Markov chain simulation. We have done some preliminary inspection on the stability of the Markov chain by monitoring the convergence of several quartiles of simulation runs as suggested in Tanner and Wong (1987), the sampler seems to have rapid convergence after a couple of iterations, as demonstrated in Figure 1. For the results in Table 2, we have tried different Gibbs cycles, c , from 5 to 30. All the TSSMEs of β hardly have any bias. Except for the SOLS approach with small S s, the estimates of δ_1 and δ_2 have very small biases. These estimates are sensitive to neither S nor c . The estimates of δ_1 from the SOLS approach show some biases with $S = 30$ or 60 . As we have pointed out before, the SOLS approach will not give consistent estimates with a fixed finite S , which translates into small S in this context. For the SIV-1 approach, even though the simulator for the bias correction term is biased, the Monte Carlo results show that small $S = 30$ does not create a substantial bias for these TSSMEs. Estimates from the SIV-2 approach have the largest SD among all five approaches. The estimates from the SIV-1, SIV-Gibbs-1 and SIV-Gibbs-2 have similar SDs. But the relative computing time for SIV-1 can be much less than those of SIV-Gibbs-1 and SIV-Gibbs-2. Among the TSSMEs, the SIV-1 approach is preferable.

6.3 Monte Carlo Results on SML Estimation

Tables 3 and 4 report SMLEs for Model-1 where x is independent of z_1 and z_2 . Table 3 reports results (SML-1) for the SL (4.5). Results (SML-2) with the other SL (4.3) are reported in Table 4. Models with or without restrictions on parameters are estimated. A model with one restriction refers to one where unit variance restrictions are imposed, and a model with two restrictions refers to one where both unit variances and $\delta_2 = 0$ are assumed known. All these SMLEs with various n s and S s provide unbiased estimates of β , γ and b 's. There are downward biases in the SML-1 estimates of δ_3 with the unrestricted model. Such biases are relatively small in the SML-2 estimates. But the latter approach has relatively large biases for δ_1 with small $S = 30$. Those biases become small when the constraint $\delta_2 = 0$ is imposed. They also decrease as n increases. Overall, these two different SLs are compatible to each other in Mean and SD statistics with a moderate $S = 60$ or 100 . The SMLEs do not seem to be very sensitive to S except for the SML-2 approach for some of the δ s. There are some changes in Mean and SD of the SML-2 estimates of δ_1 and δ_3 when S is

increased from 30 to 60. The time cost is approximately linear in S or n . The Est.SDs are underestimated in some cases but, overall, are not too far off. The SMLEs of the SEL model are expected to be efficient relative to the SMLEs of the choice model. Comparing the SMLEs in Tables 3 and 4 with the SMLEs of the choice model in Table 1 and TSSMEs of the outcome equation in Table 2, we find some slight improvements in the SMLEs of parameters of utility functions but not much difference in the estimates of β . The major improvements are on the precision of estimates of δ_1 and δ_2 . The SDs of the SMLEs of δ_1 and δ_2 in Tables 3 and 4 are much smaller than those of the TSSMEs in Table 2.

To investigate whether asymptotic normal distributions provide accurate approximations to finite sample distributions of the SMLEs for $n = 200$ and 500 , Figures 2 to 5 are graphical presentations of normal probability plots (Q-Q plots) for the SMLEs. These plots are derived from the statistical package S-Plus. When the sampling distribution is close to a normal distribution, the plotted quantiles should mostly lie on a straight line (Chambers et al. 1983). Figure 2 shows that for $n = 200$, except for the SML-1 estimate of δ_3 , finite sample distributions seem to be approximately normal. For the SML-1 estimate of δ_3 , its finite sample distribution has a thicker right-hand tail. This might be due to the fact that the δ_3 is a positive parameter in a Cholesky decomposition and its estimates have been restricted to be non-negative in estimation. Estimates of δ_3 in some replications are very close to zero. For the SML-2 estimates in Figure 3, the corresponding distribution of the estimate of δ_3 is better approximated than that of the SML-1 estimate. For larger sample size, $n = 500$, normality approximates seem reasonably adequate.

To investigate effects of correlated regressors, Table 5 reports results for the SML estimation of Model-2 and Model-3. x and z_1 are perfectly correlated in Model-3. The results in Table 5 can be compared with corresponding SMLEs in Tables 3 and 4. For both alternative models, the SMLEs of β in Table 5 have larger SDs but the SDs of the SMLEs of γ can be slightly smaller. The SDs of β for Model-2 turn out to be larger than those for Model-3. But the reduction of SD for β in Model-3 is accompanied by increased SDs for the SMLEs of δ_2 and $b(2, 1)$. Overall, stronger correlation among regressors in the outcome equation and utility functions does not have substantial effects on SMLEs for P-SEL models.

More detailed comparisons of the two different SLs in (4.3) and (4.5) will investigate their variances due to sample variation and due to simulation. Related figures in Figure 6 provide numerical evidence on the better performance of (4.5) over (4.3). The figures are drawn for one sample of size $n = 200$ of Model-1 and are derived with 100 replications of simulation runs with $S = 30$ at each value of a parameter vector.

Each figure reports the mean and two SDs due to simulation variations from a cross section of an SL with respect to a single parameter while remaining parameters are fixed at their true values. The ‘slk-1’ refers to the log SL (4.5) and the ‘slk-2’ refers to that in (4.3). The axes, beta, gamma, b21, b22, d1, d2, and d3 refer, respectively, to the parameters β , γ , b_{21} , b_{22} , δ_1 , δ_2 , and δ_3 . It is apparent that the log SL (4.5) has much smaller simulation errors than those in (4.3). Further evidence is provided in Table 6 in terms of variance decomposition. The variance of a log SL is due to two sources of variation – simulation and sample variations. The results in Table 6 are based on SLs for 100 sample sets, of which each has $n = 200$. For each sample of size 200, the likelihood is simulated by (4.5) and (4.3), which are referred to, respectively, as SL-1 and SL-2. The number of simulation runs is $S = 30$. The likelihood is evaluated at the true parameter vector. For each n , 100 independent sets of simulation runs with $S = 30$ are replicated. This allows us to investigate the variance of an SL due to both its MCV and sample variation. In Table 6, results are reported for log SLs instead of SLs in (4.6). This is because SL values are numerically too small to have a clear visual inspection. The log SL function that involves a nonlinear transformation of the SL, however, will no longer be unbiased. So for this reason, the second variance term in (4.6) may be slightly different for various unbiased likelihood simulators. Table 6 reports sample evaluations of the two variance components of log SLs for all three models at the true parameter vector. The ‘sdcm1’ refers to the empirical SD of the first variance term on the right-hand side of (4.6), the ‘sdcm2’ refers to the SD of the second variance term, and the ‘SD’ refers to the overall SD. The ‘Mean’ refers to the overall mean of SL value for a model. The sdcm1 due to simulation for the SL (4.3) is three times larger than that of the SL (4.5). The sdcm1 component is a small fraction relative to the sdcm2 component for these two SLs for all three models. The overall Means and SDs are similar in magnitude for these two SLs, even though those of (4.3) are slightly larger than those of (4.5).

