The Stata Journal

The *Stata Journal*, electronic version (ISSN 1536-8734) is a publication of Stata Press, and Stata is a registered trademark of StataCorp LP.

The Stata Journal (2006)
6, Number 2, pp. 256–272

Maximum simulated likelihood estimation of random-effects dynamic probit models with autocorrelated errors

Mark Stewart
Economics Department
University of Warwick
Coventry, UK
mark.stewart@warwick.ac.uk

Abstract. This paper investigates using maximum simulated likelihood (MSL) estimation for random-effects dynamic probit models with autocorrelated errors. It presents and illustrates a new Stata command, `redpace`, for this estimator. The paper also compares using pseudorandom numbers and Halton sequences of quasirandom numbers for MSL estimation of these models.

Keywords: st0106, redpace, simulation estimation, maximum simulated likelihood, Halton sequences, autocorrelated errors

1 Introduction

This paper examines estimation of dynamic probit models, specifically models in which the outcome probability depends on the outcome in the previous period. The presence of omitted individual heterogeneity, in the form of individual-specific effects, causes an "initial conditions" problem and renders the standard random-effects (RE) 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 (1981). 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 RE probit estimators, such as that in Stata's `xtprobit` command. I also provide a new Stata routine for the Heckman estimator of this model.

If the error terms are autocorrelated, the Heckman estimator is inconsistent. Extending it to the autocorrelated case involves higher-dimensional integrals. Maximum simulated likelihood (MSL) is a natural estimator to use. This paper describes and implements in Stata an MSL estimator for the RE dynamic probit model with autocorrelated errors. I give an empirical illustration and investigate the issue of the required number of simulations.

2 An RE 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 follows,

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

and where x_{it} is a vector of explanatory variables and $u_{it} \sim N(0, \sigma^2_u)$. The subscript i indexes individuals and the subscript t indexes 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 because of the individual-specific time-invariant α_i terms. The individual-specific RE specification adopted implies equicorrelation between the v_{it} in any two (different) periods:

$$\lambda = \text{Corr}(v_{it}, v_{is}) = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_u} \qquad\qquad t, s = 2, \ldots, T; t \neq s$$

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

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

$$\Pr\left(y_{it} | x_{it}, y_{it-1}, \alpha_i\right) = \Phi\left\{(\gamma y_{it-1} + x'_{it}\beta + \alpha_i)(2y_{it} - 1)\right\}$$

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 very strong. For example, in the union membership illustration used below, even if each individual were observed from labor market entry onward, the assumption would require

an individual's union status in the first job to be unrelated to her individual-specific factor α_i. Under this assumption, a standard RE 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 usually be expected, this method of estimation will tend to overstate the degree of state dependence, γ.

3 Heckman's estimator for 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,

$$y^*_{i1} = z'_{i1}\pi + \eta_i$$

$(i = 1, \ldots, N)$, 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, we can write it as follows:

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

$(\theta > 0)$, with α_i and u_{i1} independent of one another. We also assume 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 period is therefore specified as

$$y^*_{i1} = z'_{i1}\pi + \theta\alpha_i + u_{i1}$$

$(i = 1, \ldots, N)$, where z will include period 1 values of x variables, typically together with presample 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

$$\Phi\left\{(z'_{i1}\pi + \theta\alpha_i)(2y_{i1} - 1)\right\} \prod_{t=2}^{T} \Phi\left\{(\gamma y_{it-1} + x'_{it}\beta + \alpha_i)(2y_{it} - 1)\right\}$$

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

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

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 (forthcoming) for an application of the estimator in investigating 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 web site, http://www2.warwick.ac.uk/fac/soc/economics/staff/faculty/stewart/.

1. Wooldridge (2005) has proposed an alternative conditional maximum-likelihood estimator for the case of serially independent errors being considered in this section.

4 Models with autocorrelated errors

If the error term u_{it} is autocorrelated, reflecting, for example, correlation between transitory shocks, this complicates estimation considerably. Extending the Heckman estimator to this case involves T-dimensional integrals of normal densities. Simulation estimators provide a feasible way to address this problem. I present here an MSL estimator (Gouriéroux and Monfont 1996 and Cameron and Trivedi 2005), based on the GHK algorithm of Geweke, Hajivassiliou, and Keane (Keane 1994). 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 speaking, 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 a first-order autoregressive [AR(1)] process or a first-order moving-average [MA(1)] process. The variance–covariance matrix, Ω, of $v_i = (v_{i1}, \ldots, v_{iT})'$, will now be a function of λ, θ, and one more 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 y's can be written as the product of recursively defined conditional probabilities.

