Centre for Structural Econometrics

Discussion paper No. 09/01

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

Michael P. Keane
University of Technology Sydney
and
Arizona State University

Robert M. Sauer
University of Bristol

October 2009

Centre for Structural Econometrics
University of Bristol Department of Economics
8 Woodland Road, Bristol BS8 1TN, UK.
http://www.efm.bris.ac.uk/cse/
econ-cse@bristol.ac.uk

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

Michael P. Keane and Robert M. Sauer[1]

Abstract

This paper develops a simulation estimation algorithm that is particularly useful for estimating dynamic panel data models with unobserved endogenous state variables. Repeated sampling experiments on dynamic probit models with serially correlated errors indicate the estimator has good small sample properties. We apply the estimator to a model of female labor supply and show that the rarely used Polya model fits the data substantially better than the popular Markov model. The Polya model also produces far less state dependence and race effects, and much stronger effects of education, young children and husband's income on female labor supply decisions.

Keywords: Initial Conditions, Missing Data, Classification Error, Simulated Maximum Likelihood, Female Labor Supply

JEL Codes: C15,C23,C25,J13,J21

Shortened Title: A Simulation Estimation Algorithm

[1]This research is partially supported by the Australian Research Council, through a grant to Michael Keane (ARC grant numbers FF0561843 and DP0774247), and the Economic and Social Research Council of the United Kingdom, through a grant to Robert Sauer (ESRC grant number RES-000-22–1529).

1 Introduction

The problem of unobserved endogenous state variables arises frequently in estimation of dynamic discrete choice models. It arises when there are unobserved initial conditions, i.e., the choice process begins prior to the first period of observed data. It also arises if data on some choices is missing *during* the sample period. In either case, consistent estimation requires "integrating out" all possible choice sequences that the individual *may* have followed. However, as the length of the panel grows and the choice set becomes larger, the "integrating out" solution begins to require very high dimensional integrations, often rendering it computationally impractical.

In this paper, we assess the performance of a new simulated maximum likelihood (SML) algorithm that is particularly useful for estimating dynamic panel data models with unobserved endogenous state variables. This novel technique was introduced by Keane and Wolpin (2001) (KW) to estimate a discrete choice dynamic programming model with both unobserved initial conditions and missing choices during the sample period. However, the algorithm has wide applicability beyond the special case that KW considered. In fact, it can be used to simulate the likelihood in any context where it is tractable to perform *unconditional* simulations of data from the model.

The appeal of the new SML algorithm lies in the fact that there are many contexts where conditional simulation of data from a model is extremely difficult, while unconditional simulation is straightforward. Simulation of the likelihood in dynamic models often involves simulation of choice probabilities conditional on past history, but when history is not fully observed this is often computationally infeasible.[2]

In this study, we first describe how the SML algorithm developed by KW, which only requires unconditional simulations, can be extended to cases beyond the specific discrete dynamic programming problem they considered. In particular, we assess the

[2]For example, the GHK algorithm (see Keane (1994)) builds up the likelihood of a choice history via a series of conditional simulations. This may be infeasible in some cases (like that in KW) where part of the history is unobserved. We discuss cases where GHK has trouble in Section 5.

performance of the estimator on panel probit models with a time-varying exogenous covariate, lagged endogenous variables and serially correlated errors. These models have served as a leading case in past discussions of dynamic models with unobserved initial conditions (see Heckman (1981a)). Specification of panel probit models, rather than dynamic programs, allows us to focus on and further develop the estimation technique. The results of a series of repeated sampling experiments show that the SML estimator with the new algorithm has good small sample properties.

We then apply the algorithm to dynamic probit models of the employment status of married women, using PSID data from 1994-2003. A serious missing data problem arises because, in addition to the usual initial conditions problem, respondents were not interviewed in 1998, 2000 and 2002. Hyslop (1999) also used the PSID to estimate panel probit models of female employment status. Using the new algorithm, we extend his model to allow for classification error and missing data. This allows us to include the post-1994 data, and adopt a more general Polya model of state dependence. The Polya model fits the data much better than the traditional Markov model.

The rest of the paper is organized as follows. Section 2 reviews different approaches to the problem of unobserved endogenous state variables. Section 3 describes the dynamic panel probit model used in the repeated sampling experiments. Section 4 shows how we incorporate classification error in outcomes into the panel probit model, as this plays a key role in the algorithm. Section 5 describes our algorithm in detail. Sections 6 and 7 present our Monte-Carlo test results. Section 8 applies the algorithm to a model of female employment decisions. Section 9 concludes.

2 Background

Several solutions to the initial conditions problem, a special case of the problem of unobserved endogenous state variables, have been proposed. Heckman (1981a) showed how, in dynamic discrete choice models, one can derive the marginal probability of the initial state given stationarity. As stationarity is often problematic, Heckman (1981a) also considered estimation of fixed effects models. But he concluded it works better to approximate the probability of the initial state by a separate probit function.[3] More recently, Wooldridge (2003) proposed an alternative approximate solution to the initial conditions problem. (His approach is essentially identical to that adopted by Keane and Wolpin (1997) in the context of a dynamic programming model). Below, we compare the Heckman and Wooldridge methods to the "exact" solution obtained by using our algorithm to simulate from the start of the stochastic process.

In contrast to the initial conditions problem, the problem of missing data *during* the sample period has been less extensively explored. But missing data problems frequently arise in data sets used by economists, such as the National Longitudinal Survey of Youth (NLSY) and the Panel Study of Income Dynamics (PSID).

One method for dealing with missing data during the sample period is the EM algorithm (Dempster, Laird and Rubin (1977)). However, in EM it is often difficult to compute the conditional distribution required for the E (expectation) step (see Ruud (1991)). Another potential solution is the Gibbs-sampling algorithm. Geweke and Keane (2000) used this approach to deal with unobserved initial conditions and missing data in dynamic earnings models. But in Gibbs, as in EM, the distribution of a missing value conditional on all other information can be quite complex (see Geweke and Keane (2001)). Also, Geweke and Keane (2000) noted that Gibbs sometimes exhibited instability when integrating over long pre-sample histories.

Given these computational difficulties, applied economists frequently resort to case deletion or ad hoc imputation to handle missing data. Case deletion can cause large amounts of information to be lost, resulting in inefficient estimates. It can also introduce biases if complete histories differ systematically from censored histories. Imputation of missing values by ad hoc methods is also problematic. For instance, imputing averages tends to bias estimated variances and covariances toward zero.

[3] This approach still produced biases of more than 10% in repeated sampling experiments.

In contrast to the previous literature, the SML algorithm we propose offers a systematic unified "solution" to both the initial conditions and missing data problems. It does not involve case deletion or ad hoc imputations, yet it is computationally simple. It is simple because it does not require calculation of the initial state probability, or probabilities of events at each date t conditional on the history up through $t-1$, which is the usual approach to constructing the likelihood in dynamic models. In our algorithm, *unconditional* simulations of the model are used to form the likelihood.

The key assumption required to form the likelihood using unconditional simulations is that choices are measured with error. Then one can use unconditional *frequency* simulation to calculate probabilities of choice histories, as the usual problem that an impractically large number of draws is needed to obtain non-zero frequencies of low probability events is avoided. Furthermore, the assumption that choices are measured with error is certainly valid in the vast majority of data sets.

Prior work on the importance of classification error includes Poterba and Summers (1986, 1995) and Flinn (1997). Poterba and Summers (1986) estimate that in the CPS the probability an employed person falsely reports being unemployed or out-of-the-labor-force is 1.5%, while the probability an unemployed person falsely reports being employed is 4.0% (our calculations based on their Table II). If misclassification is present and ignored in the analysis, maximum likelihood estimation can lead to severely biased parameter estimates (Hausman, Abrevaya and Scott-Morton (1998)).[4]

The classification error process that we adopt simply specifies a probability the reported choice is the true choice, and a probability it is not. This is without loss of generality, as the investigator is free to specify the details of the process. All that is required is that one can obtain tractable expressions for the probability of observed choices conditional on true choices. We illustrate the flexibility of the algorithm by considering two very different models of classification error in our experiments.

[4]Repeated sampling experiments in Hausman et al. (1998) find considerable biases, in the range of 15% to 25%, in ordinary probit models that fail to incorporate classification error into the likelihood.

3 The Panel Data Probit Model

In the panel probit model, the utility of the first option, for individual i at time t, is denoted as u_{it}, and that of the second option is normalized to zero. The individual chooses the option with highest utility. The researcher observes the choice but not the utility. We consider models of the general form

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

where x_{it} is a strictly exogenous covariate and d_{it} is the indicator function

$$d_{it} = \begin{cases} 1 \text{ if } u_{it} \geq 0 \\ 0 \text{ otherwise.} \end{cases} \tag{2}$$

Note that specification (1) allows the entire history of past choices to affect current utility. It is, therefore, more general than the familiar first-order Markov process.[5] Depreciation in the importance of past choices is captured through the weights ρ_τ. The theoretical start of the process is, by definition, $d_{i0} = 0$.

We assume the error term ε_{it} in (1) is serially correlated. Hence, lagged choices are endogenous. With serially independent errors, lagged choices are exogenous, and the problems we consider in this paper do not arise. Our approach can accommodate very flexible forms of serial correlation, but we consider three leading cases in our experiments:

First, we consider time-invariant individual random effects, i.e.,

$$\varepsilon_{it} = \mu_i + \eta_{it} \tag{3}$$

where $\mu_i \sim N(0, \sigma_\mu^2)$ and $\eta_{it} \sim N(0, \sigma_\eta^2)$. Second, we consider an $AR(1)$ process,

$$\varepsilon_{it} = \phi_1 \varepsilon_{i,t-1} + \eta_{it} \tag{4}$$

[5] Higher order processes have not been widely applied in the economics literature. We suspect that this is due, in part, to difficulty in dealing with missing data. But higher order processes are quite standard in marketing. See, e.g., Keane (1997) and Erdem and Keane (1997).

where $\eta_{it} \sim N(0, \sigma_\eta^2)$. Third, we consider a random effects plus $AR(1)$ process:

$$\begin{aligned} \varepsilon_{it} &= \mu_i + \xi_{it} \\ \xi_{it} &= \phi_1 \xi_{i,t-1} + \eta_{it} \end{aligned} \quad (5)$$

where $\eta_{it} \sim N(0, \sigma_\eta^2)$.

While the model of (1)-(5) is restrictive, our algorithm can accommodate many alternative specifications. For example, KW considered a multinomial choice setting where the error contained both a nonparametric individual effect and a multivariate normal disturbance. Also, while we only consider a scalar process in (1), extension to vectors of discrete and mixed discrete/continuous outcomes (as in KW) is straightforward. Our goal here is to focus on relatively simple processes, so that repeated sampling experiments are feasible. Furthermore, the relatively simple processes we consider have been widely used in the literature, and have been the focus of prior work on the initial conditions problem (see Heckman (1981a) and Wooldridge (2003)).

4 Classification Error

In our approach, we assume that all discrete outcomes are measured subject to classification error (which is quite realistic in nearly all data sets used by economists). Our approach can be implemented for any classification error process, provided one can obtain a tractable expression for the probability of observed choices conditional on true choices. Letting d_{it}^* denote the reported choice, we define the four probabilities:

$$\begin{aligned} \pi_{11t} &= \Pr\left(d_{it}^* = 1 | d_{it} = 1\right), \pi_{10t} = 1 - \pi_{11t} \\ \pi_{01t} &= \Pr\left(d_{it}^* = 1 | d_{it} = 0\right), \pi_{00t} = 1 - \pi_{01t} \end{aligned} \quad (6)$$

where π_{11t} is the probability that option one is reported to be chosen ($d_{it}^* = 1$) given that it is the true choice ($d_{it} = 1$), and π_{01t} is the probability that option one is falsely reported ($d_{it}^* = 1$) given that option two is the true choice ($d_{it} = 0$). π_{00t} and π_{10t} are the corresponding conditional probabilities for option two ($d_{it}^* = 0$).

The investigator has a great deal of leeway in specifying the classification error rates π_{01t} and π_{10t}. In our Monte Carlo analysis we consider cases where error *rates* are dependent on true choices. For example, Poterba and Summers (1995) and Hausman et. al. (1998) find evidence that workers who change jobs mis-report more often than workers who do not. Similarly, Flinn (1997) finds that mis-reporting of dismissals in the NLSY is an increasing function of the true dismissal rate.

Covariate-dependent misclassification could also be easily incorporated into the model. However, if the measurement error process is a sufficiently flexible function of covariates and lagged choices, one would lose identification of the structural parameters in (1). Moreover, economic theory provides guidance for specification of the decision model but not necessarily for the model of misclassification. Thus, we prefer to focus on fairly simple specifications of the classification error process. We consider specifications distinguished by whether classification error is biased or unbiased, and whether there is dynamic mis-reporting.

4.1 Unbiased Classification Error

The assumption that classification error is unbiased imposes a very simple structure on the conditional probabilities in (6). Unbiasedness in this context means that the probability a person is observed to choose an option is equal to the true probability that he/she chooses that option, or $\Pr(d_{it}^* = 1) = \Pr(d_{it} = 1)$. The assumption of unbiased classification error is appealing because it forces the structural parameters of the model to fit the conditional choice frequencies in each period, as opposed to allowing classification error to "sop up" the errors of the structural model.

Unbiased classification error implies that the conditional probabilities in (6) are linear in the true choice probability. To see this, note that by definition,

$$\begin{aligned}\Pr(d_{it}^* = 1) &= \Pr(d_{it}^* = 1 | d_{it} = 1) \Pr(d_{it} = 1) \\ &\quad + \Pr(d_{it}^* = 1 | d_{it} = 0) \Pr(d_{it} = 0)\end{aligned} \tag{7}$$

where, in writing $\Pr(d^*_{it} = 1)$ and $\Pr(d_{it} = 1)$, we suppress the obvious dependence of these probabilities on x_{it} and lagged true choices in order to conserve on notation. If we write the conditional probabilities as the following linear functions of $\Pr(d_{it} = 1)$,

$$\begin{aligned} \Pr(d^*_{it} = 1 | d_{it} = 1) &= E + (1 - E)\Pr(d_{it} = 1) \\ \Pr(d^*_{it} = 1 | d_{it} = 0) &= (1 - E)\Pr(d_{it} = 1), \end{aligned} \tag{8}$$

where E is a parameter on $(0, 1]$, then these expressions can be substituted into (7) to yield $\Pr(d^*_{it} = 1) = \Pr(d_{it} = 1)$.