6.4 Finite Sample Variance Decomposition of TSSM and SML Estimates

A superior likelihood simulator with a relatively small MCV component in (4.6) does not necessarily imply that its corresponding SMLE of an unknown parameter will have smaller variance in a finite sample setting. The variance of a parameter estimator (without simulation) is determined by the randomness of sample observations. In simulation estimation, it depends also on the randomness of simulation draws. It is known that the SMLE is asymptotically efficient only if S increases at a rate faster than $\sqrt{n}$ (see, e.g., Lee 1992, and Gourieroux and Monfort 1993). Thus, under the latter design, the asymptotic variance of

an SMLE may be determined solely by the randomness of sample observation. However, the randomness of simulation draws may still have an important impact on the overall variance of a simulated parameter estimator in a finite sample with a moderate S . Let $\hat{\theta}(y, \xi)$ be an estimator of an unknown parameter θ , where y is a sample vector and ξ is a vector of random draws. The variance of $\hat{\theta}$ is a sum of two terms:

$$\text{var}(\hat{\theta}(y, \xi)) = E[\text{var}(\hat{\theta}(y, \xi)|y)] + \text{var}[E(\hat{\theta}(y, \xi)|y)]. \quad (6.3)$$

The $\text{var}(\hat{\theta}(y, \xi)|y)$ is the MCV of the estimator conditional on a sample. The first term in (6.3) is the mean of MCVs over sample sets. The second term captures, in general, not only the variance due to sample variation but also an effect due to simulation. As $\theta(y, \xi)$ depends on ξ nonlinearly in a complex implicit form, $E(\hat{\theta}(y, \xi)|y)$ of y may be different with various unbiased SLs. As SLs may have different shapes from various simulation approaches, the second term may also be different. Unless the conditional mean functions of y are the same, the comparison of MCVs does not necessarily provide the adequate ranking of simulation approaches in parameter estimation.

To evaluate the two effects of simulation, we report results on variance decomposition in Table 7. The reported results are the sample average of the conditional SD due to random draws (sdcm1), i.e., $E[\text{var}(\hat{\theta}(y, \xi)|y)]^{1/2}$, the sample SD of the conditional mean function (sdcm2), i.e., $\text{var}^{1/2}[E(\hat{\theta}(y, \xi)|y)]$, and the overall SD, based on 100 sets of random draws with $S = 30$ for each $n = 200$ and a total number of sample sets of 100. Compare first the TSSMEs of the outcome equation. In terms of descending sdcm1, the ranking is SIV-2, SIV-1, SIV-Gibbs-1 (Gib-1), SIV-Gibbs-2 (Gib-2), and SOLS. In terms of sdcm2, the corresponding ranking is SIV-2, Gib-1, Gib-2, SIV-1, and SOLS. The SIV-2 has the largest variance. The rankings among the SIV-1, Gib-1, and Gib-2 should not be exaggerated since their differences on these variance components are very small. Overall, these three approaches are compatible with each other. The SOLS turns out to have smaller variance components than the various SIV estimates. This may be due to the additional independent simulation of IVs for bias-correction terms with the various SIV approaches. In comparison, the SML-1 approach with the SL (4.5) has smaller sdcm1s than those of the SML-2 approach for all parameters in the model. This means that MCVs are smaller when the SML-1 approach is used. However, it is interesting to note that for the estimation of δ s, the SML-2 estimates have smaller sdcm2s than those of the SML-1 estimates. This means that the corresponding conditional mean functions of SML-2 estimators have smaller variance. For the latter, the reduction of the sdcm2 for δ_3 is large enough that its overall SD can even be smaller than the corresponding SD of the SML-1 approach. The shares of the sdcm1 over the

SD for the β , γ and b s of both SML approaches are very small. The SML-1 and SML-2 estimates with parameter constraints imposed (SML-1R and SML-2R, respectively) have similar properties. Comparing the SMLEs of the outcome equation with TSSMEs, the SMLEs have smaller variance components than those of TSSMEs.

6.5 Effects of Distributional Misspecification on SMLEs

To investigate whether the SMLEs are sensitive to misspecified distributions, Model-1 in Section 6.1 is used except that ξ_1 , ξ_2 and ξ_3 in the Cholesky decomposition are no longer normally distributed. The normal distribution is symmetric around zero and has a unique mode. To investigate possible effects of skewness and bimodals, we considered several alternatives. The design ‘Gumbel’ refers to the case that ξ s are independent Gumbel-distributed random variables and are normalized to have zero means and unit variances. The second design ‘Mixture-1’ refers to the case that each of ξ s is a 50-50 mixture of two normal variables $N(-3, 1)$ and $N(3, 1)$. These ξ s are independent and normalized to have zero means and unit variances. Since $u_1 = b_{11}\xi_1$, $u_2 = b_{21}\xi_1 + b_{22}\xi_2$, and $\epsilon = \delta_1\xi_1 + \delta_2\xi_2 + \delta_3\xi_3$, ϵ is a linear combination of three normal mixtures while u_1 is a single normal mixture. On the contrary, the third design ‘Mixture-2’ has ϵ being a single normal mixture and both u_1 and u_2 are linear combinations of normal mixtures. To be specific, permute (u_1, u_2, ϵ) to (ϵ, u_1, u_2) . Let CC' be the Cholesky decomposition of the variance matrix of (ϵ, u_1, u_2) . Given the true values of b s and δ s, they determine the variance of (ϵ, u_1, u_2) and, hence, C . With C , one has the representation that $\epsilon = c_{11}\zeta_1^*$, $u_1 = c_{21}\zeta_1^* + c_{22}\zeta_2^*$ and $u_2 = c_{31}\zeta_1^* + c_{32}\zeta_2^* + c_{33}\zeta_3^*$. The ζ^* s are designed to be independently distributed standardized 50-50 mixtures of $N(-3, 1)$ and $N(3, 1)$.

Table 8 reports SMLEs based on (4.5) under the normal disturbance specification while data are generated by a ‘Gumbel’, a ‘Mixture-1’ or a ‘Mixture-2’ design. These results shall be compared with those in Table 3 (last column) where the SL is correctly specified. For the ‘Gumbel’ samples, the SMLEs of γ and b_{21} are similar to those of the correctly specified normal SL. The estimate of β shows no bias. Only the estimates of δ s show relatively large biases. So the SMLE with normal SL is not very sensitive to the ‘Gumbel’ misspecification. For samples generated from ‘Mixture-1’ disturbances, the SMLE of γ has shown obvious downward bias. The bias does not change with a larger $S - 200$ or 400 . It can not be reduced even when n increases from 200 to 500. On the contrary, the estimates of β and δ_1 are not biased. The downward bias in the SMLE of δ_3 is larger than that in Table 3. In conclusion, the SMLEs of the coefficients in the utility functions are sensitive, but not the coefficients of the outcome equation, to the ‘Mixture-1’ distri-

bution. When samples are generated from ‘Mixture-2’ disturbances, all the parameters show little biases. Comparing the latter SMLEs with those in Table 3, we find that the biases of both estimates of γ are only about 2 percents. The bias in Table 3 is upward but that in Table 8 is downward.

