# Simulation estimation for panel data models with limited dependent variables

Keane, Michael

1993

G. S. Maddala, C. R. Rao and H. D. Vinod, eds., *Handbook of Statistics, Vol. 11*


20

# Simulation Estimation for Panel Data Models with Limited Dependent Variables

*Michael P. Keane**

## 1. Introduction

Simulation estimation in the context of panel data, limited dependent-variable (LDV) models poses formidable problems that are not present in the cross-section case. Nevertheless, a number of practical simulation estimation methods have been proposed and implemented for panel data LDV models. This paper surveys those methods and presents two empirical applications that illustrate their usefulness.

The outline of the paper is as follows. Section 2 reviews methods for estimating panel data models with serial correlation in the linear case. Section 3 describes the special problems that arise when estimating panel data models with serial correlation in the LDV case. Section 4 presents the essential ideas of method of simulated moments (MSM) estimation, as developed by McFadden (1989) and Pakes and Pollard (1989), and explains why MSM is difficult to apply in the panel data case. Section 5 describes computationally practical simulation estimation methods for the panel data probit model. Section 5.1 describes an efficient algorithm for the recursive simulation of probabilities of sequences of events. This algorithm is at the heart of all the simulation estimators that have proven feasible for panel data LDV models. Section 5.2 describes the simulation estimators for panel data probit models that are based on such recursive simulation of probabilities. Section 5.3 describes some alternative estimators that are based on conditional simulation of the latent variables in the probit model via similar recursive methods. Section 6 discusses issues that arise in simulation estimation of models more complex than the probit model. In Section 7, I use the simulation estimation methods presented in Sections 5 to 6 to estimate probit employment equations and selection

* The Institute for Empirical Macroeconomics and the Alfred P. Sloan Foundation have supported this research. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

bias-adjusted wage equations on panel data from the national longitudinal survey of young men. Section 8 concludes.

Throughout the exposition in Sections 2–6, I assume strict exogeneity of the regressors. I do this in order to focus on the special problems that arise due to simulation itself. Thus, I ignore the important issues that arise when the regressors are endogenous or predetermined rather than strictly exogenous. For discussions of these issues, the reader is referred to the excellent surveys of Heckman (1981) and Chamberlain (1985).

I also ignore simulation estimation in the context of discrete dynamic programming models. This is despite the facts that the first important econometric application of simulation estimation was in this area (Pakes, 1986), the area continues to be a fertile one (see, e.g., Berkovec and Stern, 1991; Hotz and Miller, 1991; and Geweke, Slonim and Zarkin, 1992), and that much of my current research is in this area (Keane and Wolpin, 1992; Erdem and Keane, 1992). This omission stems from my desire to focus on the special problems that arise in the simulation of probabilities of sequences of events, excluding those additional problems that arise when the solution of a dynamic programming problem must also be simulated.

## 2. Methods for estimating panel data models with serial correlation in the linear case

Since the pioneering work of Balestra and Nerlove (1966), the importance of controlling for serial correlation in panel data models has been widely recognized. There are many situations where, if an agent is observed over several time periods, we would expect the errors for that agent to be serially correlated. For instance, in wage data, those workers who have wages that are high at a point in time (after conditioning on the usual human capital variables like education and experience) tend to have persistently high wages over time. As Balestra and Nerlove pointed out, failure to account for such serial correlation when estimating linear regressions on panel data leads to bias in estimates of the standard errors of the regressor coefficients. To deal with this problem, they proposed the random effects model, in which the existence of a time-invariant individual effect, uncorrelated with the regressors and distributed with zero mean in the population, is postulated.

The random effects model produces an error structure that is equicorrelated. That is, if the true model is

$$y_{it} = X_{it}\beta + \varepsilon_{it} \,, \tag{1}$$

for $t = 1, T$ and $I = 1, N$, where $y_{it}$ is the dependent variable for person $i$ at time $t$, $X_{it}$ is a vector of strictly exogenous regressors, and $\varepsilon_{it}$ is the error term, and if

$$\varepsilon_{it} = \mu_i + \omega_{it} \tag{2}$$

where $\mu_i$ is a time-invariant random effect and $\omega_{it}$ is iid, then the covariance structure of the $\varepsilon_{it}$ is

$$\mathrm{E}\varepsilon_{it}\varepsilon_{i,t-j} = \begin{cases} \sigma_\varepsilon^2 & \text{for } j = 0 , \\ \rho\sigma_\varepsilon^2 & \text{for } j \neq 0 . \end{cases} \tag{3}$$

Here $\rho$ is the fraction of the variance of $\varepsilon$ due to the individual random effect. Thus, the correlation between the errors $\varepsilon_{it}$ for any two different time periods is $\rho$ regardless of how far apart the time periods are.

This equicorrelation assumption is obviously unrealistic in many situations. Its virtue lies in the fact that estimation of the random effects model is extremely convenient. The model (1)–(2) may be estimated using a simple two-step GLS procedure that produces consistent and asymptotically efficient estimates of the model parameters and their standard errors. If the equicorrelation assumption is incorrect, the estimates of $\beta$ remain consistent but the estimated standard errors are biased.

In cases where equicorrelation does not hold, it is simple to replace (2) with a general covariance structure and apply the same two-step GLS procedure. In the first step, obtain a consistent estimate of $\beta$ under the assumption that the $\varepsilon$ are iid and use the residuals to estimate the covariance matrix $\Sigma = \mathrm{E}\varepsilon_i\varepsilon_i'$, where $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$ is a $T \times 1$ column vector. Then, letting $\hat{\Sigma}$ denote the estimate of $\Sigma$, take the Cholesky decomposition $\hat{\Sigma} = \hat{A}\hat{A}'$, where $A$ is a lower-triangular matrix, and premultiply the $y_i$ and $X_i$ vectors by $\hat{A}'$. In a second step, estimate a regression of $\hat{A}'y_i$ on $\hat{A}'X_i$ to produce consistent and asymptotically efficient estimates of all model parameters and their standard errors. (See Amemiya and McCurdy, 1986 or Keane and Runkle, 1992.) Note that, with missing data, estimation of an unrestricted $A$ matrix would be problematic. However, restricted structures where $\Sigma$ is parameterized as, say, having random effects and ARMA error components pose no problem.

## 3. The problem of estimating LDV models with serial correlation

In sharp contrast to the linear case, estimation of LDV models with serial corelation poses difficult problems. As a leading case, consider the panel data probit model. This model is obtained if we do not observe $y_{it}$ in equation (1), but only observe the indicator function $d_{it}$, where

$$d_{it} = \begin{cases} 1 & \text{if } y_{it} \geq 0 , \\ 0 & \text{otherwise} , \end{cases} \tag{4}$$

and if we further assume that the error terms have a normal distribution, $\varepsilon_i \sim \mathrm{N}(0, \Sigma)$. Given this structure, we can write $\varepsilon_i = A\eta_i$, where $\eta_i = (\eta_{i1}, \ldots, \eta_{iT})'$ and $\eta_i \sim \mathrm{N}(0, I)$. Define $\theta$ as the vector consisting of elements of $\beta$ and the parameters determining the error covariance structure $\Sigma$. Further, define $J_{it} = \{d_{i1}, \ldots, d_{it}\}$ as the set of choices made by person $i$ through period

$t$, and $\mathrm{Prob}(J_{it} \mid X_i, \theta)$ as the probability of this set, where $X_i = (X_{i1}, \ldots, X_{iT})'$. Then the log-likelihood function evaluated at a trial parameter estimate $\hat{\theta}$ is

$$\mathscr{L}(\hat{\theta}) = \sum_{i=1}^{N} \ln \mathrm{Prob}(J_{iT} \mid X_i, \hat{\theta}) . \tag{5}$$

The difficulty inherent in evaluating this log-likelihood depends on the error structure. If the $\varepsilon_{it}$ are iid, then

$$\mathrm{Prob}(J_{it} \mid X_i, \hat{\theta}) = \prod_{l=1}^{t} \mathrm{Prob}(d_{il} \mid X_{il}, \hat{\theta}) .$$

Thus, only univariate integration is necessary to form the log-likelihood. If there are random effects, as in (2), then

$$\mathrm{Prob}(J_{it} \mid X_i, \hat{\theta}) = \int_{-\infty}^{\infty} \prod_{l=1}^{t} \mathrm{Prob}(d_{il} \mid X_{il}, \hat{\theta}) f(\mu) \, \mathrm{d}\mu .$$

Here, bivariate integration is necessary. If $f(\cdot)$ is the normal density, such bivariate integrations can be evaluated simply using the Gaussian quadrature procedure described by Butler and Moffitt (1982). Unfortunately, for more general error structures, the order of integration necessary is $T$. This makes maximum-likelihood (ML) estimation infeasible for $T \geqslant 4$.

Results in Robinson (1982) indicate that, regardless of the correlation structure of the $\varepsilon_{it}$, if the $\varepsilon_{it}$ are assumed iid, then the resultant misspecified model produces consistent estimates of $\beta$. Such an estimator is inefficient and produces biased estimates of the standard errors. However, a covariance matrix correction is available. Given these results, the value in having a capability to deal with complex serial correlation patterns in LDV models resides in four things. First, there is a potential for efficiency gain in estimating models with richer correlation structures. Second, no proof is available that misspecification of the correlation structure of $\varepsilon_{it}$ results in a consistent estimator of $\beta$ for cases other than that in which $\varepsilon_{it}$ is specified to be iid. Third, in the presence of lagged dependent variables, consistent estimation requires that the serial correlation structure be properly specified. Fourth, and most importantly, allowing for more complex serial correlation patterns can potentially improve out-of-sample prediction of agents' future choice behavior.

## 4. MSM estimation for LDV models

A natural alternative to ML estimation for LDV models is simulation-based estimation, recently studied by McFadden (1989) and Pakes and Pollard (1989). McFadden developed the MSM estimator for the probit model. To motivate the MSM estimator, it is useful to first construct the method of moments (MOM) estimator for the panel data probit model.

To construct the MOM estimator, let $k = 1$, $K$ index all possible choice sequences $J_{iT}$. Let $D_{ik} = 1$ if agent $i$ chooses sequence $k$ and $D_{ik} = 0$ otherwise. Then, following McFadden (1989), the score of the log-likelihood can be written

$$\nabla_{\hat{\theta}}\mathscr{L}(\hat{\theta}) = \sum_{i=1}^{N} \sum_{k=1}^{K} W_{ik}[D_{ik} - \mathrm{Prob}(D_{ik} = 1 \mid X_i, \hat{\theta})] , \tag{6a}$$

where $\hat{\theta}$ is a particular trial parameter estimate and

$$W_{ik} = \frac{\nabla_{\hat{\theta}} \mathrm{Prob}(D_{ik} = 1 \mid X_i, \hat{\theta})}{\mathrm{Prob}(D_{ik} = 1 \mid X_i, \hat{\theta})} . \tag{6b}$$

Note that (6a) has the form of mean zero moments $[D_{ik} - \mathrm{Prob}(D_{ik} = 1 \mid X_i, \hat{\theta})]$ times orthogonal weights $W_{ik}$. Thus, it can be used to form the first-order conditions (FOCs) of an MOM estimator for $\theta$. The MOM estimator, $\hat{\theta}_{\mathrm{MOM}}$, sets the FOC vector in (6a) equal to the zero vector. If the optimal weights $W_{ik}$ are used, this MOM estimator is asymptotically as efficient as ML. Other choices of weights that are asymptotically correlated with the $W_{ik}$ and orthogonal to the residuals produce consistent and asymptotically normal but inefficient MOM estimators. Of course, for general specifications of the error structure, this MOM estimator is not feasible because the choice probabilities are $T$-variate integrals.

