DISCUSSION PAPER SERIES

IZA DP No. 4054

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

Michael P. Keane
Robert M. Sauer

March 2009

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 Southampton
and IZA

Discussion Paper No. 4054
March 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 4054
March 2009

ABSTRACT

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

This paper develops a simulation estimation algorithm that is particularly useful for estimating dynamic panel data models with unobserved endogenous state variables. The new approach can easily deal with the commonly encountered and widely discussed "initial conditions problem," as well as the more general problem of missing state variables during the sample period. Repeated sampling experiments on dynamic probit models with serially correlated errors indicate that the estimator has good small sample properties. We apply the estimator to a model of married women's labor force participation decisions. The results show that the rarely used Polya model, which is very difficult to estimate given missing data problems, fits the data substantially better than the popular Markov model. The Polya model implies far less state dependence in employment status than the Markov model. It also implies that observed heterogeneity in education, young children and husband income are much more important determinants of participation, while race is much less important.

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

Keywords: initial conditions, missing data, simulation, female labor force participation

Corresponding author:

Robert M. Sauer
University of Southampton
School of Social Sciences
Southampton SO17 1BJ
United Kingdom
E-mail: R.M.Sauer@soton.ac.uk

1 Introduction

The problem of unobserved endogenous state variables arises frequently in the estimation of dynamic discrete choice models. It is present whenever 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 assesses the performance of and empirically implement a new simulated maximum likelihood (SML) estimation algorithm that is particularly useful for estimating dynamic panel data models with unobserved endogenous state variables. The novel estimation technique was recently introduced by Keane and Wolpin (2001) (KW) to estimate the parameters of a discrete choice dynamic programming problem with both unobserved initial conditions and missing choices during the sample period. However, the algorithm has a much wider 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 computational advantage of the new SML estimation algorithm lies in the fact that in contexts where performing conditional simulations of data from a model would be extremely difficult, unconditional simulation is often straightforward. Simulation of the likelihood in dynamic models often involves conditional simulation (of choice probabilities conditional on past history), but when past history is not fully observed, conditional simulation is often computationally infeasible.[1]

In this study, we first describe how the SML algorithm developed by KW, which

[1] 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.

only requires unconditional simulations, can be extended to a number of cases beyond the specific discrete choice dynamic programming problem they considered. In particular, we assess the performance of the estimator on panel data probit models with a time-varying exogenous covariate, lagged endogenous variables and serially correlated errors. Such panel probit models have been a leading case in past discussions of dynamic panel data models with unobserved initial conditions (see Heckman (1981*a*)). Specification of panel probit models, rather than discrete choice 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 female labor force participation using PSID data from 1994-2003. A serious missing data problem naturally arises in these data because, in addition to the usual initial conditions problem, respondents were not interviewed at all in 1998, 2000 and 2002. Hyslop (1999) also used the PSID to estimate dynamic probit models of female labor force participation, and to test for endogeneity of fertility and nonlabor income in models that include complex error structures. Using the new algorithm, we extend his results to allow for classification error and missing data. This enables us to include the post-1994 data, as well as to consider a more general specification of state dependence (i.e., the Polya model). In contrast to the results in Hyslop (1999), we reject the null hypothesis that fertility and nonlabor income are exogenous in these more general models.

The rest of this paper is organized as follows. Section 2 reviews the literature on different approaches to the problem of unobserved endogenous state variables, and places our algorithm in context. Section 3 describes the dynamic panel data probit model used in the repeated sampling experiments. Section 4 develops two different models of classification error that are incorporated into the estimation technique. Classification error in discrete outcomes is a key feature of the algorithm. Section 5 describes our algorithm in detail. Sections 6 and 7 present Monte-Carlo test results under two models of classification error. Section 8 applies the algorithm to a model of female labor force participation. Section 9 summarizes and 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 (1981*a*) showed how, in dynamic discrete choice models, the assumption of stationarity allows one to derive the marginal probability of the initial state. As stationarity is often problematic, Heckman (1981*a*) 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 (which depends on initial period covariates, and whose error is correlated with errors during the sample period).[2] More recently, Wooldridge (2003) proposed an alternative approximate solution to the initial conditions problem. 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

[2] This approximate solution performed better than fixed effects probit, but still produced biases of more than 10% in repeated sampling experiments.