6.6 Monte Carlo Results on Semiparametric Estimation

The density of (ξ_1, ξ_2, ξ_3) in the Cholesky decomposition $u_1 = b_{11}\xi_1$, $u_2 = b_{21}\xi_1 + b_{22}\xi_2$, and $\epsilon = \delta_1\xi_1 + \delta_2\xi_2 + \delta_3\xi_3$, is expanded with some finite order Hermite polynomials. The case ‘Hermite’ refers to the SL formulated with the density expansion:

$$f(\xi) = \frac{(1 + c_{100}\xi_1 + c_{010}\xi_2 + c_{001}\xi_3)^2}{(1 + c_{100}^2 + c_{010}^2 + c_{001}^2)} \prod_{l=1}^3 \phi(\xi_l). \quad (6.4)$$

This expansion has used the first-order Hermite polynomial. However, because the Hermite expansion involves a square power, it has implicitly utilized some second-order polynomial terms. This expansion has introduced three additional free parameters c s in addition to the original parameters. The formulation ‘Hermite-ind-1’ has the SL formulated with

$$f(\xi) = \frac{(1 + c_{100}\xi_1)^2}{1 + c_{100}^2} \frac{(1 + c_{010}\xi_2)^2}{1 + c_{010}^2} \frac{(1 + c_{001}\xi_3)^2}{1 + c_{001}^2} \prod_{l=1}^3 \phi(\xi_l).$$

This expansion has also three additional parameters. It differs from the ‘Hermite’ case in that independence of ξ_1 , ξ_2 and ξ_3 is imposed. The ‘Hermite-ind-2’ case extends the latter case to have a higher-order expansion:

$$f(\xi) = \frac{(1 + c_{100}\xi_1 + c_{200}(\xi_1 - 1)^2)^2}{1 + c_{100}^2 + 2c_{200}^2} \frac{(1 + c_{010}\xi_2 + c_{020}(\xi_2 - 1)^2)^2}{1 + c_{010}^2 + 2c_{020}^2} \frac{(1 + c_{001}\xi_3 + c_{002}(\xi_3 - 1)^2)^2}{1 + c_{001}^2 + 2c_{002}^2} \prod_{l=1}^3 \phi(\xi_l). \quad (6.6)$$

This expansion has introduced a total of six more parameters.

Tables 9A-9B report the SSMLEs based on (6.4)-(6.6). The interesting parameters in a semiparametric model are β and γ . Table 9A reports results where sample data are generated from Model-1 with either normal or Gumbel disturbances. When sample data are generated by the normal distribution, the SSMLEs of β are unbiased and the biases of γ have magnitudes similar to those in Table 3. As all the three SSLs contain explicitly the normal SL as a special case, the SSLs correspond to unconstrained estimation for the normally distributed sample. For the latter, as expected, the SDs of the SSMLEs can be larger than those of Table 3. The SDs of ‘Hermite-ind-2’ SSMLEs γ are two times larger than those in Table 3. However, except for that, SDs for the others are not much larger. For data generated by Gumbel disturbances, the SSMLEs are similar to those in Table 8. Table 9B reports results where sample data are generated by either

‘Mixture-1’ or ‘Mixture-2’ disturbances. For the ‘Mixture-1’ sample, neither ‘Hermite’ nor ‘Hermite-ind-1’ improves the bias of the SMLE of γ in Table 8 (column 2). The 10 percent estimated bias remains. Only ‘Hermite-ind-2’ appears to have corrected some of the biases in γ . For the ‘Mixture-2’ sample, the ‘Hermite-ind-2’ SSMLs of β have some smaller SDs, but those of γ are much larger than those in Table 8 (column 3).

6.7 Test for Normality – Finite Sample Level of Significance and Power

The SLs formulated in (6.4)-(6.6) may also be valuable for testing normality in disturbances. Since the SL of normal distribution is nested in any of the SSLs, we may investigate the possible use of SL ratio test statistics. Table 10 reports the level of significance and power of those statistics. The nominal levels of significance are based on the chi-square distribution with three degrees-of-freedom for ‘Hermite’ and ‘Hermite-ind-1’ expansions. For ‘Hermite-ind-2’ expansion, the chi-square distribution with six degrees-of-freedom is used. The ‘Gumbel’, ‘Mixture-1’ and ‘Mixture-2’ distributions are used to investigate the power of the test statistics.

The exact levels of significance of the three test statistics tend to be larger than the nominal ones. But they are not too much larger. Instead of a nominal level of 5 percent, the exact level of significance of the ‘Hermite-ind-2’ statistic is, on average, about 8 percent. The ‘Hermite’ and ‘Hermite-ind-1’ statistics are slightly more accurate than the ‘Hermite-ind-2’ statistic. However, the ‘Hermite’ and ‘Hermite-ind-1’ statistics have weaker power against the ‘Mixture-1’ alternative. They have also only moderate power against the Gumbel alternative. The ‘Hermite-ind-2’ statistic tends to have relatively better power. In particular, it has good power to test against the ‘Mixture-2’ alternative.

7. Conclusions

Simulation estimation of SEL models with a finite number of choice alternatives are considered. Conditional outcome equations can be simulated by various biased or unbiased simulators. A recursive simulator that generalizes the GHK approach for the DC model to the SEL model is possible. Unbiased simulators can be derived via GS. A variation reduction simulation technique based on conditioning can also be used. The outcome equation can be estimated by SIV methods, which are essentially MSM. The likelihood function can also be simulated with various generalizations of the GHK type simulator.

To compare the finite sample performance of various simulators and simulation estimation procedures, Monte Carlo experiments are reported. Monte Carlo experiments demonstrate that the TSSMEs derived

with various biased or unbiased simulators for the outcome equation provide similar finite sample properties. Unbiased simulators via the Gibbs sampler are not difficult to use but they are relatively computationally expensive compared to the recursive simulator, even though the latter may be biased. The SMLEs based on different density factorizations are also similar, even though variance decompositions in finite samples show interesting differences. They are relatively efficient compared to TSSMEs. Limited Monte Carlo results show that SMLEs based on the normal distributional specification may not be very sensitive to certain distributional misspecifications. But for certain normal mixture cases, some parameters may have substantial downward biases. SSL generalization based on series expansions with second-order Hermite polynomials can have some use for reducing biases. But the semiparametric estimates may have much larger variances. The SSL ratio test based on second-order Hermite polynomials has good power against some non-normal alternatives.

Table 1: SML Estimation of Discrete Choice Models
True parameters: $\gamma = 1$, $b(2, 1) = 0.6$, and $b(2, 2) = 0.8$

	no restriction			unit variance restriction		
	Mean	SD	Est.SD	Mean	SD	Est.SD
<hr/>						
n=200, S=30						
γ	1.0326	(0.1738)	[0.1613]	1.0182	(0.1384)	[0.1279]
$b(2, 1)$	0.5917	(0.1796)	[0.1808]	0.5736	(0.1467)	[0.1430]
$b(2, 2)$	0.8087	(0.1901)	[0.1815]			
	F=-151.062			F=-151.779, T=4.27		
<hr/>						
n=200, S=60						
γ	1.0321	(0.1738)	[0.1610]	1.0190	(0.1392)	[0.1263]
$b(2, 1)$	0.5922	(0.1811)	[0.1806]	0.5738	(0.1492)	[0.1432]
$b(2, 2)$	0.8059	(0.1906)	[0.1809]			
	F=-150.972			F=-151.699, T=8.38		
<hr/>						
n=200, S=100						
γ	1.0317	(0.1731)	[0.1609]	1.0187	(0.1388)	[0.1279]
$b(2, 1)$	0.5929	(0.1816)	[0.1804]	0.5750	(0.1470)	[0.1429]
$b(2, 2)$	0.8050	(0.1884)	[0.1806]			
	F=-150.964			F=-151.670, T=13.88		
<hr/>						
n=500, S=50						
γ	1.0121	(0.0939)	[0.0993]	1.0017	(0.1023)	[0.0828]
$b(2, 1)$	0.5970	(0.1103)	[0.1116]	0.5853	(0.0964)	[0.0899]
$b(2, 2)$	0.8067	(0.1074)	[0.1121]			
	F=-380.859			F=-382.256, T=18.51		

