Econometrica, Vol. 66, No. 4 (July, 1998), 863–896

THE METHOD OF SIMULATED SCORES FOR THE ESTIMATION OF LDV MODELS

BY VASSILIS A. HAJIVASSILIOU AND DANIEL L. MCFADDEN[1]

The method of simulated scores (MSS) is presented for estimating limited dependent variables models (LDV) with flexible correlation structure in the unobservables. We propose simulators that are continuous in the unknown parameter vectors, and hence standard optimization methods can be used to compute the MSS estimators that employ these simulators. The first continuous method relies on a recursive conditioning of the multivariate normal density through a Cholesky triangularization of its variance-covariance matrix. The second method combines results about the conditionals of the multivariate normal distribution with Gibbs resampling techniques. We establish consistency and asymptotic normality of the MSS estimators and derive suitable rates at which the number of simulations must rise if biased simulators are used.

KEYWORDS: Limited dependent variable models, simulation estimation, Gibbs resampling.

1. INTRODUCTION

IT HAS LONG BEEN KNOWN that classical estimation of limited dependent variable (LDV) models with flexible correlation structure in the unobservables poses formidable computational problems because of a concomitant need for high dimensional numerical integration. Examples of such models are multiperiod (panel) probit and Tobit models, as well as multinomial discrete choice models with varying substitutability between available alternatives. Recently investigators have shown that simulation estimation methods that approximate generalized moment conditions by unbiased simulators provide consistent and asymptotically normal parameter estimates for a *finite* number of simulations (McFadden (1989), Pakes and Pollard (1989)).[2] A common property of the early formulations of methods of simulating moment conditions (MSM) is that they yield criterion functions that are discontinuous in the unknown parameter vectors to be estimated. As a result, establishing their asymptotic properties requires the theory of empirical processes. Moreover, their implementation poses difficult computational problems, because standard methods for numerical

[1] We are indebted to three anonymous referees, Don Andrews, Jacques Drèze, John Geweke, John Ham, Rosa Matzkin, Michel Mouchart, Whitney Newey, Peter Robinson, Paul Ruud, and Herman van Dijk for very helpful comments and suggestions. Any remaining errors and shortcomings are, of course, our own responsibility. Yoosoon Chang, Dan Nelson, and Robert Stoddard provided expert research assistance. Hajivassiliou gratefully acknowledges financial support from the National Science Foundation through Grant No. SES-88-13630.

[2] These methods are in contrast to simulation estimation methods that simulate nonlinear expressions in criterion functions and hence require an unbounded number of simulations to achieve consistency and asymptotic normality. See, inter alia, Lerman and Manski (1981), van Praag and Hop (1987), Laroque and Salanié (1989). For an extensive review of the literature on simulation-based classical estimation, see Hajivassiliou and Ruud (1994).

optimization assume continuity of the optimand (and several also require twice continuous differentiability of the criterion function).

In this paper, we exposit and operationalize a *method of simulated scores* (MSS), which simulates directly the logarithmic derivatives corresponding to maximum likelihood estimation (MLE), and establish its asymptotic properties. We show that the MSS provides a unified treatment of a broader class of estimation problems compared to other simulation estimation methods. We discuss three simulators to be used in conjunction with MSS estimation. The first, which is a discontinuous function of the unknown parameters, generalizes acceptance-rejection methods and provides unbiased simulation of the scores. We prove that the MSS estimator using the generalized acceptance-rejection simulator is consistent uniformly asymptotically normal (CUAN) for a finite number of simulations. We show that computational problems may be eased if smooth but biased simulators are used for the MSS estimator. We develop two such simulation methods. The first employs a recursive conditioning of the normal multivariate density through a Cholesky triangularization of its variance-covariance matrix; it thus provides unbiased simulation for likelihood contributions and asymptotically unbiased simulation of the scores, and is continuous in the unknown parameters. We establish that when this method is used to simulate the scores, the resulting MSS estimator is CUAN provided the number of simulations grows faster than the square root of the number of i.i.d. observations, N. The second smooth and asymptotically unbiased simulator relies on results about the conditionals of a multivariate normal distribution and employs Gibbs resampling (Geman and Geman (1984)).[3] It then follows that for an MSS estimator based on this simulator to be CUAN, the number of resamplings used for each simulation must grow with the sample size at the (much slower) rate log N, while the number of simulations may be finite. Hence this paper shows for the first time that Gibbs resampling techniques can be useful for classical inference, while to our knowledge past applications of these techniques have been only for Bayesian estimation problems.[4]

It should be noted that several investigators in the past have proposed consistent simulation of the score as a method of estimation. See, inter alia, van Praag and Hop (1987). The MSS estimators we discuss here, when used in conjunction with the three simulators developed in this paper, have several advantages over such existing simulation estimation methods. A particular development below follows on a suggestion by Ruud (1986) that the score for the general linear exponential model can be written as conditional expectations, which might be simulated directly. This provides the first major advantage of MSS in that it is applicable to any LDV model that can be written as a set of

[3] For a good introduction to Markov-Chain Monte-Carlo methods (MCMC), of which Gibbs is a leading method, see Chib (1995). Hajivassiliou et al. (1996) discuss alternative simulators and compare their properties to the ones given here.

[4] For example, see McCulloch and Rossi (1994) for a Bayesian estimator for the multinomial probit model based on Gibbs resampling.

linear inequality constraints on the underlying latent variables, the distribution of which belongs to the linear exponential class. Hence, the method does not require the development of ad hoc simulation techniques for each type of LDV model that is under consideration. Second, since MSS simulates directly the scores, it corresponds to MSM where the *optimal* (for asymptotic efficiency) *instruments* are used. Hence, the efficiency of the MSS estimator among the class of simulation estimation methods is guaranteed. Third, when simulating functions that are continuous in the parameters are employed, certain computational complexities of MSM are avoided.

In Section 2 we explain the intractability of MLE for LDV models with correlated unobservables. We then introduce in Section 3 the simulated scores estimation method that is applicable to LDV models with flexible correlation structures in the unobservables. Such LDV models include the probit and Tobit models with panel data time-dependence, as well as multinomial choice models without restrictive assumptions on the substitutability of different alternatives such as the independence of irrelevant alternatives assumption (McFadden (1973)).

In Section 4 we establish the CUAN properties of the MSS estimator under general conditions. In Section 5 we develop a method of generating unbiased simulations of LDV probabilities and two methods for generating draws from conditional normal distributions. The first two methods make the MSS estimator continuous in the unknown parameter vectors, and are respectively based on a recursive conditioning of the normal multivariate density through a Cholesky triangularization of its variance-covariance matrix and on Gibbs resampling. The third simulator is based on refinements of acceptance-rejection methods. Software implementing the simulators and simulation-based estimators discussed in this paper is publicly available over the Internet through the Web page http://econ.lse.ac.uk/~vassilis.

Section 6 investigates the CUAN properties of MSS estimators based on each of the three simulators we introduce. We show in that section that MSS using the acceptance-rejection simulator is CUAN for a finite number of simulations; MSS using the continuous recursive conditioning simulator is CUAN provided the number of simulations rises faster than the square root of the number of observations; and finally, we obtain the result that the MSS estimator based on Gibbs resampling with R (finite) simulations is CUAN as long as the number of resamplings used for each simulation grows at a rate faster than $\log N$.

TABLE I

MSS ESTIMATORS

Estimator	Underlying Simulator	Description	Asymptotic Properties
General MSS	—	Section 3	Section 4
MSS-SRC	SRC: Smooth Recursive Conditioning	Section 5.1	Section 6
MSS-GRS	GRS: Gibbs Resampling	Section 5.2	Section 6
MSS-SAR	SAR: Sophisticated Acceptance-Rejection	Section 5.3	Section 6

Section 7 gives a simple illustrative example to highlight the relations between the MSS estimation method and other simulation methods in the literature. Section 8 concludes with a summary.

2. THE CANONICAL LDV MODEL

Consider a sample of N economic agents, assumed to be random. A data array (y_i, X_i) is observed, where X_i is an $m_i \times K$ array of exogenous variables, and y_i is an $m_i \times 1$ vector of limited dependent variables. We assume y_i is an indirect observation on a latent vector y_i^* according to a many-to-one mapping $y_i = \tau(y_i^*)$, with y_i^* given by a linear model

$$y_i^* = X_i \beta + \epsilon_i. \tag{1}$$

We assume the disturbance vector ϵ_i is multivariate normal, independent of X_i, with the structure

$$\epsilon_i = \Gamma_i \eta_i, \tag{2}$$

where Γ_i is a $m_i \times S_i$ parametric array of rank m_i, and η_i is a $S_i \times 1$ vector of independent standard normal variates. Let $\Omega_i = \Gamma_i \Gamma_i'$. Note that this notation encompasses also panel data models, since in that case y_i^* would be the stacked vector of the values of the latent variables for agent i in each period.

Define

$$D(y_i) = \{y_i^* | y_i = \tau(y_i^*)\}. \tag{3}$$

Then the likelihood of the observation is

$$\ell_i(\theta; y_i) = \int_{D(y_i)} n(y_i^* - X_i \beta, \Omega_i)\, dy_i^*, \tag{4}$$

with derivative vector $\ell_{\theta i}(\cdot)$, where β, Ω_i are functions of a $k \times 1$ deep parameter vector θ, and

$$n(\epsilon, \Omega) = (2\pi)^{-T/2} |\Omega|^{-1/2} \exp\left[-\frac{1}{2}\epsilon' \Omega^{-1} \epsilon\right] \tag{5}$$

is the multivariate normal density. The asymptotically optimal parametric maximum likelihood estimator (MLE) is defined as the argument that solves the score equations, i.e.,

$$\hat{\theta}_{MLE} \underset{\theta}{\text{solves}} \left\{ \frac{1}{N} \sum_i s_i(\theta; y_i) \equiv \frac{1}{N} \sum_i \ell_{\theta i}(\theta; y_i) / \ell_i(\theta; y_i) = 0 \right\}. \tag{6}$$

The classic cross-sectional LDV model belonging to the class we describe here is the multinomial discrete choice.

MODEL 1—Multinomial Probit: *Consider an independent sample of N individuals, with typical individual i choosing among m_i alternatives with observed attributes x_{ij}. Alternative j yields the (random) utility*

$$y_{ij}^* = x_{ij}\beta + \epsilon_{ij} \qquad (j = 1, \ldots, J)$$

and individual i chooses alternative k that satisfies

$$\underline{D(y_i)}: \quad -\infty < y_{ik}^* < \infty, \quad 0 < y_{ik}^* - y_{ij}^* < \infty, \quad \forall j \neq k. \tag{7}$$

The analyst observes the indicator $y_i \equiv \arg\max_j\{y_{i1}^*, \ldots, y_{ij}^*, \ldots, y_{iJ}^*\}$, say k. It follows that in this case, the linear restrictions on the elements of y_i^* correspond to the matrix

$$A_k = \{-I_J \text{ with column } k \text{ replaced by a vector of } 1\text{'s}\}.$$

A related model is the following.

MODEL 2—Multivariate Rank Ordered Probit: *As with model 1, individual i chooses alternative k that offers the highest utility y_{ik}^*. The analyst, however, observes the full ranking of the J alternatives in terms of the utility they yield, i.e., the analyst observes the J-dimensional vector of indices*

$$y_i \equiv (k_1, \ldots, k_J)'$$

such that

$$\underline{\mathbf{D}(\mathbf{y_i})}: \quad y_{ik_1}^* \leq y_{ik_2}^* \leq \cdots \leq y_{ik_J}^*. \tag{8}$$

For panel (longitudinal) data sets, the following models are useful:

MODEL 3—Multiperiod Probit:

$$\begin{array}{cc} \underline{\mathbf{D}(\mathbf{y_i}):} & \underline{\mathbf{y_i} = \tau(\mathbf{y_i}^*)} \\ 0 \leq y_{it}^* < \infty & y_{it} = 1 \\ -\infty < y_{it}^* \leq 0 & y_{it} = 0. \end{array} \tag{9}$$

MODEL 4—Multiperiod Tobit:

$$\begin{array}{cc} \underline{\mathbf{D}(\mathbf{y_i}):} & \underline{\mathbf{y_i} = \tau(\mathbf{y_i}^*)} \\ 0 < y_{it}^* & y_{it} = y_{it}^* > 0 \\ -\infty < y_{it}^* \leq 0 & y_{it} = 0. \end{array} \tag{10}$$

In view of (4) and (5), classical estimation by the method of maximum likelihood of either the panel binomial discrete response model (9), the panel Tobit model (10), or the MNP models (7,8) is computationally intractable when m_i, the number of time periods per individual, or the number of alternatives in

the choice set, exceeds 3 or 4, the variance-covariance matrix Ω_i of the error vector ϵ_i is left unrestricted, and conventional numerical integration (e.g., multivariate quadrature) is used. A traditional approach in obtaining ML estimates is to restrict heavily the structure of Ω_i in such a way as to make the evaluation of (4) and its derivatives computationally feasible. One extreme is to assume that the errors are independent across individuals and across time periods for a given individual, i.e.,

$$\Omega_i^{IID} = E\epsilon_i\epsilon_i' = \gamma_1^2 I_{m_i}, \tag{11}$$

where γ_1^2 is a variance parameter to be estimated.

