Maximum Simulated Likelihood Estimation of Random Effects Dynamic Probit Models with Autocorrelated Errors

Mark B. Stewart*
University of Warwick

April 2006

Abstract

This paper investigates the use of Maximum Simulated Likelihood estimation for random effects dynamic probit models with autocorrelated errors. It presents a new Stata command, **redpace**, for this estimator and illustrates its usage. The paper also compares the use of pseudo-random numbers and Halton sequences of quasi-random numbers for the MSL estimation of these models.

1 Introduction

This paper examines the estimation of dynamic probit models, specifically models in which the outcome probability is dependent on the outcome in the previous time period. The presence of omitted individual heterogeneity, in the form of individual-specific effects, results in an "initial conditions" problem and renders the standard Random Effects Probit estimator inconsistent.

If the latent equation time-varying error terms are serially uncorrelated, the model can under certain conditions, be estimated consistently by a Maximum Likelihood estimator proposed by Heckman. Evaluation of the likelihood in this case can be based on the same Gauss-Hermite quadrature approximation for the resulting integral as is used in standard Random Effects Probit estimators such as that in Stata's **xtprobit** command. A new Stata routine for the Heckman estimator of this model is presented en route below.

If the error terms are autocorrelated, the Heckman estimator too is inconsistent. Extending it to the autocorrelated case results in the need to evaluate higher dimensional integrals. Maximum Simulated Likelihood is a natural estimator to use in this case. This paper describes, and implements in Stata, a Maximum Simulated Likelihood estimator for the Random Effects Dynamic Probit model with autocorrelated errors. An empirical illustration is provided and the issue of the required number of simulations investigated.

*Address for correspondence: Mark Stewart, Economics Department, University of Warwick, Coventry CV4 7AL, UK. Tel: (44/0)-24-7653-3043. Fax: (44/0)-24-7652-3032.
E-mail: Mark.Stewart@warwick.ac.uk

2 A random effects dynamic probit model

The equation for the latent dependent variable is specified as

$$y_{it}^{*} = \gamma y_{it-1} + x_{it}'\beta + \alpha_i + u_{it} \tag{1}$$

$(i = 1, \ldots, N; t = 2, \ldots, T)$, where y_{it}^{*} is the latent dependent variable and y_{it} is the observed binary outcome variable, defined as:

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^{*} \geq 0 \\ 0 & \text{else} \end{cases} \tag{2}$$

and where x_{it} is a vector of explanatory variables and $u_{it} \sim N(0, \sigma_u^2)$. The subscript i indexes individuals and the subscript t indexes time periods. N is taken to be large, but T is typically small and regarded as fixed, so that asymptotics are on N alone. Even when the errors u_{it} are assumed serially independent, the composite error term, $v_{it} = \alpha_i + u_{it}$, will be correlated over time due to the individual–specific time–invariant α_i terms. The individual-specific random effects specification adopted implies equi-correlation between the v_{it} in any two (different) periods:

$$\lambda = Corr(v_{it}, v_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2} \qquad t, s = 2, \ldots, T; t \neq s \tag{3}$$

The standard (uncorrelated) random effects model also assumes α_i uncorrelated with x_{it}. Alternatively, following Mundlak (1978) and Chamberlain (1984), correlation between α_i and the observed characteristics in the model can be allowed for by assuming a relationship between α and either the time means of the x-variables or a combination of their lags and leads, e.g.: $\alpha_i = \overline{x}_i' a + \zeta_i$, where $\zeta_i \sim iid$ Normal and independent of x_{it} and u_{it} for all i, t. This simply has the effect of adding time means or lags and leads to the set of explanatory variables. To simplify notation the original form (1) will be used here with the understanding that these additional terms are subsumed into the x-vector in the case of the correlated random effects model.

Since y is a binary variable, a normalization is required. A convenient one is that $\sigma_u^2 = 1$. If u_{it} is normally distributed, the transition probability for individual i at time t, given α_i, is then given by

$$P\left[y_{it} | x_{it}, y_{it-1}, \alpha_i\right] = \Phi\left[(\gamma y_{it-1} + x_{it}'\beta + \alpha_i)(2y_{it} - 1)\right]. \tag{4}$$

Estimation of the model requires an assumption about the initial observations, y_{i1}, and in particular about their relationship with the α_i. The assumption giving rise to the simplest form of model for estimation is to take the initial conditions, y_{i1}, to be exogenous. However even if the start of the y-process coincides with the start of the observation period for each individual and the entire history of the y-process is observed, which is not generally the case, the assumption of independence between y_{i1} and α_i is a very strong one. For example, in the context of the union membership illustration used below, even if each individual were observed from labour market entry onwards, the assumption would require an individual's union status in first job to be unrelated to their individual-specific factor α_i. Under this assumption a standard Random Effects Probit program (such as **xtprobit**) can be used, since the likelihood can be decomposed into two independent factors and the joint probability for $t = 2, \ldots, T$ maximized without reference to that for $t = 1$. If, however, the initial conditions are correlated with the α_i, as would be expected in most situations, this method of estimation will tend to overstate the degree of state dependence, γ.