The idea of the MSM estimator is to replace the intractable integrals $\mathrm{Prob}(D_{ik} = 1 \mid X_i, \hat{\theta})$ in (6a) by unbiased Monte Carlo probability simulators. The most basic method for simulating the choice probabilities is to draw, for each individual $i$, a set of iid error vectors $(\eta_{i1}, \ldots, \eta_{iT})$ using a univariate normal random number generator and to count the percentage of these vectors that generate $D_{ik} = 1$. This is called the *frequency simulator*. More accurate probability simulators will be discussed below.

Because the simulation error enters linearly into the MSM FOCs, it will tend to cancel over observations. As a result, the MSM estimator based on an unbiased probability simulator is consistent and asymptotically normal in N for a fixed simulation size. If the frequency simulator is used, $\hat{\theta}_{\mathrm{MSM}}$ has an asymptotic covariance matrix that is $(1 + S^{-1})$ times greater than that of $\hat{\theta}_{\mathrm{MOM}}$, where $S$ is the number of draws used in the simulation. Use of more accurate probability simulators improves relative efficiency. If consistent independent simulators of the optimal weights are used, then $\hat{\theta}_{\mathrm{MSM}}$ is asymptotically (in $N$ and $S$) as efficient as ML.

Unfortunately, the MSM estimator in (6a) is not practical to implement. The source of the problem is that $K$ grows large quickly with $T$. In the binomial probit case, $K = 2^T$. Thus, for reasonably large $T$ construction of (6a) requires a very large number of calculations. If a simple frequency simulator is used, such calculations can be done quickly. However, according to McFadden and Ruud (1987), frequency simulation does not appear to work well for this problem. One difficulty is that the FOCs based on frequency simulation are not smooth functions. This makes it impossible to use gradient-based optimization methods. This problem can, however, be dealt with by use of the simplex algorithm. A more serious problem is that the denominators of the optimal weights in (6b) are the probabilities of choice sequences. These probabilities will tend to become very small as $T$ gets large, and frequency simulators based on reasonable numbers of draws will therefore tend to produce simulated probabilities of zero for many choice sequences. This makes it quite difficult to form good approximations to the optimal weights, so that the MSM estimator based on frequency simulation will tend to be very inefficient.

The natural solution to this problem is to use more efficient probability simulators that can accurately simulate small probabilities. Such simulators, based on importance sampling techniques, are considerably more expensive to construct than crude frequency simulators. Thus, it is not practical to use them in conjunction with (6a) to form the FOCs of an MSM estimator. In the next section, I describe a highly efficient algorithm for simulating probabilities of sequences of events and describe practical simulation estimators for panel data probit models based on this algorithm.

## 5. Practical simulation estimators for the panel data probit model

Recently, Keane (1990) and Hajivassiliou and McFadden (1990) have developed computational practical simulation estimators for panel data LDV models. Both methods rely on a highly accurate recursive algorithm for simulating probabilities of sequences of events that I describe in Section 5.1. In Section 5.2, I explain how these simulators can be used to construct practical simulation estimators for the panel data probit model. In Section 5.3, I describe some alternative estimators that are based on conditional simulation of the latent variables in the probit model via similar recursive methods.

### *5.1. Recursive simulation of probabilities of sequences of events*

In Keane (1990), I developed a highly accurate algorithm for simulating the probabilities of choice sequences in panel data probit models. To see the motivation for this method, first observe that the choice $d_{it}=1$ occurs if $\varepsilon_{it} \geqslant -X_{it}\beta$ while the choice $d_{it}=0$ occurs if $-\varepsilon_{it} > X_{it}\beta$. Thus, the boundary of the $\varepsilon_{it}$ distribution conditional on $d_{it}$ is

$$(2d_{it}-1)\varepsilon_{it} \geqslant (1-2d_{it})X_{it}\beta \, .$$

Since $\varepsilon_i = A\eta_i$, this constraint may be written

$$(2d_{it}-1)\eta_{it} \geqslant \frac{(1-2d_{it})X_{it}\beta - (2d_{it}-1)(A_{t1}\eta_{i1} + \cdots + A_{t,t-1}\eta_{i,t-1})}{A_{tt}} \, .$$

Recall that $J_{it} = \{d_{i1}, \ldots, d_{it}\}$ denotes the set of choices made by person $i$ in periods 1 through $t$. Further define

$$\eta(J_{i1}) = \{\eta_{i1} \mid (2d_{i1} - 1)\eta_{i1} > (1 - 2d_{i1})X_{i1}\beta\} ,$$

$$\eta(J_{it}) = \Big\{\eta_{i1}, \ldots, \eta_{it} \mid (2d_{is} - 1)\eta_{is} > \frac{(1 - 2d_{is})X_{is}\beta - (2d_{is} - 1)(A_{s1}\eta_{i1} + \cdots + A_{s,s-1}\eta_{i,s-1})}{A_{ss}} \text{ for all } s \leq t\Big\} . \quad (7)$$

These are the sets of $\eta_i$ vectors that are consistent with the set of choices made by person $i$ in periods 1 through $t$. The probability of a choice sequence $\text{Prob}(J_{it} \mid X_i, \hat{\theta})$ can be factored into a first-period unconditional choice probability times transition probabilities as follows:

$$\begin{aligned}
&\text{Prob}(J_{it} \mid X_i, \hat{\theta}) \\
&\quad = \text{Prob}(\eta_{i1}, \ldots, \eta_{it} \in \eta(J_{it})) \\
&\quad = \text{Prob}(\eta_{i1} \in \eta(J_{i1}))\, \text{Prob}(\eta_{i1}, \eta_{i2} \in \eta(J_{i2}) \mid \eta_{i1} \in \eta(J_{i1})) \\
&\qquad \times \cdots \times \text{Prob}(\eta_{i1}, \ldots, \eta_{it} \in \eta(J_{it}) \mid \eta_{i1}, \ldots, \eta_{i,t-1} \in \eta(J_{i,t-1})) . \quad (8)
\end{aligned}$$

An unbiased simulator of this probability may be obtained by the following sequential procedure:

(1) Draw an $\eta_{i1}$ from the truncated univariate normal distribution such that $\eta_{i1} \in \eta(J_{i1})$. Call the particular value that is drawn $\eta_{i1}^*$.

(2) Given $\eta_{i1}^*$, there is a range of $\eta_{i2}$ values such that

$$(2d_{i2} - 1)\eta_{i2} > [(1 - 2d_{i2})X_{i2}\beta - (2d_{i2} - 1)A_{21}\eta_{i1}^*]/A_{22} .$$

Using the notation of (7), I denote this set of $\eta_{i2}$ values by $\{\eta_{i2} \mid \eta_{i1}^*, \eta_{i2} \in \eta(J_{i2})\}$. Draw an $\eta_{i2}$ from a truncated univariate normal distribution such that $(\eta_{i1}^*, \eta_{i2}) \in \eta(J_{i2})$. Call the particular value that is drawn $\eta_{i2}^*$.

(3) Continue in this way until a vector $(\eta_{i1}^*, \ldots, \eta_{i,T-1}^*) \in \eta(J_{i,t-1})$ is obtained.

(4) Form the simulator:

$$\begin{aligned}
\widehat{\text{Prob}}(J_{it} \mid X_i, \hat{\theta}) &= \text{Prob}(\eta_{i1} \in \eta(J_{i1})) \times \text{Prob}(\eta_{i2} \in \eta(J_{i2}) \mid \eta_{i1}^*) \\
&\quad \times \cdots \times \text{Prob}(\eta_{it} \in \eta(J_{it}) \mid \eta_{i1}^*, \ldots, \eta_{i,t-1}^*) . \quad (9)
\end{aligned}$$

This probability simulator has been named the *Geweke–Hajivassiliou–Keane* or *GHK simulator* by Hajivassiliou, McFadden and Ruud (1992) because related independent work by Geweke (1991a) and Hajivassiliou led to the development of the same method. In an extensive study of alternative probability simulators, they find that the GHK simulator is the most accurate of all those considered.

Note that simulators of the transition probabilities in (8) can be obtained by the same method. In Keane (1990), I showed that, for $t \geqslant 3$, an unbiased simulator of the transition probabilities is given by

$$\widehat{\operatorname{Prob}}(d_{it} \mid J_{i,t-1}, X_i, \hat{\theta})$$
$$= \operatorname{Prob}(\eta_{it} \in \eta(J_{it}) \mid \eta_{i1}^*, \ldots, \eta_{i,t-1}^*)\omega(\eta_{i1}^*, \ldots, \eta_{it-1}^*)\,, \tag{10a}$$

where

$$\omega(\eta_{i1}^*, \ldots, \eta_{i,t-1}^*)$$
$$= \frac{\omega(\eta_{i1}^*, \ldots, \eta_{i,t-2}^*)\operatorname{Prob}(\eta_{i,t-1} \in \eta(J_{i,t-1}) \mid \eta_{i1}^*, \ldots, \eta_{i,t-2}^*)}{\operatorname{Prob}(\eta_{i,t-1} \in \eta(J_{i,t-1}) \mid \eta_{i1}, \ldots, \eta_{i,t-2} \in \eta(J_{i,t-2}))}$$
$$= \frac{\operatorname{Prob}(\eta_{i,t-1} \in \eta(J_{i,t-1}) \mid \eta_{i1}^*, \ldots, \eta_{i,t-2}^*) \cdots \operatorname{Prob}(\eta_2 \in \eta(J_{i2}) \mid \eta_{i1}^*)\operatorname{Prob}(\eta_{i1} \in \eta(J_{i1}))}{\operatorname{Prob}(\eta_{i1}, \ldots, \eta_{i,t-1} \in \eta(J_{i,t-1}))} \tag{10b}$$

This procedure may be interpreted as importance sampling where the transition probability is simulated conditional on the draw $\eta_{i1}^*, \ldots, \eta_{i,t-1}^*$ from the importance sampling density defined by steps (1)–(3) and $\omega(\eta_{i1}^*, \ldots, \eta_{i,t-1}^*)$ is the importance sampling weight. The form of the weight is the ratio of (1) the probability of event sequence $d_{i1}, \ldots, d_{i,t-1}$ as simulated by the GHK method using the draw $\eta_{i1}^*, \ldots, \eta_{i,t-1}^*$ to (2) the actual probability of the event sequence $d_{i1}, \ldots, d_{i,t-1}$.

