COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 967

THE METHOD OF SIMULATED SCORES FOR THE ESTIMATION OF LDV MODELS WITH AN APPLICATION TO EXTERNAL DEBT CRISES

Vassilis A. Hajivassiliou and Daniel L. McFadden

December 1990

The Method of Simulated Scores for the Estimation of LDV Models With An Application to External Debt Crises

by

Vassilis A. Hajivassiliou
Cowles Foundation for Research in Economics, Yale University

and

Daniel L. McFadden
Department of Economics, M.I.T.

Revised December 1990

Abstract

The method of simulated scores (MSS) is presented for estimating LDV models 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. 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. The estimation method is applied to analyze a model in which the incidence and the extent of debt repayments problems of LDC's are viewed as optimized choices of the central authorities of the countries in a framework of credit rationing. The econometric implementation of the resulting multi-period probit and Tobit models avoids the need for high dimensional integration. Our findings show that the restrictive error structures imposed by past studies may have led to unreliable econometric results.

* We are indebted to Don Andrews, Jacques Dreze, John Geweke, John Ham, Rosa Matzkin, Michel Mouchart, Whitney Newey, Paul Ruud, and Herman van Dijk for very helpful comments and suggestions. Any remaining errors and shortcomings are of course our own responsibility. Dan Nelson, Anne Royalty, and Robert Stoddard provided expert research assistance. Hajivassiliou gratefully acknowledges financial support from the National Science Foundation through grant No. SES—88—13630.

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)).[1] A common property of the methods of simulating moment conditions (MSM) is that they yield criterion functions that are discontinuous in the unknown parameter vectors to be estimated.[2] 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 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, and establish its asymptotic properties. We show that the MSS opens up a broader class of simulation techniques compared to other simulation estimation methods. In this paper 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 can then

---

[1] These methods are in contrast to simulation estimation methods that simulate non-linear 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 (1988), Laroque and Salanie (1989).

[2] A leading exception where a smooth simulator is proposed for discrete choice probabilities is the method of Stern (1988).

prove that the MSS estimator using this simulator is consistent and asymptotically normal (CAN) 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 triangularization of the normal multivariate density; 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 CAN 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)). It then follows that for an MSS estimator based on this simulator to be CAN, the number of resamplings used for each simulation must grow with the sample size at the (much slower) rate logN.[3]

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, Lerman and Manski (1981) and Hop and van Praag (1988). 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. The development in this paper 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 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

[3] Hajivassiliou, McFadden, and Ruud (1990) discuss alternative simulators and compare their properties to the ones given here.

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.[4] Third, when simulating functions that are continuous in the parameters are employed, certain computational complexities of MSM are avoided.

We employ MSS estimation to analyze econometrically the mounting external debt repayments problems of developing countries. These problems have received much attention recently, both in academic and policy circles, and in the media. The attention is well-deserved since even crude measures of external indebtedness and of repayments difficulties are steadily rising and lie well above historical standards. In this paper we offer an econometric analysis of the incidence and extent of external debt repayments problems, and attempt to quantify the impact of various factors that theory and past empirical findings suggest are precursors to such problems. Though the main modelling approach here follows McFadden et al. (1985) and Hajivassiliou (1987, 1989a), the method of simulated scores that we develop allows us for the first time to introduce in the unobservables of our models a flexible temporal correlation structure, which could not be accommodated with traditional maximum likelihood estimation methods because of the concomitant need for high-dimensional numerical integration.

Section 2 discusses the main issues from the theoretical and empirical literature on external debt and describes the theoretical approach of credit rationing in the market for international lending. Several econometric problems with the existing literature are also discussed. In this Section, we describe the data used in this study and some issues specific to the longitudinal nature of our data set. In particular, the problems of persistent unobserved heterogeneity and state dependence are discussed, and past empirical evidence is reviewed. We then present the econometric models that we estimate and explain the intractability of maximum likelihood estimation methods for our limited dependent variable models with panel data.

---

[4] This point was also noted by Ruud (1986).

In Section 3 we describe 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)).

Section 4 gives a simple illustrative example to highlight the relations between the MSS estimation method and other simulation methods in the literature. In Section 5 we discuss the empirical implementation of debt repayments crises models and analyze our empirical findings. Section 6 concludes with a summary of our results and an evaluation of the MSS methodology in analyzing the empirical problem of debt repayment crises.

Appendix 1 gives matrix differentiation results required in Section 3. Three methods for generating draws from a conditional normal distribution are developed in Appendix 2. The first two methods make the MSS estimator continuous in the unknown parameter vectors, and are respectively based on a recursive triangularization of the covariance structure and on Gibbs resampling. The third simulator relies on acceptance-rejection methods. Appendix 3 establishes the CAN properties of the MSS estimator for each of the three simulators introduced in this paper. We also show in Appendix 3 that MSS using the acceptance-rejection simulator is CAN for a finite number of simulations; that consistency and asymptotic normality of MSS using the continuous recursive triangularization simulator requires that the number of simulations rise faster than the square root of the number of observations; and finally, we obtain the result that for consistency and asymptotic normality, when the simulator based on Gibbs resampling with on R (finite) simulations is employed for MSS estimation, the number of resamplings used for each simulation need only grow at a rate faster than logN. Data sources, definitions and descriptive statistics are relegated to Appendix 4.

2. The Economic and Econometric Issues

The mounting external debt repayment problems in the Third World are very serious. Figures 1 and 2 suggest that we may now be experiencing a substantial world debt crisis, since the external debt repayment problems in the Third World have been accelerating. Figure 1 shows that the gap between obligations and repayments that are falling into arrears is widening alarmingly; the fraction of debt-servicing obligations that are in arrears in each year exhibits explosive growth. Figure 2 presents a similarly bleak picture by showing that the proportion of countries under analysis that are experiencing a repayments problem of some type (for example, obligations in arrears, or a need to request IMF assistance or a rescheduling of repayments) exhibits the same deteriorating pattern. In this paper, we analyze and model the determinants of external debt repayment problems of the developing countries within a framework of credit rationing and use the method of simulated scores to estimate the econometric models. We claim that the specific cost charged to a country by the international bankers (in the form of a "spread" over the London interbank offer rate (LIBOR)) does not perform the key role of clearing the market for international loans. Instead the allocation of scarce credit among third world countries is fundamentally carried out through quantity offers and requests. The hypothesis that the spreads are exogenously determined is formally tested in Hajivassiliou (1987) using the approach of Hajivassiliou (1986a), and it is not rejected.[5]

A number of other studies, e.g., Eaton and Gersovitz (1980, 1981), have proceeded along disequilibrium lines and applied the standard switching regimes apparatus, which allows for the separate identification of supply and demand parameters. One of the

[5] Empirical evidence (Edwards (1984)) confirms that the spreads perform only a minor role in allocation of international credit, since they do not respond very significantly to usual indicators of creditworthiness. Theoretical reasons explaining why the interest rate cannot function as a pure price in this context are given in Hajivassiliou (1987). Although here we will not offer a full theoretical justification for the exogeneity assumption, such an assumption may be motivated by recent game theoretic work on the bargaining problem with a "shrinking pie" as time goes by (see Binmore and Herrero (1984) and Shaked and Sutton (1984) for results and references), which implies that the eventual division will tend to strongly favour the short side of the market.

problems with existing studies is that they neglect information on the classification of countries as supply constrained or demand constrained that is provided, for example, by the observation of a rescheduling. To solve this problem we examine models that use the actual incidence of repayments problems to classify regimes into constrained and unconstrained periods.

Information about debt obligations in arrears is also valuable in assessing the severity of a lending constraint. Meeting all obligations promptly so that arrears are zero implies an absence of credit rationing, as the "notional" demand for new loans by a country including loans to "roll over" debt is less than the maximal supply of loans by bankers. A second possibility is a positive level of excess demand, a situation in which the country is constrained by the maximal new loans the bankers are willing to supply and tries to fill the excess demand gap for credit by letting its debt obligations fall into arrears. A rescheduling or IMF conditionality-related programs may also be necessary, depending on whether the bankers are willing to tolerate the required arrears. In McFadden et al. (1985) and Hajivassiliou (1987, 1989a), a credit-rationing model with three regimes was introduced to combine information on arrears, which is valuable in assessing the severity of a lending constraint, and qualitative information about the incidence of repayments problems, like requests for reschedulings or involvement of the International Monetary Fund. This 3-regime model simultaneously exhibits (a) a <u>probit</u> structure, since an indicator variable identifies the first regime of no debt repayments problems from the repayments problems regimes 2 and 3; (b) a <u>tobit</u> structure, in that the observed level of arrears can be either 0 or positive; (c) a <u>switching regressions</u> aspect, as the new flow of lending to a country can be either equal to the notional demand for new funds in regime 1 or to the bankers' notional supply in regime 2; and finally (d) an <u>endogenously missing data</u> structure since, when regime 3 is observed no attempts are made to identify the level of arrears and the new funds flowing to this economy. This model may be derived from a formal model of optimization subject to credit constraints.

In this paper we offer instead limited information reduced-form models that attempt to isolate the Probit and the Tobit structures. Econometric analysis of the occurrence of external debt crises in the developing world using reduced-form models is desirable for many reasons. First, such analysis can provide a forecasting system which can be used to judge the creditworthiness of a given country in terms of future repayment problems. Second, the resolution of recent policy debates as to the desirability of alternative debt relief measures should incorporate empirical evidence. Third, the likely impact of newly developed institutions in international capital markets, as for example the recently established secondary market for external debt, can be evaluated more reliably by the use of econometric evidence on the determinants and precursors of repayment problems.[6]

Consider a sample of N countries, assumed to be random. A country i is observed over T_i periods, $t=1,\ldots,T_i$. A data array (y_i,X_i) is observed, where X_i is a $T_i \times K$ array of exogenous variables, and y_i is a $T_i \times 1$ vector of limited dependent variables. For simplicity we drop the i index. We assume y is an indirect observation on a latent vector y^* according to a many-to-one mapping $y = \tau(y^*)$, with y^* given by a linear model

(II.1) $$y^* = X\beta + \epsilon .$$

We assume the disturbance vector ϵ is multivariate normal, independent of X, with the structure

(II.2) $$\epsilon = \Gamma\eta ,$$

where Γ is a $T_i \times S_i$ parametric array of rank T_i, and η is a $S_i \times 1$ vector of independent standard normal variates. Let $\Omega = \Gamma\Gamma'$.

Define

(II.3) $$D(y) = \{y^* \mid y = \tau(y^*)\} .$$

Then the likelihood of the observation is

(II.4) $$\ell(\theta;y) = \int_{D(y)} n(y^* - X\beta,\Omega)dy^* ,$$

[6] See Hajivassiliou (1989a) for attempts in this direction.

where β,Ω are functions of a k×1 deep parameter vector θ, and

(II.5) $$n(\epsilon,\Omega) = (2\pi)^{-T/2}|\Omega|^{-1/2}\exp[-\tfrac{1}{2}\,\epsilon'\Omega^{-1}\epsilon]$$

is the multivariate normal density. The asymptotically optimal parametric Maximum Likelihood estimator (MLE) would be defined as the argument that solves the score equations $\frac{1}{N}\sum_i s_i(\theta;y_i) \equiv \frac{1}{N}\sum_i \ell_{\theta i}(\theta;y_i)/\ell_i(\theta;y_i) = 0$, where $\ell_{\theta i}(\cdot)$ are the derivative vectors of the likelihood contribution ℓ_i.

We confine our attention to two specific mappings, the binomial discrete response model

(II.6) $$y_t = \mathrm{sgn}(y_t^*) ,$$

and the Tobit or censored regression model

(II.7) $$y_t = \max(0,y_t^*) .$$

In view of (II.4) and (II.5), classical estimation by the method of maximum likelihood of either the binomial discrete response model (II.6) or the Tobit model (II.7), is computationally intractable when the number of time periods per individual, T_i, exceeds 3 or 4, the variance-covariance matrix Ω of the error vector ϵ 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 in such a way as to make the evaluation of (II.4) computationally feasible. One extreme is to assume that the errors are independent across countries and across time periods for a given country, i.e.,

(II.8) $$\Omega_{IID} = E\epsilon\epsilon' = \gamma_1^2 I_T ,$$

where γ_1^2 is a variance parameter to be estimated. Despite its computational simplicity, such an assumption is often very inappropriate for a panel set of data. This issue has been neglected in most previous work on LDC debt performance, the implicit assumption being that country-year shocks are all independently and identically distributed.

In a panel model, temporal dependence can arise in at least two ways and can be a source of serious misspecification. First, heterogeneity that persists over time appears *a priori* important since countries differ in terms of colonial history, and political, religious and financial institutions, all of which may affect a country's attitude toward borrowing and defaulting and the lenders' attitudes toward the borrowing country. Such heterogeneity, which introduces serial correlation, seems an inevitable result of modelling debt performance as a function of a small number of macroeconomic variables. Second, serial correlation may be induced by learning processes that rely on a history of past repayments crises as a good predictor of future debt crises; by the role asset accumulation plays in the problem; or by our failure to address questions about the duration (actual or anticipated) of debt crises. In the models we estimate in this paper, assuming erroneously that the error-terms are i.i.d. over time for a given country will in general yield inconsistent parameter estimates because of significant state-dependence found in such models by previous investigators (McFadden et al. (1985), Hajivassiliou (1987, 1989a)).[7]

Another commonly used assumption, which allows some temporal dependence, is the one–factor analytic structure:

$$\Omega_{RE} = \gamma_1^2 I_T + \gamma_2^2 \, i_T i_T' \,, \tag{II.9}$$

where i_T is the $T\times 1$ vector of one's, and γ_1^2, γ_2^2 are variance parameters to be estimated. This implies that the integral in (II.4) can be written as a univariate integral of a product

[7] Note that in case state dependence is not present and therefore only exogenous variables appear as regressors in our models, then under additional appropriate conditions, an erroneous imposition of this i.i.d. structure will give parameter estimates that are consistent up to scale and inefficient. See Hajivassiliou (1985, 1986b). For example, if the error–components are normally distributed and i.i.d. across individuals, ϵ will also be normal with $E\epsilon_t\epsilon_s=0$ for $|t-s|>T$, since the only serial correlation in that case arises because of the persistence of the error components over all periods of obervation for a given country. This would satisfy the weak dependence conditions of Ruud (1981) and White and Domovitz (1984) for consistency of misspecified MLE. An alternative consistent and inefficient approach in such a case would be to follow the conditional ML procedures of Andersen (1970) and Chamberlain (1980) for distributions that belong to the linear exponential family. A semiparametric alternative estimation method was given by Manski (1987). But the *a priori* important state–dependence in repayment crises models, confirmed in past work, violates the exogenous regressor assumption and makes correct modelling of the serial correlation structure important for consistency and not just for efficiency.

of cumulative normal distributions, which can be evaluated very efficiently through Gaussian quadrature methods (see Heckman (1981a), Butler and Moffit (1982), and Hajivassiliou (1984)). This assumption is made for example in Hajivassiliou (1987, 1989a).

