COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125 Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1007

Simulation Estimation Methods for Limited Dependent Variable Models

by

Vassilis A. Hajivassiliou

December 1991

Revised 1992

To Appear as a Chapter in

G.S. Maddala, C.R. Rao and H.D. Vinod (editors)

Handbook of Statistics, Vol.11 (Econometrics), North-Holland

Simulation Estimation Methods for Limited Dependent Variable Models

by

Vassilis A. Hajivassiliou*

Cowles Foundation for Research in Economics
Yale University

December 1991

ABSTRACT

This Chapter discusses simulation estimation methods that overcome the computational intractability of classical estimation of limited dependent variable models with flexible correlation structures in the unobservable stochastic terms. These difficulties arise because of the need to evaluate accurately very high dimensional integrals. The methods based on simulation do not require the exact evaluation of these integrals and hence are feasible using computers of even moderate power. I first discuss a series of ideas that had been used in efforts to circumvent these computational problems either by considering more restrictive econometric models or by employing standard numerical analysis approximation methods. I then show how simulation techniques solve the computational problems without the need to resort to either untenable restrictions on the correlation structures or to rely on generally unsatisfactory numerical approximations. All currently known simulation algorithms are then compared in terms of theoretical properties and practical performance.

* Research support through NSF grant No. SES 88—13630 is gratefully acknowledged. I am indebted to Yannis Ioannides for very useful comments. Yoosoon Chang and Carmela Quintos provided expert research assistance.

SIMULATION ESTIMATION METHODS FOR LIMITED DEPENDENT VARIABLE MODELS

1. Introduction

This Chapter discusses estimation methods for Limited Dependent Variable (LDV) models that employ Monte Carlo simulation techniques to overcome numerical intractabilities of such models. These difficulties arise because high dimensional integral expressions need to be calculated repeatedly. In the past, investigators were forced to restrict attention to special classes of LDV models that are computationally tractable. The simulation estimation methods make it possible to estimate LDV models that are otherwise computationally intractable using classical estimation methods even on modern-day supercomputers.

Discussions of the estimation of various limited dependent variable models under restrictive parametric assumptions can be found in Maddala (1983), McFadden (1984), and Amemiya (1984). Typically, the estimation of many of these models involves the evaluation of multivariate integrals. This is generally the case with discrete choice models where the assumption is that each individual i evaluates all J available choices and selects the one that gives the maximum expected utility. Alternative j has observable attributes X_j and yields (random) utility $y_j^* = X_j\beta_j + \epsilon_j$. McFadden (1973) solved the estimation problem by making the restrictive assumption that the error terms ϵ_i in the random utility model are i.i.d. with an extreme value distribution. Under this assumption, the multinomial discrete choice model has the convenient closed form expression:

$$\text{(1.1)} \qquad \text{Prob}(y_i=k) = \exp(X_{ik}\beta_k) / \sum_{j=1}^{J} \exp(X_{ij}\beta_j).$$

Similar distributional assumptions enabled Dubin and McFadden (1984) to derive closed form expressions for a discrete/continuous model of choice of energy appliances and consumption of energy.

The drawback of the i.i.d. extreme-valued distributional assumption is that it severely restricts the pattern of discrete choices predicted by the model, which may be untenable in many realistic circumstances. This restriction, known as the independence of irrelevant alternatives (IIA), implies that the various alternatives are equally substitutable, neglecting for example that when selecting a mode of transportation between a car, a blue bus, and a red bus, the last two options are for all intents and purposes essentially perfect substitutes. See McFadden (1981). A restriction similar to the IIA is also exhibited by a multinomial independent probit model derived by Hausman and Wise (1978) by assuming that the errors ϵ_i are independent $N(0,\sigma_i^2)$.

Heckman (1981) attempted to trade off some of the computational simplicity of these models for a less restrictive structure on the unobservables, thus overcoming frequently undesirable restrictions like the IIA property. He achieved this by imposing a factor-analytic structure on the error vector

$$\epsilon = \Gamma\nu + S\zeta \,, \tag{1.2}$$

where ν is an $n\times 1$ vector of independent random "factors," $(E\nu=0,\ E\nu\nu'=\sigma_\nu^2 I_n)$, Γ is a $J\times n$ matrix of factor loadings, S is a $J\times J$ diagonal matrix, and ζ is a $J\times 1$ i.i.d., random vector $(E\zeta=0,\ E\zeta\zeta'=\sigma_\zeta^2 I_J)$. This implies that $E\epsilon\epsilon' = \sigma_\nu^2\Gamma\Gamma' + \sigma_\zeta^2 SS$, which reduces the dimension of the integral necessary in calculating choice probabilities. Sometimes there exists a natural economic interpretation of such a factor analytic structure, thus further increasing the appeal of this model. For example, ν can denote a small number of unmodelled factors affecting consumers' choice among airline carriers, like safety, comfort, on-time performance, etc. In a panel data context, ν can denote unobserved agent-specific and time-period-specific shocks, which make up the structure (1.2).

Another model that introduced some further computational complexity in exchange for avoiding the IIA property, is the nested multinomial model (NMNL) due to McFadden (1984). By introducing a hierarchical structure in the choice process, corresponding to blocks of 0's in the covariance matrix of ϵ, this model remains consistent with the random

utility maximization hypothesis, exhibits the IIA property only for specific stages of the hierarchical decision making, and still avoids the need to calculate integrals of high dimension.

Avery et al. (1983), Chamberlain (1984), and Poirier and Ruud (1988) noted that in certain LDV models, notably multiperiod binary response and censored models with normally distributed disturbances, incorrect specification of the error covariance structure results in inefficient but still consistent and asymptotically normal estimators. Hence, these authors developed estimation techniques that rely on period-by-period orthogonality conditions, which are computationally tractable. As already noted, however, these methods do not generalize to most LDV models, in which the covariance structure of the error terms determines crucially the joint as well as the marginal distributions of the observed limited dependent variables. Moreover, as Hajivassiliou (1986) points out, normality of the disturbances is a crucial requirement for these methods, since for general non-normal random variables, the joint, conditional, and marginal distributions all belong to different classes.

The other direction followed by researchers was to investigate numerical analysis techniques for approximating the J-dimensional integrals. Hausman and Wise (1978) showed that methods due to Divgi (1979) for the case J=3 and Steck (1958) for J=4 provided reasonably fast and accurate approximations. Clark (1961) and Daganzo et al (1977) offered some approximations for multivariate normal integrals. Horowitz et al (1981) investigated the accuracy of the Clark approximation and found that it can be quite poor.

The simulation methods discussed in this chapter offer a better solution than all of the alternative ways to address these computational problems. Section 2 discusses the key developments that let to simulation techniques suitable for the estimation of LDV models. Section 3 describes the leading methods analytically. Section 4 examines all known simulators for likelihood contributions and derivatives that are necessary for implementing the simulation estimation methods. Section 5 concludes with a summary and suggestions

for future research.

2. Developments of Simulation Techniques for the Estimation of LDV Models

An important innovation occurred in 1981, in the form of the following idea due to Lerman and Manski (1981). These authors noted that Monte Carlo simulation methods are used in numerical analysis to approximate high dimensional integrals. [See, for example, Hammersley and Handscomb (1964).] They used simulation to evaluate the choice probabilities for the multinomial discrete choice probit model. In this model, the utility derived from alternative j is assumed to be $y_j^* \sim \mathcal{N}(X_j\beta_j,\Omega_j)$. Lerman and Manski used the empirical frequency of observing $y_k^{*r} = \max\{y_1^{*r},\ldots,y_J^{*r}\}$ to approximate $P_k \equiv \text{prob}(y=k;\beta,\Omega)$, where y^{*r} is a Monte Carlo draw of a latent vector $y^* \sim \mathcal{N}(X\beta,\Omega)$, r=1,...,R. I will term this the *crude frequency simulation* (CFS) method. Their experience with this idea was disappointing; a huge number of simulations R was needed before the empirical frequency $\tilde{P}_k$, defined by $\tilde{P}_k \equiv \frac{1}{R}\sum_{r=1}^{R} 1(y_k^{*r}=\max\{y_1^{*r},\ldots,y_J^{*r}\})$, provided a good approximation of P_k, particularly when P_k was close to 0 or 1.[1] This of course can be explained by the fact that since $1(\cdot)$ is a Bernoulli random variable with probability of success P_k, the variance of the simulator $\tilde{P}_k$ is $P_k(1-P_k)/R$, which falls only linearly with R. Moreover, the frequency simulator $\tilde{P}_k$ takes the value 0 with positive probability. Hence, attempts to estimate a discrete choice model by using the method of Simulated Maximum Likelihood (SML) to maximize iteratively $\sum_i \log\tilde{P}_{ki}$ yielded very unsatisfactory estimates.

The breakthrough came with an idea by McFadden, first presented in 1986 and eventually published as McFadden (1989). McFadden's contribution was to recognize that using simulation to approximate choice probabilities in the context of estimation is a viable procedure even for a small number of simulations, R, provided the following conditions are

[1] The indicator function 1(A) of event A takes the value 1 if the event is true and 0 if false.

satisfied. (a) An unbiased simulator is used for the choice probabilities (which is satisfied by the crude frequency simulator of Lerman and Manski (1981)). (b) The functions to be simulated appear linearly in the conditions defining the estimator. (c) The same set of random draws is used to simulate the model at different trial parameter values in the process of iteratively searching for the estimator. Condition (b), violated by the SML procedure since the P_k's appear inside the $\log(\cdot)$ function, is crucial in ensuring that the independent simulation errors made in approximating each P_{ik} average out over the N observations to $E(P_{ik}-\tilde{P}_{ik})$, which is 0 by requirement (a).[2] Conditions for simulation-based estimators to be consistent and asymptotically normal were also derived in another seminal paper by Pakes and Pollard (1989), which appeared at about the same time. The method that these three authors developed satisfies requirements (a), (b), and (c) above and became known as the Method of Simulated Moments (MSM). The key point is that the choice probabilities appear linearly in the moment conditions, since the expected value of a Bernoulli dependent variable is the probability of its taking the value 1.