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.

Due to the computational difficulties in solving the missing data problem, applied economists frequently resort to the simpler methods of case deletion and imputation. Case deletion can cause large amounts of information to be lost, resulting in inefficient estimates. It can also introduce biases to the extent that 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.

In contrast to the previous literature, the SML estimation algorithm that we propose offers a systematic unified "solution" to both the initial conditions problem and the problem of missing data during the sample period. The algorithm 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 the probabilities of events at each date t conditional on the state at the start of time t, 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 in dynamic models using only unconditional simulations is that reported choices are measured with error. This allows one to simulate probabilities of choice histories using unconditional frequency simulation, as it avoids the usual problem in frequency simulation that an impractically large number of simulations is necessary to obtain non-zero probabilities of low probability events. Furthermore, the assumption that choices are measured with error is certainly valid in the vast majority of data sets that economists use.

Prior work showing the importance of classification error includes Poterba and Summers (1986, 1995) and Flinn (1997). For example, 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 the figures in their Table II). If misclassification is present and not included in the analysis, maximum likelihood estimation leads to biased and inconsistent parameter estimates (Hausman, Abrevaya and Scott-Morton (1998)).[3]

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.

3 The Panel Data Probit Model

In the panel data probit model, the utility of the first option, for individual i at time t, is denoted as u_{it}, and the utility of the second option is normalized to zero. Utility is unobserved by the researcher, but the individual is assumed to choose the option that gives greatest utility. We will 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 the specification in (1) allows the entire history of past choices to affect current utility. It is, therefore, more general than the familiar first-order Markov

[3] 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.

process.[4] Depreciation in the importance of past choices is captured through the weights ρ_τ. The theoretical start of the process is, by definition, $d_{i0} = 0$.

The error term ε_{it} in (1) is assumed to be serially correlated. Thus, lagged choices are endogenous. In the simple case of serially independent errors, lagged choices are exogenous, and the problems we consider in this paper do not arise. Although our approach is very flexible in terms of the nature of the serial correlation that can be accommodated, we consider three leading cases in our experiments. First, the source of serial correlation could be time-invariant random individual effects, i.e.,

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

where μ_i is normally distributed with zero mean and variance σ_μ^2, and η_{it} is normally distributed with zero mean and variance σ_η^2. Second, serial correlation could derive from an $AR\,(1)$ process,

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

where η_{it} has the same distribution as in (3). Third, serial correlation could arise from a combination of time-invariant random individual effects and an $AR(1)$ process, i.e.,

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

where η_{it} has the same distribution as in (3).

Although the model of (1)-(5) is restrictive, the estimation procedure can easily accommodate a wide range of alternative specifications and distributions of the error term. For example, KW employ a variant of the algorithm in a multinomial choice setting with an error term that contains both a nonparametric individual effect and a multivariate normal disturbance contemporaneously correlated across choices. Also,

[4] More general processes than first-order Markov have not been widely used in the economics literature. We suspect that this is due, in part, to the difficulty in dealing with missing data. But, more general models are quite standard in marketing. See, e.g., Erdem and Keane (1996).

while we only consider a scalar process in (1), extension to vectors of discrete and mixed discrete/continuous outcomes (as in KW) is straightforward. We emphasize that 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. In most contexts in applied economics this is a sensible assumption. Moreover, our approach can be implemented given any assumed 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, the general model of misclassification that we consider is characterized by four probabilities,

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

where π_{11t} is the probability that option one is reported to be chosen ($d_{it}^* = 1$) given that it is the true choice ($d_{it} = 1$); 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 the Monte Carlo analysis of our algorithm we consider cases where classification error rates are dependent on true choices, but not on covariates. Error rates would depend on the true choice if, for example, workers who change jobs mis-report more often than workers who do not. Poterba and Summers (1995) and Hausman et. al. (1998) find evidence of this type of misclassification in the CPS and

PSID. 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 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). Identification of structural parameters will be stronger the more parsimonious is the model of misclassification. Moreover, economic theory provides guidance for specification of the decision model but not necessarily for the model of misclassification. Thus, we 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 drive model fit.