Unfortunately, for $t-1 \geqslant 3$ it is not feasible to numerically evaluate the object $\operatorname{Prob}(\eta_{i1}, \ldots, \eta_{i,t-1} \in \eta(J_{i,t-1}))$ that appears in the denominator of the importance sampling weights. However, this probability may itself be simulated by the GHK method. If this is done, a denominator bias is induced, and the resultant transition probability simulator will be asymptotically unbiased as the number of draws used to form the GHK simulator becomes large.

Let $S$ be the number of draws used to simulate the choice sequence and transition probabilities by the GHK method. Letting $(\eta_{i1s}^*, \ldots, \eta_{i,T-1,s}^*)$ be the $s$-th sequence drawn in the GHK procedure, one obtains, for the simulated sequence probabilities,

$$\widehat{\operatorname{Prob}}(J_{it} \mid X_i, \hat{\theta}) = \frac{1}{S} \sum_{s=1}^{S} \operatorname{Prob}(\eta_{i1} \in \eta(J_{i1}))\operatorname{Prob}(\eta_{i2} \in \eta(J_{i2}) \mid \eta_{i1s}^*)$$
$$\times \cdots \times \operatorname{Prob}(\eta_{it} \in \eta(J_{it}) \mid \eta_{i1s}^*, \ldots, \eta_{i,t-1,s}^*) \tag{11}$$

and for the simulated transition probabilities,

$$\widehat{\text{Prob}}(d_{it} \mid J_{i,t-1}, X_i, \hat{\theta}) \\ = \frac{1}{S} \sum_{s=1}^{S} \text{Prob}(\eta_{it} \in \eta(J_{it}) \mid \eta^*_{i1s}, \ldots, \eta^*_{i,t-1,s}) \hat{\omega}(\eta^*_{i1s}, \ldots, \eta^*_{i,t-1,s}) , \tag{12a}$$

where

$$\hat{\omega}(\eta^*_{i1s}, \ldots, \eta^*_{i,t-1,s}) \\ = \frac{\text{Prob}(\eta_{i,t-1} \in \eta(J_{i,t-1}) \mid \eta^*_{i1s}, \ldots, \eta^*_{i,t-2,s}) \cdots \text{Prob}(\eta_{i2} \in \eta(J_{i2}) \mid \eta^*_{i1s})}{S^{-1} \sum_{r=1}^{S} \text{Prob}(\eta_{i,t-1} \in \eta(J_{i,t-1}) \mid \eta^*_{i1r}, \ldots, \eta^*_{i,t-2,r}) \cdots \text{Prob}(\eta_{i2} \in \eta(J_{i2}) \mid \eta^*_{i1r})} \tag{12b}$$

for $t \geqslant 3$ and

$$\widehat{\text{Prob}}(d_{i2} \mid J_{i1}, X_i, \hat{\theta}) = \frac{1}{S} \sum_{s=1}^{S} \text{Prob}(\eta_{i2} \in \eta(J_{i2}) \mid \eta^*_{i1s})$$

for $t = 2$. Note that if the importance sampling weights are simulated as in (12b), they are constrained to sum to one by construction. Constraining importance sampling weights to sum to one is a standard variance reduction technique often recommended in the numerical analysis literature. In (12b), the simulation error in the numerator is positively correlated with that in the denominator, so in some cases a variance reduction in simulation of the ratio may be achieved by use of the simulated rather than the true denominator.

### 5.2. *Practical simulation methods for panel data probit models based on recursive simulation of probabilities*

Three classical methods of estimation for panel data probit models have been implemented in the literature, all based on the GHK method for simulation of sequence and transition probabilities. In Keane (1990), I expressed the log-likelihood function as a sum of transition probabilities

$$\mathscr{L}(\hat{\theta}) = \sum_{i=1}^{N} \ln \text{Prob}(J_{iT} \mid X_i, \hat{\theta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \ln \text{Prob}(d_{it} \mid J_{i,t-1}, X_i, \hat{\theta})$$

and proceed to express the score as

$$\nabla_{\hat{\theta}} \mathscr{L}(\hat{\theta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \{W^1_{it}[d_{it} - \text{Prob}(d_{it} = 1 \mid J_{i,t-1}, X_i, \hat{\theta})] \\ + W^0_{it}[(1 - d_{it}) - \text{Prob}(d_{it} = 0 \mid J_{i,t-1}, X_i, \hat{\theta})]\} , \tag{13}$$

where the weights $W_{it}^1$ and $W_{it}^0$ have the form

$$W_{it}^1 = \frac{\nabla_{\hat\theta} \operatorname{Prob}(d_{it} = 1 \mid J_{i,t-1}, X_i, \hat\theta)}{\operatorname{Prob}(d_{it} = 1 \mid J_{i,t-1}, X_i, \hat\theta)},$$

$$W_{it}^0 = \frac{\nabla_{\hat\theta} \operatorname{Prob}(d_{it} = 0 \mid J_{i,t-1}, X_i, \hat\theta)}{\operatorname{Prob}(d_{it} = 0 \mid J_{i,t-1}, X_i, \hat\theta)}. \tag{14}$$

Note that (13) has the form of mean zero moments times orthogonal weights. Thus, it can be used to form the FOCs of an MOM estimator for $\theta$, where $\hat\theta_{\mathrm{MOM}}$ sets (13) to zero and the optimal weights are given by (14).

In Keane (1990), I formed an MSM estimator by substituting the simulated transition probabilities given by (12) into equation (13) and using independent simulations of the transition probabilities to simulate the optimal weights in (14). Using results in McFadden and Ruud (1991), I showed in Keane (1992) that the resultant MSM estimator is consistent and asymptotically normal if $S/\sqrt{N} \to \infty$ as $N \to \infty$. In a series of repeated sampling experiments on models with random effects plus AR(1) error components, setting $S = 10$, $N = 500$, and $T = 8$, I also showed that the bias in this MSM estimator is negligible, even when the degree of serial corelation is very strong.

The generalization of this method to more than two alternatives is straightforward and is discussed in Keane (1990). Elrod and Keane (1992) successfully applied this MSM estimator to detergent choice models with eight alternatives and up to 30 time periods per household. By allowing for a complex pattern of serial correlation, Elrod and Keane were able to produce more accurate out-of-sample forecasts of agents' future choices than could be obtained with simpler models. This is a good illustration of why the ability to estimate LDV models with complex patterns of serial correlation is important.

Hajivassiliou and McFadden (1990) expressed the score of the log-likelihood as

$$\nabla_{\hat\theta} \mathscr{L}(\hat\theta) = \sum_{i=1}^{N} \frac{\nabla_{\hat\theta} \operatorname{Prob}(J_{iT} \mid X_i, \hat\theta)}{\operatorname{Prob}(J_{iT} \mid X_i, \hat\theta)}. \tag{15}$$

They implemented a method of simulated scores (MSS) estimator by using the GHK probability simulator in (11) to simulate the numerator and denominator of (15). $\hat\theta_{\mathrm{MSS}}$ is obtained by setting the simulated score vector to zero. Hajivassiliou and McFadden showed that $\hat\theta_{\mathrm{MSS}}$ is consistent and asymptotically normal if $S/\sqrt{N} \to \infty$ as $N \to \infty$.

A third alternative is simply to implement a simulated maximum likelihood (SML) estimator by using the GHK probability simulator to simulate the log-likelihood function (5) directly. $\hat\theta_{\mathrm{SML}}$ maximizes the simulated log-likelihood function. By construction, $\hat\theta_{\mathrm{SML}}$ is also a root of the simulated score expression (15), provided the same smooth probability simulators (with the same draws) are used in both. Thus the MSS estimator given by applying the

GHK simulator to (15) is identical to the SML estimator obtained by applying the GHK simulator to (4). $\hat{\theta}_{\mathrm{SML}}$ is also consistent and asymptotically normal if $S/\sqrt{N} \rightarrow \infty$ as $N \rightarrow \infty$. See Gourieroux and Monfort (1991) for a proof.

Hajivassiliou and McFadden (1990) reported good results using the MSS procedure based on the GHK simulator with 20 draws to estimate panel data probit models in which the existence of repayment problems for less-developed countries is the dependent variable. Börsch-Supan, Hajivassiliou, Kotlikoff and Morris (1991) used the SML approach based on the GHK simulator to estimate panel data probit models where choice of living arrangements is the dependent variable.

In Keane (1992), I reported repeated sampling experiments for SML based on the GHK simulator, using the same experiment design I used to study the MSM estimator. In this experiment, SML based on GHK with $S = 10$ exhibits negligible bias when the degree of serial correlation is not extreme. However, in experiments on a model with AR(1) errors and an individual effect, with $\rho = 0.20$ and the AR(1) parameter set to 0.90, the SML estimator greatly overstates the fraction of variance due to the individual effect and understates the AR(1) parameter. The MSM estimator based on the GHK simulator does not exhibit this problem.

Finally, McFadden (1992) observed that the FOCs used in Keane (1990, 1992) can be rewritten in such a way that they have the form of weights times mean zero moments. The FOCs used in Keane (1990, 1992), obtained by substituting the simulators (12) into equation (13), have the form

$$
\begin{aligned}
\mathrm{FOC}(\hat{\theta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \Bigg\{ \hat{W}_{it}^{1} \Bigg[ d_{it} - \frac{1}{S} \sum_{s=1}^{S} & \mathrm{Prob}(d_{it} = 1 \mid \eta_{i1s}^{*}, \ldots, \eta_{i,t-1,s}^{*}) \\
& \times \omega(\eta_{i1s}^{*}, \ldots, \eta_{i,t-1,s}^{*}) \Bigg] \\
+ \hat{W}_{it}^{0} \Bigg[ (1 - d_{it}) - \frac{1}{S} \sum_{s=1}^{S} & \mathrm{Prob}(d_{it} = 0 \mid \eta_{i1s}^{*}, \ldots, \eta_{1,t-1,s}^{*}) \\
& \times \omega(\eta_{i1s}^{*}, \ldots, \eta_{i,t-1,s}^{*}) \Bigg] \Bigg\},
\end{aligned}
$$

where the importance sampling weights $\omega(\eta_{i1s}^{*}, \ldots, \eta_{i,t-1,s}^{*})$ are given in (12b) and the weights $\hat{W}_{it}^{1}$ and $\hat{W}_{it}^{0}$ are simulations by the GHK method of the optimal weights given in (14).