Note that as the true choice probability, $\Pr(d_{it} = 1)$, approaches one, the probability of a correct classification, $\Pr(d^*_{it} = 1 | d_{it} = 1)$, also approaches one, which must be the case to preserve unbiasedness. Further, as $\Pr(d_{it} = 1)$ approaches zero, $\Pr(d^*_{it} = 1 | d_{it} = 1)$ approaches E. Thus, E can be interpreted as a "baseline" classification rate. That is, very low probability events have a probability equal to E of being classified correctly. The probability of a correct classification increases linearly from E toward one as the true choice probability approaches one.

In terms of the original notation, the conditional probabilities in (6) can be written:

$$\begin{aligned} \pi_{11t} &= E + (1 - E)\Pr(d_{it} = 1) \\ \pi_{01t} &= (1 - E)\Pr(d_{it} = 1). \end{aligned} \tag{9}$$

Note the great parsimony that unbiasedness imposes on the error process. It depends on the single parameter E, which we estimate. One could generalize the process to let E depend on covariates. Then, one obtains unbiasedness *conditional* on covariates.

4.2 Biased Classification Error

Any classification error scheme that does not impose the linear relationships in (8) will lead to a biased process in which $\Pr(d^*_{it} = 1) \neq \Pr(d_{it} = 1)$. The biased process that we consider as an alternative to (8) is characterized by the following index function,

$$l_{it} = \gamma_0 + \gamma_1 d_{it} + \gamma_2 d^*_{it-1} + \omega_{it} \tag{10}$$

where d_{it}^{*} denotes the reported choice and ω_{it} is a stochastic term. If $l_{it} > 0$ then $d_{it}^{*} = 1$, while $d_{it}^{*} = 0$ otherwise. Note that (10) allows the probability of reporting a particular choice to depend on the true choice. It also allows for dynamics: the greater is γ_2 (the coefficient on d_{it-1}^{*}), the greater is the persistence in mis-reporting.

Assuming ω_{it} is logistic yields tractable expressions for classification probabilities:

$$\begin{aligned} \pi_{11t} &= \Pr\left(d_{it}^{*} = 1 | d_{it} = 1\right) = \frac{e^{\gamma_0+\gamma_1+\gamma_2 d_{it-1}^{*}}}{1+e^{\gamma_0+\gamma_1+\gamma_2 d_{it-1}^{*}}} \\ \pi_{01t} &= \Pr\left(d_{it}^{*} = 1 | d_{it} = 0\right) = \frac{e^{\gamma_0+\gamma_2 d_{it-1}^{*}}}{1+e^{\gamma_0+\gamma_2 d_{it-1}^{*}}}. \end{aligned} \tag{11}$$

In the next section, we outline the SML estimation algorithm for any specification of the classification error process in (6), as well as for the two specific classification error processes (biased and unbiased) described above in (9) and (11).

4.3 Identification

Hausman, Abrevaya and Scott-Morton (1998) (HAS) discuss identification of discrete choice models with classification error. Note that the unconditional probability that outcome one is observed is:

$$\begin{aligned} P\left(d_{it}^{*} = 1\right) &= \pi_{11t} P\left(d_{it} = 1\right) + \pi_{01t} P\left(d_{it} = 0\right) \\ &= \left(1 - \pi_{10t}\right) P\left(d_{it} = 1\right) + \pi_{01t}\left(1 - P\left(d_{it} = 1\right)\right) \\ &= \pi_{01t} + \left(1 - \pi_{10t} - \pi_{01t}\right) P\left(d_{it} = 1\right). \end{aligned}$$

HAS point out that identification of a fully parametric discrete choice model given classification error requires (i) that the probability a choice is reported be monotonically increasing in the probability it is the true choice, and (ii) that the discrete choice model satisfies index sufficiency. Here, the monotonicity assumption is met if $\pi_{10t} + \pi_{01t} < 1$, which means that the probability of an observed "1" is increasing in the probability of a true "1". This basically means that classification error can't be so severe that people mis-report their state more often than they report correctly (certainly a mild requirement).[6] In (10) this is equivalent to $\gamma_1 > 0$.

Interestingly, in the model with unbiased classification error, we can use equation (9) to obtain $\pi_{10t} + \pi_{01t} = 1 - E$. Thus, identification requires that $E > 0$, which means that even very low probability events must have some positive probability of being classified correctly. To further clarify this point, note that, in equation (8), if $E = 0$ then the probability of observing choice "1" is simply $P(d_{it} = 1)$, *regardless* of whether the true choice is one or two. Hence, when $E = 0$, the probability of observing "1" is no greater when it is the true choice than when it is not.

5 The SML Estimation Algorithm

Suppose the data consist of $\{D_i^*, x_i\}_{i=1}^N$ where $D_i^* = \{d_{it}^*\}_{t=1}^T$ is the history of reported choices for individual i, $x_i = \{x_{it}\}_{t=1}^T$ is the history of the exogenous covariate for individual i, and N is the number of individuals in the sample. For ease of exposition, assume that $\{x_{it}\}_{t=1}^T$ is fully observed for each individual i, and that $t = 1$ is the first period of observed data. Since there may be missing choices during the sample period, let $I(d_{it}^* \in D_i^*)$ be an indicator equal to one if d_{it}^* is observed, and zero otherwise. Under these conditions, simulation of the likelihood function requires constructing M simulated choice histories for each $\{x_{it}\}_{t=1}^T$ history as follows:

1. For each individual i, draw M sequences of errors from the joint distribution of $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ to form $\left\{\left\{\{\varepsilon_{it}^m\}_{t=1}^T\right\}_{i=1}^N\right\}_{m=1}^M$.
2. Given $\left\{\{x_{it}\}_{t=1}^T\right\}_{i=1}^N$ and the error sequences $\left\{\left\{\{\varepsilon_{it}^m\}_{t=1}^T\right\}_{i=1}^N\right\}_{m=1}^M$, construct M simulated choice histories for each individual i $\left\{\left\{\{d_{it}^m\}_{t=1}^T\right\}_{i=1}^N\right\}_{m=1}^M$ according to (1) and the decision rule (2).

[6] HAS also note that extreme values of $X'\beta$ convey important information about error rates. No matter how large is $X'\beta$, the probability of an observed "1" cannot exceed $1-\pi_{10t}$. Similarly, no matter how small is $X'\beta$, the probability of an observed "0" cannot exceed $1-\pi_{01t}$.

3. Construct the conditional probabilities $\left\{\left\{\widehat{\pi}_{jkt}^{m}\right\}_{t=1}^{T}\right\}_{m=1}^{M}$ for each individual i, where j denotes the simulated choice and k denotes the reported choice. The procedure to do this depends on the assumed classification error process, as we discuss below in steps ($3a$) and ($3b$).

4. Form a simulator of the likelihood contribution for each individual i as:

$$\widehat{P}\left(D_i^*|\theta, x_i\right) = \frac{1}{M}\sum_{m=1}^{M}\prod_{t=1}^{T}\left(\sum_{j=0}^{1}\sum_{k=0}^{1}\widehat{\pi}_{jkt}^{m} I\left[d_{it}^{m} = j, d_{it}^{*} = k\right]\right)^{I\left(d_{it}^{*}\in D_i^*\right)} \tag{12}$$

where θ is the vector of model parameters. This simulator is unbiased.

Step ($3a$):

In the special case of unbiased classification error, the $\widehat{\pi}_{jkt}^{m}$'s in step (3) depend on the true choice probability $\Pr\left(d_{it} = 1\right)$ (see equation (9)). Therefore, $\Pr\left(d_{it} = 1\right)$ must also be simulated. $\Pr\left(d_{it} = 1\right)$ can be approximated by the unbiased simulator

$$\widehat{P}\left(d_{it} = 1|H_{it}^{m}\right) = \frac{1}{M}\sum_{m=1}^{M}\Pr\left(\varepsilon_{it} \leq \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}^{m}\rho_{\tau}\right) \tag{13}$$

where $H_{it}^{m} = \left\{\left\{x_{i\tau}\right\}_{\tau=1}^{t}, \left\{d_{i\tau}^{m}\right\}_{\tau=1}^{t-1}\right\}$ is the history of the exogenous covariate and the simulated lagged endogenous covariate through time t.[7]

Then $\widehat{\pi}_{11t}^{m}$ and $\widehat{\pi}_{01t}^{m}$ are:

$$\begin{aligned}\widehat{\pi}_{11t}^{m} &= \Pr\left(d_{it}^{*} = 1|d_{it}^{m} = 1\right) = E + (1-E)\widehat{P}\left(d_{it} = 1|H_{it}^{m}\right) \qquad (14)\\ \widehat{\pi}_{01t}^{m} &= \Pr\left(d_{it}^{*} = 1|d_{it}^{m} = 0\right) = (1-E)\widehat{P}\left(d_{it} = 1|H_{it}^{m}\right)\end{aligned}$$

Step ($3b$):

For the biased classification error process given by (11), the $\widehat{\pi}_{jkt}^{m}$'s in step (3) depend on the reported lagged choice $d_{i,t-1}^{*}$. If $d_{i,t-1}^{*}$ is missing, it must be simulated.

[7]If ε_{it} is distributed i.i.d. $N(0,\sigma_{\varepsilon}^{2})$, the probability in the summation is $\Phi\left(a\right)$ where $a = \beta' x/\sigma_{\varepsilon}$, $\beta' x = \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}^{m}\rho_{\tau}$, and Φ is the standard normal c.d.f.. If ε_{it} is serially correlated, then the probability in (13) must, of course, be conditional on $\left\{\left\{\varepsilon_{i\tau}\right\}_{\tau=0}^{t-1}\right\}$.

This is done using (10). Let the simulated $d_{i,t-1}^{*}$ be denoted $d_{i,t-1}^{*m}$, and let $d_{i,t-1}^{*(m)} = I\left(d_{i,t-1}^{*} \in D_i^{*}\right) d_{i,t-1}^{*} + \left(1 - I\left(d_{i,t-1}^{*} \in D_i^{*}\right)\right) d_{i,t-1}^{*m}$. Then $\widehat{\pi}_{11t}^{m}$ and $\widehat{\pi}_{01t}^{m}$ are:

$$\widehat{\pi}_{11t}^{m} = \frac{e^{\gamma_0+\gamma_1+\gamma_2 d_{it-1}^{*(m)}}}{1+e^{\gamma_0+\gamma_1+\gamma_2 d_{it-1}^{*(m)}}}, \widehat{\pi}_{01t}^{m} = \frac{e^{\gamma_0+\gamma_2 d_{it-1}^{*(m)}}}{1+e^{\gamma_0+\gamma_2 d_{it-1}^{*(m)}}} \tag{15}$$

The simulation algorithm described in steps (1) to (4) builds the likelihood contribution for each individual by averaging, over M simulated choice histories, the product of the appropriate classification probabilities $\left\{\widehat{\pi}_{jkt}^{m}\right\}_{t=1}^{T}$ needed to reconcile the simulated choice history $\{d_{it}^{m}\}_{t=1}^{T}$ and the observed history $\{d_{it}^{*}\}_{t=1}^{T}$. In step (4) the indicator $I\left[d_{it}^{m} = j, d_{it}^{*} = k\right]$ "picks out" the appropriate classification probability by comparing d_{it}^{*} to d_{it}^{m}. If d_{it}^{*} is unobserved, $I\left(d_{it}^{*} \in D_i^{*}\right)$ is zero, so there is no contribution to the likelihood (i.e., simply enter one in the product in period t).[8]

Note that any observed choice history has non-zero probability conditional on any simulated choice history. This reflects the fact that any simulated history can generate any observed history when there is classification error. It is also important to note that (12) builds the likelihood using *unconditional* simulations of the model. Simulation of conditional probabilities like $P\left(d_{it}|H_{it}\right)$ is completely avoided, circumventing the severe computational problems that may arise if H_{it} is not fully observed. In the unconditional approach, the state space is updated according to previous *simulated* choices, rather than previous *reported* choices, which greatly simplifies the problem.[9]

[8] If choices are not missing at random, the probability that the choice is not observed can be incorporated into the product. A similar correction can be made to handle endogenous attrition.

[9] It is important to understand when the GHK algorithm has problems in dynamic models. GHK builds up the likelihood of a choice history using period-by-period conditional simulations. In a panel probit with serial correlation but *no state dependence*, missing choices pose no problem. To simulate a choice probability at time t requires a draw for the lagged stochastic terms consistent with observed choices up through t-1. Thus, in periods when choices are missing, one simply draws from the *unconditional* distribution of the stochastic terms. But GHK runs into problems in three cases: (i) with state dependence one must condition on lagged *simulated choices* in periods when the actual choice is missing. As one iterates on the model parameters, a simulated choice may change, leading

The asymptotic properties of the SML estimator described here are the same as in Lee (1992) and Pakes and Pollard (1989). Consistency and asymptotic normality require $\frac{M}{\sqrt{N}} \to \infty$ as $N \to \infty$. Our estimator is just a special case of SML, differentiated from past approaches only by the algorithm used to simulate the likelihood. But the importance of this should not be underestimated. Past Monte Carlo work has shown that within the class of SML estimators that share common asymptotic properties, finite sample performance hinges critically on the quality of the algorithm used to simulate choice probabilities (see Geweke and Keane (2001) for a review).

5.1 Missing Covariates and Initial Conditions

The algorithm needs only slight modification to handle missing exogenous covariates and/or an initial conditions problem. In the case of missing covariates, each missing x_{it} is simulated according to the assumed process generating the x_{it}'s. For example, suppose the x_{it}'s are time-varying and stochastic and follow the $AR(1)$ process,

$$x_{it} = \phi_2 x_{i,t-1} + \nu_{it} \tag{16}$$

where $\nu_{it} \sim N(0, \sigma_v^2)$, and where $x_{i0} = 0$. If x_{it-1} is observed and x_{it} is missing, the missing x_{it} is replaced by $\widehat{x}_{it}^m$ which equals $\phi_2 x_{it-1}$ plus a draw from the ν_{it} distribution. A new draw for ν_{it} is taken for each simulated choice history m.

The likelihood contribution for each individual i in this case becomes

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

$$\frac{1}{M}\sum_{m=1}^{M}\prod_{t=1}^{T} f_m\left(x_{it}\right)^{I(x_{it}\in x_i)}\left(\sum_{j=0}^{1}\sum_{k=0}^{1}\widehat{\pi}_{jkt}^m I\left[d_{it}^m = j, d_{it}^* = k\right]\right)^{I\left(d_{it}^*\in D_i^*\right)} \tag{17}$$

---

to discontinuities in the simulated likelihood. A possible solution is to integrate over all possible missing choices, but this becomes infeasible as the number of periods with missing choices grows large; (ii) if, as in KW, there is more then one choice variable, and only a subset is observed, drawing from the conditional distribution of the stochastic terms given that subset can be very difficult; (iii) if choices are subject to classification error, then drawing stochastic terms from their conditional distribution given the (possibly misclassified) observed choice can be very difficult.

where $f_m(x_{it})$ is the density of the exogenous covariate under draw sequence m, and $I(x_{it} \in x_i)$ is an indicator for x_{it} being observed.