Based on a suggestion by Ruud (1986), Hajivassiliou and McFadden (1990) noted that another simulation estimation method can be devised by simulating directly the scores of the likelihood function. This method of simulated scores (MSS) estimator is thus defined by

$$\hat{\theta}_{MSS} = \underset{\theta}{\text{argsolves}} \{ \frac{1}{N} \sum_i \tilde{s}_i(\theta; y_i) = 0 \}, \tag{2.1}$$

and it will be consistent, uniformly asymptotically normal (CUAN) provided unbiased simulators $\tilde{s}_i(\cdot)$ of the score vector $s_i(\cdot)$ are used. Van Praag and Hop (1987) were the first to propose the use of simulation to overcome the computational difficulties in the evaluation of likelihood scores of LDV models and also noted that this approach is applicable to all classical LDV models (and not just discrete choice problems,) though they did not use the name "Method of Simulated Scores." Their method relied on approximating the score expression by simulating both the linear derivative in the numerator and the likelihood in the denominator and taking the ratio as an approximation. It should be noted that, given the non-linearity of this expression, consistency and asymptotic normality of the resulting estimator require that the simulations of the denominator expression be based on an infinite number of draws. By contrast, the approach by Ruud, Hajivassiliou, and McFadden uses unbiased simulation to guarantee consistency and asymptotic normality for a finite number of simulations. The method proposed by van Praag and Hop (1987) has been applied in van Praag (1989), van Praag, Hop, and Eggink (1991), and Eggink, Hop, and van Praag (1992).

The next simulation-based estimation method, also to be discussed in greater detail

---

[2] The original SML procedure of Lerman and Manski also violated condition (c).

later, is the method of simulated pseudo ML (SPML), developed by Laroque and Salanié (1989) for the estimation of disequilibrium models with correlation over time. This method is defined as follows: Suppose a model stipulates that a dependent variable y_i has conditional moments, given the values of the explanatory variables, $E(y_i|X_i) = g_1(X_i,\theta_1)$ and $V(y_i|X_i) = g_2(X_i,\theta_2)$. In the context of the LDV models we are discussing, the $g(\cdot)$ functions involve integrals of high dimensions, so that they are difficult to evaluate. The SPML method simulates R times the dependent variable y_i at the current trial parameter values θ_1 and θ_2, and then uses the empirical mean and variance of $\{y_i^{(r)}\}$ as simulating functions $\tilde{g}_1(\theta_1;R)$ and $\tilde{g}_2(\theta_2;R)$. By construction, these functions are consistent estimators for $g(\cdot)$ as $R \to \infty$. The SPML estimator, $\hat{\theta}_{SPML}$, is thus defined by

$$\hat{\theta}_{SPML} = \underset{\theta_1,\,\theta_2}{\operatorname{argmax}} \{-\sum_i \left\{ [y_i-\tilde{g}_1(\theta_1;R)]^2 / \tilde{g}_2(\theta_2;R) + \log[\tilde{g}_2(\theta_2;R)] \right\} \}. \tag{2.2}$$

This amounts to a combination of simulation of integral expressions together with pseudo-maximum likelihood estimation. The method implicitly employs the assumption that the limited dependent variable y_i has a distribution in the linear exponential class. In this specific example (2.2) the distribution is assumed to be normal.[3] It should be noted that SPML is closer in spirit to the SML of Lerman and Manski (1981), in that consistency and asymptotic normality requires $R \to \infty$.

Finally, a method that theoretically also requires $R \to \infty$ is the smooth simulated maximum likelihood (SSML) method due to Börsch-Supan and Hajivassiliou (1990). These authors showed that the linearity requirement that apparently caused the unsatisfactory performance of the SML estimator was not critical if extremely accurate simulators are used that are smooth functions of the underlying parameters. In particular, they showed that using the GHK/SRC simulator explained in Section 4 below makes the resulting SSML estimator perform excellently, even for numbers of simulations R as small as 5. Apart from smoothness, this simulator differs critically from the CFS used by Lerman and

[3] See Gourieroux et al (1984a, 1984b) for the PML method without simulation.

Manski in that it is bounded away from 0 and 1 and has extremely low variance, even when the true probability is very close to either 0 or 1.

Evaluation of high-dimensional integrals has also been considered by applied physicists. For example, Dutt (1973 and 1976) discusses the special case of multivariate normal orthant probabilities. Deák (1980a and 1980b) proposes Monte Carlo integration as a way of approximating such probabilities. Another field in which fast and accurate calculation of multivariate integrals is important is Bayesian statistics, where it is necessary to calculate posterior density expressions in the context of Bayesian inference. Selected work from this voluminous literature that has investigated Monte Carlo simulation of integrals is Bauwens (1984), van Dijk (1987), Kloek and van Dijk (1978), Geweke (1989a), and West (1990).

Finally, a third field that faces intractable integration problems is the estimation of econometric models of dynamic optimization. See Rust (1989) for an approach that circumvents some of the problems by using convenient distributional assumptions and an ingenious nested fixed point algorithm. Recently Hotz and Miller (1989), Hotz and Sanders (1990), and Hotz, Miller, Sanders, and Smith (1991) have proposed introducing simulation techniques in the estimation of such problems.[4] These authors' preliminary work uses the simulated method of moments estimation framework to recover individual valuations of discrete alternatives by inverting observed conditional choice probabilities. They apply this methodology to the dynamic model of bus replacement considered by Rust (1989) to overcome some of the computational difficulties with the estimation of structural dynamic discrete choice models. In this work, however, no proofs of the asymptotic properties of the technique are yet available. Based on the Hotz et al. approach, Rust (1991) develops the *simulated value function* estimator, which is consistent and asymptotically normal, as well as computationally tractable.

---

[4] Two reviews that describe the state of the art in estimating dynamic optimization models are Rust (1991) and Pakes (1991).

3. Simulation Estimation Methods for LDV Models

We shall discuss first the leading simulation estimation methods for the multinomial discrete choice model. This does not entail any loss of generality, because more general discrete/continuous LDV models can be decomposed into a continuous part and a discrete choice part, conditional on the continuous part. Specifically, suppose that for a given individual, the observed limited dependent variable vector y has some discrete elements, denoted by y_d, and some fully observed (continuous) ones, y_c. For example, consider the discrete/continuous model, where individual i chooses alternative j which provides the highest level of utility, and simultaneously takes a continuous decision y^*_{ij} which depends on the discrete choice. The observed limited dependent variables (w_i, y_i) are then defined by:

$$
\begin{aligned}
(3.1) \qquad & w^*_{ij} = Z_{ij}\gamma_j + \epsilon_{ij} \qquad\qquad j=1,\ldots,J \\
& y^*_{ij} = X_{ij}\beta_j + u_{ij} \\
& w_i = \underset{j}{\operatorname{argmax}}(w^*_{i1},\ldots,w^*_{ij},\ldots,w^*_{iJ}) \\
& y_i = y^*_{i w_i}.
\end{aligned}
$$

In this case, y_c corresponds to the J elements of the y^*_{ij} vector, while y_d to the J elements of w^*_{ij}.

Let $y=(y_d,y_c)'$ and let θ be the vector of unknown parameters in the model. Then

$$(3.2) \qquad \ell(y;\theta,X) = \ell(y_d \mid y_c;\theta,X)\cdot\ell(y_c;\theta,X)$$

and so

$$(3.3.1) \qquad \ln\ell = \ln\ell_{d|c} + \ln\ell_c$$

and

$$(3.3.2) \qquad \partial\ln\ell/\partial\theta = \partial\ln\ell_{d|c}/\partial\theta + \partial\ln\ell_c/\partial\theta\,.$$

The parts with only continuous components pose no difficulties, since they involve only (multivariate) continuous densities, which can be evaluated directly. The terms involving the probability $\ell_{d|c} \equiv \text{Prob}(y_d^* \in \{y_d^* : y_d = \underset{j}{\text{argmax}}(y_1^*,\ldots,y_j^*)\} \mid y_c^*)$ are integral expressions over a conditional density, of dimension equal to the number of discrete elements in y, which simulation estimation methods approximate.