Figure 1

Table 2: Two-Stage Simulation Estimation of Sample Selection EquationsTrue parameters: $\beta = 1$, $\delta_1 = 1$, and $\delta_2 = 0$ sample size before selection $n = 200$

	β		δ_1		δ_2		T
	Mean	SD	Mean	SD	Mean	SD	
SIV-1							
s=30	0.9930	(0.1382)	1.0084	(0.4020)	-0.0066	(0.6405)	0.16
s=60	0.9960	(0.1377)	1.0154	(0.3925)	0.0017	(0.6030)	0.28
s=100	0.9931	(0.1389)	1.0188	(0.3768)	0.0157	(0.5893)	0.44
SIV-2							
s=30	0.9961	(0.1545)	1.0306	(0.6222)	0.0111	(1.0789)	0.28
s=60	0.9942	(0.1530)	1.0028	(0.5473)	-0.0207	(0.8877)	0.51
s=100	0.9957	(0.1530)	1.0108	(0.4993)	-0.0114	(0.8079)	0.82
SOLS							
s=30	0.9948	(0.1402)	0.9280	(0.3142)	-0.1027	(0.4553)	0.10
s=60	0.9943	(0.1379)	0.9586	(0.3303)	-0.0661	(0.4933)	0.16
s=100	0.9942	(0.1384)	0.9837	(0.3488)	-0.0307	(0.5199)	0.24
SIV-Gibbs-1							
s=30, c=5	0.9926	(0.1389)	1.0435	(0.4345)	0.0571	(0.6728)	0.92
s=30, c=15	0.9959	(0.1404)	1.0111	(0.4258)	0.0040	(0.6852)	2.85
s=30, c=30	0.9924	(0.1403)	1.0120	(0.4054)	0.0158	(0.6653)	5.65
s=60, c=10	0.9940	(0.1394)	1.0296	(0.3954)	0.0364	(0.6059)	3.73
s=100, c=10	0.9949	(0.1367)	1.0228	(0.3673)	0.0189	(0.5653)	6.34
SIV-Gibbs-2							
s=30, c=5	0.9923	(0.1388)	1.0385	(0.4342)	0.0413	(0.6748)	0.95
s=30, c=15	0.9955	(0.1398)	1.0088	(0.4164)	-0.0002	(0.6661)	2.87
s=30, c=30	0.9923	(0.1393)	1.0133	(0.4059)	0.0174	(0.6613)	5.71
s=60, c=10	0.9940	(0.1388)	1.0267	(0.3970)	0.0297	(0.6072)	3.82
s=100, c=10	0.9940	(0.1370)	1.0202	(0.3662)	0.0176	(0.5673)	6.44

NOTE: The sample size for the alternative 1 after selection is, on average, 65.65 with a minimum of 50 and a maximum of 83 observations.

Table 3: SML Estimation of Sample Selection Models — SML-1
True parameters: $\beta = 1$, $\delta_1 = 1$, $\delta_2 = 0$, $\delta_3 = 1$, $\gamma = 1$, $b(2, 1) = 0.6$, and $b(2, 2) = 0.8$

	no restriction			one restriction			two restrictions		
	Mean	SD	Est.SD	Mean	SD	Est.SD	Mean	SD	Est.SD
<hr/>									
n=200, S=30									
β	0.9959	(0.1364)	[0.1288]	0.9945	(0.1362)	[0.1302]	0.9941	(0.1340)	[0.1341]
δ_1	0.9978	(0.2975)	[0.2306]	0.9975	(0.2981)	[0.2315]	1.0206	(0.1996)	[0.1877]
δ_2	-0.0287	(0.4641)	[0.3494]	-0.0259	(0.4478)	[0.3510]	—		
δ_3	0.8098	(0.2825)	[0.1958]	0.8187	(0.2784)	[0.1962]	0.9352	(0.1779)	[0.1535]
γ	1.0299	(0.1705)	[0.1385]	1.0202	(0.1296)	[0.1149]	1.0203	(0.1262)	[0.1197]
$b(2, 1)$	0.5902	(0.1779)	[0.1639]	0.5726	(0.1472)	[0.1358]	0.5743	(0.1439)	[0.1407]
$b(2, 2)$	0.8087	(0.1897)	[0.1679]	—			—		
	F=-256.329, con# 398			F=-256.761			F=-257.312, T=10.26		
<hr/>									
n=200, S=60									
β	0.9924	(0.1459)	[0.1307]	0.9948	(0.1356)	[0.1322]	0.9941	(0.1339)	[0.1334]
δ_1	1.0018	(0.2923)	[0.2367]	0.9989	(0.2904)	[0.2348]	1.0200	(0.2002)	[0.1872]
δ_2	-0.0128	(0.4605)	[0.3669]	-0.0282	(0.4411)	[0.3619]	—		
δ_3	0.8048	(0.2932)	[0.2032]	0.8282	(0.2582)	[0.2059]	0.9358	(0.1778)	[0.1536]
γ	1.0257	(0.1766)	[0.1395]	1.0197	(0.1279)	[0.1164]	1.0208	(0.1267)	[0.1195]
$b(2, 1)$	0.5779	(0.2179)	[0.1659]	0.5746	(0.1466)	[0.1369]	0.5756	(0.1457)	[0.1404]
$b(2, 2)$	0.8251	(0.4016)	[0.1743]	—			—		
	F=-256.754, con# 395			F=-256.753, con# 397			F=-257.229, T=20.20		
<hr/>									
n=200, S=100									
β	0.9944	(0.1441)	[0.1335]	0.9939	(0.1347)	[0.1325]	0.9942	(0.1339)	[0.1338]
δ_1	1.0035	(0.2926)	[0.2384]	0.9997	(0.2908)	[0.2363]	1.0200	(0.2001)	[0.1875]
δ_2	-0.0150	(0.4612)	[0.3698]	-0.0230	(0.4445)	[0.3660]	—		
δ_3	0.8126	(0.2703)	[0.2106]	0.8284	(0.2499)	[0.2112]	0.9357	(0.1783)	[0.1532]
γ	1.0248	(0.1745)	[0.1399]	1.0194	(0.1278)	[0.1171]	1.0205	(0.1264)	[0.1201]
$b(2, 1)$	0.5903	(0.1819)	[0.1657]	0.5769	(0.1446)	[0.1377]	0.5764	(0.1437)	[0.1405]
$b(2, 2)$	0.8038	(0.1866)	[0.1713]	—			—		
	F=-256.614, con# 395			F=-256.648, con# 398			F=-257.225, T=33.66		
<hr/>									
n=500, S=30									
β	0.9970	(0.0899)	[0.0853]	0.9971	(0.0901)	[0.0856]	0.9964	(0.0897)	[0.0853]
δ_1	0.9838	(0.1760)	[0.1550]	0.9845	(0.1741)	[0.1574]	1.0014	(0.1208)	[0.1189]
δ_2	-0.0312	(0.2764)	[0.2455]	-0.0307	(0.2686)	[0.2447]	—		
δ_3	0.9459	(0.1418)	[0.1189]	0.9480	(0.1444)	[0.1199]	0.9767	(0.1068)	[0.0969]
γ	1.0113	(0.0920)	[0.0875]	1.0062	(0.0729)	[0.0737]	1.0065	(0.0728)	[0.0754]
$b(2, 1)$	0.5956	(0.1103)	[0.1062]	0.5874	(0.0882)	[0.0863]	0.5867	(0.0880)	[0.0873]
$b(2, 2)$	0.8075	(0.1081)	[0.1075]	—			—		
	F=-646.527, T=36.19			F=-647.035, T=30.412			F=-647.543, T=25.62		

NOTES:

One restriction — unit variance for choice equation disturbances;

Two restrictions — unit variance for all choice equation disturbances and $\delta_2 = 0$.