3 Heckman's estimator for the case of serially independent errors

The approach to the initial conditions problem proposed by Heckman (1981) involves specifying a linearized reduced form equation for the initial value of the latent variable:

$$y_{i1}^* = z_{i1}'\pi + \eta_i \tag{5}$$

$(i = 1, \ldots, N)$, where z_{i1} is a vector of exogenous instruments (and includes x_{i1}) and η_i is correlated with α_i, but uncorrelated with u_{it} for $t \geq 2$. Using an orthogonal projection, it can be written as:

$$\eta_i = \theta\alpha_i + u_{i1} \tag{6}$$

$(\theta > 0)$, with α_i and u_{i1} independent of one another. It is also assumed that u_{i1} satisfies the same distributional assumptions as u_{it} for $t = 2, \ldots, T$. (Any change in error variance will also be captured in θ.) The linearized reduced form for the latent variable for the initial time period is therefore specified as

$$y_{i1}^* = z_{i1}'\pi + \theta\alpha_i + u_{i1} \tag{7}$$

$(i = 1, \ldots, N)$, where z will include period 1 values of x-variables, typically together with pre-sample variables as instruments.[1]

The joint probability of the observed binary sequence for individual i, given α_i, in the Heckman approach, assuming serially independent u_{it}, is thus:

$$\Phi\left[(z_{i1}'\pi + \theta\alpha_i)(2y_{i1} - 1)\right] \prod_{t=2}^{T} \Phi\left[(\gamma y_{it-1} + x_{it}'\beta + \alpha_i)(2y_{it} - 1)\right]. \tag{8}$$

For a random sample of individuals the likelihood to be maximized is then given by

$$\prod_i \int_{\alpha^*} \left\{ \Phi\left[(z_{i1}'\pi + \theta\sigma_\alpha\alpha^*)(2y_{i1} - 1)\right] \prod_{t=2}^{T} \Phi\left[(\gamma y_{it-1} + x_{it}'\beta + \sigma_\alpha\alpha^*)(2y_{it} - 1)\right] \right\} dF(\alpha^*) \tag{9}$$

where F is the distribution function of $\alpha^* = \alpha/\sigma_\alpha$. Under the normalization used, $\sigma_\alpha = \sqrt{\lambda/(1-\lambda)}$. If α is taken to be normally distributed, the integral over α^* can be evaluated using Gaussian–Hermite quadrature (Butler and Moffitt, 1982). See Stewart (2005) for an application of the estimator in the context of an investigation of the dynamics of the conditional probability of unemployment. A Stata program for this estimator of the random effects dynamic probit model, **redprob**, can be downloaded from the author's website.

4 Models with autocorrelated errors

If the error term u_{it} is autocorrelated, reflecting for example correlation between transitory shocks, this complicates estimation considerably. Extension of the Heckman estimator to this case requires the evaluation of T-dimensional integrals of Normal

[1] An alternative Conditional Maximum Likelihood estimator for the serially independent errors case has been proposed by Wooldridge (2005).

densities. Simulation estimators provide a feasible way to address this problem. A Maximum Simulated Likelihood (MSL) estimator (see for example Gourieroux and Monfort, 1996, and Cameron and Trivedi, 2005), based on the GHK algorithm of Geweke, Hajivassiliou and Keane (see for example Keane, 1994) is presented here. The MSL estimation routine provides a consistent estimator of the vector of parameters as the number of simulation draws tends to infinity (and is asymptotically equivalent to the ML estimator). Strictly for the simulation error to disappear asymptotically, the number of simulation draws needs to increase at a rate greater than the square root of the sample size.

The model used in the program described here is as described above but with u_{it} following either an AR(1) process or an MA(1) process. Ω, the variance-covariance matrix of $v_i = (v_{i1}, \ldots, v_{iT})'$ will now be a function of λ, θ and one additional parameter. The error vector can be written as $v_i = C\eta_i$ with $\eta_i \sim N(0, I)$ and C the lower-triangular Cholesky decomposition of Ω (i.e. such that $CC' = \Omega$). The GHK algorithm then uses the fact that the probability of an observed sequence of ys can be written as the product of recursively defined conditional probabilities.

Using the Cholesky decomposition, the latent equations can be written as

$$y_{it}^* = \mu_{it} + \sum_{j=1}^{t} c_{tj}\eta_{ij} \tag{10}$$

where $\mu_{it} = \gamma y_{it-1} + x_{it}'\beta$ for $t \geq 2$ and $\mu_{i1} = z_{i1}'\pi$. The probability of an observed sequence of ys is given by

$$\begin{aligned} P_i = \Phi((y_{i1}-1)a_{i1}) \quad &\times \quad \int_{L_{i1}}^{U_{i1}} \Phi((y_{i2}-1)a_{i2})\phi(\eta_{i1})d\eta_{i1} \\ &\times \quad \int_{L_{i1}}^{U_{i1}} \int_{L_{i2}}^{U_{i2}} \Phi((y_{i3}-1)a_{i3})\phi(\eta_{i1})\phi(\eta_{i2})d\eta_{i2}d\eta_{i1} \\ &\times \quad \ldots \end{aligned} \tag{11}$$

