A Computationally Practical Simulation Estimation Algorithm for Dynamic Panel Data Models with Unobserved Endogenous State Variables

Michael P. Keane
University of Technology Sydney
Arizona State University

Robert M. Sauer
University of Bristol

Introduction

- Further develop a new SML algorithm for dynamic discrete choice models that deals with:

    - missing endogenous state variables during the sample period

    - the initial conditions problem

    - serially correlated errors

- Advantages of the algorithm

    - computationally simple

      uses *unconditional* simulations in construction of likelihood

    - incorporates classification error

      required to form likelihood from unconditional simulations

- We show the estimator has good small sample properties

    - repeated sampling experiments on Markov and Polya models

    - Polya models allow for more general dynamics/state dependence

- We apply the estimator to a model of female labour supply and show:

    - the Polya model fits the data better than the standard Markov model

    - allowing for more general state dependence effects is important

        - get far less state dependence and smaller race effects

        - get stronger effects of education, children and non-labour income

Outline

- Specify Panel Data Probit Model
- Discuss Models of Classification Error (CE)
- Describe the SML Algorithm
- Show Monte Carlo Results
- Application to model of Female Labor Supply

Panel Probit Model

$$u_{it} = \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau + \varepsilon_{it}$$

$$d_{it} = \begin{cases} 1 \text{ if } u_{it} \geq 0 \\ 0 \text{ otherwise.} \end{cases}$$

$$\begin{aligned} \varepsilon_{it} &= \mu_i + \xi_{it} \\ \xi_{it} &= \phi_1 \xi_{i,t-1} + \eta_{it} \\ x_{it} &= \phi_2 x_{i,t-1} + \nu_{it} \end{aligned}$$

$$\begin{aligned} \mu_i &\sim N(0, \sigma_\mu^2) \\ \eta_{it} &\sim N(0, \sigma_\eta^2) \\ \nu_{it} &\sim N(0, \sigma_\nu^2) \end{aligned}$$

simulate data from model so algorithm can easily handle wider range of distributions

Classification Error (CE)

- General model of CE we consider

$$\pi_{10t} = \Pr\left(d_{it}^{*} = 0 \mid d_{it} = 1\right)$$
$$\pi_{01t} = \Pr\left(d_{it}^{*} = 1 \mid d_{it} = 0\right)$$
$$\pi_{00t} = 1 - \pi_{01t}$$
$$\pi_{11t} = 1 - \pi_{10t}$$

  as in Poterba and Summers (1986, 1995) and HAS (1998)

- HAS (1998) develops identification conditions (more later)

- Only need tractable expression for π_{jkt}'s to form likelihood

CE Model 1: Unbiased Classification Error

- imposes unconditional prob report an option equals true prob

$$\Pr(d_{it}^* = 1) = \Pr(d_{it} = 1)$$

generates tractable linear expressions for π_{jkt}'s

$$\begin{aligned} \pi_{11t} &= \Pr(d_{it}^* = 1 | d_{it} = 1) \\ &= E + (1 - E)\Pr(d_{it} = 1) \\ \pi_{01t} &= \Pr(d_{it}^* = 1 | d_{it} = 0) \\ &= (1 - E)\Pr(d_{it} = 1) \end{aligned}$$

- E is estimable parameter where
  - low prob events have prob equal to E of being classified correctly
  - prob of correct classification increases linearly in E
- $\Pr(d_{it} = 1)$ is easily simulated

CE Model 2: Biased Classification Error

- Do not impose $\Pr(d_{it}^* = 1) = \Pr(d_{it} = 1)$ rather assume:

$$l_{it} = \gamma_0 + \gamma_1 d_{it} + \gamma_2 d_{it-1}^* + \omega_{it}$$

$$d_{it}^* = \begin{cases} 1 \text{ if } l_{it} \geq 0 \\ 0 \text{ otherwise.} \end{cases}$$

$$\omega_{it} \sim \text{logistic}$$

- Also get tractable expressions for π_{jkt}'s:

$$\pi_{11t} = \Pr\left(d_{it}^* = 1 \mid d_{it} = 1\right) = \frac{e^{\gamma_0 + \gamma_1 + \gamma_2 d_{it-1}^*}}{1 + e^{\gamma_0 + \gamma_1 + \gamma_2 d_{it-1}^*}}$$

$$\pi_{01t} = \Pr\left(d_{it}^* = 1 \mid d_{it} = 0\right) = \frac{e^{\gamma_0 + \gamma_2 d_{it-1}^*}}{1 + e^{\gamma_0 + \gamma_2 d_{it-1}^*}}$$