In the discrete choice model, J mutually exclusive and exhaustive alternatives yield a vector of utilities or payoffs $y_i^* = X_i\beta + \epsilon_i$. The disturbance vector ϵ_i is i.i.d., mean-independent of X_i, with density function $f_\epsilon(\epsilon_i;\Omega_i(\sigma))$. Individual i chooses the alternative k which yields the highest payoff. The observed choice is characterized by the J×1 dummy variable vector y_i, with $y_{ij}=1$ for j=k, and $y_{ij}=0$ otherwise. In this case, the vector of unknown parameters is $\theta=(\beta',\sigma')'$. Given the non-linearity of this model, all estimation methods we consider will involve iterative search. Consider the trial parameter vector $\theta^{(n)}$ at iteration n. A maximum likelihood estimation algorithm requires the evaluation of $L(\theta^{(n)})$

$$(3.4) \qquad L(\theta^{(n)}) = \frac{1}{N} \cdot \sum_{i=1}^{N} \ln \ell_i(\theta^{(n)};y_i) \,.$$

A method of scoring seeks to evaluate

$$(3.5) \qquad S(\theta^{(n)}) = \frac{1}{N} \cdot \sum_{i=1}^{N} s_i(\theta^{(n)};y_i) = \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{\ell_{i\theta}(\theta^{(n)};y_i)}{\ell_i(\theta^{(n)};y_i)} \,.$$

A method of moments algorithm calculates

$$(3.6) \qquad M(\theta^{(n)}) = \frac{1}{N} \cdot \sum_{i=1}^{N} m_i(\theta^{(n)};y_i) = \frac{1}{N} \cdot \sum_{i=1}^{N} w(\theta^{(n)};X_i)'(y_i - g_{1i}(\theta^{(n)};X_i)) \,,$$

where $g_{1i}(\cdot) \equiv E(y_i;\theta^{(n)})$ and $w(\cdot)$ is an instrument function. Finally, a SPML routine evaluates the quadratic form

$$(3.7) \qquad Q(\theta^{(n)}) = \frac{1}{N} \cdot \sum_{i=1}^{N} \left\{ [y_i - g_{1i}(\theta^{(n)};X_i)]^2 \,/\, g_{2n}(\theta^{(n)};X_i) + \log[g_{2n}(\theta^{(n)};X_i)] \right\},$$

where $g_{1i}(\cdot)\equiv E(y_i;\theta^{(n)})$ and $g_{2i}(\cdot)\equiv V(y_i;\theta^{(n)})$.

If the expressions $\ell_i(\theta^{(n)};y_i)$, $\ell_{i\theta}(\theta^{(n)};y_i)$, $s_i(\theta^{(n)};y_i)$, $g_{1i}(\theta^{(n)};y_i)$, and $g_{2i}(\theta^{(n)};y_i)$ are analytically or numerically tractable, computer routines can be written to evaluate these expressions as functions of any possible trial parameter vector $\theta^{(n)}$. Estimation by simulation relies on simulating routines we shall denote by $\tilde{\ell}_i(\theta^{(n)};y_i,R)$, $\tilde{\ell}_{i\theta}(\theta^{(n)};y_i,R)$, $\tilde{s}_i(\theta^{(n)};y_i,R)$, $\tilde{g}_{1i}(\theta^{(n)};y_i,R)$, and $\tilde{g}_{2i}(\theta^{(n)};y_i,R)$ respectively. These routines are defined as follows: Draw a set of R uniform J-dimensional random vectors $\tilde{u}_r^1,\ldots,\tilde{u}_r^r,\ldots,\tilde{u}_r^R$. By the assumptions of this model, the disturbance vector ϵ_i is i.i.d. with density function $f_\epsilon(\epsilon_i;\theta^0)$, with $E(\epsilon_i|X_i)=0$ and $E(\epsilon_i\epsilon_i'|X_i)=\Omega_i(\sigma)$.

Consider the trial parameter vector $\theta^{(n)}=(\beta^{(n)},\sigma^{(n)})'$. Using the inverse of the cumulative distribution function of ϵ, $F_\epsilon^{-1}(\cdot)$, the uniform draws u_i^r can be transformed into a set of R $\tilde{\epsilon}_i$'s,

$$\tilde{\epsilon}_i^r(\sigma^{(n)}) = F_\epsilon^{-1}(\tilde{u}_i^r;\Omega(\sigma^{(n)})), \tag{3.8}$$

which will imply a set of R simulated latent vectors $\tilde{y}_i^{*r}(\theta^{(n)})$, using the specification $y_{ij}=1$ for $y_{ij}=\max(y_{1i},\ldots,y_{iJ})$ and $y_{ij}=0$ otherwise. From the R simulated $\tilde{y}_i^r$ vectors, we calculate the empirical counterparts of the $\ell_i(\cdot)$ etc. functions and thus define the simulators $\tilde{\ell}_i(\theta^{(n)};y_i,R)$, $\tilde{\ell}_{i\theta}(\theta^{(n)};y_i,R)$, $\tilde{s}_i(\theta^{(n)};y_i,R)$, $\tilde{g}_{1i}(\theta^{(n)};y_i,R)$, and $\tilde{g}_{2i}(\theta^{(n)};y_i,R)$. For example, $\tilde{\ell}_i(\cdot)$ and $\tilde{g}_{1i}(\cdot)$ will correspond to the vectors of empirical frequencies with which alternative j yielded the highest simulated payoff. Keeping the same uniform random variates, $\tilde{u}_i^r$, a new trial parameter vector $\theta^{(n+1)}$ will imply a new set of simulated $\tilde{\epsilon}_i^r(\sigma^{(n+1)})$, leading to new $\tilde{y}_i^*$ and hence new values for $\tilde{\ell}_i$, etc. The iterative search algorithms will keep trying different parameter vectors θ, always using the same $\tilde{u}_i^r$'s, to satisfy the relevant criterion, i.e., simulated ML will attempt to maximize $\tilde{L}(\theta)$, MSS will set $\tilde{S}(\theta)$ and MSM will set $\tilde{M}(\theta)$ as close to 0 as possible, while SPML will attempt to minimize $\tilde{Q}(\theta)$.

The main theoretical properties of such estimators can be summarized following arguments in Hajivassiliou and McFadden (1990). Suppose $\hat{\theta}$ is a simulation estimator that

solves $0 = N^{1/2}E_N\tilde{q}(\hat{\theta},\eta)$, where $\tilde{q}$ is an approximation (involving Monte Carlo elements η) to a function $\gamma(\theta)$ of ℓ and its derivatives that has expectation zero at the true parameter θ^o. For example, for SML and MSS, $\gamma(\theta)$ is the logarithmic score of the likelihood function, $S(\theta)$, whereas for MSM $\gamma(\theta)$ is the orthogonality condition $m(\theta)$. E_N denotes empirical expectation over an independent sample of size N. Then, one can write

$$(3.9)\qquad 0 = N^{1/2}E_N\tilde{q}(\hat{\theta},\eta) \equiv N^{1/2}E_N\gamma(\theta^o) + N^{1/2}E_N[\tilde{q}(\theta^o,\eta) - \gamma(\theta^o)]$$
$$+ N^{1/2}E_N[\gamma(\hat{\theta}) - \gamma(\theta^o)] + N^{1/2}E_N[\tilde{q}(\hat{\theta},\eta) - \gamma(\hat{\theta}) - \tilde{q}(\theta^o,\eta) + \gamma(\theta^o)] .$$

Under standard regularity conditions, the first term is asymptotically normal, reflecting the noise in the observations, and the third term is proportional to $\sqrt{N}(\hat{\theta} - \theta^o)$.[5] The last term will be of order $o_p(1)$ for simulators that satisfy a "stochastic equicontinuity condition".[6]

Let $\tilde{q}_i(\theta,\eta)$ denote the simulated value of $q_i(\theta)$, for a sample of independently, identically distributed observations i=1,...,N. Define a simulation bias,

$$(3.10)\qquad B_N(\theta) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[E_\eta\tilde{q}_i(\theta,\eta) - q_i(\theta)] ,$$

where E_η denotes an expectation with respect to the simulation process, given the observation. Following the method of McFadden (1989) and Pakes and Pollard (1989),

---

[5] When q is a smooth function of crude frequency simulators of ℓ, $\partial\ell/\partial\theta$, and $\partial\ln\ell/\partial\theta$ obtained using R Monte Carlo draws, Hajivassiliou and McFadden (1990) show that the second term will behave like $\sqrt{N/R}$ times an expression that is asymptotically normal, so that it will be comparable in magnitude to the first term when R and N are proportional. If, in addition, there is any averaging out of simulation noise across observations, the second term may be of order $o_p(1)$ when R and N are proportional, or comparable in magnitude to the first term for fixed R.

[6] The functions $\{\zeta_N(\cdot)\}$ are stochastically equicontinuous at $\Theta_1 \subseteq \Theta$ if for each $\epsilon>0$ and $\lambda>0$, there exists $\delta > 0$ and N_o such that for $N \geq N_o$,

$$\text{Prob}\Big(\sup_{\substack{|\theta - \theta'|<\delta \\ \theta'\in\Theta,\,\theta\in\Theta_1}} |\zeta_N(\theta) - \zeta_N(\theta')| > \epsilon\Big) < \lambda .$$

Hajivassiliou and McFadden (1990) show that assumptions on the simulation bias plus regularity assumptions, are sufficient for the simulation estimator $\hat{\theta}_N$ that solves $\sum_{i=1}^{N} \tilde{q}_i(\hat{\theta}_N,\eta) = 0$ to be consistent and asymptotically normal.[7]

Theorem. (Hajivassiliou and McFadden (1990)) *Assume that the parameter θ is contained in a compact set Θ, and that the true value θ^o is in the interior of Θ. Assume that the criterion $q_i(\theta)$ is continuously differentiable on Θ. Assume that the criterion and its derivatives are dominated by a function independent of θ with finite first and second order moments. Assume that $\mathbf{E}_i q_i(\theta) = 0$ if and only if $\theta = \theta^o$, and that $J = -\mathbf{E}_i q_{i\theta}(\theta^o)$ is positive definite, where $\mathbf{E}_i$ denotes expectation with respect to the distribution of the observations. Assume that the observations and simulators are independently identically distributed across observations. Assume that (i) the simulation bias converges to zero in probability, uniformly in θ, and (ii) the simulation residual process[8] is stochastically equicontinuous. Assume that a MSS estimator solving $0 = \sum_{i=1}^{N} \tilde{q}_i(\hat{\theta}_N,\eta)$ exists for each N.[9] Then, the estimator satisfies*

$\hat{\theta}_N \xrightarrow{p} \theta^o$ *and*

$\sqrt{N}(\hat{\theta}_N - \theta^o) \xrightarrow{d} Z \sim \mathcal{N}(0, J^{-1}+J^{-1}VJ^{-1})$,

where $V = \mathbf{E}[\tilde{q}_i(\theta^o,\eta) - \mathbf{E}_\eta \tilde{q}_i(\theta^o,\eta)][\tilde{q}_i(\theta^o,\eta) - \mathbf{E}_\eta \tilde{q}_i(\theta^o,\eta)]'$.