where $a_{i1} = \mu_{i1}/c_{11}$, $a_{i2} = (\mu_{i2} + c_{21}\eta_{i1})/c_{22}$, $a_{i3} = (\mu_{i3} + c_{31}\eta_{i1} + c_{32}\eta_{i2})/c_{33}$, etc. and $(L_{it}, U_{it}) = (-a_{it}, \infty)$ if $y_{it} = 1$ or $(-\infty, -a_{it})$ if $y_{it} = 0$. For a sequence of length T this probability will be the product of T such terms. As can be seen from this expression, simulation of the probabilities requires draws from a truncated Normal. If ξ_{it} is a draw from a standard uniform distribution, then the required draws from a truncated Normal are constructed as $\Phi^{-1}[(1-\xi_{it})\Phi(L_{it}) + \xi_{it}\Phi(U_{it})]$.

The steps in the GHK simulator for this model are therefore as follows (for the rth draw):

1. Calculate $a_{i1} = \mu_{i1}/c_{11}$.
2. Draw ξ_{i1}^r from a standard uniform and calculate $\eta_{i1}^r = \Phi^{-1}[(1-\xi_{i1}^r)\Phi(L_{i1}) + \xi_{i1}^r\Phi(U_{i1})]$ and $a_{i2}^r = (\mu_{i2} + c_{21}\eta_{i1}^r)/c_{22}$.
3. Draw ξ_{i2}^r from a standard uniform and calculate $\eta_{i2}^r = \Phi^{-1}[(1-\xi_{i2}^r)\Phi(L_{i2}^r) + \xi_{i2}^r\Phi(U_{i2}^r)]$ and $a_{i3}^r = (\mu_{i3} + c_{31}\eta_{i1}^r + c_{32}\eta_{i2}^r)/c_{33}$.
4. Repeat this step successively for the remaining time periods.

The simulated likelihood is then given by

$$L^* = \prod_{i=1}^{N} \left\{ \frac{1}{R} \sum_{r=1}^{R} \left[\Phi((2y_{i1}-1)a_{i1}) \prod_{t=2}^{T} \Phi((2y_{it}-1)a_{it}^r) \right] \right\} \tag{12}$$

Two models with first-order autocorrelation are considered here. In the first of these, as considered by Hyslop (1999) and Stewart (2005), u_{it} is specified as a first-order autoregressive process:

$$u_{it} = \rho u_{it-1} + \varepsilon_{it} \tag{13}$$

with $-1 < \rho < 1$. In this model the variance-covariance matrix of v_i is given by:

$$\Omega = \begin{bmatrix} \theta^2\sigma_\alpha^2 + 1 & & & & \\ \theta\sigma_\alpha^2 + \rho & \sigma_\alpha^2 + 1 & & & \\ \theta\sigma_\alpha^2 + \rho^2 & \sigma_\alpha^2 + \rho & \sigma_\alpha^2 + 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \theta\sigma_\alpha^2 + \rho^{T-1} & \sigma_\alpha^2 + \rho^{T-2} & \sigma_\alpha^2 + \rho^{T-3} & \cdots & \sigma_\alpha^2 + 1 \end{bmatrix} \tag{14}$$

where $\sigma_\alpha^2 = \lambda/(1-\lambda)$. The model is estimated by MSL as outlined above.

In the second model u_{it} is specified as a first-order moving average process:

$$u_{it} = \varepsilon_{it} - \mu\varepsilon_{it-1} \tag{15}$$

In this case the variance-covariance matrix of v_i is given by:

$$\Omega = \begin{bmatrix} \theta^2\sigma_\alpha^2 + 1 & & & & \\ \theta\sigma_\alpha^2 - \omega & \sigma_\alpha^2 + 1 & & & \\ \theta\sigma_\alpha^2 & \sigma_\alpha^2 - \omega & \sigma_\alpha^2 + 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \theta\sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + 1 \end{bmatrix} \tag{16}$$

where $\sigma_\alpha^2 = \lambda/(1-\lambda)$ and $\omega = \mu/(1+\mu^2)$. Again the model is estimated by MSL.

The basic estimator uses R pseudo-random number draws (from a standard uniform), which are assumed to be independent. Efficiency of the estimator can be improved by using variance reduction methods which use dependent draws. The best known is antithetic sampling. In this case, for each draw ξ from the standard uniform distribution, we also use $(1-\xi)$. Hence the R random draws consist of $R/2$ antithetic pairs, $\{\xi^r, 1-\xi^r\}$. Antithetic sampling can reduce the variance of the MSL estimator, but this is not guaranteed. This idea can be extended by using segments of the unit interval in conjunction with the antithetics. The sub-division provided by this "symmetric systematic sampling" gives more even coverage, but there is a trade-off between coverage and randomness. (See Train, 2003, pp. 221-4.)

