# DEPARTMENT OF ECONOMICS

# WORKING PAPER SERIES

2002-13

# McMASTER UNIVERSITY

Department of Economics
Kenneth Taylor Hall 426
1280 Main Street West
Hamilton, Ontario, Canada
L8S 4M4

http://socserv.socsci.mcmaster.ca/~econ/

# USING SIMULATION-BASED INFERENCE WITH PANEL DATA IN HEALTH ECONOMICS

Paul Contoyannis[a], Andrew M. Jones[b] and Roberto Leon-Gonzalez[c]

[a]*Centre for Health Economics and Policy Analysis, McMaster University*
[b]*Department of Economics and Related Studies,University of York*
[c]*Centre for Health Economics & Department of Economics and Related Studies,University of York,*

http://www.york.ac.uk/res/herc/yshe
September 2002

## ABSTRACT

Panel datasets provide a rich source of information for health economists, offering the scope to control for individual heterogeneity and to model the dynamics of individual behaviour. However the qualitative or categorical measures of outcome often used in health economics create special problems for estimating econometric models. Allowing a flexible specification of the autocorrelation induced by individual heterogeneity leads to models involving higher order integrals that cannot be handled by conventional numerical methods. The dramatic growth in computing power over recent years has been accompanied by the development of simulation-based estimators that solve this problem. This review uses binary choice models to show what can be done with conventional methods and how the range of models can be expanded by using simulation methods. Practical applications of the methods are illustrated using data on health from the British Household Panel Survey (BHPS).

**Keywords:** econometrics, panel data, simulation methods, determinants of health

**Acknowledgements:** This paper is based on the preconference session 'Using panel data econometrics in health economics' presented at the 3rd iHEA Conference, University of York, 21st-25th July 2001. Data from the *British Household Panel Survey (BHPS)* were supplied by the ESRC Data Archive. Neither the original collectors of the data nor the Archive bear any responsibility for the analysis or interpretations presented here. Funding was provided by the ESRC award no: R000238169 – 'Simulation-based econometric approaches to investigating the interaction of lifestyle and health'. The authors would like to thank Nigel Rice for his advice and two anonymous referees for their comments.

## Introduction

Panel datasets, such as the European Community Household Panel (ECHP) and the U.S. PSID and Monitoring the Future panels, provide a rich source of information for health economists. Panel data offer the scope to control for individual heterogeneity and to model the dynamics of individual behaviour. However the measures of outcome used in health economics are often qualitative or categorical. These create special problems for estimating econometric models. Allowing a flexible specification of the autocorrelation induced by individual heterogeneity leads to models involving higher order integrals that cannot be handled by conventional numerical methods. The dramatic growth in computing power over recent years has been accompanied by the development of simulation estimators that solve this problem. This review uses binary choice models to show what can be done with conventional methods and how the range of models can be expanded by using simulation methods. Practical applications of the methods are illustrated using data on health from the British Household Panel Survey (the BHPS).

Section 1 gives an overview of binary choice models for panel data and introduces our empirical application to BHPS data for a binary measure of health. It discusses the interpretation of individual effects in panel data models and shows how these can be modelled using the random effects probit model, the conditional logit model and by parameterising the individual effect. Extensions of the random effects model, to allow for serial correlation, can be dealt with by simulation-based inference. Section 2 introduces classical simulation methods. These are designed to approximate higher order integrals and they include the GHK simulator for the truncated multivariate normal distribution. We focus on a particular method of estimation, Maximum Simulated Likelihood (MSL). We present some empirical results and use these to discuss issues in that arise in practical applications of MSL. The section concludes with a brief overview of other methods of estimation (MSM, MSS). Section 3 moves to Bayesian MCMC methods. It begins with an introduction to the Bayesian approach to inference before introducing the concept of Markov Chain Monte Carlo (MCMC). Implementation of MCMC involves the use of Gibbs sampling and Metropolis-Hastings algorithms, along with the use of data augmentation to deal with latent variables. This section concludes with an overview of convergence analysis and methods for model selection and testing.

# 1. Binary choice models for panel data

## 1.1 A brief introduction to our model

To illustrate the methods reviewed in this paper we use a panel data model for a binary measure of health applied to data drawn from the British Household Panel Survey (BHPS). The BHPS is a longitudinal survey of private households in Great Britain, with the same respondents questioned each year. The survey contains data on socio-demographic, income and health variables. It is an annual survey of each adult household member (aged 16 and over). The survey was designed to be a nationally representative sample of over 5,000 households, giving approximately 10,000 individual interviewees. The first wave was carried out between 1st September 1990 and 30th April 1991.

We present two sets of results:

1. The first are based on data for the binary measure of health problems that are artificially generated from our BHPS sample, using models with known parameter values. This provides us with a benchmark to illustrate how the different methods of estimation perform against the "true" values of the parameters. These results are not intended to provide a comprehensive Monte Carlo experiment but they do show how the different estimators perform with our dataset and they provide a context for the empirical application.
2. We also present results for the actual data on health problems, to show how the methods perform in a real empirical application.

Our model applies to a binary dependent variable ("does health limit your daily activities?"). There are repeated measurements for each wave (t=1,…., T) for a sample of n individuals (i=1,…..,n), and the binary dependent variable $y_{it}$ can be modelled in terms of a continuous latent variable $y^*_{it}$,

$$(1) \qquad y_{it} = 1(y^*_{it} > 0) = 1(X'_{it}\beta + u_{it} > 0)$$

where 1(.) is a binary indicator function. In our empirical application with actual data, X includes variables to capture "permanent" and "transitory" income, measured by the mean of household

income across all waves of the panel and deviations around that mean respectively, along with marital status, education and household composition. In our empirical application to the BHPS we restrict the analysis to a sub-sample of 2,715 men (full details of the sample and variables are given in Contoyannis, Jones and Rice [1]).

The error term $u_{it}$ could be allowed to be freely correlated over time or the correlation structure could be restricted. A common specification is the error components model which splits the error into a time-invariant individual random effect (RE), $\alpha_i$, and a time-varying idiosyncratic random error, $\varepsilon_{it}$,

$$(2) \qquad y_{it} = 1(y^*_{it} > 0) = 1(X'_{it}\beta + \alpha_i + \varepsilon_{it} > 0)$$

The idiosyncratic error term could be autocorrelated, for example following an AR(1) process, $\varepsilon_{it}=\rho\varepsilon_{it-1} + \eta_{it}$, or it could be independent over t (giving the random effects model). The simplest possible specification is to assume that the $u_{it}$ are independent over t.

## 1.2 Individual effects in panel data

To understand the role of individual effects in panel data models, consider the standard linear panel data regression model, in which there are repeated measurements (t=1,…., T) for a sample of n individuals (i=1,…..,n),

$$(3) \qquad y_{it} = X'_{it}\beta + u_{it} = X'_{it}\beta + \alpha_i + \varepsilon_{it}$$

The presence of $\alpha_i$ implies clustering within individuals so that a random effects specification can improve the efficiency of the estimates of $\beta$. This stems from the structure imposed on the variance-covariance matrix of the error term,

$$(4) \qquad Var[u_{it}] = E[u_{it}u_{is}] = \sigma_\alpha^2 + \sigma_\varepsilon^2, \quad t=s$$

$$E[u_{it}u_{is}] = \sigma_\alpha^2, \quad t\neq s$$

These efficiency gains can be exploited by using (4) to construct a generalised least squares (GLS) estimator.

Consistency of the GLS estimator rests on the assumption that the error term is independent of the regressors. Failure to account for correlation between the unobservable individual effects ($\alpha$) and the regressors (X) will lead to inconsistent estimates of the $\beta$s. The least squares dummy variable approach (LSDV) gets around this by conditioning on the individual effects, including a dummy variable for each individual, but this may be prohibitive if there are a large number of cross section observations. Alternatively, the individual effects can be swept from the equation by transforming variables into deviations from their within-group means. Applying least squares to the transformed equation gives the covariance or within-groups estimator of $\beta$ (CV). Similarly, the model could be estimated in first differences to eliminate the time-invariant individual effects. Identification of $\beta$ rests on there being sufficient variation within groups. In practice, fixed effects may only work well when there are many observations and much variation within groups.

Now consider a nonlinear model, for example the binary choice model based on the latent variable specification in Equation (2). Assume that the distribution of $\varepsilon_{it}$ is symmetric with distribution function F(.). Then,

$$(5) \qquad P(y_{it}=1) = P(\varepsilon_{it} > -X'_{it}\beta - \alpha_i) = F(X'_{it}\beta + \alpha_i)$$

This illustrates the so-called problem of incidental parameters. As $n \to \infty$ the number of parameters to be estimated ($\beta$, $\alpha_i$) also grows. In linear models the estimators $\hat{\beta}$ and $\hat{\alpha}$ are asymptotically independent, which means that taking mean deviations or differencing the data allows the derivation of estimators for $\beta$ whose limits do not depend on $\hat{\alpha}$. In general, this is not possible in nonlinear models and the inconsistency of estimates of $\alpha$ carries over into the estimates of $\beta$. Setting the incidental parameter problem aside, the fixed effect probit model can be estimated by including a dummy variable for each individual. Heckman [2] presents Monte Carlo evidence that suggests that the small sample bias in the estimates of $\beta$ is relatively small for values of T of 8 and over. More recently, Greene [3] has championed the use of this 'brute force' approach to fixed effects estimation of nonlinear models.

## 1.3 Random effects probit model

Assuming that $\alpha$ and $\varepsilon$ are normally distributed and independent of $X$ gives the random effects probit model (REP). In this case $\alpha$ can be integrated out to give the sample log-likelihood function,

$$(6) \qquad lnL = \sum_{i=1}^{n} \{ ln \int_{-\infty}^{+\infty} \prod_{t=1}^{T} (\Phi[d_{it}(X'_{it}\beta + \alpha)] ) f(\alpha) d\alpha \}$$

where $d_{it} = 2y_{it} - 1$. This expression contains a univariate integral which can be approximated by Gauss-Hermite quadrature. Assuming $\alpha \sim N(0,\sigma_\alpha^2)$, the contribution of each individual to the sample likelihood function is,

$$(7) \qquad L_i = \int_{-\infty}^{+\infty} (1/\sqrt{2\pi\sigma_\alpha^2}) \exp(-\alpha^2/2\sigma_\alpha^2) \{ g(\alpha) \} d\alpha ,$$

where $g(\alpha) = \prod_{t=1}^{T} \Phi[d_{it}(X'_{it}\beta + \alpha)]$. Use the change of variables, $\alpha = (\sqrt{2\sigma_\alpha^2})z$, to give,

$$(8) \qquad L_i = (1/\sqrt{\pi}) \int_{-\infty}^{+\infty} \exp(-z^2) \{ g((\sqrt{2\sigma_\alpha^2})z) \} dz$$

As it takes the generic form $\int_{-\infty}^{+\infty} \exp(-z^2) f(z) dz$, this expression is suitable for Gauss-hermite quadrature and can be approximated as a weighted sum,

$$(9) \qquad L_i \approx (1/\sqrt{\pi}) \sum_{j=1}^{m} w_j \, g( (\sqrt{2\sigma_\alpha^2})a_j)$$

where the weights ($w_j$) and abscissae ($a_j$) are tabulated in standard mathematical references and $m$ is the number of nodes or quadrature points (see e.g., Butler and Moffitt [4], Pudney [5]).