In this paper, we consider a third model for ϵ. This is the natural generalization of (II.9) that adds an autoregressive structure:

(II.10) $\epsilon_t = \alpha + \xi_t \, , \, \xi_t = \rho\xi_{t-1} + \nu_t \quad t = 1,\ldots,T$

$\nu_t \sim N(0, \sigma_\nu^2) \, , \, \xi_0 \sim N(0, \sigma_0^2) \, , \, \sigma_0^2 = \sigma_\xi^2 = \sigma_\nu^2/(1-\rho^2)$ by stationarity.

$\alpha \sim N(0, \sigma_\alpha^2)$, α and ξ_t independent.

This one–factor plus AR(1) structure, with a variance-covariance matrix denoted by Ω_{AR1RE} , implies that (II.4) will involve a T-dimensional integral, thus rendering efficient classical estimation methods infeasible.[8] Hence, we turn to the *method of simulated scores,* which avoids the need for multidimensional integration.

[8] Computationally tractable but inefficient methods are available under special circumstances that are not satisfied in this paper — see the previous footnote.

3. MSS Estimation of LDV Panel Data with Serial Correlation

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 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 Appendix 2. Two of those techniques make MSS continuous in the unknown parameters.

Using the matrix differentiation results of Appendix 1 and dropping the subscript i for simplicity, the derivatives of the likelihood (II.4) of a typical observation with respect to the parameters β, Γ can be shown to satisfy

$$\text{(III.1)} \qquad \ell_\beta(\theta;y) \equiv \frac{\partial \ell(\theta;y)}{\partial \beta} = \ell(\theta;y)\, X'\Omega^{-1} E\{y^* - X\beta \,|\, y^* \in D(y)\}$$

$$\text{(III.2)} \qquad \ell_\Gamma(\theta;y) \equiv \frac{\partial \ell(\theta;y)}{\partial \Gamma} = -\ell(\theta;y)\, \Omega^{-1}\, [I - E\{(y^* - X\beta)(y^* - X\beta)' \,|\, y^* \in D(y)\}\, \Omega^{-1}]\, \Gamma.$$

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

$$\text{(III.3)} \qquad \ell_\theta(\theta;y) \equiv \frac{\partial \ell(\theta;y)}{\partial \theta} = \ell(\theta;y) E\{h(y^* - X\beta) \,|\, y^* \in D(y)\},$$

where $h(u)$ is a vector of terms that are linear or quadratic in $u \equiv y^* - X\beta$, and depend on X and the mapping from the deep parameters θ to β and Γ. In our case $\theta = (\beta', (\text{vec}\,\Gamma)')'$ are directly the deep parameters. Then

$$\text{(III.4)} \qquad h(u) = \begin{bmatrix} X'\Omega^{-1}u \\ -\Omega^{-1}[I - uu'\Omega^{-1}]\,\Gamma_1 \\ \vdots \\ -\Omega^{-1}[I - uu'\Omega^{-1}]\,\Gamma_S \end{bmatrix},$$

where $\Gamma_1,\ldots,\Gamma_S$ are the columns of Γ.[9]

[9] More generally, $h(u)$ will be the vector (III.4) premultiplied by the array of derivatives of (β, vec Γ) with respect to the deep parameters.

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

$$\text{(III.5)} \qquad s(\theta;y) \equiv \frac{\partial \ln \ell(\theta;y)}{\partial \theta} = \ell_\theta / \ell = E\{h(y^* - X\beta) \mid y^* \in D(y)\}\ .$$

The set D(y) in three leading cases of LDV models corresponds to a set of linear inequality constraints on the elements of the latent vector y^*, as follows[10]:

Case 1: multiperiod probit

$$\text{(III.6)} \qquad \begin{array}{ll} 0 \le y_t^* < \infty & y_t=1 \\ -\infty < y_t^* \le 0 & y_t=0\ . \end{array}$$

Case 2: multiperiod Tobit[11]

$$\text{(III.7)} \qquad \begin{array}{ll} y_t \le y_t^* \le y_t & y_t>0 \\ -\infty < y_t^* \le 0 & y_t=0\ . \end{array}$$

Case 3: multinomial probit

Consider an independent sample of N individuals, with typical individual i choosing among J alternatives with observed attributes x_j. Alternative j yields the (random) utility

$$y_j^* = x_j\beta + \epsilon_j \qquad j=1,\ldots,J\ .$$

Individual i chooses alternative k that satisfies

$$\text{(III.8)} \qquad -\infty < y_k^* < \infty\ ,\ 0 < y_k^* - y_j^* < \infty\ .$$

In this case, the linear restrictions on the elements of y^* correspond to the matrix

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

---

[10] 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:

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

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

[11] The restrictions on y^* in this case are described by $\{y \le y^* \le y$ when $y>0$, $-\infty<y^*\le 0$ when $y=0\}$ instead of the more customary way of $\{y^*=y$ when $y>0$, $-\infty<y^*\le 0$ when $y=0\}$ in order to highlight the point that in the LDV models analyzed here the set D(y) describes a set of linear inequality constraints on y^*.

In view of the assumption that the observations are i.i.d. across countries, the maximum likelihood estimator is a root of the sum of scores (III.5) over subjects, i.e.,

(III.9) $\quad \hat{\theta}_{MLE} \ \underset{\theta}{\text{solves}} \ \{ \frac{1}{N} \sum_i [\ell_{i\theta}/\ell_i] = 0 \}$.

Recall that by (III.5), at the true parameter vector θ^* , $E\{\frac{\partial \ln \ell(\theta^*)}{\partial \theta}\} = E\{\ell_\theta/\ell\} = E\{h(y^*-X\beta^*)|D(y)\} = 0$.

Consider a simulator, $\tilde{h} \equiv \tilde{h}(X_i\beta,\Omega)$, for the score function $h(\cdot)$, satisfying the set of restrictions $y_i^* \in D(y_i)$. Also consider a simulator $\bar{\tilde{h}}_R \equiv \bar{\tilde{h}}(X_i\beta,\Omega,R) \equiv \frac{1}{R} \sum_r \tilde{h}_r$, which averages R independent simulations $\tilde{h}_r$. The MSS estimator we propose here replaces hard-to-compute conditional expectation terms in the logarithmic score with simulators $\bar{\tilde{h}}_R$:[12]

(III.10) $\quad \hat{\theta}_{MSS} \ \underset{\theta}{\text{solves}} \ \{ \frac{1}{N} \sum_i \bar{\tilde{h}}_R = 0 \}$.

In case $\bar{\tilde{h}}_R$ 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 Appendix 2 as simulator (3), and is based on acceptance-rejection arguments (Devroye (1986)). In practice, we show that one obtains computationally more tractable MSS estimators by employing biased simulators that are continuous in θ. The first such simulator, simulator (1) in Appendix 2, is based on a recursive triangularization of the multivariate normal density. A second continuous method, simulator (2), is also presented in Appendix 2. This simulator employs Gibbs resampling methods, which improves the rate at which the bias vanishes. As a result, for MSS to be CAN as $N \longrightarrow \infty$, if one uses the first smooth simulator to construct the MSS estimator with R simulations, R must rise at a rate faster than $\sqrt{N}$, whereas if the simulator based on Gibbs-resampling is employed with

[12] 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.

R simulations, then n, the number of resamplings used to generate each simulation, must grow faster than the (much) slower rate logN.

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:

$\hat{\theta}_{LM}=\underset{\theta}{\operatorname{argmax}}\ \frac{1}{N}\sum_i \ln\{\frac{1}{R}\sum_i \tilde{\ell}_{ir}\}$, such that the log-likelihood contributions ℓ_i are simulated unbiasedly ($E\tilde{\ell}_{ir}=\ell_i$) and consistently with R ($\tilde{\ell}_r \xrightarrow{p} \ell$). Lerman and Manski proposed using 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 MSS estimators when used in conjunction with the three simulators developed in this paper have several additional advantages over existing simulation estimation methods. The fact that MSS relies on the idea in Ruud (1986) 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 that can be written as a set of linear inequality constrains on the underlying latent variables, the distribution of which belongs to the linear exponential class. Three illustrations were given in (III.6)–(III.8). 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 $C = \{-1, +1\}^{Ti}$, with 2^{Ti} possible patterns of choice over time. The fact that T_i is fairly large in typical applications[13], however, renders intractable simple frequency simulators for choice-probabilities in the moment conditions. Moreover, the standard MSM approach is not readily applicable to other LDV models that have both discrete and continuous features.[14]

A further considerable advantage of MSS estimators is that because they simulate directly the conditional expectation expressions that appear linearly in 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 (1989b): for satisfactory efficiency, MSM estimation requires good approximations to optimal instruments, which in general is difficult to achieve.

Let us now describe the three simulation methods we propose in Appendix 2 to use in conjunction with MSS estimators. The simulator (3) is based on Ruud's (1986) suggestion to use an unbiased simulator of the conditional expectation $E\{h(y^*-X\beta|D(y))\}$ which appears in the logarithmic score, by drawing standard normal vectors η sequentially until R values of $y^* = X\beta + \Gamma\eta \in D(y)$ are observed, where R is fixed in advance, then forming a sample average of $h(y^* - X\beta)$ for the y^* drawn that are in $D(y)$.[15] Define $\delta_{D(y)}=1$ if $y^* \in D(y)$, $=0$ otherwise. If $\ell(\theta;y) = E\delta_{D(y)}(y^*)$ is small, as should be expected in realistic cases with a large number of alternative choices or choice patterns over time,

---

[13] For most countries in our sample, the number of time periods with available data is 17.

[14] For example, in the multiperiod Tobit or censored regression model, one has $y_{tn} = \max(0, y^*_{tn})$. Define

$$I_n = I_n(y_n) = \{t | y_{tn} = 0,\ t = 1,\ldots,T\}$$
$$J_n = J_n(y_n) = \{t | y_{tn} > 0,\ t = 1,\ldots,T\}\ .$$

The likelihood for a respondent is

$$\ell(y,\theta) = \int_{y^*_I \le 0} n(y^*_I - X_I\beta\,,\, y_J - X_J\beta\,,\, \Omega)dy^*_I\ ,$$

where y^*_I is the subvector of y^* with components in I. But $n(y^*_I - X_I\beta, y_J - X_J\beta, \Omega)$ $= n(y_J - X_J\beta, \Omega_{JJ}) \cdot n(y^*_I - \mu_I, \tilde{\Omega}_{II})$, with $\mu_I \equiv E(y^*_I | y_J)$ and $\tilde{\Omega}_{II} \equiv E(y^*_I | y_J)$.
The log–likelihood for a respondent then consists of a term that has a closed form expression and a second term which is a multinomial probability that all components of y^*_I are non–positive.

[15] See Ruud (1990) for combining these ideas with the EM algorithm.

then a large simulation sample is required to obtain the simulator. Simulator (3) 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 non-uniform 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 CAN.

Simulator (1) in Appendix 2 is based on the observation that the conditional expectation expression that appears in (III.5) can be written as:

(III.11) $$E(h(y^* - X\beta) \mid D(y)) = \frac{\int h(y^* - X\beta)\,\delta_{D(y)}(y^*)\ \ n(y^* - X\beta, \Omega)dy^*}{\int \delta_{D(y)}(y^*)\ \ n(y^* - X\beta, \Omega)dy^*} = \ell_\theta / \ell.$$

Then the numerator of this expression can be simulated in an unbiased fashion. Appropriate simulators can then be used for the conditioning probability in the denominator. Hence, another MSS estimator can be defined by:

(III.12) $$\tilde{\theta}_{FMSS} = \underset{\theta}{\text{solves}} \left\{ \frac{1}{N} \sum_i \left[\sum_r \tilde{\tilde{g}}_r \Big/ \sum_r \tilde{\tilde{\ell}}_r\right] = 0 \right\},$$

$$\text{such that } E\tilde{\tilde{g}}_r = \ell_\theta,\ E\tilde{\tilde{\ell}}_r = \ell, \text{ and } \tilde{\tilde{\ell}}_r \xrightarrow{P} \ell \text{ with R.}$$

A method in the literature that works along these lines is due to van Praag and Hop (1988). Their method employs independent simulators of the numerator and denominator of (III.12). Unfortunately, such an approach suffers from two drawbacks, when it uses a frequency simulator for the denominator expression. First, since the frequency simulator is not bounded away from 0, the number of simulations used for approximating the denominator probability must be very large for satisfactory performance. Second, the approach yields discontinuous optimization problems. Both shortcomings can be overcome using the recursive triangularization simulator (1) discussed in Appendix 2, which is smooth and bounded away from 0 and 1.[16] Appendix 3 proves that the MSS estimator

---

[16] Another simulator of LDV probabilities which is smooth and bounded away from 0 and 1 is due to Stern (1988). Extensive Monte Carlo evidence in Hajivassiliou (1989b) shows that simulator (1) strictly

based on simulator (1) is CAN provided R rises faster than $\sqrt{N}$.

Finally, introducing the Markovian updating scheme known as Gibbs resampling we obtain simulator (2) which estimates the complete score function. An MSS estimator based on simulator (2) using a finite number R of terminal simulations is CAN, provided n, the number of Gibbs resamplings used for each simulation, grows faster than logN. This is a very satisfactory rate, given the smoothness and the computational simplicity of this simulator.

4. 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,...,N, for which classical estimation is of course computationally very straightforward.

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

Define

$$\text{(IV.2a)} \qquad \ln \ell_i = \ln \Phi(y_i \cdot x_i'\beta)$$

$$\text{(IV.2b)} \qquad = d_i \cdot \ln \Phi(x_i'\beta) + (1-d_i)\cdot \ln(1-\Phi(x_i'\beta))$$

and

$$\text{(IV.3a)} \qquad s_i = \ell_{i\theta}/\ell_i = x_i \cdot \frac{\phi(y_i \cdot x_i'\beta)}{\Phi(y_i \cdot x_i'\beta)} \cdot y_i = x_i \cdot E(\epsilon_i \mid y_i^* \in D(y_i))$$

$$\text{(IV.3b)} \qquad = x_i \cdot \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)(1-\Phi(x_i'\beta))} \cdot (d_i - \Phi(x_i'\beta)) = w_i(\theta) \cdot (d_i - \Phi(x_i'\beta))$$

In this case, $\theta=\beta$. Then maximum likelihood estimator solves the first order conditions

---

dominates the Stern simulator in terms of simulation MSE.

