University of Wisconsin-Madison

January, 1999 **Staff Paper No. 421**

Maximum Simulated Likelihood:
A Brief Introduction for Practitioners

By

Carlos Arias
and
Thomas L. Cox

AGRICULTURAL &
APPLIED ECONOMICS

STAFF PAPER SERIES

MAXIMUM SIMULATED LIKELIHOOD:
A BRIEF INTRODUCTION FOR PRACTITIONERS

Carlos Arias and Thomas L. Cox [1]

Abstract:

This paper discusses the increasing importance of probability simulation methods in the context of Maximum Simulated Likelihood. Three probability simulators are analyzed following their chronological order of appearance. This analysis provides an intuitive approach to the basic idea behind probability simulation, the successive improvements and probable future developments. The paper pays special attention to the role of simulation noise in Maximum Simulated Likelihood.

Key words: Maximum Simulated Likelihood, Probability Simulation Methods, High-Dimensional Integrals, Partially Observable Endogenous Variables, Censoring, Discrete Choice Models.

[1] Carlos Arias is an assistant professor at the Department of Economics, University of Oviedo, Spain. E-mail: arias@econo.uniovi.es. Thomas L. Cox is a professor at the Department of Agricultural and Applied Economics, University of Wisconsin-Madison. E-mail: cox@aae.wisc.edu.

MAXIMUM SIMULATED LIKELIHOOD:
A BRIEF INTRODUCTION FOR PRACTITIONERS

1. Introduction

There are numerous examples in economics in which the likelihood function of the model contains high-dimensional integrals without a closed form solution. The computational cost of such integrals increases very fast with the dimension of the problem. As a result, the estimation of models with these characteristics was considered unfeasible until recently.

These integrals appear generally as the result of some endogenous variables being only partially observable. A non-exhaustive list of economic models with difficult to compute definite integrals in the likelihood function includes:[2]

1) *Systems of equations with several censored endogenous variables*. For example, a systems of demand equations where some individuals choose not to consume several of the goods in the system (Wales and Woodland, 1983; Lee and Pitt, 1986). In general, the dimension of the definite integral to compute in these models increases with the number of products not consumed.
2) *Panel data Tobit models with a flexible temporal correlation in the unobservable variables.* Hajivassiliou (1993) applies this methodology to the analysis of external debt crisis. Feenberg and Skinner (1994) analyze the persistence of health care expenditures.
3) *Discrete choice models that analyze the choice between multiple alternatives as a function of individual characteristics*. In this case, the dimension of the definite integrals is given by the number of different options that individuals face. These models are behind many early attempts to find a feasible way of computing higher dimensional integrals. A leading example is the paper by MacFadden, 1989.

[2] A detailed list of examples of economic models with high dimension integrals can be found in Gourieroux and Monfort (1996 p. 8-14) and Stern (1997).

4) *Discrete inter-temporal choice models*. These models analyze the relationship between individual characteristics and decisions over time in a dynamic programming framework. In this case, the dimension of the definite integrals is given by the number of periods analyzed. These models are used to analyze retirement decisions, child spacing, international currency and asset choices, and patent renewal decisions.
5) *Unobserved heterogeneity*. This problem can be overcome by integrating over the probability distribution of unobserved individual characteristics.

Historically, the estimation of these models has relied on the use of quite strong simplifying assumptions. For clarity of exposition we distinguish between:

a) ***Assumptions about the structure of the model***. For example, the aggregation of goods helps to estimate a system of demand equations where individuals report zero consumption of several goods in the system in two ways. First, by aggregating, one will be dealing with a smaller number of censored variables. Second, sometimes the zero consumption simply disappears with aggregation. The problem is that aggregation can hide some important characteristics of the model and even make the estimation useless for some research objectives.
b) ***Assumptions about the disturbances of the model***. In this case, we assume that the random disturbances of the model follow a probability distribution defined by easy to integrate probability density functions. This is the case in problems of discrete choice where the assumption of an Extreme Value Distribution leads to the Multinomial Logit Model or the assumption of a Generalized Extreme Value distribution leads to the Nested Logit Model. The Multinomial Logit model distributional assumption makes the results suffer from the well know problem of independence of irrelevant alternatives. In the Nested Logit Model, it is necessary to make arbitrary assumptions about the order in which decisions are made.