Define $\omega(\eta_{i1s}^{*}, \ldots, \eta_{i,t-1,s}^{*}) = \omega_{Ai,t-1,s}^{*} / \omega_{Bi,t-1,S}^{*}$, where $\omega_{Ai,t-1,s}^{*}$ is the numerator of (12b) and $\omega_{Bi,t-1,S}^{*}$ is the denominator. Then the FOCs can be rewritten as

$$
\begin{aligned}
\mathrm{FOC}(\hat{\theta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \Bigg\{ \frac{\hat{W}_{it}^{1}}{\omega_{Bi,t-1,S}^{*}} \Bigg[ & d_{it} \omega_{Bi,t-1,S}^{*} \\
& - \frac{1}{S} \sum_{s=1}^{S} \mathrm{Prob}(d_{it} = 1 \mid \eta_{i1s}^{*}, \ldots, \eta_{i,t-1,s}^{*}) \omega_{Ai,t-1,s}^{*} \Bigg]
\end{aligned}
$$

$$+\frac{\hat{W}^0_{it}}{\omega^*_{Bi,t-1,S}}\left[(1-d_{it})\omega^*_{Bi,t-1,S}\right.$$
$$\left.-\frac{1}{S}\sum_{s=1}^{S}\text{Prob}(d_{it}=0\,|\,\eta^*_{i1s},\ldots,\eta^*_{i,t-1,s})\omega^*_{Ai,t-1,s}\right]\Bigg\}.$$

The terms in brackets are now mean zero residuals, so an MSM estimator based on these FOCs is consistent and asymptotically normal for fixed $S$. The potential drawback of this procedure is that, while the optimal weights $W^1_{it}$ and $W^0_{it}$ for the Keane (1990, 1992) estimator have transition probabilities in the denominator, the optimal weights $W^1_{it}/\omega^*_{Bi,t-1}$ and $W^0_{it}/\omega^*_{Bi,t-1}$ for this new estimator (where $\omega^*_{Bi,t-1}=\text{E}(\omega^*_{Bi,t-1,S})$) have sequence probabilities in the denominator. Since sequence probabilities will generally be very small relative to transition probabilities, the denominator bias in simulation of the optimal weights will tend to become more severe, and efficiency relative to ML may deteriorate. An important avenue for future research is to explore the small-sample properties of this estimator.

### 5.3. *Alternative methods based on conditional simulation of the latent variables in the LDV model*

Hajivassiliou and McFadden (1990) discussed a fourth classical method for estimating panel data probit models that has not yet been implemented in the literature. This is based on the idea, due to Van Praag and Hop (1987) and Ruud (1991), that the score can be written in terms of the underlying latent variables of the model as follows:

$$\nabla_{\beta}\mathscr{L}(\hat{\theta})=\sum_{i=1}^{N}X_i'\hat{E}^{-1}\text{E}[y_i-X_i\hat{\beta}\,|\,J_{iT}]\,,$$
$$\nabla_{\hat{A}}\mathscr{L}(\hat{\theta})=\sum_{i=1}^{N}\{-\hat{\Sigma}^{-1}\hat{A}+\hat{\Sigma}\text{E}[(y_i-X\hat{\beta})(y_i-X\hat{\beta})'\,|\,J_{iT}]\hat{\Sigma}^{-1}\hat{A}\}\,. \quad (16)$$

Unbiased simulators of this score expression can be obtained if the error terms $\varepsilon_i=y_i-X_i\beta$ can be drawn from the conditional distribution determined by $J_{iT}$, $X_i$, and $\hat{\theta}$ as in equation (7). Given such draws, unbiased simulators of the conditional expectations in (16) may be formed. An MSS estimator that sets the resultant simulated score vector to zero is consistent and asymptotically normal for fixed $S$. The first application of this MSS procedure was by Van Praag and Hop (1987). They used MSS to estimate a cross section tobit model, for which it is feasible to draw error vectors from the correct conditional distribution.

Of course, it is difficult to draw the $\varepsilon_i$ directly from complex conditional distributions such as that given by (7). One method, investigated by Albert and Chib (1993), Geweke (1991a,b) and McCulloch and Rossi (1992), is Gibbs sampling. The Gibbs sampling procedure is related to the GHK sampling scheme described earlier in that it requires recursive draws from univariate normals. Steps (1)–(3) of the GHK procedure generate a vector $\eta_{i1}^*, \ldots, \eta_{iT}^*$ that is drawn from an importance sampling distribution rather than from the true multivariate distribution of $\eta_i$ conditional on $J_{iT}$, $X_i$, and $\hat{\theta}$. However, under mild conditions, the Gibbs sampling procedure produces, asymptotically, draws from the correct distribution. To implement the Gibbs procedure, first implement steps (1)–(3) of the GHK procedure to obtain a starting vector $(\eta_{i1}^*, \ldots, \eta_{iT}^*)$. Note that step (3) must be amended so the $\eta^*$ vector is extended out completely to time $T$. Then perform the following steps:

(1) Starting with an initial vector $(\eta_{i1}^*, \ldots, \eta_{iT}^*)$, drop $\eta_{i1}^*$ and draw a new $\eta_{i1}$ from the truncated univariate normal distribution such that $(\eta_{i1}, \eta_{i2}^*, \ldots, \eta_{i3}^*) \in \eta(J_{iT})$. Replace the old value of $\eta_{i1}^*$ with the new draw for $\eta_{i1}$.

(2) Starting with the vector $(\eta_{i1}^*, \ldots, \eta_{iT}^*)$ from step (1), drop $\eta_{i2}^*$ and draw a new $\eta_{i2}$ from the truncated univariate normal distribution such that $(\eta_{i1}^*, \eta_{i2}, \eta_{i3}^*, \ldots, \eta_{iT}^*) \in \eta(J_{iT})$. Replace the old value of $\eta_{i2}^*$ with the new draw for $\eta_{i2}$.

(3) Continue in this way until a complete new vector $(\eta_{i1}^*, \ldots, \eta_{iT}^*) \in \eta(J_{iT})$ is obtained.

(4) Return to step (1) and, using the $\eta_i^*$ vector from step (3) as the new initial $\eta$ vector, obtain a new draw for $\eta_{i1}$, etc.

Steps (1)–(3) are called a *cycle* of the Gibbs sampler. Suppose that steps (1)–(3) are repeated $C$ times, always beginning step (1) with the $\eta_i^*$ vector that was obtained from the previous cycle. Gelfand and Smith (1990) showed that, under mild conditions, as $C \to \infty$ the distribution of $(\eta_{i1}^*, \ldots, \eta_{iT}^*)$ converges to the true conditional distribution at a geometric rate. Hajivassiliou and McFadden (1990) showed that using Gibbs sampling to simulate the score expression (16) results in an estimator that is consistent and asymptotically normal if $C/\log N \to \infty$ as $N \to \infty$.

The drawback of the Gibbs sampling approach to simulating the score expression (16) is that each time the trial parameter estimate $\hat{\theta}$ is updated in the search for $\hat{\theta}_{\text{MSS}}$ the Gibbs sampler must converge. I am not aware of applications in cross-section or panel data settings. (Recall that Hajivassiliou and McFadden, 1990 actually implemented $\hat{\theta}_{\text{MSS}}$ based on (15) in their work.)

An alternative GHK-like approach may also be used to simulate the conditional expectations in (16). As both Van Praag and Hop (1987) and Keane (1990) noted, weighted functions of the $(\eta_{i1}^*, \ldots, \eta_{iT}^*)$ vectors obtained by steps (1)–(3) of the GHK procedure, with importance sampling weights of the form (10b), give unbiased estimators of the conditional expectations in (16). That is, given a set of vectors $(\eta_{i1s}^*, \ldots, \eta_{iTs}^*)$ for $s = 1, S$ obtained by steps (1)–(3a) of the GHK procedure, one obtains unbiased simulators

$$\hat{\mathrm{E}}[\varepsilon_{it} \mid J_{iT}] = \frac{1}{S} \sum_{s=1}^{S} \{A_{t1}\eta_{i1s}^* + \cdots + A_{tt}\eta_{its}^*\}\omega(\eta_{i1s}^*, \ldots, \eta_{i,T-1,s}^*) ,$$

$$\begin{aligned}\hat{\mathrm{E}}[\varepsilon_{it}\varepsilon_{iu} \mid J_{iT}] = \frac{1}{S} \sum_{s=1}^{S} &\{A_{t1}\eta_{i1s}^* + \cdots + A_{tt}\eta_{its}^*\} \\ &\times \{A_{u1}\eta_{i1s}^* + \cdots + A_{uu}\eta_{i1s}^*\}\omega(\eta_{i1s}^*, \ldots, \eta_{i,T-1,s}^*) . \quad (17)\end{aligned}$$

Of course, as was discussed above, it is not feasible to construct the exact weights when $T-1 \geqslant 3$. In that case the weights must be simulated as in (12b), and the resultant conditional expectation simulators will only be asymptotically unbiased in $S$. An estimator based on substituting the expectation simulators (17) into the score expression (16) has not been tried in the panel data case.

Albert and Chib (1993), Geweke (1991b) and McCulloch and Rossi (1992) have observed that Gibbs sampling may be used as a Bayesian inference procedure, rather than merely as a computational device for simulating the conditional expectations in (16). This procedure has the following steps.

(1) Given a starting parameter value $\hat{\theta} = (\hat{\beta}_0, \hat{A}_0)$ and an initial vector $\hat{\varepsilon}_0$, use steps (1)–(3) of the Gibbs sampler described above to obtain a draw $\hat{\varepsilon}_1$ from the distribution of $\varepsilon$ conditional on $J_T$, $X$, $\hat{\beta}_0$, and $\hat{A}_0$.

(2) Construct $\hat{y}_1 = X\hat{\beta}_0 + \hat{\varepsilon}_1$. Regress $\hat{y}_1$ on $X$, using a seemingly unrelated regression framework to account for the cross-equation correlations determined by $\hat{A}_0$. The resultant point estimates and variance–covariance matrix for the $\beta$ vector give the normal distribution of $\beta$ conditional on $J_T$, $X$, $\hat{y}_1$, and $\hat{A}_0$. Draw $\hat{\beta}_1$ from this conditional distribution.

(3) Given $\hat{y}_1$ and $\hat{\beta}_1$, we may form the residuals from the regression. These residuals determine an inverse Wishart distribution of $\Sigma$ conditional on $J_T$, $X$, $\hat{y}_1$, and $\hat{\beta}_1$. Draw $\hat{\Sigma}_1$ from this conditional distribution, and form $\hat{A}_1$.

(4) Return to step (1), using $\hat{\varepsilon}_1$ as the new initial $\varepsilon$ vector, and obtain a new draw $\hat{\varepsilon}_2$ from the distribution of $\varepsilon$ conditional on $J_T$, $X$, $\hat{\beta}_1$, and $\hat{A}_1$.