Using the Cholesky decomposition, we can write the latent equations as

$$y^*_{it} = \mu_{it} + \sum_{j=1}^{t} c_{tj}\eta_{ij}$$

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 y's is given by

$$\begin{aligned}\Pr_i = \Phi\{(y_{i1}-1)a_{i1}\} &\times \int_{L_{i1}}^{U_{i1}} \Phi\{(y_{i2}-1)a_{i2}\}\phi(\eta_{i1})d\eta_{i1} \\ &\times \int_{L_{i1}}^{U_{i1}} \int_{L_{i2}}^{U_{i2}} \Phi\{(y_{i3}-1)a_{i3}\}\phi(\eta_{i1})\phi(\eta_{i2})d\eta_{i2}d\eta_{i1} \\ &\times \cdots\end{aligned}$$

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 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 periods.

The simulated likelihood is then given by

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

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

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

with $-1 < \rho < 1$. Here the variance–covariance matrix of v_i is given by

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

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:

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

Here the variance–covariance matrix of v_i is given by

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

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 pseudorandom number draws (from a standard uniform), which are assumed to be independent. Efficiency of the estimator can be improved by using variance reduction methods that use dependent draws. The best known is antithetic sampling. Here 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 reduction is not guaranteed. This idea can be extended by using segments of the unit interval in conjunction with the antithetics. The subdivision provided by this "symmetric systematic sampling" gives more even coverage, but there is a tradeoff between coverage and randomness (see Train 2003, 221–224).

Another approach to improving the efficiency of the estimator is to replace the pseudorandom numbers with quasirandom 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, using these sequences produces variance reduction and hence improves efficiency. To put this another way, a particular level of efficiency can be achieved with fewer 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 using Halton sequences. However, Train (2003) warns that Halton draw results need to be viewed with caution and describes some anomalous findings. The program described below gives both pseudorandom draws (with and without antithetics) and Halton draws as options. The comparison is discussed further in the context of the empirical illustration in section 6.

5 The redpace command

5.1 Syntax

`redpace` *depvar varlist* (*varlist*$_{\text{init}}$) [*if*] [*in*] [, `i(`*varname*`)` `t(`*varname*`)`
`rep(#)` `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*. The user must ensure that both this variable and *depvar* are binary 0/1 variables. *varlist* should also contain the variables in x. *varlist*$_{\text{init}}$ 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 balance is checked by the program.

The parameters λ, θ, and ρ (in the AR(1) model) are constrained by using 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

$(1/2)\ln\{(1+\rho)/(1-\rho)\}$. 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 pseudorandom 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 `seed(81234567)`, but this option 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 `seg(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 quasirandom sequences be used rather than pseudorandom 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. This option is ignored if `halton` is not specified. The numbers specified must be integers. The user must ensure that they are primes.

`drop(#)` specifies the number of initial elements of the Halton sequence to be dropped for burn in. This option is ignored if `halton` is not specified. The default is `drop(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. It also can be used to reduce the 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 option 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 using the command and the output produced from the `union` data (http://www.stata-press.com/data/r9/union.dta) used in [XT] **xtprobit** to model the probability of union membership. The data are for U.S. young women and are from the National Longitudinal Survey of Youth. A subsample of the dataset is used: (1) only data from 1978 onward are used, (2) the data for 1983 are dropped, and (3) only those individuals observed in each of the remaining six waves are kept.

```
. use http://www.stata-press.com/data/r9/union
(NLS Women 14-24 in 1968)
. sort idcode year
. drop if year<78
(10136 observations deleted)
. drop if year==83
(2194 observations deleted)
. by idcode: gen nwaves=_N
. keep if nwaves==6
(9076 observations deleted)
```

This code gives $N = 799$ individuals observed in each of $T = 6$ waves and hence a sample size of $NT = 4{,}794$. Here the observations for 85 and 87 are implicitly treated as if they were for 84 and 86, respectively, which would give six 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 also 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 pseudorandom numbers without antithetics and $R = 500$ to fit 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)
. mat bstart1 = ( 1.322202, -.0234323, -.0363382,
> -.3695182,  .0803734,  .0108658, -.0133834, -.7548028,
> -.4195162, -.8910983,  .0756969, -.3512991,  .2041694)
. mat colnames bstart1 = union:Lunion union:age union:grade union:south
> union:_cons rfper1:age rfper1:grade rfper1:south
> rfper1:not_smsa rfper1:_cons logitlam:_cons atar1:_cons
> ltheta:_cons
```

```
. 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: # of replications = 500
Seed set to  945430778
Standard sampling
```

(*output omitted*)

```
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.