To avoid technical difficulties, Hajivassiliou and McFadden (1990) assume that the

---

[7] The proof of the following Theorem can be found in Appendix 3 of Hajivassiliou and McFadden (1990).

[8] See Hajivassiliou and McFadden (1990) for precise definitions.

[9] It is sufficient to define $\hat{\theta}_N$ to be an approximate solution satisfying

$$\mathcal{O}(1) = \sum_{i=1}^{N} \tilde{q}_i(\hat{\theta}_N) ;$$

such an estimator always exists.

multivariate distribution y^* is truncated to a large compact rectangle.[10] The simulators $\tilde{\ell}_i(\theta)$, $\tilde{\ell}_{i\theta}(\theta)$, and $\tilde{s}_i(\theta)$ can be formed in several methods. These are discussed in the following section. Hajivassiliou and McFadden (1990) give general sufficient conditions for assumption (i) of asymptotic unbiasedness and assumption (ii) of stochastic equicontinuity in the Theorem. Specifically, they show that if the simulation process is unbiased, as for example in MSM and MSS with unbiased simulators $\tilde{\ell}_i(\theta)$ and $\tilde{s}_i(\theta)$, or when the bias in an observation is dominated by a positive function independent of θ whose expectation is of order $a(1/\sqrt{N})$, then the simulation bias converges to 0. This results in consistent, asymptotically normal simulation estimators. McFadden and Ruud (1991) give further theoretical results along these lines. The suggestion to simulate the likelihood scores directly using the conditional distribution $Z_i \equiv \{y_i^* \mid y_i^* \epsilon D(y_i)\}$, where $D(y) \equiv \{y^* : y = \underset{j}{\text{argmax}}(y_1^*,\ldots,y_J^*)\}$, is due to Ruud (1986).

It is important to note that the method of simulated scores differs in a very significant way from the method of simulated moments. MSS implicitly uses the *optimal* set of instruments, since as $R\to\infty$ it corresponds to MLE which is statistically efficient. On the other hand, the efficiency of MSM rests crucially on the choice of the instrument function $w(\theta;X_i)$. To highlight this issue, consider as a simple example the binary probit model for an independent cross-section of individuals, i=1,...,N, for which classical estimation is, of course, computationally very straightforward.

$$
\begin{aligned}
(3.11) \qquad & y_i^* = x_i'\beta + \epsilon_i \quad \epsilon_i \sim N(0,1) \\
& y_i = 1 \qquad d_i = 1 \text{ if } y_i^* > 0 \quad (y_i = 2d_i - 1) \\
& \quad = -1 \qquad = 0 \text{ if } y_i^* \le 0 .
\end{aligned}
$$

Define

[10] This does not entail any essential loss of empirical generality, since for distributions with unbounded support, like the multivariate normal, one can restrict attention to the same distribution truncated to the square defined by the limits of computing machine representation of floating-point numbers.

$$\text{(3.12a)} \qquad \ln \ell_i = \ln \Phi(y_i \cdot x_i'\beta)$$

$$\text{(3.12b)} \qquad = d_i \cdot \ln \Phi(x_i'\beta) + (1-d_i)\cdot \ln(1-\Phi(x_i'\beta))$$

and

$$\text{(3.13a)} \qquad s_i = \ell_{i\theta}/\ell_i = x_i \cdot \frac{\phi(y_i \cdot x_i'\beta)}{\Phi(y_i \cdot x_i'\beta)} \cdot y_i = x_i \cdot E(\epsilon_i \mid y_i^* \in D(y_i))$$

$$\text{(3.13b)} \qquad = x_i \cdot \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)(1-\Phi(x_i'\beta))} \cdot (d_i - \Phi(x_i'\beta)) = w_i(\theta)\cdot(d_i - \Phi(x_i'\beta))$$

In this case, $\theta=\beta$. Then maximum likelihood estimator solves the first order conditions $L_{N\theta}(\hat{\theta}) = \frac{1}{N}\sum_i s_i(\hat{\theta}) = 0$.

Equation (3.13b) for the score of observation i highlights a method-of-moments interpretation of maximum likelihood estimation when the optimal instruments $w_i(\theta)$, defined in (3.13b), are used. Simulating the conditional expectation expressions in equation (3.13a) corresponds to the method of scoring. It should be noted that the basic consistency requirement that $E(s_i(y_i;\theta^*)|x_i) = 0$ is satisfied; in equation (3.13a) it is satisfied because $P(y_i|\theta^*,x_i) = \Phi(y_i \cdot x_i'\theta^*)$ and in equation (3.13b) because $E(d_i|\theta^*,x_i) = \Phi(x_i'\theta^*)$.

The original method of simulated moments (McFadden (1989) and Pakes and Pollard (1989)) substitutes an unbiased simulator, $\tilde{\Phi}(x_i'\beta)$, for $\Phi(x_i'\beta)$ and exploits the linearity in $(d_i-\Phi(\cdot))$ of the score expression (3.13b). For high efficiency this method requires that consistent estimators for the optimal instruments, $w_i(\theta^*)$, be used. This is confirmed by the theoretical and Monte Carlo results of Hajivassiliou (1990), who reached the following conclusions. First, the optimal instrument function w(.) in (3.13b) (which of course in more realistic cases is intractable to calculate) yields considerable mean-square-error advantages over the simpler choice x_i, which choice also satisfies the theoretical requirements for consistency and asymptotic normality. Second, the SML method of Lerman and Manski (1981) that uses the crude frequency simulators for the choice probabilities (SML/CFS) offers satisfactory performance only when the number of

simulations employed is very large. This number grows faster than linearly with the complexity of the LDV model under analysis.[11] The method that simulates separately the denominator of the scores by frequency methods performs unsatisfactorily, and it is easily dominated by all the other methods tried, primarily because frequency simulators are not bounded away from 0 and 1. Before barely satisfactory performance is achieved, a huge number of simulations for the denominator expressions in (3.13) must be employed. These problems are significantly alleviated once a smooth simulator, bounded away from 0, like the GHK/SRC simulator described in the next section, is used for the denominator expression. In all the cases investigated, the MSS based on the GHK/SRC simulator performs impressively; it approaches the (optimal) performance of MLE with even 2 simulations per dimension of the underlying latent variable vector. Moreover, the method is found to be numerically stable, which was to be expected given its continuity in the underlying parameters. Efficiency properties of various simulation estimation methods are also discussed in Lee (1990).

The method of simulated scores bypasses the issue of searching for the optimal instruments and simulates instead *directly* the expression $E(\epsilon_i^* \mid y_i^* \in D(y_i))$ which implies that the optimal instruments are now available automatically in the form of x_i. In other words, MSS uses simulators for the expression $E(\epsilon_i^* \mid y_i^* \in D(y_i))$, say $\tilde{E}(\epsilon_i^* \mid y_i^* \in D(y_i))$. To see the relation of MSS to MLE, recall that $x_i \cdot E(\epsilon_i^* \mid y_i^* \in D(y_i)) = x_i \cdot \frac{\phi(y_i \cdot x_i'\beta)}{\Phi(y_i \cdot x_i'\beta)} = s_i(y_i, \beta; x_i)$. The Lerman and Manski (1981) method uses unbiased and consistent frequency simulators of $\Phi(x_i'\beta)$ directly in the likelihood function (3.12a). Van Praag and Hop (1987) simulate both the linear derivative in the numerator and the likelihood in the denominator in (3.13a) and take the ratio as an approximation. They use this approach in several applications: Van Praag (1989), van Praag, Hop, and Eggink (1991), and Eggink, Hop, and van Praag (1992).

Gourieroux and Monfort (1990) discuss simulation estimation techniques for models with heterogeneity in terms of consistency, asymptotic normality, convergence rates, and

[11] As theory suggests, the SML/CFS is improved significantly by maintaining the same set of underlying random variates while iterating the optimization algorithm to convergence.

asymptotic bias. A distinction is made between the case where simulations are the same across observations as opposed to different. Several studies have attempted to generalize the theoretical results outlined in this section about the properties of simulation estimators by relaxing the random sample assumption we have been employing. Duffie and Singleton (1989) use similar results to develop a simulation estimation method for models for financial data with temporal dependence. Laroque and Salanié (1989) propose simulation estimation methods for multi-market disequilibrium using SPML. Laroque and Salanié (1990) carry out an extensive Monte Carlo study to evaluate the relative performance of the various algorithms for disequilibrium models. This study finds that the SPML methods are less liable to spurious maxima. Moreover, the use of simulation does not seem to entail substantial efficiency losses compared to full ML, which is considerably less tractable computationally. Finally, Lee and Ingram (1991) propose simulation to overcome estimation problems of dynamic optimization under uncertainty. The main differences from McFadden (1989) and Pakes and Pollard (1989) are that Lee and Ingram allow for disturbances to be correlated across units of observations, while they require the criterion functions defining the estimators be continuous functions of the underlying parameters.

As we shall see in the next section, some simulators like the Crude Frequency Simulator (CFS) are not continuous functions of the unknown parameter vector or of the underlying Monte Carlo draws. Such a feature necessitates the use of derivative-free methods like the non-linear simplex algorithm of Nelder and Meade (1964), since standard derivative-based optimization methods like Gauss-Newton or Newton-Raphson[12], cannot be used. Ruud (1991) showed that the EM algorithm appears promising to overcome some of the speed problems of MSS. For certain types of LDV models, Berry and Pakes (1990) report satisfactory numerical stability of simulation estimators, even when based on non-smooth simulation algorithms.