Another approach to improving the efficiency of the estimator that can be used is to replace the pseudo-random numbers by quasi-random numbers. These systematic sequences are constructed to provide better coverage of the domain of the distribution. One of the simplest is Halton sequences. A detailed treatment of these is given in Train (2003). See Cappellari and Jenkins (2006) for a Stata program. In certain situations use of these produces variance reduction and hence improves efficiency. To put this another way, a particular level of efficiency can be achieved with a lower number of draws and hence reduce computer time. The evidence of Train (2003) and others

for mixed logit models suggests that the number of draws required can be reduced considerably by the use of Halton sequences. However Train (2003) warns that Halton draw results need to be viewed with caution and describes some anomalous findings. Both pseudo-random draws (with and without antithetics) and Halton draws are provided as options in the program described below. The comparison is discussed further in the context of the empirical illustration in Section 6 below.

5 The redpace command

5.1 Syntax

`redpace` *depvar varlist* (*varlist*$_{init}$) [`if` *exp*] [`in` *range*] [, `i`(*varname*) `t`(*varname*) `seed`(#) `seg`(#) `halton` `primes`(*matname*) `drop`(#) `from`(*matname*) `mavg` `noauto`]

The lagged dependent variable must be constructed by the user and must appear as the first variable in *varlist*. It is the user's responsibility to ensure that both this variable and *depvar* are binary 0/1 variables. *varlist* should additionally contain the variables in x. *varlist*$_{init}$ should contain the variables in z.

`redpace` requires a balanced panel, that is the number of "time" observations, T, for each cross-section unit must be the same. This is checked by the program.

The parameters, λ, θ, and ρ (in the AR(1) model) are constrained by use of transformations. Logit, log and arc-hyperbolic tangent transformations are used respectively. The transformed parameters that are estimated are therefore: $\ln(\lambda/(1-\lambda))$, $\ln(\theta)$, and $\frac{1}{2}\ln((1+\rho)/(1-\rho))$. The estimates of the original parameters and their approximate asymptotic standard errors are presented after convergence using `_diparm`.

5.2 Options

`i`(*varname*) specifies the variable name that contains the cross-section identifier, corresponding to index i.

`t`(*varname*) specifies the variable name that contains the time-series identifier, corresponding to index t.

`rep`(#) specifies the number of replications, or draws, R.

`seed`(#) specifies the initial value of the pseudo-random number seed to be used by the `uniform()` function. Use the `seed` option to ensure reproducibility of results. The number specified must be an integer. The default is 81234567. `seed` is ignored if `halton` is specified.

`seg`(#) specifies symmetric systematic sampling (with antithetics) to be used and specifies the number of segments of the unit interval that are to be used. `seg(2)` provides antithetic sampling. The default value is 1 (i.e. standard sampling). If `seg` is specified, the number of replications, R, must be a multiple of the number of segments, and the number of segments must be a multiple of 2.

`halton` specifies that Halton quasi-random sequences are to be used rather than pseudo-random numbers, which is the default. This option requires the program `mdraws` (Cappellari and Jenkins, 2006).

`primes`(*matname*) specifies a $1 \times (T-1)$ matrix containing the prime numbers to be used for the Halton sequences. Ignored if `halton` not specified. The numbers

specified must be integers. It is the user's responsibility to ensure that they are primes.

`drop(#)` specifies the number of initial elements of the Halton sequence to be dropped for burn in. Ignored if `halton` not specified. The default is 0.

`from(`*matname*`)` specifies a matrix containing starting values for the parameters of the model. Use this option to check that a global maximum has been found. Also use to reduce required number of iterations or to restart a previously halted run. The default uses a pooled probit for $t \geq 2$ and separate probit for the initial period reduced form.

`mavg` specifies that the first-order moving average model should be used. The default is the first-order autoregressive model.

`noauto` specifies that the model without autocorrelated errors (i.e. the Heckman model of Section 3) is to be estimated (by MSL). This is useful for comparing with the Gaussian-Hermite quadrature estimates of the same model to see if the value of R chosen is large enough in that case.

6 Example using Stata Reference Manual data

This section gives an example of the use of the command and the output produced using the union data (http://www.stata-press.com/data/r9/union.dta) used in [R] **xtprobit** to model the probability of union membership. The data are for US young women and are from the NLSY. A subsample of the dataset is used: (1) only data from 1978 onwards are used, (2) the data for 1983 are dropped, and (3) only those individuals observed in each of the remaining 6 waves are kept:

```
drop if year<78
drop if year==83
by idcode: gen nwav=_N
keep if nwav==6
```

This gives $N = 799$ individuals observed in each of $T = 6$ waves and hence a sample size of $NT = 4,794$. In the example here the observations for 85 and 87 are implicitly treated as if they were for 84 and 86 respectively, which would give 6 waves at regular 2-year intervals.

In addition to the lagged dependent variable, the model used includes `age` (age in current year), `grade` (years of schooling completed), and `south` (1 if resident in south). These variables are contained in x in the specification of the model given above. The vector z additionally contains the variable `not_smsa` (1 if living outside a standard metropolitan statistical area). This variable has a significant negative effect on the probability of union membership in the initial period reduced form, whether estimated as a separate probit or as part of the full model.