$L_{N\theta}(\hat{\theta}) = \frac{1}{N} \sum_i s_i(\hat{\theta}) = 0.$

Equation (IV.3b) 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 (IV.3b), are used. Simulating the conditional expectation expressions in equation (IV.3a) corresponds to the method of scoring. It should be noted that the basic consistency requirement that $E(s_i(y_i;\theta^*)|x_i) = 0$ is satisfied; in equation (IV.3a) it is satisfied because $P(y_i|\theta^*,x_i) = \Phi(y_i \cdot x_i'\theta^*)$ and in equation (IV.3b) because $E(d_i|\theta^*,x_i) = \Phi(x_i'\theta^*)$.

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 (IV.3b) 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, follows Ruud (1986) and simulates instead *directly* the expression $E(\epsilon_i^* \mid y_i^* \in D(y_i))$ which implies that the optimal instruments are now available automatically in the form of x_i. In other words, MSS uses simulators for the expression $E(\epsilon_i^* \mid y_i^* \in D(y_i))$, say $\tilde{E}(\epsilon_i^* \mid y_i^* \in D(y_i))$. To see the relation of MSS to MLE, recall that $x_i \cdot E(\epsilon_i^* \mid y_i^* \in D(y_i)) = x_i \cdot \frac{\phi(y_i \cdot x_i'\beta)}{\Phi(y_i \cdot x_i'\beta)} = s_i(y_i,\beta;x_i)$.

The Lerman and Manski (1981) method uses unbiased and consistent frequency simulators of $\Phi(x_i'\beta)$ directly in the likelihood function (IV.2a);[17] vanPraag and Hop (1988) use independent simulations of the numerator and denominator expressions in (IV.3a).

Hajivassiliou (1989b) contrasts the method of simulated scores to the other simulation estimation methods available in the literature via Monte-Carlo. The results in that study support the following conclusions: first, the choice of instrument functions in the methods that simulate generalized moment conditions can be critical. Employing the

---

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

ideal instrument function w(.) in (IV.3b) (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. The method that simulated separately the denominator of 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 (1) of Appendix 2, was used for the denominator expression. In all the cases investigated, the method of simulated scores based on simulator (1) performed impressively; it approached the (optimal) performance of MLE with even 2 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.

5. Empirical Implementation of Debt Crises Models

The dependent variables for the models we estimate are as follows: for the probit case, a dummy variable for a repayments problem was defined to take the value 1 if IMF support was requested (either in the form of a standby agreement of second or higher tranche or use of the IMF Extended Fund Facility), if the bankers were approached to organize a rescheduling (including Paris Club, commercial banks, and aid-consortia renegotiations), or if a country let its external debt obligations (in principal or interest repayments) fall in arrears. This information was compiled from our own country-by-country investigations and from published and unpublished IMF sources. The date of rescheduling was selected to reflect the key economic developments precipitating it. In the Tobit model, the dependent variable was the total external debt obligations of a country in arrears. Figures were obtained from confidential files at the World Bank.[18] See the Data Appendix for data sources and variable definitions.

We employed exogenous variables, already identified in the literature as possible determinants of the incidence and extent of repayments problems. See Feder and Just (1977), Feder, Just and Ross (1981), Cline (1983), McFadden et al. (1985), and Hajivassiliou (1987, 1989a).[19] Since we estimate reduced form models of excess credit demand, the signs of the coefficients are difficult to predict. We begin with factors that are important in determining the creditworthiness of a country and hence the supply of lending, such as the ratio of outstanding debt to exports. This measures the extent to which exports, the main source of foreign exchange, can cover the external indebtedness of the country.

The ratio of reserves to imports is a measure of how long an economy can finance its imports by using its stock of reserves without seeking refuge in higher levels of external

---

[18] Arrears on principal of smaller than 1 percent of disbursed debt, and interest arrears of less than 0.1 percent of debt were treated as "cosmetic" and hence set to zero.

[19] To alleviate possible endogeneity issues, we lagged all explanatory variables by one year. If significant serial correlation exists in the unobservables, this procedure will not be sufficient to overcome the endogeneity problem.

borrowing. This ratio may both indicate high creditworthiness and low demand for new loans, *ceteris paribus*, since existing stocks of reserves can be used to do such financing.

The ratio of debt service due over exports is considered as a further creditworthiness indicator, since it describes the ability of an economy to finance its yearly interest and principal obligations that are a pressing short run concern. Separation of interest from principal repayments is undertaken because the on-going "liquidity vs. solvency" controversy predicts different impact of principal and interest obligations in precipitating debt crises.[20]

Real GNP per capita reflects both aid motivations by the suppliers of new lending and the degree of financial well-being of a country. As a measure of openness of the economy we employ the ratio of the current account balance relative to GNP. A high exports/GNP ratio may be viewed as an undesirable characteristic by international bankers, because it reflects vulnerability to price shocks and to falling demand for its exported goods. On the other hand, the planners of a country with a highly open economy are more likely to be disciplined in their international financial dealings and less likely to repudiate, recognizing the severe losses from a drying-up of international credit.

Past repayments problems reflected in IMF arrangements, reschedulings or significant arrears outstanding could be strong indicators of a lack of creditworthiness. It is important to attempt to identify whether the significance of such past problems manifests learning by creditors in the face of uncertainty or whether they spuriously appear statistically significant if one fails to model satisfactorily temporal dependence in the unobservables. Alternative measures based on the number of all past problems beginning from 1971 were tried to examine whether bankers have "long memories."

The method of simulated scores was employed to estimate multiperiod binary probit

[20] According to the first view, the international capital markets are not frictionless, so that a debt crisis might be induced by a lack of liquidity to a financially sound borrower. The "solvency" view maintains that credit crises are manifestations of insolvency. Presumably, prompt receipt of interest payments from a country reflects solvency.

and Tobit models under the three correlation structures described by equations (II.8)–(II.10) respectively. The results are contrasted with maximum likelihood estimation for (II.8) and (II.9), in which cases the numerical integration problems involved are manageable. Table 1 presents probit MLE and MSS results for three different versions, each of which includes different independent variables.[21] As already explained, the signs of the coefficients are difficult to predict *a priori*, since they correspond to reduced form models of excess credit demand. Our first finding is that, *ceteris paribus*, a country is more likely to request a rescheduling of its debt obligations, let its obligations go in arrears, or ask for IMF assistance, the greater the number of similar problems it has encountered in the previous year and the higher its outstanding stock of debt relative to its exports. We also find that countries with high foreign reserves relative to imports are less likely to get into debt repayments problems. Bankers are seen to have "short memories" in the sense that the incidence of a debt problem in the immediately preceding period is a stronger predictor of similar problems in the future, compared to the cumulated number of problems in the whole observed past history. In general, we confirm results in past studies (McFadden et al. (1985), Hajivassiliou (1987, 1989a)), that there is strong evidence of persistent unobservable country heterogeneity, which is attenuated but not eliminated by the inclusion of variables measuring the occurrence of problems in the preceding year.

Proceeding to Table 2, we present similar results that measure the severity and extent of credit constraints through the use of the multiperiod Tobit model. The coefficient estimates are somewhat better determined than in the probit case, presumably confirming the high informational content of the confidential arrears variable. Note in particular the statistically strong, negative sign of the coefficient of the current account to GNP ratio. The cumulated number of past problems is now statistically significant, suggesting stronger temporal dependence in the the severity of crises.

The ratio of interest service due to exports has a significant and positive effect on

[21] All MSS results reported in the Tables employed simulator (1), which operationally seemed to offer the best compromise in terms of speed and accuracy.

the propensity to encounter debt repayments problems, while the ratio of principal service due to exports, even though generally less significant, has a negative sign. This evidence mildly favours the "solvency" hypothesis. Since the ratio of total outstanding debt relative to exports appears strongly significant, the liquidity hypothesis would predict positive coefficients for both interest and principal due, because of the implied deleterious effect of shortening average maturities. That we typically find a *negative* effect of principal due lends weak support to the "solvency" view by rejecting the liquidity view. It is interesting to note that attempting to pool interest and principal repayments due into a single debt-service–due variable is statistically very strongly rejected; debt service due appears to be insignificant.[22]

The last three columns of each table are obtained through the MSS methodology. The first two are given as a benchmark comparison to MLE, since in the first two cases, i.i.d. errors and factor–analytic correlation structure, ML estimation is computationally tractable. The results are quite reassuring since the MSS estimates are very close to the MLE ones with slightly lower accuracy, as predicted by asymptotic theory. The main novelty is the fifth column of each table, since in that case, the random-effects plus AR(1) error structure (II.10) renders infeasible estimation by classical ML methods. Note the significant positive autocorrelation in the error (significant in both models, more strongly so in the Tobit case). Moreover, some of the variables that were found statistically not

---

[22] Hajivassiliou (1989b) examines other issues that have important policy implications: For example, is an overvalued exchange rate one of the fundamental causes of external financing problems, or is overvaluation a very costly distortion that *arises from* very high levels of external indebtedness? Measures of overvaluation were constructed to investigate this issue, based on a discrepancy between official and black market exchange rates. Another issue analyzed in that study is the importance of world economic factors, which are exogenous to a developing country, in explaining the occurrence of external debt repayments problems. Such factors include the volume of import demand by industrialized countries, inflation in the OECD nations, and world interest rates. The findings have important policy implications on LDC "adjustment efforts" to stave off external financing crises. Finally, the models were investigated for the possibility of structural breaks occurring in the processes determining repayment problems over time. One popular view attributes part of the blame for the LDC repayments problems to the glut of "petrodollars" after the first 1973 major oil-shock. Only very weak evidence that such structural breaks occurred in the estimated relationships was found. Another possible structural break occurred after the much greater institutional involvement that followed the onset of defaults beginning in 1982. Some evidence that the probabilities of repayments problems worsened after 1981 was found.

significant with the more restrictive correlated structures (e.g., CA/GNP), now become very important. The statistical significance of observed history of problems is in most cases considerably reduced, whereas the random effects hardly loose their significance. This suggests that the persistence arises more from the unobservables of the model (e.g., the magnitude of excess demand for credit in the previous period) rather than the observed incidence of a repayments problem. Note that these estimates suffer from the long-standing problem of initial conditions in dynamic limited dependent variable models (see Heckman (1981b)). The reported results were obtained using Heckman's approximate solution, which assumes the same functional form for the distribution of the initial condition. Given that our date set is a "long" panel with approximately 14 years of observation per country, we do not expect the treatment of the initial conditions to make a substantial difference in estimation.[23]

6. Conclusion

In this paper we presented the method of simulated scores (MSS) and developed three simulators to use for the likelihood scores. In contrast to most simulation estimation methods proposed in the literature, the MSS estimators based on simulators (1) and (2) 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 (1) is CAN when the number of simulations used rises at the square root of the number of observations available. Using simulator (2) 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. Use of simulator (3) guarantees

---

[23] In a separate line of research, we are currently analysing this issue and show that it can be addressed by adapting flexible functional form and semiparametric estimation methods, recently developed by Gallant and Nychka (1987). Simple tests of the adequacy of the distributional assumption for the initial condition can be devised by employing non-parametric density estimation methods.

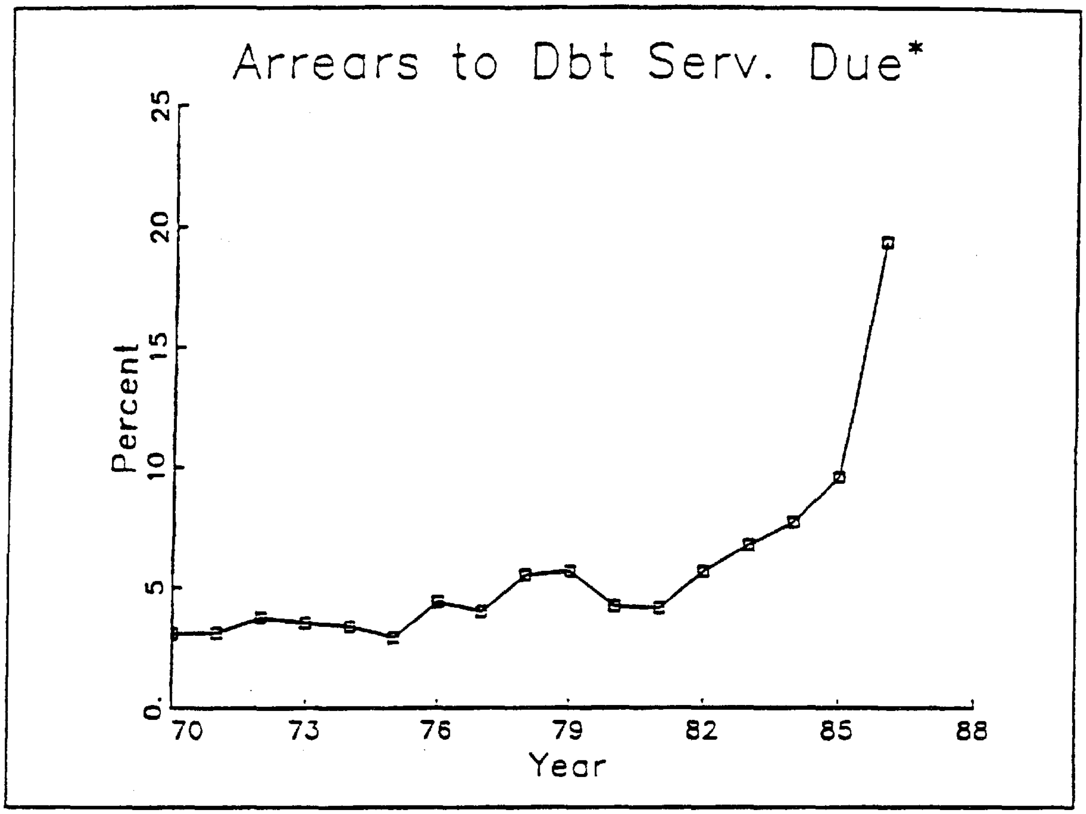
that MSS will be CAN for a *finite* number of simulations.

We offered an econometric analysis of the incidence and extent of external debt repayments problems using MSS estimation. This allowed us for the first time to introduce a flexible correlation structure over time in the unobservables of our LDV models, which can not be handled by traditional maximum likelihood estimation methods because of high-dimensional integrals.