The list of applications of simulation estimation techniques to estimate modern

[12] See Quandt (1983).

econometric models is growing rapidly. Hajivassiliou and McFadden (1990) apply the method of simulated scores to a model of external debt repayments problems using a panel data set of developing countries. Bolduc and Kaci (1991) use both SML and MSM to estimate a discrete choice model among five local residential telephone services. Keane (1990) develops a variant of the GHK/SRC simulator for multi-period choice probabilities and uses it to study temporal dependence in employment and wages. Berkovec and Stern (1991) use the method of simulated moments to investigate retirement decisions of older men and Stern (1991) also uses MSM to study the choice of transportation mode by elderly and disabled people. Hajivassiliou and Ioannides (1991) estimate by the method of simulated scores a generalized switching Euler equation model, to study consumption and labor supply behavior in the face of borrowing constraints using a large longitudinal data set of households. Bloemen and Kapteyn (1991) estimate a labour supply function jointly with a wage equation using smooth simulated maximum likelihood estimation following Börsch-Supan and Hajivassiliou (1990). Börsch-Supan et al (1991) also apply SSML to investigate a multi-period, multinomial model of housing choices by the elderly.

4. Simulators for ℓ, $\partial\ell/\partial\theta$, and $\partial\ln\ell/\partial\theta$

In the canonical LDV model analyzed here, the latent vector $y_i^* \in \mathbb{R}^M$ gives rise to the observed limited dependent variable vector $y_i \equiv \tau(y_i^*)$, such that $\tau(\cdot)$ partitions $\mathbb{R}^M$ into regions $d_i=1,\ldots,M$. Let $D(X_i,\theta,d)$ denote the set of y_i^* that map into d_i. These regions typically correspond to a set of linear inequality constraints on the elements of y_i^* of the form

$$(4.1) \qquad \{ y_i^* \in D(X_i,\theta,d_i) \} \equiv \{ a(X_i,\theta,d_i) \le W(X_i,\theta,d_i).y^* \le b(X_i,\theta,d_i) \} ,$$

where $W(\cdot)$ is a positive definite matrix and the boundary vectors $a(\cdot)$ and $b(\cdot)$ are allowed to have infinite elements. It is not difficult to see that the leading LDV models considered in the literature fit into this framework.

The estimation methods discussed in the previous section require fast and accurate simulation of the likelihood contribution, its derivatives, and the score, respectively:

$$(4.2.1) \qquad \ell(D_i,\theta) = \int 1(y_i^* \epsilon D_i)\cdot f(y_i^* - \mu_i;\Omega_i(\sigma))dy_i^*$$

$$(4.2.2) \qquad \nabla_\theta \ell(D_i,\theta) = \int 1(y_i^* \epsilon D_i)\cdot \nabla_\theta f(y_i^* - \mu_i;\Omega_i(\sigma))dy_i^*$$

$$(4.2.3) \qquad \nabla_\theta \ln \ell(D_i,\theta) \equiv s(D_i,\theta) =$$

$$\int 1(y_i^* \epsilon D_i)\cdot \nabla_\theta f(y_i^* - \mu_i;\Omega_i(\sigma))dy_i^* \;/\; \int 1(y_i^* \epsilon D_i)\cdot f(y_i^* - \mu_i;\Omega_i(\sigma))dy_i^*$$

where $\mu_i \equiv X_i\beta$ and $f(\cdot)$ denotes the p.d.f. of the latent vector y^*. In the remainder of this section methods are discussed for simulating expressions (4.2).

It should be noted that if we assume further that the distribution of y^* is a member of the linear exponential family, such as the normal distribution, the score of observation i can be written

$$(4.3) \qquad s_i(D_i,\theta) = \mathbf{E}_{y^*}(h(y^*;\theta,X_i)\,|\,y^* \epsilon D(X_i,\theta,d_i)) \,,$$

where $h(y^*;\cdot)$ is a vector of polynomials in y^*. See Ruud (1986) for an illustration of this result. Hence, this argument shows that for the successful implementation of MSS it is useful to devise ways to generate draws from the truncated distribution $z \equiv \{y^* | y^* \epsilon D\}$, since such draws $\tilde{z}^r$ could then be used to simulate the score unbiasedly by

$$(4.4) \qquad \tilde{s}_i(D_i,\theta,\eta) \equiv \frac{1}{R}\sum_{r=1}^{R} h(\tilde{z}^r) \,.$$

The simulation method that is perhaps the most intuitive to demonstrate is the following. Write the random vector y^* as

$$(4.5) \qquad y^* = \mu + \Gamma\eta \,,$$

where η is an independent standard normal vector of dimension M and $\Gamma(\sigma)$ is a lower triangular Choleski factor of $\Omega(\sigma)$, that is, $\Omega(\sigma) = \Gamma(\sigma)\Gamma(\sigma)'$. A simple approach to approximating (4.2.1) is to make repeated Monte Carlo draws for η, use (4.5) to calculate

y^* for each parameter vector, and then form empirical analogs of the expectations in (4.2.1) and (4.2.2). This approach defines the *crude frequency simulator* (CFS) of $\ell(D;\theta)$ and its derivatives. Similarly, a crude frequency simulator for $\nabla_\theta \ln \ell$ can be formed by rejecting draws of y^* that do not satisfy the conditioning event $y^* \in D$, and then calculating an empirical analog of the *conditional* expectation in (4.2.3) using the accepted draws.

The CFS are quick to compute and ideal for parallel processing. However, they are *not* continuous in parameters or in the underlying Monte Carlo draws, exhibiting jumps at different θ and η values. This complicates iterative parameter search. A second serious shortcoming of CFS is that they take the value 0 with a positive probability. This explains to a large extent the unsatisfactory performance of SML when CFS was used by Lerman and Manski (1981) to approximate ℓ_i's, since the SML criterion function is $\Sigma_i \ln \tilde{\ell}_i$.

Börsch-Supan and Hajivassiliou (1990) overcome these problems by developing the method of Smooth SML (SSML), which uses instead a simulator for ℓ_i that is (a) smooth in the unknown parameters and (b) is bounded away from 0. For the normal case $y^* \sim N(\mu,\Omega)$, a simulation method that possesses both these properties and also is extremely fast to compute is the method termed *GHK* by Hajivassiliou et al (1991).[13]

To discuss GHK and other simulation techniques, it is useful to introduce the following notation. For a vector of indices (1, ..., n), we use "<j" to denote the subvector (1, ..., j–1), and "–j" to denote the subvector that excludes component j. Thus, for a matrix Γ, $\Gamma_{j,<j}$ denotes a vector containing the first j–1 elements of row j, and $\Gamma_{-j,-j}$ denotes the subarray excluding row j and column j. For a vector η, $\eta_{<j}$ is the subvector of the first j–1 components, and η_{-j} is the subvector excluding component j.

Consider the triangularizing transformation $y^* = \mu + \Gamma\eta$, where Γ is the Choleski factor of Ω. The indicator $1(y^* \in D)$ then becomes $1(\mu + \Gamma\eta \in D)$, which can be written recursively as the product of indicators of the events

---

[13] This stands for Geweke, Hajivassiliou, and Keane. I will use the same acronyms as in Hajivassiliou et al (1991) to refer to the various simulators.

$$(4.6) \qquad D_j(\eta_{<j}) = \{\eta_j | (a_j - \mu_j - \Gamma_{j,<j}\eta_{<j})/\Gamma_{jj} < \eta_j < (b_j - \mu_j - \Gamma_{j,<j}\eta_{<j})/\Gamma_{jj}\}$$

for $j = 1, ..., M$. Define $\phi(\eta_j | D_j(\eta_{<j})) = \phi(\eta_j)1(\eta_j \in D_j(\eta_{<j}))/\Phi(D_j(\eta_{<j}))$, the conditional distribution of η_j given the event $D_j(\eta_{<j})$. A recursive scheme of this type was suggested by van Praag and Hop (1987) and by Geweke (1989b). Define a weight

$$(4.7) \qquad \omega(\eta) = \prod_{j=1}^{M} \Phi(D_j(\eta_{<j})).$$

Then it follows that

$$(4.8.1) \qquad \ell = \int f(\mu+\Gamma\eta)\omega(\eta) \prod_{j=1}^{M} \phi(\eta_j | D_j(\eta_{<j}))d\eta$$

and

$$(4.8.2) \qquad \nabla_\theta \ell = \int \nabla_\theta f(\mu+\Gamma\eta)\omega(\eta) \prod_{j=1}^{M} \phi(\eta_j | D_j(\eta_{<j}))d\eta \,.$$

An unbiased simulator of ℓ is an average of $f(\mu+\Gamma\eta)\omega(\eta)$ and an unbiased simulator of $\nabla_\theta \ell$ is an average of $\nabla_\theta f(\mu+\Gamma\eta)\omega(\eta)$, where $\omega(\eta)$ is the weighting function (4.7), over draws constructed recursively from the one-dimensional conditional densities $\phi(\eta_j | D_j(\eta_{<j}))$. Drawing η_j's from this density can be achieved by the following method, which is continuous in the unknown parameters and the Monte Carlo draws:

$$(4.9) \qquad \eta_j = \Phi^{-1}(\zeta_j \Phi((a_j - \mu_j - \Gamma_{j,<j}\eta_{<j})/\Gamma_{jj}) + (1 - \zeta_j)\Phi((b_j - \mu_j - \Gamma_{j,<j}\eta_{<j})/\Gamma_{jj})) \,.$$

The ζ_j are draws from the uniform [0,1] density.[14]

---

[14] This is a simple application of the probability integral transform result (see Feller (1971)). Let X be distributed according to the univariate uniform distribution on [0,1]. Then

$$Z \equiv G^{-1}(X) = \Phi^{-1}[\,(\Phi(b) - \Phi(a)) \cdot X + \Phi(a)\,]$$

is distributed N(0,1) s.t. $a \leq Z \leq b$. The proof follows by recognizing that the corresponding c.d.f. is

$$G(z) = \frac{\Phi(z) - \Phi(a)}{\Phi(b) - \Phi(a)},$$

where Φ denotes the univariate normal cumulative distribution function. Note that Z is a continuous and differentiable function of the parameters a and b.