## 1.4 Extensions and alternatives for the random effects probit

The random effects probit model has two important limitations: it relies on the assumptions that the error components have a normal distribution and that errors are not correlated with the regressors. Normality can be relaxed by using a finite mixture model. The possibility of correlated effects can be dealt with by using conditional (fixed effects) approaches or by parameterising the effect.

*The finite mixture model*

Deb [6] applies a random effects probit model in which the distribution of the individual effect is approximated by a discrete density. This is an example of a finite mixture model (see e.g., Heckman and Singer [7]). In this case the sample log-likelihood is approximated by,

$$(10) \qquad lnL = \sum_{i=1}^{n} ln \left( \sum_{j=1}^{C} \pi_j \{ \prod_{t=1}^{T} \Phi[d_{it}(X'_{it}\beta + \alpha_j)] \} \right) , 0 \le \pi_j \le 1 , \sum_{j=1}^{C} \pi_j = 1$$

Deb uses Monte Carlo experiments to assess the small sample properties of the estimator. These show that only 3-4 points of support are required for the discrete density to mimic normal and chi-square densities sufficiently well so as to provide approximately unbiased estimates of the structural parameters and the variance of the individual effect.

*The conditional logit estimator.*

The conditional logit estimator uses the fact that $\sum_t y_{it}$ is a sufficient statistic for $\alpha_i$ (see e.g., Chamberlain [8]). This means that conditioning on $\sum_t y_{it}$ allows a consistent estimator for $\beta$ to be derived. Using the logistic function,

$$(11) \qquad P(y_{it} = 1) = F(X'_{it}\beta + \alpha_i) = \exp(X'_{it}\beta + \alpha_i)/(1 + \exp(X'_{it}\beta + \alpha_i))$$

it is possible to show that,

$$(12) \qquad P[(0,1) \mid (0,1) \text{ or } (1,0)] = \exp((X_{i2} - X_{i1})'\beta)/(1 + \exp((X_{i2} - X_{i1})'\beta))$$

This implies that a standard logit model can be applied to differenced data and the individual effect is swept out. In practice, conditioning on those observations that make a transition – (0,1) or (1,0) – and discarding those that do not – (0,0) or (1,1) – means that identification of the models relies on those observations where the dependent variable changes over time.

*Parameterising the individual effect*

Another approach to dealing with individual effects that are correlated with the regressors is to specify $E(\alpha|X)$ directly. For example, in dealing with a random effects probit model Chamberlain [8,9] suggests using,

$$\alpha_i = X'_i\alpha + u_i \quad , \quad u_i \sim iid\ N(0, \sigma^2) \tag{13}$$

where $X_i=( X_{i1},....,X_{iT})$, the values of the regressors for every wave of the panel, and $\alpha=(\alpha_1,....,\alpha_T)$. Then, by substituting, the distribution of $y_{it}$ conditional on X but marginal to $\alpha_i$ has the probit form,

$$P(y_{it}=1) = \Phi[(1+\sigma^2)^{-1/2}(X'_{it}\beta + X'_i\alpha)] \tag{14}$$

The model could be estimated as a random effects probit to retrieve the parameters of interest $(\beta,\sigma)$. Recently Wooldridge [10] has shown that this approach can also be applied in a random effects probit model with state dependence. In this case the initial values of the dependent variable are also included in Equation (13) in order to deal with the problem that the initial conditions are correlated with the individual effect (the so-called 'initial conditions' problem, see Heckman [2]).

## 1.5 Empirical applications

Our empirical results are in two parts. First, results are presented for simulated data and, then, for the actual BHPS data. The simulated data are designed to match some of the key features of the actual BHPS sample and to show how the different estimators perform against the true

values of the parameters. The actual data for the BHPS income variable is used as a starting point from which to generate observations for a binary dependent variable. This is done for three different data generating processes:

Model 1: Based on a RE+AR(1) specification of the error term.
Model 2: Again with RE+AR(1) but building in a correlation between the error term and the income variable, giving a "correlated effects" specification.
Model 3: The final specification uses uncorrelated effects but allows for a more general covariance structure on the error term. This is simulated using a RE+ARMA(2,1) process.
The full specification of the simulated models is shown in Box 1.

Table 1(a) shows the results of the pooled probit and random effects probit estimators applied to simulated data from the three models. Table 1(b) shows a summary of results for the pooled and random effects probit models applied to our actual binary measure of health problems in the BHPS (for brevity we only report the coefficients on the income variables). The pooled probit model treats the data as a single cross section and ignores the fact that there are repeated observations for each individual. These estimates provide a useful benchmark for the random effects model. It has been shown that the pooled probit (pseudo-) ML estimator gives consistent estimates of the $\beta$s, irrespective of whether the assumed error structure is correct (Robinson [11]). Of course the pooled probit model does not provide an estimate of $\sigma_\alpha^2$ and, therefore, information about the structure of the error term and the relative importance of the individual effect.

**BOX 1: Data generating process for simulated data:**

$$y_{it}^{*} = c + \beta \times \ln(income)_{it} + \alpha_i + \varepsilon_{it}$$

$$y_{it} = 1 \quad \text{iff} \quad y_{it}^{*} > 0$$

$$\alpha_i \sim N(0,\sigma_{\alpha}^{2}) \qquad \sigma_{\alpha}^{2} = 0.8$$

$$c = 1 \quad \beta = 1$$

$$\sigma_{v}^{2} = \sigma_{\alpha_i}^{2} + \sigma_{\varepsilon_{it}}^{2}$$

ln(income) is the log of transitory income in deviations with respect to its mean.

**MODEL 1: PROBIT WITH RANDOM EFFECTS PLUS AR(1).**

$$E\left(\alpha_i, \ln(income)_{it}\right) = 0$$

$$\varepsilon_{it} = \rho\varepsilon_{i(t-1)} + \eta_{it}$$

$$\eta_{it} \sim N(0,0.102)$$

$$\rho = 0.7$$

$$Var(\alpha_i + \varepsilon_{it}) = 1$$

The implied variance-covariance matrix for $(\varepsilon_{i1} + \alpha_i, \varepsilon_{i2} + \alpha_i, \ldots, \varepsilon_{iT} + \alpha_i)$ is:

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 0.94 | 0.898 | 0.8686 | 0.84802 | 0.833614 | 0.82353 |
| 0.94 | 1 | 0.94 | 0.898 | 0.8686 | 0.84802 | 0.833614 |
| 0.898 | 0.94 | 1 | 0.94 | 0.898 | 0.8686 | 0.84802 |
| 0.8686 | 0.898 | 0.94 | 1 | 0.94 | 0.898 | 0.8686 |
| 0.84802 | 0.8686 | 0.898 | 0.94 | 1 | 0.94 | 0.898 |
| 0.833614 | 0.84802 | 0.8686 | 0.898 | 0.94 | 1 | 0.94 |
| 0.82353 | 0.833614 | 0.84802 | 0.8686 | 0.898 | 0.94 | 1 |

**MODEL 2: PROBIT WITH RANDOM EFFECTS PLUS AR(1) INDIVIDUAL EFFECTS CORRELATED WITH THE REGRESSORS.**

$$E\left(\alpha_i | \ln(income)_{it}\right) = 8.1 \times \ln(income)_{i1} + 8.2 \times \ln(income)_{i2} + \ldots + 8.7 \times \ln(income)_{i7}$$

$$\varepsilon_{it} = \rho\varepsilon_{i(t-1)} + \eta_{it}$$

$\eta_{it} \sim N(0,0.102)$

$\rho = 0.7$

$Var(\alpha_i + \varepsilon_{it}) = 1$

**MODEL 3: PROBIT WITH RANDOM EFFECTS PLUS ARMA(2,1).**

$E\left(\alpha_i, \ln(income)_{it}\right) = 0$

$\varepsilon_{it} = \rho_1 \varepsilon_{i(t-1)} + \rho_2 \varepsilon_{i(t-2)} + \eta_{it} - \theta_1 \eta_{i(t-1)}$

$\eta_{it} \sim N(0,0.0416)$

$\rho_1 = 0.7 \quad \rho_1 = 0.25 \quad \theta_1 = 0.8$

$Var(\alpha_i + \varepsilon_{it}) = 1$

The implied Variance-covariance matrix of $(\varepsilon_{i1} + \alpha_i, \varepsilon_{i2} + \alpha_i, \ldots, \varepsilon_{iT} + \alpha_i)$ is:

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 0.928455 | 0.939919 | 0.930057 | 0.92602 | 0.920728 | 0.916014 |
| 0.928455 | 1 | 0.928455 | 0.939919 | 0.930057 | 0.92602 | 0.920728 |
| 0.939919 | 0.928455 | 1 | 0.928455 | 0.939919 | 0.930057 | 0.92602 |
| 0.930057 | 0.939919 | 0.928455 | 1 | 0.928455 | 0.939919 | 0.930057 |
| 0.92602 | 0.930057 | 0.939919 | 0.928455 | 1 | 0.928455 | 0.939919 |
| 0.920728 | 0.92602 | 0.930057 | 0.939919 | 0.928455 | 1 | 0.928455 |
| 0.916014 | 0.920728 | 0.92602 | 0.930057 | 0.939919 | 0.928455 | 1 |

Note that due to the MA term, the lag 2 correlation is larger than the lag 1 correlation. Correlations diminish slowly after the second lag.