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

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

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 $\Pr(d_{it}=1)$,

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

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, $\Pr(d_{it}^{*}=1 \mid d_{it}=1)$, also approaches one, which must be the case to preserve unbiasedness. Further, as $\Pr(d_{it}=1)$ approaches zero, $\Pr(d_{it}^{*}=1 \mid d_{it}=1)$ approaches E. E can thus be interpreted as a "baseline" classification rate. In other words, 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:

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

Note the great parsimony that unbiasedness imposes on the classification error process. It depends on the single parameter E, which is treated as a free parameter in estimation. One could generalize this specification by letting E depend on covariates. In that case, 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, in general, lead to a biased classification error process in which $\Pr(d_{it}^{*}=1) \neq \Pr(d_{it}=1)$. The biased classification error scheme 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. Notice that the specification in (10) allows the probability of reporting a particular choice to differ by the true choice, and allows for dynamic mis-reporting, since d_{it-1}^* appears in the index function. The greater in magnitude is γ_2 (the coefficient on d_{it-1}^*), the more likely is persistent mis-reporting.

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

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

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).[5] In (10) this is equivalent to $\gamma_1 > 0$.

Interestingly, in our 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.[6] 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 $I(d_{it}^*\ observed)$ 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:

[5] 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 $1 - \pi_{01t}$.

[6] A recent paper by Gould (2007) claims to implement our algorithm using $E = 0$, but, as we see here, this is not possible. What Gould actually did is set $P(d_{it}^* = 1|X_{it}) = P(d_{it} = 1)$, i.e., set the choice probability conditional on a person's state X_{it} equal to the unconditional choice probability in the population. Hence, any parameters capturing dynamics in his model are not identified.

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).

3. Construct the conditional probabilities $\left\{\{\widehat{\pi}_{jkt}^m\}_{t=1}^T\right\}_{m=1}^M$ 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:

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

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

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

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

[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 $\left\{\{\varepsilon_{i\tau}\}_{\tau=0}^{t-1}\right\}$.

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

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

Step $(3b)$:

For the biased classification error process given by (11), the $\widehat{\pi}_{jkt}^{m}$'s in step (3) depend on the reported choice in the previous period $d_{i,t-1}^{*}$. If $d_{i,t-1}^{*}$ is missing, it must be simulated. This can be easily done using (10). Let the simulated $d_{i,t-1}^{*}$ be denoted $d_{i,t-1}^{*m}$, and let $d_{i,t-1}^{*(m)} = I\left(d_{i,t-1}^{*}\ observed\right) d_{i,t-1}^{*} + \left(1 - I\left(d_{i,t-1}^{*}\ observed\right)\right) d_{i,t-1}^{*m}$. 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, $I\left(d_{it}^{*}\ observed\right)$ is zero, and 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 choice history can generate any observed choice history when there is classification error. It is also important to note that (12) builds the likelihood using *unconditional* simulations of the model. The simulation of conditional probabilities like $P\left(d_{it} \mid H_{it}\right)$ 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

[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.

to previous *simulated* choices, rather than previous *reported* choices, which greatly simplifies the problem.[9]

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 estimation procedure described above need only be slightly modified to accommodate 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

[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 simple panel probit model with serial correlation but no state dependence, missing choices present no problem for GHK. To simulate a choice probability at time t, one needs a draw for the lagged stochastic terms that is 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. However, GHK runs into problems in three cases: (i) with state dependence one must also condition on lagged simulated *choices* in periods when the actual choice is missing. As one iterates on the model parameters, the simulated choice may change, leading to discontinuities in the simulated likelihood. A possible solution is to integrate over all possible missing choices (weighting by the probability of each), 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 the subset of observed choices can be extremely 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 extremely difficult.

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 ν_{it} is normally distributed with zero mean and variance σ_v^2, and where $x_{i0} = 0$. If x_{it-1} is observed and x_{it} is missing, then 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 from the ν_{it} distribution is taken for each simulated choice history m.

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

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

where $f_m(x_{it})$ is the density of the exogenous covariate.