The output from the `redpace` program using pseudo-random numbers without antithetics and $R = 500$ to estimate the AR(1) model is as follows:

```
. sort idcode year

. by idcode: gen tper = _n

. by idcode: gen Lunion = union[_n-1]
(799 missing values generated)

. redpace union Lunion age grade south (age grade south not_smsa), /*
> */  i(idcode) t(tper) rep(500) seed(945430778) from(bstart1)
```

```
Pseudo-random number draws
Seed set to  945430778
# of replications = 500
Standard sampling

(output deleted)

RE Dynamic Probit Model with AR1 errors           Number of obs   =       4794
                                                  Wald chi2(4)    =     155.53
Log likelihood = -1854.0621                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
       union |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
union        |
      Lunion |   1.322202   .1537599     8.60   0.000     1.020838    1.623566
         age |  -.0234323    .008068    -2.90   0.004    -.0392452   -.0076194
       grade |  -.0363382   .0199431    -1.82   0.068    -.0754259    .0027495
       south |  -.3695182   .0991797    -3.73   0.000    -.5639069   -.1751295
       _cons |   .0803735    .400068     0.20   0.841    -.7037454    .8644924
-------------+----------------------------------------------------------------
rfper1       |
         age |   .0108658    .024266     0.45   0.654    -.0366946    .0584262
       grade |  -.0133834   .0333403    -0.40   0.688    -.0787291    .0519623
       south |  -.7548028   .1667058    -4.53   0.000     -1.08154   -.4280655
    not_smsa |  -.4195162   .1659528    -2.53   0.011    -.7447778   -.0942547
       _cons |   -.891098   .8482112    -1.05   0.293    -2.553561    .7713655
-------------+----------------------------------------------------------------
   /logitlam |    .075697     .28295     0.27   0.789    -.4788748    .6302688
      /atar1 |  -.3512991   .0650661    -5.40   0.000    -.4788263   -.2237719
     /ltheta |   .2041693   .1742898     1.17   0.241    -.1374323     .545771
-------------+----------------------------------------------------------------
      lambda |   .5189152   .0706363     7.35   0.000     .3825179    .6525504
         ar1 |  -.3375271   .0576535    -5.85   0.000    -.4453032   -.2201101
       theta |   1.226506   .2137674     5.74   0.000     .8715933    1.725939
------------------------------------------------------------------------------
```

The estimate of γ is positive and highly significant. The three x-variables all have negative effects on the conditional probability of union membership. The effects of `age` and `south` are significant at the 1% level, that of `grade` at the 10% level. The estimate of λ implies that 52% of the composite error variance is attributed to that in the individual-specific effects. The estimate of θ is significantly greater than 0 and insignificantly different from 1. The estimate of ρ is significantly negative, implying that successive realizations of u_{it} are negatively correlated.

Alternative estimators for the model are presented in Table 1. The first estimates presented are simple pooled probit estimates. These ignore the cross-correlation between the composite error term in different time periods for the same individual. This gives an estimate of γ of 1.88, larger than for the other estimators in the table, but the estimates of the β-coefficients are all smaller and less significant than for the other estimators. However care needs to be taken with such comparisons. The random effects models and the pooled probit model use different normalizations. The random effects models use a normalization of $\sigma_u^2 = 1$, while the pooled probit estimator uses $\sigma_v^2 = 1$. When making comparison with pooled probit estimates, random effects model estimates need to be multiplied by an estimate of $\sigma_u/\sigma_v = \sqrt{1-\lambda}$.

The **xtprobit** estimates in the next column allow individual-specific effects, but take the initial condition to be exogenous. This results in a considerable reduction in the estimate of γ. Allowing for the different normalizations, the scaled estimate of the coefficient on lagged union membership is 0.7949, less than half the pooled probit estimate. The estimated coefficients on the x-variables are all more significant than the pooled probit estimates and, even after rescaling for the different normalizations, larger (in absolute value). The restriction $\lambda = 0$ reduces this random effects probit specification to the pooled probit without the random individual-specific effects. This restriction, i.e. the absence of individual-specific effects, can be tested by comparing

columns [1] and [2]. However adjustment to the standard test is required since the parameter vector under the null hypothesis lies on the boundary of the parameter space, and the standard asymptotic theory of classical tests is not applicable in such boundary situations. (See Godfrey, 1988, and Andrews, 2001.) The quasi-likelihood-ratio statistic for this restriction equals 20.92. The limiting distribution of the test statistic is equal to one half of the distribution function of a $\chi^2(1)$ random variable plus one half. Thus for example the 1% asymptotic critical value for the quasi-likelihood-ratio test is given by the 98th percentile of a $\chi^2(1)$ random variable, 5.412 (display *invchi2(1,0.98)*). The p-value for this quasi-likelihood-ratio test statistic can be calculated as *chi2tail(1,q)/2*, where q is the test statistic. In this case the p-value is zero to several decimal places. The hypothesis $\lambda = 0$ is strongly rejected.