Table 1(a): Estimates for the pooled (INP) and random effects (REP, with 30 point quadrature) probits: simulated data.

| | Model 1: RE+AR(1) | | Model 2: Correlated effects | | Model 3: RE+ARMA(2,1) | |
|---|---|---|---|---|---|---|
| | INP | REP | INP | REP | INP | REP |
| $\beta$ (=1) | 1.021 | 1.023 | 0.795 | 0.376 | 1.002 | 1.080 |
| | (0.031) | (0.036) | (0.010) | (0.018) | (0.011) | (0.030) |
| $\sigma_\alpha^2$ (= 0.8) | | 0.888 | | 0.964 | | 0.917 |
| | | (0.007) | | (0.002) | | (0.004) |

Table 1(b): Estimates for the pooled and random effects probits: actual data.

| | Pooled probit | Random effects probit (24 point quadrature) |
|---|---|---|
| Ln('permanent income') | -0.573 | -0.526 |
| | (0.061) | (0.048) |
| Ln('transitory income') | -0.115 | -0.053 |
| | (0.030) | (0.023) |
| $\sigma_\alpha^2$ | - | 0.784 |
| | | (0.012) |
| LnL | -6263.5 | -4291.2 |

Table 1(a) shows that, as expected, the pooled probit gives an estimate of $\beta$ that is very close to the true value in models 1 and 3, despite ignoring the autocorrelation in the error term. However the presence of correlated effects, in model 2, means that both the pooled probit and the random effects probit are inconsistent estimators of $\beta$; in this example there is serious downwards bias in the estimates. This has important implications for the empirical application with actual data. If the income variables are endogenous, due to correlation between income and the individual effect, estimates of the income effects based on the assumption of uncorrelated effects may be seriously biased. This possibility cannot be ruled out and the illustrative empirical estimates should be treated with caution.

The pooled probit specification does not provide estimates of the variance of the individual effect $\sigma_\alpha^2$. The estimates of $\sigma_\alpha^2$ from the random effects probit show some upwards bias. This is

because the estimated values are picking up the AR(1) and ARMA(2,1) processes and the estimated values give an average of the implied covariance terms shown in Box 1. In other words, the estimate of $\sigma_\alpha^2$ is not only picking up the variance of the individual effect but also persistence in the error term $\varepsilon$. For model 1 the estimate lies between the lag 1 covariance of 0.94 and the lag 6 covariance of 0.82. For model 3 it lies between the lag 2 covariance of 0.939 and the lag 6 covariance of 0.916.

For the actual data, in Table 1(b), the income effects are negative, suggesting that those with higher household income are less likely to report limiting health problems. The estimates of the coefficient on the log of permanent income are quite similar for the pooled and random effects probits, although the size of the permanent effect relative to the transitory effect is smaller in the pooled probit. The estimate of the variance of the individual effect in the random effects specification, $\sigma_\alpha^2$, is 0.784. Since the overall error variance has to be set equal to one in order to identify the probit model, this can also be interpreted as the proportion of the overall error variance that is explained by the time invariant individual effect, indicating a high degree of persistence in health problems.

## 2. Classical simulation methods

### 2.1 Simulation-based estimation

The random effects probit model only involves a univariate integral. More complex models, for example where the error term $\varepsilon_{it}$ is assumed to follow an AR(1) process lead to sample log-likelihood functions that involve higher order integrals. Monte Carlo simulation techniques can be used to deal with the computational intractability of nonlinear models, such as the panel probit model and the multinomial probit. Popular methods of simulation-based inference include classical Maximum Simulated Likelihood (MSL) estimation, and Bayesian Markov Chain Monte Carlo (MCMC) estimation. This section introduces the classical approach.

Recall that the general version of our model is,

$$(15) \qquad y_{it} = 1(y^*_{it} > 0) = 1(X'_{it}\beta + u_{it} > 0)$$

This implies that the probability of observing the sequence $y_{i1}$ …….$y_{iT}$ for a particular individual is,

$$(16) \qquad Prob(y_{i1},\ldots,y_{iT}) = \int_{ai1}^{bi1} \cdots \int_{aiT}^{biT} f(u_{i1},\ldots,u_{iT})du_{iT},\ldots,du_{i1}$$

with $a_{it} = -X'_{it}\beta$, $b_{it}=\infty$ if $y_{it}=1$ and $a_{it}=-\infty$, $b_{it} = -X'_{it}\beta$ if $y_{it}=0$. The sample likelihood $L$ is the product of these integrals, $L_i$, over all n individuals. In certain cases, such as the random effects probit model, $L_i$ can be evaluated by quadrature. In general, the T-dimensional integral $L_i$ cannot be written in terms of univariate integrals that are easy to evaluate. Gaussian quadrature works well with low dimensions but computational problems arise with higher dimensions. Multivariate quadrature uses the Cartesian product of univariate evaluation points and the number of evaluation points increases exponentially (see Judd[12, p.275] for a method of evaluation that does not use the Cartesian product). Instead we can use Monte Carlo (MC) simulation to approximate integrals that are numerically intractable. This includes numerous models derived from the multivariate normal distribution (the panel probit, multinomial and multivariate probit, panel ordered probit and interval regression, panel Tobit, etc.). MC approaches use pseudo-random selection of evaluation points and computational cost rises less rapidly than with quadrature.

The principle behind simulation-based estimation is to replace a population value by a sample analogue. This means that we can use laws of large numbers (LLNs) and central limit theorems (CLTs) to derive the statistical properties of the estimators. The basic problem is to evaluate an integral of the form,

$$(17) \qquad \int_a^b \{h(u)\}f(u)\,du = E_u[\,h(u)]$$

This can be approximated using draws from f$(u)$, $u_r$, $r=1,\ldots,R$,

*(18)* $$(1/R)\sum_{r=1}^{R} h(u_r)$$

This is the direct MC estimate of $E_u[\ h(u)]$. Direct MC estimators are usually unbiased and consistent in R (due to the LLN) and asymptotically normal (due to the CLT).

*The Crude Frequency Simulator*

Lerman and Manski [13] proposed a MC algorithm for the evaluation of multivariate normal (MVN) probabilities such as those in the panel probit model. This is rarely used in practical applications but it illustrates a simple way of simulating MVN probabilities directly. The CFS for the probability, $P_i$, of a sequence of binary outcomes in the panel probit model works as follows:

## The CFS algorithm

1. Generate a T vector of pseudo-random independent standard normal variates.
2. Convert this into a $N(0,\Sigma)$ vector, where $\Sigma$ is the covariance matrix of $f(u_{i1},\ldots, u_{iT})$.
3. Determine whether this vector matches the conditions for the observed sequence of outcomes $y_{i1}, \ldots, y_{iT}$.
4. Repeat these steps a large number, R, times.
5. Evaluate the relative frequency of draws that are consistent with the observed outcomes. This gives an approximate value for $P_i$.

The CFS is computationally simple and cheap. But it has problems. It can easily return zero for $P_i$. This leads to computational problems when taking logs or ratios. It is discontinuous in the parameters creating a problem for derivative-based optimisation routines. The CFS has higher variance than other unbiased and consistent simulators for MVN probabilities.

An alternative is the GHK (Geweke-Hajivassiliou-Keane) simulator, which simulates probabilities under the multivariate normal distribution (see Part II of the structured bibliography in the Appendix). The GHK is a smooth recursive conditioning simulator (SRC). The GHK algorithm draws recursively from truncated univariate normals. This relies on the decomposition,

*(19)* $$f(u_1,\ldots,u_T) = f(u_1)f(u_2\,|\,u_1)\ldots\ldots f(u_{T-1}\,|\,u_{T-2},\ldots u_1)f(u_T\,|\,u_{T-1},\ldots u_1)$$

along with the fact that the conditional (in our case truncated) normal density can be written as a univariate normal. The GHK simulator produces probability estimates that are bounded between 0 and 1. The estimates are continuous and differentiable with respect to $(\beta,\Sigma)$, because each contribution is continuous and differentiable. It has a smaller variance than the CFS, because each element is bounded between 0 and 1. The GHK appears to be the most accurate simulator available for a given computation time. Box 2 provides Hajivassiliou's [14] GAUSS code for the GHK algorithm, a full description of this program can be found in the reference.

**BOX 2: GAUSS code for obtaining probability estimates using the GHK algorithm (source: Hajivassiliou [14])**

```
proc ghk(m,mu,w,wi,c,a,b,r,u);
local j,ii,ta,tb,tt,wgt,v,p;
j=1;
ii=1;
ta=cdfn((a[1,1]-mu[1,1])/(c[1,1]+1.e-100))*ones(1,r);
tb=cdfn((b[1,1]-mu[1,1])/(c[1,1]+1.e-100))*ones(1,r);
tt=cdfni(u[1,.].*ta+(1-u[1,.]).*tb);
wgt=tb-ta;
do while j<m;
j=j+1;
ta=cdfn(((a[j,1]-mu[j,1])*ones(1,r)-c[j,ii]*tt)/(c[j,j]+1.e-100));
tb=cdfn(((b[j,1]-mu[j,1])*ones(1,r)-c[j,ii]*tt)/(c[j,j]+1.e-100));
tt=tt| cdfni(u[j,.].*ta+(1-u[j,.]).*tb);
ii=ii|j;
wgt=wgt.*(tb-ta);
endo;
p=sumc(wgt')/r;
retp(p);
endp;
```

2.2 Maximum Simulated Likelihood (MSL)

This is a simple extension of classical maximum likelihood estimation (MLE) and is useful in many cases when the log-likelihood function involves high dimensional integrals. This includes the panel probit with RE+AR(1). The idea is to replace the likelihood function $L_i$ with a sample average over random draws,

$$l_i = (1/R) \sum_{r=1}^{R} l(u_{ir}) \tag{20}$$

where $l(u_{ir})$ is an unbiased simulator of $L_i$. The MSL estimates are the parameter values that maximize,

$$Lnl = \sum_{i=1}^{n} Lnl_i \tag{21}$$

In practice, antithetics can be used to reduce the variance of the simulator. These are based on,

$$l_i = (1/2R) \sum_{r=1}^{2R} l(u_{ir}) \tag{22}$$

where $u_{ir} = u_{i1}$ ,….., $u_{i2R}$ and $u_j = -u_{j-R}$ for j=R+1,….2R. Antithetics reduce the variance by using symmetric draws. If the probability simulator is linear in the draws, this approach reduces the variance to zero.

Having an unbiased simulator $l_i$ of $L_i$ (from CFS or GHK) does not imply an unbiased simulator of $lnL_i$ or the overall sample log-likelihood function (as $E[lnl_i] \neq ln(E[l_i])$). Of course MLE is, in general, biased due to nonlinearity. But, unlike MLE, the MSL estimator is not consistent solely in n. This is because the simulator is biased downwards for all individuals and the bias depends on $\beta$. Consistency and asymptotic unbiasedness can be obtained by reducing the error in the simulated sample log-likelihood to zero as $R \to \infty$ at a sufficient rate with n. Hajivassiliou and Ruud [15] show that a sufficient rate is $R/\sqrt{n} \to \infty$ as $n \to \infty$. Hajivassiliou and Ruud also show that this is sufficient for the usual MLE estimate of the covariance matrix to be used without any correction.