Table 4: SML Estimation of Sample Selection Models — SML-2
True parameters: $\beta = 1$, $\delta_1 = 1$, $\delta_2 = 0$, $\delta_3 = 1$, $\gamma = 1$, $b(2, 1) = 0.6$, and $b(2, 2) = 0.8$

	no restriction			one restriction			two restrictions		
	Mean	SD	Est.SD	Mean	SD	Est.SD	Mean	SD	Est.SD
<hr/>									
n=200, S=30									
β	0.9935	(0.1407)	[0.1333]	0.9919	(0.1398)	[0.1342]	0.9943	(0.1350)	[0.1353]
δ_1	0.9253	(0.3028)	[0.2355]	0.9298	(0.2970)	[0.2325]	0.9840	(0.2049)	[0.1927]
δ_2	-0.1240	(0.4333)	[0.3233]	-0.1204	(0.4289)	[0.3183]	—		
δ_3	0.9220	(0.2169)	[0.1629]	0.9260	(0.2098)	[0.1666]	0.9931	(0.1600)	[0.1457]
γ	1.0304	(0.1720)	[0.1431]	1.0182	(0.1288)	[0.1177]	1.0188	(0.1264)	[0.1208]
$b(2, 1)$	0.5908	(0.1835)	[0.1692]	0.5719	(0.1509)	[0.1404]	0.5728	(0.1463)	[0.1429]
$b(2, 2)$	0.8112	(0.1936)	[0.1758]	—			—		
	F=-256.606, T=17.41			F=-257.178			F=-257.903		
<hr/>									
n=200, S=60									
β	0.9932	(0.1391)	[0.1307]	0.9947	(0.1387)	[0.1301]	0.9953	(0.1345)	[0.1334]
δ_1	0.9655	(0.2874)	[0.2274]	0.9688	(0.2922)	[0.2236]	1.0039	(0.2054)	[0.1871]
δ_2	-0.0629	(0.4360)	[0.3298]	-0.0555	(0.4419)	[0.3245]	—		
δ_3	0.8734	(0.2515)	[0.1697]	0.8710	(0.2581)	[0.1685]	0.9614	(0.1745)	[0.1478]
γ	1.0278	(0.1706)	[0.1403]	1.0199	(0.1291)	[0.1160]	1.0201	(0.1271)	[0.1208]
$b(2, 1)$	0.5908	(0.1799)	[0.1676]	0.5726	(0.1513)	[0.1385]	0.5750	(0.1466)	[0.1420]
$b(2, 2)$	0.8057	(0.1929)	[0.1722]	—			—		
	F=-256.160, T=35.65			F=-256.772			F=-257.479		
<hr/>									
n=200, S=100									
β	0.9961	(0.1393)	[0.1299]	0.9937	(0.1423)	[0.1269]	0.9942	(0.1339)	[0.1332]
δ_1	0.9749	(0.2899)	[0.2316]	0.9817	(0.2838)	[0.2247]	1.0063	(0.2014)	[0.1885]
δ_2	-0.0569	(0.4483)	[0.3440]	-0.0337	(0.4704)	[0.3298]	—		
δ_3	0.8518	(0.2710)	[0.1773]	0.8407	(0.2864)	[0.1750]	0.9579	(0.1741)	[0.1500]
γ	1.0264	(0.1703)	[0.1380]	1.0175	(0.1281)	[0.1141]	1.0198	(0.1264)	[0.1201]
$b(2, 1)$	0.5924	(0.1804)	[0.1650]	0.5713	(0.1526)	[0.1359]	0.5761	(0.1440)	[0.1413]
$b(2, 2)$	0.8039	(0.1883)	[0.1700]	—			—		
	F=-256.054, T=61.28			F=-256.660			F=-257.486		
<hr/>									
n=500, S=30									
β	0.9971	(0.0904)	[0.0853]	0.9970	(0.0904)	[0.0851]	0.9969	(0.0900)	[0.0852]
δ_1	0.9013	(0.1809)	[0.1518]	0.9040	(0.1792)	[0.1539]	0.9637	(0.1222)	[0.1227]
δ_2	-0.1409	(0.2543)	[0.2134]	-0.1358	(0.2525)	[0.2152]	—		
δ_3	1.0247	(0.1068)	[0.0968]	1.0249	(0.1073)	[0.1024]	1.0289	(0.0993)	[0.0920]
γ	1.0132	(0.0921)	[0.0886]	1.0055	(0.0725)	[0.0720]	1.0055	(0.0729)	[0.0750]
$b(2, 1)$	0.5988	(0.1107)	[0.1068]	0.5872	(0.0883)	[0.0861]	0.5854	(0.0890)	[0.0883]
$b(2, 2)$	0.8120	(0.1089)	[0.1087]	—			—		
	F=-647.799, T=41.97			F=-648.306			F=-649.022		

Figure 2: Normal probability plot -- SML-1; $n=200$, $S=30$

Figure 3: Normal probability plot -- SML-1; $n=500$, $S=30$

Figure 4: Normal probability plot -- SML-2; $n=200$, $S=30$

Figure5: Normal probability plot -- SML-2; $n=500$, $S=30$

Table 5: SML Estimation of Sample Selection Models – Strongly Correlated Regressors
True parameters: $\beta = 1$, $\delta_1 = 1$, $\delta_2 = 0$, $\delta_3 = 1$, $\gamma = 1$, $b(2, 1) = 0.6$, and $b(2, 2) = 0.8$
Sample size $n=200$, $S=30$