Despite its computational simplicity, such an assumption is often very inappropriate for realistic applications. In panel data models, for example, temporal dependence can arise because of unobserved heterogeneity that persists over time. Conversely, in the cross-sectional multinomial discrete choice case, assuming the (11) covariance structure leads to potentially problematic implications of the so-called *independence of irrelevant alternatives* variety.[5]

Another commonly used assumption, which allows some dependence across the elements of ϵ_i, is the one-factor analytic structure:

$$\Omega_i^{RE} = \gamma_1^2 I_{m_i} + \gamma_2^2 i_{m_i} i_{m_i}', \tag{12}$$

where i_{m_i} is the $m_i \times 1$ vector of one's, and γ_1^2, γ_2^2 are variance parameters to be estimated. This implies that the integral in (4) can be written as a univariate integral of a product of cumulative normal distributions, which can be evaluated very efficiently through Gaussian quadrature methods (see Heckman (1981), Butler and Moffit (1982), and Hajivassiliou (1984)).

In many applications it is useful to consider a third model for ϵ_i. This is the natural generalization of (12) that adds an autoregressive structure:

$$\epsilon_{ij} = \alpha_i + \xi_{ij}, \quad \xi_{ij} = \rho\xi_{i,j-1} + \nu_{ij} \qquad (j = 1, \ldots, J), \tag{13}$$

$\nu_{ij} \sim N(0, \sigma_\nu^2)$, $\xi_{i0} \sim N(0, \sigma_0^2)$, $\sigma_0^2 = \sigma_\xi^2 = \sigma_\nu^2/(1-\rho^2)$ by stationarity; $\alpha_i \sim N(0, \sigma_\alpha^2)$, α_i and ξ_{ij} independent.

This one-factor plus AR(1) structure, with a variance-covariance matrix denoted by Ω_i^{AR1RE}, implies that (4) will involve an m_i-dimensional integral, thus rendering efficient classical estimation methods infeasible. This is also true in the most flexible case, when Ω_i is allowed to be a general covariance matrix (i.e., positive definite and symmetric) with additional assumptions imposed to achieve identification. Hence, we turn to the *method of simulated scores*, which avoids the need for multidimensional integration.

3. MSS ESTIMATION OF LDV MODELS WITH CORRELATED ERRORS

In this section we present the method of simulated scores and show that it is applicable to the class of LDV models that can be written as sets of linear

[5] See McFadden (1973) for an explanation of this point.

inequality constraints on the underlying latent variables, the distribution of which belongs to the linear exponential class. This approach builds on an idea by Ruud (1986). Three simulation techniques to use in conjunction with MSS estimation are presented in Section 5. Two of those techniques make MSS continuous in the unknown parameters.

Using standard matrix differentiation results, the derivatives of the likelihood (4) of a typical observation with respect to the parameters β, Γ_i can be shown to satisfy

$$\ell_\beta(\theta; y_i) \equiv \frac{\partial \ell(\theta; y_i)}{\partial \beta} = \ell(\theta; y_i) X' \Omega_i^{-1} E\{y_i^* - X_i \beta | y_i^* \in D(y_i)\}, \tag{14}$$

$$\begin{aligned} \ell_{\Gamma_i}(\theta; y_i) &\equiv \frac{\partial \ell(\theta; y_i)}{\partial \Gamma_i} \\ &= -\ell_i(\theta; y_i) \Omega_i^{-1} \\ &\quad \times \left[I_{m_i} - E\{(y_i^* - X_i \beta)(y_i^* - X_i \beta)' | y_i^* \in D(y_i)\} \Omega_i^{-1} \right] \Gamma_i. \end{aligned} \tag{15}$$

It will be useful for later analysis to write the derivative of (4) with respect to θ as

$$\ell_{i\theta}(\theta; y_i) \equiv \frac{\partial \ell_i(\theta; y_i)}{\partial \theta} = \ell_i(\theta; y_i) E\{h(y_i^* - X_i \beta) | y_i^* \in D(y_i)\}, \tag{16}$$

where $h(u)$ is a vector of terms that are linear or quadratic in $u \equiv y_i^* - X_i \beta$, and depend on X_i and the mapping from the deep parameters θ to β and Γ_i. In the general case with unrestricted Ω_i, the deep parameters θ consist of β and the unique elements of Ω. In the most restrictive covariance model (11), $\theta = [\beta', \gamma_1]'$; in the random-effects model (12), $\theta = [\beta', \gamma_1, \gamma_2]'$; and in the random-effects-plus-AR(1) case (13), $\theta = [\beta', \sigma_\nu^2, \sigma_0^2, \rho, \sigma_\alpha^2]'$.

In general, $h(u)$ will be the vector

$$\begin{bmatrix} X_i' \Omega_i^{-1} u \\ -\Omega_i^{-1} \left[I_{m_i} - uu' \Omega_i^{-1} \right] \Gamma_i \end{bmatrix}, \tag{17}$$

premultiplied by the array of derivatives of $(\beta, \text{vec } \Gamma_i)$ with respect to the deep parameters.

For the general LDV model, the score of a subject is

$$\begin{aligned} s_i(\theta; y_i) &\equiv \frac{\partial \ln \ell_i(\theta; y_i)}{\partial \theta} = E\{h(y_i^* - X_i \beta) | y_i^* \in D(y_i)\} \\ &= \ell_{i\theta} / \ell_i = \frac{\int_{D(y_i)} h(z, X_i, \beta, \Omega_i) n(z - X_i \beta, \Omega_i)\, dz}{\ell_i}. \end{aligned} \tag{18}$$

The set $D(y_i)$ in the four leading cases of LDV models we consider here corresponds to a set of linear inequality constraints on the elements of the latent vector y_i^*, as described in the previous section (equations (7)–(10)).[6]

In view of our assumption that the observations are i.i.d. across individuals, the maximum likelihood estimator is a root of the sum of scores (18) over subjects—see equation (6), implying that at the true parameter vector θ^*,

$$E\left\{\frac{\partial \ln \ell_i(\theta^*)}{\partial \theta}\right\} = E\{\ell_{i\theta}/\ell_i\} = E\{h(y_i^* - X_i\beta^*)|D(y_i)\} = 0.$$

Consider a simulator, $\tilde{h}_{ir} \equiv \tilde{h}_r(X_i\beta, \Omega_i)$, for the score function $h(\cdot)$, satisfying the set of restrictions $y_i^* \in D(y_i)$. Also consider a simulator $\tilde{h}_{iR} \equiv \tilde{h}(X_i\beta, \Omega_i, R) \equiv (1/R)\Sigma_r \tilde{h}_{ir}$, which averages R independent simulations $\tilde{h}_{ir}$. The MSS estimator we propose here replaces hard-to-compute conditional expectation terms in the logarithmic score with simulators $\tilde{h}_{iR}$:[7]

$$\hat{\theta}_{MSS} \underset{\theta}{\text{solves}} \left\{\frac{1}{N}\sum_i \tilde{h}_{iR} = 0\right\}. \tag{19}$$

In case $\tilde{h}_{iR}$ is an unbiased simulator of the score, $\hat{\theta}_{MSS}$ is consistent and asymptotically normal for a finite number of simulations R. Such a simulation method is discussed in Section 5 as simulator SAR, based on acceptance-rejection arguments (Devroye (1986)). The MSS estimator that results by employing simulator SAR will be labelled MSS-SAR.

We further show that one may obtain computationally more tractable MSS estimators by employing (possibly biased) simulators that are continuous in θ. An MSS estimator that uses biased simulations of the score function relies on the fact that the score $s_i(\theta; y_i) = \ell_{i\theta}(\theta; y_i)/\ell_i(\theta; y_i)$, and uses unbiased simulators for $\ell_{i\theta}(\theta; y_i)$ and $\ell_i(\theta; y_i)$. It is thus defined by

$$\hat{\theta}_{MSSB} \underset{\theta}{\text{solves}} \left\{\frac{1}{N}\sum_i \tilde{\tilde{h}}_{i\theta R} = 0\right\}, \tag{20}$$

[6] A similar set of linear inequality constraints on the latent dependent vector can also be defined for the canonical disequilibrium model with T markets, which are observed to be demand- or supply-constrained:

$$\begin{aligned} y_{i1} &= \min(y_{i1}^*, y_{i2}^*), \\ y_{i2} &= \min(y_{i3}^*, y_{i4}^*), \\ &\cdots \\ y_{iT/2} &= \min(y_{i,T-1}^*, y_{iT}^*), \end{aligned}$$

where y_{ij}^* denotes notional demands if j is odd, and notional supplies if j is even.

[7] It is important to point out that all the asymptotic properties we will establish will require that the same underlying random variates used to simulate the $h(\cdot)$ functions must be used throughout the iterative search for the solution to the simulated scores.

where

$$(21) \qquad \tilde{\bar{h}}_{i\theta R} \equiv \frac{1}{R}\sum_{r=1}^{R}\frac{\tilde{\ell}_{i\theta r}(\theta; y_i)}{\tilde{\ell}_{ir}(\theta; y_i)},$$

such that $E\tilde{\ell}_{\theta ir} = \ell_{\theta i}$, $E\tilde{\ell}_{ir} = \ell_i$, and $\tilde{\ell}_{ir} \xrightarrow{p} \ell_i$ with R.

It is important to note that in general it is advantageous to simulate the numerator and denominator expressions *simultaneously* as in (21) rather than separately $((1/R_n)\sum_{r=1}^{R_n}\tilde{\ell}_{i\theta r}/(1/R_d)\sum_{s=1}^{R_d}\tilde{\ell}_{is})$. This is because by using the same underlying draw to simulate numerator and denominator instead of one draw for the numerator and an independent one for the denominator, a positive correlation is built into the numerator and denominator simulations, which reduces the variance of the simulator for the ratio.[8]

In Section 5 we give two simulators that are continuous in θ. The first such simulator, simulator SRC, is based on a recursive conditioning of the normal multivariate density through a Cholesky triangularization of its variance-covariance matrix and allows us to define MSS-SRC along the lines of (20)–(21). A second continuous method, simulator GRS, is also presented in Section 5 and employs Gibbs resampling methods. The resulting estimator, MSS-GRS, is an implementation of (19). As we show in Section 6, for MSS-SRC based on R simulations to be CUAN as $N \to \infty$, R must rise at a rate faster than $\sqrt{N}$, whereas if the simulator based on Gibbs-resampling (MSS-GRS) is employed with R simulations, then n_G, the number of Gibbs resamplings used to generate each simulation, must grow faster than the (much) slower rate $\log N$. When, in addition, R grows without bound (at any rate), MSS-GRS achieves the full asymptotic efficiency of MLE.

These features are a marked improvement over the properties of the first simulation estimation method for LDV models developed by Lerman and Manski (1981). These authors explored the use of simulation in the context of estimating the classic discrete choice model and proposed the estimator

$$(22) \qquad \hat{\theta}_{LM} = \arg\max_{\theta}\left\{\frac{1}{N}\sum_{i}\frac{1}{R}\sum_{r}\ln\tilde{\ell}_{ir}\right\},$$

such that the likelihood contributions ℓ_i are simulated unbiasedly ($E\tilde{\ell}_{ir} = \ell_i$) and consistently with R ($\tilde{\ell}_{ir} \xrightarrow{p} \ell_i$). Lerman and Manski used the empirical choice probabilities as the simulating function $\tilde{\ell}$. This estimator is a discontinuous function of the parameters and it is not bounded away from 0 and 1. Hence, because of these problems Lerman and Manski found that their estimator required a very large number of simulations for satisfactory performance.

The fact that MSS relies on the idea that the score for the general linear exponential model can be written as conditional expectations, which might be simulated directly, implies that MSS is generally applicable to any LDV model

[8] This is confirmed by the Monte-Carlo experiments in Hajivassiliou (1996).

that can be written as a set of linear inequality constraints on the underlying latent variables, the distribution of which belongs to the linear exponential class. Four illustrations were given in (7)–(10). Hence, the method does not require the development of ad hoc simulation techniques for each type of LDV model that is under consideration. This generality of the MSS estimator improves on existing estimation methods of simulated moments (MSM) which require specialized arguments for different classes of LDV models. See, for example, the MSM approach developed by McFadden (1989) for the special case of the multinomial probit model. The case of multiperiod binary discrete response can be thought of as a multinomial probit model over the choice set $C = \{-1, +1\}^{m_i}$, with 2^{m_i} possible patterns of choice over time. The fact that m_i is fairly large in typical applications[9] renders intractable simple frequency simulators for choice-probabilities in the moment conditions. A similar difficulty with the MSM approach arises in the rank order probit model (model 4 above) because with J available alternatives in the choice set there exist $J!$ possible orderings that would require that their probabilities be calculated.