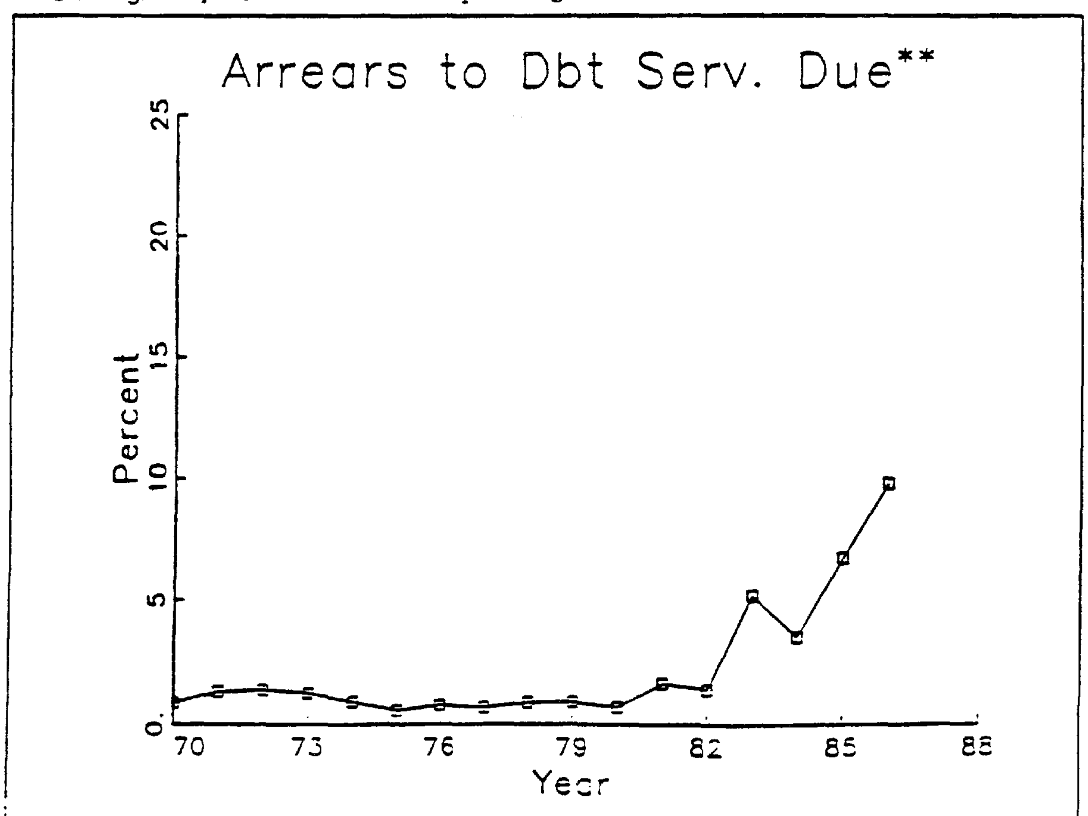
The main theoretical approach adopted was one of credit rationing in the market for international lending. Several econometric problems with the existing literature were discussed, in particular the issues of persistent unobserved heterogeneity and state dependence which are specific to the longitudinal nature of our data set. This analysis attempted to quantify the impact of various factors that because of theoretical arguments and past empirical findings are believed to act as precursors to such problems. We have shown that the restrictive correlation structures imposed by past studies, necessary to render feasible the classical method of MLE, was giving unreliable econometric results.

We conclude that the simulation estimation techniques are likely to prove very useful in carrying out econometric analyses of limited dependent variables models with theoretically more appropriate correlation structures.

Figure 1



* Unweighted; 93 countries reporting



** Weighted by Total Outstanding Debt; 93 countries reporting

Figure 2

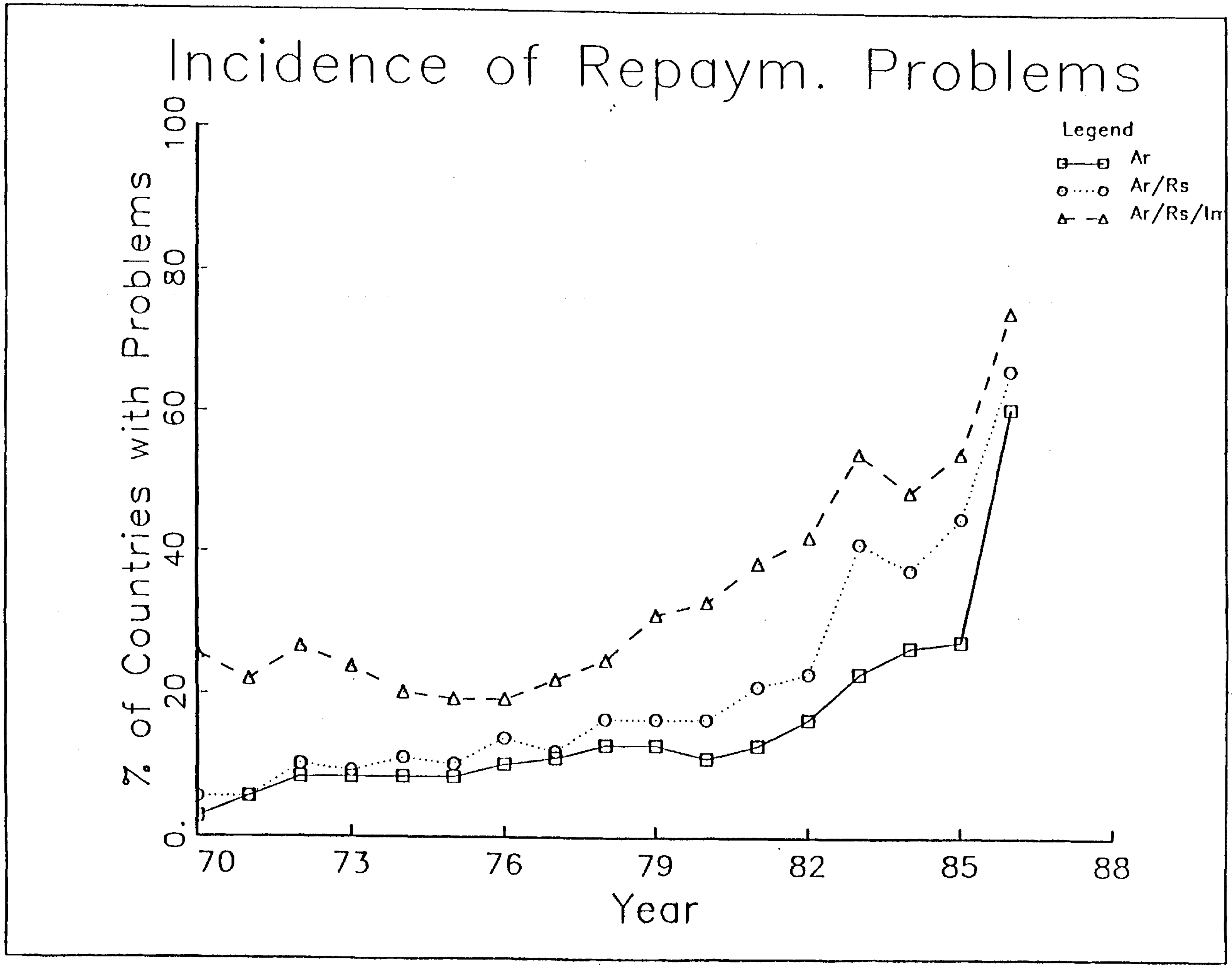

Incidence of Repaym. Problems
% of Countries with Problems
0.
20
40
60
80
100
70
73
76
79
82
85
88
Year
Legend
Ar
Ar/Rs
Ar/Rs/In

Table 1
Probit MLE and MSS Results
Dependent variable: PARIF (1 if IMF involvement, rescheduling, or sign. arrears).
Asymptotic t-statistics in brackets.
1338 available observations (91 countries, observed for the 17 years 1971–1987)

Method	MLE	MLE	MSS	MSS	MSS
	Probit	Probit with RE	Probit	Probit with RE	Probit with RE and AR1
	(Ω_{IID})	(Ω_{RE})	(Ω_{IID})	(Ω_{RE})	(Ω_{AR1RE})
constant	–1.16	–1.23	–1.34	–1.35	–2.31
	[–10.17]	[–11.27]	[–10.01]	[–11.18]	[–10.27]
Debt to	1.46e–3	1.59e–3	1.49e–3	1.53e–3	1.37e–3
Exports	[2.76]	[2.33]	[2.87]	[2.47]	[2.95]
Reserves	–7.92e–2	–7.81e–2	–7.63e–2	–7.45e–2	–3.21e–2
to Imports	[–4.21]	[–3.88]	[–4.43]	[–3.25]	[–5.75]
Interest	6.94e–2	5.77e–2	6.47e–2	5.28e–2	3.57e–2
Service Due	[4.43]	[4.99]	[4.71]	[4.38]	[3.21]
Principal	–2.37e–2	–2.52e–2	–2.37e–2	–2.18e–2	–3.44e–2
Service Due	[–2.35]	[–2.59]	[–2.35]	[–2.26]	[–3.92]
Per Capita	–5.39e–4	–5.57e–4	–4.38e–4	–4.88e–4	–1.22e–4
GDP (80 $)	[–1.45e–2]	[–1.27e–2]	[–1.31e–2]	[–1.27e–2]	[–1.57e–3]
Curr. Acc.	4.40e–2	4.56e–2	3.84e–2	3.82e–2	7.68e–2
to GNP	[0.11]	[0.29]	[0.22]	[0.32]	[2.59]
Past Signif.	1.91	1.95	1.78	1.93	2.01
Int Arrears	[7.02]	[7.59]	[6.22]	[7.38]	[6.47]
Past Signif.	1.59	1.21	1.47	1.32	2.14
Pr Arrears	[5.34]	[5.27]	[4.78]	[5.13]	[3.53]
Past Resch.	1.45	1.01	1.36	1.21	0.67
or IMF inv.	[12.05]	[10.11]	[11.64]	[9.57]	[8.78]
Cumulated	3.00e–2	2.77e–2	3.27e–2	2.46e–2	3.61e–2
Signf I Arr	[0.62]	[0.53]	[0.48]	[0.32]	[0.46]
Cumulated	6.12e–2	6.55e–2	6.26e–2	6.51e–2	3.29e–2
Signf P Arr	[0.79]	[0.88]	[0.69]	[0.38]	[0.32]
Cumulated	4.27e–2	4.01e–2	4.17e–2	4.35e–2	2.98e–2
Resc or IMF	[1.73]	[1.12]	[1.47]	[1.47]	[1.59]
γ_2 (RE)	—	0.32	—	0.33	0.22
		[3.26]		[2.98]	[2.03]
ρ	—	—	—		.203
					[2.47]
loglikelihood					
at optimum	–498.54	–497.61	—	—	—
constrained	–927.43	–927.43	—	—	—
Pseudo R^2	0.463	0.469	0.463	0.469	0.496
% corr. prd	84.60	84.70	84.60	84.70	89.42

Notes:
1. γ_1 = st. dev. of i.i.d. error (=1), γ_2 = st. dev. of random effect, ρ = AR1 coefficient
2. Simulator (1) with 20 simulations was used for the MSS estimations

Table 2
Tobit MLE Results
Dependent variable: SARF (Significant Arrears deflated by exports).
Asymptotic t–statistics in brackets.
1338 available observations (91 countries, observed for the 17 years 1971–1987)

Method	MLE	MLE	MSS	MSS	MSS
	Tobit (Ω_{IID})	Tobit with RE (Ω_{RE})	Tobit (Ω_{IID})	Tobit with RE (Ω_{RE})	Tobit with RE and AR1 (Ω_{AR1RE})
constant	–2.40e–2 [–11.99]	–2.53e–2 [–8.08]	–2.24e–2 [–10.35]	–2.36e–2 [–7.42]	–3.21e–2 [–6.83]
Debt to Exports	6.65e–6 [1.56]	8.53e–6 [1.73]	6.36e–6 [1.74]	8.15e–6 [1.27]	4.53e–4 [2.49]
Reserves to Imports	–3.35e–4 [–1.64]	–2.99e–4 [–1.35]	–3.18e–4 [–1.83]	–2.64e–4 [–1.39]	–2.43e–4 [–2.92]
Interest Service Due	2.05e–4 [1.56]	2.53e–4 [1.61]	2.27e–4 [1.93]	2.36e–4 [1.92]	2.42e–4 [2.31]
Principal Service Due	–2.90e–5 [–0.26]	–1.80e–5 [–0.16]	–2.53e–5 [–0.35]	–1.64e–5 [–0.28]	–2.04e–5 [–1.83]
Per Capita GDP (80 $)	8.01e–4 [1.64]	9.70e–4 [1.69]	8.26e–4 [1.42]	9.85e–4 [1.93]	7.24e–4 [1.44]
Curr. Acc. to GNP	–1.46e–2 [–2.86]	–1.39e–2 [–2.03]	–1.18e–2 [–2.36]	–1.49e–2 [–2.49]	–2.04e–2 [–3.25]
Past Signif. Int Arrears	1.04e–2 [5.41]	8.98e–3 [5.12]	1.23e–2 [5.27]	8.53e–3 [5.43]	7.42e–3 [4.21]
Past Signif. Pr Arrears	8.82e–3 [4.02]	8.24e–3 [3.83]	8.73e–3 [4.37]	8.34e–3 [3.48]	5.82e–3 [2.47]
Past Resch. or IMF inv.	1.76e–3 [1.25]	1.24e–3 [0.81]	1.17e–3 [1.58]	1.26e–3 [0.36]	1.69e–3 [1.81]
Cumulated Signf I Arr	1.06e–3 [3.25]	9.33e–4 [2.89]	1.24e–3 [3.37]	9.39e–4 [2.47]	6.28e–4 [2.37]
Cumulated Signf P Arr	–9.57e–4 [–1.54]	–1.27e–3 [–2.03]	–9.26e–4 [–1.11]	–1.53e–3 [–2.39]	–0.25e–3 [–2.04]
Cumulated Resc or IMF	9.93e–4 [3.81]	1.31e–3 [4.16]	9.27e–4 [3.23]	1.43e–3 [4.27]	2.63e–3 [3.45]
Signif Past I Arrears	1.68 [3.14]	1.82 [3.13]	1.35 [3.56]	1.59 [3.45]	1.52 [2.63]
Signif Past P Arrears	0.84 [8.80]	0.80 [8.32]	0.81 [7.31]	0.67 [8.32]	0.47 [6.92]
γ_1	1.26e–2 [93.00]	1.20e–2 [14.96]	1.18e–2 [95.21]	1.12e–2 [13.42]	1.01e–2 [7.42]
γ_2	—	3.96e–3 [4.37]	—	3.54e–3 [4.74]	2.36e–3 [3.73]
ρ	—	—	—		0.289 [2.93]
loglikelihood:					
at optimum	554.880	556.399	—	—	—
constrained	227.560	227.560	—	—	—