	no restriction			one restriction			two restrictions		
	Mean	SD	Est.SD	Mean	SD	Est.SD	Mean	SD	Est.SD
SMLE-1				Model-2					
β	1.0014	(0.1626)	[0.1403]	1.0013	(0.1623)	[0.1427]	0.9964	(0.1496)	[0.1390]
δ_1	0.9836	(0.2847)	[0.2477]	0.9857	(0.2783)	[0.2520]	1.0157	(0.2018)	[0.1885]
δ_2	-0.0607	(0.4743)	[0.3695]	-0.0543	(0.4645)	[0.3818]	—		
δ_3	0.8145	(0.2967)	[0.1971]	0.8245	(0.2865)	[0.2055]	0.9396	(0.1674)	[0.1527]
γ	1.0097	(0.1517)	[0.1355]	0.9995	(0.1172)	[0.1129]	0.9989	(0.1153)	[0.1183]
$b(2,1)$	0.6096	(0.1691)	[0.1613]	0.5890	(0.1309)	[0.1313]	0.5884	(0.1298)	[0.1379]
$b(2,2)$	0.8024	(0.1791)	[0.1665]	—			—		
	F=-256.815, con# 399			F=-257.279, T=13.64			F=-257.927, T=10.27		
SMLE-2									
β	1.0095	(0.1644)	[0.1421]	1.0119	(0.1643)	[0.1446]	0.9945	(0.1523)	[0.1401]
δ_1	0.9010	(0.2833)	[0.2356]	0.8967	(0.2809)	[0.2389]	0.9804	(0.2046)	[0.1951]
δ_2	-0.1604	(0.4500)	[0.3248]	-0.1712	(0.4381)	[0.3341]	—		
δ_3	0.9243	(0.2218)	[0.1603]	0.9338	(0.2120)	[0.1647]	0.9939	(0.1531)	[0.1464]
γ	1.0087	(0.1494)	[0.1397]	0.9975	(0.1176)	[0.1155]	0.9976	(0.1159)	[0.1183]
$b(2,1)$	0.6061	(0.1655)	[0.1663]	0.5885	(0.1303)	[0.1357]	0.5872	(0.1306)	[0.1394]
$b(2,2)$	0.8034	(0.1786)	[0.1733]	—			—		
	F=-257.205, T=17.81			F=-257.723, T=14.63			F=-258.505, T=12.21		
SMLE-1				Model-3					
β	1.0007	(0.1495)	[0.1437]	0.9975	(0.1465)	[0.1416]	1.0020	(0.1339)	[0.1374]
δ_1	1.0016	(0.2785)	[0.2379]	0.9987	(0.2802)	[0.2351]	1.0187	(0.2009)	[0.1880]
δ_2	-0.0298	(0.4938)	[0.3693]	-0.0347	(0.4931)	[0.3664]	—		
δ_3	0.8103	(0.2767)	[0.2061]	0.8167	(0.2678)	[0.2088]	0.9500	(0.1581)	[0.1523]
γ	1.0229	(0.1657)	[0.1437]	1.0113	(0.1249)	[0.1164]	1.0115	(0.1238)	[0.1216]
$b(2,1)$	0.5935	(0.1925)	[0.1630]	0.5758	(0.1507)	[0.1336]	0.5766	(0.1500)	[0.1384]
$b(2,2)$	0.8051	(0.1881)	[0.1704]	—			—		
	F=-258.552			F=-259.094			F=-259.713		
SMLE-2									
β	0.9984	(0.1550)	[0.1416]	0.9965	(0.1507)	[0.1431]	1.0115	(0.1364)	[0.1389]
δ_1	0.9387	(0.2896)	[0.2254]	0.9360	(0.2896)	[0.2312]	0.9851	(0.2034)	[0.1949]
δ_2	-0.1190	(0.4634)	[0.3138]	-0.1174	(0.4574)	[0.3211]	—		
δ_3	0.9118	(0.2144)	[0.1630]	0.9162	(0.2153)	[0.1664]	1.0007	(0.1472)	[0.1476]
γ	1.0209	(0.1714)	[0.1468]	1.0096	(0.1259)	[0.1201]	1.0114	(0.1237)	[0.1220]
$b(2,1)$	0.5949	(0.1955)	[0.1665]	0.5750	(0.1526)	[0.1378]	0.5753	(0.1509)	[0.1391]
$b(2,2)$	0.8051	(0.1918)	[0.1759]	—			—		
	F=-258.930			F=-259.476			F=-260.269		

Figure 6: Simulated Likelihood

slk-1

-290 -270 -250 -230

slk-1

-290 -270 -250 -230

slk-2

-300 -280 -260 -240

slk-2

-290 -270 -250 -230

slik-1

slik-1

slik-2

slik-2

Table 6: Variance Decomposition of Simulated Likelihood Functions
number of samples=100; simulation replications=100/sample; n=200; S=30

model	Mean	sdcm1	sdcm2	SD
SL-1				
Model-1	-261.3	0.4759	13.55	13.56
Model-2	-262.4	0.4784	13.92	13.93
Model-3	-263.1	0.4727	14.53	14.54
SL-2				
Model-1	-261.9	1.238	13.72	13.78
Model-2	-263.0	1.239	14.01	14.07
Model-3	-263.8	1.272	14.66	14.71

Table 7: Finite Sample Variance Decomposition of Coefficient Estimates
number of samples=100; simulation replications=100/sample; n=200; S=30; c=10

	β	δ_1	δ_2	δ_3	γ	$b(2, 1)$	$b(2, 2)$
<hr/>							
SML-1							
sdcm1	0.0083	0.0393	0.1046	0.1049	0.0163	0.0284	0.0240
sdcm2	0.1392	0.2456	0.3980	0.2415	0.1330	0.1532	0.1780
SD	0.1394	0.2487	0.4115	0.2633	0.1340	0.1558	0.1796
<hr/>							
SML-1R							
sdcm1	0.0010	0.0048	—	0.0043	0.0066	0.0159	—
sdcm2	0.1473	0.2014	—	0.1873	0.1160	0.1349	—
SD	0.1473	0.2014	—	0.1873	0.1162	0.1359	—
<hr/>							
SML-2							
sdcm1	0.0371	0.1729	0.2964	0.1506	0.0329	0.0500	0.0564
sdcm2	0.1480	0.2116	0.2887	0.1529	0.1411	0.1591	0.1781
SD	0.1525	0.2733	0.4138	0.2146	0.1448	0.1668	0.1868
<hr/>							
SML-2R							
sdcm1	0.0237	0.0581	—	0.0585	0.0111	0.0180	—
sdcm2	0.1463	0.1950	—	0.1541	0.1156	0.1377	—
SD	0.1482	0.2034	—	0.1648	0.1161	0.1389	—
<hr/>							
SIV-1							
sdcm1	0.0222	0.2136	0.3821				
sdcm2	0.1501	0.3175	0.5001				
SD	0.1517	0.3827	0.6294				
<hr/>							
SIV-2							
sdcm1	0.0228	0.4096	0.7512				
sdcm2	0.1646	0.4138	0.6710				
SD	0.1662	0.5822	1.0070				
<hr/>							
SOLS							
sdcm1	0.0174	0.1457	0.2494				
sdcm2	0.1497	0.2520	0.3525				
SD	0.1507	0.2911	0.4318				
<hr/>							
Gib-1							
sdcm1	0.0219	0.2100	0.3790				
sdcm2	0.1503	0.3261	0.5170				
SD	0.1519	0.3879	0.6410				
<hr/>							
Gib-2							
sdcm1	0.0185	0.2053	0.3671				
sdcm2	0.1502	0.3258	0.5169				
SD	0.1513	0.3851	0.6340				

Table 8: SMLE of Misspecified Sample Selection Models
True parameters: $\beta = 1$, $\delta_1 = 1$, $\delta_2 = 0$, $\delta_3 = 1$, $\gamma = 1$, $b(2, 1) = 0.6$, and $b(2, 2) = 0.8$