It should also be noted that the simulation estimators considered here apply also to general LDV models that have both discrete and continuous features by writing the joint likelihood as the product of the marginal distribution of the continuous part times the conditional distribution of the discrete part conditional on the continuous. Typically, analytic expressions will exist for the continuous part, while simulation methods will be applied to the conditional discrete part. As an illustration, consider the multiperiod Tobit or censored regression model, where $y_{it} = \max(0, y_{it}^*)$.

Define

$$I = I_i(y_i) = \{t | y_{it} = 0, t = 1, \ldots, T\},$$

$$J = J_i(y_i) = \{t | y_{it} > 0, t = 1, \ldots, T\}.$$

The likelihood for a respondent is

$$\ell(y_i, \theta) = \int_{y_I^* \le 0} n(y_I^* - X_I\beta, y_J - X_J\beta, \Omega)\, dy_I^*,$$

where y_I^* is the subvector of y_i^* with components in I. But

$$n(y_I^* - X_I\beta, y_J - X_J\beta, \Omega) = n(y_J - X_J\beta, \Omega_{JJ}) \cdot n\left(y_I^* - \mu_I, \tilde{\Omega}_{II}\right),$$

with $\mu_I \equiv E(y_I^* | y_J)$ and $\tilde{\Omega}_{II} \equiv \operatorname{var}(y_I^* | y_J)$.

The log-likelihood for a respondent then consists of a term that has a closed form expression and a second term that is a multinomial probability that all components of y_I^* are nonpositive.

A further considerable advantage of MSS estimators is that because they simulate directly the conditional expectation expressions that appear linearly in

[9] For most countries in the sample of Hajivassiliou (1994), the number of time periods with available data is 17.

the scores, they implicitly employ the optimal instrument functions in a generalized method of moments context. This issue is found to be critical in the Monte-Carlo study of Hajivassiliou (1996): for satisfactory efficiency, MSM estimation requires good approximations to optimal instruments, which in general is difficult to achieve. In contrast, MSS automatically uses the optimal instruments for asymptotic efficiency.

Let us now outline the three simulation methods we propose in Section 5 to use in conjunction with MSS estimators. The simulator SAR is based on a suggestion in Ruud (1986) to use an unbiased simulator of the conditional expectation $E\{h(y_i^* - X_i\beta | D(y_i))\}$ that appears in the logarithmic score, by drawing standard normal vectors η sequentially until R values of $y_i^* = X_i\beta + \Gamma_i\eta \in D(y_i)$ are observed, where R is fixed in advance, then forming a sample average of $h(y_i^* - X_i\beta)$ for the y_i^* drawn that are in $D(y_i)$.[10] Define $\delta_{D(y_i)} = 1$ if $y_i^* \in D(y_i)$, $\delta_{D(y_i)} = 0$ otherwise. If $\ell_i(\theta; y_i) = E\delta_{D(y_i)}(y_i^*)$ is small, as should be expected in realistic cases with a large number of alternative choices or choice patterns over time, then a large simulation sample is required to obtain the simulator. Simulator SAR is an alternative method based on acceptance-rejection arguments that is computationally much more efficient. (See Press et al. (1986) and Devroye (1986) for using the acceptance-rejection method to generate nonuniform random variates.) These approaches also yield discontinuous estimators but have the advantage that a finite number of terminal simulations, R, is needed for MSS to be CUAN.

Further, the SRC simulator in Section 5 is motivated by the observation that the conditional expectation expression that appears in (18) can be written as

$$E(h(y_i^* - X_i\beta)|D(y_i)) = \frac{\int h(y_i^* - X_i\beta)\,\delta_{D(y_i)} n(y_i^* - X_i\beta, \Omega_i)\,dy_i^*}{\int \delta_D(y_i) n(y_i^* - X_i\beta, \Omega_i)\,dy_i^*} = \ell_{i\theta}/\ell_i. \tag{23}$$

Hence implementation of MSS-SRC in the form of (20) and (21) requires simulators with good properties for the derivative vectors $\ell_{i\theta}$ and the probabilities ℓ_i. Note, in particular, that SRC is by design bounded away from 0 and 1, which is crucial given that ℓ_i appears in the denominator of the score expression.

As we discussed earlier, one can think of an alternative MSS estimator that simulates the numerator and denominator of the score ratio *separately* of one another in the form

$$\tilde{\theta}_{MSSBR} \underset{\theta}{\text{solves}} \left\{ \frac{1}{N}\sum_i \left[\frac{1}{R_n}\sum_r^{R_n} \ell_{\theta ir} \Big/ \frac{1}{R_d}\sum_s^{R_d} \ell_{is} \right] = 0 \right\}. \tag{24}$$

A method in the literature that works along these lines is due to van Praag and Hop (1987). Their method employs independent simulators of the numerator

[10] See Ruud (1991) for combining these ideas with the EM algorithm.

and denominator of (23).[11] As we explained already, our (20)–(21) implementation is preferable because it reduces simulation variance by simulating simultaneously the score as a ratio. In either case, it is imperative that one not use a frequency simulator for the denominator expression because the frequency simulator is not bounded away from 0. The recursive conditioning simulator SRC we discuss in Section 5 is by design bounded away from 0 and 1 and, in addition, is smooth, which avoids a second difficulty with the frequency simulator.[12]

Finally, introducing the Markovian updating scheme known as Gibbs resampling we obtain simulator GRS which estimates the complete score function as in (19). An MSS estimator based on simulator GRS R terminal simulations, each based on n_G resamplings (or "burn-in" cycles), is CUAN for any R, provided n_G grows faster than log N. In addition, MSS-GRS will achieve the full asymptotic efficiency of MLE as R grows without bound at any rate. These are very satisfactory rates, especially given the smoothness and the computational simplicity of this simulator.

4. ASYMPTOTIC DISTRIBUTION OF MSS ESTIMATORS

Let $s_i(\theta) = \ell_{i\theta}(\theta)/\ell_i(\theta)$ denote the score for observation i, and let $\tilde{s}_i(\theta)$ denote the simulated value of $s_i(\theta)$, for a sample of independently, identically distributed observations $i = 1, \ldots, N$. Define a simulation bias,

$$B_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [\boldsymbol{E}_i \tilde{s}_i(\theta) - s_i(\theta)], \tag{25}$$

where $\boldsymbol{E}_i$ denotes an expectation with respect to the simulation process, given the observation. Define a simulation residual process,

$$\zeta_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i(\theta), \tag{26}$$

$$\text{with} \quad \xi_i(\theta) \equiv [\tilde{s}_i(\theta) - \boldsymbol{E}_i \tilde{s}_i(\theta) - \tilde{s}_i(\theta^*) + \boldsymbol{E}_i \tilde{s}_i(\theta^*)].$$

Following the method of McFadden (1989) and Pakes and Pollard (1989), we show that assumptions on the simulation bias and simulation residual process, plus regularity assumptions, are sufficient for the MSS estimator $\hat{\theta}_N$ that solves $\sum_{i=1}^{N} \tilde{s}_i(\hat{\theta}_N) = 0$ to be consistent and asymptotically normal.

THEOREM 1: *Assume that the parameter θ is contained in a compact set Θ, and that the true value θ^* is in the interior of Θ. Assume that the score $s_i(\theta)$ is continuously differentiable on Θ. Assume that the score and its derivatives, and the*

[11] See equation (22) on p. 19 of their paper.

[12] Another simulator of LDV probabilities which is smooth and bounded away from 0 and 1 is due to Stern (1992). Extensive Monte Carlo evidence in Hajivassiliou et al. (1996) shows that simulator SRC strictly dominates the Stern simulator in terms of simulation MSE.

simulated score, are dominated by a function independent of θ *with finite first and second order moments. Assume that* $E_i s_i(\theta) = 0$ *if and only if* $\theta = \theta^*$, *and that* $J = -E_i s_{i\theta}(\theta^*)$ *is positive definite, where* 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 is stochastically equicontinuous.*[13] *Assume that a MSS estimator solving* $0 = \sum_{i=1}^{N} \tilde{s}_i(\hat{\theta}_N)$ *exists for each* N.[14] *Then, the estimator satisfies* $\hat{\theta}_N \xrightarrow{p} \theta^*$ *and* $\sqrt{N}(\hat{\theta}_N - \theta^*) \xrightarrow{d} Z \sim \mathcal{N}(0, J^{-1} + J^{-1}QJ^{-1})$, *where* $Q = E[\tilde{s}_i(\theta^*) - E_i\tilde{s}_i(\theta^*)][\tilde{s}_i(\theta^*) - E_i\tilde{s}_i(\theta^*)]'$.

PROOF: The defining equations for the estimator can be written, by adding and subtracting terms, as

$$0 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} s_i(\hat{\theta}_N) \equiv A_N + C_N(\hat{\theta}_N) + \zeta_N(\hat{\theta}_N) + B_N(\hat{\theta}_N), \tag{27}$$

with

$$A_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} s_i(\theta^*) + \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [\tilde{s}_i(\theta^*) - E_i\tilde{s}_i(\theta^*)],$$

$$C_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [s_i(\theta) - s_i(\theta^*)].$$

The i.i.d. assumption on the observations and simulation, the dominance condition that implies the existence of moments, and the condition $Es_i(\theta^*) = 0$, imply by the Lindberg-Levy central limit theorem that A_N is asymptotically normal with mean zero and covariance matrix $J + Q$. Then $A_N/\sqrt{N} = o_p(1)$. The stochastic equicontinuity assumption (ii) implies that ζ_N is uniformly stochasti-

[13] 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_0 such that for $N \geq N_0$,

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

If ζ_N is stochastically equicontinuous at Θ, with Θ compact and convex, and $\zeta_N(\theta^\circ)$ is stochastically bounded for some $\theta^\circ \in \Theta$, then ζ_N is uniformly stochastically bounded on Θ. This follows by noting that at most $2M/\delta$ points less than a distance δ apart are required on a line segment between θ° and any $\theta \in \Theta$, where M bounds the diameter of Θ. Then,

$$\sup_{\theta \in \Theta} |\zeta_N(\theta) - \zeta_N(\theta^\circ)| \leq (2M/\delta) \sup_{|\theta' - \theta''| < \delta} |\zeta_N(\theta') - \zeta_N(\theta'')|,$$

implying that given $\epsilon, \lambda > 0$, there exists $\delta > 0$ such that $\text{prob}(\sup_{\theta \in \Theta} |\zeta_N(\theta) - \zeta_N(\theta^\circ)| > 2M\epsilon/\delta) < \lambda$. This works for the simulation residual process in this section with $\theta^\circ = \theta^*$ since $\zeta_N(\theta^*) = 0$.

[14] It is sufficient to define $\hat{\theta}_N$ to be an approximate solution satisfying $\mathcal{O}(1) = \sum_{i=1}^{N} \tilde{s}_i(\hat{\theta}_N)$; such an estimator always exists.

cally bounded, and hence that $\zeta_N(\hat{\theta}_N)/\sqrt{N} = o_p(1)$, and assumption (i) on the simulation bias implies that $B_N(\hat{\theta}_N) = o_p(1)$. The continuous differentiability of s_i on Θ and the moment conditions imply that $C_N(\theta)/\sqrt{N}$ satisfies a Uniform Law of Large Numbers, converging to a continuously differentiable function $\psi(\theta)$ that is bounded away from zero when $\theta \neq \theta^*$, with $\psi_\theta(\theta^*) = -J$.[15] From (27), one then has

$$0 = C_N(\hat{\theta}_N)/\sqrt{N} + o_p(1) = \psi(\hat{\theta}_N) + o_p(1),$$

implying that $\hat{\theta}_N \xrightarrow{p} \theta^*$.

Next, $\sqrt{N}(\hat{\theta}_N - \theta^*)$ is shown to be stochastically bounded. The asymptotic normality of A_N, and assumptions (i) and (ii), imply

$$\mathcal{O}_p(1) = C_N(\hat{\theta}_N) = \left(\frac{1}{N}\sum_{i=1}^{N}\left[s_{i\theta}(\theta^*) + \mathcal{O}(\hat{\theta}_N - \theta^*)\right]\right)\sqrt{N}(\hat{\theta}_N - \theta^*),$$

with the second equality following from a Taylor's expansion of $s_i(\hat{\theta}_N)$ about θ^* using the differentiability and dominance assumptions. Then, $Es_{i\theta}(\theta^*)$ nonsingular and $\hat{\theta}_N \xrightarrow{p} \theta^*$ imply $\sqrt{N}(\hat{\theta}_N - \theta^*) = \mathcal{O}_p(1)$. Note that this result and assumption (ii) imply that $\zeta_N(\hat{\theta}_N) = o_p(1)$.