Under the assumption that ν_{it} is distributed normally, the density of x_{it} according to draw sequence m is,

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

where $\widehat{x}_{it-1}^{(m)} = I(x_{i,t-1}\ observed)\, x_{it-1} + (1 - I(x_{i,t-1}\ observed))\, \widehat{x}_{it-1}^{m}$ and ϕ is the standard normal p.d.f.. Note that in periods in which x_{it} is missing, the density does not affect the likelihood. $f_m(x_{it})$ enters the likelihood only when x_{it} is observed. The parameters 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]

[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.

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

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

The only difference between (17) and (19) is that here the first d_{it}^* is observed at $t = \widetilde{\tau}$. In Heckman's approximation method, one would specify a distribution for $d_{i\widetilde{\tau}}^*$. In our method, it is not necessary to construct a marginal distribution for the initial state. The distribution of the initial state in period $\widetilde{\tau}$ is implicitly determined by the simulated 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 might just set $\widetilde{\tau}$ large enough so that estimates are not sensitive to further increases. Alternatively, if the theoretical start of the process can not be determined, one could easily 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 explicitly considered below.

5.2 Importance Sampling

Non-smoothness of the simulated likelihood function based on (19) arises 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 estimation procedure can be easily modified to take advantage of importance sampling techniques that smooth the likelihood and enable the use of standard 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.

[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.

We then hold the $\{d_{it}^m(\theta_0)\}_{t=1}^T$ sequences fixed as we vary θ. Each simulated choice sequence then has an associated importance sampling weight, $W_m(\theta)$, that varies with θ. The basic idea of importance sampling is that, when we change θ, sequences that are more (less) likely under the new θ receive increased (reduced) weight. Thus:

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

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 vector of trial parameters 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 function is to construct, at the initial θ_0, simulated choice histories $\{d_{it}^m(\theta_0)\}_{t=1}^T$ and the latent variable sequences $\{U_{it}^m(\theta_0)\}_{t=1}^T$ that generate $\{d_{it}^m(\theta_0)\}_{t=1}^T$, where $U_{it}^m(\theta_0) = \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}^m \rho_\tau + \varepsilon_{it}$. One then holds 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,

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

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

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

Note that (19) is just a special case of (24) with $W_m = 1$ for each simulated choice 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.

6 Monte-Carlo Tests - Unbiased Misclassification

This section reports Monte-Carlo tests of the SML estimator with unbiased classification error. The algorithm used to generate artificial data sets with unbiased classification error is described in Appendix A. Subsections 6.1 and 6.2 present results for the random effects and $AR(1)$ specifications of the error term, respectively. In each repeated sampling 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.

Each estimation on the 50 different panels $\{D_i^*, x_i\}_{i=1}^{N}$ uses a different random number generator seed to generate the M unconditional simulations for each individ-

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

ual in the sample. For each repeated sampling experiment, the true parameters, the mean, the median, the empirical standard deviations, the root mean square error of the estimates, and the t-statistics for the statistical significance of the biases, based on the empirical standard deviations, are reported.[13]

6.1 Random Effects Model

In the random effects model, the error term ε_{it} follows the components of variance structure in (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 for this model is $\theta = \{\beta_0, \beta_1, \rho, \alpha, \phi_2, \sigma_v, \sigma_\mu, E\}$. 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 type of model (a generalized Polya process with decay) are discussed in Heckman (1981*b*).

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 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 in Table 1 shows that, over time, an increasing proportion of individuals choose the first option. At $t = 1$ just under 50 percent of the sample have $d_{it} = 1$. At $t = 10$, the proportion reaches 85 percent. 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 $\beta' x$

[13] 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.

column displays the mean and variance of $\beta' x = \beta_0 + \beta_1 x_{it} + \rho \sum_{\tau=0}^{t-1} e^{-\alpha(t-\tau-1)} d_{i\tau}$ and the Mean ε_{it} column displays the mean and variance of the composite error term. 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 the variance of ε_{it} by the third period.