- Easily incorporates dynamic (persistent) misreporting

- d_{it-1}^* can be simulated from l_{it} model if missing

Identification

- rewrite $\Pr(d_{it}^* = 1)$ as

$$\pi_{11t}\Pr(d_{it} = 1) + \pi_{01t}\Pr(d_{it} = 0)$$
$$= (1 - \pi_{10t})\Pr(d_{it} = 1) + \pi_{01t}(1 - \Pr(d_{it} = 1))$$
$$= \pi_{01t} + (1 - \pi_{10t} - \pi_{01t})\Pr(d_{it} = 1)$$

- shows need non-linear $\Pr(d_{it} = 1)$ and monotonicity $\pi_{10t} + \pi_{01t} < 1$

- otherwise standard identification issues

  - SD effects from causal effect of lagged X's
  - SD effects sensitive to modeling of serial correlation (RE or AR(1), etc.)

- in biased CE model, lagged reported choice identified because not perfectly correlated with lagged X's

The SML Estimation Algorithm

Data: $\{D_i^*, x_i\}_{i=1}^N$, $D_i^* = \{d_{it}^*\}_{t=1}^T$, $x_i = \{x_{it}\}_{t=1}^T$

1. Draw M times from the ε_{it} distribution to form $\left\{\{\{\varepsilon_{it}^m\}_{t=1}^T\}_{i=1}^N\right\}_{m=1}^M$

2. Given $\{\{x_{it}\}_{t=1}^T\}_{i=1}^N$ and $\left\{\{\{\varepsilon_{it}^m\}_{t=1}^T\}_{i=1}^N\right\}_{m=1}^M$, construct $\left\{\{\{d_{it}^m\}_{t=1}^T\}_{i=1}^N\right\}_{m=1}^M$ according to model

3. Construct conditional probs $\left\{\{\hat{\pi}_{jkt}^m\}_{t=1}^T\right\}_{m=1}^M$

4. Form a simulator likelihood contribution:

$$\hat{P}\left(D_i^* \mid \theta, x_i\right) =$$

$$\frac{1}{M} \sum_{m=1}^{M} \prod_{t=1}^{T} \left(\sum_{j=0}^{1} \sum_{k=0}^{1} \hat{\pi}_{jkt}^m I[d_{it}^m = j, d_{it}^* = k] \right)^{I\left(d_{it}^* \text{ observed}\right)}$$

Important Things to Note

- any observed choice history has non-zero prob conditional on any simulated history

- *unconditional* simulation method completely avoids usual integrating out problem

- consistency and asymptotic normality require that $\frac{M}{\sqrt{N}} \to \infty$ as $N \to \infty$ (as in Pakes and Pollard (1989) and Lee (1992))

- but still need to check small sample properties of estimator

What if Missing X's and/or Initial Conditions Problem?

- if missing x_{it}'s, can simulate them and modify likelihood to include density of x_{it}'s
- if initial conditions problem can simulate model from $t = 0, \ldots, T$ with $d_{i0} = x_{i0} = 0$
- or can imbed Heckman's solution - specify marginal distribution for $d_{i\tilde{\tau}}$, then simulate from $t = \tilde{\tau}, \ldots, T$
- or can imbed Wooldridge's solution - random effect function of $d_{i\tilde{\tau}}$ and covariates, simulate from $t = \tilde{\tau} + 1, \ldots, T$

Importance Sampling (Smooth) Version of Algorithm

- construct $\{d_{it}^m(\theta_0)\}_{t=1}^T$ and $\{U_{it}^m(\theta_0)\}_{t=1}^T$ and hold fixed as vary θ

- calculate importance sampling weights as vary θ

$$W_m(\theta) = \frac{g\left(U_{i1}^m(\theta_0),\ldots,U_{iT}^m(\theta_0) \mid \theta, x_i\right)}{g\left(U_{i1}^m(\theta_0),\ldots,U_{iT}^m(\theta_0) \mid \theta_0, x_i\right)}$$

$$g(U_i^m(\theta_0)|\theta, x_i) = \prod_{t=1}^{T} \frac{1}{\sigma_\varepsilon}\phi\left(\frac{a}{\sigma_\varepsilon}\right)$$

$$a = U_{it}^m(\theta_0) - \beta_0 - \beta_1 x_{it} - \sum_{\tau=0}^{t-1} d_{i\tau}^m(\theta_0)\rho_\tau$$