The third column gives the results from using the Heckman estimator of Section 3 (using the program **redprob**). This allows for the endogeneity of the initial conditions, but assumes no autocorrelation in the u_{it}. This results in a reduction in the estimate of γ compared to the **xtprobit** estimates, and a further increase in absolute value in the β slope coefficients. Exogeneity of the initial conditions in the random effects model can be viewed as resulting from imposing $\theta = 0$ on this model. Here too, testing this hypothesis must allow for the fact that it is on the boundary of the parameter space. The quasi-likelihood-ratio statistic, comparing columns [2] and [3], is 222.76. The p-value, calculated as above, is zero to several decimal places. The hypothesis $\theta = 0$ is strongly rejected.

The fourth column gives the corresponding estimates using Maximum Simulated Likelihood estimation instead of Gaussian-Hermite quadrature, and using the same pseudo-random number draws as used for the AR(1) model in the final column. There is reasonably close agreement with the previous column, suggesting that 500 replications is sufficient for the MSL estimator at least in the no autocorrelation case.

The final column of the table reproduces the estimates from the Stata output given above for the AR(1) model. Compared with the model without autocorrelation, the estimate of γ doubles. The increase in $\hat{\gamma}$ is to be expected since $\hat{\rho} < 0$. When the estimates are scaled relative to σ_v rather than σ_u, the increase is even more marked. The estimates of the β slope coefficients decline slightly in absolute value when autocorrelation is introduced and are close to the corresponding **xtprobit** estimates. The hypothesis of no autocorrelation is strongly rejected. The transformed parameter **atar1** has a z-ratio of -5.40 and hence gives a Wald $\chi^2(1)$ test statistic of 29.2. The likelihood-ratio test statistic (also $\chi^2(1)$) is 11.8. The random effects probit model with AR(1) errors in column [5] clearly dominates the other columns of Table 1.

To investigate the number of replications required for the MSL estimator, the same model was estimated (with the same initial seed) for different values of R. The estimates of the main parameters of interest in the model (γ, λ, ρ) and the value of the maximised log simulated likelihood are given in Table 2. The estimates change relatively little once R is greater than about 100. The same is true for the maximised log simulated likelihood. Note that the maximised log simulated likelihood is not monotonic in R. How much accuracy one demands of the parameter estimates is a matter of judgement. On the basis of this particular seed the 500 replications used initially seems sufficient.

On a 2.4GHz Pentium 4 Windows XP machine running Stata/SE 9.0 this model with 500 replications took 15 hours 37 minutes for 4 iterations to convergence. Run times are close to proportional to R.

Another useful approach to investigating the number of replications required is to examine the MSL estimates based on different seeds. Of course this type of exercise is almost limitless. The limited exercise conducted here examined the MSL estimates for 10 different seeds (themselves randomly chosen). The 10 seeds were selected, to avoid overlaps of the sequences, as follows:

```
set obs 10
set seed 987654321
gen long s=int((uniform()+10-_n)*100000000)
```

The resulting 10 seeds are listed in Table 3. The estimates of the main parameters of interest in the AR(1) model (γ, λ, ρ) and the value of the maximised log simulated likelihood using $R = 500$ based on each of these 10 seeds are given in Table 3. The estimate of γ varies between 1.301 and 1.333, a range of 0.032, slightly over 2% of the mean estimate. The estimates of λ have a range of 0.015 and those of ρ a range of 0.012. These seem relatively small. However the maximised log simulated likelihood varies between -1853.55 and -1855.72 across the ten seeds. This gap of 2.17 is probably rather more than one would want to see.

The variation in the estimates of the key parameters across the ten seeds for different values of R is shown in Table 4. Required accuracy is a matter of judgement, but the gaps between the minimum and maximum estimates for $R \leq 100$ might be viewed as greater than acceptable.

The same model was also estimated using Halton sequences of quasi-random numbers. To investigate the number of replications required for the MSL estimator in this case, the model was estimated using 10 different sets of primes. (Each run requires 5 primes since $T = 6$.) The 10 sets of primes used are given in Table 5 together with the estimates, using $R = 100$, of the main parameters of interest (γ, λ, ρ) and the value of the maximised log simulated likelihood. While not essential, the primes 2 and 5 are not used with this value of R to avoid perfect correlations due to cycles. The estimates of γ vary between 1.302 and 1.324, a range of 0.022. The estimates of λ have a range of 0.011 and those of ρ a range of 0.009. These are all slightly less than the corresponding ranges for pseudo-random numbers with $R = 500$. The maximised log simulated likelihood has a range of 0.97, which is less than half the range for 500 pseudo-random numbers in Table 3.

The variation in the estimates of the key parameters across the ten sets of primes for different values of R is shown in Table 6. Although again subject to judgement, the variation for $R < 100$ might be viewed as greater than acceptable, while that for $R = 100$ compares favourably with the results based on 500 pseudo-random numbers.

It may be debated whether comparing the variation in MSL estimates using Halton sequences of quasi-random numbers across different sets of primes is comparable with comparing the variation in MSL estimates using pseudo-random numbers across different seeds. However, subject to this, there is evidence here of the use of Halton sequences reducing the number of replications required. Further investigation of this issue is part of ongoing research.