It should be noted that the GHK simulator defined by (4.6–4.9) is applicable also to some non-normal distributions, as long as the univariate draws from the conditional one-dimensional distributions correspond to the multivariate distribution of ϵ in the latent variable specification $y^*=X\beta+\epsilon$. This useful property is not shared by many other simulators discussed here that rely on the very special additivity property of the normal distribution.

Another way to obtain a smooth simulator is to begin from CFS and introduce the *Kernel smoothed frequency simulator* (KFS), suggested by McFadden (1989). KFS replaces the indicator function $1(y^* \in D)$ in the crude frequency simulator with the function

$$(4.10) \qquad \mathcal{K}_{D,\omega}(y^*) \equiv \mathcal{K}((y^*-b)/\omega) - \mathcal{K}((y^*-a)/\omega) ,$$

where $\mathcal{K}(w)$ is a smooth kernel function from $\mathbb{R}^M$ onto [0,1] with $\mathcal{K}(-\infty) = 1$ and $\mathcal{K}(+\infty) = 0$, and ω is a window width parameter. This simulator is a biased estimate of ℓ and $\nabla_\theta \ell$ for positive ω, but in statistical applications one can shrink ω as sample size increases.

Another important simulation principle is *Acceptance/Rejection sampling* (ARS). Methods based on this principle provide a mechanism for drawing from a conditional density when practical exact transformations from uniform or standard normal variates are not available. The crudest form of ARS is to sample from the unconditional distribution of y^* using (4.5), reject points not in D until R accepted points are found, and form an empirical average of $h(y^*)$ over the accepted points. The following refinement (see Devroye (1986) or Rubinstein (1981)) permits improvement in the "yield" of the method by sampling from a comparison distribution that puts little or no weight outside D and has the property that the ratio of the comparison density to the target density is uniformly bounded above by a small number.[15]

Suppose $f(x)$ is an M-dimensional density, and one wishes to sample from the conditional density $f(\cdot \mid A)$ given the event $x \in A$. Suppose $g(x)$ is a density from which it

[15] The "yield" of an ARS procedure is defined to be the proportion of random variates that are accepted out of all those drawn.

Acceptance/Rejection

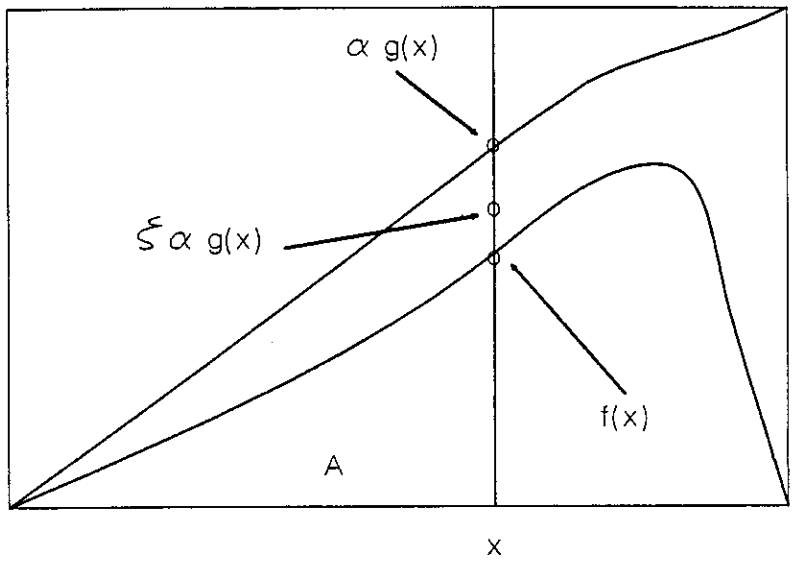

Figure 1

is practical to sample, with the property that $\sup_A f(x)/g(x) \leq \alpha < +\infty$. Assume that either the support of g is **A**, or that it is practical to test if $x \in \mathbf{A}$; that it is practical to calculate f(x) and g(x); and that it is practical to calculate a bound α. Draw x from g and ζ from a uniform density on [0,1], repeat this process until a pair satisfying $x \in \mathbf{A}$ and $f(x) \geq \zeta\alpha g(x)$ is observed, and accept the associated x. Then, it follows, through a simple application of Bayes' theorem, that the accepted points have density $f(x|\mathbf{A}) \equiv f(x)/f(\mathbf{A})$. The expected yield of this method is $1/\alpha$. In the specific example in Figure 1, the particular x drawn would be rejected. Possible comparison distributions for ARS are independent exponential or truncated normal distributions in y* space, but greater yields can be obtained using the recursive truncated normal distribution (4.6) employed in the GHK simulator. Lemmas in McFadden (1989) and Hajivassiliou and McFadden (1990) give protocols for use of independent exponential or recursive truncated normal comparison distributions. To prevent lengthy computations, ARS may be modified to incorporate a censoring rule, such as if the first r trials do not yield an acceptance, then the last draw from g(x) is accepted unconditionally. This method will be biased, but Hajivassiliou et al

(1991) show that the bias is bounded at a geometric rate in r.

Note that ARS is not a continuous function of the unknown parameters. To see this, note that a point $y^* = \mu + \Gamma\eta$ may, for given η, move from the rejection to the acceptance region with small changes in the parameters. Hajivassiliou and McFadden (1990) show that the ARS nevertheless satisfies a stochastic equicontinuity condition that enables its use in simulation estimation applications.

An important case where bias may be a major issue is simulation of $1/\ell$ in the score expressions $\nabla_\theta \ell/\ell$. One technique for achieving an unbiased simulator is based on the observation made by Ruud that $1/\ell$ is the expectation of the number of independent draws R from (4.5) required to yield $y^* \in D$; this can be simulated by drawing sequentially from (4.5) until $y^* \in D$ is observed. See McFadden and Ruud (1991) for details. This method is termed the *sequentially unbiased simulator* (SUS).

Another general principle for devising simulators is *importance sampling.* See Hammersley and Handscomb (1964). Suppose the integrands in ℓ and $\nabla_\theta \ell$ can, with multiplication and division by a factor if necessary, be written as the product of a density $\gamma(y^*)$, whose support coincides with or contains D, from which it is easy to sample, and the remainder. Then, ℓ and $\nabla_\theta \ell$ can be written as

$$(4.11.1) \qquad \ell = \int \{1(v \in D) f(v-\mu,\Omega)/\gamma(v)\}\gamma(v)dv \equiv \mathbf{E}_\gamma 1(v \in D) f(v-\mu,\Omega)/\gamma(v) ,$$

and

$$(4.11.2) \qquad \nabla_\theta \ell = \int \{1(v \in D)\nabla_\theta f(v-\mu,\Omega)/\gamma(v)\}\gamma(v)dv \equiv \mathbf{E}_\gamma 1(v \in D)\nabla_\theta f(v-\mu,\Omega)/\gamma(v) ,$$

where $v \equiv y^*$, and $f(\cdot)$ denotes the p.d.f. of y^*. An empirical expectation using draws from $\gamma(v)$ gives an unbiased simulator that is smooth in parameters. For the $y^* \sim \mathcal{N}(\mu,\Omega)$ case, this is termed *normal importance sampling* (NIS).

For fast computation, choose γ so that the components are independent, or are obtained as simple transformations of independent variates. Two possible choices of γ are considered in Hajivassiliou et al (1991). First, the NISE method uses independent

exponential densities,

$$(4.12) \qquad \gamma(v) = \prod_{i=1}^{M} \exp((v_i - b_i)/c_i)/c_i,$$

where c_i are parameters that can be set as part of the simulation. Draws from this density are easily computed using $v_i = b_i + c_i \cdot \log \zeta_i$, where ζ_i is a uniform [0,1] variate. An alternative that is more likely to concentrate probability for γ in the same region as the multivariate normal is the product of truncated normals,

$$(4.13) \qquad \gamma(v) = \prod_{i=1}^{M} \phi((v_i - \alpha_i)/c_i)/[\Phi((b_i - \alpha_i)/c_i) - \Phi((a_i - \alpha_i)/c_i)], \quad v \in D,$$

with $\alpha_i = \mu_i$ and $c_i = \sqrt{\Omega_{ii}}$. This defines the NIST simulator. One can sample from this distribution using

$$(4.14) \qquad v_i = \alpha_i + c_i \Phi^{-1}(\zeta_i \Phi((b_i - \alpha_i)/c_i) + (1 - \zeta_i)\Phi((a_i - \alpha_i)/c_i)),$$

with ζ_i a uniform [0,1] variate.[16]

A promising method, particularly well suited to be used in conjunction with the MSS estimator, is the *Gibbs sampler simulator* for $\nabla_\theta \ell$. This simulator is based on a Markov chain that utilizes computable univariate truncated normal densities to construct transitions, and has the desired truncated multivariate normal as its limiting distribution.[17] The simulator was developed by Hajivassiliou, starting from stochastic relaxation methods studied by Geman and Geman (1984). This simulator is defined by the following Markovian updating scheme.[18] Suppose D is finite. Start from any $v^{(0)} \in D$. Define a recursive procedure with steps $i = 1, ..., M$ in rounds $j = 1, ..., r$. Suppose at step i in round j, $v^{(j-1)}$ and $v_{<i}^{(j)}$ have been determined. Define

---

[16] A third choice for $\gamma(\cdot)$ that in practice seems to be less successful than (4.13), is due to Moran (1984). West (1990) develops and investigates an adaptive Monte Carlo method, based on importance sampling and density estimation techniques using kernels. He finds that these methods possess the potential to develop automatic routines for Bayesian estimation methodology.