- likelihood contribution becomes

$$\widehat{P}\left(D_i^* \mid \theta, x_i\right) = \frac{1}{M}\sum_{m=1}^{M} W_m(\theta)\prod_{t=\tilde{\tau}}^{T} f_m(x_{it})^{I\left(x_{it}\ observed\right)}$$
$$\left(\sum_{j=0}^{1}\sum_{k=0}^{1}\widehat{\pi}_{jkt}^m I[d_{it}^m = j, d_{it}^* = k]\right)^{I\left(d_{it}^*\ observed\right)}$$

Repeated Sampling Experiments

- RE+AR(1) Polya Model with Exponential Decay:

$$
\begin{aligned}
u_{it} &= \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau + \varepsilon_{it} \\
\rho_\tau &= \rho e^{-\alpha(t-\tau-1)} \\
\varepsilon_{it} &= \mu_i + \xi_{it} \\
\xi_{it} &= \phi_1 \xi_{i,t-1} + \eta_{it} \\
\eta_{it} &\sim N(0, (1-\sigma_\mu^2)(1-\phi_1^2)) \\
x_{it} &= \phi_2 x_{i,t-1} + \nu_{it}, \nu_{it} \sim N(0, \sigma_\nu^2)
\end{aligned}
$$

- Sample Size: $N = 500,\ T = 10$
- Replications: $R = 50$
- Simulated histories per person: $M = 1000$
- Also do experiments on Markov model

Table 3

Repeated Sampling Experiments
Random Effects Polya Model
Unbiased Classification Error
(No Initial Conditions Problem)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	*RMSE*	t-Stat
20% Missing Choices and X's $(t = 1, \ldots, 10)$						
β_0	-.1000	-.1051	-.1023	.0436	.0439	-.83
β_1	1.0000	1.0167	1.0191	.0611	.0634	1.92
ρ	1.0000	1.0479	1.0446	.0444	.0653	7.63
α	.5000	.4977	.5031	.0656	.0657	-.24
ϕ_2	.2500	.2520	.2505	.0176	.0177	.80
σ_v	.5000	.5015	.5016	.0057	.0059	1.86
σ_μ	.8000	.8056	.8017	.0287	.0292	1.38
E	.7500	.7428	.7430	.0172	.0187	-2.95
60% Missing Choices and X's $(t = 1, \ldots, 10)$						
β_0	-.1000	-.0997	-.1116	.0542	.0543	.05
β_1	1.0000	1.034	1.0258	.0894	.0924	1.85
ρ	1.0000	1.0401	1.0512	.0682	.0791	4.15
α	.5000	.4957	.4973	.0721	.0722	-.42
ϕ_2	.2500	.2507	.2498	.0372	.0373	.13
σ_v	.5000	.5011	.5017	.0089	.0090	.88
σ_μ	.8000	.8096	.8044	.0421	.0432	1.61
E	.7500	.7493	.7440	.0288	.0288	-.16

Table 4

Repeated Sampling Experiments
Random Effects Polya Model
Unbiased Classification Error
(No Missing Choices or X's, Initial Conditions Problem)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
Simulate from start of process with $d_{i0} = 0$ $(t = 11, \ldots, 20)$						
β_0	-.1000	-.1001	-.1022	.0295	.0295	-.02
β_1	1.0000	1.0286	1.0337	.0454	.0537	4.46
ρ	1.0000	1.0298	1.0253	.0324	.0440	6.51
α	.5000	.5044	.5004	.0320	.0323	.98
ϕ_2	.2500	.2501	.2526	.0135	.0135	.05
σ_v	.5000	.5015	.5025	.0042	.4985	2.56
σ_μ	.8000	.8130	.8145	.0245	.0277	3.74
E	.7500	.7450	.7410	.0193	.0199	-1.82
Assume process starts with $d_{i,10} = 0$ $(t = 11, \ldots, 20)$						
β_0	-.1000	.9367	.9513	.0543	1.0381	135.05
β_1	1.0000	.2966	.2844	.0938	.7096	-53.01
ρ	1.0000	.9543	.9333	.3278	.3310	-.99
α	.5000	.4187	.3995	.2957	.3067	-1.94
σ_μ	.8000	.9905	.9923	.0090	.1907	149.11
E	.7500	.7144	.7125	.0230	.0424	-10.96