Specifically under the assumption that $\nu_{it} \sim N(0, \sigma_v^2)$, we have that:

$$f_m(x_{it}) = \frac{1}{\sigma_v}\phi\left(\frac{x_{it} - \phi_2 \widehat{x}_{it-1}^{(m)}}{\sigma_v}\right) \tag{18}$$

where $\widehat{x}_{it-1}^{(m)} = I(x_{i,t-1} \in x_i)\, x_{it-1} + (1 - I(x_{i,t-1} \in x_i))\, \widehat{x}_{it-1}^{m}$ and ϕ is the standard normal p.d.f.. Note that $f_m(x_{it})$ only enters the likelihood in periods when x_{it} is observed. ϕ_2 and σ_v now become part of the parameter vector θ.

In the case of an initial conditions problem, $t = 1$ is not the first period of observed data. Let $t = \widetilde{\tau}$ be the first period of observed data where $\widetilde{\tau} > 1$. Simulated choice histories are still constructed from the theoretical start of the process, i.e., from $t = 0$ with $d_{i0} = x_{i0} = 0$, irrespective of the value of $\widetilde{\tau}$. If the x_{it}'s are also missing, the path of x_{it}'s must be simulated from $t = 1$ until $t = \widetilde{\tau}$.[10]

The likelihood contribution for each individual i in this case takes the form
$\widehat{P}(D_i^*, x_i|\theta) =$

$$\frac{1}{M}\sum_{m=1}^{M}\prod_{t=\widetilde{\tau}}^{T} f_m(x_{it})^{I(x_{it}\in x_i)}\left(\sum_{j=0}^{1}\sum_{k=0}^{1}\widehat{\pi}_{jkt}^{m} I\left[d_{it}^{m} = j, d_{it}^{*} = k\right]\right)^{I\left(d_{it}^{*}\in D_i^{*}\right)} \tag{19}$$

The only difference between (17) and (19) is that here the first d_{it}^* is observed at $t = \widetilde{\tau}$. In Heckman's method, one specifies an approximate marginal distribution for $d_{i\widetilde{\tau}}^*$. In our method, this is not necessary. We integrate over the correct distribution for $d_{i\widetilde{\tau}}^*$ by simulating the choice and covariate history from $t = 1$ through $t = \widetilde{\tau} - 1$.

In some applications, the process has a natural start date (e.g., age 16 for decisions to stay in school or enter the labor force). In others, all that can be known reliably is that the process started well before the observation period. In that case, one may set

[10] If the first period of observed data is individual specific, simply replace $\widetilde{\tau}$ with $\widetilde{\tau}_i$. Note that if the model before $\widetilde{\tau}_i$ is different from the model after $\widetilde{\tau}_i$ (e.g., due to non-stationarity), one would simply simulate outcomes from the appropriate model.

$\widetilde{\tau}$ large enough so that estimates are not sensitive to further increases. Alternatively, one could nest Heckman's approximation method inside our algorithm, as a simple way to handle the initial period, while using our approach to handle missing data during the sample period. Such "hybrid" approaches will be considered below.

5.2 Importance Sampling

The simulated likelihood function in (19) is not smooth because, holding the draw sequence $\{\varepsilon_{it}^m\}_{t=1}^T$ fixed, a change in θ can induce discrete changes in the $\{d_{it}^m\}_{t=1}^T$ sequence. However, the algorithm can be modified to use importance sampling techniques to smooth the likelihood, allowing use of gradient methods of optimization.[11] We smooth the likelihood by first constructing simulated choice histories $\{d_{it}^m(\theta_0)\}_{t=1}^T$ at an initial θ_0. We hold these $\{d_{it}^m(\theta_0)\}_{t=1}^T$ sequences *fixed* as we vary θ. Each simulated choice sequence has an associated importance sampling weight, $W_m(\theta)$, that varies with θ. Then, as we change θ, sequences that are more (less) likely under the new θ receive increased (reduced) weight. Thus:

$$W_m(\theta) = \frac{\Pr\left(d_{i1}^m(\theta_0), ..., d_{iT}^m(\theta_0)|\theta, x_i\right)}{\Pr\left(d_{i1}^m(\theta_0), ..., d_{iT}^m(\theta_0)|\theta_0, x_i\right)} \tag{20}$$

where the numerator is the joint probability that simulated choice history m occurs given the current trial parameter vector θ, while the denominator is the joint probability that simulated choice history m occurs given the initial trial parameter vector θ_0. For example, the joint probability of simulated choice history m in the dynamic probit model with serially independent errors is simply:

$$\prod_{t=1}^{T} \Pr\left(\varepsilon_{it} \leq \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}^m \rho_\tau\right). \tag{21}$$

An alternative way to smooth the likelihood is as follows: Again we construct simulated choice histories $\{d_{it}^m(\theta_0)\}_{t=1}^T$ at the initial θ_0. But now we save the latent

[11] The non-smooth version of the estimation algorithm considered until now necessitates the use of (relatively slow) non-gradient methods of optimization such as the simplex method.

variable sequences $\{U_{it}^m(\theta_0)\}_{t=1}^T$ that generated them, where $U_{it}^m(\theta_0) = \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}^m \rho_\tau + \varepsilon_{it}$. We then hold both the $\{d_{it}^m(\theta_0)\}_{t=1}^T$ and $\{U_{it}^m(\theta_0)\}_{t=1}^T$ sequences fixed as θ varies. In this approach, each simulated choice sequence receives an importance sampling weight, $W_m(\theta)$, that takes the form,

$$W_m(\theta) = \frac{g\left(U_{i1}^m(\theta_0), ..., U_{iT}^m(\theta_0)|\theta, x_i\right)}{g\left(U_{i1}^m(\theta_0), ..., U_{iT}^m(\theta_0)|\theta_0, x_i\right)} \tag{22}$$

where $g(\cdot)$, the joint density of simulated latent variable sequence m, is the product of standardized $U_{it}^m(\theta_0)$ densities. For example, in the case of serially independent errors, the joint density of simulated choice history m in (22) is

$$g\left(U_i^m(\theta_0)|\theta, x_i\right) = \prod_{t=1}^T \frac{1}{\sigma_\varepsilon}\phi\left(\frac{1}{\sigma_\varepsilon}\left[U_{it}^m(\theta_0) - \beta_0 - \beta_1 x_{it} - \sum_{\tau=0}^{t-1} d_{i\tau}^m(\theta_0)\rho_\tau\right]\right) \tag{23}$$

where ϕ is the standard normal p.d.f.. The weights in (22) are easier to calculate than the weights in (20) in some contexts. In the repeated sampling experiments reported below, and in the empirical application, we use the weights in (22).

The likelihood contribution for agent i in the smooth version of the algorithm is $\widehat{P}(D_i^*, x_i|\theta) =$

$$\frac{1}{M}\sum_{m=1}^M W_m(\theta)\prod_{t=\widetilde{\tau}}^T f_m(x_{it})^{I(x_{it}\in x_i)}\left(\sum_{j=0}^1\sum_{k=0}^1 \widehat{\pi}_{jkt}^m I\left[d_{it}^m = j, d_{it}^* = k\right]\right)^{I\left(d_{it}^*\in D_i^*\right)} \tag{24}$$

Note that (19) is a special case of (24) with $W_m = 1$ for each simulated history m.[12]

An important computational advantage of the re-weighting scheme over the implicit equal weighting scheme in (19) is that it requires simulated choice histories to be generated only once for each individual, with an initial vector of trial parameters θ_0, as opposed to constructing simulated choice histories at each vector of trial parameters θ. KW used this smooth version of the algorithm to construct standard errors (with weights as in (20)), but used the non-smooth version in estimation (using a simplex algorithm). Ackerberg (2001) describes an analogous use of importance sampling and has a good discussion of how his approach differs from ours.

[12] Efficiency of importance sampling is often improved by normalizing weights to sum to one.

6 Monte-Carlo Tests - Unbiased Misclassification

Here we report Monte-Carlo tests of the SML estimator with unbiased classification error. Keane and Sauer (2009) describes the algorithm used to generate artificial data. Sections 6.1 and 6.2 present results for the random effects and $AR(1)$ specifications of the error term, respectively. In each experiment, a vector of true model parameters is chosen and used to create 50 Monte-Carlo data sets which differ in the realizations of the stochastic terms. Parameter estimates are then obtained for each data set.[13] For each experiment, we report the true parameters, as well as the mean, median, empirical standard deviation, and root mean square error of the estimates. t-statistics for the biases, based on the empirical standard deviations, are also reported.[14]

6.1 Random Effects Model

In the random effects model, the error term ε_{it} is given by (3). The true start of the process is $d_{i0} = 0$. The exogenous covariate x_{it} is generated by the $AR(1)$ process in (16). The depreciation weights ρ_τ are assumed to follow an exponential decay process, $\rho_\tau = \rho e^{-\alpha(t-\tau-1)}$. The parameter α captures the "speed" of depreciation. The vector of estimable parameters is thus $\theta = \{\beta_0, \beta_1, \rho, \alpha, \phi_2, \sigma_v, \sigma_\mu, E\}$. However, in the special case of no initial conditions problem and no missing exogenous covariates, ϕ_2 and σ_v need not be estimated. Identification conditions for this model (a generalized Polya process with decay) are discussed in Heckman (1981b).

Table 1 reports summary statistics, by time period, for a representative data set. The number of individuals N is set to 500, the number of periods T is set to 10, there are no missing choices or missing exogenous covariates, and the vector of true

[13] For each data set, we use a different seed for the random number generator that generates the M unconditional simulations for each individual in the sample.

[14] We do not compare true average partial effects to estimated average partial effects. The reason is that, in dynamic models, there are a multitude of average partial effects that could be calculated.

parameters is set at $\theta = \{-.10, 1.00, 1.00, .50, .25, .50, .80, .75\}$. To identify the scale of utility, the variance of ε_{it} is normalized to one, so $\sigma_\mu^2 + \sigma_\eta^2 = 1$. Thus, individual effect accounts for 64 percent of the variance in ε_{it} (as σ_μ is set to .80).

The Mean d_{it} column shows that, over time, an increasing proportion of individuals choose the first option (i.e., 48 percent at $t = 1$ vs. 85 percent at $t = 10$). The Mean d_{it}^* column shows that the proportion that report choosing the first option closely tracks the true proportion. This is a consequence of unbiased classification error. The Mean $\beta' x$ column displays the mean and variance of $\beta' x = \beta_0 + \beta_1 x_{it} + \rho \sum_{\tau=0}^{t-1} e^{-\alpha(t-\tau-1)} d_{i\tau}$. Over time, the mean of $\beta' x$ increases at a decreasing rate, reflecting both the increasing proportion of $d_{it} = 1$ over time and the relatively strong depreciation of past choices. The variance of $\beta' x$ is roughly comparable to that of ε_{it} by the third period.

The Mean π_{11t} and Mean π_{00t} columns present probabilities of correct classification.[15] The probability of a correct match of $d_{it} = 1$ and $d_{it}^* = 1$, π_{11t}, is .863 at $t = 1$, increasing to .956 at $t = 10$, while that of a correct match of $d_{it} = 0$ and $d_{it}^* = 0$, π_{00t}, is .887 at $t = 1$, decreasing to .794 at $t = 10$. This pattern emerges because π_{11t} is an increasing function of the proportion choosing $d_{it} = 1$, and π_{00t} is a decreasing function of the same proportion (see (9)). The base classification rate E is set to .75, so even low probability events have a fairly high chance of being classified correctly.

6.1.1 Non-Smooth SML Algorithm

Table 2 reports the results of four repeated sampling experiments using the non-smooth SML algorithm. The experiments differ in the proportion of randomly missing choices during the sample period. The four panels correspond to data generating processes (DGPs) with no missing choices, and 20%, 40% and 60% missing choices, respectively. There are no missing exogenous covariates. In estimation, we set the number of simulated choice histories per individual, M, equal to 1000. For starting values, each parameter is bumped 20% away from the true values.

[15] We use $\widetilde{M} = 1000$ to calculate the error rates. Hence they vary slightly due to simulation noise.

As we see in Table 2, the SML estimator produces biases, but they are negligible in magnitude. Bias in the estimate of ρ is statistically significant in all four panels; however, the magnitude of the bias never exceeds 5.1 percent. Biases in estimates of β_1 and E are sometimes significant but never exceed 2 percent. Mean and median parameter estimates are quite close, suggesting the sampling distributions are symmetric. The empirical standard errors of the estimates generally increase with increased incidence of missing choices; but the point estimates are little affected, as choices are missing at random. Thus, the effect of more missing choices is only to reduce the effective sample size (t-statistics for significant biases generally decrease because the biases are mostly unaffected and the empirical standard errors increase).

The biases in Table 2 also appear small given that biases on the order of 5-8% often arise even in panel data models estimated by classical maximum likelihood (see Heckman (1981*a*)). But the models in Table 2, even that in the first panel (i.e., no missing choices and no initial conditions problem), are very difficult to estimate by classical maximum likelihood. This is because conditional choice probabilities are hard to construct when only lagged *reported* choices are known and not lagged *true* choices. Missing choice data amplifies the problem.

The small sample biases in Table 2 do not appear to be due to simulation error. Doubling the number of simulated choice histories M to 2000 or reducing it to $M = 500$ has little affect on the results. Mean time to convergence over the 50 repetitions in the second panel of Table 2 (20% missing choices) was 3.73 hours with a standard deviation of .92. With $M = 500$ this falls to 1.46 hours with a standard deviation of .34. The experiments were run on a desktop with two 1.0 GHz processors.

Table 3 reports the results of three repeated sampling experiments analogous to those in Table 2, except for a modified DGP where the exogenous covariate is also missing when the choice is missing. The parameters of the exogenous covariate process, ϕ_1 and σ_v, are estimated jointly with the other model parameters. As the results in Table 3 illustrate, adding missing covariates does not change the general conclusions from Table 2. The bias in the estimate of ρ is statistically significant but is still negligible in magnitude. The maximum bias over all parameters is only 4.8%.