[17] GSS can in principle be generalized to non-normal distributions, provided the corresponding univariate distributions are easy to draw from.

[18] This description follows Hajivassiliou and McFadden (1990).

$$(4.15) \qquad v_i^{(j)} = \kappa_{ij} + \sigma_i \Phi^{-1}(\zeta_{ij}\Phi((b_i - \kappa_{ij})/\sigma_i - (1 - \zeta_{ij})\Phi((a_i - \kappa_{ij})/\sigma_i) ,$$

where the ζ_{ij} are independent uniform [0,1] variates,

$$(4.16) \qquad \kappa_{ij} = \mu_i + \Omega_{i,-i}\Omega_{-i,-i}^{-1}\begin{bmatrix} v_{<i}^{(j)} - \mu_{<i}^{(j)} \\ v_{>i}^{(j-1)} - \mu_{>i}^{(j-1)} \end{bmatrix},$$

and

$$(4.17) \qquad \sigma_i = \left[\Omega_{ii} - \Omega_{i,-i}\Omega_{-i,-i}^{-1}\Omega_{-i,i}\right]^{1/2}.$$

Note that $v \in D$ by construction. Repeat this process r "Gibbs resampling rounds." Then, as Hajivassiliou and McFadden (1990) prove, the random draws obtained by this simulator have a distribution which converges at a geometric rate to the true distribution $z \equiv \{y^* | y^* \epsilon D\}$ as the number of Gibbs resampling rounds r grows to infinity. They also show that an MSS estimator based on GSS with a finite number of terminal draws R will be consistent and asymptotically normal provided r grows at the rate logN, where N is the sample size. This is a very satisfactory rate. In contrast, consistency and asymptotic normality of a MSS estimator based on AUS or SUS require that R grows at rate $\sqrt{N}$.

Geweke (1991) implements the Gibbs sampler for efficient simulation from the truncated multivariate normal and Student–t distributions, by developing a speedy algorithm for generating univariate truncated normal variates, compared to the algorithm I will denote by RNDTRN that is given in footnote 14. Geweke denotes his algorithm by GGRNRM. I have performed simple Monte Carlo experiments which gave the following comparison. To draw 20,000 $x \sim N(0,1)$ truncated on $-1<x<1$, on a SUN4/80 workstation took 13.23 seconds using GGRNRM and 15.23 using RNDTRN. On the other hand, when the constraint was one-sided and far from the mean ($-10<x<5$), then the respective times were almost equal (22.2 vs. 20.0). This follows from the fact that Geweke's GGRNRM algorithm is a mixture of RNDTRN and a speedier method, which is only used in regions close to the unconditional mean.

Finally, mention must be made of three simulators for ℓ, $\nabla_\theta \ell$, and $\nabla_\theta \ln \ell$ that were developed explicitly for the $y^* \sim \mathcal{N}(\mu,\Omega)$ case, and that may be difficult or impossible to generalize to non-normal y^*. First, consider the *Stern decomposition simulator* (SDS), suggested by Stern (1990). This method writes $y^* \sim \mathcal{N}(\cdot;\mu,\Omega)$ as a sum $y^* = Y+W$, with $Y \sim \mathcal{N}(\cdot;\, \mu,\, \lambda^2 I)$ and $W \sim \mathcal{N}(\cdot;\, 0,\, \Omega-\lambda^2 I)$. That is, y^* equals the sum of a "small" independently distributed normal vector and a second normal vector that carries the information on the covariance matrix of y^*. Then, by the law of iterated expectations,

$$\ell = \int_w \left\{ \int_{y=a}^{b} \left[\prod_{i=1}^{M} \phi((y_i - w_i - \mu_i)/\lambda)/\lambda \right] dy \right\} \nabla_\theta n(w,\, \Omega-\lambda^2 I) dw, \tag{4.18}$$

where $n(\cdot)$ denotes the normal p.d.f. The term in braces can be integrated analytically; then the SDS averages this interior integral over r Monte Carlo draws $w = [\Omega-\lambda^2 I]^{1/2}\eta$, where η is a standard normal vector and $[\Omega-\lambda^2 I]^{1/2}$ is a Choleski factor of $\Omega-\lambda^2 I$. The SDS provides an unbiased smooth simulator. This simulator is fast to compute, but it can be computationally burdensome to determine λ such that $\Omega - \lambda^2 I$ is positive definite, and accuracy fails when M is large and the eigenvalues of Ω are uneven.

Another simulator, developed explicitly for the $y^* \sim \mathcal{N}(\mu,\Omega)$ case, can be interpreted as importance sampling with the uniform distribution on the unit sphere intersecting the negative orthant as the comparison density. This simulator was suggested by McFadden (1989). A related method is *Deák's chi–square simulator* (DCS). See Deák (1980a, 1980b). The DCS simulator is obtained by drawing an antithetic random grid of points on the unit sphere and then using a spherical transformation about the mean of the multivariate normal distribution.[19] It is unbiased for ℓ and $\nabla_\theta \ell$ and smooth in parameters.

---

[19] The principle of *antithetic variates* is to introduce negative correlation between successive Monte Carlo draws in order to reduce the variance of simulation sample averages. See Hendry (1984) for a discussion. Hajivassiliou et al (1991) generalize the principle to the multivariate case by selecting a regular grid of points whose location is random. This technique can be employed in simulators of ℓ, $\nabla \ell$, and $\nabla \ln \ell$ when the method has sufficient symmetry. Another way to reduce simulation variance in Monte Carlo methods is through the use of *control variates*. These are random variables with analytic expectations that are positively correlated with the random variable whose expectation is to be simulated. Then, simulation variance can be reduced by simulating only the difference between the expectations of the variate of interest and the control variate. See Hendry (1984) for details.

Note that all the simulators for ℓ, $\nabla_\theta \ell$, and $\nabla_\theta \ln \ell$ discussed in this section have been implemented in GAUSS and in FORTRAN computer code in Hajivassiliou et al (1991). The code can be requested from these authors. The main conclusion of this study is that the GHK simulator is the best of all the 13 simulation algorithms considered. It consistently has the best (or among the best) mean and median bias, standard deviation, and RMSE characteristics. For details, readers can refer to Hajivassiliou et al (1991). The main conclusion confirms other findings in the literature by Bolduc and Kaci (1989) and Mühleisen (1991), who considered GHK against a handful of alternative simulators. This result is found to hold in Hajivassiliou et al (1991) against *all* other 12 simulators presented here.

5. Conclusion

This Chapter discussed estimation methods for Limited Dependent Variable (LDV) models that employ Monte Carlo simulation techniques to overcome numerical intractabilities of such models. These difficulties arise because high dimensional integral expressions need to be calculated repeatedly. In the past, investigators were forced to restrict attention to special classes of LDV models that are computationally tractable. The simulation estimation methods offer dramatic computational advantages over classical methods. In a typical example, computations that would require 4 months to estimate on a modern super-computer with classical methods can be carried out overnight on a high-end Personal Computer. Hence, simulation estimation methods now make it possible to estimate LDV models that are computationally intractable using classical estimation methods even on state-of-the-art supercomputers.

The main simulation estimation methods for LDV models developed in the econometrics literature, namely SML, MSM, MSS, SPML, and SSML, have been discussed, and results have been presented about the asymptotic properties of such simulation-based estimators. Specific simulation algorithms to use in conjunction with these five estimation

methods were also described. The leading simulation estimation methods require the simulation of one or more of the following expressions: probabilities of the limited dependent variables, derivatives of such probabilities with respect to underlying parameters, and derivatives of the (natural) logarithm of the probabilities of the dependent variables.

There are two main areas in which future research in this field should prove particularly fruitful. The first is to design simulation algorithms that work satisfactorily for the case of non-normal distributions. The second is to introduce simulation methods into semi-parametric estimation approaches. This would be especially important in freeing the leading simulation estimation methods for LDV models discussed in this Chapter from their very restrictive fully parametric framework.

REFERENCES

Amemiya, T. (1984). Tobit models: A survey. J. Econometrics 24, 3–61.

Avery, R., L. Hansen and V. Hotz (1983). Multiperiod probit models and orthogonality condition estimation. Internat. Econom. Rev. 24, 21–35.

Bauwens, L. (1984). Bayesian Full Information Analysis of Simultaneous Equation Models using Integration by Monte Carlo. Springer-Verlag, Berlin.

Berkovec, J. and S. Stern (1991). Job exit behavior of older men. Econometrica 59, 189–210.

Berry, S. and A. Pakes (1990). The performance of alternative simulation estimators. Mimeo, Yale University.

Bloemen, H. and A. Kapteyn (1991). The joint estimation of a non-linear labour supply function and a wage equation using simulated response probabilities. Mimeo, Tilburg University.

Bolduc, D. and M. Kaci (1989). On the estimation of probabilities in multinomial probit models: An empirical comparative study of existing techniques. Mimeo, Université Laval.

Bolduc, D. and M. Kaci (1991). Multinomial probit models with factor-based autoregressive errors: A computationally efficient estimation approach. Mimeo, Université Laval.