The undesirable properties of these simplifying assumptions motivated the search for a method of evaluation of high-dimensional integrals that keeps a reasonable balance between accuracy and computational costs. The idea proposed is both simple and powerful: the integrals of interests are probabilities of a certain event in a random process. We can simulate that random process and use the empirical probability of the event as an approximation to the value of the integral of interest (Lerman and Manski, 1981)[3]. This initial idea and its sequels have been labeled in the literature as Probability Simulation Methods. In particular, the Method of Simulated Likelihood consists in simulating rather than computing high dimensional integrals in the likelihood function.

These two approaches (restrictive specification with a closed form solution versus a more general specification that requires a numerical solution) reflect an ongoing methodological problem in economics. Often, we accept quite restrictive assumptions in order to get closed form solutions for the model. An alternative, less restrictive specification would lead to a model with a numerical solution (Judd, 1998). Developments in computational hardware and software have fostered the use of numerical methods in many areas of economic research.

There is a booming literature on Probability Simulation Methods that improve on the basic idea (Stern, 1992; Borsch-Saupan and Hajivassiliou, 1993) and analyze the theoretical properties of the proposed simulators (Hajivassiliou *et al*., 1996). At the same time there are a number of papers that use Probability Simulation Methods in empirical economic analysis.[4] The objective of the present paper is to review some of the more "user friendly" probability simulation methods for use in Maximum Simulated Likelihood. A number of results scattered through the literature are summarized using some intuition

[3] This idea, to simulate rather than calculate a probability, had been used before in the field of computer simulation. However, Lerman and Manski can be credited with the pioneering use of probability simulation in the empirical analysis of economic problems.

[4] The Review of Economics and Statistics (Vol. 67, November 1994) contains several seminal papers in the use of probability simulation methods in empirical applications.

gained while using these techniques in applied work. We hope this effort provides an easy way to become aware of the usefulness and relative simplicity of using Probability Simulation Methods in applied work.

The structure of the paper is as follow: Section 2 reviews three well known probability simulators. Section 3 discusses the role of simulation noise in Maximum Simulated Likelihood and provides and heuristic explanation for a statistical test for the relevance of simulation noise. Section 4 ends the paper with some concluding remarks.

2. Probability Simulation Methods

The numerical methods commonly used to approximate the value of a definite integral are known as quadrature methods. In these numerical techniques, the integrand is substituted by an approximating polynomial of degree k. For the one-dimensional case, the interpolation of this approximating polynomial needs of k+1 evaluations of the integrand. The accuracy of the approximation increases with the order of the polynomial. In these quadrature methods, the number of evaluations increases exponentially with the dimension of the integral. There are some results from numerical analysis that reduce the number of evaluations needed for a given level of accuracy (Judd, 1998). But, in any case, the computational costs of quadrature methods increases very fast with the dimension of the problem. An accessible introduction to quadrature methods can be found in Pudney (1989).

For high dimensional problems, probability simulation methods are an attractive alternative to the costly quadrature methods. As mentioned before, these methods are based on the fact that the integral of interest represents the probability of an event in a population. This integral is then approximated by simulation rather than computing the probability of interest. In this section we review three proposals that can be helpful in understanding the advantages and limitations of probability simulation methods: the Crude Frequency Simulator, the Stern Simulator and the Geweeke-Hajivassiliou-Keane (GHK) simulator.