Table 14

Repeated Sampling Experiments
Random Effects Polya Model
Biased Classification Error
Smooth Algorithm
(20% Missing Choices and X's, No Initial Conditions Problem)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	*RMSE*	t-Stat
β_0	-.1000	-.0795	-.0686	.0685	.0714	2.12
β_1	1.0000	1.0265	1.0330	.0833	.0874	2.25
ρ	1.0000	.9466	.9374	.1410	.1508	-2.68
α	.5000	.4409	.4360	.1038	.1195	-4.02
ϕ_2	.2500	.2480	.2472	.0153	.0155	-.91
σ_v	.5000	.5019	.5027	.0048	.0052	2.76
σ_μ	.8000	.8211	.8225	.0321	.0384	4.65
γ_0	-3.5000	-3.3313	-3.2996	.2606	.3104	4.58
γ_1	5.0000	4.7243	4.7334	.3014	.4084	-6.47
γ_2	2.0000	2.1031	2.0794	.2372	.3185	3.07

Table 17

Repeated Sampling Experiments
Polya Model with Random Effects and $AR(1)$ Errors
Biased Classification Error
Smooth Algorithm
(20% Missing Choices and X's, No Initial Conditions Problem)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
β_0	-.1000	-.0823	-.0824	.0513	.0543	2.44
β_1	1.0000	1.0215	1.0082	.0907	.0932	1.67
ρ	1.0000	.9782	.9948	.1459	.1475	-1.06
α	.5000	.4709	.4931	.1092	.1130	-1.89
ϕ_2	.2500	.2477	.2487	.0154	.0155	-1.04
σ_ν	.5000	.5020	.5028	.0048	.0052	2.89
σ_μ	.8000	.8267	.8280	.0372	.0458	5.07
ϕ_1	.4000	.3892	.4114	.1223	.1228	-.62
γ_0	-3.5000	-3.3261	-3.2815	.2645	.3165	4.65
γ_1	5.0000	4.7020	4.7290	.3270	.4424	-6.44
γ_2	2.0000	2.1233	2.1126	.2316	.3495	3.76

Empirical Application

- Estimate Polya and Markov models of female labor force participation using PSID data 1994-2003

- Missing data problem because PSID went biannual after 1997 (missing choices and X's in 1998, 2000 and 2002)

- Embed AR(1) process for nonlabor income

- Simulate choices from age 16 (theoretical start of process)

- Test for endogeneity of fertility and nonlabor income (as in Chamberlain (1984), Jakubson (1988) and Hyslop (1999))

Table 18
Sample Characteristics
PSID Calendar Years 1994-2003
Missing Years 1998, 2000, and 2002
(N=1310)

	Mean (1)	Std. Dev. (2)
Participation	.816	.291
(avg. over 7 years)	(.008)	
Husband's Annual Earnings	46.40	41.18
(avg. over 7 years)	(11.38)	
($1000 1994)		
No. Children aged 0-2 years	.135	.231
(avg. over 10 years)	(.006)	
No. Children aged 3-5 years	.181	.254
(avg. over 10 years)	(.007)	
No. Children aged 6-17 years	.937	.864
(avg. over 10 years)	(.024)	
Age	36.93	8.00
(1994)	(.221)	
Education	13.56	2.10
(maximum over 10 years)	(.06)	
Race	.198	.398
(1=Black)	(.011)	