## 2.3 Application of MSL

Table 2 compares the estimates of the income effects and the variance of the individual effect for the random effects probit model computed using both quadrature (with 24 points) and MSL (with 150 replications) with the actual data. Simulation is not required for this model but it provides a useful test of the reliability of the simulation approach. It is clear from the table that the estimates are very similar.

Table 2: Comparison of quadrature and MSL for the random effects probit:actual data.

| | MLE (24 point) | MSL (R=150) |
|---|---|---|
| Ln('permanent income') | -0.526 | -0.510 |
| | (0.048) | (0.049) |
| Ln('transitory income') | -0.053 | -0.052 |
| | (0.023) | (0.023) |
| $\sigma_\alpha^2$ | 0.784 | 0.788 |
| | (0.012) | (0.012) |
| LnL | -4291.2 | -4290.5 |

Simulation becomes necessary to move beyond the simple random effects (RE) specification. Table 3(a) shows how the MSL estimates of models that allow for an RE+AR(1) error structure and for an unrestricted covariance matrix perform with the simulated data. The estimates are based on 150 replications. As with the random effects probit estimator, the MSL estimators do a good job of reproducing the the true value of $\beta$ for models 1 and 3 and they also perform poorly when the individual effect is correlated with the regressor in model 2. For the model 1 data, the RE+AR(1) estimator over-estimates the size of $\sigma_\alpha^2$ and under-estimates the size of $\rho$, suggesting that the are problems in separately identifying the variance components with these data. The negative value of $\rho$ for the model 3 data seems to be capturing the fact that, due to the ARMA error term, the lag 2 correlation is larger than the lag 1 correlation. Table 3(b) presents estimates of the income effects for the actual data with the models ranging from the RE probit, through the RE+AR(1), to an unrestricted covariance matrix. The results show that the income effects are largely unaffected by moving to more flexible specifications of the covariance matrix..

However these more flexible specifications do provide information about the variance components.

Table 3(a): MSL estimates (R=150): simulated data.

| | Model 1: RE+AR(1) | | Model 2: Correlated effects | | Model 3: RE+ARMA(2,1) | |
|---|---|---|---|---|---|---|
| | RE+AR(1) | Unres | RE+AR(1) | Unres | RE+AR(1) | Unres* |
| $\beta$ (=1) | 0.993 | 0.999 | 0.206 | 0.200 | 1.005 | 1.006 |
| | (0.030) | (0.031) | (0.013) | (0.013) | (0.031) | (0.022) |
| $\sigma_\alpha^2$ (= 0.8) | 0.826 | | 0.991 | | 0.937 | |
| | (0.019) | | (0.002) | | (0.005) | |
| $\rho$ (=0.7) | 0.668 | | 0.645 | | -0.046 | |
| | (0.041) | | (0.082) | | (0.051) | |

* Due to problems with convergence with R=150 these results are presented for R=40.

Table 3(b): Estimated income effects under alternative covariance structures: actual data

| | RE (MSL) | RE+AR(1) | Unrestricted |
|---|---|---|---|
| Ln('permanent income') | -0.510 | -0.509 | -0.511 |
| | (0.049) | (0.049) | (0.049) |
| Ln('transitory income') | -0.052 | -0.057 | -0.055 |
| | (0.023) | (0.025) | (0.024) |

To assess the overall statistical performance of the models on the actual data Table 4 shows various model selection criteria. The most general model, with an unrestricted covariance matrix, has the largest log-likelihood function. But the model has many more parameters than the RE+AR(1) and RE specifications. Information criteria can be used to penalise the measure of goodness of fit for the loss of degrees of freedom. The unrestricted model is still preferred according to the Akaike information criterion (AIC) but, when the number of parameters is penalised more heavily with the Bayesian information criterion (BIC) and Consistent AIC (CAIC), the RE+AR(1) specification is favoured.

Table 4: Model selection criteria using MSL: actual data.

| | LnL | AIC | BIC | CAIC |
|---|---|---|---|---|
| Unrestricted | -4188.07 | 8421.12 | 8731.92 | 8820.00 |
| RE+AR(1) | -4214.83 | 8455.66 | 8635.23 | 8626.82 |
| RE | -4290.55 | 8606.10 | 8778.76 | 8851.31 |

Table 5 assesses the impact of simulation bias on the results by comparing estimates based on different numbers of replications. Table 5(a) shows that only 10 replications are required to get estimates of $\beta$ close to the true value, using simulated data from model 1. However many more replications are required to get the estimates of the variance components, in particular the value of $\rho$, to converge towards their true values. This reflects the problem of separately identifying the variance components. Table 5(b) shows that the estimates from the actual data are robust for values of R in the range 75-150. Simulation bias can be assessed more formally by using the test statistic proposed by Hajivassiliou [14] (see Contoyannis, Jones and Rice [1] for an application of this test).

Table 5(a): Estimates for different values of R - MSL RE+AR(1) estimator applied to model 1: simulated data.

| | $\beta$ (=1.0) | $\sigma_\alpha^2$ *(=0.8)* | $\rho$ (=0.7) |
|---|---|---|---|
| R= 10 | 1.019 | 0.822 | 0.596 |
| | (0.029) | (0.014) | (0.036) |
| R= 20 | 0.982 | 0.846 | 0.605 |
| | (0.029) | (0.014) | (0.039) |
| R= 40 | 0.999 | 0.830 | 0.636 |
| | (0.030) | (0.017) | (0.040) |
| R = 60 | 0.996 | 0.830 | 0.645 |
| | (0.030) | (0.017) | (0.041) |
| R = 100 | 0.989 | 0.832 | 0.658 |
| | (0.030) | (0.018) | (0.041) |
| R=150 | 0.993 | 0.826 | 0.668 |
| | (0.030) | (0.019) | (0.041) |

Table 5(b): Estimates for different values of R: actual data.

| | 'permanent' | 'transitory' | $\sigma_\alpha^2$ | $\rho$ |
|---|---|---|---|---|
| R=75 | -0.515 | -0.057 | 0.698 | 0.521 |
| R=100 | -0.515 | -0.059 | 0.695 | 0.528 |
| R=150 | -0.510 | -0.057 | 0.691 | 0.541 |

*Some guidelines*

Theoretical considerations and experience of applying these methods suggests that the following guidelines should be taken into account when putting MSL into practice:-

1. Use fixed draws: The random draws should only be drawn once and not varied as the optimisation algorithm searches. If the random variates do change, the values of $(\beta,\Sigma)$ which maximize the simulated log-likelihood will change and the optimisation routine may never converge. This is a general requirement for any simulation-based estimation that uses an iterative optimisation routine. Furthermore, the asymptotic theory for these estimators is based on a given set of draws.

2. Use a smooth and bounded simulator: It is important to use a simulator which is smooth in $\beta$ (such as the GHK algorithm) so that derivative based optimisation routines may be used. It is also important to use simulators which are bounded by 0 and 1 (e.g. GHK), so the simulated sample log-likelihood can always be evaluated. When using numerical derivatives it is advisable to use alternative step sizes until the estimates of the derivatives are stable, starting with a relatively large step size and working down

3. Scale the data and use good initial values: The tails of the multivariate normal density die out very rapidly. This causes potential underflows during computation (the values are smaller than can be expressed and manipulated during computation). This problem is severe when taking logs. Hajivassiliou [14] suggests standardising regressors (to zero mean and unit variance) before estimation. More generally, reasonable starting values will help to prevent this problem. These can be obtained from simpler specifications that do not require simulation, such as the pooled or random effects probit estimators.

4. Use antithetics: Use of antithetic variates can reduce simulation variance and bias substantially.

5. Validate the algorithm: Use a simple model to validate a new program by comparing MSL and MLE.

6. Check for bias: Bias is model specific and, while there is guidance in the literature – for example many studies suggest that values of R less than 50 are sufficient - sensitivity analysis is important. As noted above, a test for asymptotic bias is available (Hajivassiliou, [14]). Bias corrections are available but are computationally difficult and may not perform well (e.g., Lee, [16]).

## 2.4 Other Classical estimators

MSL is not the only classical simulation estimator available. The Method of Simulated Scores (MSS) is a general approach based on the simulated score function. The estimator is implicitly defined as the value of $\beta$ which satisfies,

(23) $$1/n \sum_{i=1}^{n} l_{i\beta} / l_i = 0$$

where $l_{i\beta} = [1/R_1 \sum_{r=1}^{R1} l_{i\beta r}]$ is an unbiased simulator for the vector of derivatives $L_{i\beta}$ and $l_i$ is an unbiased simulator for $L_i$ based on $R_2$ simulations. In general, the MSS estimator is consistent and asymptotically normal as $n \to \infty$ and $R_2/\sqrt{n} \to \infty$. The value of $R_1$ affects the efficiency of the estimator.

The method of moments (MoM) estimator for the probit model solves the orthogonality condition,

(24) $$1/n \sum_{i=1}^{n} Q_i [y_i - \Phi(X'_i\beta)] = 0$$

for a fixed and exogenous vector of instruments Q (e.g. X). MoM is equivalent to MLE with a suitable choice of Q. If the moments cannot be evaluated analytically (as in the panel probit model) then the residual can be simulated and, for efficiency, so can the optimal instruments. This gives the Maximum Simulated Moments (MSM) estimator (McFadden [17]). Further details of the links between MSL, MSM and MSS can be found in the references given in the Appendix.

Box 3 suggests some convenient sources of software for classical simulation-based estimation.