Table 4 reports the results of three repeated sampling experiments that focus on the initial conditions problem rather than missing information during the sample period. The number of periods in the first two experiments is increased to $T = 20$. The DGP is modified so that choices and covariates are completely missing in periods $t = 1, ..., 10$ but there are no missing choices or covariates from $t = 11, ..., 20$.

The first panel of Table 4 reports the results of simulating from $t = 0$, the theoretical start of the process, and forming the likelihood for periods $t = 11$ to $t = 20$ as in equation (19). Biases in the estimates of $\beta_1, \rho, \sigma_\nu$ and σ_μ are statistically significant, but negligible in magnitude (i.e., no more than 3 percent). Simulating choices from the theoretical start of the process works quite well.

The second panel of Table 4 reports the results of ignoring the initial conditions problem by assuming the choice process starts at $t = 10$ with $d_{i,10} = 0$.[16] The biases produced by this method are generally substantial. σ_μ in particular is badly biased upwards. The incorrect treatment of the initial condition results in a substantial overestimate of the importance of individual effects.[17]

In the third panel of Table 4 we treat the initial conditions problem by constructing a proxy for the initial value of $\sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau$. The number of periods in this experiment is increased to $T = 30$. The DGP is modified so choices and covariates are completely missing from $t = 1, ..., 10$ but are observed from $t = 11, ..., 30$. Observed choices in periods $t = 11, ..., 20$ are used to proxy for $\sum_{\tau=0}^{20} d_{i\tau}\rho_\tau$ and the likelihood is constructed using data from $t = 21, ..., 30$. The latent index at $t = 21$, u_{21}, is set to:

$$u_{i21} = \beta_0 + \beta_1 x_{i21} + \rho \sum_{\tau=11}^{20} e^{-\alpha(21-\tau-1)} d_{i\tau}^* + \varepsilon_{i21}. \qquad (25)$$

[16] Pre-sample covariates are also ignored, so the x_{it} process parameters ϕ_1 and σ_ν are not estimated.

[17] In fact, almost all of the estimates of σ_μ are near the upper bound of one, so the standard deviation over the fifty estimates is very small.

The biases produced by this method are generally substantial in magnitude. Similar to the case where the initial conditions problem was ignored, there is substantial upward bias in the estimated variance of the random effect. Also, the estimate of the base classification rate parameter E is severely biased downward.

Table 5 reports the results of four repeated sampling experiments where there is an initial conditions problem and the model has a more familiar first-order Markov structure. The Markov model is nested in the Polya model by setting $\alpha = 0$ and $\tau = t - 1$ so that $u_{it} = \beta' x_{it} + \rho d_{it-1} + \varepsilon_{it}$. The first panel of Table 5 reports the results of handling the initial conditions problem by simulating from $t = 0$ and forming the likelihood using data from periods $t = 10$ to $t = 20$, as in equation (19). Simulating choices from the theoretical start of the process works quite well in the Markov model. The resulting biases are small in magnitude, never exceeding 4.1%.

The second panel of Table 5 reports the results of ignoring the initial conditions problem in the Markov model by setting $d_{i9} = 0$. The estimate of ρ in this experiment is substantially biased downward and σ_μ is substantially biased upward. The incorrect treatment of the initial condition results in estimates that imply an overly weak effect of previous choices on current utility, and an overly strong individual effect.

The third panel of Table 5 reports results of treating the initial condition as exogenous (i.e., simply substituting the observed choice in period 10 into the utility function in period 11). The biases produced by this method are generally less severe than ignoring the initial conditions problem but, as might be expected when treating the initial condition as exogenous, the estimate of ρ is biased upwards (by 14%).

The fourth panel of Table 5 applies the Heckman (1981a) method of approximating the marginal probability of the initial state using a probit model that incorporates only information on exogenous covariates. This method specifies a different latent index function, u_{it}^H, in the first period of observed data. The latent index at $t = 10$ is

$$u_{it}^H = \gamma_0 + \gamma_1 x_{it} + \varepsilon_{it}^H \tag{26}$$

where the variance of ε_{it}^H is normalized to one and the correlation between ε_{it}^H and the individual effect μ_i is $\rho_{\mu\epsilon^H}$. The parameters γ_0, γ_1 and $\rho_{\mu\epsilon^H}$ are estimated jointly with the other parameters of the model. We still use our algorithm to accommodate classification error and form the likelihood using unconditional simulations from $t = 10, ..., 20$, but at $t = 10$ we simulate from (26) instead of (1). In effect, we nest Heckman's procedure for handling the initial period within our algorithm.

The results show that nesting the Heckman method in our procedure works relatively well in the random effects model. ρ is over-estimated by only 6.4%. Although biases are not substantial for Heckman's approximate solution method (except for the constant), simulation from the theoretical start of the process, when known, is clearly preferable, as the parameter estimates are less biased and more precise.

The fifth panel of Table 5 nests the Wooldridge (2003) approach to solving the initial conditions problem within our algorithm. The Wooldridge method models the conditional mean of the random effect as a function of the initial condition and the entire path of exogenous covariates.[18] Assuming the conditional mean is linear,

$$E\left[\mu_i | d_{i0}^*, x_{i11}, ..., x_{i20}\right] = \alpha_0 + \alpha_1 d_{i10}^* + \alpha_2 x_{i11} + \cdots + \alpha_{11} x_{i20}, \tag{27}$$

the latent index in period $t = 11, ..., 20$, is

$$u_{it}^W = \widetilde{\beta}_0 + \beta_1 x_{it} + \rho d_{it-1} + \alpha_1 d_{i10}^* + \alpha_2 x_{i11} + \cdots + \alpha_{11} x_{i20} + \eta_{it} \tag{28}$$

where $\widetilde{\beta}_0 = \beta_0 + \alpha_0$. Note that β_0 and α_0 cannot be separately identified. The additional parameters that are identified in this approach are α_1 through α_{11}.

The estimation results show that nesting Wooldridge's method within our algorithm produces an estimate of ρ that is biased *downward* by 12.6%. In contrast, Heckman's method yields an estimate of ρ that is biased *upward* by 6.4%. Wooldridge's approach also produces a more significant bias in the estimate of E.[19]

[18] Keane and Woplin (1997) used essentially the same approach in the context of a DP model.

[19] The conclusions from the experiments are not sensitive to the extent of unbiased classification error in the data generating process. Similar results were obtained for E, the base classification rate, set to .25 and .50. Lower values of E correspond to a greater extent of classification error.

An interesting question is how our algorithm performs if there is in fact no (or negligible) classification error in the data. This scenario is implausible in micro datasets (e.g., in our experience, even machine generated data like that from supermarket scanners contain error, as human factors can always creep in), but it may be more plausible in certain macro contexts (e.g., a cross country panel on sovereign defaults). If classification error is not present, the assumption it exists serves simply as a tool to guarantee a non-zero likelihood given a finite simulation size, analogous to McFadden's (1989) appending of extreme value errors onto the probit model to obtain a "kernel smoothed" frequency simulator of probit choice probabilities. As there, the extra source of error leads to bias in the simulator, which diminishes as the scale of the auxiliary error goes to zero. How this affects estimates is an empirical question.

To address this issue, Table 6 reports results of three repeated sampling experiments where the true DGP has no classification error (and no initial conditions problem). The three panels display results for the random effects Polya model with 20%, 40% and 60% missing choices and covariates in each period, respectively. The results show negligible biases that never exceed 5%. The mean estimate of E tends towards the upper bound of one, so the estimated extent of classification error is very small. As these results illustrate, our algorithm is useful as a way to handle difficult likelihood function simulations even when there is no classification error in the data.

6.1.2 The Smooth SML Algorithm (Importance Sampling)

The smooth version of the algorithm differs from the non-smooth version in that we simulate choice histories only *once* for each individual in the sample, at the *initial* vector of trial parameters. Rather than simulating new histories as we iterate on the model parameters, we apply importance sampling weights to the original histories. The smooth algorithm enables us to use gradient methods of optimization, as opposed to more time consuming non-gradient methods. We again set simulation size $M = 1000$ and use starting values that are bumped 20% away from the true values.

Table 7 reports results of three experiments that use the smooth SML algorithm, with the weights in (22). These are analogous to the experiments in Table 3 using the non-smooth algorithm. The three experiments differ in the proportion of missing choices and covariates during the sample period, and assume no initial conditions problem. Like Table 3, Table 7 reveals a few statistically significant biases, but they are trivial in magnitude. Also, the smooth algorithm is roughly twice as fast. Recall that the mean time to convergence over the 50 repetitions in the second panel of Table 2 (20% missing choices) was 3.73 hours with a standard deviation of .92. But that for the first panel of Table 6 is only 1.94 hours with a standard deviation of .97.

6.2 $AR(1)$ Error Model

In the $AR(1)$ error model, the error term ε_{it} follows the process in (4). The theoretical start of the process is again $d_{i0} = 0$. As in the random effects model, the exogenous covariate x_{it} is generated by the $AR(1)$ process in (16), the depreciation weights ρ_τ follow the exponential decay process, $\rho_\tau = \rho e^{-\alpha(t-\tau-1)}$, and the variance of ε_{it} is normalized to one. The parameter vector is now $\theta = \{\beta_0, \beta_1, \rho, \alpha, \phi_2, \sigma_v, \phi_1, E\}$.

Table 8 reports summary statistics of a representative data set for the Polya model with $AR(1)$ errors. The data is generated with $N = 500$, $T = 10$, no missing choices or covariates, and the true parameter vector $\theta = \{-.10, 1.00, 1.00, .50, .25, .50, .80, .75\}$. Note that the $AR(1)$ parameter of .80 implies considerable serial correlation. A comparison of Tables 1 and 8 shows that the summary statistics produced by the $AR(1)$ error model are quite similar to those produced by the random effects model.

6.2.1 Non-Smooth SML Algorithm

The order of the experiments on the $AR(1)$ model is similar to that for the random effects model. Tables 9-11 correspond to Tables 3-5. The three panels of Table 9 report results of increasing the incidence of missing choices and covariates during the sample period, assuming no initial conditions problem. As in the random effects models, the bias for ρ is generally significant but negligible in magnitude, never exceeding 4.6%. In general, biases and standard errors of the parameter estimates are smaller in the $AR(1)$ model than in the random effects model (compare Tables 3 and 9).

Table 10 examines different solutions to the initial conditions problem. The first panel shows that simulating choices from the theoretical start of the process works quite well in the $AR(1)$ model (just as it did in the random effects model). But the second panel, where the initial conditions problem is ignored (i.e., set $d_{i,10} = 0$), reveals serious biases. The $AR(1)$ parameter (ϕ_1) is substantially over-estimated (i.e., .92 vs. .80), and the biases for ρ and α are also very large. As ρ is biased down and α is biased up, the estimates understate the importance of lagged choices.

In the third panel of Table 10 (analogous to the third panel of Table 4) we use the pre-sample history to proxy for $\sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau$. In this case biases are smaller. There is modest upward bias for ρ and α, and modest downward bias for the $AR(1)$ parameter.

Table 11 examines different solutions to the initial conditions problem in the Markov model with $AR(1)$ errors. Again, using our method to simulate from the theoretical start of the process works well. As we see in panel 2, ignoring the initial conditions problem leads to substantial downward bias in ρ. On the other hand, treating the initial condition as exogenous (panel 3) or using the Heckman method (panel 4) both lead to serious upward biases for ρ, by 23% and 20%, respectively.[20]

6.2.2 The Smooth SML Algorithm (Importance Sampling)

Table 12 reports the results of estimating the Polya model with $AR(1)$ errors, missing exogenous covariates but no initial conditions problem, and using the smooth SML algorithm with the weights in (22). As in the random effects model (see Table 7), biases are negligible in magnitude. Again, estimation is much faster using the smooth algorithm. For example, while mean time to convergence for the 50 runs in the first

[20] The Wooldridge approach is not applied in the $AR(1)$ case because it was developed specifically for a random effects model, as shown in (27).

panel of Table 9 (20% missing choices) was 3.07 hours (standard deviation .71), that for the 50 runs in the first panel of Table 12 was 1.84 hours (standard deviation .72).

7 Monte-Carlo Tests - Biased Misclassification

This section presents Monte-Carlo tests of the SML estimator with *biased* classification error, as specified in (11). Keane and Sauer (2009) describes the algorithm used to generate artificial data. Sections 7.1 and 7.2 present results for Polya models with random effects and $AR(1)$ errors, respectively. In Section 7.3, we present results for the Polya model with *both* random effects and $AR(1)$ errors.

7.1 Random Effects Model

7.1.1 Non-Smooth SML Algorithm

The three panels of Table 13 report results using the non-smooth SML algorithm on Polya models with random effects and *biased* classification error. The vector of structural parameters is the same as in our previous random effects experiments. In all three panels, 20% of the choices and exogenous covariates are missing in each period and there is no initial conditions problem. The three experiments differ in the true parameters of the classification error process, γ_0, γ_1 and γ_2. Specifically, the conditional probabilities $\pi_{11t} = \Pr(d^*_{it} = 1 | d_{it} = 1)$ and $\pi_{01t} = \Pr(d^*_{it} = 1 | d_{it} = 0)$ are $(.97, .18)$, $(.95, .27)$ and $(.95, .50)$, in the first, second and third panels, respectively. This gives "low," "medium," and "high" levels of misclassification.

The results in Table 13 reveal relatively few statistically significant biases. Only estimates of ρ and σ_v are consistently biased, but the magnitudes of these biases are negligible (less than 3 percent). In general, the algorithm seems to perform very well, both in terms of uncovering the structural parameters and in terms of uncovering the parameters of the classification error process. Note that, as the extent of classification bias increases, it leads to larger empirical standard errors. This is as expected: with more classification error, the data contain less information about the true process.[21]

7.1.2 Smooth SML Algorithm

Table 14 reports results for the random effects Polya model with biased classification error, using the smooth SML algorithm with the weights in (22). As in the first panel of Table 13, 20% of choices and covariates are missing in each period, there is no initial conditions problem and there is a relatively low degree of bias in misclassification. The results reveal slightly larger biases than when using the non-smooth algorithm (compare with Table 13 panel 1), but the biases remain small. The largest are in the estimates of ρ and α, which are biased down by 5.4% and 8.8%, respectively.

7.2 $AR(1)$ Error Model

7.2.1 Non-Smooth SML Algorithm

The three panels in Table 15 repeat the experiments of Table 13, except here we apply the non-smooth algorithm to a Polya model with $AR(1)$ errors rather than random effects. The results tell a similar story. The biases are negligible in magnitude, rarely exceeding 3 percent, and the empirical standard errors grow with the extent of bias in the true classification error process. The algorithm again does a good job of uncovering the parameters of the classification error process.