Steps (1)–(3) are a cycle of the Gibbs sampling inference procedure. Observe that $\varepsilon$, $\beta$, and $A$ have a joint conditional distribution given by $X$ and the observed choice sequences $J_T$. These can be decomposed into conditionals, and steps (1)–(3) represent sequential draws from these conditionals. Thus, the Gelfand and Smith (1990) result holds. Letting $C$ index cycles, if steps (1)–(3) are repeated $C$ times, then as $C \to \infty$, the distribution of $(\hat{\varepsilon}_C, \hat{\beta}_C, \hat{A}_C)$ for $C > C^*$ can be used to integrate the true joint distribution of $\varepsilon$, $\beta$, and $A$ by Monte Carlo. Both Geweke (1991b) and McCulloch and Rossi (1992) show how priors for $\beta$ and $A$ may be incorporated into this framework by simple modifications of the normal and inverse Wishart distributions from which $\hat{\beta}$ and $\hat{A}$ are drawn on steps (2)–(3).

This Gibbs sampling inference procedure has been applied successfully to cross-section probit problems by McCulloch and Rossi (1992) and Geweke, Keane and Runkle (1992), and to cross-section tobit models by Chib (1993) and Geweke (1991b). McCulloch and Rossi (1992) have also successfully applied the method in a panel data setting. They estimate a probit model on margarine brand choice data, allowing for random effects in the brand intercepts and in the price coefficient.

The simulated EM algorithm, due to Van Praag and Hop (1987) and Ruud (1991), is a method for obtaining $\hat{\theta}_{\mathrm{MSS}}$ that is closely related to the Gibbs sampling inference procedure. The essential difference is that, on steps (2) and (3), which correspond to the M or 'maximization' step of the EM algorithm, the point estimates for $\beta$ and $A$ are used rather than taking draws for $\beta$ and $A$ from the estimated conditional distributions. With this amendment, repetition of steps (1)–(3) results in convergence of $(\hat{\beta}_C, \hat{A}_C)$ to a consistent and asymptotically normal point estimate as $C \rightarrow \infty$. Note that step (1) of the Gibbs inference procedure corresponds to the E or 'expectation' step of the EM algorithm. Here, any method for forming the conditional expectations in (16) may be substituted for the Gibbs sampler. Applications of the simulated EM algorithm to cross-section LDV models can be found in Van Praag and Hop (1987) and in Van Praag, Hop and Eggink (1991), who draw directly from conditional distributions in the E step. To my knowledge the simulated EM algorithm has not yet been applied in a panel data case.

## 6. Extensions to more general models

In deciding which simulation estimation method to use in a particular application, it is important to recognize that there are some models that are difficult to put in an MSM framework. This point was made by McFadden and Ruud (1991). Consider the case of the selection model:

$$w_{it} = \begin{cases} X_{it}\gamma + v_{it} & \text{if } d_{it} = 1 , \\ \text{unobserved} & \text{otherwise} , \end{cases} \tag{18}$$

For $t = 1, T$, $i = 1, N$, where $w_{it}$ is a continuous variable (1), that is observed only if $d_{it} = 1$, $X_{it}$ is the same vector of exogenous regressors as in (1), $\gamma$ is the corresponding coefficient vector, and $v_{it}$ is the error term. Redefine $J_{it}$ to include the $w_{it}$, giving $J_{it} = \{d_{i1}, w_{i1}, \ldots, d_{it}, w_{it}\}$. Let $w_{it}$ have conditional density $f(w_{it} \mid J_{i,t-1}, X_i, \hat{\theta})$. Assume that $\varepsilon_i$ and $v_i$ are jointly normally distributed with covariance matrix $\Sigma$. Any exclusion restrictions in the model (i.e., variables in $X$ that affect $y$ but not $w$) are represented by restricting to zero the appropriate elements of $\gamma$.

As is discussed in Heckman (1979), OLS estimation of (18) using only observations where $d_{it} = 1$ produces biased estimates of $\beta$ when $\varepsilon_i$ and $v_i$ are correlated. Thus, equations (1), (4), and (18) must be estimated jointly. The log-likelihood function for the selection model given by (1), (4), and (18) is

$$\mathscr{L}(\hat{\theta}) = \sum_{i=1}^{N} \Big\{ \sum_{t \in U_i} \ln \operatorname{Prob}(d_{it} = 0 \mid J_{i,t-1}, X_i, \hat{\theta}) + \sum_{t \in E_i} \ln \operatorname{Prob}(d_{it} = 1 \mid J_{t,t-1}, w_{it}, X_i, \hat{\theta}) f(w_{it} \mid J_{i,t-1}, X_i, \hat{\theta}) \Big\} ,$$

where $U_i$ is the set of time periods for which $d_{it} = 0$ and $E_i$ is the set of time periods for which $d_{it} = 1$.

The score for this likelihood may be written as

$$\nabla_{\hat\theta}\mathcal{L}(\hat\theta) = \sum_{i=1}^{N}\left\{\sum_{t\in U_i}\{W_{it}^{0}[(1-d_{it}) - \text{Prob}(d_{it}=0\,|\,J_{i,t-1}, X_i, \hat\theta)] + W_{it}^{1}[d_{it} - \text{Prob}(d_{it}=1\,|\,J_{i,t-1}, X_i, \hat\theta)]\} + \sum_{t\in E_i}\{W_{it}^{2}[(1-d_{it}) - \text{Prob}(d_{it}=0\,|\,J_{i,t-1}, w_{it}, X_i, \hat\theta)] + W_{it}^{3}[d_{it} - \text{Prob}(d_{it}=1\,|\,J_{i,t-1}, w_{it}, X_i, \hat\theta)] + \nabla_{\hat\theta}\ln f(w_{it}\,|\,J_{i,t-1}, X_i, \hat\theta)\}\right\}, \tag{19a}$$

where

$$\begin{aligned} W_{it}^{0} &= \nabla_{\hat\theta}\ln \text{Prob}(d_{it}=0\,|\,J_{i,t-1}, X_i, \hat\theta)\,, \\ W_{it}^{1} &= \nabla_{\hat\theta}\ln \text{Prob}(d_{it}=1\,|\,J_{i,t-1}, X_i, \hat\theta)\,, \\ W_{it}^{2} &= \nabla_{\hat\theta}\ln \text{Prob}(d_{it}=0\,|\,J_{i,t-1}, w_{it}, X_i, \hat\theta)\,, \\ W_{it}^{3} &= \nabla_{\hat\theta}\ln \text{Prob}(d_{it}=1\,|\,J_{i,t-1}, w_{it}, X_i, \hat\theta)\,. \end{aligned} \tag{19b}$$

Notice that (19a) is not interpretable as the FOCs for an MOM estimator because the objects $[(1-d_{it}) - \text{Prob}(d_{it}=0\,|\,J_{i,t-1}, w_{it}, x_i, \hat\theta)]$ and $[d_{it} - \text{Prob}(d_{it}=1\,|\,J_{i,t-1}, w_{it}, X_i, \hat\theta)]$ are not mean zero residuals in the population for which $t\in E_i$ due to the correlation between $v_{it}$ and $\varepsilon_{it}$. Furthermore, the expression $\nabla_{\hat\theta}\ln f(w_{it}\,|\,J_{i,t-1}, X_i, \hat\theta)$ can be written in terms of objects $[w_{it} - X_{it}\beta - E(v_{it}\,|\,J_{i,t-1}, X_i, \hat\theta)]$ times weights, but these objects also have nonzero expectation in the population for which $t\in E_i$ because of the correlation between $v_{it}$ and $\varepsilon_{it}$. Thus (19a)–(19b) cannot be used to construct an MSM estimator. If the score as given by (19a)–(19b) is simulated using unbiased simulators for the choice probabilities, including those in the numerator and denominator of the $W_{it}^{j}$ for $j=0,\ 3$, then it is an MSS situation, where consistency and asymptotic normality are achieved only if $S/\sqrt{N}\to\infty$ as $N\to\infty$ because of the bias created by simulating the denominators of the $W_{it}^{j}$.

McFadden and Ruud (1991) discussed a bias correction technique that can be used to put a large class of models, including the selection model, into an MSM framework. The score contribution of person $i$ at $t$ is given by

$$\nabla_{\hat\theta}\mathcal{L}_{it}(\hat\theta) = (1-d_{it})\nabla_{\hat\theta}\ln \text{Prob}(d_{it}=0\,|\,J_{i,t-1}, X_i, \hat\theta) + d_{it}\nabla_{\hat\theta}\ln \text{Prob}(d_{it}=1\,|\,J_{i,t-1}, w_{it}, X_i, \hat\theta) \times f(w_{it}\,|\,J_{i,t-1}, X_i, \hat\theta)\,. \tag{20}$$

The expected value of this score contribution, conditional on $X_i$ and $J_{i,t-1}$, is

$$\begin{aligned}
&\mathrm{E}[\nabla_{\hat\theta}\mathscr{L}_{it}(\hat\theta)\mid J_{i,t-1},X_i] \\
&\quad= \mathrm{Prob}(d_{it}=0\mid J_{i,t-1},X_i,\hat\theta)\nabla_{\hat\theta}\ln \mathrm{Prob}(d_{it}=0\mid J_{i,t-1},X_i,\hat\theta) \\
&\qquad+ \mathrm{E}[d_{it}\nabla_{\hat\theta}\ln \mathrm{Prob}(d_{it}=1\mid J_{i,t-1},w_{it},X_i,\hat\theta) \\
&\qquad\times f(w_{it}\mid J_{i,t-1},X_i,\hat\theta)\mid J_{i,t-1},X_i]\,. 
\end{aligned}\tag{21}$$

Although the expected value of the simulated score contribution at the true parameter vector is not zero due to denominator bias in the simulation, the difference between the simulated score contribution and the expected value of the simulated score contribution conditional on $J_{i,t-1}$ and $X_i$ will have expectation zero at the true parameter vector. Thus, by subtracting (21) from (20), to obtain

$$\begin{aligned}
&\nabla_{\hat\theta}\mathscr{L}_{it} - \mathrm{E}[\nabla_{\hat\theta}\mathscr{L}_{it}(\hat\theta)\mid J_{i,t-1},X_i] \\
&\quad= [(1-d_{it}) - \mathrm{Prob}(d_{it}=0\mid J_{i,t-1},X_i,\hat\theta)] \\
&\qquad\times \nabla_{\hat\theta}\ln \mathrm{Prob}(d_{it}=0\mid J_{i,t-1},X_i,\hat\theta) \\
&\qquad+ \{d_{it}\nabla_{\hat\theta}\ln \mathrm{Prob}(d_{it}=1\mid J_{i,t-1},w_{it},X_i,\hat\theta)f(w_{it}\mid J_{i,t-1},X_i,\hat\theta) \\
&\qquad- \mathrm{E}[d_{it}\nabla_{\hat\theta}\ln \mathrm{Prob}(d_{it}=1\mid J_{i,t-1},w_{it},X_i,\hat\theta) \\
&\qquad\quad\times f(w_{it}\mid J_{i,t-1},X_i,\hat\theta)\mid J_{it-1},X_i]\}\,,
\end{aligned}\tag{22}$$

an expression is obtained that can be used to construct an MSM estimator. Both the term $[(1-d_{it}) - \mathrm{Prob}(d_{it}=0\mid J_{i,t-1},X_i,\hat\theta)$ and the term in the braces $\{\cdot\}$ are mean zero residuals. The orthogonal weight on the former term is $\nabla_{\hat\theta}\ln \mathrm{Prob}(d_{it}=0\mid J_{i,t-1},X_i,\hat\theta)$ while the weight on the latter term is simply one. Thus, substitution of unbiased simulators for all the probabilities in (22) gives an MSM estimator that is consistent and asymptotically normal for fixed simulation size. Of course, the transition probabilities in (22) are difficult objects to simulate. Using the GHK method described in Section 5.2 to simulate these probabilities would again produce an MSM estimator that is consistent and asymptotically normal if $S/\sqrt{N}\to\infty$ as $N\to\infty$.