To establish asymptotic normality, use the Taylor's expansion above of $C_N(\hat{\theta}_N)$ and assumptions (i) and (ii) in Theorem (1) to obtain

$$0 = A_N + \left(\frac{1}{N}\sum_{i=1}^{N}\left[s_{i\theta}(\theta^*) + \mathcal{O}(\hat{\theta}_N - \theta^*)\right]\right)\sqrt{N}(\hat{\theta}_N - \theta^*) + o_p(1).$$

But, $(1/N)\sum_{i=1}^{N} s_{i\theta}(\theta^*) \xrightarrow{p} J$, implying

$$\sqrt{N}(\hat{\theta}_N - \theta^*) = -J^{-1}A_N + o_p(1) \xrightarrow{d} Z \sim \mathcal{N}(0, J^{-1}(J+Q)J^{-1}). \qquad Q.E.D.$$

5. CONDITIONAL NORMAL DISTRIBUTIONS AND SIMULATION

In this section we present three simulation techniques to use with MSS estimators. The first two are continuous in the unknown parameters, and provide asymptotically unbiased simulators of the score. Asymptotic unbiasedness of simulator SRC requires the number of simulations employed to grow without bound, while simulator GRS, which uses a finite number of simulations, is asymptotically unbiased as the number of resamplings used to generate each simulation rises without bound. The third simulator, SAR, is an unbiased estimator for the score for a finite number of simulations, and is a discontinuous

[15] A U.L.L.N. states that given $\epsilon, \delta > 0$, there exists N_0 such that for $N \geq N_0$, $\text{prob}(\max_{\theta \in \Theta}|C_N(\theta) - \psi(\theta)| > \delta) < \epsilon$.

function of the unknown parameters. The consistency and asymptotic normality of MSS estimators based on any of these three simulation methods is established in Section 6 specializing the general results of Section 4.

We illustrate our methods for the leading distributional case of multivariate normality. Consider the general normal LDV model:

$$(28) \qquad y_i = \tau(y_i^*), \quad y_i^* \sim N(X_i\beta, \Omega_i) \qquad (i = 1, \ldots, N).$$

For simplicity, we will drop the i index whenever no ambiguity would arise. The MSS estimator requires simulating the $h(y^*)$ functions that appear in the scores, conditional on $y^* \in D(y)$. Hence, our general objective is to obtain random draws from the distribution y^* subject to $y = \tau(y^*)$. Then, we see from Section 3 that three types of functions need to be simulated. The first function is the likelihood contribution ℓ, given by (4). The second is the likelihood derivative ℓ_θ (equations (14) and (15)). Finally, the third function is the logarithmic score $\partial \ln \ell / \partial\theta$, given by (18).

Our general objective will be to develop unbiased simulators for these functions that are computationally very fast, and simulators that, though only asymptotically unbiased, possess biases that vanish at sufficiently fast rates as to guarantee consistency and asymptotic normality of MSS estimators that employ them.[16] The first continuous simulator is based on the idea of employing a Cholesky triangularization so as to make the constraints $y^* \in D(y)$ recursive.[17] This will make simulator SRC unbiased for the likelihood contributions and asymptotically unbiased for the logarithmic scores. The second continuous simulator employs repeated drawings from univariate truncated conditional normal distributions and applies Gibbs resampling methods (Geman and Geman (1984), Gelfand and Smith (1990), Chib and Greenberg (1995)) to ensure that the joint distribution from which we are simulating converges to the appropriate multivariate truncated normal distribution. Hence, simulation method GRS will provide unbiased drawings of likelihood contributions and scores as the number of Gibbs resamplings rises to infinity. Finally, we describe a third simulator, SAR, based on acceptance-rejection arguments, which though a discontinuous function of the underlying model parameters, provides unbiased drawings of likelihood contributions and scores for a finite number of terminal simulations used.

We first introduce some notation: For a vector of indices $(1, \ldots, J)$, we use the notation "$<j$" to denote the subvector $(1, \ldots, j-1)$, "$\leq j$" to denote the subvector $(1, \ldots, j)$, and "$-j$" to denote the subvector that excludes component j. Thus, for a matrix L, $L_{j,<j}$ denotes a vector containing the first $j-1$

[16] Unbiased and consistent simulators for the integrals appearing in expressions (4), (14), and (15) can also be obtained through importance sampling and other methods (see Moran (1984, 1985, 1986), Deák (1980), McFadden (1989), Stern (1992)). These methods cannot be used for direct unbiased simulation of the logarithmic score (18), unless an infinite number of simulations are averaged.

[17] Geweke (1989) uses this triangularization in a Bayesian context and Keane (1994) employs it in the special case of estimating by simulation a multiperiod (panel-data) binary probit model.

elements of row j, and $L_{-j,-j}$ denotes the subarray excluding row j and column j. For a vector ϵ, $\epsilon_{<j}$ is the subvector of the first $j-1$ components, and ϵ_{-j} is the subvector excluding component j.

Define $q(u,a,b) \equiv \Phi^{-1}(\Phi(a)\cdot(1-u) + \Phi(b)\cdot u)$, where $0 < u < 1$ and $-\infty \leq a < b \leq \infty$. Then q is a mapping that takes a uniform $(0,1)$ random variate into a truncated standard normal random variate on the interval $[a,b]$.

PROPOSITION 1: *Consider the multivariate normal $J \times 1$ random vector $Y \sim N(X\beta, \Omega)$ with Ω positive definite, the linear transformation $Z = MY \sim N(MX\beta, \Sigma)$, with M nonsingular and $\Sigma = M\Omega M'$, and the event $\mathbf{B} \equiv \{a^* \leq Z = MY \leq b^*\}$, with $-\infty \leq a^* < b^* \leq +\infty$. Define $P \equiv \int_B n(z; MX\beta, \Sigma)\,dz$, $a \equiv a^* - MX\beta, b \equiv b^* - MX\beta$, and let L denote the lower-triangular Cholesky factor of Σ. Let $(u_1, \ldots, u_J)$ be a vector of independent uniform $(0,1)$ random variates. Define recursively for $j = 1, \ldots, J$:*

$$e_j = q\big(u_j, (a_j - L_{j1}e_1 - \cdots - L_{j,j-1}e_{j-1})/L_{jj}, \tag{29}$$
$$(b_j - L_{j1}e_1 - \cdots - L_{j,j-1}e_{j-1})/L_{jj}\big),$$

$$Q_j \equiv \Phi\big((b_j - L_{j1}e_1 - \cdots - L_{j,j-1}e_{j-1})/L_{jj}\big) \tag{30}$$
$$- \Phi\big((a_i - L_{j1}e_1 - \cdots - L_{j,j-1}e_{j-1})/L_{jj}\big).$$

Define $e \equiv (e_1, \ldots, e_J)'$, $\tilde{Y} \equiv X\beta + M^{-1}Le$, and $Q(e) \equiv Q_1 \cdot \cdots \cdot Q_J$. Then $\tilde{Y}$ is a random vector on $\mathbf{B}$, and the ratio of the densities of $\tilde{Y}$ and Y at $y = X\beta + M^{-1}Le$, where e is any vector satisfying $a \leq Le \leq b$, is $P/Q(e)$.

PROOF: Since the ratio of densities is preserved by a common linear transformation, it is sufficient to consider the ratio of the density of the random vector e constructed recursively above and the density of a standard normal random vector ϵ conditioned on the event $a \leq L\epsilon \leq b$. But these densities differ only in their normalizing denominators, $Q(e)$ and P respectively. *Q.E.D.*

It is important to reiterate that the $\tilde{Y}$'s implied by the sequential scheme of Proposition 1 are not distributed according to the multivariate truncated normal distribution $Y \sim N(X\beta, \Omega)$ conditional on $a^* \leq MY \leq b^*$ because the recursive constraints defined by the Cholesky decomposition are not independent. This fact should be evident from a simple two-dimensional example. Suppose $b_1 = b_2 = \infty$ as is the case in the probit model, and $l_{21} > 0$, corresponding to a positive correlation between Y_1 and Y_2. Draws of e_1 according to the inequality in ((29), $j = 1$) will ignore the constraint in ((29), $j = 2$), hence will be too small on average. Given an e_1 too small, e_2, obeying the second constraint ((29), $j = 2$), will be too large on average.

Despite this fact, we can show that combining Proposition 1 together with importance-sampling arguments we obtain smooth, unbiased, and direct simulators for the likelihood contributions ℓ_i and their derivatives $\ell_{\theta i}$, and a smooth,

asymptotically unbiased simulator of the score function, termed simulator SRC. The results in Section 6 will establish that MSS estimators based on simulator SRC will be CUAN provided the number of simulations used grows faster than $\sqrt{N}$.

5.1. *SRC: A Smooth, Direct Simulator for Likelihood Scores and Contributions*

The likelihood contribution of the general LDV model examined in this paper is given by (4), which we can rewrite, after dropping index i, as

$$\ell(y, X; \beta, \Omega) = \int_{a^*(y) \le M(y)\cdot z \le b^*(y)} n(z - X\beta, \Omega)\, dz \tag{31}$$

$$= \text{prob}[a^*(y) \le M(y)\cdot Y \le b^*(y);\ Y \sim N(X\beta, \Omega)].$$

But

$$\text{prob}[a^*(y) \le M(y)\cdot Y \le b^*(y); Y \sim N(X\beta, \Omega)]$$

$$= \text{prob}[a(y, X, \beta, \Omega) \le L(y, \Omega)\cdot \nu$$

$$\le b(y, X, \beta, \Omega);\ \nu \sim N(0, I)],$$

with a, L, and b as defined in Proposition 1. Hence, the likelihood contribution becomes

$$\ell(y, X; \beta, \Omega) \tag{32}$$

$$= \text{prob}[a(y, X, \beta, \Omega) \le L(y, \Omega)\cdot \nu \le b(y, X, \beta, \Omega);\ \nu \sim N(0, I)]$$

$$= \int_{a(y, X, \beta, \Omega) \le L(y, \Omega)\cdot \nu \le b(y, X, \beta, \Omega)} \prod_{j=1}^{J} \phi(\nu_j)\, d\nu_j.$$

Now consider a $J \times 1$ vector e_r drawn according to the sequential scheme described in equations (29). Obtain R such vectors e_r''s and define the likelihood contribution simulator $\tilde{\ell}(e; y, X; \beta, \Omega; R)$:

$$\tilde{\ell}(e; y, X; \beta, \Omega; R) = \frac{1}{R} \sum_{r=1}^{R} \prod_{j=1}^{J} Q_j(e_{1r}, \ldots, e_{j-1,r}), \tag{33}$$

where $Q_1 \equiv \text{prob}(a_1/l_{11} \le e_1 \le (b_1/l_{11})$, and for $j < 1$

$$Q_j(e_1, \ldots, e_{j-1})$$

$$\equiv \text{prob}\big((a_j - L_{j,<j}\cdot e_{<j})/l_{jj} \le e_j$$

$$\le (b_j - L_{j,<j}\cdot e_{<j})/l_{jj} e_1, \ldots, e_{j-1}\big).^{18}$$

[18] Recall that since $e_j \sim N(0, 1)$, $\text{prob}(k_1 \le e_j \le k_2) = \Phi(k_2) - \Phi(k_1)$.

LEMMA 1: *The simulator* $\tilde{\ell}(e; y, X; \beta, \Omega; R)$ *defined by* (33) *is an unbiased estimator of* $\ell(y, X; \beta, \Omega)$.

PROOF: It is sufficient to show the Lemma for $R = 1$. The expected value of $\tilde{\ell}$ is $E\tilde{\ell} = \int \tilde{\ell}(e) f(e)\, de$, where $f(e)$ denotes the density that generates the (biased) sequential truncated draws e_r in Proposition 1. By (33), the definition of $\tilde{\ell}$, and result (32),

$$E\tilde{\ell} = \int_{-\infty}^{\infty} \left(\prod_{j=1}^{J} Q_j \right) \cdot \left(\prod_{j=1}^{J} \phi(e_j)/Q_j \right) de_1 \cdots de_J$$

$$= \int_{a \le L \cdot e \le b} \prod_{j=1}^{J} \phi(e_j)\, de_j$$

$$= \text{prob}(a \le L \cdot \nu \le b) = \ell(y, X; \beta, \Omega). \qquad Q.E.D.$$

The combination of the recursive conditioning method, the above Lemma, and the smooth univariate truncated variate generation algorithm produces an unbiased (for any value of R) multivariate probability simulator that is smooth, i.e., a continuous and differentiable function of the model parameters β and Ω.[19] Moreover, apart from an initial Cholesky decomposition and several matrix multiplications, most computational effort is in drawing the univariate truncated normal variates according to the steps in (29). This effort is approximately linear in J, the dimension of the probability integral, which is an extremely convenient feature of simulator SRC. The results of Börsch-Supan and Hajivassiliou (1993) confirm the excellent computational efficiency of simulator SRC. In terms of timing, we find that generating 1000 simulations $\tilde{\ell}_r$ according to this algorithm from the 10-dimensional distribution $Y \sim N(\mu, \Sigma)$ with $\Sigma = \{\sigma_{jj} = 1, \sigma_{ij} = .5$ for $i \neq j\}$ subject to $Y_j > 0$ if j is even and $Y_j < 0$ if j is odd, required 5.3 seconds on a 486/33 MHz Personal Computer, and 7.4 seconds for 20 dimensional Y vectors. In contrast, a (discontinuous) acceptance-rejection algorithm required 8 minutes for the 10 dimensional case and nearly 1 hour for the 20 dimensional one.