**BOX 3: Software for classical simulation methods**

http://econ.lse.ac.uk/~vassilis/pub/simulation

NMRSIM_G.LIB

Library of GAUSS routines for multivariate normal rectangle probabilities and their derivatives. Companion to Hajivassiliou, McFadden and Ruud [18]

SSMLMNP by A. Borsch-Supan & V. Hajivasilliou

Estimation code for simulated maximum likelihood of multinomial probit in Fortran 77

Limdep for Windows (http://www.limdep.com)

Simulator for multivariate normal CDF using GHK algorithm. By default, R=100 but can be changed. Up to m=20 variate integral

# 3. Bayesian MCMC methods

## 3.1 The Bayesian approach

In Bayesian analysis a prior density of the parameters of interest, $\pi(\theta)$, is updated with the information contained in the sample (for the RE+AR(1) panel probit model $\theta = \{\beta,\rho,\sigma^2_\alpha\}$). Given a specified sample likelihood, $\pi(y|\theta)$, the posterior density of $\theta$ is given by Bayes' theorem,

$$\pi(\theta|y) = \pi(\theta)\pi(y|\theta) / \pi(y) \tag{25}$$

where,

$$\pi(y) = \int \pi(\theta)\pi(y|\theta)\, d\theta \tag{26}$$

$\pi(y)$ is known as the predictive likelihood and it is used for model comparison. It determines the probability that the specified model is correct. The posterior density $\pi(\theta|y)$ reflects updated beliefs about the parameters. Given the posterior distribution, a 95% credible interval can be constructed that contains the true parameter with probability equal to 95%. Point estimates for the parameters can be computed using the posterior mean,

$$E(\theta|y) = \int \theta\pi(\theta|y)\, d\theta \tag{27}$$

## 3.2 Markov Chain Monte Carlo (MCMC) Methods

Bayesian estimates can be difficult to compute. For instance, the posterior mean (27) is an integral with dimension equal to the number of parameters in the model. In order to overcome the difficulties in obtaining the characteristics of the posterior density, Markov Chain Monte Carlo (MCMC) methods are used. The methods provide a sample from the posterior distribution. Posterior moments and credible intervals are obtained from this sample.

Random numbers can be easily generated for some well- known distributions. For example, Kinderman et al. [19] developed a simple method to obtain a sample of independent draws from a normal density. This or other methods are implemented in standard statistical software, such as GAUSS and STATA. However, posterior densities in econometric models have not been so extensively studied, and such simple methods do not exist.

MCMC algorithms yield a sample from the posterior density by constructing a Markov Chain which converges in distribution to the posterior density. In a Markov chain each value is drawn conditionally on the previous iteration. After discarding the initial iterations, the remaining values can be regarded as a sample from the posterior density. The algorithms explained below are different ways of appropriately constructing a Markov Chain.

*Gibbs Sampling*

To implement Gibbs sampling the vector of parameters $\theta$ is subdivided into *s* groups, $\theta = (\theta_1, ... \theta_s)$. For example, with two groups, let $\theta = (\theta_1, \theta_2)$. A draw from a distribution $\pi(\theta_1, \theta_2)$ can be obtained in two steps. First, draw $\theta_1$ from its marginal distribution $\pi(\theta_1)$. Second, draw $\theta_2$ from its conditional distribution given $\theta_1$, $\pi(\theta_2 | \theta_1)$. In many situations it is possible to sample from the conditional distribution $\pi(\theta_2 | \theta_1)$ but it is not obvious how to sample from the marginal $\pi(\theta_1)$. The Gibbs sampling algorithm solves this problem by sampling iteratively from the full conditional distributions. Even though the Gibbs sampling algorithm never draws from the marginal, after a sufficiently large number of iterations, the draws can be regarded as a sample from the joint distribution.

To implement Gibbs sampling initial values are fixed arbitrarily. Then, at each iteration in the chain, $\theta^k = (\theta_1^k, \theta_2^k)$ is drawn as follows:

*1.* $\theta_1^k$ is drawn from the distribution $\pi(\theta_1 | y, \theta_2^{k-1})$

*2.* $\theta_2^k$ is drawn from the distribution $\pi(\theta_2 | y, \theta_1^{k-1})$

The process is repeated until a sufficiently large sample is obtained.

In practice the conditional distributions must be easy to sample from. Also, the number of sub-groups in which θ is subdivided should be kept as small as possible to speed up convergence. Good starting values can substantially save in computing time in complex models.

*Data Augmentation*

Following Tanner and Wong [20], latent or missing data can be regarded as parameters belonging to θ. Although this introduces many more parameters into the model, the conditional densities belong to well-known families and there are simple methods to sample from them. This makes the use of the Gibbs sampling possible. For example, in a simple probit model,

*(28)* $$y^*_i = X'_i\beta + \varepsilon_i,\ \varepsilon \sim N(0,1),\ \theta=(\beta,y^*)$$

If the latent data is not estimated, the posterior density for beta is,

(29) $$\prod_{i=1}^{n} (\Phi[d_i(X_i'\beta)])f(\beta)/\pi(y) \qquad \text{with} \qquad \pi(y) = \int \prod_{i=1}^{n} (\Phi[d_i(X_i'\beta)])f(\beta)d\beta\ ,$$

$d_i = 2y_i - 1$, and where $f(\beta)$ is the prior density of beta. Unfortunately, there are no simple methods to generate a sample from this density. However, if the latent data is estimated the posterior density for latent data and parameters is:

(30) $$\prod_{i=1}^{n} \left\{1(y_i^* d_i > 0)\phi(y_i^* - X_i'\beta)\right\} f(\beta)/\pi(y) \text{ with}$$

$$\pi(y) = \int \prod_{i=1}^{n} \left\{1(y_i^* d_i > 0)\phi(y_i^* - X_i'\beta)dy_i^*\right\} f(\beta)d\beta$$

where $\phi$ is the density function of a standard normal, and $f(\beta)\prod_{i=1}^{n} \phi(y_i^* - X_i'\beta)$ is the prior density for latent data and β. A Gibbs sampling algorithm can be used to obtain values for β and $y^*_i$ from this density.

The conditional distribution of $y^*_i$, given β, is $N(X'_i\beta,1)$, truncated to positive values if y=1 and truncated to negative values if y=0. The conditional density of β given the latent data is that

which arises in a linear regression model. If the prior density for $\beta$ is normal, the conditional of $\beta$ given $y^*_i$ is a also a normal (Lee [21], page 167) and hence the Kinderman et al. [19] method can be used to sample from it.

*Metropolis-Hastings (M-H) Algorithms*

There are situations in which it does not seem possible to sample from a conditional density, and hence the Gibbs sampling cannot be applied directly. In these situations, Gibbs sampling can be combined with a so called Metropolis step. In the Metropolis step, values for the parameters are drawn from an arbitrary density, and accepted or rejected with some probability. Say that $\pi(\theta_1 \mid y, \theta_2)$ is not easy to sample from. Let $q(\theta_1 \mid \theta_1^{k-1}, \theta_2^{k-1})$ be a density for $\theta_1$, given $\theta_1^{k-1}$ and $\theta_2^{k-1}$, that can be sampled easily. A M-H algorithm generates $\theta^k = (\theta_1^k, \theta_2^k)$ as follows:

1. Draw a candidate value $\theta_1^*$ for $\theta_1^k$ from the distribution $q(\theta_1^* \mid \theta_1^{k-1}, \theta_2^{k-1})$.
2. Set $\theta_1^k = \theta_1^*$ with some probability $\gamma$ otherwise keep $\theta_1^k = \theta_1^{k-1}$.

The probability $\gamma$ depends on the values of the ratio $\pi(.)/q(.)$ evaluated at the new proposed value and at the previous value in the chain (see for instance Gamerman [22]). New candidates can be generated in a simple way from a normal distribution centred at the previous value in the chain and with arbitrary variance. However, if the dimension of $\theta_1$ is greater than 2, this may not work well. Alternatively, new candidates can be generated from a distribution that approximates the conditional density $\pi(\theta_1 \mid y, \theta_2^{k-1})$. Bad approximations may result in low probabilities of acceptance and slow convergence.

Once a value for $\theta_1$ has been generated, a value for $\theta_2$ will be drawn from the conditional density $\pi(\theta_2 \mid y, \theta_1^k)$. If it was not obvious how to sample from this density, a Metropolis step could also be used to generate $\theta_2$. Data augmentation can also be combined with the Metropolis algorithm. The panel probit model provides an example of how Gibbs sampling, a Metropolis step and data augmentation can be combined.

## 3.3 MCMC estimation of the panel probit

To see how MCMC works in practice consider the panel probit with autocorrelated error terms,

$$(31) \qquad y_{it} = 1(y^*_{it} > 0) = 1(X'_{it}\beta + \alpha_i + \varepsilon_{it} > 0), \quad \varepsilon_{it} = \rho\varepsilon_{it-1} + \eta_{it}, \quad \text{var}(\alpha_i) = \sigma^2_\alpha$$

The priors selected for $\beta$, $\sigma^2_\alpha$ and $\rho$ are normal, inverted gamma and uniform (-1,1) distributions respectively. The parameters and latent data are divided into 5 groups: $\theta = (y^*_{it}, \beta, \alpha_i, \sigma^2_\alpha, \rho)$. Note that data augmentation is used, since latent data and individual effects are treated as parameters to estimate. We use a Gibbs sampling algorithm, and hence parameters in each group are drawn from their conditional density given the parameters in the rest of groups. Since it is not obvious how to sample from the conditional density of $\rho$, a Metropolis step is used to generate it. The following six steps describe the algorithm:

1. Fix the initial values.
2. Generate $y^*_{it}$ from a truncated normal distribution with mean $X'_{it}\beta + \alpha_i$ and variance $\text{var}(\varepsilon_{it})$, according to the value of $y_{it}$.
3. Generate $\beta$ from a normal distribution. The mean and variance of this distribution are the Bayesian point estimates in a model in which $y^*_{it}$ and $\alpha_i$ are observed and $(\sigma^2_\alpha, \rho)$ are known.
4. Generate $\alpha_i$ from a normal distribution. The mean and variance are the point estimates in a model in which $y^*_{it}$ is observed and $(\beta, \sigma^2_\alpha, \rho)$ are known.
5. Generate $\sigma^2_\alpha$ from an inverted gamma distribution. The parameters of this inverted gamma are the same as in a model in which $y^*_{it}$ and $\alpha_i$ are observed and $(\beta, \rho)$ are known.
6. $\rho$ can be generated with a M-H step. New candidates can be generated from a normal distribution.

Box 4 sketches an outline of how this algorithm could be implemented in GAUSS. Each of the procedures in the code implements one of the steps in the algorithm. The entire code can be found in the web page address referred to in Box 5 below.