Observe that in (22) the object $\mathrm{E}[d_{it}\nabla_{\hat\theta}\ln \mathrm{Prob}(d_{it}\mid J_{i,t-1},w_{it},X_i,\hat\theta)f(w_{it}\mid J_{i,t-1},X_i,\hat\theta)\mid J_{it-1},X_i]$ must be simulated. This situation is particularly difficult because, to take the outer expectation, $w_{it}$ must be drawn from the $f(w_{it}\mid J_{i,t-1},X_i,\hat\theta)$ density, and then the term $\nabla_{\hat\theta}\ln \mathrm{Prob}(d_{it}\mid J_{i,t-1},w_{it},X_i,\hat\theta)$ must be simulated conditional on each $w_{it}$ draw. If the first term in braces, the term $\nabla_{\hat\theta}\ln \mathrm{Prob}(d_{it}\mid J_{i,t-1},w_{it},X_i,\hat\theta)$ that involves the observed $w_{it}$, is simulated using $S$ draws, then, in order for the difference in braces to have mean zero, the derivatives of the log-probabilities in the second term must also be simulated using $S$ draws *per each* $w_{it}$ draw.

Keane and Moffitt (1991) implemented the MSM estimator based on (22) in a cross-section choice problem – the choice by low-income single mothers of welfare program participation and work status – where it is feasible to construct unbiased probability simulators. Despite the fact that the MSM estimator is consistent and asymptotically normal for fixed $S$ in this problem, Keane and Moffitt found that a very large number of draws is necessary for the estimator to produce reasonable results. This stems from the difficulty of simulating the expectation over wage draws described above. Thus, MSM estimation based on (22) may not be promising in the panel data case. Keane and Moffitt (1991) also reported results based on a direct simulation of the score as expressed in (19). This MSS estimator performed at least as well as the MSM estimator, and given that it is much easier to program, it may be the preferred course for panel data selection models. As discussed by McFadden and Ruud (1991), it is also rather difficult to put the tobit model in an MSM form. But Hajivassiliou and McFadden (1990) reported good results using MSS based on the GHK method to estimate panel data tobit models in which the dependent variable is the total external debt obligation of a country in arrears.

## 7. Estimating the serial correlation structure in employment and wage data

### 7.1. *Results using NLS employment data*

In this section, I use the MSM estimator obtained by substituting the transition probability simulator (12a)–(12b) into the MSM first-order condition (13)–(14) to estimate panel data probit models that relax the equicorrelation assumption, using employment data from the national longitudinal survey of young men (NLS). The goal is to determine whether the simple random effects model with equicorrelated errors can adequately capture the pattern of temporal dependence in these data. As I discussed in Section 3, the random effects model has been the most popular specification for panel data LDV models. Prior to the advent of simulation-based inference, it was not computationally feasible to relax the equicorrelation assumption. Thus, the results in this section provide the first test of the equicorrelation assumption for labor market data.

The NLS is a U.S. sample of 5225 males aged 14–24 selected in 1966 and interviewed in 12 of the 16 years from 1966 to 1981. Data were collected on employment status and other sociodemographic characteristics. The sample used here is exactly that employed by Keane, Moffitt and Runkle (1988). The data screens and overall properties of the data are discussed there. Following data screens, the analysis sample contains 2219 males with a total of 11 886 person-year observations. The regressors used in the employment equation are a constant (CONST), the national unemployment rate (U-RATE), a time trend (TREND), years of school completed (EDUC), years of labor force experience (EXPER), the square of experience (EXPER$^2$), a white dummy (WHITE), a dummy for wife present in the home (WIFE), and number of children (KIDS).

Estimation results are reported in Table 1. The first column gives constant cross-section estimates of the regressor coefficients obtained by ML. Columns (2)–(6) contain estimates when various patterns of serial correlation are assumed. These estimates are starred if they differ significantly from the

Table 1
Estimates of probit employment equations on NLS young men

| Parameter | $\hat{\beta}$ ML | $\hat{\beta}$ ML-quadrature | | $\hat{\beta}$ MSM | | |
|---|---|---|---|---|---|---|
| | | 4 points | 16 points | Random effects | RE + AR(1) error | RE + MA(1) error |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\rho$ | – | 0.3792 | 0.3509 | 0.3577 | 0.3298 | 0.3377 |
| | | (0.0203) | (0.0192) | (0.0248) | (0.0294) | (0.0263) |
| AR(1) | – | – | – | – | 0.1901 | – |
| | | | | | (0.0437) | |
| MA(1) | – | – | – | – | – | 0.1998 |
| | | | | | | (0.0685) |
| CONST | 0.4644 | 0.5454 | 0.5713 | 0.4895 | 0.3934 | 0.5411 |
| | (0.1161) | (0.1230) | (0.1283) | (0.1551) | (0.1576) | (0.1545) |
| U-RATE | −0.0740 | −0.0678 | −0.0697 | −0.0647 | −0.0590 | −0.0664 |
| | (0.0135) | (0.0122) | (0.0124) | (0.0157) | (0.0160) | (0.0162) |
| TREND | −0.0121 | −0.0091 | −0.0124 | −0.0185 | −0.0205 | −0.0164 |
| | (0.0059) | (0.0062) | (0.0063) | (0.0080) | (0.0082) | (0.0081) |
| EDUC | 0.0664 | 0.0599 | 0.0593 | 0.0680 | 0.0720 | 0.0634 |
| | (0.0061) | (0.0070) | (0.0078) | (0.0106) | (0.0107) | (0.0105) |
| EXPER | 0.0176 | 0.0076 | 0.0113 | 0.0094 | 0.0107 | 0.0060 |
| | (0.0098) | (0.0097) | (0.0099) | (0.0118) | (0.0123) | (0.0121) |
| $\text{EXPER}^2$ | −0.1521 | −0.1260 | −0.1280 | −0.1140 | −0.1120 | −0.1040 |
| ÷ 100 | (0.0414) | (0.0400) | (0.0400) | (0.0470) | (0.0490) | (0.0490) |
| WHITE | 0.2022 | 0.2355 | 0.2453 | 0.2205 | 0.2101 | 0.2230 |
| | (0.0503) | (0.0639) | (0.0646) | (0.0637) | (0.0637) | (0.0627) |
| WIFE | 0.4548 | 0.3271*** | 0.3394*** | 0.3218*** | 0.3316** | 0.3332*** |
| | (0.0352) | (0.0363) | (0.0365) | (0.0448) | (0.0457) | (0.0453) |
| KIDS | 0.0716 | 0.0726 | 0.0712 | 0.0741 | 0.0713 | 0.0708 |
| | (0.0138) | (0.0147) | (0.0150) | (0.0195) | (0.0196) | (0.0195) |
| $\gamma(1)$ | 0.0000 | 0.3792 | 0.3509 | 0.3577 | 0.4572 | 0.4520 |
| $\gamma(2)$ | 0.0000 | 0.3792 | 0.3509 | 0.3577 | 0.3540 | 0.3377 |
| $\gamma(3)$ | 0.0000 | 0.3792 | 0.3509 | 0.3577 | 0.3344 | 0.3377 |
| Log-likelihood function | −3611 | −3419 | −3421 | −3460 | −3447 | −3448 |
| $\chi^2(9)$ | – | 24.68** | 16.77* | 15.16* | 16.57* | 15.83* |
| CPU minutes | 1.76 | 2.87 | 5.19 | 11.55 | 12.67 | 12.59 |

Note: Standard errors of the parameter estimates are in parentheses. Three stars (***) indicate that a parameter differs from the ML no-effects estimate at the 1% significance level. Two stars (**) indicate the 5% level, and one star(*) indicates the 10% level. The $\chi^2(9)$ statistic is for the null hypothesis that the regressor coefficients equal the ML no-effects estimates (the 5% critical value is 16.92 and the 10% critical value is 14.68). The data set used is the NLS survey of young men. There are observations on 2219 individuals, with a total of 11 886 person–year observations. The MSM estimates were obtained using 10 draws for the GHK simulator. Log-likelihood function values for the MSM estimators are simulated.

consistent ML estimates in column (1). A $\chi^2(9)$ test for the null that all the regressor coefficients equal the column (1) values is also reported.

Random effects estimates using approximate-ML via 4- and 16-point quadrature are reported in columns (2) and (3). There is a non-negligible change in the parameter estimates in moving from 4- to 16-point quadrature, as indicated by the fact that, for 4-point quadrature, the null that the regressor coefficients equal the cross-section estimates is rejected at the 5 percent level, while with 16-point quadrature the null is only rejected at the 10 percent level. Thus, I will concentrate on the 16-point quadrature results. Random effects estimates obtained via MSM are reported in column (4). They were obtained using $S = 10$. The MSM estimates of both the parameters and their standard errors are quite close to the 16-point ML-quadrature estimates, and the null that the regressor coefficients equal the consistent cross-section estimates is again rejected at the 10 percent but not the 5 percent level.

If the random effects assumption is correct, then both the cross-section and random effects estimates are consistent, and we would expect no significant difference in the regressor coefficients obtained via random effects and no effects estimators. The effect of the random effects estimator should be simply to adjust standard errors to account for serial correlation. In going from column (1) to columns (2), (3), or (4) there is a general rise in the estimated standard errors. However, one of the estimated coefficients, that on the WIFE variable, changes substantially. The ML-quadrature and MSM estimates both show a drop of about three standard errors for this coefficient.

Since random effects estimates may be inconsistent in the equicorrelation assumption fails and because we are interested in discovering whether the actual pattern of temporal dependence in the data is more complex, I relax the equicorrelation assumption in columns (5) and (6). Here, estimates are obtained which allow for AR(1) and MA(1) error components in addition to the random effects. Since the individuals in the data are observed for up to 12 periods, these estimates require the evaluation of 12-variate integrals. Thus, the estimation is not feasible by ML and can only be performed using the MSM estimator.