To obtain a smooth and asymptotically unbiased simulator for s_i, the logarithmic score (18), recall that $s \equiv \ell_\theta / \ell = E[h(Y - X\beta) | Y \in D(y)]$. Hence, we define

$$\tilde{s}_R \equiv \tilde{\ell}_{\theta R} / \tilde{\ell}_R, \tag{34}$$

where $\tilde{\ell}_{\theta R} \equiv (1/R) \Sigma_r \{h(M^{-1} L e_r) \cdot \Pi_j Q_j(e_{j-1,r})\}$, and $\tilde{\ell}_R \equiv (1/R) \times \Sigma_r \Pi_j Q_j(e_{j-1,r})$.[20] From Lemma 1 given above, $E\tilde{\ell}_R = \ell$; given the linear form of the likelihood derivative $h(\cdot)$ function, an exactly analogous importance

[19] See also Börsch-Supan and Hajivassiliou (1993).

[20] As discussed in Section 3, employing the same e_r's for the numerator and denominator expressions results in MSS estimators with better statistical properties compared to using independent e's for the numerator and denominator.

sampling argument as the one used in the proof of the lemma establishes that $E\tilde{l}_{\theta R} = \ell_\theta$. Hence, a standard law of large numbers implies that as $R \to \infty$, the simulator for the denominator, $\tilde{\ell}_R$, converges to $\ell = E\{\mathbf{1}(Y \in D(y))\}$, the probability of the event $Y \in D(y)$, and the simulator for the numerator, $\tilde{\ell}_{\theta R}$, converges to $\ell_\theta = E\{h(Y) \cdot 1(Y \in D(y))\}$. Thus, $\tilde{s}_R \xrightarrow{p} \ell_\theta / \ell \equiv s$. See the Monte-Carlo experiments in Hajivassiliou (1996) for an investigation of the choice of R for satisfactory performance of this simulator. Moreover, the continuity in $\tilde{u}$ and the unknown parameters makes estimators based on this simulator extremely fast. Therefore, one can afford quite high R values, because the necessary time is approximately linear in the dimension of Y and is independent of the magnitude of prob$(Y \in D(y_i))$, in sharp contrast to discontinuous simulators. Thus we see that the SRC simulator breaks a major "curse of dimensionality" of simulation in this context. A further feature of the simulator that helps explain its distinctly superior performance when used for estimation compared to (discontinuous) frequency simulators, is that, unlike the latter, the smooth, recursive conditioning simulator presented here is bounded away from 0 and 1. This is obvious from (33).

For more details on the comparative performance of simulator SRC in simulating LDV probabilities and derivatives, see, inter alia, Börsch-Supan and Hajivassiliou (1993), Hajivassiliou (1993), and Hajivassiliou et al. (1996). These studies confirm a general consensus that has developed in the recent literature that the SRC simulator is impressively accurate in approximating LDV probabilities and derivatives, especially when taking into account the low computational effort it involves. A key intuition behind these findings seems to be that the Cholesky triangularization underlining the SRC method implies an importance-sampling distribution that, while computationally extremely tractable, provides an excellent approximation to the true correlation structure of the unobservables. This well-documented ability of the SRC simulator to approximate extremely accurately LDV probabilities with low computational effort is particularly useful in selecting among multiple roots of the score equations in models where the likelihood function does not have a unique global maximum.

5.2. *GRS: An Infinite Algorithm for Generating Truncated Multivariate Normal Variates Based on Gibbs Resampling*

In this section we show that by employing Gibbs resampling techniques (Geman and Geman (1984)) we can devise another smooth simulation technique, simulator GRS, which has the correct truncated multivariate density $Y \sim N(X\beta, \Omega)$ conditional on $a^* \leq MY \leq b^*$ asymptotically with the Gibbs resampling rounds, n_G. Though the Gibbs-based simulator GRS only guarantees drawing from the correct multivariate truncated normal distribution as the number of Gibbs resamplings rises without bound, the Monte-Carlo findings in Hajivassiliou (1996) suggest that the convergence rate of this method is very rapid. This finding confirms the result in Section 6 that MSS estimators using

the Gibbs-resampling-based simulator are consistent and asymptotically normal provided that number of Gibbs resamplings (or "burn-in" cycles) grows faster than $\log N$. Hence this paper shows for the first time that Gibbs resampling techniques can be useful for classical inference.

The Gibbs sampler was developed for and has been applied to the problems of complex, large scale stochastic models, such as image reconstruction, neural networks and expert systems.[21] In these cases, direct specification of a joint distribution is typically not feasible. Instead, the full set of conditionals is specified. Consider a $J \times 1$ variate random vector Y and let

$$[Y_j | Y_{-j}] \qquad (j = 1, \ldots, J), \tag{35}$$

denote the distribution of the variable Y_j conditional on all the random variables constituting Y excluding Y_j.

For the purposes of this section, we further assume that the truncation region (a, b) of the multivariate normal distribution in (28) is compact, which is equivalent to assuming $-\infty < a < b < +\infty$. This does not entail any loss of empirical generality, since we can consider large compact rectangles defined, for example, by the limits of computing machine representation of floating point numbers. We let $\boldsymbol{B}$ denote the (compact) rectangle $[\bar{a}, \bar{b}]$.

Gibbs sampling is a Markovian updating scheme that proceeds as follows. Given an arbitrary starting set of values $Y_1^{(0)}, Y_2^{(0)}, \ldots, Y_J^{(0)}$, we draw $Y_1^{(1)} \sim [Y_1 | Y_2^{(0)}, \ldots, Y_J^{(0)}]$, then $Y_2^{(1)} \sim [Y_2 | Y_1^{(1)}, Y_2^{(0)}, \ldots, Y_J^{(0)}]$, $Y_3^{(1)} \sim [Y_3 | Y_1^{(1)}, Y_2^{(1)}, Y_3^{(0)}, \ldots, Y_J^{(0)}], \ldots$, and so on, up to $Y_J^{(1)} \sim [Y_J | Y_1^{(1)}, \ldots, Y_{J-1}^{(1)}]$. Thus each variable is "visited" in the "natural" order and a cycle in this scheme requires J random variate generations. After n_G such iterations we would arrive at $Y^{(n_G)} \equiv (Y_1^{(n_G)}, \ldots, Y_J^{(n_G)})$. Proposition 2 will establish that $Y^{(n_G)}$ will asymptotically have the true joint distribution of Y as n_G grows without bound. In our case, we let Y describe the distribution of $Y \sim N(X\beta, \Omega)$ conditional on $a^* \leq M \cdot Y \leq b^*$, and let $Y_r^{(n_G)}$ be a vector drawn according to the Gibbs scheme after n_G resamplings. By (18), the logarithmic score, s, equals the expectation of $h(Y, X, \beta, \Omega)$ over the distribution of Y. It then follows trivially that $Eh(Y_r^{(n_G)}, X, \beta, \Omega)$ converges to s as the number of Gibbs resamplings, n_G, grows to infinity. Hence, we define simulator GRS by

$$\tilde{s}(Y^{(n_G)}, y, X, \beta, \Omega, n, R) \equiv (1/R) \sum_r h(Y_r^{(n_G)}, y, X, \beta, \Omega),$$

where R is the (finite) number of terminal simulations drawn, and n_G the number of Gibbs resamplings used for each simulation. Though $\tilde{s}$ is unbiased for the true s only asymptotically with n_G, we prove in Section 6 that the MSS estimator using simulator GRS is CUAN provided n_G rises at a rate at least as fast as $\log N$. In addition, it will achieve the full efficiency of MLE as $R \to \infty$ at any rate.

[21] The relevance of Gibbs resampling methods to our problem was suggested to us by John Geweke. McCulloch and Rossi (1994) develop a Bayesian estimator for the MNP model based on the Gibbs simulator.

Geman and Geman (1984) establish various convergence results of the Gibbs resampling scheme under mild regularity conditions for a finite sites and states problem. Given our interest in the normality case, which is continuous, the Geman and Geman (1984) results are not directly applicable. We are able, however, to establish analogous results for the continuous case, by exploiting results in Orey (1971) about the behavior of general state-space Markov chains.[22] Consider a set A with positive Lebesgue measure. The following definitions are standard (for example, see Orey (1971), Futia (1982)):

DEFINITION 1: A Markov process is *uniformly recurrent* if the probability of reaching state A within n_G transitions is bounded below by a positive number, uniformly in the starting point x.

DEFINITION 2: A density $f(x)$ is an *invariant* of the Markov process if it describes the distribution of the outcomes of the process irrespective of the number of transitions.

In the Gibbs sampler application, one transition corresponds to one updating cycle: start from $(Y_1^{(0)},\ldots,Y_J^{(0)})$, draw $\tilde{Y}_1$ from $[\tilde{Y}_1|Y_2^{(0)},\ldots,Y_J^{(0)}]$, draw $\tilde{Y}_2$ from $[\tilde{Y}_2|\tilde{Y}_1,Y_3^{(0)},\ldots,Y_J^{(0)}],\ldots,$ draw $\tilde{Y}_j$ from $[\tilde{Y}_j|\tilde{Y}_1,\ldots,\tilde{Y}_{j-1},Y_{j+1}^{(0)},\ldots,Y_J^{(0)}],\ldots,$ draw $\tilde{Y}_J$ from $[\tilde{Y}_J|\tilde{Y}_1,\ldots,\tilde{Y}_{J-1}]$, where the $\tilde{Y}$'s are drawn from the correct univariate conditional normal truncated density, as described in the preliminary results preceding Proposition 1. Specifically, let $[\tilde{Y}_j|\tilde{Y}_{-j}]$ denote the conditional distribution of $\tilde{Y}_j$ conditional on the $(J-1)\times 1$ vector excluding the jth random variable. Clearly $\tilde{Y}_j|\tilde{Y}_{-j}\sim N(\mu_{j|-j},\Sigma_{j|-j})$ conditional on $a^*\le M\cdot\tilde{Y}\le b^*$, where $\mu_{j|-j}=\mu_j+\Omega_{j,-j}\cdot\Omega_{-j,-j}^{-1}\cdot(\tilde{Y}_{-j}-\mu_{-j})$, $\mu_k\equiv(X\beta)_k$, and $\Sigma_{j|-j}=\Omega_{jj}-\Omega_{j,-j}\cdot\Omega_{-j,-j}^{-1}\cdot\Omega_{-j,j}$. Then it follows that the truncated multivariate normal distribution Y conditional on the compact region $a^*\le MY\le b^*$ will be an invariant of this process, since the $[\tilde{Y}_j|\tilde{Y}_1,\ldots,\tilde{Y}_{j-1},Y_{j+1}^{(0)},\ldots,Y_J^{(0)}]$ distributions are by construction the one-dimensional conditionals of that joint distribution.

PROPOSITION 2: *For compact support* $\boldsymbol{B}\equiv[\bar{a},\bar{b}]$, $-\infty<\bar{a}<\bar{b}<\infty$, *the joint density of* $(Y_1^{(n_G)},\ldots,Y_J^{(n_G)})$ *converges in* L_1 *norm to the true joint density,* $n(z-X\beta,\Omega,\bar{a},\bar{b})$ *at a geometric rate in* n_G.