**BOX 4: GAUSS code for MCMC estimation of panel probit model**

```
iter=1;
Beta[1,.]=zeros(1,K);            /*Step 1*/
Rho[1]=0;
Var_ind[1]=1;
Do while iter<=NIT;
Latent=augment(Beta[iter,.], Effects, Var_ind[iter], Rho[iter]);          /*Step 2*/
Beta[iter+1,.]=sampleB(Latent, Effects, Var_ind[iter], Rho[iter]);        /*Step 3*/
Effects=sampleI(Latent, Beta[iter+1,.], Var_ind[iter], Rho[iter]);        /*Step 4*/
Var_ind=sampleV(Latent, Effects, Beta[iter+1,.], Rho[iter]);              /*Step 5*/
Rho[iter+1]=sampleR(Latent, Effects, Beta[iter+1], Var_ind[iter+1],
Rho[iter]);       /*Step 6 */
iter=iter+1;
endo;
```

3.4 Output and convergence analysis

Sample means, standard deviations and correlations estimated from the MCMC draws are all strongly consistent estimates of the corresponding characteristics of the posterior distribution. A sample for any transformation of the parameters $g(\theta)$ can be obtained by equivalently transforming the sampled values for $\theta$. This allows the calculation of credible intervals and posterior means for $g(\theta)$. For example, in the panel probit model, $g(\theta)$ could represent the marginal effect of income on the probability of being ill. Since values in the chain are not independent, sample standard deviations are usually biased. Geweke [23] proposes an alternative method that is implemented in standard packages.

Tables 6(a) and 6(b) show that the MCMC algorithm produces very similar point estimates and standard errors to those estimated by MSL, for the full range of specifications. Figure 1 shows the results of the MCMC estimation for the autocorrelation parameter $\rho$ and the variance of the individual effect $\sigma^2_\alpha$ in the RE+AR(1) model. The left-hand panels show the output of the Markov chain for successive iterations. The right-hand panels compare kernel density estimates of the prior and posterior densities for the two parameters.

Table 6(a): Comparison of MSL(R=150) and MCMC: model 1 simulated data.

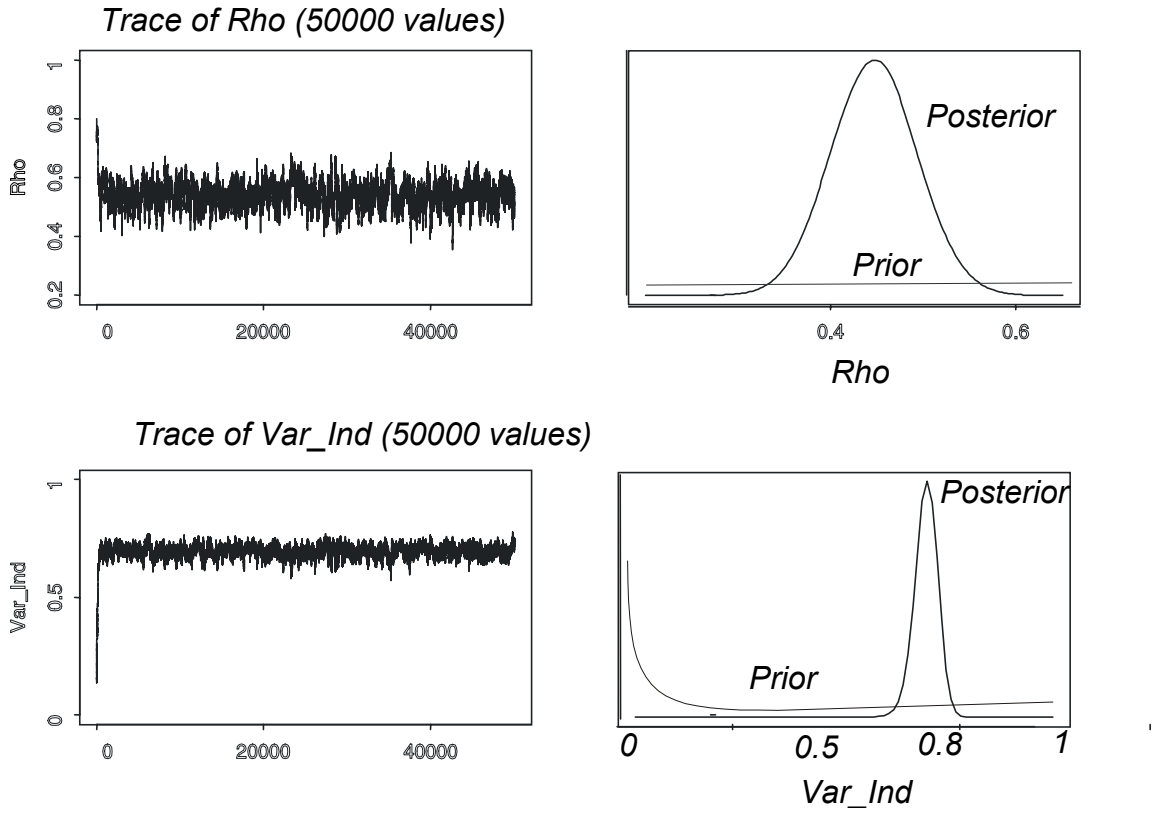
| | Model 1: RE+AR(1) | |
|---|---|---|
| | MSL | MCMC |
| $\beta$ (=1) | 0.993 | 0.992 |
| | (0.030) | (0.031) |
| $\sigma_\alpha^2$ (= 0.8) | 0.826 | 0.825 |
| | (0.019) | (0.020) |
| $\rho$ (=0.7) | 0.668 | 0.667 |
| | (0.041) | (0.041) |

Table 6(b): Comparison of MSL and MCMC estimates for the panel probit models: actual data

| | RE + AR(1) | | Un- restricted | |
|---|---|---|---|---|

| | MSL | MCMC | MSL | MCMC |
|---|---|---|---|---|
| Ln('permanent income') | -0.510 | -0.504 | -0.511 | -0.481 |
| | (0.049) | (0.049) | (0.049) | (0.051) |
| Ln('transitory income') | -0.057 | -0.057 | -0.054 | -0.051 |
| | (0.025) | (0.024) | (0.024) | (0.022) |
| $\sigma_\alpha^2$ | 0.691 | 0.696 | | |
| | (0.025) | (0.025) | | |
| $\rho$ | 0.541 | 0.541 | | |
| | (0.040) | (0.042) | | |

Figure 1: MCMC parameter estimates and prior and posterior densities: actual data.



Independently of the initial values, after a sufficient number of iterations, the algorithm should converge to the posterior distribution. An informal way of checking for convergence is to subdivide the chain in several parts and compare the average and standard deviations for each part. Alternatively, averages and standard deviations of chains started at different initial values may be compared. High serial correlation in the chain values indicates that a longer chain will be necessary to obtain precise estimates. Cowles and Carlin [24] review some formal tests of convergence. The freely available software CODA includes the following tests: the Geweke test for convergence; the Heidelberger and Welch test for convergence which also tests whether the posterior mean has been estimated with a given degree of accuracy; and the Raftery and Lewis test gives the length of the chain necessary to calculate a credible interval with a specified degree of accuracy.

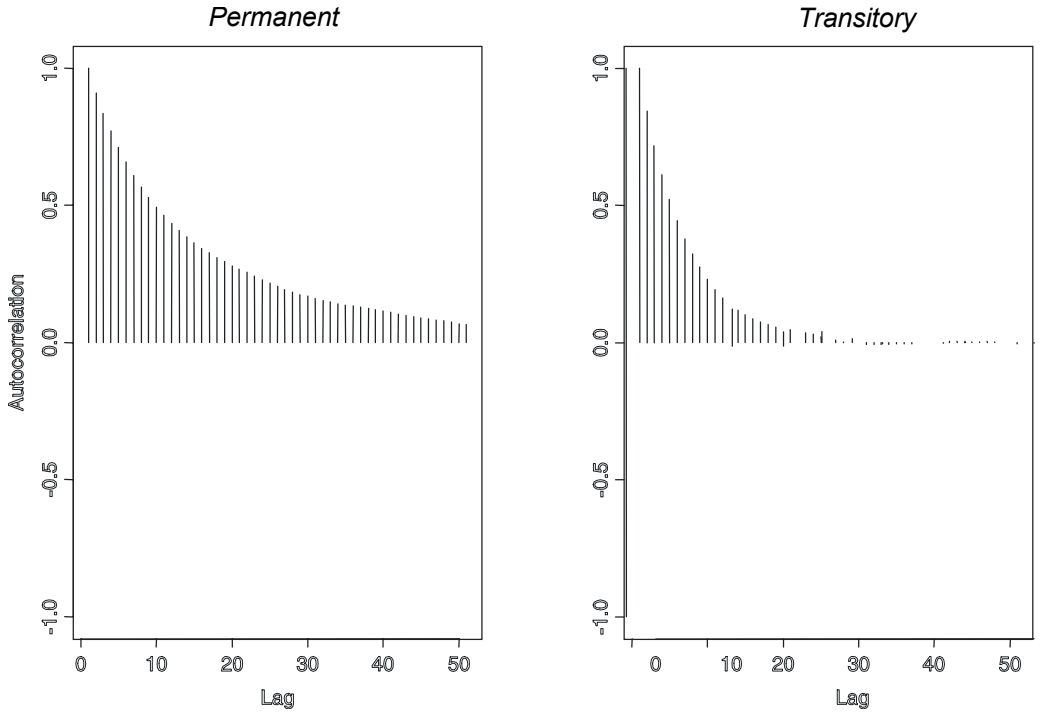
High serial correlation in the chain indicates that a longer chain will be necessary to obtain precise estimates. When there are large autocorrelations, the chain recovers the parameter space slowly, and hence more iterations are necessary to obtain a representative sample from the posterior density. If the serial correlations are too large, it might be infeasible to run the algorithm for long enough, and an alternative algorithm should be designed. Liu [25] describes some strategies that can be followed in order to construct a faster algorithm.