Turning to the MSM results, first note that the time requirements for the MSM estimations are quite modest – the timings being about 12.6 cpu minutes on an IBM 3083 (compared to 5.2 for 16-point quadrature on the random effects model). Second, note that the equicorrelation assumption does fail. In column (5), the estimated AR(1) parameter is 0.1901 with a $t$-statistic of 4.4. In column (6), the estimated MA(1) parameter is 0.1998 with a $t$-statistic of 2.9. The $\gamma(j)$ reported in the table are the $j$-th lagged autocorrelations implied by the estimated covariance parameters. The first lagged autocorrelation is about 30 percent larger for the model with AR(1) components than it is for the models with random effects alone (0.46 vs. 0.35). Thus, the random effects model would overestimate the probability of a transition from employment to unemployment because it underestimates short-run persistence.

Although these results show a significant departure from equicorrelation, relaxing the equicorrelation assumption has little effect on the parameter estimates. Furthermore, the 10- to 11-point improvements in the simulated log-likelihood with inclusion of MA(1) or AR(1) components is not particularly great. Thus, it appears that false imposition of equicorrelation does not lead to substantial parameter bias or deterioration of fit in models of male employment patterns.

### 7.2. *Temporal dependence in wages and the movement of real wages over the business cycle*

In this section, I consider an application of the MSS estimator to nonrandom-sample selection models of the type described by Heckman (1979). In these models, a probit is estimated jointly with a continuous dependent-variable equation, where the dependent variable is only observed for the chosen state. Because of the truncation of the error term in the equation for the continuous dependent variable, OLS estimates of that equation are biased, and the residuals from the OLS regression produce biased estimates of the error structure for the continuous variable. Thus, joint estimation is necessary to obtain consistent estimates. As I described in Section 6, it is difficult to estimate such models by MSM. Instead, I implement an MSS estimator by simulating the score for the selection model as written in (19).

The particular application considered here is the estimation of selection bias-adjusted wage equations. Keane, Moffitt and Runkle (1988) used selection models with random effects in order to estimate the cyclical behavior of real wages in the NLS. Their estimates controlled for the cross-correlation of permanent and transitory error components in wage and employment equations. By controlling for these cross-correlations, they hoped to control for systematic movements of workers with high or low unobserved wage components in and out of the labor force over the business cycle. By so doing, they could obtain estimates of cyclical real wage movement holding labor force quality constant. Keane, Moffitt and Runkle found that real wage movements were procyclically biased by quality variation, with high-wage workers the most likely to become unemployed in a recession. It is possible that the Keane, Moffitt and Runkle results may be biased due to false imposition of the equicorrelation assumption. Thus, it is important to examine robustness of their results to the specification of the error structure.

The NLS data used in this analysis were already described in Section 7.1 and used in the employment equation estimates presented there. The only new variable is the wage, which is the hourly straight time real wage (deflated by the consumer price index) at the interview date. The log wage is the dependent variable.

Estimation results are reported in Table 2. The first column gives consistent cross-section estimates obtained by ML. Columns (2)–(6) contain estimates obtained assuming various patterns of serial correlation. These estimates are starred if they differ significantly from the consistent estimates in column (1).

Table 2
Estimates of selection model on NLS young men

| Parameter | $\hat{\beta}$ ML | $\hat{\beta}$ ML-quadrature | | $\hat{\beta}$ MSM | | |
|---|---|---|---|---|---|---|
| | | 4 points | 9 points | Random effects | RE + AR(1) error | RE + MA(1) error |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| *Wage equation* | | | | | | |
| U-RATE | −0.0039 | −0.0063 | −0.0055 | −0.0057 | −0.0095** | −0.0066 |
| | 0.0034 | (0.0023) | (0.0023) | (0.0027) | (0.0026) | (0.0028) |
| TIME | 0.0073 | 0.0125*** | 0.0105** | 0.0119** | 0.0124** | 0.0119* |
| | (0.0015) | (0.0013) | (0.0015) | (0.0021) | (0.0021) | (0.0021) |
| EDUC | 0.0606 | 0.0520*** | 0.0487*** | 0.0500*** | 0.0529** | 0.0516*** |
| | (0.0016) | (0.0018) | (0.0023) | (0.0032) | (0.0033) | (0.0032) |
| EXPER | 0.0263 | 0.0242 | 0.0260 | 0.0242 | 0.0276 | 0.0256 |
| | (0.0024) | (0.0015) | (0.0016) | (0.0027) | (0.0030) | (0.0028) |
| $\text{EXPER}^2$ | −0.0736 | −0.0780 | −0.0750 | −0.0720 | −0.0870 | −0.0780 |
| ÷ 100 | (0.0108) | (0.0060) | (0.0050) | (0.0100) | (0.0130) | (0.0110) |
| WHITE | 0.1923 | 0.1767 | 0.1829 | 0.1936 | 0.1934 | 0.1932 |
| | (0.0134) | (0.0150) | (0.0236) | (0.0235) | (0.0234) | (0.0231) |
| CONSTANT | 0.0494 | 0.0967* | 0.1510*** | 0.1234* | 0.0923 | 0.1054 |
| | (0.0297) | (0.0284) | (0.0371) | (0.0450) | (0.0459) | (0.0450) |
| *Employment equation*: | | | | | | |
| U-RATE | −0.0646 | −0.0699 | −0.0693 | −0.0648 | −0.0589 | −0.0570 |
| | (0.0133) | (0.0131) | (0.0126) | (0.0157) | (0.0160) | (0.0160) |
| TIME | −0.0126 | −0.0118 | −0.0169 | −0.0200 | −0.0208 | −0.0226 |
| | (0.0058) | (0.0058) | (0.0062) | (0.0081) | (0.0081) | (0.0082) |
| EDUC | 0.0610 | 0.0578 | 0.0655 | 0.0664 | 0.0645 | 0.0707 |
| | (0.0060) | (0.0058) | (0.0076) | (0.0106) | (0.0104) | (0.0104) |
| EXPER | 0.0014 | 0.0067 | 0.0084 | 0.0017 | −0.0020 | 0.0022 |
| | (0.0095) | (0.0094) | (0.0099) | (0.0120) | (0.0122) | (0.0122) |
| $\text{EXPER}^2$ | −0.1034 | −0.1020 | −0.1130 | −0.0920 | −0.0740 | −0.0860 |
| ÷ 100 | (0.0401) | (0.0390) | (0.0400) | (0.0480) | (0.0500) | (0.0490) |
| WHITE | 0.1961 | 0.2148 | 0.2493 | 0.2345 | 0.2229 | 0.2223 |
| | (0.0492) | (0.0500) | (0.0637) | (0.0635) | (0.0632) | (0.0635) |
| WIFE | 0.4597 | 0.3393** | 0.3770** | 0.3550** | 0.3664** | 0.3498** |
| | (0.0323) | (0.0341) | (0.0359) | (0.0446) | (0.0449) | (0.0448) |
| KIDS | 0.1151 | 0.0621*** | 0.0895* | 0.0930 | 0.0869 | 0.0942 |
| | (0.0127) | (0.0131) | (0.0142) | (0.0198) | (0.0199) | (0.0199) |
| CONSTANT | 0.4922 | 0.6446 | 0.4936 | 0.5147 | 0.5083 | 0.4146 |
| | (0.1142) | (0.1117) | (0.1274) | (0.1554) | (0.1551) | (0.1536) |
| *Covariance parameters*: | | | | | | |
| $\rho_{\text{wage}}$ | – | 0.6073 | 0.5995 | 0.5449 | 0.4547 | 0.4807 |
| | | (0.0069) | (0.0075) | (0.0131) | (0.0162) | (0.0132) |
| $\rho_{\text{employment}}$ | – | 0.2364 | 0.3275 | 0.3548 | 0.3090 | 0.3285 |
| | | (0.0136) | (0.0182) | (0.0243) | (0.0266) | (0.0257) |
| $\text{AR}(1)_{\text{wage}}$ | – | – | – | – | 0.4803 | – |
| | | | | | (0.0165) | |
| $\text{AR}(1)_{\text{employment}}$ | – | – | – | – | 0.2538 | – |
| | | | | | (0.0442) | |
| $\text{MA}(1)_{\text{wage}}$ | – | – | – | – | – | 0.2426 |
| | | | | | | (0.0851) |

Table 2 (*Continued*)

| Parameter | $\hat{\beta}$ ML | $\hat{\beta}$ ML-quadrature | | $\hat{\beta}$ MSM | | |
|---|---|---|---|---|---|---|
| | | 4 points | 9 points | Random effects | RE + AR(1) error | RE + MA(1) error |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $\mathrm{MA}(1)_{\mathrm{employment}}$ | – | – | – | – | – | 0.2003 (0.0564) |
| Correlation of permanent parts | – | 0.4330 (0.0132) | −0.2798 (0.0213) | −0.2002 (0.0434) | −0.1859 (0.0508) | −0.2075 (0.0466) |
| Correlation of transitory parts | −0.6947 (0.0228) | −0.2651 (0.0775) | −0.3339 (0.0654) | −0.3491 (0.0636) | −0.3861 (0.0566) | −0.3944 (0.0609) |
| $\sigma_{\mathrm{wage}}$ | 0.4301 (0.0031) | 0.4162 (0.0027) | 0.4109 (0.0025) | 0.4055 (0.0030) | 0.4049 (0.0030) | 0.4052 (0.0031) |
| Log-likelihood function | −8984 | −6312 | −6216 | −6321 | −6004 | −6244 |
| $\chi^2(16)$ | – | 362** | 164** | 75** | 78** | 72** |
| CPU minutes | 2.09 | 12.94 | 40.09 | 42.68 | 47.38 | 47.38 |

Note: Standard errors of the parameter estimates are in parentheses. Three stars (***) indicate that a parameter differs from the ML no-effects estimate at the 1% significance level. Two stars (**) indicate the 5% level, and one star (*) indicates the 10% level. The $\chi^2(16)$ statistic is for the null hypothesis that the regressor coefficients equal the ML no-effects estimates. The 5% critical value is 26.30. The data set used is the NLS survey of young men. There are observations on 2219 individuals, with a total of 11 886 person–year observations. The MSS estimates were obtained using 10 draws for the GHK simulator. Log-likelihood function values for the MSS estimators are simulated.

A $\chi^2$ test for the null that all the regressor coefficients equal the column (1) values is also reported.

Random effects estimates via approximate ML with 4 and 9 quadrature points are reported in columns (2) and (3). Clearly, there is very strong persistence in the wage equation errors, as 60 percent of the wage error variance is accounted for by random effects. Observe that the ML-quadrature estimates are quite far from the cross-section estimates. Particularly noticeable is the coefficient on EDUC in the wage equation, which is from 4.8 to 5.2 standard errors below the cross-section estimate. The $\chi^2$ tests overwhelmingly reject the null that the random effects estimates equal the consistent cross-section estimates.