True data	Gumbel			Mixture-1			Mixture-2		
	Mean	SD	Est.SD	Mean	SD	Est.SD	Mean	SD	Est.SD
<hr/>									
n=200, S=30									
β	0.9962	(0.1488)	[0.1410]	0.9888	(0.1314)	[0.1296]	0.9884	(0.1381)	[0.1349]
δ_1	1.2279	(0.2108)	[0.1950]	0.9984	(0.1567)	[0.1768]	0.9712	(0.1351)	[0.1872]
δ_3	0.8908	(0.1985)	[0.1795]	0.9130	(0.1252)	[0.1458]	0.9669	(0.1408)	[0.1422]
γ	1.0256	(0.1219)	[0.1200]	0.8998	(0.1033)	[0.1100]	0.9774	(0.1192)	[0.1139]
$b(2, 1)$	0.5760	(0.1383)	[0.1336]	0.6015	(0.1310)	[0.1374]	0.5490	(0.1413)	[0.1471]
	F=-260.039			F=-262.866			F=-259.700		
<hr/>									
n=200, S=60									
β	0.9962	(0.1487)	[0.1408]	0.9888	(0.1313)	[0.1295]	0.9885	(0.1381)	[0.1354]
δ_1	1.2275	(0.2106)	[0.1962]	0.9982	(0.1571)	[0.1774]	0.9708	(0.1347)	[0.1874]
δ_3	0.8912	(0.1985)	[0.1819]	0.9130	(0.1253)	[0.1462]	0.9671	(0.1406)	[0.1418]
γ	1.0262	(0.1225)	[0.1200]	0.9000	(0.1029)	[0.1101]	0.9782	(0.1196)	[0.1143]
$b(2, 1)$	0.5752	(0.1405)	[0.1347]	0.6022	(0.1305)	[0.1374]	0.5493	(0.1399)	[0.1472]
	F=-259.995			F=-262.828			F=-259.648		
<hr/>									
n=200, S=100									
β	0.9962	(0.1488)	[0.1412]	0.9889	(0.1313)	[0.1299]	0.9884	(0.1381)	[0.1351]
δ_1	1.2268	(0.2106)	[0.1955]	0.9982	(0.1570)	[0.1775]	0.9704	(0.1347)	[0.1878]
δ_3	0.8916	(0.1986)	[0.1814]	0.9131	(0.1252)	[0.1463]	0.9674	(0.1405)	[0.1416]
γ	1.0265	(0.1222)	[0.1200]	0.9004	(0.1033)	[0.1102]	0.9782	(0.1198)	[0.1138]
$b(2, 1)$	0.5749	(0.1409)	[0.1346]	0.6019	(0.1308)	[0.1375]	0.5501	(0.1391)	[0.1471]
	F=-259.955			F=-262.788			F=-259.588		
<hr/>									
Mixture-1									
	n=200, S=200			n=200, S=400					
β	0.9889	(0.1312)	[0.1294]	0.9889	(0.1313)	[0.1296]			
δ_1	0.9980	(0.1568)	[0.1771]	0.9979	(0.1567)	[0.1771]			
δ_3	0.9132	(0.1252)	[0.1465]	0.9133	(0.1251)	[0.1461]			
γ	0.9006	(0.1032)	[0.1107]	0.9007	(0.1029)	[0.1102]			
$b(2, 1)$	0.6015	(0.1308)	[0.1379]	0.6020	(0.1305)	[0.1377]			
	F=-262.787			F=-262.756					
<hr/>									
	n=500, S=60			n=500, S=100					
β	0.9946	(0.0805)	[0.0822]	0.9945	(0.0805)	[0.0822]			
δ_1	1.0083	(0.1019)	[0.1123]	1.0083	(0.1020)	[0.1120]			
δ_3	0.9332	(0.0795)	[0.0932]	0.9332	(0.0795)	[0.0932]			
γ	0.8995	(0.0643)	[0.0690]	0.8995	(0.0645)	[0.0693]			
$b(2, 1)$	0.6056	(0.0867)	[0.0872]	0.6063	(0.0869)	[0.0870]			
	F=-660.418			F=-660.294					

Table 9A: SSML Estimation of Sample Selection Models
True parameters: $\beta = 1$, $\delta_1 = 1$, $\delta_2 = 0$, $\delta_3 = 1$, $\gamma = 1$, $b(2, 1) = 0.6$, and $b(2, 2) = 0.8$

	Hermite		Hermite-ind-1		Hermite-ind-2	
	Mean	SD	Mean	SD	Mean	SD
Data:	Normal disturbances					
n=200, S=30						
β	0.9947	(0.1371)	0.9947	(0.1371)	0.9969	(0.1420)
γ	1.0250	(0.1377)	1.0244	(0.1380)	0.9993	(0.2293)
	F=-255.724		F=-255.729		F=-253.808	
n=200, S=60						
β	0.9938	(0.1383)	0.9936	(0.1386)	0.9928	(0.1415)
γ	1.0242	(0.1407)	1.0233	(0.1399)	1.0145	(0.2559)
	F=-255.676		F=-255.675		F=-253.807	
n=200, S=100						
β	0.9935	(0.1378)	0.9933	(0.1377)	0.9949	(0.1423)
γ	1.0226	(0.1398)	1.0231	(0.1394)	1.0453	(0.2828)
	F=-255.622		F=-255.629		F=-253.821	
Data:	Gumbel disturbances					
n=200, S=30						
β	0.9947	(0.1490)	0.9954	(0.1493)	0.9957	(0.1447)
γ	1.0235	(0.1347)	1.0197	(0.1321)	1.0115	(0.2205)
	F=-257.662		F=-257.759		F=-255.409	
n=200, S=60						
β	0.9945	(0.1477)	0.9947	(0.1475)	0.9957	(0.1478)
γ	1.0255	(0.1361)	1.0189	(0.1337)	1.0371	(0.2331)
	F=-257.539		F=-257.602		F=-255.424, con# 399	
n=200, S=100						
β	0.9959	(0.1470)	0.9962	(0.1475)	0.9956	(0.1468)
γ	1.0252	(0.1386)	1.0175	(0.1371)	1.0609	(0.2552)
	F=-257.309		F=-257.402		F=-255.196, con# 398	

Table 9B: SSML Estimation of Sample Selection Models
True parameters: $\beta = 1$, $\delta_1 = 1$, $\delta_2 = 0$, $\delta_3 = 1$, $\gamma = 1$, $b(2, 1) = 0.6$, and $b(2, 2) = 0.8$

	Hermite		Hermite-ind-1		Hermite-ind-2	
	Mean	SD	Mean	SD	Mean	SD
Data:	Mixture-1					
n=200, S=30						
β	0.9895	(0.1335)	0.9894	(0.1334)	0.9916	(0.1309)
γ	0.9044	(0.1171)	0.9046	(0.1171)	0.9197	(0.2670)
	F=-261.040		F=-261.048		F=-258.325	
n=200, S=60						
β	0.9894	(0.1334)	0.9894	(0.1335)	0.9903	(0.1201)
γ	0.9051	(0.1172)	0.9029	(0.1166)	1.0077	(0.3367)
	F=-260.918		F=-260.933		F=-257.641, con# 399	
n=200, S=100						
β	0.9894	(0.1331)	0.9896	(0.1335)	0.9901	(0.1214)
γ	0.9088	(0.1196)	0.9050	(0.1163)	1.0843	(0.3991)
	F=-260.859		F=-260.888		F=-257.238, con# 397	
n=500, S=60						
β					0.9964	(0.0711)
γ					0.9690	(0.2838)
					F=-653.506 con# 398	
n=500, S=100						
β					0.9974	(0.0691)
γ					1.0520	(0.3649)
					F=-651.885, con# 398	
Data:	Mixture-2					
Hermite-ind-2;						
	n=200,	S=30	n=200,	S=60	n=200,	S=100
β	1.0001	(0.0946)	0.9963	(0.0861)	0.9965	(0.0840)
γ	0.9551	(0.2806)	1.0170	(0.3502)	1.0809	(0.3934)
	F=-239.638, con# 398		F=-239.092, con# 398		F=-238.507, con# 399	