Notes:
1. γ_1 = st. dev. of i.i.d. error, γ_2 = st. dev. of random effect, ρ = AR1 coefficient
2. Simulator (1) with 20 simulations was used for the MSS estimations

Appendix 1

Matrix Differentiation Results

For the purposes of Section 3, we give the following formulas for matrix differentiation. If A is a n×n square matrix, not necessarily symmetric, with determinant $|A|$, then[24]

(A1.1) $$\frac{\partial \ln |A|}{\partial A} = (A')^{-1}$$

(A1.2) $$\frac{\partial A^{-1}}{\partial A} = A^{-1} \otimes A^{-1}$$

(A1.3) $$\frac{\partial (z' A^{-1} z)}{\partial A} = -A^{-1} zz' A'^{-1} .$$

Using these formulas, the derivatives of a multivariate normal density $n(z - \mu, \Omega)$, with variance-covariance matrix $\Omega = \Gamma\Gamma'$, are

(A1.4) $$\frac{\partial \ln\ n(z - \mu,\ \Omega)}{\partial \mu} = \Omega^{-1}(z - \mu)$$

(A1.5) $$\frac{\partial \ln\ n(z - \mu,\ \Omega)}{\partial \Omega} = -\frac{1}{2}\,\Omega^{-1} + \frac{1}{2}\,\Omega^{-1}\,(z - \mu)(z - \mu)'\,\Omega^{-1} .$$

Also,

(A1.6) $$\frac{\partial (z' \Omega^{-1} z)}{\partial \Gamma} = -2\ \Omega^{-1} zz' \Omega^{-1} \Gamma$$

and

(A1.7) $$\frac{\partial \ln |\Omega|}{\partial \Gamma} = 2\ \Omega^{-1}\Gamma ,$$

implying

(A1.8) $$\frac{\partial \ln\ n(z - \mu, \Omega)}{\partial \Gamma} = -\Omega^{-1}\Gamma + \Omega^{-1}(z - \mu)(z - \mu)'\Omega^{-1}\Gamma .$$

---

[24] Note that we use $\frac{\partial g(A)}{\partial A}$, with typical element $\frac{\partial g(A)}{\partial a_{ij}}$, to denote the partial derivatives of a function $g(\cdot)$ with respect to element a_{ij}, *without regard to possible symmetry of matrix A*. In other words, for the purposes of differentiation, we let $\frac{\partial g(A)}{\partial A}$ denote the partial derivative of g(A), holding constant all other elements of A, even though for a symmetric A we have, of course, $a_{ji}=a_{ij}$. This is to prevent double-counting when we apply the matrix chain rule of differentiation in (A1.1–8) below.

Appendix 2

Methods for Generating Draws from Conditional Normal Distributions

In this Appendix 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 (1) requires the number of simulations employed to grow without bound, while simulator (2), 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 is an unbiased estimator for the score for a finite number of simulations, and is a discontinuous function of the unknown parameters. The consistency and asymptotic normality of MSS estimators based on any of these three simulation methods is established in Appendix 3.

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

$$\text{(A2.1)} \qquad y_i = \tau(Y_i^*)\,, \qquad Y_i^* \sim N(X_i\beta,\Omega), \qquad i=1,\ldots,N.$$

The MSS estimator requires simulating the $h(Y_i^*)$ functions that appear in the scores, conditional on $Y_i^* \in D(y_i)$. Hence, our general objective is to obtain random draws from the distribution Y_i^* subject to $y_i = \tau(Y_i^*)$. Then we see from Section 3 that three types of functions need to be simulated. The first function is the likelihood contribution ℓ_i :

$$\text{(A2.2)} \qquad \ell_i \equiv \text{Prob}(D(y_i)) = \int_{D(y_i)} n(z-X_i\beta,\Omega)\,dz\,.$$

The second is the likelihood derivative $\ell_{i\theta}$:

$$\text{(A2.3)} \qquad \ell_{i\theta} \equiv \int_{D(y_i)} h(z,X_i,\beta,\Omega)\,n(z-X_i\beta,\Omega)\,dz\,.$$

Finally, the third function is the logarithmic score $\frac{\partial \ln \ell_i}{\partial \theta}$:

$$s_i \equiv \ell_{i\theta}/\ell_i = \int_{D(y_i)} h(z,X_i,\beta,\Omega)\,\frac{n(z-X_i\beta,\Omega)}{\ell_i}\,dz\,. \tag{A2.4}$$

For simplicity, we will drop the i index whenever no ambiguity would arise.

Our general objective will be to develop unbiased simulators for these functions, that are computationally very fast; and simulators that though only asymptotically unbiased, their bias vanishes at sufficiently fast rates as to guarantee consistency and asymptotic normality of MSS estimators that employ them.[25] The first continuous simulator is based on the idea of employing a Choleski triangularization so as to make the constraints $Y^* \in D(y)$ recursive.[26] This will make simulator (1) 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 (1988)) to ensure that the joint distribution we are simulating from converges to the appropriate multivariate truncated normal distribution. Hence, simulation method (2) will provide unbiased drawings of likelihood contributions and scores for a finite number of terminal simulations, as the number of Gibbs resamplings rises to infinity. Finally, we describe a third simulator 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 give the following preliminary result for the univariate case.

[25] Unbiased and consistent simulators for the integrals appearing in expressions (A2.2)–(A2.3) can also be obtained through importance sampling and other methods (see Moran (1984, 1985, 1986), Deak (1980), McFadden(1989), Stern (1988)). These methods cannot be used for direct unbiased simulation of the logarithmic score (A2.4), unless an infinite number of simulations is averaged.

[26] Geweke (1989) uses this triangularization in a Bayesian context and Keane (1990) employs it in the special case of estimating by simulation a multiperiod (panel-data) binary probit model.

Proposition 1:

Suppose that the random variable U *has a uniform* (0,1) *distribution, and* $\Phi(\cdot)$ *is the standard normal* N(0,1) *cumulative distribution function. Define the random variable* $Q \equiv \Phi^{-1}((\Phi(b) - \Phi(a)) \cdot U + \Phi(a))$. *Then* Q *will be distributed* N(0,1) *conditional on* $a \leq Q \leq b$.

Proof:

Since U is distributed uniformly on (0,1), $V \equiv (\Phi(b) - \Phi(a)) \cdot U + \Phi(a)$ will be distributed uniformly on $(\Phi(a),\Phi(b))$. It then follows by the probability integral transform result (see Feller (1971)) that $Q \equiv \Phi^{-1}(V)$ will be distributed as N(0,1) conditional on $a \leq Q \leq b$, since the implied c.d.f. for Q is $G(q) = \frac{\Phi(q) - \Phi(a)}{\Phi(b) - \Phi(a)}$. Note that this result yields random variates that are continuous functions of the parameters of the distribution, a and b. □

For a vector of indices (1,...,J), we use the notation "<j" to denote the subvector (1,...,j–1), "≤j" to denote the subvector (1,...,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 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. Employing this notation and Proposition 1, we can now establish the following:

Proposition 2:

Consider the $J \times 1$ *random variate vector* Y^* *distributed as* $N(X\beta,\Omega)$ *conditional on* $a^* \leq MY^* \leq b^*$, *where* $-\infty \leq a^* < +\infty$, $-\infty < b^* \leq +\infty$, $a^* < b^*$, *the matrix* M *is non-singular, and the matrix* Ω *is positive definite. Define* $a \equiv a^* - MX\beta$, $b \equiv b^* - MX\beta$, $\bar{a} \equiv M^{-1}a^*$, $\bar{b} \equiv M^{-1}b^*$, *and let* L *be the (lower-triangular) Choleski decomposition of* $S \equiv M\Omega M' \equiv LL'$. *The density of this random vector* Y^* *is*

$$(A2.5)\qquad n(y^*-X\beta,\Omega,\bar{a},\bar{b}) \equiv n(y^*-X\beta,\Omega)/P, \textit{ where } P = \int_{\bar{a}}^{\bar{b}} n(z-X\beta,\Omega)\,dz,\ \bar{a} \le y^* \le \bar{b}.$$

Then the following results hold:

(a) The conditional density of Y_j^* *from* $n(y_j^*;\ X\beta,\Omega,\bar{a},\bar{b})$ *given* $Y_{-j}^*=y_{-j}^* \in (\bar{a}_{-j},\bar{b}_{-j})$, *is univariate truncated normal, with density*

$$(A2.6)\qquad \frac{1}{\sigma_j}\phi((y_j^*-\mu_j)/\sigma_j)/[\Phi((\bar{b}_j-\mu_j)/\sigma_j)-\Phi((\bar{a}_j-\mu_j)/\sigma_j)],$$

where $\bar{a}_j \le y_j^* \le \bar{b}_j$,

$\mu_j = (X\beta)_j + \Omega_{j,-j}\cdot(\Omega_{-j,-j})^{-1}\cdot(y_{-j}^*-(X\beta)_{-j})$, *and*

$\sigma_j = [\Omega_{jj} - \Omega_{j,-j}\cdot(\Omega_{-j,-j})^{-1}\cdot\Omega_{-j,j}]^{\frac{1}{2}}$.

(b) *Draw sequentially* $e_1 \sim N(0,1)$ *s.t.* $a_1 \le l_{11}\cdot e_1 \le b_1$, $e_2 \sim N(0,1)$ *s.t.* $a_2 \le l_{21}\cdot e_1 + l_{22}\cdot e_2 \le b_2$, ... , *and* $e_J \sim N(0,1)$ *s.t.* $a_J \le l_{J1}\cdot e_1 + \dots + l_{JJ}\cdot e_J \le b_J$. *Let* $e \equiv (e_1,\dots,e_J)'$. *Then the joint probability density function of the vector e is given by*

$$(A2.7)\qquad f(e) = \prod_{j=1}^{J} \phi(e_j)/Q_j(e_1,\dots,e_{j-1}) \quad \textit{on the set } a \le L\cdot e \le b,$$
$$= 0 \quad \textit{otherwise},$$

where $Q_1 \equiv \text{Prob}(a_1/l_{11} \le e_1 \le b_1/l_{11})$,

$Q_j(e_1,\dots,e_{j-1}) \equiv \text{Prob}((a_j-L_{j,<j}\cdot e_{<j})/l_{jj} \le e_j \le (b_j-L_{j,<j}\cdot e_{<j})/l_{jj} \mid e_1,\dots,e_{j-1})$.

Define $\tilde{y} \equiv M^{-1}L\cdot e + X\beta$. *Then* $\tilde{y}$ *has a distribution conditional on* $a+MX\beta \equiv a^* \le M\tilde{y} \le b^* \equiv b+MX\beta$, *which in general is different from the distribution of* Y^* *conditional on* $a^* \le MY^* \le b^*$.

<u>Proof</u>:

(a) Write $n(y^*-X\beta,\Omega)/P = n(y_j^*-\mu_j,\sigma_j^2)\cdot n(y_{-j}^*-(X\beta)_{-j},\Omega_{-j,-j})/P$, and divide this joint probability distribution function by its integral in y_j^* over $\bar{a}_j \le y_j^* \le \bar{b}_j$ to get the result.

(b) Consider $Y^* \sim N(X\beta,\Omega)$ conditional on $a^* \le MY^* \le b^*$ and define

$\nu \equiv L^{-1}M(Y^*-X\beta)$. Then the event $a^* \leq MY^* \leq b^*$ is equivalent to $a^*-MX\beta \equiv a \leq L\cdot\nu \leq b \equiv b^*-MX\beta$. Since L is lower triangular, the implied constraints on ν are recursive,

$$\text{(A2.8)} \qquad \begin{array}{lcll} a_1/l_{11} & \leq \nu_1 \leq & b_1/l_{11} & \\ (a_j-L_{j,<j}\cdot\nu_{<j})/l_{jj} & \leq \nu_j \leq & (b_j-L_{j,<j}\cdot\nu_{<j})/l_{jj}\,, & j=2,\ldots,n\,. \end{array}$$

Though recursive, these constraints on ν are interdependent, and therefore there is no convenient way of generating ν vectors with the distribution $\nu \sim N(0,I)$ conditional on the constraints (A2.8).

Consider instead the random vector e defined by the following *sequential* procedure, satisfying constraints exactly analogous to (A2.8):

$$\text{(A2.8}'\text{)} \qquad \begin{array}{lcll} a_1/l_{11} & \leq e_1 \leq & b_1/l_{11} & \\ (a_j-L_{j,<j}\cdot e_{<j})/l_{jj} & \leq e_j \leq & (b_j-L_{j,<j}\cdot e_{<j})/l_{jj}\,, & j=2,\ldots,n\,. \end{array}$$

Draw first an $e_1 \sim N(0,1)$ satisfying constraint (A2.8′,j=1); then, *given* the value of e_1, draw $e_2 \sim N(0,1)$ satisfying constraint (A2.8′,j=2); ... ; finally, given the values of $e_1,\ldots,e_{J-1}$, draw $e_J \sim N(0,1)$ satisfying constraint (A2.8′,j=J). In other words, we draw sequentially e_1, $e_2|e_1$, $e_3|e_1,e_2$, ... , $e_J|e_1,\ldots,e_{J-1}$ satisfying (A2.8). It should then be clear that the p.d.f. of this sequentially drawn e vector is given by (A2.7).

Finally, define $\tilde{y} \equiv M^{-1}\cdot L\cdot e + X\beta$. Since the distribution of e satisfying constraints (A2.8′) *sequentially* is not the same as the distribution of ν satisfying constraints (A2.8) *jointly*, neither is the distribution of $\tilde{y}$ the same as the distribution of Y^* conditional on $a^* \leq MY^* \leq b^*$. □

It is important to reiterate that $\tilde{y}$'s implied by the sequential scheme of Proposition 2 are not distributed according to the multivariate truncate normal distribution $Y^* \sim N(X\beta,\Omega)$ conditional on $a^* \leq M\tilde{y} \leq b^*$ because the recursive constraints defined by the Choleski decomposition are not independent. The main point here is that while

$Y^* \sim N(X\beta,\Omega)$ conditional on $a^* \leq M \cdot Y^* \leq b^*$ is a linear transformation of $\nu \sim N(0,I)$ conditional on $a \leq L \cdot \nu \leq b$, the distribution of ν is not the same as the distribution of the sequentially drawn e. Therefore, neither is the distribution of $\tilde{y}$ defined on $a^* \leq M \cdot \tilde{y} \leq b^*$ the same as the distribution of Y^* defined on $a^* \leq MY^* \leq b^*$. 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 (A2.8′,j=1) will ignore the constraint in (A2.8′,j=2), hence will be too small on average. Given an e_1 too small, e_2, obeying the second constraint (A2.8′,j=2), will be too large on average.