7.2.2 Smooth SML Algorithm (Importance Sampling)

Table 16 reports a similar experiment to that in Table 14, except we apply the smooth algorithm to a Polya model with $AR(1)$ errors rather than random effects. Biases are once again negligible in magnitude and noticeably smaller than in the random effects specification. Estimates of ρ and α are biased by only 2.1% and 2.3%, respectively. Biases for the classification error parameters are significant but small in magnitude.

[21] Estimation times were similar for models with biased vs. unbiased classifcation error.

7.3 Random Effects and $AR(1)$ Errors

Finally, we consider a model with *both* random effects and $AR(1)$ errors. Here, ε_{it} follows the process in (5). The true σ_μ is set to .80 while the $AR(1)$ parameter ϕ_1 is set to .40. To conserve on space we report results only for the Polya model, and only using the smooth algorithm (with the weights in (22)). The results are reported in Table 17. As in Tables 14 and 16, there are 20% missing choices and covariates in each period and low classification error bias. Some biases are significant but they are small in magnitude. In particular, the biases in the estimates of ρ and α are only 2.2% and 6%, respectively. Recall that the biases in these parameters in the random effects only model are 5.4% and 8%, respectively (see Table 14), and in the $AR(1)$ errors only model they are are 2.1% and 2.3%, respectively (see Table 16). It is interesting that the algorithm seems to have little difficulty disentangling the various sources of persistence in the data $\{\rho, \alpha, \sigma_\mu, \phi_1\}$.[22]

8 Application to Female Labor Supply

In this section, we use our algorithm to estimate dynamic probit models of married women's employment status, using PSID data from 1994-2003. As respondents were not interviewed every year during the sample period, the data contain both missing choices (missing endogenous state variables) and missing covariates, in addition to an initial conditions problem. Thus, it would be extremely difficult to simulate the likelihood using alternative approaches. This is particular true in the Polya model.

[22]For instance, compared to the RE only model (Table 14), the increases in RMSE for ρ, α, and σ_μ are very modest when ϕ_1 is added. RMSEs are considerably greater than in the $AR(1)$ only model (Table 16), but that is not the result of having random effects plus $AR(1)$ errors. RMSEs for ρ, α, and σ_μ are already considerably larger in the models with RE (Table 14) than in models with $AR(1)$ errors (Table 16).

8.1 The Data

The data are drawn from the 2004 PSID, including both the core random sample and the nonrandom Survey of Economic Opportunity. Restricting the sample to 1994-2003 produces a panel of the same length as in the repeated sampling experiments. A serious missing data problem arises because the PSID switched from annual to biannual surveys after the 1997 wave. Hence, PSID families were not interviewed in 1998, 2000, and 2002.[23] Even in the seven years when labor force participation is reported, it is likely to be measured with error.[24]

We build a panel from the PSID that has $N = 1310$ women and $T = 10$ years. We include women who are between the ages of 18 and 60 in 1995, are continuously married during the period, and whose husbands were labor force participants in each of the seven actual survey years. These are typical sample selection criteria in the literature on female labor force participation (see, e.g., Hyslop (1999)).

Table 18 presents descriptive statistics for the estimation sample. The mean of the dummy variable for employment status is .82. The remaining variables are covariates we use to predict employment. These are average annual husband's earnings (the proxy for nonlabor income), the number of children aged 0-2, 3-5 and 6-17, age,

[23] Respondents were asked a series of questions related to their activities in the "off-years" of the PSID. However, we treat retrospective responses as missing. There is *no* retrospective information collected on husband's annual earnings (non-labor income).

[24] For example, Poterba and Summers (1986) used the "CPS reconciliation data" to assess the extent of classification error in employment in the CPS. In these data, Census re-interviews a household a week after its original interview. In the event of disagreement about employment status, they try to assess true status. Figures in Poterba and Summers Table II imply the probability an employed person reports being unemployed or out-of-the-labor-force is 1.5%, while the probability an unemployed person reports being employed is 4.0%. Unfortunately, there is no direct evidence on classification error in the PSID. The PSID validation study (see Bound et al. (1994)) only covered respondents who worked for a single large firm. As all subjects were employed, the data cannot be used to assess the probability of misreporting employment when ones true state is unemployed.

highest grade completed over the sample period (which is then held constant from 1994-2003), and race (equal to one if black). All covariates except nonlabor income are available for the full ten years because they are either not time-varying (education, race), vary in a known way (age) or can be re-constructed from information in the 2003 panel (e.g., the fertility variables). We assume an $AR(1)$ process for the (sometimes) missing time-varying covariate (nonlabor income).

8.2 The Model

We fit two models (Markov and Polya) to married women's employment decisions:

$$\begin{aligned}
\text{Markov} &: \quad u_{it} = \beta_0 + \beta_1 \ln(y_{it}) + \beta_2' X_{it} + \rho d_{i,t-1} + \varepsilon_{it} \\
\text{Polya} &: \quad u_{it} = \beta_0 + \beta_1 \ln(y_{it}) + \beta_2' X_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau + \varepsilon_{it}, \;\; \rho_\tau = \rho e^{-\alpha(t-\tau-1)} \\
d_{it} &= 1 \text{ if } u_{it} \geq 0, \;\; 0 \text{ otherwise}, \; d_{i0} = 0 \\
\ln(y_{it}) &= \phi_2 \ln(y_{i,t-1}) + \nu_{it}, \;\; \nu_{it} \sim N\left(0, \sigma_\nu^2\right) \\
\varepsilon_{it} &= \mu_i + \xi_{it} \\
\xi_{it} &= \phi_1 \xi_{it-1} + \eta_{it}, \;\; \eta_{it} \sim N(0, (1-\sigma_\mu^2)(1-\phi_1^2)) \\
l_{it} &= \gamma_0 + \gamma_1 d_{it} + \gamma_2 d_{it-1}^* + \omega_{it} \\
\mu_i &= \sum_{t=1}^{T} \delta_t' W_{it} + \sigma_\mu \zeta_i, \;\; \zeta_i \sim N(0,1)
\end{aligned} \tag{29}$$

where y_{it} is the husband's earnings in year t, and X_{it} is a vector containing the fertility, race and education covariates, as well as year effects. The error structure for both the Markov and Polya models is random effects plus $AR(1)$ errors. We assume that the probability of reporting a particular employment state depends on true employment status as well as lagged reported status (to allow for persistence in misreporting).

The model in (29) is more general than those in the repeated sampling experiments because we allow for correlated random effects (see the last line in (29)) or "CRE". That is, following Chamberlain (1982, 1984), the random effects are allowed to be correlated with the vector W_{it} which contains $\ln(y_{it})$ and the three fertility variables.

A test of the null hypothesis H_0: $\delta_t = 0$ is a test for whether fertility and nonlabor income are exogenous in the sense they are uncorrelated with the individual effects.[25]

To estimate (29), we assume the choice process begins at age 16. Of course, most women in the sample are not observed until later. To deal with this initial conditions problem, we simulate employment and nonlabor income from age 16 onward. We estimate the model using the smooth algorithm with the importance sampling weights in (22). The number of simulated histories for for each person, M, is set to 250.[26]

8.3 Estimation Results

8.3.1 The Markov Model

Table 19 reports results for four different versions of the Markov model in (29). Column (1) reports point estimates and asymptotic standard errors for a restricted version with random effects only (i.e., $\phi_1 = 0$ and $\delta_t = 0$). The results show precisely measured effects of nonlabor income, fertility, age, race and education. And the signs and relative magnitudes of the effects are all in the expected directions.

The estimate $\widehat{\rho} = 2.31$ in Column (1) implies strong positive state dependence in employment status. Permanent unobserved heterogeneity is also an important source of persistence. The estimate $\widehat{\sigma}_\mu = .89$ implies that 89% of the total error variance is due to the individual effect. The $AR(1)$ coefficient in the nonlabor income process, $\widehat{\phi}_2 = .999$, implies husband's income is essentially a random walk.

[25] Note that consistency (in N with T fixed) of the conventional RE model requires strict exogeneity of the covariates. The CRE model relaxes this by letting μ_i be correlated with the time-varying covariates W_{it} for all $t = 1, ..., T$. However, it still imposes that time-varying covariates W_{it} are uncorrelated with lagged values of the time-varying errors ξ_{it}. This would be violated if, for instance, current shocks to employment status affect future utility.

[26] Setting M as high as 1000, as in the repeated sampling experiments, is computationally burdensome with N=1310. Thus, we performed additional experiments with N=1310 and M=250. We found that biases remain negligible, although standard errors are higher.

The estimates of the classification error process, γ_0, γ_1 and γ_2, imply classification error is important, and that there is considerable persistence in misclassification. The estimates imply that $\widehat{\pi}_{01t} = .299$ and $\widehat{\pi}_{10t} = .073$ when $d^*_{i,t-1} = 0$, and $\widehat{\pi}_{01t} = .677$ and $\widehat{\pi}_{10t} = .016$ when $d^*_{i,t-1} = 1$. Thus, the probability of mis-reporting a one (employment) when the true state is zero (not employed) increases from 29.9% to 67.7% if employment was reported in the previous period. Similarly, if employment is reported in the previous period, the probability of mis-reporting non-employment when the true state is employment falls from 7.3% to 1.6%. On average, the probability of misreporting one's state is 2.01%. Note that this is in the ballpark of the figures obtained by Poterba and Summers (1986) for the CPS. Finally, the χ^2 goodness of fit statistic has a p-value of .1024, so the model is not rejected at the 10% level.[27]

Column (2) reports results for the correlated random effects version of the model. The CRE model produces qualitatively similar estimates to those obtained in Column (1). However, the log-likelihood improves 22 points leading to rejection of H_0: $\delta_t = 0$. Specifically, the χ^2 likelihood ratio statistic is 44.58 with 27 degrees of freedom, giving a p-value of .0243. Thus, we find clear evidence that fertility and nonlabor income are not exogenous in a random effects probit model with first order state dependence.

The models in Columns (3) and (4) expand those in Columns (1) and (2) by adding $AR(1)$ transitory errors, introducing the additional parameter ϕ_1. In Column (3), the point estimate of $\widehat{\phi}_1$ is .608, and it is precisely estimated, suggesting that $AR(1)$ serial correlation is a significant source of persistence in married women's employment status. However, estimates of other parameters are little affected, and the likelihood improvement is modest. Column (4) reports results for the CRE version of the Markov model with random effects and $AR(1)$ errors. Results are qualitatively similar to those in Column (3). However, the log-likelihood improves by 31 points

[27]The Pearson chi-squared statistic is calculated by comparing actual and predicted frequencies of employment paths over the seven years of observed choices in the panel. To avoid small cell problems, we reduce the number of paths from $(2^7) = 128$ to 48 by combining "similar" paths.

leading to clear rejection of the null hypothesis $\delta_t = 0$ (i.e., the likelihood ratio χ^2 has a p-value of .0002). Thus we again reject exogeneity of fertility and non-labor income. The χ^2 goodness of fit statistic for this model is 56.40 with a p-value of .1637.

8.3.2 The Polya Model

Table 20 presents results for four versions of the Polya model (corresponding to the four versions of the Markov model in Table 19). If we look at the most general model in Column (4), we see that the Polya model implies much greater effects of husband income, young children and education on female labor supply than does the Markov model. It also implies a much smaller effect of race. Interestingly, the estimated variance of the random effect is similar in the Markov and Polya models, but the $AR(1)$ parameter is somewhat smaller in the Polya model (i.e., .46 vs. .61).

The Polya process estimates, $\widehat{\rho}$ and $\widehat{\alpha}$, imply that past employment is an important determinant of current employment, but that the influence of past choices falls quickly over time. For example, in Column (1), u_{it} increases by .6363 (the point estimate of ρ) when $d_{i,t-1} = 1$, holding all else constant. This is in contrast to an increase in u_{it} of 2.3 to 2.5 in the Markov models. In the Markov model $d_{i,t-2} = 1$ has no effect on u_{it}, while in the Polya model, setting $d_{i,t-2} = 1$ increases u_{it} by .0959. Moving further into the past, u_{it} increases by only .0145 when $d_{i,t-3} = 1$. Further lags have negligible effects. The sum of the lag coefficients is .75. Thus, the degree of true state dependence implied by the Polya models is much less than that implied by the Markov models. Instead, the Polya models ascribe more of the persistence in choices to observable heterogeneity (husband income, young children, education).

The Polya models imply only slightly lower classification error rates than the Markov models. For example, in Column (1), the estimated classification error rates in the Polya model are $\widehat{\pi}_{01t} = .246$ and $\widehat{\pi}_{10t} = .059$ when $d^*_{i,t-1} = 0$, and $\widehat{\pi}_{01t} = .630$ and $\widehat{\pi}_{10t} = .012$ when $d^*_{i,t-1} = 1$. The average probability of misreporting is 1.89% compared to 2.01% in Table 19, Column (1).

The Polya models fit the data noticeably better than the Markov models. For example, comparing the full models in Columns (4) of Tables 19-20, the improvement in the log-likelihood is 105 points with the addition of only one parameter (α).[28] Also, the Pearson chi-squared statistic is 51.02 with a p-value of .3186, compared to 56.40 with a p-value of .1637 in the Markov model.

Finally, the null hypothesis of exogenous fertility and nonlabor income is once again rejected in the Polya models. In the model with only random effects(Column (2)), the χ^2 statistic for H_0: $\delta_t = 0$ is 46.42 with a p-value of .0158. In the model that adds $AR(1)$ errors (Column (4)) it is 59.62 with a p-value of .0005.

Our findings contrast with those of Hyslop (1999), who cannot reject exogeneity of fertility and husband's income in models very similar to our Markov model. Our PSID sample differs from his because, using our estimation algorithm, we are able to depart from having a balanced panel and include women with missing data. But, as we show in Keane and Sauer (2006), the discrepancy in results is mainly due to the fact that we correct for classification error within our SML estimation algorithm. Allowing for classification error leads to an inference that there is more true persistence in labor supply choices (since our model interprets some transitions as spurious - arising due to misclassification of employment state). This, in turns, leads to estimates that imply a greater importance of individual effects, and, in turn, greater covariance of the individual effects with fertility and husband's income.

9 Conclusion

This paper assesses the performance of a new computationally practical SML estimation algorithm for dynamic discrete choice panel data models with unobserved endogenous state variables. The estimation technique offers a unified approach to the initial conditions problem and the problem of missing data during the sample

[28]Note that the Markov model is nested in the Polya model by setting $\alpha = 0$.

period. The computational advantage of the estimation algorithm lies in the fact that it requires only *unconditional* simulation of data from the model to form the likelihood. Performing unconditional simulations is often straightforward in contexts where performing conditional simulations is computationally infeasible. Therefore, in such contexts, our algorithm may have a significant advantage over algorithms such as GHK, MCMC and EM that require conditional simulation.