Table 10: Likelihood Ratio Test - Level of Significance and Power

	N	S	10%	5%	2%	1%
H_0 : Normal		True: Normal				
Hermite	200	30	13.00	6.25	3.50	2.00
Hermite	200	60	12.00	6.50	3.00	1.75
Hermite	200	100	13.00	6.75	2.50	1.25
Hermite-ind-1	200	30	12.75	5.75	3.50	2.00
Hermite-ind-1	200	60	11.00	6.50	3.00	1.75
Hermite-ind-1	200	100	13.25	7.00	2.75	1.50
Hermite-ind-2	200	30	15.50	6.75	3.75	2.25
Hermite-ind-2	200	60	15.75	9.50	4.75	2.50
Hermite-ind-2	200	100	14.75	8.00	3.75	2.75
H_0 : Normal		True: Gumbel				
Hermite	200	30	23.75	16.75	9.75	7.00
Hermite	200	60	24.81	17.04	10.03	6.77
Hermite	200	100	30.75	22.00	13.75	9.00
Hermite-ind-1	200	30	24.00	15.50	7.50	4.75
Hermite-ind-1	200	60	24.75	16.50	9.25	6.00
Hermite-ind-1	200	100	28.75	21.25	12.25	7.25
Hermite-ind-2	200	30	30.50	22.75	14.50	10.25
Hermite-ind-2	200	60	30.08	20.80	13.53	9.27
Hermite-ind-2	200	100	33.17	23.62	14.82	9.30
H_0 : Normal		True: Mixture-1				
Hermite	200	30	17.00	8.75	3.50	1.75
Hermite	200	60	18.75	9.50	4.75	3.00
Hermite	200	100	18.25	9.50	5.75	3.75
Hermite-ind-1	200	30	16.50	8.75	4.00	1.75
Hermite-ind-1	200	60	18.50	9.25	4.00	3.00
Hermite-ind-1	200	100	17.75	8.50	5.50	4.00
Hermite-ind-2	200	30	28.00	18.75	10.50	7.50
Hermite-ind-2	200	60	41.85	27.57	16.04	12.53
Hermite-ind-2	200	100	44.58	33.00	23.68	16.62
Hermite-ind-2	500	60	55.28	46.73	35.18	28.39
Hermite-ind-2	500	100	66.58	55.53	45.73	41.21
H_0 : Normal		True: Mixture-2				
Hermite-ind-2	200	30	67.59	66.08	64.07	63.82
Hermite-ind-2	200	60	75.88	72.86	69.35	67.84
Hermite-ind-2	200	100	80.20	76.19	71.43	69.42

References

- Albert, J. and S. Chib, 1993, Bayesian analysis of binary and polychotomous data, *Journal of the American Statistical Association* 88, 669-679.
- Amemiya, T. 1974, Multivariate regression and simultaneous equation models when the dependent variables are truncated normal, *Econometrica* 42, 999-1012.
- , 1978, The estimation of a simultaneous equation generalized probit model, *Econometrica* 46, 1193-1205.
- , 1984, Tobit models: a survey, *Journal of Econometrics* 24, 3-61.
- Borsch-Supan, A., and V.A. Hajivassiliou, 1993, Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models, *Journal of Econometrics* 58, 347-368.
- Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey, 1983, *Graphical methods for data analysis*, Wadsworth and Brooks/Cole Pub. Co. (Pacific Grove, CA).
- Danielsson, J. and J.F. Richard, 1993, Accelerated gaussian importance sampler with application to dynamic latent variable models, *Journal of Applied Econometrics* 8, S153-S173.
- Duan, N., W.G. Manning, Jr., C.N. Morris, and J.P. Newhouse, 1984, Choosing between the sample selection model and the multi-part model, *Journal of Business and Economic Statistics* 2, 283-289.
- Dubin, J. and D. McFadden, 1984, An econometric analysis of residential electric appliance holdings and consumption, *Econometrica* 52, 345-362.
- Gallant, A.R. and D. Nychka, 1987, Semi-nonparametric maximum likelihood estimation, *Econometrica* 55, 363-390.
- Geman, S. and D. Geman, 1984, Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721-741.
- Geweke, J., 1989, Bayesian inference in econometric models using Monte Carlo integration, *Econometrica* 24, 1317-1339.
- , 1991, Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints, in: *Computer science and statistics: Proceedings of the twenty-third symposium on the interface* (American statistical association, Alexandria, VA) 571-578.
- Geweke, J., M. Keane, and D. Runkle, 1994, Alternative computational approaches to inference in the multinomial probit model, *The Review of Economics and Statistics* 76, 609-632.
- , 1997, Statistical inference in the multinomial multiperiod probit model, *Journal of Econometrics* 80, 125-165.

- Gourieroux, C., and A. Monfort, 1993, Simulation-based inference: a survey with special reference to panel data models, *Journal of Econometrics* 59, 5-33.
- Hajivassiliou, V., 1993, Simulation estimation methods for limited dependent variable models, in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics*, Vol. 11 (Amsterdam, North Holland), 519-543.
- and D. McFadden, 1990, The method of simulated scores for the estimation of LDV models: with an application to the problem of external debt crises, Cowles Foundation discussion paper no. 967, Yale University (New Haven, CT).
- and P. Ruud, 1994, Classical estimation methods for LDV models using simulation, Chapter 40 in: R.F. Engle and D.L. McFadden, ed., *Handbook of Econometrics*, Vol.4, (Amsterdam, North Holland), 2383-2441.
- , D. McFadden, and P. Ruud, 1996, Simulation of multivariate normal orthant probabilities: theoretical and computational results, *Journal of Econometrics* 72, 85-134.
- Hausman, J.A., and D.A. Wise, 1978, A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences, *Econometrica* 46, 403-426.
- Hay, J. and R.J. Olsen, 1984, Let them eat cake: a note on comparing alternative models of the demand for medical care, *Journal of Business and Economic Statistics* 2, 279-282.
- Heckman, J.J., 1990, Varieties of selection bias, *American Economic Association Papers and Proceedings*, 313-318.
- Hendry, D.F. and J.F. Richard, 1992, Likelihood evaluation for dynamic latent variables models. In: H.M. Amman, D.A. Belsley and L.F. Pau (eds), *Computational Economics and Econometrics* (Kluwer, Amsterdam).
- Keane, M.P., 1993, Simulation estimation methods for panel data limited dependent variable models, in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics*, Vol. 11, (Amsterdam, North-Holland).
- , 1994, A computationally practical simulation estimator for panel data, *Econometrica* 62, 95-116.
- Lee, L.F. 1982, Some approaches to the correction of selectivity bias, *Review of Economic Studies* 49, 355-372.
- , 1983, Generalized econometrics models with selectivity, *Econometrica* 51, 507-512.
- , 1992, On efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models, *Econometric Theory* 8, 518-552.
- , 1995, The computation of opportunity costs in polychotomous choice models with selectivity, *The Review of Economics and Statistics* 77, 423-435.
- Maddala, G.S., 1983, *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

- McCulloch, R. and P.E. Rossi, 1994, An exact likelihood analysis of the multinomial probit model, *Journal of Econometrics* 64, 207-240.
- McFadden, D., 1989, A method of simulated moments for estimation of discrete choice models without numerical integration, *Econometrica* 57, 995-1026.
- and P.A. Ruud, 1994, Estimation by simulation, *The Review of Economics and Statistics* 76, 591-608.
- Newey, W., J.J. Powell, and J.R. Walker, 1990, Semiparametric estimation of selection models: some empirical results, *American Economic Review Papers and Proceedings* 80, 324-328.
- Pakes, A., and D. Pollard, 1989, Simulation and the asymptotics of optimization estimators, *Econometrica* 54, 755-785.
- Schmertmann, C.P., 1994, Selectivity bias correction methods in polychotomous sample selection models, *Journal of Econometrics* 60, 101-132.
- Tanner, M.A. and W.H. Wong, 1987, The calculation of posterior distribution by data augmentation, *Journal of the American Statistical Association* 82, 528-550.