a) The Crude Frequency Simulation/Estimator

Lerman and Manski (1981) propose generating a pseudo-random sample of observations from the relevant population and using the relative frequency of the event in the sample to approximate the integral of interest. This simulation method is called a "Crude Frequency Simulator". This method is intuitive and helps to introduce the idea of probability simulation. The integral to compute is:

$$\Pr(\mathbf{a} < \mathbf{u} < \mathbf{b}) = \int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{u})d\mathbf{u} \qquad (1)$$

where, $\mathbf{u}$ is a random vector distributed multivariate normal with mean $\mathbf{0}$ and variance $\mathbf{W}$, g is the probability density function of the random vector $\mathbf{u}$ and $\mathbf{a}$ and $\mathbf{b}$ are vectors that define the event of interest. A crude frequency simulator can be constructed as follows:

1. Draw a pseudo-random sample[5] of R vectors from the population defined by the probability density function g($\mathbf{u}$) in (1).
2. Calculate how many observations in this sample lay in the integration interval, or in other words, compute the occurrence of the event of interest in the pseudo-random sample.
3. Approximate the probability or interest by the relative frequency of the event $\mathbf{a} < \mathbf{u} < \mathbf{b}$.

The computational cost of this estimator can be considered linear in the number of replications (R). The statistical properties of this probability simulator can be better understood by interpreting the probability as the expected value of an indicator function.

$$\Pr(\mathbf{a} < \mathbf{u} < \mathbf{b}) = \int_{-\infty}^{+\infty} D_{ab}(\mathbf{u})g(\mathbf{u})d\mathbf{u} = E[D_{ab}(\mathbf{u})] \qquad (2)$$

where D_{ab} takes the value of one if and only if $\mathbf{u}$ is in the interval ($\mathbf{a}$,$\mathbf{b}$). Using the crude frequency simulator to approximate the probability (1) can be seen as using the sample

[5] The sample is labeled as pseudo-random because the computer process used to generate it is actually deterministic. It only appears to be random. For a discussion along these lines see Davidson and Mackinnon (1993, p. 734).

mean of the indicator to approximate the expected value of the indicator. This interpretation of the probability simulator as a sample mean in random sampling to estimate a population expectation gives the simulator well known statistical properties such as strong consistency, minimum variance and unbiasedness (Lerman and Manski, 1981).

This method has two important drawbacks. First, if the true probability is smaller than one, then for any finite number of draws, the probability of zero frequency count is positive. Therefore, the frequency simulator yields consistent likelihood estimates only if both the sample size and the number of draws per observation go to infinity. From a practical point of view this means that a large number of draws is needed to obtain reasonably accurate estimates of small probabilities. Second, the frequency simulator is not a continuous function of the parameters of interest. As a consequence, the proof of consistency and asymptotic normality of the simulation/estimation needs a much more complex asymptotic theory. From a practical point of view, this discontinuity precludes the use of standard gradient methods of optimization for maximum likelihood estimation.

The difficulties of using this simulator in Maximum Simulated Likelihood spurred the search for alternative approaches. For example, McFadden (1989) proposes the "Method of Simulated Moments". In this method, the response probability function in the sample moments is replaced by the Crude Frequency Simulator. However, in the sample moments the simulated probabilities are summed over observations. In this case, simulation noise cancels out by the law of large numbers and therefore the accuracy requirements of the simulator are lower in the Method of Simulated Moments.

The performance of the crude frequency simulator method can be improved using some variance reduction techniques such as antithetic variates or control variates (Ross, 1988). The use of antithetic variates consists of introducing a negative correlation between successive pseudo random draws. In this manner, it is possible to reduce the variance of the sample mean used as a probability simulator. Control variates are random variables with analytical expectations that are positively correlated with the random variable whose

probability we want to simulate. Instead of simulating the variable of interest, the difference between that variable and the control variates is simulated, which reduces the simulation variance.