PROOF: Define $p(n_G,x,y)$ for $(x,y)\in\boldsymbol{B}$ to be the density of $Y^{(n_G)}$ starting from $Y^{(0)}=x$; this is given constructively by the Gibbs updating scheme we described. Also by construction, p is continuous on $\boldsymbol{B}$, $p(1,x,y)>0$, and $p(n_G,x,y)=\int p(n_G-1,x,z)\cdot p(1,z,y)\,dz>0$ for $n_G>1$. Since, by assumption, $\boldsymbol{B}$ is compact, $p(1,x,y)$ is bounded positive on $\boldsymbol{B}$. This implies in turn that the process is uniformly recurrent, since the probability of never reaching a set A of

[22] The reader is also referred to Nummelin (1984), Tierney (1991), and Roberts and Polson (1994). Our results differ from those in the cited studies in that we establish conditions for *geometric* rates of convergence of the Gibbs sampler.

positive measure in n_G rounds is bounded above by $[1 - \gamma\mu(A)]^{n_G}$, from any starting point, where γ is the positive lower bound on $p(1, x, y)$ for $(x, y) \in B$, and $\mu(A)$ is the Lebesgue measure of A. One can verify by substitution that the truncated multivariate normal with density $n(z - X\beta, \Omega, \bar{a}, \bar{b})$ is an invariant of the Gibbs process. Then, Theorem 7.2 in Orey (1971) implies that the L_1 distance $\| p(n_G, Y^{(0)}, y) - n(z - X\beta, \Omega, \bar{a}, \bar{b})\|$ converges to 0 as $n_G \to \infty$ at a geometric rate; in other words, there exists $M > 0$ and $\lambda \in (0, 1)$ such that from any initial $Y^{(0)}$, one has $\int p(n_G, Y^{(0)}, y) - n(z - X\beta, \Omega, \bar{a}, \bar{b})\, dy \leq M\lambda^{n_G}$. *Q.E.D.*

It should be noted that this result of geometric convergence rate does not hold for a multivariate distribution with noncompact support. For example, consider the one-factor model $\epsilon_i = \alpha\zeta + u_i$, $i = 1, 2$, where ζ and u_i are independent standard normal variates, and α is a parameter. In this model, as $\alpha \to \infty$, $\text{corr}(\epsilon_1, \epsilon_2) = \alpha^2/(\alpha^2 + 1) \to 1$ and the rate of convergence of the Gibbs sampler from an initial density to the limiting density becomes slower and slower. The convergence problems would be alleviated if $\epsilon_i \in [\underline{b}, \bar{b}]$ with $\underline{b}$ and $\bar{b}$ finite. Hence, we will proceed with considering only large compact rectangles (for example, defined by the limits of computing machine representation of floating point numbers).

It should be noted also that, like simulator SRC above, simulator GRS is by construction continuous in the distributional parameters, β, Ω, $\bar{a}$, M, and $\bar{b}$. As found in Hajivassiliou et al. (1996), it is computationally tractable and the convergence rate of the Gibbs resamplings is very fast. Hence, the MSS-GRS estimator possesses desirable properties in terms of computational performance. These findings confirm our result below that consistency and asymptotic normality of the MSS estimator, based on simulator GRS using a finite number of terminal simulations, requires that the number of Gibbs resamplings used to generate each draw rise only faster than log N.

5.3. *SAR: Acceptance-Rejection Algorithms for Generating Truncated Multivariate Normal Variates*

We finally present a third simulation method, which generates draws $\tilde{Y}_r$ directly from multivariate normal distributions conditional on linear inequality regions, based on acceptance-rejection arguments. Then, a direct simulator of the score defined by $(1/R)\Sigma_r h(\tilde{Y}_r - X\beta)$ will be unbiased for any number of terminal simulations, R. We are therefore able to prove in Section 6 that the MSS estimator that uses simulator SAR will be CUAN for *any* (finite or infinite) number of simulations.

The idea for this algorithm is based on the following:

PROPOSITION 3: *In order to generate draws from a density $f(z) = c \cdot g(z) \cdot \psi(z)$, where $c > 1$, g is a convenient density, and ψ is $[0, 1]$ valued, generate Z from g and U uniform $[0, 1]$. Accept Z only if $U \leq \psi(Z)$; otherwise, continue trying with new pairs of Z and U. An accepted Z will have density $f(z)$.*

PROOF: For illustrative purposes, we give a proof from first principles. Let $\tilde{x}$ be drawn from $g(x)$ with support D, and $\tilde{u}$ from uniform $[0,1]$. Consider the c.d.f. of the truncated r.v. Y where

$$Y \equiv \begin{cases} \tilde{x} & \text{if } \tilde{u} \le \psi(\tilde{x}), \\ \text{not observed} & \text{otherwise.} \end{cases} \tag{36}$$

The random variable Y describes the distribution of an accepted draw according to the acceptance-rejection scheme of this theorem. Then,

$$\begin{aligned} F_Y(y) &= \text{prob}(\tilde{x} \le y | \tilde{x} \text{ accepted}) \\ &= \text{prob}(\tilde{x} \le y, \tilde{u} \le \psi(\tilde{x}))/\text{prob}(\tilde{u} \le \psi(\tilde{x})) \\ &= \frac{\int_{-\infty}^{y} \text{prob}(\tilde{u} \le \psi(x)) g(x)\, dx}{\int_D \text{prob}(\tilde{u} \le \psi(x)) g(x)\, dx} \\ &= \frac{\int_{-\infty}^{y} \psi(x) g(x)\, dx}{\int_D \psi(x) g(x)\, dx} \\ &= \int_{-\infty}^{y} \frac{1}{c} \cdot f(x)\, dx \Big/ \int_D \frac{1}{c} \cdot f(x)\, dx. \end{aligned} \tag{37}$$

Hence, the p.d.f. of Y is $f_Y(y) = f(y)/\int_D f(z)\, dz$ as required. Note that in this procedure the expected number of trials before the first acceptance is equal to c. *Q.E.D.*

In our case, let $f(z)$ denote the p.d.f. of the vector $Y \sim N(X\beta, \Omega)$ conditional on $D(y) \equiv \{a^* \le M \cdot Y \le b^*\}$. This density is given by

$$f(z) = \begin{cases} \dfrac{n(z - X\beta, \Omega)}{\int_{a^* \le Mz \le b^*} n(z - X\beta, \Omega)\, dz} & \text{if } z \in D(y), \\ 0 & \text{otherwise.} \end{cases} \tag{38}$$

Hence, the objective will be to devise convenient densities $g(z)$ from which to draw, satisfying $f(z) = c \cdot g(z) \cdot \psi(z)$, with implied large expected acceptance rates, $1/c$. We propose two such choices of convenient densities $g(\cdot)$:

Acceptance-Rejection Method (a): Consider the independent truncated normal density

$$Z \sim N(X\beta, \Lambda) \quad \text{conditional on} \quad D(y) \equiv \{a(y) \le Z \le b(y)\}, \tag{39}$$

where Λ is a diagonal positive definite matrix, with diagonal elements λ_j. This is a "convenient" density for simulation, with p.d.f. denoted by $g(z)$, because sequential sampling from it is straightforward using the method discussed in Section 5 to generate univariate normal truncated random variates, and be-

cause, given the independence of the elements of Z, the probability of the conditioning event $D(y)$ is also simple to calculate, since it is equal to

$$\text{(40)} \qquad \text{prob}(Z \in D(y)) = \text{prob}(a \le Z \le b) = \prod_{j=1}^{J} \{\Phi[b_j/\lambda_j] - \Phi[a_j/\lambda_j]\}.$$

Hence, the density of Z conditional on $D(y)$ is

$$\text{(41)} \qquad g(z) = n(z - X\beta, \Lambda)/\text{prob}(Z \in D(y)).$$

Choose Λ so that $\Lambda - \Omega$ is positive definite.[23] Then $\Lambda \ge \Omega$, and

$$\text{(42)} \qquad \max_z \frac{n(z - X\beta, \Omega)}{n(z - X\beta, \Lambda)} = \Lambda^{\frac{1}{2}}/\Omega^{\frac{1}{2}} \equiv \gamma \ge 1.$$

Draw a variate $\tilde{z}$ according to the $g(\tilde{z})$ density and a $\tilde{u}$ from uniform $(0, 1)$, and accept $\tilde{z}$ if and only if

$$\text{(43)} \qquad \tilde{u} \le \frac{n(\tilde{z} - X\beta, \Omega)}{n(\tilde{z} - X\beta, \Lambda) \cdot \gamma} \equiv \psi(\tilde{z}) \le 1.$$

By simple inspection, we then see that we have written $f(z)$ as $c \cdot g(z) \cdot \psi(z)$, where $g(z)$ is given in (41), $\psi(z)$ in (43), and the constant c determining the expected number of draws before the first acceptance is

$$\text{(44)} \qquad c \equiv \gamma \cdot \frac{\text{prob}(Z \in D(y))}{\text{prob}(Y \in D(y))},$$

where $\text{prob}(Z \in D(y))$ is given by (40). Hence, by Proposition 3, the accepted $\tilde{z}$'s will have density (39) as required. The acceptance rate $1/c$ can be maximized given Ω by choosing Λ suitably.

Acceptance-Rejection Method (b): We have shown that by defining a, b, and L as in Proposition 1, the density $f(z)$ in (39) can be written over its support as

$$\text{(45)} \qquad f(z) = n(z, I) \Big/ \int_{a \le Lz \le b} n(z, I)\, dz.$$

Draw a vector $\tilde{e}$ using the sequential scheme of Proposition 1, which has the (convenient) density

$$\text{(46)} \qquad g(e) = n(e, I) \Big/ \prod_{j=1}^{J} Q_j(e_{j-1})$$

with support

$$\text{(47)} \qquad \{a \le L \cdot e \le b\}.$$

[23] For example, choose Λ so that $\lambda_j^2 \ge \Sigma_{s=1}^{J} \Omega_{js}$, implying that $\Lambda - \Omega$ has a weakly dominant positive diagonal.

Consider a bound $\boldsymbol{B}$ such that $\boldsymbol{B} \geq \text{prob}(y_i^* \in D(y)) = \int_{a \leq I, z \leq b} n(z, I)\, dz$ and $\boldsymbol{B} \geq \text{prob}(a \leq L \cdot e \leq b) = \prod_{j=1}^{J} Q_j(e_{j-1})$.[24] The acceptance-rejection scheme (b) is then to compare the sequentially drawn $\tilde{e}$ to a uniform $(0, 1)$ variate $\tilde{u}$ and accept $\tilde{e}$ if and only if

$$\tilde{u} \leq \prod_{j=1}^{J} Q_j(\tilde{e}_{j-1})/B \equiv \psi(\tilde{e}) \leq 1. \tag{48}$$

Thus, we have written density $f(z)$ in (45) as $c \cdot g(z) \cdot \psi(z)$, where (41) gives $g(z)$, (48) gives $\psi(z)$, and $c \equiv B/\text{prob}(y_i^* \in D(y))$. Therefore, by Proposition 3, acceptance-rejection method (b) generates accepted $\tilde{e}$'s with density (45), which is equivalent to the desired density (38). The method will have an expected acceptance rate of $1/c = \text{prob}(y_i^* \in D(y))/B$, which is larger the closer the bound is to the true conditioning probability. This bound is tight for positively correlated elements of y_i^*, and becomes less so for negatively correlated y_j^*'s. This is confirmed by the Monte-Carlo results in Hajivassiliou (1993) and Hajivassiliou et al. (1996).

Though this method is not continuous in the parameters of the underlying distribution, the results in Hajivassiliou (1996) suggest that simulator SAR exhibits quite satisfactory performance in practice when an optimization method is used that does not require differentiability of the optimand, such as the nonlinear simplex algorithm of Nelder and Mead (1964).

Software implementing the simulators and simulation-based estimators discussed in this paper is publicly available over the Internet through the Web page http://econ.lse.ac.uk/~vassilis.

6. ASYMPTOTIC DISTRIBUTION OF MSS-SRC, MSS-GRS, AND MSS-SAR

This paper is concerned with the case of LDV models formed from a vector of exogenous variables x, a parameter θ, and a standard normal latent vector $v \in \boldsymbol{R}^m$. A finite series of hyperplanes, of the form $\{v \in \boldsymbol{R}^m | v \cdot p_k(x, \theta) = c_k(x, \theta)\}$, with p_k a normal vector of unit length, partition $\boldsymbol{R}^m$ into regions $d = 1, \ldots, M$. There may also be a linear mapping from v to a continuous vector y that depends on x, θ and d: $y = a(x, \theta, d) + B(x, \theta, d)v$. Let $D(x, \theta, d)$ denote the set of v that maps into d. Then, the score of observation i from an independently, identically distributed sample of size N can be written

$$s_i(\theta) = E_v(h(v, \theta, x_i, d_i, y_i) | v \in D(x_i, \theta, d_i)), \tag{49}$$

where h is a vector of polynomials in v. As explained in the previous section, to avoid technical difficulties we assume for the LDV special case (without any essential loss of empirical generality) that the multivariate normal distribution v is truncated to a large compact rectangle. We make the regularity assumptions that the functions $p_k(x, \theta), c_k(x, \theta), a(x, \theta)$, and $B(x, \theta)$ are all continuously

[24] Such a bound can be constructed as the probability of $L \cdot z$ lying in the smallest rectangular region containing the support $\{a \leq L \cdot z \leq b\}$, where $z \sim N(0, I)$. This bound is easy to calculate given that the region defining it is rectangular and z_j is i.i.d. $N(0, 1)$.

differentiable in θ, and that these functions and their derivatives are dominated by a square-integrable function $m(x)$. The simulator $\tilde{s}_i(\theta)$ will be formed by one of the following methods, corresponding to simulators SRC, GRS, and SAR in Section 5:

1. (SRC) Simulate the numerator and denominator of

$$s_i(\theta) = \frac{E_v(h(v,\theta,x_i,d_i,y_i)\cdot \mathbf{1}(v\in D(x_i,\theta,d_i))}{E_v \mathbf{1}(v\in D(x_i,\theta,d_i))}, \tag{50}$$

employing fixed sequences of random generators v. One approach is to use an unbiased simulator with one or more draws for the numerator and an unbiased simulator with R_N independent draws that is uniformly bounded positive, with $R_N/\sqrt{N}\to\infty$, for the denominator. (For example, simulators based on (33) and (34) meet these requirements.) As already explained, an alternative approach resulting in an MSS estimator with improved statistical properties is to use the same R_N draws for the numerator and denominator again with $R_N/\sqrt{N}\to\infty$.