Table 19
Markov Models
Biased Classification Error
Smooth Algorithm

	Random Effects (1)	Correlated Random Effects (2)	Random Effects + $AR(1)$ Errors (3)	Correlated Random Effects + $AR(1)$ Errors (4)
$ln(y_{it})$	-.1669 (.0020)	-.1510 (.0035)	-.1697 (.0013)	-.1646 (.0024)
$\#kids0\text{-}2_t$	-.6433 (.0036)	-.5382 (.0046)	-.6659 (.0031)	-.4271 (.0038)
$\#kids3\text{-}5_t$	-.3342 (.0033)	-.3524 (.0043)	-.3650 (.0026)	-.3379 (.0032)
$\#kids6\text{-}17_t$	-.0845 (.0015)	-.0830 (.0028)	-.0808 (.0011)	0.0734 (.0019)
$age_t/10$	.6676 (.0105)	.5818 (.0129)	.6887 (.0101)	.6792 (.0112)
$age_t^2/100$	-.1438 (.0012)	-.1364 (.0014)	-.1525 (.0010)	-.1565 (.0011)
$race_i$	.5547 (.0034)	.5467 (.0040)	.4518 (.0025)	.4533 (.0031)
$education_i$	.0501 (.0076)	.0407 (.0081)	.0581 (.0059)	.0392 (.0062)
ρ	2.3148 (.0256)	2.3582 (.0263)	2.4047 (.0243)	2.5099 (.0251)
ϕ_2	.9993 (.0052)	.9993 (.0058)	.9992 (.0047)	.9993 (.0049)
σ_v	.2719 (.0061)	.2718 (.0063)	.2758 (.0060)	.2755 (.0061)
σ_μ	.8947 (.0012)	.8949 (.0014)	.8877 (.0011)	.8905 (.0013)
γ_0	-.8535 (.0428)	-.9716 (.0521)	-0.8346 (.0419)	-.9454 (.0495)
γ_1	3.3974 (.0589)	3.4328 (.0625)	3.6335 (.0544)	3.5653 (.0583)
γ_2	1.5943 (.0923)	1.6178 (.0968)	1.7012 (.0915)	1.6734 (.0937)
ϕ_1	-	-	.6084 (.0079)	.6136 (.0085)
Log-Like	-12673.61	-12651.32	-12668.19	-12637.15
χ^2 $(\delta = 0)$	-	44.58 (.0243)	-	62.08 (.0002)
χ^2 (GOF)	59.62 (.1024)	57.15 (.1474)	58.32 (.1245)	56.40 (.1637)
N	1310	1310	1310	1310

Table 20
Polya Models
Biased Classification Error
Smooth Algorithm

	Random Effects (1)	Correlated Random Effects (2)	Random Effects + $AR(1)$ Errors (3)	Correlated Random Effects + $AR(1)$ Errors (4)
$ln(y_{it})$	-.3089 (.0019)	-.3111 (.0024)	-.3066 (.0015)	-.3040 (.0018)
$\#kids0\text{-}2_t$	-.5964 (.0043)	-.6000 (.0047)	-.6495 (.0035)	-.6339 (.0042)
$\#kids3\text{-}5_t$	-.3648 (.0034)	-.3565 (.0039)	-.3325 (.0032)	-.3466 (.0038)
$\#kids6\text{-}17_t$	-.0145 (.0015)	-.0123 (.0021)	-.0211 (.0012)	-.0225 (.0014)
$age_t/10$	.7527 (.0110)	.7387 (.0112)	.7081 (.0109)	.7263 (.0111)
$age_t^2/100$	-.1310 (.0012)	-.1274 (.0014)	-.1262 (.0010)	-.1280 (.0013)
$race_i$	.3083 (.0033)	.2272 (.0035)	.2945 (.0031)	.2684 (.0033)
$education_i$	.0652 (.0074)	.0558 (.0081)	.0630 (.0069)	.0611 (.0077)
ρ	.6363 (.0087)	.7281 (.0095)	.6758 (.0084)	.6979 (.0089)
α	1.8924 (.0763)	1.9502 (.0821)	2.1278 (.0712)	2.1457 (.0759)
ϕ_2	.9994 (.0055)	.9994 (.0057)	.9994 (.0054)	.9994 (.0055)
σ_v	.2743 (.0072)	.2736 (.0073)	.2742 (.0066)	.2736 (.0069)
σ_μ	.8949 (.0015)	.8970 (.0016)	.8952 (.0013)	.8960 (.0015)
γ_0	-1.1203 (.0498)	-.8962 (.0510)	-.9940 (.0482)	-.9404 (.0491)
γ_1	3.8880 (.0610)	3.6738 (.0625)	3.6809 (.0600)	3.7190 (.0611)
γ_2	1.6520 (.0981)	1.5320 (.0989)	1.5658 (.0979)	1.6096 (.0980)
ϕ_1	-	-	.4606 (.0091)	.4596 (.0098)
Log-Like	-12568.10	-12544.89	-12561.69	-12531.88
χ^2 $(\delta = 0)$	-	46.42 (.0158)	-	59.62 (.0005)
χ^2 (GOF)	54.62 (.2075)	51.90 (.2887)	53.32 (.2442)	51.02 (.3186)
N	1310	1310	1310	1310

Concluding Remarks

- New SML algorithm may have a significant computational advantage over others (e.g., GHK, MCMC and EM) in certain missing data contexts

- Algorithm requires assumption of CE but we believe this is reasonable in almost all empirical applications in economics

- Estimator has good small sample properties

- It produces sensible results in empirical applications of female labour supply behaviour

- Estimator is very easy to implement