Despite this fact, we can show that combining Proposition 2 together with importance-sampling arguments we can define smooth, unbiased, and direct simulators for the likelihood contributions ℓ_i and their derivatives $\ell_{\beta i}$, and a smooth, asymptotically unbiased simulator of the score function, termed simulator (1). The results in Appendix 3 will establish that MSS estimators based on simulator (1) will be CAN provided the number of simulations used grows at rate $\sqrt{N}$.

Simulator (1):

A Smooth, Direct Simulator for Likelihood Scores and Contributions

The likelihood contribution of the general LDV model examined in this paper is given by (A2.2), which we can rewrite as

$$\text{(A2.2}')\qquad \ell(y,X;\beta,\Omega) = \int_{a^*(y) \leq M(y)\cdot Y^* \leq b^*(y)} n(y^* - X\beta,\Omega)\, dy^*$$

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

But

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

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

with a, L, and b as defined in Proposition 2. Hence, the likelihood contribution becomes

(A2.9) $$\ell(y,X;\beta,\Omega) = \text{Prob}[\, a(y,X,\beta,\Omega) \le L(y,\Omega)\cdot\nu \le b(y,X,\beta,\Omega);\ \nu \sim N(0,I)\,]$$

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

Now consider a J×1 vector e_r drawn according to the sequential scheme described in equations (A2.8′), with p.d.f. given by (A2.7). Obtain R such vectors e_r's and define the likelihood contribution simulator $\tilde{\ell}(e;\, y,X;\, \beta,\Omega;\, R)$

(A2.10) $$\tilde{\ell}(e;\, y,X;\, \beta,\Omega;\, R) = \frac{1}{R}\sum_{r=1}^{R}\prod_{j=1}^{J} Q_j(e_{1r},\dots,e_{j-1,r})\,,$$

where $Q_1 \equiv \text{Prob}(a_1/l_{11} \le e_1 \le (b_1/l_{11})$, and
$Q_j(e_1,\dots,e_{j-1}) \equiv \text{Prob}((a_j - L_{j,<j}\cdot e_{<j})/l_{jj} \le e_j \le (b_j - L_{j,<j}\cdot e_{<j})/l_{jj} \mid e_1,\dots,e_{j-1})$.[27]

<u>Lemma</u>:

The simulator $\tilde{\ell}(e;\, y,X;\, \beta,\Omega;\, R)$ *defined by (A2.10) is an unbiased estimator of* $\ell(y,X;\, \beta,\Omega)$.

<u>Proof</u>:

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 (A2.7) that generates the (biased) sequential truncated draws e_r. By (A2.10), the definition of $\tilde{\ell}$, and result (A2.9),

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

$$= \int_{a \le L\cdot e \le b} \prod_{j=1}^{J} \phi(e_j)\, de_j = \text{Prob}(a \le L\cdot\nu \le b) = \ell(y,X;\, \beta,\Omega)\,. \quad \square$$

---

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

The combination of the Geweke 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 for the likelihood contribution (A2.2) that is smooth, i.e., a continuous and differentiable function of the model parameters β and Ω. Moreover, apart from an initial Choleski decomposition and several matrix multiplications, most computational effort is in drawing the univariate truncated normal variates according to the steps in (A2.8′). This effort is linear in J, the dimension of the probability integral, which is an extremely convenient feature of simulator (1). The results of Hajivassiliou (1989b) and Börsch-Supan and Hajivassiliou (1990), confirm the excellent computational efficiency of simulator (1). Hajivassiliou (1989b) shows that generating 1000 simulations $\tilde{\ell}_r$ according to this algorithm from the 10 dimensional distribution $Y^* \sim N(\mu,\Sigma)$ with $\Sigma=\{\sigma_{jj}=1,\ \sigma_{ij}=.5 \text{ for } i\neq j\}$ subject to $Y^*_j>0$ if j even and <0 if j odd, required 14.3 seconds on a NEC 386/16MHz Personal Computer, and 34.4 seconds for 20 dimensional Y^* vectors. In contrast, a sophisticated (discontinuous) acceptance-rejection algorithm required 40 minutes for the 10 dimensional case and in excess of 3 hours for the 20 dimensional one.

To obtain a smooth and asymptotically unbiased simulator for s_i, the logarithmic score (A2.4), recall that $s_i \equiv \ell_{i\theta}/\ell_i = E[h(Y^*-X\beta)\,|\,Y^*\epsilon D(y)]$. Hence, we define

$$\text{(A2.11)}\qquad \mathfrak{LS}_R \equiv \tilde{\ell}_{\theta R}/\tilde{\ell}_R\ ,$$

where $\tilde{\ell}_{\theta R} \equiv \frac{1}{R}\sum_r \{\, h(M^{-1}Le_r)\cdot \prod_j Q_j(e_{j-1,r})\,\}$, and $\tilde{\ell}_R \equiv \frac{1}{R}\sum_s \prod_j Q_j(e_{j-1,s})$. From the Lemma given above, $E\tilde{\ell}_R = \ell$; given the linear form of the likelihood derivative $h(\cdot)$ function, an exactly analogous importance sampling argument as the one used in the proof of the Lemma establishes that $E\tilde{\ell}_{\theta R} = \ell_\theta$. Given these facts, a standard law of large numbers then implies that, as $R \longrightarrow \infty$, the simulator for the denominator, $\tilde{\ell}_R$, converges to $\ell_i=E\{1(Y^*_i\epsilon D(y_i))\}$, the probability of the event $Y^*_i \in D(y_i)$, and the simulator for the numerator, $\tilde{\ell}_{\theta R}$, converges to $\ell_{i\theta}=E\{h(Y^*_i)\cdot 1(Y^*_i\epsilon D(y_i))\}$. Hence, $\mathfrak{LS}_R \xrightarrow{P} \ell_{i\theta}/\ell_i \equiv s_i$. In

Monte-Carlo experiments we found that R=100 was large enough for satisfactory sampling performance even with n=64 dimensions. The proportional bias in the parameter estimates remained below 10%. Moreover, the continuity in $\tilde{e}$ and the unknown parameters makes this simulator extremely fast. Hence, 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 $\text{Prob}(Y^* \in D(y))$, in sharp contrast to discontinuous simulators. A further feature of this simulator that apparently causes 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. For details on the comparative performance of simulator (1), see Börsch-Supan and Hajivassiliou (1990).

Next, we show that by employing Gibbs resampling techniques (Geman and Geman (1984)) we can devise another smooth simulator, simulator (2), 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. Though the Gibbs-based simulator (2) only guarantees drawing from the correct multivariate truncated normal distribution as the number of Gibbs resamplings rises without bound, Monte-Carlo experience in Hajivassiliou (1989b) suggests that the convergence rate of this method is very rapid. This finding confirms the result in Appendix 3 that MSS estimators using the Gibbs-resampling-based simulator are consistent and asymptotically normal for a *finite* number of terminal simulations, R, as the number of Gibbs resamplings used, n, grows at rate log N.

Simulator (2):

An Infinite Algorithm for Generating Truncated Multivariate Normal Variates, Based on Gibbs Resampling

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.[28] 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

$$\text{(A2.12)} \qquad [Y_j \mid Y_{-j}] \qquad\qquad j=1,\ldots,J\,,$$

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 $(\bar{a},\bar{b})$ of the multivariate normal distribution in (A2.1) is compact, which is equivalent to assuming $-\infty < \bar{a} < \bar{b} < +\infty$. This does not entail any loss of empirical generality, since we can consider a large compact rectangle defined by the limits of computing machine representation of floating point numbers. We let **B** denote the (compact) rectangle $[\bar{a},\bar{b}]$.

Gibbs sampling is a Markovian updating scheme which proceeds as follows. Given an arbitrary starting set of values $Y_1^{(0)},Y_2^{(0)},\ldots,Y_J^{(0)}$, we draw $Y_1^{(1)}\sim[Y_1 \mid Y_2^{(0)},\ldots,Y_J^{(0)}]$, then $Y_2^{(1)}\sim[Y_2 \mid Y_1^{(1)},Y_2^{(0)}\ldots,Y_J^{(0)}]$, $Y_3^{(1)}\sim[Y_3 \mid Y_1^{(1)},Y_2^{(1)},Y_3^{(0)}\ldots,Y_J^{(0)}]$, ... , and so on, up to $Y_J^{(1)}\sim[Y_J \mid Y_1^{(1)},\ldots,Y_{J-1}^{(1)}]$. Thus each variable is "visited" in the "natural" order and a cycle in this scheme requires J random variate generations. After n such iterations we would arrive at $Y^{(n)} \equiv (Y_1^{(n)},\ldots,Y_J^{(n)})$. Proposition 3 will establish that $Y^{(n)}$ will asymptotically have the true joint distribution of Y as n 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 $Y_r^{(n)}$ be a vector drawn according to the Gibbs scheme after n resamplings. By (A2.4), the

[28] The relevance of Gibbs resampling methods to our problem was suggested to us by John Geweke.

logarithmic score, s_i, equals the expectation of $h(Y,X,\beta,\Omega)$ over the distribution of Y. It then follows trivially that $E\,h(Y_r^{(n)},X,\beta,\Omega)$ converges to s_i as the number of Gibbs resamplings, n, grows to infinity. Hence, we define simulator (2) by $\tilde{s}_i(Y^{(n)},y,X,\beta,\Omega,n,R) \equiv \frac{1}{R}\sum_r h(Y_r^{(n)},y,X,\beta,\Omega)$, where R is the (finite) number of terminal simulations drawn, and n the number of Gibbs resamplings used for each simulation. Though $\tilde{s}_i$ is unbiased for the true s_i only asymptotically with n, we prove in Appendix 3 that the MSS estimator using simulator (2) is CAN provided n rises at a rate at least as fast as logN.

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 Markov chains. Consider a set A with positive Lebesgue measure. We give five definitions from Orey (1971): (i) A Markov process is *irreducible* if the probability that the process, starting from any x, ever visits the set A is positive; (ii) a Markov process is *recurrent* if the probability that it ever visits the set A is 1, starting from any point x; (iii) a Markov process is *aperiodic* if with positive probability it goes from x to set A in one move; (iv) a Markov process is *uniformly recurrent* if the probability of reaching state A within n transitions is bounded below by a positive number, uniformly in the starting point x; (v) 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 $[\tilde{Y}_2 | \tilde{Y}_1,Y_3^{(0)},\ldots,Y_J^{(0)}]$, ..., draw $\tilde{Y}_j$ from $[\tilde{Y}_j | \tilde{Y}_1,\ldots,\tilde{Y}_{j-1},Y_{j+1}^{(0)},\ldots,Y_J^{(0)}]$, ..., 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 Proposition 2, part (a). These drawings are done according to the scheme of 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 j–th random variable. From Proposition 2(a), $\tilde{Y}_j|\tilde{Y}_{-j} \sim N(\mu_{j|-j},\Sigma_{j|-j})$ conditional on $a^* \leq M\cdot\tilde{Y} \leq b^*$, where $\mu_{j|-j} = \mu_j + \Omega_{j,-j}\cdot\Omega^{-1}_{-j,-j}\cdot(\tilde{Y}_{-j}-\mu_{-j})$, $\mu_k \equiv (X\beta)_k$, and $\Omega_{j|-j} = \Omega_{jj} - \Omega_{j,-j}\cdot\Omega^{-1}_{-j,-j}\cdot\Omega_{-j,j}$. Then it follows that the truncated multivariate normal distribution Y^* conditional on the compact region $a^* \leq MY^* \leq b^*$ will be an invariant of this process, since the $[\tilde{Y}_j|\tilde{Y}_1,\ldots,\tilde{Y}_{j-1},Y^{(0)}_{j+1},\ldots,Y^{(0)}_J]$ distributions are by construction the one-dimensional conditionals of that joint distribution.

Proposition 3 (Convergence at a Geometric rate):

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

Proof:

Define $p(n,x,y)$ for $(x,y) \in B$ to be the density of $Y^{(n)}$ starting from $Y^{(0)}=x$; this is given constructively by the Gibbs updating scheme we described. Also by construction, p is continuous on B, $p(1,x,y) > 0$, and $p(n,x,y) = \int p(n-1,x,z)\cdot p(1,z,y)\, dz > 0$ for $n > 1$. Since by assumption B is compact, $p(1,x,y)$ is bounded positive on B. This implies in turn that the process is uniformly recurrent, since the probability of never reaching a set A of positive measure in n rounds is bounded above by $[1-\gamma\mu(A)]^n$, 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 $n(y^*-X\beta,\Omega,\bar{a},\bar{b})$ is an invariant of the Gibbs process. Then, Theorem 7.2 in Orey (1971) implies that the L_1 distance $\|p(n,Y^{(0)},y) - n(y^*-X\beta,\Omega,\bar{a},\bar{b})\|$ converges to 0 as $n \to \infty$ 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 $\int |p(n,Y^{(0)},y) - n(y^*-X\beta,\Omega,\bar{a},\bar{b})|\, dy \leq M\lambda^n$. □

It should be noted that, like simulator (1) above, simulator (2) is by construction continuous in the distributional parameters, β, Ω, a^*, M, and b^*. As found in Hajivassiliou

(1989b), it is computationally tractable and the convergence rate of the Gibbs resamplings is very fast. Hence, the MSS estimator based on it possesses desirable properties in terms of computational performance. These findings confirm the result of Appendix 3 that consistency and asymptotic normality of the MSS estimator based on simulator (2) using a finite number of terminal simulations, requires that the number of Gibbs resamplings used to generate each draw rises only at the rate logN.

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 $\frac{1}{R}\sum_r h(\tilde{Y}_r - X\beta)$ will be unbiased for any number of terminal simulations, R. We are therefore able to prove in Appendix 3 that the MSS estimator that uses simulator (3) will be CAN for *any* (finite or infinite) number of simulations. Though this method is not continuous in the parameters of the underlying distribution, the results in Hajivassiliou (1989b) suggest that simulator (3) 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).