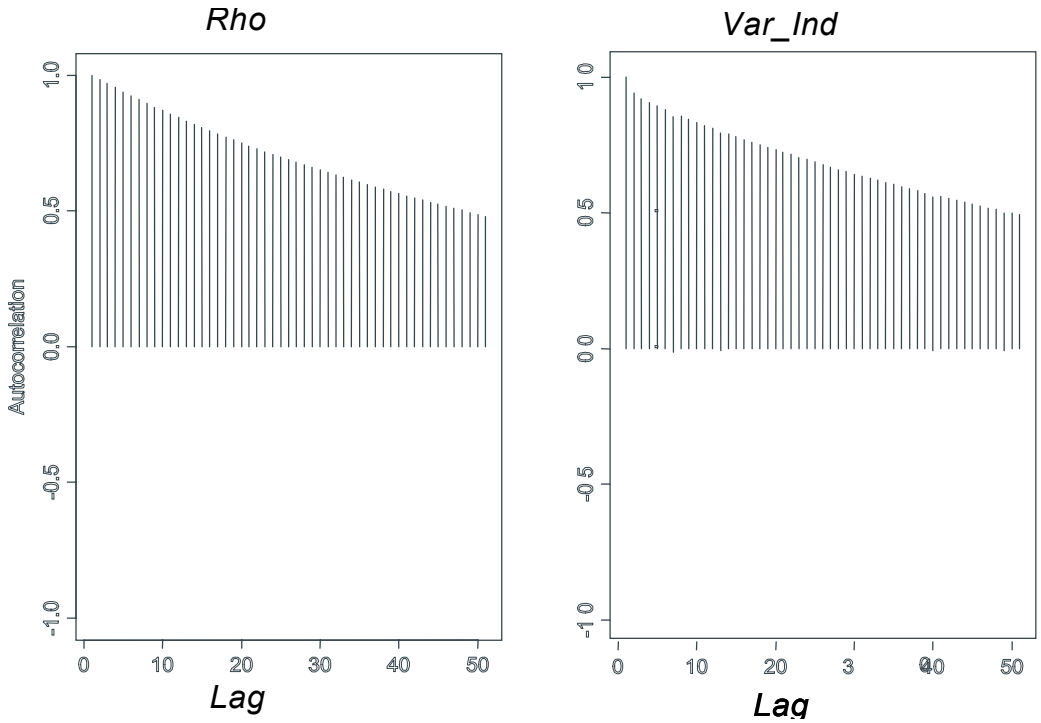
Results of applying the CODA software to our MCMC algorithm for the RE+AR(1) model are shown in Table 7. These show that a longer chain is required to get convergence in the estimates for $\sigma_\alpha^2$ and $\rho$ than for the income effects. This finding is reflected in the plots of the autocorrelation functions, shown in Figure 2, these 'die-out' for the two income parameters but persist for $\sigma_\alpha^2$ and $\rho$.

Table 7: Results of convergence analysis: actual data.

| | Geweke | Convergence (H-W) | Posterior mean (H-W) | Length of chain required |
|---|---|---|---|---|
| Ln('permanent income') | 0.108 | passed | passed | 11,000 |
| Ln('transitory income') | 0.255 | passed | passed | 8,000 |
| $\sigma_\alpha^2$ | -0.782 | passed | passed | 35,000 |
| $\rho$ | 0.420 | passed | passed | 21,000 |

Figure 2: Autocorrelations for key parameters: actual data.




3.5 Testing and model selection

Bayesian inference offers a unified approach to testing, which can be used to discriminate between any type of models. This contrasts with likelihood ratio and Wald tests, which are often not appropriate to discriminate between non-nested models or to test parameter values on the boundary of the parameter space. Testing in the Bayesian approach consists of comparing the probabilities of different hypotheses. The probability of each model is determined by the predictive likelihood, $\pi(y)$, which is the normalising constant in the denominator of Bayes' theorem,

$$\pi(y \mid M_i) = \int \pi(\theta)\pi(y \mid \theta, M_i)\, d\theta \tag{32}$$

Notice that, in comparison to (26), the predictive likelihood is now explicitly conditional on the model specification $M_i$. Given $m$ possible models $\{M_i\}$, and prior probabilities for each model, $\pi(M_i)$, the posterior probability for model $M_i$ *is,*

$$\pi(M_i \mid y) = \pi(M_i)\pi(y \mid M_i) \,/\, \sum_{j=1}^{m} \pi(M_j)\pi(y \mid M_j) \tag{33}$$

Although the posterior probability depends on the number of models $m$, which is determined a priori, the ratio of the probabilities of two different models does not depend on $m$. In the case of equal prior probabilities for each model this ratio is known as the Bayes factor (B),

$$B_{i,j} = \pi(y \mid M_i)/\pi(y \mid M_j) \tag{34}$$

If a model is to be selected, it should be the model with the largest value for the predictive likelihood.

Chib and Jeliazkov [26] present a method to calculate the predictive likelihood, $\pi(y)$. From Bayes' theorem,

$$\pi(\theta \mid y) = \pi(\theta)\pi(y \mid \theta) \,/\, \pi(y) \tag{35}$$

It follows that any particular value, $\theta^*$, of the parameters satisfies the identity,

(36) $$ln\pi(y) = ln\pi(\theta^*) + ln\pi(y|\theta^*) - ln\pi(\theta^*|y)$$

Chib and Jeliazkov [26] propose a method to estimate the posterior ordinate $ln\pi(\theta^*|y)$. The method requires running the algorithm for additional iterations. In order to assess the accuracy of the calculation, they also provide the standard deviation of the estimated value for $ln\pi(\theta^*|y)$. If evaluation of the likelihood, $\pi(\theta^*|y)$, involves multiple integrals it can be computed using the methods described in Section 2.

To illustrate the use of Bayes factors for our models let:
$M_1$: Random effects and independent time variant errors
$M_2$: Random effects and AR(1) errors
$M_3$: Unrestricted variance-covariance matrix
This gives values of the Bayes factors $Pr(M_1)/Pr(M_2)$ and $Pr(M_3)/Pr(M_2)$ that are very close to zero. These results support the RE+AR(1) specification against the simpler RE model and against the more general unrestricted model. This parallels the finding based on the BIC and CAIC criteria in the classical analysis.

*Testing hypotheses about* $\theta$

When the hypothesis of interest is of the type $\theta_1=k$ , it is possible to use Verdinelli and Wasserman's [27] method. Unlike the Chib and Jeliazkov method, in many situations this method does not require any additional computations. Their procedure gets more complicated in terms of computing time when the normalising constant of $\pi(\theta_1|y,\theta_2)$ is not known, or when $\theta_1$ and $\theta_2$ are not independent a priori.

*Model Averaging.*

The posterior probabilities for each model lead to a procedure to deal with uncertainty about the appropriate model to use. The posterior density for $\theta$ takes into account the different possible specifications,

(37) $$\pi(\theta|y) = \sum_{j=1}^{m} \pi(\theta|y,M_j)\pi(M_j|y)$$

The posterior mean for θ is a weighted average of the posterior means in each model, the weighted by the posterior probability of each model.

*(38)* $$E(\theta \mid y) = \sum_{j=1}^{m} E(\theta \mid y, M_j)\pi(M_j \mid y)$$

Box 5 lists sources of software for doing Bayesian analysis using MCMC. It also provides a reference for the GAUSS code used to estimate the panel probit models.

**BOX 5: Software for Bayesian MCMC analysis**

BACC freely available at www.econ.umn.edu/~bacc
Includes, among other models, univariate latent models with flexible distributions for the error term (mixture of normals and student-t distributions). Allows the user to implement more complex procedures building on simpler models. This requires knowledge of C language.

BUGS and CODA freely available at www.mrc-bsu.cam.ac.uk/bugs/
BUGS allows users to easily specify their own MCMC sampling algorithm
CODA provides checks of convergence for the MCMC output

MLWIN is commercial software which includes Bayesian estimation of multi-level models

The GAUSS code in Box 3 can be downloaded from http://www.york.ac.uk/res/herc/yshe

## Overview

This review illustrates the scope for using simulation methods to allow for flexible specifications of heterogeneity in nonlinear models for panel data. It uses binary choice models to show what can be done with conventional methods and how the range of models can be expanded by using classical and Bayesian simulation methods. Practical applications of the methods are illustrated using data on self-reported health from the British Household Panel Survey (the BHPS). Our aim is to provide a brief introduction to simulation methods and to show their relevance for applied analysis in health economics. To provide some guidance for readers who would like to pursue the topic in more detail the Appendix provides a structured bibliography that includes key references to methods and some examples of applications in health economics.

## REFERENCES


1. Contoyannis P, Jones AM, Rice N. Simulation-based inference in dynamic panel probit models: an application to health. http://www.york.ac.uk/res/herc/yshe [29 May 2002].

2. Heckman J. Statistical models for discrete panel data. In *Structural Analysis of discrete data with econometric applications*, Manski CF, McFadden D (eds). MIT Press: Cambridge Mass., 1981; 114-178.

3. Greene W. Fixed and random effects in nonlinear models. http://www.stern.nyu.edu/~wgreene [19 August 2002]

4. Butler J, Moffitt R. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* 1982; **50**:761-764.

5. Pudney S. *Modelling individual choice: The econometrics of corners, kinks and holes.* Appendix 3-Computation of probability integrals. Blackwell: Oxford, 1989.

6. Deb P. A discrete random effects probit model with application to the demand for preventive care. *Health Economics* 2001; **10**: 371-383.

7. Heckman JJ, Singer B. A method of minimizing the distributional impact in econometric models for duration data. *Econometrica 1984;* **52**: 271-230.

8. Chamberlain G. Analysis of covariance with qualitative data. *Review of Economic Studies 1980;* **47**: 225-238.

9. Chamberlain G. Panel data. In *Handbook of Econometrics Volume 1,* Griliches Z, Intrilligator M. (eds). Elsevier: Amsterdam, 1984; 1247-1318.

10. Wooldridge JM. The initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. Mimeo, Michigan State University, 2000.

11. Robinson P. On the asymptotic properties of estimators of models containing limited dependent variables. *Econometrica* 1982; **50**: 27-41.

12. Judd KL. *Numerical Methods in Economics.* The MIT Press: Cambridge MA, 1998.

13. Lerman S, Manski C. On the use of simulated frequencies to approximate choice probabilities. In *Structural Analysis of discrete data with econometric applications*, Manski CF, McFadden D. (eds). MIT Press: Cambridge Mass., 1981; 305-319.

14. Hajivassiliou V. Some practical issues in maximum simulated likelihood. In *Simulation-Based inference in econometrics: Methods and Applications.* Mariano R, Schuermann T, Weeks MJ. (eds). Cambridge University Press: Cambridge, 2000; 71-99.

15. Hajivassiliou V, Ruud P. Classical Estimation methods using simulation. In *Handbook of Econometrics Volume 4*. Engle R, McFadden D. (eds). Elsevier: Amsterdam, 1994; 2383-2441.

16. Lee L-F. Simulated maximum likelihood estimation of dynamic discrete choice statistical models. *Journal of Econometrics* 1997; **82**:1-35.

17. McFadden D. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 1989; **57**:995-1026.

18. Hajivassiliou V, McFadden D, Ruud P. Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics 1996,* **72**:85-134.

19. Kinderman AJ, Ramage JG. Computer Generation of Normal random Numbers. *Journal of the American Statistical Association* 1976; **71**: 893-896.