In order to make it feasible to simulate the likelihood using unconditional simulations, a classification error process in discrete choices must be assumed. However, the assumption that reported choices are misclassified is a reasonable one in almost all empirical applications in economics. The estimation technique can also accommodate a wide range of classification error processes, as long as it is possible to write a tractable expression for the classification error rates. The extent of classification error in the data can be estimated jointly with the structural model parameters, or, if good prior information is available, specified a priori.

The SML estimation algorithm was tested via a series of repeated sampling experiments on a panel data probit model with a time-varying exogenous covariate, lagged endogenous variables, serially correlated errors, and two different classification error processes (i.e., biased and unbiased). The estimator was shown to have good small sample properties. That is, we found that biases are generally negligible in magnitude even for high amounts of missing information in the data. We also found that the algorithm did a good job of uncovering both structural parameters of the model and the parameters of the classification error process.

The algorithm also performed well in experiments where the data are not subject to classification error. In these cases the assumed misclassification process is merely a device for smoothing the simulated likelihood.

The new SML estimation algorithm can easily handle initial conditions problems via integration from the theoretical start of the decision process. Or it can be combined with either Heckman's (1981*a*) or Wooldridge's (2003) approximate solution methods for the initial conditions problem. Such a hybrid approach may be appealing when there is no natural starting point to the process, but one still wishes to use our method to deal with missing data *during* the sample period. Heckman's method was found to work better than Wooldridge's in our experiments with a random effects model. But, Heckman's method worked less well in our experiments with an $AR(1)$ error model (i.e., we found a 20% upward bias in the coefficient on the lagged choice). It is preferable to simulate choices from the theoretical start of the process if possible.

Interestingly, our SML algorithm seems to perform a bit better (in terms of consistently producing negligible bias) for models with biased as opposed to unbiased classification error. In order to impose the constraint that classification error be unbiased, one must specify that error rates are functions of true choice probabilities. This means error rates must themselves be simulated, inducing additional noise into the likelihood simulation as well as additional computation time. In contrast, with biased classification error, one can specify that error rates are closed form functions of true choices (and perhaps also lagged observed choices and covariates), avoiding one component of simulation error and computation time.

We also apply the algorithm to panel data probit models of married women's employment status using PSID data from 1994-2003. A serious missing data problem arises in these data because (i) respondents were not interviewed in 1998, 2000 and 2002, (ii) there is nonresponse in interview years, and (iii) the average age at which women are first observed is 37, creating an initial conditions problem. We solve the initial conditions problem by simulating employment status and nonlabor income realizations from the theoretical start of the process, assumed to be age 16. We estimate both Markov and Polya models assuming biased classification error.

The utility of the algorithm was revealed in two ways: First, we found that the Polya model, which is more difficult to estimate using conventional methods than the much more commonly used Markov model (since missing data creates greater problems), provides a substantially better fit to the data. It also leads to substantially different economic results - i.e., state dependence is far less important as a source of persistence in labor supply, while observed heterogeneity is more important.

Second, the ability to accommodate classification error enables the algorithm to adjust for the impact of spurious transitions on the estimated degree of persistence in true choices. This implies greater importance of individual random effects, and higher covariance of these with observed characteristics. As a result, in contrast to results in Hyslop (1999), we find strong evidence that husband's income and fertility are endogenous in dynamic probit models of women's labor force participation.

Future research will examine the small sample properties of the estimation technique in more complex dynamic models. For example, observed continuous outcomes, such as wages, can be incorporated into estimation by specifying measurement error densities that enter the likelihood. The estimation method can also be extended to handle cases in which the missing data are not missing at random, there is endogenous attrition, or there is feedback from past choices to future covariates.

Michael P. Keane, University of Technology Sydney and Arizona State University, and Robert M. Sauer, University of Bristol.

References

[1] Ackerberg, D., “A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation,” Department of Economics, UCLA, June 2001.

[2] Bound, John; Brown, Charles; Duncan, Greg J. and Rodgers, Willard, “Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data,” Journal of Labor Economics, 12 (1994), 345-368.

[3] Chamberlain, G., "Multivariate Regression Models for Panel Data," Journal of Econometrics, 18 (1982), 5-46.

[4] Chamberlain, G., “Panel Data," in Z. Griliches and M.D. Intrilligator, eds., Handbook of Econometrics, Volume 2, (Amsterdam: Elsevier Science Publishers B.V., 1984), 1247-1318.

[5] Dempster, A.P., N.M. Laird and D.B. Rubin, “Maximum Likelihood From Incomplete Data via the EM Algorithm,” Journal of the Royal Statistical Society, B39 (1977), 1-38.

[6] Erdem, T. and M.P. Keane, “Decision Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets,” Marketing Science, 15 (1996), 1-20.

[7] Flinn, C.J., “Equilibrium Wage and Dismissal Processes,” Journal of Business and Economic Statistics, 15 (1997), 221-236.

[8] Geweke, J. and M.P. Keane, “An Empirical Analysis of Male Income Dynamics in the PSID: 1968-1989,” Journal of Econometrics, 96 (2000), 293-356.

[9] Geweke, J. and M.P. Keane, “Computationally Intensive Methods for Integration in Econometrics,” in J.J. Heckman and E. Leamer, eds., Handbook of Econometrics, Volume 5, (Amsterdam: Elsevier Science Publishers B.V., 2001), 3463-3568.

[10] Hausman, J.A., J. Abrevaya and F.M. Scott-Morton, "Misclassification of the Dependent Variable in a Discrete-Response Setting," Journal of Econometrics, 87 (1998), 239-269.

[11] Heckman, J.J., "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process," in C. Manski and D. McFadden, eds., The Structural Analysis of Discrete Data (Cambridge: MIT Press, 1981a), 179-195.

[12] Heckman, J.J., "Statistical Models for Discrete Panel Data," in C. Manski and D. McFadden, eds., The Structural Analysis of Discrete Data (Cambridge: MIT Press, 1981b), 114-178.

[13] Hyslop, D.R., "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," Econometrica, 67 (1999), 1255-1294.

[14] Keane, M.P., "A Computationally Practical Simulation Estimator for Panel Data," Econometrica, 62 (1994), 95-116.

[15] Keane, M.P., "Modeling Heterogeneity and State Dependence in Consumer Choice Behavior," Journal of Business and Economic Statistics, 15 (1997), 310-327.

[16] Keane, M.P. and R.M. Sauer, "Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior," IZA Working Paper No. 2332, September 2006.

[17] Keane, M.P. and R.M. Sauer, "A Computationally Practical Simulation Estimation Algorithm for Dynamic Panel Data Models with Unobserved Endogenous State Variables," IZA Working Paper No. 4054, March 2009.

[18] Keane, M.P. and K.I. Wolpin, "The Career Decisions of Young Men," Journal of Political Economy, 105 (1997), 473-522.

[19] Keane, M.P. and K.I. Wolpin, "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," International Economic Review, 42 (2001), 1051-1103.

[20] Lee, L.F, "On the Efficiency of Methods of Simulated Moments and Maximum Simulated Likelihood Estimation of Discrete Response Models," Econometric Theory, 8 (1992), 518-522.

[21] McFadden, D., "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," Econometrica, 57 (1989), 995-1026.

[22] Pakes A. and D. Pollard, "Simulation and the Asymptotics of Optimization Estimators," Econometrica, 57 (1989), 1027-1057.

[23] Poterba J.M. and L.H. Summers, "Reporting Errors and Labor Market Dynamics," Econometrica, 54 (1986), 1319-1338.

[24] Poterba J.M. and L.H. Summers, "Unemployment Benefits and Labor Market Transitions: A Multinomial Logit Model with Errors in Classification," Review of Economics and Statistics, 77 (1995), 207-216.

[25] Ruud, P.A., "Extensions of Estimation Methods Using the EM Algorithm," Journal of Econometrics, 49 (1991), 305-341.

[26] Wooldridge, J.M., "Simple Solutions to the Initial Conditions Problem for Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," Journal of Applied Econometrics, 20, (2005), 39-54.

Table 1

Summary Statistics
Representative Data Set
Random Effects Polya Model
Unbiased Classification Error

t	Mean d_{it}	Mean d^*_{it}	Mean $\beta' x$	Mean ε_{it}	Mean π_{11t}	Mean π_{00t}	N
1	.4800	.4800	-.0124	.0094	.8630	.8870	500
			(.2701)	(1.0147)			
2	.5780	.5780	.4909	.0149	.8947	.8553	500
			(.5601)	(1.0046)			
3	.6560	.6660	.8940	-.0116	.9142	.8359	500
			(.8547)	(.9919)			
4	.7140	.7260	1.1917	-.0005	.9264	.8236	500
			(1.0645)	(1.0102)			
5	.7460	.7440	1.4164	-.0232	.9347	.8153	500
			(1.1355)	(.9606)			
6	.7640	.7580	1.6214	-.0089	.9414	.8086	500
			(1.2164)	(1.0396)			
7	.8140	.8000	1.7812	-.0325	.9474	.8026	500
			(1.1329)	(1.020)			
8	.8120	.8100	1.8797	.0138	.9509	.7991	500
			(1.2081)	(1.0405)			
9	.8220	.8100	1.9806	.0092	.9545	.7955	500
			(1.1668)	(1.0107)			
10	.8460	.8500	1.9863	.0211	.9565	.7935	500
			(1.0949)	(.9539)			

Note: d_{it} is the true choice, d^*_{it} is the reported choice, π_{11t} and π_{00t} are the probabilities of a correct classification, and $\beta' x = u_{it} - \beta_0$. Variances are in parentheses. The frequency simulator that is used to compute the true classification error rates has $\widetilde{M}$ set to 1000. The model is:

$$\begin{aligned} u_{it} &= \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau + \varepsilon_{it} \\ d_{i0} &= 0,\ \rho_\tau = \rho e^{-\alpha(t-\tau-1)} \\ x_{it} &= \phi_2 x_{i,t-1} + \nu_{it},\ \nu_{it} \sim N\left(0, \sigma^2_\nu\right) \\ \varepsilon_{it} &= \mu_i + \eta_{it},\ \mu_i \sim N\left(0, \sigma^2_\mu\right),\ \eta_{it} \sim N\left(0, 1 - \sigma^2_\mu\right). \end{aligned}$$

Table 2

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

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
No Missing Choices ($t = 1, ..., 10$)						
β_0	-.1000	-.0975	-.0950	.0427	.0427	.42
β_1	1.0000	1.0171	1.0196	.0552	.0578	2.20
ρ	1.0000	1.0463	1.0462	.0513	.0691	6.38
α	.5000	.4912	.4926	.0499	.0506	-1.22
σ_μ	.8000	.8062	.8009	.0269	.0276	1.62
E	.7500	.7408	.7417	.0162	.0186	-3.99
20% Missing Choices ($t = 1, ..., 10$)						
β_0	-.1000	-.0995	-.1017	.0428	.0428	.08
β_1	1.0000	1.0114	1.0199	.0611	.0622	1.32
ρ	1.0000	1.0450	1.0356	.0528	.0694	6.04
α	.5000	.4864	.4985	.0719	.0731	-1.34
σ_μ	.8000	.8095	.8066	.0259	.0275	2.59
E	.7500	.7409	.7399	.0184	.0206	-3.50
40% Missing Choices ($t = 1, ..., 10$)						
β_0	-.1000	-.1025	-.1001	.0530	.0530	-.33
β_1	1.0000	1.0183	1.0265	.0612	.0648	2.09
ρ	1.0000	1.0505	1.0425	.0524	.0728	6.81
α	.5000	.4887	.4882	.0633	.0643	-1.26
σ_μ	.8000	.8047	.7989	.0339	.0343	.98
E	.7500	.7437	.7412	.0231	.0239	-1.94
60% Missing Choices ($t = 1, ..., 10$)						
β_0	-.1000	-.1070	-.1052	.0596	.0600	-.82
β_1	1.0000	1.0147	1.0161	.0860	.0872	1.21
ρ	1.0000	1.0485	1.0562	.0603	.0773	5.68
α	.5000	.4970	.4982	.0817	.0817	-.26
σ_μ	.8000	.8016	.8012	.0486	.0487	.23
E	.7500	.7477	.7426	.0287	.0288	-.55

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 1.

Table 3

Repeated Sampling Experiments
Random Effects Polya Model
Unbiased Classification Error
(Missing X's, 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 ($t = 1, ..., 10$)						
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_ν	.5000	.5015	.5016	.0057	.0059	1.86
σ_μ	.8000	.8056	.8017	.0287	.0292	1.38
E	.7500	.7428	.7430	.0172	.0187	-2.95
40% Missing Choices and X's ($t = 1, ..., 10$)						
β_0	-.1000	-.1087	-.1099	.0539	.0546	-1.15
β_1	1.0000	1.0141	1.0233	.0678	.0692	1.48
ρ	1.0000	1.0458	1.0374	.0636	.0784	5.10
α	.5000	.4953	.4949	.0600	.0602	.56
ϕ_2	.2500	.2521	.2546	.0253	.0254	.59
σ_ν	.5000	.5012	.5012	.0069	.0070	1.21
σ_μ	.8000	.8046	.8063	.0347	.0350	.94
E	.7500	.7474	.7416	.0245	.0246	-.74
60% Missing Choices and X's ($t = 1, ..., 10$)						
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_ν	.5000	.5011	.5017	.0089	.0090	.88
σ_μ	.8000	.8096	.8044	.0421	.0432	1.61
E	.7500	.7493	.7440	.0288	.0288	-.16

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 1.

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 $d_{i0} = 0$ $(t = 11, ..., 20)$						
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_ν	.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 $d_{i,10} = 0$ $(t = 11, ..., 20)$						
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
Use reported data from $t = 11, ..., 20$ to proxy for initial condition at $t = 21$ $(t = 11, ..., 30)$						
β_0	-.1000	-.5239	-.4859	.3039	.5216	-9.86
β_1	1.0000	.4742	.4671	.1788	.5553	-20.80
ρ	1.0000	1.0522	1.1064	.3076	.3120	1.20
α	.5000	.5839	.6139	.2299	.2448	2.58
σ_μ	.8000	.9388	.9758	.0811	.1608	12.10
E	.7500	.5795	.5714	.0615	.1812	-19.61

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 1.