2. (GRS) Carry out Gibbs resampling as in subsection 5.2 for n_G rounds, employing a fixed sequence of random generators v, with $n_G/(\log N)\to\infty$. Form the simulator by averaging h over a fixed number of terminal draws, R.

3. (SAR) Average h over draws of v from its conditional distribution, where these draws are obtained by acceptance-rejection methods that employ a fixed sequence of random generators v, as in subsection 5.3.

We give some general sufficient conditions for assumption (i) of asymptotic unbiasedness and assumption (ii) of stochastic equicontinuity in Theorem 1 of Section 4. We show that these sufficient conditions are satisfied in our special case of LDV models for each of the simulation methods (1)–(3). The hypotheses of Theorem 1 other than (i) and (ii) are assumed to continue to hold in the following corollaries.

COROLLARY 1: *If the simulation process is unbiased, or if the bias in an observation is dominated by a positive function independent of θ whose expectation is of order $(1/\sqrt{N})$, then the simulation bias converges to zero. In our LDV special case, this result holds for the simulator SAR that is unbiased, and holds for simulators SRC or GRS with the stated sampling rates.*

PROOF: The result holds trivially for unbiased simulators such as simulator SAR. When the simulation bias in an observation is dominated by a function with expectation of order $o(1/\sqrt{N})$, the result follows from Markov's inequality:

$$\begin{aligned} &P\Big(\sup_{\theta\in\Theta}|B_N(\theta)|>\epsilon\Big) \\ &\quad< \frac{1}{\sqrt{N}\,\epsilon}\sum_{i=1}^{N} E_i \sup_{\theta\in\Theta}|E_i\tilde{s}_i(\theta)-s_i(\theta)| \\ &\quad= E_i\sqrt{N}\sup_{\theta\in\Theta}|E_i\tilde{s}_i(\theta)-s_i(\theta)|/\epsilon\to 0. \end{aligned}$$

We denote by E_i expectation with respect to the observed data and by $\boldsymbol{E}_i$ expectation with respect to the simulation sequence.
For simulation method SRC, one has

$$|\boldsymbol{E}_i\tilde{s}_i(\theta) - s_i(\theta)| = |\ell_{i\theta}|\cdot\left|\boldsymbol{E}_i\left(1/\tilde{\ell}_i\right) - 1/\ell_i\right|$$
$$\leq |\ell_{i\theta}|\cdot\boldsymbol{E}_i|1/\tilde{\ell}_i - 1/\ell_i| \leq |\ell_{i\theta}|\cdot\boldsymbol{E}_i|\tilde{\ell}_i - \ell_i|/\ell_i k.$$

where k is a positive lower bound on the simulator of ℓ_i.

But the dominance conditions and the assumption that the simulator in the denominator uses R_N draws implies that $E_i\boldsymbol{E}_i|\tilde{\ell}_i - \ell_i| = \mathcal{O}_p(1/\sqrt{R_N})$, and $R_N/\sqrt{N} \to \infty$ gives the result. For simulation method GRS, one has

$$|\boldsymbol{E}_i\tilde{s}_i(\theta) - s_i(\theta)|$$
$$= |E_v(h|v \in D(y_i), v \sim f^{n_{G_N}}) - E_v(h|v \in D(y_i), v \sim f)|$$
$$\leq M'\|f^{n_{G_N}} - f\| \leq Me^{-\lambda n_{G_N}},$$

where $f^{n_{G_N}}$ denotes the distribution of the Gibbs sampler after n_{G_N} rounds, f denotes the true distribution of the latent variable, M', M, and λ are positive constants, and $\|\cdot\|$ is the L_1 norm. The first inequality follows from the compactness of the support of v, the second from Proposition 2 which states that when the support is compact, $\|f^{n_{G_N}} - f\|$ converges to zero at a geometric rate. Then, taking $n_{G_N} > (\log N)/2\lambda$ yields the result. *Q.E.D.*

To obtain a sufficient condition for stochastic equicontinuity, we employ a theorem of Ossiander (1987) that extends results of Dudley (1978). Some preliminary definitions and assumptions are necessary. Let $(\boldsymbol{H}, V, \mu)$ denote a probability space, Θ a compact subset of $\boldsymbol{R}^k$, and $\xi(v, \theta)$ a measurable function on $\boldsymbol{H} \times \Theta$. Assume that ξ is dominated by a square-integrable function ν on $\boldsymbol{H}$. Assume that $E\xi(V, \theta) \equiv 0$, and let $\sigma^2 \equiv E\nu(V)^2$. Consider a sequence of nested partitions of Θ into N_j regions, for $j = 1, 2, \ldots$. Let Θ_j be a finite set containing one point from each region of partition j, and define $\theta_j(\theta)$ to be the mapping from θ to the point in Θ_j that is in the same partition region. Define

$$\delta_j = \max_{\theta \in \Theta_j}\left[E^* \sup_{\{\theta'|\theta = \theta_j(\theta')\}} |\xi(V, \theta') - \xi(V, \theta)|^2\right]^{1/2}, \tag{51}$$

where E^* denotes outer expectation. Then, δ_j is a measure of the accuracy with which ξ can be approximated above and below by region-wise constant functions. Assume $\delta_j \to 0$. Let v_i for $i = 1, 2, \ldots$ denote independent realizations of V, and form $\zeta_N(\theta) = (1/\sqrt{N})\sum_{i=1}^N \xi(v_i, \theta)$. Ossiander (1987) establishes that $\zeta_N(\theta)$ is stochastically equicontinuous, provided an integral measuring the rate at which N_j increases as δ_j falls, is finite; an upper bound on this integral is

$$\sum_{j=2}^{\infty} [\log N_j]^{1/2}(\delta_j - \delta_{j-1}) < +\infty. \tag{52}$$

We next introduce a regularity condition on simulators that is sufficient to satisfy (52). The simulator $\tilde{s}_i(\theta)$ is *probably Lipschitz* on Θ if there exists $\delta_0 > 0$ and an integrable function $m_i \geq 1$ with a finite third moment such that $|\tilde{s}_i(\theta)| \leq m_i$ and for $0 < \delta < \delta_0$ and almost all $\theta \in \Theta$, there exists a probability $Q_{i\delta}(\theta)$ satisfying $Q_{i\delta}(\theta) \leq m_i \delta$ and the condition that $|\tilde{s}_i(\theta') - \tilde{s}_i(\theta)| \leq m_i \cdot |\theta' - \theta|$ for $|\theta' - \theta| < \delta$ with probability at least $1 - Q_{i\delta}(\theta)$. This condition allows the simulator to have discontinuities, but requires that the probability of a discontinuity within a small neighborhood of most θ be small, and that the simulator be moderately smooth except at discontinuities. A continuously differentiable simulator will clearly satisfy the condition.

COROLLARY 2: *Assume that the simulator $\tilde{s}_i(\theta)$ is probably Lipschitz on Θ. Then, the simulation residual process is stochastically equicontinuous. In the LDV special case, simulators SRC and GRS are continuously differentiable, while simulator SAR is probably Lipschitz, so that stochastic equicontinuity holds for all of the simulators.*

PROOF: Without loss of generality, assume $\Theta \subseteq [0,1]^K$. For any integer j, partition this cube into 2^{Kj} small cubes with sides of length 2^{-j}. Let Θ_j be a set containing one point selected from each cube that intersects Θ. These points can be selected so that $Q_\delta(\theta) \leq K\delta^\gamma$ for $\theta \in \Theta_j$. Define $\theta_j(\theta)$ to be the mapping from θ into the point in Θ_j that is in the same region of the partition; then $|\theta - \theta_j(\theta)| \leq 2^{-j} \equiv \beta_j < 1$.

Define the function

$$B_{ij}(\theta) = \begin{cases} m_i \beta_j & \text{if } \tilde{s}_i \text{ is Lipschitz on the cube containing } \theta_j(\theta), \\ 2m_i & \text{otherwise,} \end{cases}$$

and note that this function is regionwise constant on partition j. Using the Lipschitz hypothesis, one has $|\tilde{s}_i(\theta) - \tilde{s}_i(\theta_j(\theta))| \leq B_{ij}(\theta)$. Also, for j large enough so that $\beta_j < \delta_0$,

$$\begin{aligned} EB_{ij}(\theta)^2 &\leq E_i\left\{\left(1 - Q_{\beta_j}(\theta_j(\theta))\right)m_i^2\beta_j^2 + Q_{\beta_j}(\theta_j(\theta))2m_i^2\right\} \\ &\leq E_i\left\{m_i^2\beta_j^2 + 2m_i^3\beta_j\right\} \leq 3\beta_j E_i m_i^3 \equiv \delta_j^2. \end{aligned}$$

Define $\delta_j^2 = 2E_i m_i^3$ for $\beta_j \geq \delta_0$. Then,

$$\begin{aligned} &\sum_{j=2}^{\infty} [\log N_j]^{1/2}(\delta_{j-1} - \delta_j) \\ &\quad = \sum_{j=2}^{\infty} [Kj \log 2]^{1/2}(2^{-(j-1)} - 2^{-j}) \cdot 3E_i m_i^3 < +\infty. \end{aligned}$$

Then, the condition for the Ossiander result holds, and stochastic equicontinuity follows.

Consider the special case of LDV models. The simulators SRC and GRS are continuously differentiable on Θ, so they are Lipschitz with probability one, and the result follows.

Now consider simulator SAR. Given a fixed sequence of random generators v_r for $r = 1, 2, \ldots$, the acceptance-rejection procedure can be described as one in which trials are rejected until the criterion $v_r \in D(x_i, \theta, d_i)$ is met, then $\tilde{s}_i(\theta) = h(v_r, \theta, x_i, d_i, y_i)$ for the accepted v_r. Given $\theta \in \Theta$ and $\delta > 0$, let $N_\delta(\theta)$ denote a δ-neighborhood of θ. Let $R_\delta(x_i, \theta, d_i)$ denote the probability that a trial will lead to rejection for all $\theta' \in N_\delta(\theta)$, equal to the integral of the truncated standard normal density over the intersection of $D(x_i, \theta', d_i)^c$ for θ' in the neighborhood. Let $A_\delta(x_i, \theta, d_i)$ denote the probability that a trial will lead to acceptance for all $\theta' \in N_\delta(\theta)$, equal to the integral of the truncated standard normal density over the intersection of $D(x_i, \theta', d_i)$ for θ' in the neighborhood. The probability of acceptance on the same trial for all $\theta' \in N_\delta(\theta)$ is then $A_\delta(x_i, \theta, d_i)/(1 - R_\delta(x_i, \theta, d_i))$.

Suppose that $p_k(x, \theta) \cdot v \le c_k(x, \theta)$ for $k = 1, \ldots, K$ defines the set $D(x, \theta, d)$. The compactness of the support of v, the continuous differentiability of p_k and c in θ, and the dominance assumption, imply by Taylor's expansions that

$$|p_k(x, \theta') \cdot v - p_k(x, \theta) \cdot v| \le m(x) \cdot |\theta' - \theta| \le m(x) \cdot \delta,$$

$$|c(x, \theta') - c(x, \theta)| \le m(x) \cdot |\theta' - \theta| \le m(x) \cdot d.$$

Then,

$$\begin{aligned} A_\delta(x, \theta, d) &= P\left(\left\{v \middle| \sup_{\theta' \in N_\delta(\theta)} \delta^{(\theta)}(p_k(x, \theta') \cdot v - c_k(x, \theta')) \le 0, k = 1, \ldots, K\right\}\right) \\ &\ge P(\{v | p_k(x, \theta) \cdot v - c_k(x, \theta) \le -2m(x) \cdot \delta, k = 1, \ldots, K\}). \end{aligned}$$

Similarly,

$$\begin{aligned} R_\delta(x, \theta, d) &= P\left(\left\{v \middle| \inf_{\theta' \in N_\delta(\theta)} \delta^{(\theta)}(p_k(x, \theta') \cdot v - c_k(x, \theta')) > 0, k = 1, \ldots, K\right\}\right) \\ &\ge P(\{v | p_k(x, \theta) \cdot v - c_k(x, \theta) > 2m(x) \cdot \delta, k = 1, \ldots, K\}), \end{aligned}$$

and

$$R_\delta(x, \theta, d) \le P(\{v | p_k(x, \theta) \cdot v - c_k(x, \theta) > 0, k = 1, \ldots, K\}).$$

Then, the probability that the simulator has a discontinuity in $N_\delta(\theta)$ satisfies

$$\begin{aligned} Q_{i\delta}(\theta) &= 1 - A_\delta(x_i, \theta, d_i)/(1 - R_\delta(x_i, \theta, d_i)) \\ &\le \frac{P(\{v | p_k(x, \theta) \cdot v - c_k(x, \theta)| \le 2m(x) \cdot \delta, k = 1, \ldots, K\})}{P(\{v | p_k(x, \theta) \cdot v - c_k(x, \theta) \le 0, k = 1, \ldots, K\})}. \end{aligned}$$

But $p_k(x,\theta)\cdot v$ is standard normal, implying

$$Q_{i\delta}(\theta) \leq \frac{\sum_{k=1}^{K}[\Phi(c_k(x_i,\theta)+2m(x_i)\delta)-\Phi(c_k(x_i,\theta)-2m(x_i)\cdot\delta)]}{\prod_{k=1}^{K}\Phi(c_k(x_i,\theta))}.$$

The denominator of this ratio is bounded positive, and the numerator is bounded by $4Km(x_i)\delta$. This fact, together with the observation that the simulator is continuously differentiable, with a dominated derivative, when it does not have a discontinuity on $N_\delta(\theta)$, establishes that the simulator is probably Lipschitz. This argument is unchanged if the direction of some of the inequalities defining the sets $D(x_i,\theta,d_i)$ is reversed. Therefore, the corollary is proved for all cases of simulator SAR. *Q.E.D.*

The performance in practice of the three MSS estimators discussed here is compared through extensive Monte-Carlo experiments in Hajivassiliou (1996). In addition, Hajivassiliou (1997) develops (i) tests of the adequacy of particular values of R for MSS-SRC and of R and n_G for MSS-GRS in relation to the sample size N; and (ii) diagnostics for model specification in LDV models based on MSS estimators.