<u>Simulator (3):</u>

<u>An Acceptance–Rejection Algorithm for Generating Truncated Multivariate Normal Variates</u>

The idea for this algorithm is based on the following:

Proposition 4 (Theorem 3.3 in Devroye (pp.47–48)):

In order to generate draws from a density

$f(z)=c\cdot g(z).\psi(z)$, *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* $U \le \psi(Z)$ *; 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

(A2.13) $$Y \equiv \begin{cases} \tilde{x} & \text{if } \tilde{u} \le \psi(\tilde{x}) \\ \text{not observed otherwise} \end{cases} .$$

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

(A2.14) $$F_Y(y) = \text{Prob}(\tilde{x} \le y \,|\, \tilde{x} \text{ accepted}) = \text{Prob}(\tilde{x} \le y, \tilde{u} \le \psi(\tilde{x})) / \text{Prob}(\tilde{u} \le \psi(\tilde{x}))$$

$$= \frac{\int_{-\infty}^{y} \text{Prob}(\tilde{u} \le \psi(x))\ g(x)\ dx}{\int_D \text{Prob}(\tilde{u} \le \psi(x))\ g(x)\ dx} = \frac{\int_{-\infty}^{y} \psi(x)\ g(x)\ dx}{\int_D \psi(x)\ g(x)\ dx} = \int_{-\infty}^{y} \frac{1}{c} \cdot f(x)\, dx \Big/ \int_D \frac{1}{c} \cdot f(x)\, dx .$$

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. □

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

(A2.15) $$f(z) = \frac{n(z - X\beta, \Omega)}{\int_{a^* \le Mz \le b^*} n(z - X\beta, \Omega)\, dz} \qquad \text{if } z \in D(y),$$

$$= 0 \qquad \text{otherwise.}$$

Hence, the objective will be to devise convenient densities g(z) to draw from, 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

(A2.16) $\quad Z \sim N(X\beta,\Lambda) \quad$ conditional on $D(y) \equiv \{\, M(y)^{-1}a^*(y) \le Z \le M(y)^{-1}b^*(y)\,\}$,

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 of Proposition 1 to generate univariate normal truncated random variates, and because, 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

(A2.17) $\quad \mathrm{Prob}(Z\epsilon D(y))$

$$= \mathrm{Prob}(M^{-1}a^* \le Z \le M^{-1}b^*) = \prod_{j=1}^{J}\{\Phi[(M^{-1}b^*)_j/\lambda_j] - \Phi[(M^{-1}a^*)_j/\lambda_j]\}\,.$$

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

(A2.18) $\quad g(z) \;= n(z-X\beta,\Lambda)\;/\;\mathrm{Prob}(Z\epsilon D(y)).$

Choose Λ so that $\Lambda-\Omega$ is positive definite.[29] Then $|\Lambda| \ge |\Omega|$, and

(A2.19) $$\max_z \frac{n(z-X\beta,\Omega)}{n(z-X\beta,\Lambda)} = |\Lambda|^{\frac{1}{2}}/|\Omega|^{\frac{1}{2}} \equiv \gamma \ge 1.$$

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

(A2.20) $$\tilde{u} \le \frac{n(\tilde{z}-X\beta,\Omega)}{n(\tilde{z}-X\beta,\Lambda)\cdot\gamma} \equiv \psi(\tilde{z}) \le 1.$$

By simple inspection, we then see that we have written f(z) as $c\cdot g(z)\cdot\psi(z)$, where

---

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

g(z) is given in (A2.18), $\psi(z)$ in (A2.20), and the constant c determining the expected number of draws before the first acceptance is

$$\text{(A2.21)} \qquad c \equiv \gamma \cdot \frac{\text{Prob}(Z \in D(y))}{\text{Prob}(Y^* \in D(y))},$$

where Prob(Z∈D(y)) is given by (A2.17). Hence, by Theorem 3.3 the accepted $\tilde{z}$'s will have density (A2.15) as required. The acceptance rate 1/c can be maximized given Ω by choosing Λ suitably.

Finally we give

<u>Acceptance–Rejection Method (b)</u>:

We have shown that by defining a, b, and L as in Proposition 2, the density f(z) in (A2.15) can be written over its support as

$$\text{(A2.22)} \qquad f(z) \;= n(z,I) \Big/ \int_{a \le Lz \le b} n(z,I)\, dz\,.$$

Draw a vector $\tilde{e}$ using the sequential scheme of Proposition 2, which has the (convenient) density

$$\text{(A2.23)} \qquad g(e) \;= n(e,I) \Big/ \prod_{j=1}^{J} Q_j(e_{j-1}) \qquad \text{with support } \{\, a \le L\cdot e \le b \,\}.$$

Consider a bound B such that $B \ge \text{Prob}(Y^* \in D(y)) = \int_{a \le Lz \le b} n(z,I)\, dz$ and $B \ge \text{Prob}(\, a \le L\cdot e \le b\,) = \prod_{j=1}^{J} Q_j(e_{j-1})$.[30] 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

$$\text{(A2.24)} \qquad \tilde{u} \le \prod_{j=1}^{J} Q_j(\tilde{e}_{j-1}) \,/\, B \equiv \psi(\tilde{e}) \le 1.$$

Thus, we have written density f(z) in (A2.22) as $c\cdot g(z)\cdot \psi(z)$, where (A2.23) gives

---

[30] Such a bound can be constructed as the probability of $L\cdot z$ lying in the smallest rectangular region containing the support $\{\, a \le L\cdot z \le b \,\}$, 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).

$g(z)$, (A2.24) gives $\psi(z)$, and $c \equiv B / \text{Prob}(Y^* \epsilon D(y))$. Therefore, by Theorem 3.3, acceptance-rejection method (b) generates accepted $\tilde{e}$'s with density (A2.22), which is equivalent to the desired density (A2.15). The method will have an expected acceptance rate of $1/c = \text{Prob}(Y^* \epsilon D(y)) / B$, which is larger the closer the bound is to the true conditioning probability. This bound is tight for positively correlated elements of Y^*, and becomes less so for negatively correlated Y^*_j's. This is confirmed by the Monte-Carlo results in Hajivassiliou, McFadden, and Ruud (1990).

Appendix 3

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,

$$\text{(A3.1)} \qquad B_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [E_{\sim}\tilde{s}_i(\theta) - s_i(\theta)] \,,$$

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

$$\text{(A3.2)} \qquad \zeta_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i(\theta) \,, \ \text{with} \ \ \xi_i(\theta) \equiv [\tilde{s}_i(\theta) - E_{\sim}\tilde{s}_i(\theta) - \tilde{s}_i(\theta^*) + E_{\sim}\tilde{s}_i(\theta^*)] \,.$$

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 estimation $\hat{\theta}_N$ that solves $\sum_{i=1}^{N} \tilde{s}_i(\hat{\theta}_N) = 0$ to be consistent and asymptotically normal.

Theorem. 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 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.*[31] *Assume that a MSS estimator solving* $0 = \sum_{i=1}^{N} \tilde{s}_i(\hat{\theta}_N)$ *exists for each N.*[32] *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 $Q = \mathbf{E}[\tilde{s}_i(\theta^*) - \mathbf{E}_{\sim}\tilde{s}_i(\theta^*)][\tilde{s}_i(\theta^*) - \mathbf{E}_{\sim}\tilde{s}_i(\theta^*)]'$.

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

---

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

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

If ζ_N is stochastically equicontinuous at Θ, with Θ compact and convex, and $\zeta_N(\theta^o)$ is stochastically bounded for some $\theta^o\in\Theta$, 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^o and any $\theta\in\Theta$, where M bounds the diameter of Θ. Then,

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

implying that given $\epsilon,\lambda>0$, there exists $\delta>0$ such that

$$\text{Prob}(\sup_{\theta\in\Theta} |\zeta_N(\theta) - \zeta_N(\theta^o)| > 2M\epsilon/\delta) < \lambda .$$

This works for the simulation residual process in this Appendix with $\theta^o = \theta^*$ since $\zeta_N(\theta^*)=0$.

[32] 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.

$$\text{(A3.3)} \qquad 0 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} s_i(\hat{\theta}_N) \equiv A_N + C_N(\hat{\theta}_N) + \zeta_N(\hat{\theta}_N) + B_N(\hat{\theta}_N) ,$$

with,

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

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

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 stochastically 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 $\psi_\theta(\theta^*) = -J$.[33] From (A3.3), 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,

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

with the second equality following from a Taylor's expansion of $s_i(\hat{\theta}_N)$ about θ^* using

[33] A U.L.L.N. states that given ϵ, $\delta > 0$, there exists N_o such that for $N \geq N_o$,

$$\text{Prob}\left(\max_{\theta \in \Theta} |C_N(\theta) - \psi(\theta)| > \delta \right) < \epsilon .$$

the differentiability and dominance assumptions. Then, $\mathbf{E}s_{i\theta}(\theta^*)$ non-singular 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 (A3.3) 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,

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

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

This paper is concerned with the <u>special case</u> of LDV models formed from a vector of exogenous variables x, a parameter θ, and a standard normal latent vector $v \in \mathbb{R}^m$. A finite series of hyperplanes, of the form $\{v \in \mathbb{R}^m | v \cdot p_k(x,\theta) = c_k(x,\theta)\}$, with p_k a normal vector of unit length, partition $\mathbb{R}^m$ 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 $\mathbf{D}(x,\theta,d)$ denote the set of v that map into d. Then, the score of observation i from an independently, identically distributed sample of size N can be written

$$\text{(A3.4)} \qquad s_i(\theta) = \mathbf{E}_v(h(v,\theta,x_i,d_i,y_i) | v \in \mathbf{D}(x_i,\theta,d_i))\,,$$

where h is a vector of polynomials in v. To avoid technical difficulties, we assume for the special case (without any essential loss of empirical generality) that the multivariate normal distribution v is truncated to a large compact rectangle.[34] We make the regularity

[34] For example, the density of v is multivariate normal, truncated to the square defined by the limits of computing machine representation of floating-point numbers.

assumptions that the functions $p_k(x,\theta)$, $c_k(x,\theta)$, $a(x,\theta)$, and $B(x,\theta)$ are all continuously 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 for the special case by one of the following methods, corresponding to simulators (1)–(3) in Appendix 2:

(1) Simulate independently the numerator and denominator of

$$\text{(A3.5)}\qquad s_i(\theta) = \frac{E_v(h(v,\theta,x_i,d_i,y_i)\cdot 1(v\in D(x_i,\theta,d_i))}{E_v 1(v\in D(x_i,\theta,d_i))},$$

employing fixed sequences of random generators v. An unbiased simulator with one or more draws is used for the numerator. An unbiased simulator with R_N independent draws that is uniformly bounded positive, with $R_N/\sqrt{N} \longrightarrow \infty$, is used for the denominator. (For example, simulators based on (A2.10) and (A2.11) meet these requirements.)

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

(3) 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.

We give some general sufficient conditions for assumption (i) of asymptotic unbiasedness and assumption (ii) of stochastic equicontinuity in the Theorem. We show that these sufficient conditions are satisfied in the <u>special case</u> for each of the simulation methods (1)–(3). The hypotheses of the Theorem 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* $o(1/\sqrt{N})$*, then the simulation bias converges to zero. In the special case, this result holds for the simulator (3) that is unbiased, and holds for simulators (1) or (2) with the stated sampling rates.*

Proof: The result holds trivially for unbiased simulators such as simulator (3). 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:

$$P(\sup_{\theta\in\Theta} |B_N(\theta)| > \epsilon)$$

$$< \frac{1}{\sqrt{N}\epsilon} \sum_{i=1}^{N} \mathbf{E}_i \sup_{\theta\in\Theta} |\mathbf{E}_{\sim}\tilde{s}_i(\theta) - s_i(\theta)|$$

$$= \mathbf{E}_i\sqrt{N} \sup_{\theta\in\Theta} |\mathbf{E}_{\sim}\tilde{s}_i(\theta) - s_i(\theta)|/\epsilon \longrightarrow 0 .$$

For simulation method (1), one has

$$|\mathbf{E}_{\sim}\tilde{s}_i(\theta) - s_i(\theta)| = |\ell_{i\theta}| \cdot |\mathbf{E}_{\sim}(1/\tilde{\ell}_i) - 1/\ell_i| ,$$

$$\leq |\ell_{i\theta}| \cdot \mathbf{E}_{\sim}|1/\tilde{\ell}_i - 1/\ell_i| ,$$

$$\leq |\ell_{i\theta}| \cdot \mathbf{E}_{\sim}|\tilde{\ell}_i - \ell_i|/\ell_i m,$$

where m 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 $\mathbf{E}_i\mathbf{E}_{\sim}|\tilde{\ell}_i - \ell_i| = \mathcal{O}_p(1/\sqrt{R_N})$, and $R_N/\sqrt{N} \longrightarrow \infty$ gives the result.

For simulation method (2), one has

$$|\mathbf{E}_{\sim}\tilde{s}_i(\theta) - s_i(\theta)|$$

$$= |\mathbf{E}_v(h|v\epsilon D(y_i),\ v \sim f^n) - \mathbf{E}_v(h|v\epsilon D(y_i),\ v \sim f)|$$

$$\leq M' \ \|f^n - f\| \leq Me^{-\lambda n} ,$$

where f^n denotes the distribution of the Gibbs sampler after 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 3 which states that when the support is compact, $\|f^n - f\|$ converges to zero at a geometric rate. Then, taking $n_N > (\log N)/2\lambda$ yields the result. □

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 $(\mathbb{H},\mathcal{V},\mu)$ denote a probability space, Θ a compact subset of $\mathbb{R}^k$, and $\xi(v,\theta)$ a measurable function on $\mathbb{H}\times\Theta$. Assume that ξ is dominated by a square-integrable function ν on $\mathbb{H}$. Assume that $\mathbf{E}\xi(V,\theta) \equiv 0$, and let $\sigma^2 \equiv \mathbf{E}\nu(V)^2$. 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

$$\text{(A3.6)} \qquad \delta_j = \max_{\theta\epsilon\Theta_j} \left[\mathbf{E}^* \sup_{\{\theta' | \theta=\theta_j(\theta')\}} |\xi(V,\theta') - \xi(V,\theta)|^2 \right]^{1/2} ,$$

where $\mathbf{E}^*$ 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 $\delta_j \longrightarrow 0$. Let v_i for $i = 1,2,\ldots$ denote independent realizations of V, and form $\zeta_N(\theta) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi(v_i,\theta)$. Ossiander 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

$$\text{(A3.7)} \qquad \sum_{j=2}^{\infty} [\log N_j]^{1/2}(\delta_j - \delta_{j-1}) < +\infty .$$

We next introduce a regularity condition on simulators that is sufficient to satisfy (A3.7). The simulator $\tilde{s}_i(\theta)$ is probably Lipschitz on Θ if there exists $\delta_o > 0$ and an integrable function $m_i \geq 1$ with a finite third moment such that $|\tilde{s}_i(\theta)| \leq m_i$ and for $0 < \delta < \delta_o$ 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 special case, simulators (1) and (2) are continuously differentiable, while simulator (3) 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 region-wise constant on partition j. Using the Lipschitz hypothesis, one has $|\tilde{s}_i(\theta) - \tilde{s}_i(\theta_j(\theta))| \le B_{ij}(\theta)$. Also, for j large enough so that $\beta_j < \delta_o$,

$$\mathrm{E}\, B_{ij}(\theta)^2 \le \mathrm{E}_i\{(1 - Q_{\beta_j}(\theta_j(\theta))\, m_i^2\, \beta_j^2 + Q_{\beta_j}(\theta_j(\theta))2m_i^2\}$$

$$\le \mathrm{E}_i\{m_i^2\, \beta_j^2 + 2m_i^3\, \beta_j\} \le 3\beta_j \mathrm{E}_i m_i^3 \equiv \delta_j^2 .$$

Define $\delta_j^2 = 2\mathrm{E}_i m_i^3$ for $\beta_j \ge \delta_o$. Then,

$$\sum_{j=2}^{\infty} [\log N_j]^{1/2}(\delta_{j-1} - \delta_j) = \sum_{j=2}^{\infty} [kj \log 2]^{1/2}(2^{-(j-1)} - 2^{-j})\cdot 3\mathrm{E}_i m_i^3 < +\infty .$$

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

Consider the special case. The simulators (1) and (2) are continuously differentiable on Θ, so they are Lipschitz with probability one, and the result follows.