Börsch-Supan, A. and V. Hajivassiliou (1990). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. Cowles Foundation Discussion Paper No. 960, Yale University.

Börsch-Supan, A., V. Hajivassiliou, L. Kotlikoff, and J. Morris (1991). Health, children, and elderly living arrangements: A multi-period multinomial probit model with unobserved heterogeneity and autocorrelated errors. In Topics in the Economics of Aging (D. Wise, editor), pp. 79–107. University of Chicago Press, Chicago.

Chamberlain, G. (1984). Panel data. In Handbook of Econometrics, Vol. 2 (Griliches, Z. and M. Intriligator, editors), pp. 1247–1318. North-Holland, Amsterdam.

Clark, C. (1961). The greatest of a finite set of random variables. Oper. Res. 9, 145–162.

Daganzo, C., F. Bouthelier, and Y. Sheffi (1977). Multinomial probit and qualitative choice: A computationally efficient algorithm. Transportation Sci. 11, 338–358.

Deák, I. (1980a). Three digit accurate multiple normal probabilities. Numer. Math. 35, 369–380.

Deák, I. (1980b). Fast procedures for generating stationary normal vectors. J. Statist. Comput. Simulation 10, 225–242.

Devroye, L. (1986). Non–Uniform Random Variate Generation. Springer-Verlag, New York.

Divgi, D. (1979). Calculation of univariate and bivariate normal probability functions. Ann. Statist., 903–910.

Dubin, J. and D. McFadden (1984). An econometric analysis of residential electric appliance holdings and consumption. Econometrica 52, 345–362.

Duffie, D. and K. Singleton (1989). Simulated moments estimation of Markov models of asset prices. Mimeo, Stanford University.

Dutt, J. (1973). A representation of multivariate normal probability integrals by integral transforms. Biometrika 60, 637–645.

Dutt, J. (1976). Numerical aspects of multivariate normal probabilities in econometric models. Annals of Economic and Social Measurement 5, 547–562.

Eggink, E., B. van Praag, and J. Hop (1992). A symmetric approach to the labor market with the household as unit of observation. Erasmus University, Working Paper, presented at the Rotterdam International Conference on Simulation and at the ESEM Brussels.

Feller, W. (1971). An Introduction to Probability Theory and its Applications. Wiley, New York.

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6, 721–741.

Geweke, J. (1989a). Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339.

Geweke, J. (1989b). Efficient simulation from the multivariate normal distribution subject to linear inequality constraints and the evaluation of constraint probabilities. Mimeo, Duke University.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. Forthcoming in Computing Science and Statistics: Proceedings of the Twenty–Third Symposium.

Gourieroux, C. and A. Monfort (1990). Simulation based inference in models with heterogeneity. Mimeo, INSEE.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1984a). Pseudo maximum likelihood methods: Theory. Econometrica 52, 681–700.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1984b). Pseudo maximum likelihood methods: Applications to Poisson models. Econometrica 52, 701–720.

Hajivassiliou, V. (1986). Serial correlation in limited dependent variable models: Theoretical and Monte Carlo results. Cowles Foundation Discussion Paper No. 803, Yale University.

Hajivassiliou, V. (1990). The method of simulated scores: A presentation and comparative evaluation. Mimeo, Cowles Foundation for Research in Economics, Yale University.

Hajivassiliou, V. and Y. Ioannides (1991). Switching regressions models of the Euler equation: Consumption, labor supply, and liquidity constraints. Mimeo, Cowles Foundation for Research in Economics, Yale University.

Hajivassiliou, V. and D. McFadden (1990). The method of simulated scores, with application to models of external debt crises. Cowles Foundation Discussion Paper No. 967, Yale University.

Hajivassiliou, V., D. McFadden, and P. Ruud (1991). Simulation of multivariate normal orthant probabilities: Methods and programs. Mimeo, Cowles Foundation for Research in Economics, Yale University.

Hammersley, J. and D. Handscomb (1964). Monte Carlo Methods. Methuen, London.

Hausman, J. and D. Wise (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica 46, 403–426.

Heckman, J. (1981). Dynamic discrete models. In Structural Analysis of Discrete Data with Econometric Applications (Manski, C. and D. McFadden, editors), pp. 179–195. MIT Press, Cambridge.

Hendry, D. (1984). Monte Carlo experimentation in econometrics. In Handbook of Econometrics, Vol. 2 (Griliches, Z. and M. Intriligator, editors), pp. 937–976. North Holland, Amsterdam.

Horowitz, J., J. Sparmonn, and C. Daganzo (1981). An investigation of the accuracy of the Clark approximation for the multinomial probit model. Transportation Sci. 16, 382–401.

Hotz, V.J. and R. Miller (1989). Conditional choice probabilities and the estimation of dynamic programming models. GSIA Working Paper 88–89–10.

Hotz, V.J., R. Miller, S. Sanders, and J. Smith (1991). A simulation estimator for dynamic discrete choice models. Mimeo, NORC, University of Chicago.

Hotz, V.J. and S. Sanders (1990). The estimation of dynamic discrete choice models by the method of simulated moments. NORC, University of Chicago.

Keane, M. (1990). A computationally efficient practical simulation estimator for panel data, with applications to estimating temporal dependence in employment and wages. Mimeo, University of Minnesota.

Kloek, T. and H. van Dijk (1978). Bayesian estimates of equation system parameters: An application of integration by Monte Carlo. Econometrica 46, 1–20.

Laroque, G. and B. Salanié (1989). Estimation of multi-market disequilibrium fix-price models: An application of pseudo maximum likelihood methods. Econometrica 57, 831–860.

Laroque, G. and B. Salanié (1990). The properties of simulated pseudo-maximum likelihood methods: The case of the canonical disequilibrium model. Working Paper No. 9005, CREST–Departement de la Recherche, INSEE.

Lee, B.–S. and B. Ingram (1991). Simulation estimation of time–series models. J. Econometrics 47, 197–205.

Lee, L.-F. (1990). On the efficiency of methods of simulated moments and maximum simulated likelihood estimation of discrete response models. Center for Economic Research, Discussion Paper No. 260, University of Minnesota.

Lerman, S. and C. Manski (1981). On the use of simulated frequencies to approximate choice probabilities. In Structural Analysis of Discrete Data with Econometric Applications (C. Manski and D. McFadden, editors), pp. 305–319. MIT Press, Cambridge.

Maddala, G.S. (1983). Limited Dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In Frontiers in Econometrics (P. Zarembka, editor), pp. 105–142. Academic Press, New York.

McFadden, D. (1981). Econometric models of probabilistic choice. In Structural Analysis of Discrete Data with Econometric Applications (C. Manski and D. McFadden, editors), pp. 198–272. MIT Press, Cambridge.

McFadden, D. (1984). Econometric analysis of qualitative response models. In Handbook of Econometrics, Vol. 2 (Griliches, Z. and M. Intriligator, editors), pp. 1395–1457. North Holland, Amsterdam.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57, 995–1026.

McFadden, D. and P. Ruud (1991). Estimation by simulation. Mimeo, M.I.T.

Moran, P. (1984). The Monte Carlo evaluation of orthant probabilities for multivariate normal distributions. Austral. J. Statist. 26, 39–44.

Mühleisen, M. (1991). On the use of simulated estimators for panel models with limited-dependent variables. Mimeo, University of Munich.

Nelder, J. and R. Meade (1964). A simplex method for function minimization. Comput. J. 7, 308–313.

Pakes, A. (1991). Estimation of dynamic structural models: Problems and prospects Part II: Mixed Continuous-Discrete Controls and Market Interactions. Mimeo, Yale University.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. Econometrica 57, 1027–1057.

Poirier, D. and P. Ruud (1988). Probit with dependent observations. Rev. Econom. Stud. 55, 593–614.

Quandt, R. (1983). Computational problems in econometrics. In Handbook of Econometrics, Vol. 1 (Griliches, Z. and M. Intriligator, editors), pp. 1395–1457. North Holland, Amsterdam.

Rubinstein, R. (1981). Simulation and the Monte Carlo Method. Wiley, New York.

Rust, J. (1989). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. Econometrica 55, 999–1033.

Rust, J. (1991). Estimation of dynamic structural models: Problems and prospects Part II: discrete decision processes. SSRI Working Paper #9106, University of Wisconsin at Madison.

Ruud, P. (1986). On the method of simulated moments for the estimation of limited dependent variable models. Mimeo, University of California at Berkeley.

Ruud, P. (1991). Extensions of estimation methods using the EM algorithm. J. Econometrics 49, 305–341.

Steck, G.P. (1958). A table for computing trivariate normal probabilities. Annals of Mathematical Statistics 29, 780–800.

Stern, S. (1991). Discrete choice models with unobserved heterogeneity: Demand for transportation by elderly and disabled people. University of Virginia, Working Paper.

Stern, S. (1992). A method for smoothing simulated moments of discrete probabilities in multinomial probit models. Econometrica 60, 943–952.

van Dijk, H.K. (1987). Some advances in Bayesian estimation methods using Monte Carlo integration. In Advances in Econometrics, Vol. 6 (Fomby, T.B. and G.F. Rhodes, editors), pp. 100–120. JAI Press: Connecticut.

van Praag, B., and J. Hop (1987). Estimation of continuous models on the basis of set-valued observations. Erasmus University, Working Paper, presented at the ESEM Copenhagen.

van Praag, B., J. Hop, and E. Eggink (1991). A symmetric approach to the labor market by means of the simulated EM–algorithm with an application to married females. Erasmus University, Working Paper, presented at the ESEM Cambridge.

West, M. (1990). Bayesian computations: Monte-Carlo density estimation. Discussion Paper #90–A10, Duke University.