Other techniques that can improve simulation/estimation performance are the use of importance sampling and kernel smoothing. Importance sampling consists of sampling from a more convenient population. This can be done by writing the probability in (1) as:

$$\Pr(\mathbf{a} < \mathbf{u} < \mathbf{b}) = \int_{-\infty}^{+\infty} D_{ab}(\mathbf{u}) \frac{g(\mathbf{u})}{h(\mathbf{u})} h(\mathbf{u}) d\mathbf{u} . \tag{3}$$

Now, we sample from the population defined by the probability density function h(**u**) and compute the relative frequency of:

$$D_{ab}(\mathbf{u}) \frac{g(\mathbf{u})}{h(\mathbf{u})} . \tag{4}$$

The advantages of this method are that sampling from h can be easier than sampling from g. Furthermore, when the ratio g(**u**)/h(**u**) is relatively constant the sampling variance of the importance sampling simulator tends to be small.[6]

The use of a smoothed kernel consists of replacing the indicator function D_{ab} in the crude frequency simulator with a smooth function K_{ab} with the following properties:

$$\begin{aligned} &K_{ab} = K\left(\frac{u - b}{w}\right) - K\left(\frac{u - a}{w}\right) \\ &K(-\infty) = 1 \\ &K(+\infty) = 0 \end{aligned} \tag{5}$$

where w is a window parameter that makes K go to the indicator function as w goes to zero.

Importance sampling and kernel smoothing techniques have proved to be fruitful paths for improving the performance of a probability simulator. In fact, the probability

[6] See Stern (1997) and Vijverberg (1997) for a detailed discussion.

simulators discussed in the next two sections can be seen as straightforward applications of these techniques.

b) The Stern Simulation/Estimator

The "Crude Frequency Simulator" was improved in several subsequent papers. Stern (1992) proposes an unbiased and smooth simulator of a multivariate probability. The smoothness of the probability simulator is important for reducing the computational cost for a given a level of accuracy. Borsh-Saupan and Hajivassiliou (1993) provide a heuristic explanation for the importance of smoothness. With only one draw, the crude frequency simulator can map the parameters of the population in two outcomes: zero and one. Only the increase in the number of draws serves to improve the smoothness of the simulator. In contrast, each draw of a smooth simulator can map the value of the parameters in the [0,1] interval, and no additional smoothing is necessary. So, smoothness reduces the number of draws needed for a given level of accuracy. Moreover, smoothness permits the application of standard optimization algorithms and standard asymptotic theory.

The Stern estimator is designed to compute probabilities from a multivariate normal population. Unlike the crude frequency simulator, this estimator is distribution specific at least to the extent that it needs the additive property of the normal distribution family. The simulator is constructed following these steps. First, assume $\mathbf{u}$ is distributed multivariate normal with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Omega}$. Second, partition $\mathbf{u}$ into two components. One of them is made of independent random variables and the other has a covariance matrix as small as possible in the positive definite sense, that is:

$$\begin{aligned} &\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 \\ &\mathbf{w}_1 \sim N(0, D) \quad \mathbf{w}_2 \sim N(\mu, \Omega - D) \end{aligned} \tag{6}$$

where $D=\text{diag}(d_1,...,d_n)$. The probability of the event can be written as:

$$\Pr(a < u < b) = \Pr(a < w_1 + w_2 < b) = \Pr(a - w_2 < w_1 < b - w_2) \tag{7}$$

where:

$$\Pr(\ a - w_2 < \ \ w_1 < b - w_2) = \int H(w_2) f(w_2) dw_2 \tag{8}$$

and:

$$H(w_2) = \prod_{i=1}^{k} \Phi\left[\left(\frac{b_i - w_{2i}}{d_i}\right) - \Phi\left(\frac{a_i - w_{2i}}{d_i}\right)\right]. \tag{9}$$

H can be simulated by drawing $\mathbf{w}_2$ from a normal distribution with mean $\boldsymbol{\mu}$ and variance $(\mathbf{W} - \mathbf{D})$. The probability of interest can be simulated as:

$$\Pr(\mathbf{a} < \mathbf{u} < \mathbf{b}) \approx \sum_{r=1}^{R} \frac{H(w_{2r})}{R} \tag{10}$$

where R is the number of draws from the pseudo random vector.