7. AN ILLUSTRATIVE EXAMPLE OF ALTERNATIVE SIMULATION ESTIMATION METHODS

To illustrate the method of simulated scores and contrast it to other simulation estimation methods that have been proposed in the literature, consider the simple binary probit model for an independent cross-section of individuals, $i=1,\ldots,N$, for which classical estimation is, of course, computationally very straightforward:

$$\text{(53)}\qquad \begin{aligned} &y_i^* = x_i'\beta+\epsilon_i, \quad \epsilon_i \sim N(0,1),\\ &y_i = 1,\ d_i = 1, \quad \text{if } y_i^* > 0 \quad (y_i = 2d_i - 1),\\ &y_i = -1,\ d_i = 0, \quad \text{if } y_i^* \le 0. \end{aligned}$$

Define

$$\text{(54)}\qquad \ln \ell_i = \ln \Phi(y_i\cdot x_i'\beta),$$

$$\text{(55)}\qquad \ln \ell_i = d_i\cdot \ln\Phi(x_i'\beta)+(1-d_i)\cdot\ln(1-\Phi(x_i'\beta)),$$

and

$$\text{(56)}\qquad s_i = \ell_{i\theta}/\ell_i = x_i\cdot\frac{\phi(y_i\cdot x_i'\beta)}{\Phi(y_i\cdot x_i'\beta)}\cdot y_i = x_i\cdot E(\epsilon_i | y_i^* \in D(y_i)),$$

$$(57)\qquad s_i = x_i \cdot \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)(1-\Phi(x_i'\beta))} \cdot [d_i - \Phi(x_i'\beta)]$$
$$= w_i(\theta) \cdot [d_i - \Phi(x_i'\beta)].$$

In this case, $\theta = \beta$. Then maximum likelihood estimator solves the first order conditions $L_{N\theta}(\hat{\theta}) = (1/N)\sum_{i=1}^{N} s_i(\hat{\theta}) = 0$. Equation (57) 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 (57), are used. Focusing on the conditional expectation expressions in equation (56) 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 both cases; in equation (56) it is satisfied because $P(y_i|\theta^*, x_i) = \Phi(y_i \cdot x_i'\theta^*)$ and in equation (57) because $E(d_i|\theta^*, x_i) = \Phi(x_i'\theta^*)$, where θ^* denotes the true parameter value.

The original method of simulated moments (McFadden (1989) and Pakes and Pollard (1989)) proposed substituting an unbiased simulator, $\tilde{\Phi}(x_i'\beta)$, for $\Phi(x_i'\beta)$ and exploiting the linearity of the score expression (57) in $(d_i - \Phi(\cdot))$. For high efficiency this method requires that consistent estimators for the optimal instruments, $w_i(\theta^*)$, be used. The method of simulated scores we discuss in this paper simulates instead either *directly* (using simulators SAR or GRS) the expression $E(\epsilon_i^*|y_i^* \in D(y_i))$ or *indirectly* the components of the ratio $E(\epsilon_i^*|y_i^* \in D(y_i)) = \int_D (\partial\epsilon_i^*/\partial\theta)/\int_D \epsilon_i^*$ (using simulator SRC). This fact implies that the optimal instruments are now available automatically in the form of x_i. To see the relation of MSS to MLE, recall that $x_i \cdot E(\epsilon_i^*|y_i^* \in D(y_i)) = x_i \cdot [\phi(y_i \cdot x_i'\beta)/\Phi(y_i \cdot x_i'\beta)] = s_i(y_i, \beta; x_i)$. The Lerman and Manski (1981) method uses unbiased and consistent frequency simulators of $\Phi(x_i'\beta)$ directly in the likelihood function (54);[25] van Praag and Hop (1987) use independent simulations of the numerator and denominator expressions in (56), as in (20).

Hajivassiliou (1996) contrasts the method of simulated scores to the other simulation estimation methods available in the literature via Monte-Carlo. The results there support the following conclusions: first, the choice of instrument functions in the methods that simulate generalized moment conditions can be critical, as we argued above. Employing the ideal instrument function $w(\cdot)$ in (57) (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 simulated MLE method of Lerman and Manski (1981) offers satisfactory performance only when the number of simulations employed is large, if frequency simulators are used. This number grows faster than linearly with the complexity of the LDV model under analysis. As theory suggests, the Lerman and Manski method is improved significantly by maintaining the same set of underlying random variates while iterating the optimization algorithm to convergence. Third, the method that simulated separately the denominator of

[25] A similar method has been proposed by Laroque and Salanié (1989) to tackle the numerical integration problems in multimarket disequilibrium problems.

the scores by frequency methods performed unsatisfactorily, and it was easily dominated by all the other methods tried, primarily because frequency simulators are not bounded away from 0 and 1. Before barely satisfactory performance was achieved, a huge number of simulations for the denominator expressions had to be employed. These problems were significantly alleviated once a smooth simulator, bounded away from 0, like simulator SRC of Section 5, was used for the denominator expression. Finally, in all the cases investigated, the method of simulated scores based on simulator SRC performed impressively; it approached the (optimal) performance of MLE with even ten simulations per dimension of the underlying latent variable vector. Moreover, the method was found to be numerically stable, which was to be expected given its continuity in the underlying parameters.

8. CONCLUSION

In this paper we presented the method of simulated scores (MSS), established its asymptotic properties, and developed three simulators to use for the likelihood scores. In contrast to many simulation estimation methods proposed in the literature, the MSS estimators based on simulators SRC and GRS are continuous in the unknown parameter vectors and hence standard optimization methods can be employed. Furthermore, we showed that the MSS estimator based on simulator SRC is CUAN when the number of simulations used rises as the square root of the number of observations available, while use of simulator SAR guarantees that MSS will be CUAN for a *finite* number of simulations. Finally, using simulator GRS for MSS estimation with a finite number of simulations requires instead that the number of Gibbs resamplings used for each simulation rise only as the logarithm of the number of observations. Our MSS estimator based on Gibbs sampling shows for the first time that Monte-Carlo Markov-Chain techniques can be useful for not only Bayesian but also classical inference.

We conclude that simulation estimation techniques make feasible econometric analyses of limited dependent variables models with theoretically more appropriate correlation structures.

Dept. of Economics, London School of Economics, Houghton St., London, WC2A 2AE, England; vassilis@lse.ac.uk; http://econ.lse.ac.uk/ ~ vassilis

and

Dept. of Economics, University of California at Berkeley, 787 Evans Hall, Berkeley, CA 94720, U.S.A.; mcfadden@econ.berkeley.edu; http://elsa.berkeley.edu/users/mcfadden/index.html

Manuscript received January, 1994; final revision received May, 1997.

REFERENCES

BÖRSCH-SUPAN, A., AND V. HAJIVASSILIOU (1993): "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics*, 58, 347–368.

BUTLER, J. S., AND R. MOFFIT (1982): "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," *Econometrica*, 50, 761–764.

CHIB, S., AND E. GREENBERG (1995): "Understand the Metropolis-Hastings Algorithm," *American Statistician*, 49, 327–335.

DEÁK, I. (1980): "Three Digit Accurate Multiple Normal Probabilities," *Numerische Mathematik*, 35, 369–380.

DEVROYE, L. (1986): *Non-Uniform Random Variate Generation*. New York: Springer.

DUDLEY, R. (1978): "Central Limit Theorems for Empirical Measures," *Annals of Probability*, 6, 899–929 ("Correction," *Annals of Probability*, 7 (1979), 909–911).

FUTIA, F. (1982): "Invariant Distributions and the Limiting Behavior of Markovian Economic Models," *Econometrica*, 50, 377–408.

GELFAND, A., AND A. F. M. SMITH (1990): "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

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. (1989): "Efficient Simulation from the Multivariate Normal Distribution Subject to Linear Inequality Constraints and the Evaluation of Constraint Probabilities," Duke University Working Paper.

HAJIVASSILIOU, V. (1984): "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model: A Comment," MIT Mimeo.

——— (1993): "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: The Effects of Vectorization," *International Journal of Supercomputer Applications*, 7, 231–253.

——— (1994): "A Simulation Estimation Analysis of External Repayments Problems of Developing Countries," *Journal of Applied Econometrics*, 9, 109–132.

——— (1996): "A Monte-Carlo Comparison of Leading Simulation Estimators for LDV Models," Mimeo, Department of Economics, London School of Economics.

——— (1997): *Simulation-Based Inference and Diagnostic Tests: Some Practical Issues*. Cambridge: Cambridge University Press.

HAJIVASSILIOU, V., AND P. RUUD (1994): "Classical Estimation Methods Using Simulation," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle and D. McFadden. Amsterdam: North-Holland, pp. 2383–2441.

HAJIVASSILIOU, V., D. MCFADDEN, AND P. RUUD (1996): "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results," *Journal of Econometrics*, 72, 85–134.

HECKMAN, J. (1981): "Dynamic Discrete Models," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden. Cambridge, MA: M.I.T. Press.

KEANE, M. (1994): "A Computationally Efficient Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116.

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.

LERMAN, S., AND C. MANSKI (1981): "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden. Cambridge, MA: MIT Press, pp. 305–319.

MCCULLOCH, R., AND P. ROSSI (1994): "An Exact Likelihood Analysis of the Multinomial Probit Model," *Journal of Econometrics*, 64, 207–240.

MCFADDEN, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York: Academic Press, pp. 105–142.

——— (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

MORAN, P. A. P. (1984): "The Monte Carlo Evaluation of Orthant Probabilities for Multivariate Normal Distributions," *Australian Journal of Statistics*, 26, 39–44.

——— (1985): "Calculation of Multivariate Normal Probabilities—Another Special Case," *Australian Journal of Statistics*, 27, 60–67.

——— (1986): "Orthant Probabilities and Gaussian Markov Processes," *Journal of Applied Probability*, 43, 413–417.

NELDER, J. A., AND R. MEAD (1964): "A Simplex Method for Function Minimization," *Computer Journal*, 7, 308–313.

NUMMELIN, E. (1984): *General Irreducible Markov Chains and Non-negative Operators*. Cambridge: Cambridge University Press.

OREY, S. (1971): *Limit Theorems for Markov Chain Transition Probabilities*. London: van Nostrand.

OSSIANDER, M. (1987): "A Central Limit Theorem under Metric Entropy with Bracketing," *Annals of Probability*, 15, 897–919.

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057.

PRESS, W., B. FLANNERY, S. TEUKOLSKY, AND W. VETTERLING (1986): *Numerical Recipes*. Cambridge: Cambridge University Press.

ROBERTS, G. O., AND N. G. POLSON (1994): "On the Geometric Convergence of the Gibbs Sampler," *Journal of the Royal Statistical Society, Series B*, 56, 171–182.

RUUD, P. A. (1986): "On the Method of Simulated Moments for the Estimation of Limited Dependent Variable Models," University of California at Berkeley, mimeo.

——— (1991): "Extensions of Estimation Methods Using the EM Algorithm," *Journal of Econometrics*, 49, 305–341.

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

TIERNEY, L. (1991): "Markov Chains for Exploring Posterior Distributions," Technical Report, School of Statistics, University of Minnesota.

VAN PRAAG, B. M. S., AND J. P. HOP (1987): "Estimation of Continuous Models on the Basis of Set-Valued Observations," Erasmus University Working Paper.