Now consider simulator (3). 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, implies 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,

$$A_\delta(x,\theta,d) = P(\{v|\ \sup_{\theta'\in N_\delta(\theta)} (p_k(x,\theta')\cdot v - c_k(x,\theta')) \le 0,\ k = 1,\dots,K\})$$
$$\ge P(\{v\,|\,p_k(x,\theta)\cdot v - c_k(x,\theta) \le -2m(x)\cdot\delta,\ k = 1,\dots,K\})\,.$$

Similarly,

$$R_\delta(x,\theta,d) = P(\{v|\ \inf_{\theta'\in N_\delta(\theta)} (p_k(x,\theta')\cdot v - c_k(x,\theta')) > 0,\ k = 1,\dots,K\})$$
$$\ge P(\{v\,|\,p_k(x,\theta)\cdot v - c_k(x,\theta) > 2m(x)\cdot\delta,\ k = 1,\dots,K\})\,,$$

and

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

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

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

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

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

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 sense 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 (3). □

Appendix 4
Data Sources, Description of Variables, and Descriptive Statistics

PART 1: Data Sources

Abbreviations for Data Sources (Computer Tapes)

BOP World Bank, World Tables, economic data sheet 2, balance of payments (1987)
ERP U.S. Council of Economic Advisers, 1985 Economic Report of the President
IMF IMF Annual Reports of the Director, various issues.
IFS International Monetary Fund, International Financial Statistics (1987)
WB World Bank, World Tables, economic data sheet 1 (1987)
WDT World Bank, World Debt Tables (1987)
WCY World Currency Yearbook, various issues.

All series consist of 1853 country-year observations, on 109 countries over the 1970–1986 period. All conversions between dollar and local currency values employed the period average exchange rate from IFS.

PART 2: Construction of Variables and Descriptive Statistics

Indicators of Repayments Problems	mean	standard deviation
PSArI	(0.126)	(0.354)

Presence of "Significant" Arrears in Interest, 1970–1986, WB. "Significant" defined as greater than .001 of Total External Debt.

PSArP	(0.069)	(0.262)

Presence of "Significant" Arrears in Principal, 1970–1986, WB. "Significant" defined as greater than .01 of Total External Debt.

PRSSIMF	(0.293)	(0.541)

Occurrence of a Rescheduling Arrangement and/or IMF involvement, 1970–1986, IMF. IMF involvement defined as IMF support. IMF support is defined by an IMF standby agreement of second or higher tranche or use of the IMF Extended Fund Facility. Reschedulings include Paris Club, commercial banks, and aid-consortia renegotiations. This information was compiled from our own country-by-country investigations, and from published and unpublished IMF sources. The date of rescheduling was selected to reflect the key economic developments precipitating the rescheduling.

SArI	(0.0003)	(0.0011)

Level of "Significant" Arrears in Interest, 1970–1986, WB. "Significant" defined as greater than .001 of Total External Debt.

SArP	(0.0008)	(0.0050)

Level of "Significant" Arrears in Principal, 1970–1986, WB. "Significant" defined as greater than .01 of Total External Debt.

SAr	(0.0011)	(0.006)

"Significant" Total Arrears in Principal and Interest, 1970–1986, WB. See above.

Crisis3F	(0.723)	(1.154)

"Severity of Crisis" Indicator: 0=no repayments problem, 1=significant arrears only, 2=IMF or RSS, 1971–1987.

PArIF (0.419) (0.647)
Binary Indicator, 0=no repayments problem, 1=significant arrears, IMF involvement, or rescheduling agreement, 1971–1987.

SArF (0.002) (0.008)
"Significant" Total Arrears in Principal and Interest, 1971–1987, WB. See above.

CumPSArI (0.727) (2.096)
Cumulated number of past years with significant arrears in interest present.

CumPSArP (0.315) (1.064)
Cumulated number of past years with significant arrears in principal present.

CumRorI (1.960) (3.204)
Cumulated number of past years with a rescheduling or an IMF agreement in effect.

<u>**Explanatory Variables**</u>

PCGDP80 (1.261) (1.848)
Per Capita GDP, 1980 US$, 1970–1986, WB.

DbttoExp (176.991) (232.814)
Total External Debt Relative to Exports, 1970–1986, WDT, IFS. Total debt includes public and private debt outstanding and disbursed, short–term debt, and use of IMF credit.

REStoImp (3.343) (4.585)
International Reserves (Excl. Gold) Relative to Imports, 1970–1986, WDT.

DSDtoExp (11.994) (15.403)
Total Debt Service Due Relative to Exports, 1970–1986, WDT, WB. Debt service due defined as interest and principal paid (TDS from WDT) plus outstanding interest and principal arrears.

ISDtoExp (4.871) (6.763)
Interest Service Due Relative to Exports, 1970–1986, WDT, WB. Interest service due defined as interest paid (INT from WDT) plus outstanding interest arrears.

PSDtoExp (7.123) (9.421)
Principal Service Due Relative to Exports, 1970–1986, WDT, WB. Principal service due defined as principal paid plus outstanding principal arrears.

DSPTOEXP (11.989) (15.396)
Total Debt Service Paid Relative to Exports, 1970–1986, WDT.

ISPtoExp (4.870) (6.762)
Total Interest Service Paid Relative to Exports, 1970–1986, WDT.

PSPtoExp (7.118) (9.415)
Total Principal Service Paid Relative to Exports, 1970–1986, WDT.

CAtoGNP (–0.103) (0.159)
Current Account Balance (Exports – Imports) Relative to GNP, 1970–1986, WDT.

PART 3: **Panel Summary Statistics**

Means of variables over individuals in a given year ($\frac{1}{N}\sum_i x_{it}$)

Variable	1971	1972	1973	1974	1975	1976	1977	1978	1979
Number of Countries with available Data	91	90	89	89	89	88	85	82	81
PSArI	0.044	0.067	0.090	0.101	0.090	0.091	0.071	0.085	0.111
PSArP	0.022	0.022	0.034	0.034	0.022	0.034	0.047	0.024	0.049
PRSSIMF	0.231	0.233	0.202	0.191	0.135	0.125	0.152	0.182	0.173
SArI	6.73e–5	3.17e–5	4.64e–5	6.32e–5	5.76e–5	7.21e–5	7.23e–5	9.92e–5	1.46e–4
SArP	5.85e–5	6.92e–5	9.81e–5	1.11e–4	1.28e–4	1.12e–4	3.24e–4	1.54e–4	4.54e–4
SAr	1.25e–4	1.07e–4	1.44e–4	1.74e–4	1.84e–4	1.84e–4	3.96e–4	2.53e–4	5.96e–4
Crisis3F	0.538	0.489	0.472	0.348	0.360	0.410	0.435	0.463	0.593
PArIF	0.297	0.289	0.281	0.213	0.236	0.261	0.259	0.293	0.346
SArF	1.43e–4	2.43e–4	1.74e–4	1.84e–4	2.83e–4	4.02e–4	2.44e–4	5.89e–4	9.92e–4
CumPSArI	0.055	0.111	0.169	0.270	0.360	0.432	0.471	0.573	0.691
CumPSArP	0.022	0.044	0.079	0.112	0.135	0.159	0.212	0.244	0.296
CumRorI	0.341	0.511	0.708	0.899	1.034	1.114	1.247	1.451	1.593
PCGDP80	1.032	1.066	1.119	1.151	1.197	1.228	1.300	1.312	1.341
DbttoExp	111.730	123.420	124.590	113.710	109.670	117.490	126.250	161.470	185.860
RestoImp	3.056	3.200	3.710	4.025	3.929	3.285	3.565	3.689	3.512
DSDtoExp	7.589	8.150	8.614	8.936	8.145	9.309	9.611	10.259	13.290
ISDtoExp	2.427	2.641	2.677	2.817	2.787	3.333	3.490	3.635	4.508
PSDtoExp	5.162	5.509	5.938	6.119	5.358	5.975	6.121	6.624	8.782
ISPtoExp	2.426	2.641	2.676	2.817	2.786	3.333	3.489	3.634	4.507
PSPtoExp	5.160	5.507	5.936	6.117	5.356	5.974	6.119	6.622	8.778
CAtoGNP	–0.078	–0.093	–0.092	–0.090	–0.095	–0.118	–0.100	–0.104	–0.123

Panel Summary Statistics (continued)

	1980	1981	1982	1983	1984	1985	1986	1987
Number of Countries with available Data	78	77	74	72	69	68	65	51
PSArI	0.103	0.104	0.135	0.167	0.203	0.280	0.277	0.275
PSArP	0.039	0.054	0.125	0.029	0.059	0.185	0.569	0.051
PRSSIMF	0.377	0.446	0.472	0.580	0.515	0.554	0.471	0.243
SArI	9.92e–5	2.02e–4	4.68e–4	4.63e–4	8.48e–4	1.27e–3	8.27e–3	1.43e–4
SArP	1.78e–4	8.04e–4	1.06e–3	2.25e–4	5.54e–4	2.39e–3	1.01e–2	4.23e–4
SAr	2.77e–4	1.01e–3	1.53e–3	6.88e–4	1.40e–3	3.66e–3	1.09e–2	5.63e–4
Crisis3F	0.909	1.041	1.236	1.174	1.191	1.308	1.255	0.808
PArIF	0.481	0.581	0.681	0.652	0.647	0.815	0.784	0.436
SArF	9.67e–4	1.69e–4	9.34e–4	1.47e–3	3.50e–3	1.39e–3	1.09e–2	5.27e–4
CumPSArI	0.831	1.000	1.194	1.406	1.662	2.015	1.560	0.821
CumPSArP	0.364	0.432	0.556	0.507	0.574	0.785	1.137	0.346
CumRorI	2.078	2.595	3.139	3.855	4.324	4.831	5.314	1.744
PCGDP80	1.433	1.462	1.405	1.269	1.274	1.217	1.430	1.390
DbttoExp	186.850	207.730	257.240	277.880	273.120	311.100	282.520	186.500
RestoImp	3.512	2.834	2.613	2.700	2.619	2.797	3.465	3.827
DSDtoExp	12.53	14.03	15.617	16.23	16.27	19.31	21.774	13.65
ISDtoExp	5.332	6.251	7.692	8.326	8.106	9.141	10.397	5.024
PSDtoExp	7.200	7.777	?.925	7.902	8.163	10.164	11.377	8.630
ISPtoExp	5.331	6.250	7.690	8.323	8.101	9.134	10.394	5.023
PSPtoExp	7.197	7.772	7.917	7.893	8.151	10.146	11.347	8.626
CAtoGNP	−0.118	−0.127	−0.127	−0.110	−0.086	−0.094	−0.078	−0.112

REFERENCES

Andersen, E., 1970, Asymptotic Properties of Conditional Maximum Likelihood Estimators, Journal of the Royal Statistical Society, Series B, 32, 283–301.

Binmore, K. and M. Herrero, 1984, Frictionless Non Walrasian Markets, ICERD discussion paper, London School of Economics.

Börsch-Supan, A. and V. Hajivassiliou, 1990, Smooth Unbiased Multivariate Probability Simulators for ML Estimation of LDV, Cowles Foundation Discussion Paper.

Butler, J.S. and R. Moffit, 1982, A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model, Econometrica 50, 761–764.

Chamberlain, G., 1980, Analysis of Covariance with Qualitative Data, Review of Economic Studies, 47, 225–238.

Cline, W, 1983, International Debt and the Stability of the World Economy, Washington DC: Institute for International Economics.

Deak, I, 1980, Three digit accurate Multivariate Normal Probabilities, Numerische Mathematik, 35, 369–380.

Devroye, L., 1986, Non–Uniform Random Variable Generation. New York: Springer-Verlag.

Dudley, R., 1978, Central Limit Theorems for Empirical Measures, Annals of Probability, 6, 899–929 (correction, *ibid* 7 (1979), 909–911).

Eaton, J. and M. Gersovitz, 1980, LDC Participation in International Financial Markets: Debt and Reserves, Journal of Development Economics 7, 1–21.

Eaton, J. and M. Gersovitz, 1981, Debt with Potential Repudiation: Theoretical and Empirical Analysis, Review of Economic Studies 48, 289–309.

Edwards, S., 1984, LDC Foreign Borrowing and Default Risk: An Empirical Investigation (1976–1980), American Economic Review 74, 726–734.

Feder, G., and R. Just, 1977, A study of Debt Servicing Capacity Applying Logit Analysis, Journal of Development Economics 4, 25–38.

Feder, G., R. Just, and K. Ross, 1981, Projecting Debt Servicing Capacity of Developing Countries, Journal of Financial and Quantitative Analysis 16, 651–669.

Feller, W., 1971, An Introduction to Probability Theory and Its Applications, Vol.II. Wiley: New York.

Gallant, R.A., and D.W. Nychka, 1987, Semi–Nonparametric Maximum Likelihood Estimation, Econometrica 55, March: 363–390.

Gelfand, A.E., and A.F.M. Smith, 1988, Sampling Based Approaches to Calculating Marginal Densities, University of Connecticut mimeo.

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.