Table 5

Repeated Sampling Experiments
Random Effects Markov 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 $d_{i0} = 0$ $(t = 10, ..., 20)$						
β_0	-.1000	-.1127	-.1086	.0391	.0411	-2.30
β_1	1.0000	1.0379	1.0364	.0324	.0500	8.25
ρ	1.0000	1.0330	1.0319	.0386	.0508	6.04
ϕ_2	.2500	.2496	.2511	.0136	.0136	-.19
σ_ν	.5000	.5014	.5011	.0045	.4986	2.17
σ_μ	.8000	.8137	.8133	.0294	.0324	3.29
E	.7500	.7293	.7294	.0150	.0256	-9.75
Assume process starts with $d_{i9} = 0$ $(t = 10, ..., 20)$						
β_0	-.1000	.1598	.1594	.0775	.2712	23.70
β_1	1.0000	.9126	.9171	.0693	.1115	-8.92
ρ	1.0000	.6396	.6171	.1025	.3747	-24.87
σ_μ	.8000	.8823	.8948	.0369	.0902	15.80
E	.7500	.7218	.7226	.0222	.0395	-8.99
Treat $d_{i,10}$ as exogenous						
β_0	-.1000	-.1882	-.1867	.0771	.1171	-8.09
β_1	1.0000	1.0328	1.0480	.0595	.0679	3.90
ρ	1.0000	1.1369	1.1465	.1024	.1710	9.45
σ_μ	.8000	.7838	.7843	.0460	.0488	-2.49
E	.7500	.7240	.7262	.0233	.0349	-7.91

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The Markov model replaces $\sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau$ in Table 1 with $\rho d_{i,t-1}$.

Table 5 (continued)

Repeated Sampling Experiments
Random Effects Markov 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
Use Heckman's approximation method to proxy for initial condition at $t = 10$ ($t = 10, ..., 20$)						
β_0	-.1000	-.1721	-.1705	.0728	.1025	-7.01
β_1	1.0000	.9868	.9831	.0616	.0630	-1.52
ρ	1.0000	1.0637	1.0673	.1074	.1249	4.20
σ_μ	.8000	.7735	.7767	.0472	.0542	-3.97
E	.7500	.7438	.7456	.0181	.0191	-2.44
γ_0		.3819	.3843	.0757		
γ_1		.6857	.6799	.1008		
$\rho_{\mu\epsilon^H}$		.6565	.6589	.0627		
Use Wooldridge's method of conditioning the distribution of the unobserved effect ($t = 11, ..., 20$)						
$\widetilde{\beta}_0$	-.1000	-.3276	-.3045	.0872	.2438	-18.46
β_1	1.0000	.9520	.9611	.0628	.0790	-5.40
ρ	1.0000	.8734	.8741	.0712	.1453	-12.57
σ_μ	.8000	.8034	.7988	.0478	.0479	.50
E	.7500	.7046	.7064	.0308	.0549	-10.43
α_1		.4522	.4314	.1124		
α_2		-.0137	-.0132	.0700		
α_3		-.0055	.0009	.0741		
α_4		.0162	.0234	.0761		
α_5		.0124	.0009	.0852		
α_6		.0042	.0058	.0617		
α_7		-.0043	-.0053	.0714		
α_8		.0125	.0021	.0683		
α_9		-.0022	-.0076	.0794		
α_{10}		.0094	.0061	.0708		
α_{11}		.0124	.0132	.0815		

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The Markov model replaces $\sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau$ in Table 1 with $\rho d_{i,t-1}$.

Table 6

Repeated Sampling Experiments
Random Effects Polya Model
No Classification Error in DGP
(Missing X's, 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 ($t = 1, ..., 10$)						
β_0	-.1000	-.0856	-.0852	.0460	.0482	2.21
β_1	1.0000	1.0219	1.0220	.1113	.1135	1.39
ρ	1.0000	1.0177	1.0223	.0745	.0766	1.68
α	.5000	.5015	.4918	.0633	.0633	.16
ϕ_2	.2500	.2377	.2441	.0697	.0708	-1.24
σ_ν	.5000	.4972	.4979	.0142	.0144	-1.38
σ_μ	.8000	.8005	.8009	.0465	.0465	.07
E	1.0000	.9249	.9290	.0566	.0937	-9.39
40% Missing Choices and X's ($t = 1, ..., 10$)						
β_0	-.1000	-.0921	-.0850	.0833	.0837	.67
β_1	1.0000	1.0207	1.0250	.1159	.1177	1.26
ρ	1.0000	1.0403	1.0185	.1072	.1146	2.66
α	.5000	.4864	.5139	.1010	.1019	-.95
ϕ_2	.2500	.2415	.2351	.1197	.1200	-.50
σ_ν	.5000	.4963	.4992	.0270	.0272	-.97
σ_μ	.8000	.8045	.8096	.0614	.0615	.52
E	1.0000	.9180	.9230	.0496	.0955	-11.69
60% Missing Choices and X's ($t = 1, ..., 10$)						
β_0	-.1000	-.0678	-.0805	.0893	.0949	2.55
β_1	1.0000	.9929	1.0300	.1795	.1797	-.28
ρ	1.0000	1.0280	1.0361	.1139	.1173	1.74
α	.5000	.4685	.4938	.1208	.1249	-1.84
ϕ_2	.2500	.2432	.2431	.1030	.1032	-.46
σ_ν	.5000	.4945	.4961	.0230	.0236	-1.68
σ_μ	.8000	.8055	.7908	.0694	.0696	.56
E	1.0000	.9366	.9341	.0698	.0922	-7.25

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 1.

Table 7

Repeated Sampling Experiments
Random Effects Polya Model
Unbiased Classification Error
Smooth Algorithm
(Missing X's, 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 ($t = 1, ..., 10$)						
β_0	-.1000	-.0900	-.0926	.0656	.0664	1.07
β_1	1.0000	.9974	.9927	.0962	.0962	-.19
ρ	1.0000	1.0347	1.0259	.1415	.1457	1.73
α	.5000	.5219	.5026	.1275	.1294	1.22
ϕ_2	.2500	.2512	.2494	.0162	.0163	.54
σ_ν	.5000	.5014	.5021	.0055	.0057	1.80
σ_μ	.8000	.8174	.8201	.0356	.0396	3.46
E	.7500	.7414	.7410	.0167	.0188	-3.65
40% Missing Choices and X's ($t = 1, ..., 10$)						
β_0	-.1000	-.0951	-.0832	.0682	.0684	.51
β_1	1.0000	1.0193	1.0146	.1046	.1064	1.31
ρ	1.0000	1.0627	1.0371	.1583	.1703	2.80
α	.5000	.5526	.5167	.1612	.1696	2.31
ϕ_2	.2500	.2498	.2536	.0246	.0246	-.05
σ_ν	.5000	.5124	.5023	.0792	.0802	1.10
σ_μ	.8000	.8162	.8168	.0343	.0380	3.34
E	.7500	.7453	.7408	.0220	.0225	-1.52
60% Missing Choices and X's ($t = 1, ..., 10$)						
β_0	-.1000	-.0956	-.0783	.0933	.0934	.33
β_1	1.0000	1.008	1.0093	.1596	.1598	.35
ρ	1.0000	1.0546	1.0652	.2215	.2281	1.74
α	.5000	.5488	.5637	.1854	.1917	1.86
ϕ_2	.2500	.2506	.2515	.0383	.0382	.11
σ_ν	.5000	.5011	.5015	.0084	.0085	.91
σ_μ	.8000	.8115	.8077	.0439	.0454	1.84
E	.7500	.7498	.7472	.0270	.0270	-.05

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 1.

Table 8

Summary Statistics
Representative Data Set
Polya Model with $AR(1)$ Errors
Unbiased Classification Error

t	Mean d_{it}	Mean d_{it}^*	Mean $\beta' x$	Mean ε_{it}	Mean π_{11t}	Mean π_{00t}	N
1	.4600	.4580	-.0125	-.0330	.8622	.8878	500
			(.2701)	(1.0164)			
2	.5740	.5700	.4709	-.0220	.8935	.8565	500
			(.5272)	(1.0525)			
3	.6340	.6280	.8778	-.0146	.9128	.8372	500
			(.8917)	(.9698)			
4	.6940	.6800	1.1514	-.0055	.9265	.8235	500
			(1.1668)	(.8593)			
5	.7380	.7420	1.3771	.0504	.9367	.8133	500
			(1.2028)	(.8507)			
6	.7700	.7840	1.5895	.0311	.9454	.8046	500
			(1.2453)	(.8962)			
7	.8000	.7960	1.7679	.0392	.9537	.7963	500
			(1.1408)	(.9582)			
8	.8360	.8620	1.8576	.0142	.9588	.7912	500
			(1.1427)	(.9893)			
9	.8480	.8260	1.9912	.0086	.9640	.7860	500
			(1.1048)	(1.0212)			
10	.8600	.8720	2.0187	.0233	.9677	.7823	500
			(.9955)	(.9182)			

Note: d_{it} is the true choice, d_{it}^* is the reported choice, π_{11t} and π_{00t} are the probabilities of a correct classification, and $\beta' x = u_{it} - \beta_0$. Variances are in parentheses. The frequency simulator that is used to compute the true classification error rates has $\widetilde{M}$ set to 1000. The model is:

$$
\begin{aligned}
u_{it} &= \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau + \varepsilon_{it} \\
d_{i0} &= 0,\ \rho_\tau = \rho e^{-\alpha(t-\tau-1)} \\
x_{it} &= \phi_2 x_{i,t-1} + \nu_{it},\ \nu_{it} \sim N\left(0, \sigma_\nu^2\right) \\
\varepsilon_{it} &= \phi_1 \varepsilon_{i,t-1} + \eta_{it},\ \eta_{it} \sim N(0, 1-\phi_1^2)
\end{aligned}
$$

Table 9

Repeated Sampling Experiments
Polya Model with $AR(1)$ Errors
Unbiased Classification Error
(Missing X's, 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 ($t = 1, ..., 10$)						
β_0	-.1000	-.1042	-.0981	.0391	.0394	-.76
β_1	1.0000	1.0021	1.0060	.0519	.0519	.29
ρ	1.0000	1.0444	1.0393	.0424	.0614	7.40
α	.5000	.5057	.5058	.0423	.0428	1.12
ϕ_2	.2500	.2521	.2486	.0181	.0183	.83
σ_ν	.5000	.5018	.5024	.0057	.0060	2.21
ϕ_1	.8000	.7996	.8003	.0264	.0264	-.12
E	.7500	.7473	.7486	.0174	.0176	-1.08
40% Missing Choices and X's ($t = 1, ..., 10$)						
β_0	-.1000	-.1052	-.1014	.0400	.0403	-.92
β_1	1.0000	1.0036	1.0011	.0566	.0567	.45
ρ	1.0000	1.0460	1.0400	.0446	.0640	7.30
α	.5000	.5018	.5053	.0405	.0405	.32
ϕ_2	.2500	.2522	.2531	.0261	.0262	.61
σ_ν	.5000	.5019	.5026	.0067	.0070	1.98
ϕ_1	.8000	.8002	.7989	.0301	.0301	.05
E	.7500	.7504	.7524	.0251	.0251	.12
60% Missing Choices and X's ($t = 1, ..., 10$)						
β_0	-.1000	-.1041	-.0996	.0524	.0526	-.55
β_1	1.0000	1.0003	1.0124	.0748	.0748	.03
ρ	1.0000	1.0433	1.0372	.0610	.0748	5.03
α	.5000	.5047	.5077	.0621	.0623	.54
ϕ_2	.2500	.2521	.2514	.0384	.0385	.39
σ_ν	.5000	.5007	.5018	.0086	.0086	.61
ϕ_1	.8000	.7988	.8019	.0364	.0364	-.23
E	.7500	.7514	.7514	.0346	.0348	.77

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 8.

Table 10

Repeated Sampling Experiments
Polya Model with $AR(1)$ Errors
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 $d_{i0} = 0$ $(t = 11, ..., 20)$						
β_0	-.1000	-.0896	-.0925	.0265	.0285	2.77
β_1	1.0000	1.0224	1.0221	.0479	.0529	3.31
ρ	1.0000	1.0194	1.0148	.0298	.0356	4.60
α	.5000	.5121	.5128	.0238	.0267	3.59
ϕ_2	.2500	.2511	.2531	.0138	.0139	.56
σ_ν	.5000	.5011	.5013	.0047	.0049	1.58
ϕ_1	.8000	.8071	.8100	.0280	.0289	1.80
E	.7500	.7420	.7455	.0261	.0273	-2.16
Assume process starts with $d_{i,10} = 0$ $(t = 11, ..., 20)$						
β_0	-.1000	.9503	.9682	.0605	1.0520	122.84
β_1	1.0000	.1699	.3883	.4544	.9463	-12.92
ρ	1.0000	.5849	.5266	.2792	.5003	-10.51
α	.5000	.7102	.7385	.3180	.3812	4.67
ϕ_1	.8000	.9221	.9259	.0316	.1261	27.33
E	.7500	.7656	.7485	.1323	.1332	.83
Use reported data from $t = 11, ..., 20$ to proxy for initial condition at $t = 21$ $(t = 11, ..., 30)$						
β_0	-.1000	-.0862	-.0812	.0617	.0632	1.58
β_1	1.0000	.9406	.9781	.0932	.1105	-4.50
ρ	1.0000	1.0445	1.0219	.0924	.1026	3.41
α	.5000	.5908	.5674	.0737	.1170	8.72
ϕ_1	.8000	.7562	.7749	.0828	.0937	-3.74
E	.7500	.7348	.7378	.0288	.0325	-3.73

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 8.