7 Acknowledgements

Thanks to Stephen Jenkins for useful discussions and to him and an anonymous referee for helpful comments on an earlier draft of the paper.

8 References

Andrews, D.W.K. 2001. "Testing when a parameter is on the boundary of the maintained hypothesis". *Econometrica* **69**: 683–734.

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

Cameron, A.C. and P.K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Cappellari, L. and S.P. Jenkins. 2006. "Calculation of multivariate normal probabilities by simulation". In this issue of *Stata Journal*.

Chamberlain, G. 1984. "Panel data", in Z. Griliches and M. Intrilligator (eds.). *Handbook of Econometrics*. Amsterdam: North-Holland.

Godfrey, L.G. 1988. *Misspecification Tests in Econometrics*. Cambridge: Cambridge University Press.

Gourieroux, C. and A. Monfort. 1996. *Simulation-Based Econometric Methods*. Oxford: Oxford University Press.

Heckman, J.J. 1981. "The incidental parameters problem and the problem of initial conditions in estimating a discrete time - discrete data stochastic process", in C.F. Manski and D. McFadden (eds.). *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge: MIT Press.

Hyslop, D.R. 1999. "State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women". *Econometrica* **67**: 1255–94.

Keane, M. 1994. "A computationally practical simulation estimator for panel data". *Econometrica* **62**: 95–116.

Mundlak, Y. 1978. "On the pooling of time series and cross section data". *Econometrica* **46**: 69–85.

Stewart, M.B. 2005. "The inter-related dynamics of unemployment and low-wage employment". Forthcoming *Journal of Applied Econometrics*.

Train, K.E. 2003. *Discrete Choice Models with Simulation*. Cambridge: Cambridge University Press.

Wooldridge, J.M. 2005. "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity", *Journal of Applied Econometrics*, **20**, 39–54.

Table 1
Estimates for different models

	Pooled probit	RE probit (**xtprobit**)	**redprob** quadrature	**redpace** no auto MSL	**redpace** with AR(1) MSL
	[1]	[2]	[3]	[4]	[5]
γ	1.8849	1.1507	0.6344	0.6350	1.3222
	(0.0525)	(0.1421)	(0.0983)	(0.0985)	(0.1538)
age	-0.0087	-0.0240	-0.0286	-0.0285	-0.0234
	(0.0058)	(0.0086)	(0.0092)	(0.0092)	(0.0081)
grade	-0.0145	-0.0387	-0.0539	-0.0532	-0.0363
	(0.0103)	(0.0207)	(0.0269)	(0.0268)	(0.0199)
south	-0.1685	-0.3692	-0.4883	-0.4903	-0.3695
	(0.0519)	(0.1034)	(0.1239)	(0.1241)	(0.0992)
constant	-0.6986	0.1788	0.5633	0.5609	0.0804
	(0.2474)	(0.4183)	(0.4799)	(0.4795)	(0.4001)
λ		0.5228	0.6996	0.6991	0.5189
		(0.0730)	(0.0345)	(0.0346)	(0.0706)
θ			0.8641	0.8613	1.2265
			(0.1095)	(0.1089)	(0.2138)
ρ					-0.3375
					(0.0577)
No. obs.	3995	3995	4794	4794	4794
Model lnL	-1573.64	-1563.18	-1860.22	-1859.96	-1854.06
Total lnL	-1982.06	-1971.60	-1860.22	-1859.96	-1854.06

Notes:

1. Standard errors in brackets.
2. **redpace** estimates given are based on 500 replications.
3. The seed used is 945430778.
4. **redprob** and **xtprobit** estimates given are based on 24 quadrature points.
5. For the Pooled probit and RE probit columns, the total log-likelihood is the sum of the model log-likelihood (for $t \geq 2$; 3995 observations) and that for a simple probit for the initial period reduced form ($t = 1$; 799 observations).

Table 2
MSL estimates for AR(1) model for different numbers of replications

R	γ	λ	ρ	Ln(SL)
20	1.376 (0.147)	0.479 (0.070)	-0.328 (0.053)	-1859.84
50	1.341 (0.154)	0.502 (0.072)	-0.329 (0.057)	-1858.15
100	1.312 (0.152)	0.521 (0.069)	-0.329 (0.057)	-1854.22
200	1.318 (0.154)	0.519 (0.070)	-0.334 (0.058)	-1853.80
500	1.322 (0.154)	0.519 (0.071)	-0.338 (0.058)	-1854.06
1000	1.316 (0.156)	0.521 (0.071)	-0.335 (0.059)	-1854.31

Notes:
1. Standard errors in brackets.
2. MSL based on pseudo-random numbers. The seed used is 945430778.