Notice that 4- and 9-point quadratures produce very different estimates of the cross-correlation of random effects. With 4 points, this is estimated as 0.4330, and with 9 points, it is estimated as −0.2798, both estimates being highly significant. The 4-point results are what Keane, Moffitt and Runkle reported. Since use of roughly 4 quadrature points is typical in the literature, these results demonstrate the need to use larger numbers of quadrature points in applied work. Increasing the number of points to 12 did not produce much change in results (the likelihood changed only from −6216 to −6206). Use of 12 points is very expensive for this model, as it required 88 cpu minutes.

These random effects results overturn the Keane, Moffitt and Runkle finding that permanent wage and employment error components are positively correlated. However, it should be noted that Keane, Moffitt and Runkle considered their preferred specification to be a semiparametric random effects model estimated using the technique of Heckman and Singer (1982), and this technique did choose the likelihood peak which has a negative cross-correlation of the random effects. Such a negative correlation, indicating that those with permanently high wage errors supply less labor, is not surprising since it can be explained by income effects. More surprising is the negative correlation between the transitory error components, implying that those with the temporarily high wages supply less labor. As was noted by Keane, Moffitt and Runkle, this appears difficult to reconcile with intertemporal substitution theories of the business cycle.

The MSS estimates of the random effects model are reported in column (4). These were obtained using the GHK simulator with 10 draws to simulate the transition probabilities. The regressor coefficient estimates are all quite close to the 9-point ML-quadrature estimates. Larger standard errors for the MSS estimates account for the smaller (but still highly significant) $\chi^2$ test for the null of equality with the no-effects estimates (75 vs. 164). Column (5) contains MSS estimates of a model that allows for random effects plus AR(1) error components. When the AR(1) components are included, the AR(1) parameter in the wage equation is a substantial 0.4803 (with standard error 0.0165) and the fraction of the wage error variance explained by the individual effects drops to 45 percent. In the employment equation, the AR(1) parameter is also highly significant (0.2538 with standard error 0.0442). Clearly, the equicorrelation assumption is overwhelmingly rejected by the data. The first four lagged autocorrelations of the wage equation error implied by the MSS estimates in column (5) are 0.72, 0.58, 0.52, 0.48 – as compared to the autocorrelation of 0.60 at all lags implied by the random effects model. The first four lagged autocorrelations of the employment equation error are 0.48, 0.35, 0.32, and 0.31 as compared to the 0.3275 at all lags implied by the random effects model. Note, also, that the computational cost of the MSM estimator that allows for this more complex error pattern (47.38 cpu minutes on an IBM 3083) is only slightly greater than the cost of ML-quadrature estimation of the random effects model (40.49 cpu minutes).

In the model with a moving-average error component (column (6)), the MA(1) parameter in the wage equation is 0.2426 (with standard error 0.0851) and that in the employment equation is 0.2003 (with standard error 0.0564). Based on the simulated log-likelihood values, this model does not seem to fit as well as the model with AR(1) error components.

Although the equicorrelation assumption is rejected by the data, the parameter estimates obtained via MSS change only slightly when AR(1) and MA(1) error components are included in the model. Thus, the divergence of random effects estimates from the consistent no-effects estimates does not appear to result from the false imposition of the equicorrelation assumption in this case. In particular, the most likely explanation for the substantial drop in the education coefficient in going from the model with no effects to the models with random effects is that the individual effect in the wage equation is correlated with the education variable. That is, the individual effect is actually a fixed effect.

I now turn to the issue of the cyclicality of the real wage. All three MSS models give estimates of the cross-correlations of the random effects in the range from −0.19 to −0.21, and estimates of the cross-correlations of the time varying error components in the range from −0.35 to −0.39. Since negative correlations imply that high-wage workers are most likely to leave work in a recession, these results imply a degree of procyclical bias in aggregate wage measures which is considerably stronger than that found by Keane, Moffitt and Runkle, who report a positive correlation of the permanent components and a −0.33 correlation for the transitory components (column (4)). Since Keane, Moffitt and Runkle's main conclusion was that aggregate wage measures are procyclically biased, this can be viewed as a strengthening of that result.

The estimated unemployment rate coefficients are −0.0039 for the no-effects model, −0.0055 for the random effects model estimated by 9-point quadrature, −0.0057 for the random effects model estimated by MSS, −0.0095 for the random effects plus AR(1) error model, and −0.0066 for the random effects plus MA(1) error model. These estimates imply that a one-percentage-point increase in the unemployment rate corresponds to a fall in the real wage of between 0.4 percent and 1 percent. Thus, Keane, Moffitt and Runkle's finding that movements in the real wage are weakly procyclical appears to be robust to relaxation for the equicorrelation assumption.

## 8. Conclusion

The application of simulation estimation techniques to panel data LDV models is clearly more difficult than the application of these methods to cross-section problems. Yet the recent development of highly accurate GHK simulators for transition and choice probabilities has made simulation estimation in the panel data LDV context feasible. Three classical methods, an MSM estimator based on using the GHK method to simulate transition probabilities, an MSS estimator based on using the GHK method to simulate the score and an SML estimator based on using GHK to simulate choice probabilities, have been successfully applied in the literature. As the empirical examples in Section 7 show, these methods allow one to estimate panel data LDV models with complex error structures involving random effects and ARMA errors in times similar to those necessary for estimation of simple random effects models by quadrature. A Bayesian method based on Gibbs sampling has also been successfully applied. An important avenue for future research is to further explore the performance of methods based on conditional simulation of the latent variables of the LDV model, such as the simulated EM and Gibbs sampling approaches, in the panel data setting, and to compare the performance of these methods to that of MSM, MSS and SML.

## References

Albert, J. and S. Chib (1993). Bayesian analysis of binary and polychotomous choice data. *J. Amer. Statist. Assoc.*, to appear.

Amemiya, T. and T. E. McCurdy (1986). Instrumental-variables estimation of an error-components model. *Econometrica* **54**, 869–80.

Balestra, P. and M. Nerlove (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica* **34**, 585–612.

Berkovec, J. and S. Stern (1991). Job exit behavior of older men. *Econometrica* **59**(1), 189–210.

Börsch-Supan, A., V. Hajivassiliou, L. Kotlikoff and J. Morris (1991). Health, children, and elderly living arrangements: A multiperiod-multinomial probit model with unobserved heterogeneity and autocorrelated errors. In: D. Wise, ed., *Topics in the Economics of Aging*. Univ. of Chicago Press, Chicago, IL.

Butler, J. S. and R. Moffitt (1982). A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica* **50**(3), 761–64.

Chamberlain, G. (1985). Panel data. In: Z. Griliches and M. D. Intriligator, eds., *Handbook of Econometrics*. North-Holland, Amsterdam.

Chib, S. (1993). Bayes inference in the tobit censored regression model. *J. Econometrics*, to appear.

Elrod, T. and M. Keane (1992). A factor-analytic probit model for estimating market structure in panel data. Manuscript, University of Alberta.

Erdem, T. and M. Keane (1992). A dynamic structural model of consumer choice under uncertainty. Manuscript, University of Alberta.

Gelfand, A. E. and A. F. M. Smith (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409.

Geweke, J. (1991a). Efficient simulation from the multivariate normal and student-*t* distributions subject to linear constraints. In: E. M. Keramidas, ed., *Computing Science and Statistics: Proc. 23rd Sympos. on the Interface*. Interface Foundation Inc., Fairfax, VA, 571–578.

Geweke, J. (1991b). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Research Department Staff Report 148, Federal Reserve Bank of Minneapolis.

Geweke, J., M. Keane and D. Runkle (1992). Alternative computational approaches to statistical inference in the multinomial probit model. Manuscript, University of Minnesota.

Geweke, J., R. Slonim and G. Zarkin (1992). Econometric solution methods for dynamic discrete choice problems. Manuscript, University of Minnesota.

Gourieroux, C. and A. Monfort (1993). Simulation based econometric methods. *J. Econometrics*, to appear.

Hajivassiliou, V. and D. McFadden (1990). The method for simulated scores for the estimation of LDV models with an application to external debt crises. Cowles Foundation Discussion Paper 967.

Hajivassiliou, V., D. McFadden and P. Ruud (1992). Simulation of multivariate normal orthant probabilities: Theoretical and computational results. Cowles Foundation Discussion Paper 1021.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153–61.

Heckman, J. (1981). Statistical models for discrete panel data. In: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. MIT Press, Cambridge, MA.

Heckman, J. and B. Singer (1982). The identification problem in econometric models for duration data. In: W. Hildenbrand, ed., *Advances in Econometrics*. Cambridge Univ. Press, Cambridge.

Hotz, J. V. and R. Miller (1993). Conditional choice probabilities and the estimation of dynamic programming models. *Rev. Econom. Stud.*, to appear.

Keane, M. (1990). Four essays in empirical macro and labor economics. Unpublished Ph.D. Dissertation, Brown University.

Keane, M. (1992). A computationally practical simulation estimator for panel data. Institute for Empirical Macroeconomics Discussion Paper 16, Federal Reserve Bank of Minneapolis.

Keane, M. and R. Moffitt (1991). A structural model of multiple welfare program participation and labor supply. Working Paper, Brown University.

Keane, M., R. Moffitt and D. Runkle (1988). Real wages over the business cycle: Estimating the impact of heterogeneity with micro data. *J. Politic. Econom.* **96**(6), 1232–66.

Keane, M. and D. Runkle (1992). On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous. *J. Business Econom. Statist.* **10**(1), 1–9.

Keane, M. and K. Wolpin (1992). Solution and estimation of discrete dynamic programming models by simulation: Monte Carlo evidence. Manuscript, University of Minnesota.

McCulloch, R. and P. E. Rossi (1992). An exact likelihood analysis of the multinomial probit model. Manuscript, Graduate School of Business, University of Chicago.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* **57**(5), 995–1026.

McFadden, D. (1992). Estimation by the method of simulated moments. Paper presented at Conference on Simulation Based Inference in Econometrics, Rotterdam, June 5–6.

McFadden, D. and P. Ruud (1987). Estimation of limited dependent variable models from the regular exponential family by the method of simulated moments. Working Paper, University of California, Berkeley.

McFadden, D. and P. Ruud (1991). Estimation by simulation. Working Paper, University of California, Berkeley.

Pakes, A. (1986). Patents as options: Some estimates of the value of holding European patent stocks. *Econometrica* **54**, 755–84.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica* **57**(5), 1027–58.

Robinson, P. (1982). On the asymptotic properties of models containing limited dependent variables. *Econometrica* **50**(1), 27–41.

Ruud, P. (1991). Extensions of estimation methods using the EM algorithm. *J. Econometrics* **49**(3), 305–41.

Van Praag, B. M. S. and J. P. Hop (1987). Estimation of continuous models on the basis of set-valued observations. Report 8705/A, Econometric Institute, Erasmus University.

Van Praag, B. M. S., J. P. Hop and E. Eggink (1991). A symmetric approach to the labor market by means of the simulated EM-algorithm with an application to married females. Report 9145/A, Econometric Institute, Erasmus University.