The Stern simulator can be interpreted as a kernel smoothed variant of the crude frequency simulator where H is the kernel. It can be interpreted as well as an importance sampling variant where $f(\mathbf{w}_2)$ is the importance sampling probability density function.

This method of simulation gives exact results when the matrix of variance-covariance is diagonal and $\mathbf{w}_2 = \mathbf{0}$. However, the quality of the approximation worsens with increasing dependence between the random variables in the vector **u**, i.e., when the matrix contains large off-diagonal terms (Borsh-Saupan and Hajivassiliou, 1993). More recently, Stern (1999) finds cases in which his simulator outperforms the GHK simulator.

c) The GHK Simulation/Estimator

Geweke (1989) and Borsh-Saupan and Hajivassiliou (1993) proposed the "Smooth Recursive Conditioning" (SRC) simulator. In more recent papers the SRC simulator is labelled as the Geweke-Hajivassiliou-Keane simulator (GHK). The literature on probability simulation has expanded in the last few years. Hajivassiliou et al. (1996) analyze the properties of a number of available probability simulators and find that the GHK probability simulator outperforms all other methods by keeping a good balance between accuracy and

computational costs. This probability simulator is relatively simple to program for different dimensions of the problem and it can be generalized to any distributional assumption. As a result, it is convenient for use in empirical applications (Hajivassiliou, 1993; Feenberg and Skinner, 1994; Chen and Cosslett, 1998). For this reason, it is worth a detailed explanation of the working of this probability simulation method.

The GHK simulates the value of the integral in (1). The starting point is that:

$$\Pr(\mathbf{a} < \mathbf{u} < \mathbf{b}) = \Pr(\mathbf{a} < \mathbf{Le} < \mathbf{b}) \tag{11}$$

where, **L** is the lower triangular Cholesky factor of $\mathbf{W}$, such that $\mathbf{LL'}=\mathbf{W}$, and **e** is a random vector of independent standard normal variables. The right hand side of expression (11) is easier to simulate than the probability in the left hand side due to the triangular structure of the constrains defined by **Le**. The intervals defining the event in the right hand side of expression (11) can be written as:

$$\begin{aligned} a_1 &< l_{11}e_1 < b_1 \\ a_2 &< l_{12}e_1 + l_{22}e_2 < b_2 \\ &\dots \\ a_n &< l_{1n}e_1 + \dots + l_{nn}e_n < b_n \end{aligned} \tag{12}$$

where l_{ij}, a_i and b_i are the corresponding elements of L, **a** and **b**. For notational convenience, arranging terms in (12) the event in the right hand side of equations (11) can be decomposed into the following events:

$$\begin{aligned} \frac{a_1}{l_{11}} < e_1 < \frac{b_1}{l_{11}} \\ \frac{a_2 - l_{12}e_1}{l_{22}} < e_2 < \frac{b_2 - l_{12}e_1}{l_{22}} \\ \frac{a_3 - l_{13}e_1 - l_{23}e_2}{l_{33}} < e_3 < \frac{b_3 - l_{13}e_1 - l_{23}e_2}{l_{33}} \\ \dots \\ \frac{a_n - l_{1n}e_1 - l_{2n}e_2 - \dots - l_{n\text{-}1n}e_{n\text{-}1}}{l_{nn}} < e_n < \frac{b_n - l_{1n}e_1 - l_{2n}e_2 - \dots - l_{n\text{-}1n}e_{n\text{-}1}}{l_{nn}} \end{aligned} \tag{13}$$

Expression (13) shows the recursive nature of the constraints that affect the random vector **e**. As a result, the probability of interest can be written as:

$$\Pr(a < Le < b) = \Pr(A_1)\Pr(A_2|A_1)\Pr(A_3|A_1, A_2)...\Pr(A_n|A_1,...,A_{n-1}). \quad (14)$$