Table 3
MSL estimates for AR(1) model for R=500 and different seeds
Pseudo-random numbers

Row	Seed	γ	λ	ρ	Ln(SL)
1	945430778	1.322 (0.154)	0.519 (0.071)	-0.338 (0.058)	-1854.06
2	862683501	1.318 (0.155)	0.520 (0.071)	-0.335 (0.059)	-1854.74
3	700921694	1.323 (0.156)	0.518 (0.072)	-0.336 (0.058)	-1855.72
4	642850439	1.301 (0.156)	0.528 (0.070)	-0.329 (0.060)	-1854.51
5	594203018	1.326 (0.155)	0.516 (0.072)	-0.337 (0.058)	-1855.58
6	480067244	1.304 (0.156)	0.527 (0.070)	-0.331 (0.060)	-1853.85
7	366110265	1.327 (0.153)	0.517 (0.070)	-0.338 (0.058)	-1855.34
8	241963761	1.322 (0.152)	0.520 (0.070)	-0.338 (0.057)	-1854.53
9	177063593	1.333 (0.154)	0.513 (0.072)	-0.341 (0.058)	-1855.51
10	80102774	1.321 (0.153)	0.520 (0.070)	-0.339 (0.058)	-1853.55

Notes:
1. Standard errors in brackets.

Table 4
Variation in MSL parameter estimates across different seeds
Pseudo-random numbers

Parameter	Mean	Min	Max	Std. Dev.
$R = \mathbf{20}$				
γ	1.388	1.249	1.489	.0646
λ	0.475	0.423	0.531	.0287
ρ	-0.334	-0.378	-0.270	.0277
$R = \mathbf{50}$				
γ	1.356	1.289	1.451	.0466
λ	0.497	0.451	0.521	.0211
ρ	-0.336	-0.372	-0.299	.0196
$R = \mathbf{100}$				
γ	1.343	1.310	1.383	.0241
λ	0.506	0.486	0.521	.0114
ρ	-0.338	-0.350	-0.329	.0075
$R = \mathbf{200}$				
γ	1.333	1.311	1.372	.0182
λ	0.513	0.492	0.524	.0091
ρ	-0.338	-0.350	-0.333	.0059
$R = \mathbf{500}$				
γ	1.320	1.301	1.333	.0100
λ	0.520	0.513	0.528	.0046
ρ	-0.336	-0.341	-0.329	.0036

Notes:
1. All statistics calculated across 10 seeds.
2. Seeds used are given in Table 3.

Table 5
MSL estimates for AR(1) model for R=100 and different sets of primes
Halton quasi-random numbers

Row	Primes	γ	λ	ρ	Ln(SL)
1	(3, 7, 11, 13, 17)	1.305 (0.156)	0.526 (0.070)	-0.330 (0.059)	-1854.89
2	(7, 11, 13, 17, 19)	1.322 (0.154)	0.519 (0.070)	-0.338 (0.058)	-1854.84
3	(3, 11, 13, 17, 19)	1.312 (0.154)	0.524 (0.070)	-0.336 (0.058)	-1854.08
4	(3, 7, 13, 17, 19)	1.322 (0.153)	0.519 (0.070)	-0.338 (0.058)	-1854.46
5	(3, 7, 11, 17, 19)	1.324 (0.154)	0.517 (0.071)	-0.339 (0.058)	-1854.79
6	(3, 7, 11, 13, 19)	1.314 (0.155)	0.522 (0.071)	-0.335 (0.059)	-1854.72
7	(7, 11, 13, 17, 23)	1.313 (0.156)	0.523 (0.071)	-0.333 (0.059)	-1855.05
8	(3, 11, 13, 17, 23)	1.302 (0.158)	0.528 (0.071)	-0.330 (0.060)	-1854.32
9	(3, 7, 13, 17, 23)	1.318 (0.156)	0.521 (0.071)	-0.335 (0.059)	-1854.65
10	(3, 7, 11, 17, 23)	1.316 (0.157)	0.521 (0.072)	-0.334 (0.060)	-1855.01

Notes:
1. Standard errors in brackets.

Table 6
Variation in MSL parameter estimates across different sets of primes
Halton quasi-random numbers

Parameter	Mean	Min	Max	Std. Dev.
$R = \mathbf{10}$				
γ	1.284	1.242	1.325	0.0283
λ	0.530	0.512	0.549	0.0118
ρ	-0.302	-0.325	-0.282	0.0141
$R = \mathbf{20}$				
γ	1.315	1.260	1.368	0.0371
λ	0.517	0.495	0.537	0.0151
ρ	-0.327	-0.350	-0.295	0.0180
$R = \mathbf{50}$				
γ	1.335	1.314	1.366	0.0158
λ	0.508	0.493	0.518	0.0082
ρ	-0.335	-0.342	-0.329	0.0039
$R = \mathbf{80}$				
γ	1.331	1.314	1.348	0.0135
λ	0.516	0.508	0.523	0.0056
ρ	-0.342	-0.350	-0.335	0.0059
$R = \mathbf{100}$				
γ	1.315	1.302	1.324	0.0073
λ	0.522	0.517	0.528	0.0033
ρ	-0.335	-0.339	-0.330	0.0032

Notes:
1. All statistics calculated across 10 sets of primes.
2. Sets of primes used are given in Table 5.