20. Tanner M, Wong W. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 1987; **82**: 528-550.

21. Lee PM. *Bayesian statistics: an introduction*. Arnold: London, 1997.

22. Gamerman D. *Markov Chain Monte Carlo*.Chapman & Hall: London, 1997.

23. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*. Bernardo JM, Berger J, Dawid AP, Smith AFM. (eds). Oxford University Press: Oxford, 1992; 169-193.

24. Cowles M, Carlin B. Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association 1996*; **91**:883-904.

25. Liu, J. S. *Monte Carlo Strategies in Scientific Computing* Springer-Verlag, New York, 2001.

26. Chib S, Jeliazkov I. *Marginal Likelihood From the Metropolis-Hastings Output. Journal of the American Statistical Association 2001*; **96**:270-281.

27. Verdinelli I, Wasserman L. Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association* 1995; **90**: 614-618.

## APPENDIX: STRUCTURED BIBLIOGRAPHY

### PART I: Methods for nonlinear panel data models


Arellano M, Honoré B. Panel data models: some recent developments. In *Handbook of Econometrics Volume 5*, Heckman JJ, Leamer E. (eds). Elsevier: Amsterdam, 2001; 3229-3296.

Baltagi B. *Econometric Analysis of Panel Data.* John Wiley: Chichester, 2001.

Butler J, Moffitt R. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* 1982; **50**:761-764.

Chamberlain G. Panel Data. In *Handbook of Econometrics Volume 1*. Griliches Z, Intriligator MD. (eds). Elsevier: Amsterdam, 1984; 1247-1318

Guilkey D, Murphy J. Estimation and testing in the random effects probit model. *Journal of Econometrics* 1993; **59**:301-317.

Heckman J. Statistical models for discrete panel data. In *Structural Analysis of discrete data with econometric applications*, edited by C. F. Manski CF, McFadden D. (eds). MIT Press: Cambridge Mass., 1981: 114-178.


### PART II: Classical simulation methods

i) Numerical integration and simulation as alternatives for solving integrals.


Geweke J. Monte Carlo simulation and numerical integration. In *Handbook of Computational Economics*, Amman HM, Kendrick DA, Rust J. (eds). Elsevier: Amsterdam, 1996: 731-800.

Pudney S. *Modelling individual choice: The econometrics of corners, kinks and holes.* Appendix 3-Computation of probability integrals. Blackwell: Oxford, 1989.

Rubinstein R. *Simulation and the Monte Carlo method.* John Wiley: Chichester, 1981.


ii) The GHK and other probability simulators


Borsch-Supan A, Hajivassiliou V. Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics* 1993; **58**:347-368

Geweke J. Efficient simulation from the multivariate normal and student- t distributions subject to linear constraints and the evaluation of constraint probabilities. *Prepared for 'Computing Science and Statistics: 23rd symposium'* Seattle, April 22-24 1991.

Hajivassiliou V. Simulating normal rectangle probabilities and their derivatives: Effects of vectorization. *The International Journal of Supercomputer Applications* 1993; **7**:231-53.

Hajivassiliou V, McFadden D, Ruud P. Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics* 1996*;* **72**:85-134.

Stern S. A method for smoothing simulated moments of discrete probabilities in multinomial probit models. *Econometrica* 1992; 60: 943-952.

Vijverberg V. Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics* 1997; **76**:281-307.

Vijverberg V. Rectangular and wedge-shaped multivariate normal probabilities. *Economics Letters* 2000; **68**:13-20.

(iii) Maximum simulated likelihood

Hajivassiliou V. Some practical issues in maximum simulated likelihood. In *Simulation-based inference in econometrics: methods and applications*. Mariano R, Schuermann T, Weeks MJ. (eds). Cambridge University Press: Cambridge, 2000; 71-99.

Hyslop D. State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica* 1999; **67**:1255-1294.

Lee L-F. Simulated maximum likelihood estimation of dynamic discrete choice statistical models. *Journal of Econometrics* 1997; **82**:1-35.

Lerman S, Manski C. On the use of simulated frequencies to approximate choice probabilities. In *Structural Analysis of discrete data with econometric applications*. C. F. Manski CF, McFadden D. (eds). MIT Press: Cambridge Mass., 1981; 305-319.

Mealli F, Pudney S. Occupational pensions and job mobility in Britain: Estimation of a random-effects competing risks model. *Journal of Applied Econometrics* 1996; **11**:293-320.

(iv) MSS and MSM

Hajivassiliou V, McFadden D. The method of simulated scores for the estimation of LDV models. *Econometrica* 1998; **66**:863-896.

Keane M. A computationally practical simulation estimator for panel data. *Econometrica* 1994; **62**: 95-116.

McFadden D. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 1989; **57**:995-1026

(v) Surveys

Contoyannis P. An introduction to simulation based estimation and inference. *University of York:* http://www.york.ac.uk/res/herc/docs/simulation.pdf [26 November 2001].

Geweke J, Keane M. Computationally intensive methods for integration in econometrics. In *Handbook of Econometrics Volume 5*, Heckman JJ, Leamer E. (eds). Elsevier: Amsterdam, 2001; 3463-3568.

Gourieroux C, Monfort A. Simulation-based inference: A survey with special reference to panel data models. *Journal of Econometrics* 1993; **59**:5-33.

Gourieroux C, Monfort A. *Simulation-based econometric methods*. Oxford University Press: Oxford, 1996.

Hajivassiliou V. Simulation estimation methods for limited dependent variable models In *Handbook of Statistics Volume 11*. Maddala GS, Rao CR, Vinod HD. (eds). Elsevier: Amsterdam, 1993; 519-543.

Hajivassiliou V, Ruud P.Classical estimation methods using simulation. In *Handbook of Econometrics Volume 4*. Engle R, McFadden D. (eds). Elsevier: Amsterdam, 1994; 2383-2441.

Keane M. Simulation estimation for panel data models with limited dependent variables. In *Handbook of Statistics Volume 11*. Maddala GS, Rao CR, Vinod HD. (eds). Elsevier: Amsterdam, 1993; 545-571.

McFadden D, Ruud P. Estimation by simulation. *The Review of Economics and Statistics* 1994*;* **76:**591-608.

Stern S. Simulation-based estimation. *Journal of Economic Literature* 1997, **35**:2006-2039.

**PART III: Bayesian simulation methods**

(i) Introduction to the Bayesian approach.

Lee PM. *Bayesian statistics: an introduction*. Arnold: London, 1997.

(ii) Introduction to MCMC methods.

Casella G, George E. Explaining the Gibbs sampler. *The American Statistician* 1992*;* **46**:167-174.

Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. *The American Statistician* 1995; **49**:327-335.

Gamerman D. *Markov Chain Monte Carlo*.Chapman & Hall: London, 1997.

Gilks WR, Richardson S, Spiegelhalter, D.J. (eds) *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London, 1996.

Smith A, Roberts G. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B* 1993; **55**:3-23.

Tanner TA, Wong WH. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* 1987; **82**: 528-550.

(iii) Bayesian MCMC methods in econometrics.

Bauwens L, Lubrano M, Richard J. *Bayesian inference in dynamic econometric models*. Oxford University Press: Oxford, 1999.

Chib S. Markov chain Monte Carlo methods: computation and inference. In *Handbook of Econometrics Volume 5*, Heckman JJ, Leamer E. (eds). Elsevier: Amsterdam, 2001; 3569-3649.

Geweke J. Using simulation methods for Bayesian econometric models: inference, development, and communication. *Econometric Reviews* 1999; **18**:1-73.

(iv) Some Bayesian MCMC algorithms for limited dependent variable models.

Albert J, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* 1993; **88**: 669-679.

Chib S. Bayesian analysis of correlated binary data. In *Generalized linear models: a Bayesian perspective*. Dey D, Ghosh S, Mallick B. (eds). Marcel-Dekker: New York, 2000; 113-131.

Chib S, Greenberg E. Analysis of multivariate probit models. *Biometrika* 1998; **85**:347-361.

Cowles K, Carlin BP, Connett JE. Bayesian Tobit modelling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of the American Statistical Association* 1996; **91**: 86-98.

Geweke J, Keane, M, and Runkle, D. (1997) Statistical inference in the multinomial multiperiod probit model. *Journal of Econometrics*, 80:125-165

McCulloch RE, Polson NG, Rossi PE. A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics* 2000; **99**: 173-193.

(v) Analysis of convergence.

Cowles M, Carlin B. Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association* 1996; **91**:883-904.

Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian Statistics 4*. Bernardo JM, Berger J, Dawid AP, Smith AFM. (eds). Oxford University Press: Oxford, 1992; 169-193.

(vi) Model uncertainty.

Chib S, Jeliazkov I. *Marginal Likelihood From the Metropolis-Hastings Output. Journal of the American Statistical Association 2001*; **96**:270-281.

Fernandez C, Steel M. Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 2001; **100**: 381-427.

Verdinelli I, Wasserman L. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association* 1995; **90**: 614-618.

**PART IV: Selected applications in health economics**

Bolduc D, Lacroix G, Muller C. The choice of medical providers in rural Benin: a comparison of discrete choice models. *Journal of Health Economics* 1996; **15**:477-498

Borsch-Supan A, Hajivassiliou VA, Kotlikoff LJ, Morris N. Health, children, and elderly living arrangements: a multiperiod multinomial probit model with unobserved heterogeneity and autocorrelated errors. In *Topics in the Economics of Aging*. Wise DA. (ed). University of Chicago Press: Chicago, 1992.

Chib S, Hamilton B. Bayesian analysis of cross-section and clustered data treatment models. *Journal of Econometrics* 2000; **97**:25-50.

Contoyannis P, Jones AM, Rice N. The dynamics of health in British households: simulation-based inference in panel probit models. http://www.york.ac.uk/res/herc/yshe [26 November 2001].

Hamilton B. The impact of HMOs on Medicare costs: Bayesian MCMC estimation of a robust panel data Tobit model with survival, *Health Economics* 1999; **8**:403-414.

Hamilton B. Estimating treatment effects in randomized clinical trials with non-compliance: The impact of maternal smoking on birth-weight. *Health Economics* 2001; **10**: 399-410.

Munkin M, Trivedi, P. Simulated maximum likelihood estimation of multivariate mixed-Poisson models, with application. *Econometrics Journal* 1999; **2**: 29-40.

Pudney S, Shields M. Gender and Racial Discrimination in pay and promotion for NHS nurses. *Oxford Bulletin of Economics and Statistics* 2000; **62** Special Issue: 801-835.

Pudney S, Shields M. Gender, race, pay and promotion in the British nursing profession: estimation of a generalized ordered probit model. *Journal of Applied Econometrics* 2000; **15**:367-399.