The idea behind the GHK simulator is that while expression (14) can be difficult to calculate, it is rather easy to simulate. The GHK simulator can be written as:

$$\tilde{\Pr}(a < Le < b) = \frac{1}{R}\sum_{r=1}^{R}\Pr(A_1)\Pr(A_2|e_{1r})\Pr(A_3|e_{1r}, e_{2r})...\Pr(A_n|e_{1r},...,e_{n-1r}) \quad (15)$$

where the e_{ir}´s are drawn sequentially from independent standard normal distributions truncated by expression (13) and R is the number of simulations. Once the e_{ir}´s are drawn the terms in the product are calculated as:

$$\Pr(A_i|\ e_{1r}, e_{2r},..., e_{i-1r}) = \Phi\left(\frac{b_i - l_{i1}e_{1r} - ...l_{i-1i}e_{i-1r}}{l_{ii}}\right) - \Phi\left(\frac{a_i - l_{i1}e_{1r} - ...l_{i-1i}e_{i-1r}}{l_{ii}}\right) \quad (16)$$

where Φ is the cumulative distribution function of a standard normal distribution function[7].

The truncated random variables e_i can be generated smoothly using the integral transform theorem (Ross, 1988, p. 389). For example, If y is distributed as a standard normal subject to the constraint (c < y < d) the cumulative distribution function of y can be written as:

$$G(y) = \frac{\Phi(y) - \Phi(c)}{\Phi(d) - \Phi(c)} = z \quad (17)$$

where z is distributed uniformly in the interval [0,1]. From (17) we can obtain an expression that relates the truncated random variable y with the uniformly distributed random variable z:

[7] Hajivassiliou (1998) provides a complete implementation of the GHK simulator in the GAUSS computer matrix language.

$$y = \Phi^{-1}\Big(\big(\Phi(d) - \Phi(c)\big)z + \Phi(c)\Big). \tag{18}$$

The GHK simulator can be seen as a member of the family of importance sampling probability simulators (Vijverberg, 1997). This approach offers a path for searching for better probability simulators.

3. Simulation Noise in Maximum Simulated Likelihood

Borsch-Saupan and Hajivassiliou (1993) proved that the probability simulator in (15) is an unbiased estimator of the true probability. However, the logarithm of the simulator is not an unbiased estimator of the logarithm of the true probability and therefore the simulated likelihood function is not unbiased. As a consequence, the estimates of the parameters by Maximum Simulated Likelihood are biased due to simulation noise. The intuition behind this result is that simulation noise cancels out when adding up the likelihood contributions or its derivatives. However, this canceling does not occur with the logarithm (or any nonlinear function) of the simulated value.

Hajivassiliou (1997) proposes a test for bias generated by simulation noise in MSL estimation. The null hypothesis of the test and its alternative can be written as:

$$\begin{aligned} H_0&: E\left[\frac{\partial \ln L^*(y,\tilde{\theta})}{\partial \theta}\right] = 0 \\ H_1&: E\left[\frac{\partial \ln L^*(y,\tilde{\theta})}{\partial \theta}\right] \neq 0 \end{aligned} \tag{19}$$

where, L* is a simulated likelihood function, y is a vector of observations and $\tilde{\theta}$ is the MSL estimate of the parameter of interest θ. The rejection of the null hypothesis is interpreted as evidence of bias due to simulation noise. The test relies on simulation of the data generating

process $y(\tilde{\theta})$ to calculate the empirical mean (m) and variance (v) of the score variable in (19). This simulation has to be independent of the simulation used for computation of the GHK simulator. The test can be written as:

$$w = NSm'v^{-1}m \tag{20}$$

where N is the number of observations in the sample and S the number of simulations per observation. Under the null hypothesis, w is distributed chi-square with degrees of freedom equal to the number of parameters.