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

Table 1: Estimates for different models

Parameter	Pooled probit (SE) [1]	RE probit (`xtprobit`) (SE) [2]	`redprob` quadrature (SE) [3]	`redpace` no auto MSL (SE) [4]	`redpace` with AR(1) MSL (SE) [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. of obs.	3,995	3,995	4,794	4,794	4,794
Model lnL	−1,573.64	−1,563.18	−1,860.22	−1,859.96	−1,854.06
Total lnL	−1,982.06	−1,971.60	−1,860.22	−1,859.96	−1,854.06

Note: `redpace` estimates given are based on 500 replications.
The seed used is 945430778.
`redprob` and `xtprobit` estimates given are based on 24 quadrature points.
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).

The `xtprobit` estimates in the next column allow individual-specific effects but take the initial condition to be exogenous. Using this estimator considerably reduces 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 RE 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 quasilikelihood-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 quasilikelihood-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 quasilikelihood-ratio test statistic can be calculated as `chi2tail(1,q)/2`, where `q` is the test statistic. Here 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 estimator allows for the endogeneity of the initial conditions but assumes no autocorrelation in the u_{it}. Using this estimator reduces the estimate of γ compared with the `xtprobit` estimates and further increases the absolute value in the β slope coefficients. Exogeneity of the initial conditions in the RE model can be viewed as resulting from imposing $\theta = 0$ on this model. Here too, testing this hypothesis must allow for its being on the boundary of the parameter space. The quasilikelihood-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 MSL estimation instead of Gaussian–Hermite quadrature and using the same pseudorandom 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 $\widehat{\gamma}$ is to be expected since $\widehat{\rho} < 0$. 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 RE 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. Table 2 gives the estimates of the main parameters of interest in the model (γ, λ, ρ) and the value of the maximized log simulated likelihood. The estimates change relatively little once R is greater than about 100. The same is true for the maximized log simulated likelihood.

The maximized log simulated likelihood is not monotonic in R. How much accuracy one demands of the parameter estimates is a matter of judgment. On the basis of this particular seed, the 500 replications used initially seems sufficient.

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

R	γ (SE)	λ (SE)	ρ (SE)	ln(SL)
20	1.376 (0.147)	0.479 (0.070)	−0.328 (0.053)	−1,859.84
50	1.341 (0.154)	0.502 (0.072)	−0.329 (0.057)	−1,858.15
100	1.312 (0.152)	0.521 (0.069)	−0.329 (0.057)	−1,854.22
200	1.318 (0.154)	0.519 (0.070)	−0.334 (0.058)	−1,853.80
500	1.322 (0.154)	0.519 (0.071)	−0.338 (0.058)	−1,854.06
1,000	1.316 (0.156)	0.521 (0.071)	−0.335 (0.059)	−1,854.31

Note: MSL based on pseudorandom numbers. The seed used is 945430778.