Table 11

Repeated Sampling Experiments
Markov Model with $AR(1)$ Errors
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 $d_{i0} = 0$ $(t = 10, ..., 20)$						
β_0	-.1000	-.1171	-.1125	.0429	.0462	-2.81
β_1	1.0000	1.0185	1.0191	.0323	.0373	4.05
ρ	1.0000	1.0354	1.0316	.0465	.0585	5.38
ϕ_2	.2500	.2511	.2509	.0139	.0140	.56
σ_ν	.5000	.5013	.5016	.0050	.0052	1.89
ϕ_1	.8000	.8081	.8077	.0266	.0278	2.15
E	.7500	.7401	.7403	.0126	.0160	-5.58
Assume process starts with $d_{i9} = 0$ $(t = 10, ..., 20)$						
β_0	-.1000	.1895	.1797	.0547	.2946	37.43
β_1	1.0000	.8189	.8025	.0727	.1951	-17.63
ρ	1.0000	.5932	.5807	.1054	.4202	-27.29
ϕ_1	.8000	.8377	.8343	.0268	.0463	9.95
E	.7500	.7539	.7544	.0164	.0168	1.68
Treat $d_{i,10}$ as exogenous						
β_0	-.1000	-.2416	-.2501	.0492	.1500	-20.36
β_1	1.0000	1.0150	1.0239	.0430	.0456	2.46
ρ	1.0000	1.2330	1.2380	.0702	.2434	23.47
ϕ_1	.8000	.7480	.7456	.0374	.0640	-9.83
E	.7500	.7322	.7316	.0151	.0234	-8.35
Use Heckman's approximation method to proxy for initial condition at $t = 11$ $(t = 10, ..., 20)$						
β_0	-.1000	-.2181	-.2206	.0538	.1298	-15.54
β_1	1.0000	1.0333	1.0315	.0471	.0577	5.00
ρ	1.0000	1.1997	1.2129	.0604	.2086	23.37
ϕ_1	.8000	.7727	.7746	.0316	.0418	-6.13
E	.7500	.7385	.7385	.0116	.0164	-7.00
γ_0		.4149	.4118	.0564		
γ_1		.6628	.6614	.0722		
$\rho_{\mu\epsilon^H}$		.7238	.7266	.0386		

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$. The Markov model replaces $\sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau$ in Table 8 with $\rho d_{i,t-1}$.

Table 12

Repeated Sampling Experiments
Polya Model with $AR(1)$ Errors
Unbiased Classification Error
Smooth Algorithm
(Missing X's, 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 ($t = 1, ..., 10$)						
β_0	-.1000	-.1007	-.0998	.0336	.0337	-.16
β_1	1.0000	.9936	.9838	.0519	.0522	-.87
ρ	1.0000	1.0336	1.0387	.0824	.0890	2.88
α	.5000	.5214	.5076	.0751	.0781	2.01
ϕ_2	.2500	.2513	.2494	.0162	.0163	.56
σ_ν	.5000	.5014	.5020	.0055	.0057	1.82
ϕ_1	.8000	.8004	.8009	.0203	.0203	.14
E	.7500	.7475	.7490	.0175	.0177	-.99
40% Missing Choices and X's ($t = 1, ..., 10$)						
β_0	-.1000	-.1041	-.1028	.0285	.0288	-1.03
β_1	1.0000	.9892	.9759	.0721	.0729	-1.05
ρ	1.0000	1.0604	1.0539	.1118	.1271	3.82
α	.5000	.5406	.5226	.0998	.1078	2.88
ϕ_2	.2500	.2517	.2532	.0248	.0248	.49
σ_ν	.5000	.5013	.5019	.0067	.0068	1.34
ϕ_1	.8000	.7984	.8004	.0193	.0194	-.60
E	.7500	.7506	.7514	.0233	.0233	.17
60% Missing Choices and X's ($t = 1, ..., 10$)						
β_0	-.1000	-.0979	-.0925	.0409	.0410	.35
β_1	1.0000	.9833	.9510	.1107	.1119	-1.07
ρ	1.0000	1.0625	1.0014	.1819	.1923	2.43
α	.5000	.5465	.5126	.1566	.1633	2.10
ϕ_2	.2500	.2537	.2515	.0364	.0366	.72
σ_ν	.5000	.5004	.5002	.0085	.0084	.30
ϕ_1	.8000	.8004	.7976	.0233	.0233	.12
E	.7500	.7524	.7527	.0314	.0315	.54

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 8.

Table 13

Repeated Sampling Experiments
Random Effects Polya Model
Biased Classification Error
(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
Low Classification Error Bias ($t = 1, ..., 10$)						
β_0	-.1000	-.0922	-.944	.0387	.0394	1.42
β_1	1.0000	1.0198	1.0131	.0531	.0567	2.63
ρ	1.0000	1.0144	1.0102	.0390	.0415	2.61
α	.5000	.5031	.5104	.0489	.0490	.45
ϕ_2	.2500	.2489	.2456	.0161	.0161	-.47
σ_ν	.5000	.5018	.5018	.0050	.0053	2.47
σ_μ	.8000	.8068	.8041	.0239	.0248	1.99
γ_0	-3.5000	-3.4867	-3.4762	.0580	.0595	1.62
γ_1	5.0000	4.9845	5.0033	.0728	.0744	-1.51
γ_2	2.0000	2.0161	2.0236	.0446	.0475	2.56
Medium Classification Error Bias ($t = 1, ..., 10$)						
β_0	-.1000	-.0941	-.0988	.0425	.0429	.98
β_1	1.0000	1.0045	1.0119	.0608	.0609	.52
ρ	1.0000	1.0222	1.0232	.0465	.0515	3.37
α	.5000	.5160	.5253	.0658	.0677	1.71
ϕ_2	.2500	.2476	.2452	.0162	.0163	-1.04
σ_ν	.5000	.5022	.5026	.0050	.0054	3.04
σ_μ	.8000	.8049	.8041	.0272	.0276	1.29
γ_0	-3.0000	-2.9902	-2.9826	.0561	.0570	1.24
γ_1	4.0000	3.98	3.9951	.0776	.0787	-1.19
γ_2	2.0000	2.0104	2.0134	.0782	.0789	.94
High Classification Error Bias ($t = 1, ..., 10$)						
β_0	-.1000	-.0988	-.0918	.0708	.0708	.12
β_1	1.0000	1.0145	1.0068	.0693	.0708	1.48
ρ	1.0000	1.0218	1.0228	.0791	.0820	1.94
α	.5000	.5088	.5328	.0993	.0997	.63
ϕ_2	.2500	.2484	.2460	.0164	.0165	-.70
σ_ν	.5000	.5021	.5028	.0051	.2980	2.90
σ_μ	.8000	.8023	.7999	.0406	.3050	.40
γ_0	-3.0000	-2.9918	-2.9983	.0638	.0643	.91
γ_1	3.0000	2.9842	2.9920	.0829	.0844	-1.34
γ_2	3.0000	3.0190	3.0371	.1018	.1036	-1.32

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 1.

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
		Low Classification Error Bias ($t = 1, ..., 10$)				
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_ν	.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

Note: The number of replications is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 1.

Table 15

Repeated Sampling Experiments
Polya Model with $AR(1)$ Errors
Biased Classification Error
(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
		Low Classification Error Bias ($t = 1, ..., 10$)				
β_0	-.1000	-.1033	-.1039	.0406	.0407	.57
β_1	1.0000	1.0176	1.0114	.0649	.0673	1.91
ρ	1.0000	1.0322	1.0325	.0385	.0502	5.92
α	.5000	.5017	.5050	.0461	.0461	.25
ϕ_2	.2500	.2496	.2502	.0165	.0165	-.16
σ_ν	.5000	.5018	.5023	.0049	.0052	2.62
ϕ_1	.8000	.7987	.7961	.0264	.0265	-.35
γ_0	-3.5000	-3.4987	-3.4809	.0664	.0665	.14
γ_1	5.0000	4.9831	5.0056	.0697	.0717	-1.72
γ_2	2.0000	2.0265	2.0196	.0451	.0513	4.15
		Medium Classification Error Bias ($t = 1, ..., 10$)				
β_0	-.1000	-.0893	-.0982	.0525	.0536	1.44
β_1	1.0000	1.0075	1.0040	.0745	.0749	.71
ρ	1.0000	1.0283	1.0364	.0534	.0604	3.75
α	.5000	.5162	.5101	.0540	.0563	2.12
ϕ_2	.2500	.2478	.2469	.0163	.0164	-.94
σ_ν	.5000	.5024	.5027	.0046	.0052	3.74
ϕ_1	.8000	.8016	.8023	.0312	.0312	.35
γ_0	-3.0000	-3.0058	-3.0009	.0716	.0718	-.57
γ_1	4.0000	3.9802	3.9803	.0735	.0761	-1.90
γ_2	2.0000	2.0151	2.0227	.0659	.0676	1.62
		High Classification Error Bias ($t = 1, ..., 10$)				
β_0	-.1000	-.0926	-.0896	.0756	.0758	.69
β_1	1.0000	1.0135	1.0201	.0778	.0790	1.23
ρ	1.0000	1.0276	1.0255	.0682	.0735	2.86
α	.5000	.5074	.5033	.0624	.0629	.83
ϕ_2	.2500	.2476	.2446	.0152	.0153	-1.10
σ_ν	.5000	.5019	.5030	.0051	.0055	2.62
ϕ_1	.8000	.7980	.8046	.0386	.0387	-.36
γ_0	-3.0000	-3.0026	-2.9870	.0823	.0824	-.23
γ_1	3.0000	2.9899	2.9807	.0680	.0687	-1.04
γ_2	3.0000	3.0186	3.0185	.0693	.0717	1.90

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$. The model is the same as in Table 8.

Table 16

Repeated Sampling Experiments
Polya Model with $AR(1)$ 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
Low Classification Error Bias ($t = 1, ..., 10$)						
β_0	-.1000	-.0958	-.0971	.0336	.0338	.89
β_1	1.0000	1.0016	.9979	.0539	.0539	.21
ρ	1.0000	1.0213	1.0224	.0746	.0775	2.02
α	.5000	.5117	.5171	.0633	.0644	1.31
ϕ_2	.2500	.2488	.2466	.0151	.0152	-.58
σ_ν	.5000	.5020	.5028	.0047	.0051	2.95
ϕ_1	.8000	.8035	.8030	.0177	.0181	1.41
γ_0	-3.5000	-3.3707	-3.3710	.2730	.3021	3.35
γ_1	5.0000	4.7756	4.7931	.2778	.3571	-5.71
γ_2	2.0000	2.1014	2.0863	.1859	.2957	3.86

Note: The number of replications is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is the same as in Table 8.

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
		Low Classification Error Bias ($t = 1, ..., 10$)				
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

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500. $Std(\widehat{\beta})$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t-statistics are calculated as $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The model is:

$$
\begin{aligned}
u_{it} &= \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau + \varepsilon_{it} \\
d_{i0} &= 0,\ \rho_\tau = \rho e^{-\alpha(t-\tau-1)} \\
x_{it} &= \phi_2 x_{i,t-1} + \nu_{it},\ \nu_{it} \sim N\left(0, \sigma_\nu^2\right) \\
\varepsilon_{it} &= \mu_i + \xi_{it} \\
\xi_{it} &= \phi_1 \xi_{it-1} + \eta_{it}, \eta_{it} \sim N(0, (1-\sigma_\mu^2)(1-\phi_1^2))
\end{aligned}
$$

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)	

Note: Means and standard errors (in parentheses) for 1310 continuously married women in the PSID between 1994 and 2003, aged 18-60 in 1994, with positive husband earnings and hours worked in each non-missing year. Earnings are in thousands of 1994 dollars. Variable definitions and sample selection criteria are the same as those chosen by Hyslop (1999) for PSID calendar years 1980-1986.

Table 19

Female Labor Force Participation Decisions
PSID Calendar Years 1994-2003
Missing Years 1998, 2000, and 2002
Markov Model with Random Effects and $AR(1)$ errors
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_ν	.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-Likelihood	-12673.61	-12651.32	-12668.19	-12637.15
χ^2 (H_0: $\delta = 0$)	-	44.58 (.0243)	-	62.08 (.0002)
χ^2 (Pearson GOF)	59.62 (.1024)	57.15 (.1474)	58.32 (.1245)	56.40 (.1637)
N	1310	1310	1310	1310

Note: The model is:

$$\begin{aligned}
u_{it} &= \beta_0 + \beta_1 \ln(y_{it}) + \beta_2' X_{it} + \rho d_{i,t-1} + \varepsilon_{it} \\
d_{i0} &= 0, \\
\ln(y_{it}) &= \phi_2 \ln(y_{i,t-1}) + \nu_{it},\ \nu_{it} \sim N\left(0, \sigma_\nu^2\right) \\
\varepsilon_{it} &= \mu_i + \xi_{it} \\
\xi_{it} &= \phi_1 \xi_{it-1} + \eta_{it},\ \eta_{it} \sim N(0, (1-\sigma_\mu^2)(1-\phi_1^2)) \\
l_{it} &= \gamma_0 + \gamma_1 d_{it} + \gamma_2 d_{it-1}^* + \omega_{it} \\
\mu_i &= \sum_{t=1}^{T} \delta_t' W_{it} + \sigma_\mu \zeta_i,\ \zeta_i \sim N(0,1)
\end{aligned}$$

y_{it} is the husband's annual earnings in year t. X_{it} contains year effects in addition to the fertility, race and education covariates that appear explicitly in the table. W_{it} contains $\ln(y_{it})$ and the three fertility variables. Standard errors are in parentheses (p-values for the LRT and Pearson GOF chi-square statistics).

Table 20

Female Labor Force Participation Decisions
PSID Calendar Years 1994-2003
Missing Years 1998, 2000, and 2002
Polya Model with Random Effects and $AR(1)$ errors
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_ν	.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-Likelihood	-12568.10	-12544.89	-12561.69	-12531.88
χ^2 (H_0: $\delta = 0$)	-	46.42 (.0158)	-	59.62 (.0005)
χ^2 (Pearson GOF)	54.62 (.2075)	51.90 (.2887)	53.32 (.2442)	51.02 (.3186)
N	1310	1310	1310	1310

Note: The model is:

$$\begin{aligned}
u_{it} &= \beta_0 + \beta_1 \ln(y_{it}) + \beta_2' X_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau + \varepsilon_{it} \\
d_{i0} &= 0,\ \rho_\tau = \rho e^{-\alpha(t-\tau-1)} \\
\ln(y_{it}) &= \phi_2 \ln(y_{i,t-1}) + \nu_{it},\ \nu_{it} \sim N\left(0, \sigma_\nu^2\right) \\
\varepsilon_{it} &= \mu_i + \xi_{it} \\
\xi_{it} &= \phi_1 \xi_{it-1} + \eta_{it},\ \eta_{it} \sim N(0, (1-\sigma_\mu^2)(1-\phi_1^2)) \\
l_{it} &= \gamma_0 + \gamma_1 d_{it} + \gamma_2 d_{it-1}^* + \omega_{it} \\
\mu_i &= \sum_{t=1}^{T} \delta_t' W_{it} + \sigma_\mu \zeta_i,\ \zeta_i \sim N(0,1)
\end{aligned}$$

y_{it} is the husband's annual earnings in year t. X_{it} contains year effects in addition to the fertility, race and education covariates that appear explicitly in the table. W_{it} contains $\ln(y_{it})$ and the three fertility variables. Standard errors are in parentheses (p-values for the LRT and Pearson GOF chi-square statistics).