The bias created by simulation noise decreases with the number of pseudo random draws used for calculating the GHK simulator. It is very important to distinguish between the number of simulations used to calculate the GHK simulator (denoted by R) and the number of simulation used to compute the simulation noise test (denoted by S). If the proposed test permits the rejection the null hypothesis of negligible bias due to simulation noise, increase the number of simulations in the GHK simulator (R) until an acceptable value of w is obtained. Then, re-estimate the model with the number of simulations that generates the result of negligible simulation noise bias.

4. Conclusions

There are a number of important economic models where empirical estimation requires fast computation of high dimensional definite integrals. Recent developments suggest that probability simulation methods are an attractive approach for the estimation of such models. This brief survey tries to help in moving these probability simulator methods from the theoretical literature to applied work. The paper surveys key literature on probability simulation methods in an intuitive but rigorous manner so that the main ideas behind these techniques can be grasped easily by the average researcher. In particular, the GHK simulator is easy to program and has good statistical/estimation properties. For these reasons, it is an ideal candidate to be in the toolbox of any applied researcher in economics.

5. References

Börsh-Saupan, A. and V. Hajivassiliou. "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models". *Journal of Econometrics*, 58(1993):347-68.

Chen, H.Z. and R.C. Cosslett. "Environmental Quality Preference and Benefit Estimation in Multinomial Probit Models: A Simulation Approach", *American Journal of Agricultural Economics*, 80(1998): 512-520.

Davidson. R. and J.G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, 1993.

Feenberg, D. and J. Skinner. "The Risk and Duration of Catastrophic Health Care Expenditures". *The Review of Economics and Statistics*, 4(1994): 633-647.

Geweke, J. "Efficient Simulation from the Multivariate Normal Distribution Subject to Linear Inequality Constraints and the Evaluation of Constraint Probabilities". Duke University, Durham, N.C: Mimeo, 1989.

Gourieroux, C. and A. Monfort, *Simulation-Based Econometric Methods*, CORE lectures, Oxford, 1996.

Hajivassiliou, W. A., " A Simulation Estimation Analysis of the External Debt Crises of Developing Countries", Cowles Foundation for Research in Economics, Working paper, 1993.

Hajivassiliou, W.A., " Some Practical Issues in Maximum Simulated Maximum Likelihood", Forthcoming in *Simulation-Based Inference in Econometrics: Methods and Applications*, Roberto Mariano, Melvyin Weeks and Til Schuermann (eds), Cambridge, University Press, 1999.

Hajivassiliou, V., D. McFadden and P. Ruud. "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational results". *Journal of Econometrics*, 72(1996):85-134.

Judd, K. *Numerical Methods in Economics*. MIT Presss, 1998.

Lee, L.F. and M. M. Pitt. "Microeconometric Demand Systems with Binding Nonnegativity Constraints: The Dual Approach". *Econometrica,* 54(1986):1237-42.

Lerman, S. and C. Manski. "On the Use of Simulated Frequencies to Approximate Choice

Probabilities". in: *Structural Analysis of Discrete Data with Econometric Applications*. Edited by Manski C.F. D. and McFadden. Cambridge-Massachusetts: The MIT press, 1981.

Pudney, S. *Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes.* New York and Oxford : Basil-Blackwell, 1989.

Ross, S.M., *A First Course in Probability*, Macmillan, New York, 1988.

Stern, S. "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models". *Econometrica,* 60(1992): 943-52.

Stern, S., "Simulation-Based Estimation", Journal of Economic Literature, 25 (1997): 2006-2039.

Stern, S., "Simulation and Estimation: Motivation and Methods". Forthcoming in *Simulation-Based Inference in Econometrics: Methods and Applications,* Roberto Mariano, Melvyin Weeks and Til Schuermann (eds), Cambridge, University Press, 1999.

Vijverberg, W.P.M., " Monte Carlo Evaluation of Multivariate Normal Probabilities", *Journal of Econometrics*, 76 (1997): 281-307.

Wales, T.J. and A. D. Woodland. "Estimation of Consumer Demand Systems with Binding Non-Negative Constraints". *Journal of Econometrics,* 21(1983):263-85.