On a 2.4-GHz Pentium 4 Windows XP machine running Stata/SE 9.0, this model with 500 replications took 15 hours, 37 minutes for four 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)
```

Table 3 gives the resulting 10 seeds. The estimates of the main parameters of interest in the AR(1) model (γ, λ, ρ) and the value of the maximized 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 more than 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 ranges seem relatively small. However, the maximized log simulated likelihood varies between −1,853.55 and −1,855.72 across the 10 seeds. This gap of 2.17 is probably rather more than one would want to see.

Table 3: MSL estimates for AR(1) model for $R = 500$ and different seeds (pseudorandom numbers)

Row	Seed	γ (SE)	λ (SE)	ρ (SE)	ln(SL)
1	945430778	1.322 (0.154)	0.519 (0.071)	−0.338 (0.058)	−1,854.06
2	862683501	1.318 (0.155)	0.520 (0.071)	−0.335 (0.059)	−1,854.74
3	700921694	1.323 (0.156)	0.518 (0.072)	−0.336 (0.058)	−1,855.72
4	642850439	1.301 (0.156)	0.528 (0.070)	−0.329 (0.060)	−1,854.51
5	594203018	1.326 (0.155)	0.516 (0.072)	−0.337 (0.058)	−1,855.58
6	480067244	1.304 (0.156)	0.527 (0.070)	−0.331 (0.060)	−1,853.85
7	366110265	1.327 (0.153)	0.517 (0.070)	−0.338 (0.058)	−1,855.34
8	241963761	1.322 (0.152)	0.520 (0.070)	−0.338 (0.057)	−1,854.53
9	177063593	1.333 (0.154)	0.513 (0.072)	−0.341 (0.058)	−1,855.51
10	80102774	1.321 (0.153)	0.520 (0.070)	−0.339 (0.058)	−1,853.55

Table 4 shows the variation in the estimates of the key parameters across the 10 seeds for different values of R. Required accuracy is a matter of judgment, but the gaps between the minimum and maximum estimates for $R \leq 100$ might be viewed as greater than acceptable.

Table 4: Variation in MSL parameter estimates across different seeds (pseudorandom numbers)

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

Note: All statistics calculated across 10 seeds.
Seeds used are given in table 3.

The same model was also estimated using Halton sequences of quasirandom 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 five 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 maximized log simulated likelihood. Although 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 numbers are all slightly less than the corresponding ranges for pseudorandom numbers with $R = 500$. The maximized log simulated likelihood has a range of 0.97, which is less than half the range for 500 pseudorandom numbers in table 3.

Table 5: MSL estimates for AR(1) model for $R = 100$ and different sets of primes (Halton quasirandom numbers)

Row	Primes	γ (SE)	λ (SE)	ρ (SE)	ln(SL)
1	(3, 7, 11, 13, 17)	1.305 (0.156)	0.526 (0.070)	−0.330 (0.059)	−1,854.89
2	(7, 11, 13, 17, 19)	1.322 (0.154)	0.519 (0.070)	−0.338 (0.058)	−1,854.84
3	(3, 11, 13, 17, 19)	1.312 (0.154)	0.524 (0.070)	−0.336 (0.058)	−1,854.08
4	(3, 7, 13, 17, 19)	1.322 (0.153)	0.519 (0.070)	−0.338 (0.058)	−1,854.46
5	(3, 7, 11, 17, 19)	1.324 (0.154)	0.517 (0.071)	−0.339 (0.058)	−1,854.79
6	(3, 7, 11, 13, 19)	1.314 (0.155)	0.522 (0.071)	−0.335 (0.059)	−1,854.72
7	(7, 11, 13, 17, 23)	1.313 (0.156)	0.523 (0.071)	−0.333 (0.059)	−1,855.05
8	(3, 11, 13, 17, 23)	1.302 (0.158)	0.528 (0.071)	−0.330 (0.060)	−1,854.32
9	(3, 7, 13, 17, 23)	1.318 (0.156)	0.521 (0.071)	−0.335 (0.059)	−1,854.65
10	(3, 7, 11, 17, 23)	1.316 (0.157)	0.521 (0.072)	−0.334 (0.060)	−1,855.01

Table 6 shows the variation in the estimates of the key parameters across the 10 sets of primes for different values of R. Although again subject to judgment, the variation for $R < 100$ might be viewed as greater than acceptable, whereas that for $R = 100$ compares favorably with the results based on 500 pseudorandom numbers.

Table 6: Variation in MSL parameter estimates across different sets of primes (Halton quasirandom numbers)

Parameter	Mean	Min	Max	Std. dev.
$R = 10$				
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
$R = 20$				
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
$R = 50$				
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
$R = 80$				
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
$R = 100$				
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

Note: All statistics calculated across 10 sets of primes.
Sets of primes used are given in table 5.

It may be debated whether comparing the variation in MSL estimates using Halton sequences of quasirandom numbers across different sets of primes is comparable to comparing the variation in MSL estimates using pseudorandom numbers across different seeds. However, subject to this comparability issue, there is evidence that using Halton sequences reduces the number of replications required. More investigation of this issue is part of ongoing research.

7 Acknowledgments

I thank Stephen Jenkins 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: 761–764.

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

Cappellari, L., and S. P. Jenkins. 2006. Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *Stata Journal* 6: 156–189.

Chamberlain, G. 1984. Panel data. In *Handbook of Econometrics*, ed. Z. Griliches and M. Intrilligator, vol. 2, 1247–1318. Amsterdam: North Holland.

Godfrey, L. G. 1988. *Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches*. Cambridge: Cambridge University Press.

Gouriéroux, C., and A. Monfont. 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 *Structural Analysis of Discrete Data and Econometric Applications*, ed. C. F. Manski and D. L. McFadden, 114–178. Cambridge, MA: MIT Press.

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

Keane, M. P. 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. Forthcoming. The inter-related dynamics of unemployment and low-wage employment. *Journal of Applied Econometrics*.

Train, K. 2003. *Discrete Choice Methods 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.

About the author

Mark Stewart is a professor of economics at the University of Warwick, Coventry, UK. His research interests are mainly in labor market economics and applied econometrics.