The Mean π_{11t} and Mean π_{00t} columns of Table 1 present the average probabilities of a correct classification.[14] The average probability of a correct match of $d_{it} = 1$ and $d_{it}^* = 1$, π_{11t}, is .863 in period 1 and increases over time to .956 in period 10. The average probability of a correct match of $d_{it} = 0$ and $d_{it}^* = 0$, π_{00t}, is .887 in period 1 and decreases to .794 in period 10. This pattern emerges because π_{11t} is an increasing linear function of the proportion choosing $d_{it} = 1$, and π_{00t} is a decreasing linear function of the same proportion, as shown in (9). The slope of the linear functions is $(1 - E)$. The base classification error rate E is set to .75, implying that even low probability events have a fairly high probability 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 difference between the four experiments is 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. The number of simulated choice histories per individual, M, is set equal to 1000, unless otherwise noted. For starting values, we use an initial parameter vector where each element is bumped 20% away from the true values.

As the figures in Table 2 illustrate, the SML estimator produces biases, but they are negligible in magnitude. The bias in the estimate of ρ is statistically significant

[14] We use $\widetilde{M} = 1000$ to calculate the classification probabilities.

in all four panels; however, the magnitude of the bias never exceeds 5.1 percent. The biases in the estimates of β_1 and E are sometimes significant but never exceed 2 percent. The medians of the parameter estimates are also quite close to the means, suggesting that the sampling distributions are symmetric. Note that the empirical standard errors of the estimates generally increase with the increased incidence of missing choices. An increased incidence of missing choices does not change the point estimates much since a higher proportion of missing choices does not substantially alter reported choice frequencies. Since choices are missing at random, the effect of a higher proportion of missing choices is only to reduce the effective sample size. The t-statistics for significant biases generally decrease because the biases are mostly unaffected and the empirical standard errors increase.

The biases we see in Table 2 are relatively small considering that biases on the order of 5-8% are quite common 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 (with 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 negligible 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 does little to change the results. Lowering M to 500 also has little effect, but is 61% faster. Mean time to convergence over the 50 repetitions in the second panel of Table 2 (20% missing choices, $M = 1000$) is 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 computer with two 1.0 GHz processors and 0.5 GHz RAM.

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. Here, 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 simply ignoring the initial conditions problem by assuming the choice process starts at $t = 10$ with $d_{i,10} = 0$. As missing pre-sample covariates are also ignored, the parameters of the exogenous covariate process, ϕ_1 and σ_ν, are not estimated. The biases produced by this method are generally substantial in magnitude. 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.[15]

The third panel of Table 4 reports the results of handling the initial conditions problem by constructing a proxy for the initial value of $\sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau$ using the observed data. The number of periods in this experiment is increased to $T = 30$. The DGP is modified so choices and covariates are completely missing in periods $t = 1, ..., 10$

[15] The variance of the composite error term is restricted to be between zero and one. Since almost all of the estimates of σ_μ are close to the upper boundary of one, the standard deviation over the fifty estimates is very small.

but are observed from $t = 11, ..., 30$. The observed choices in period $t = 11, ..., 20$ are used to form a proxy for $\sum_{\tau=0}^{20} d_{i\tau}\rho_\tau$ and the likelihood is constructed using only data from $t = 21, ..., 30$. In this method, the latent index at $t = 21$, u_{21}, is given by:

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

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 error rate parameter E is severely biased downward.

Table 5 reports the results of four repeated sampling experiments in which there is an initial conditions problem and the model has a more familiar first-order Markov structure in past choices. The Markov model is nested in the general 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 $t = 10, ..., 20$, except 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 approach (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. 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.[16]

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 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 estimation algorithm differs from the non-smooth version in that we simulate choice histories only *once* for each individual in the sample, at the

[16] 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 error rate, set to .25 and .50. Lower values of E correspond to a greater extent of classification error.

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 set of sequences. The smooth algorithm enables the use of standard gradient methods of optimization, as opposed to generally more time consuming non-gradient methods. We again set simulation size $M = 1000$ and use an initial parameter vector where each element is bumped 20% away from true values.

Table 7 reports the results of three repeated sampling experiments that use the smooth SML algorithm, with the weights specified in (22). These are analogous to the repeated sampling experiments in Table 3 that use the non-smooth algorithm. The three experiments differ in the proportion of missing choices and covariates during the sample period, assuming no initial conditions problem. Like Table 3, Table 7 reveals a few statistically significant biases, but the biases are trivial in magnitude.

It is important to note that the smooth version of the algorithm is faster. As reported earlier, the mean time to convergence over the 50 repetitions in the second panel of Table 2 (20% missing choices) is 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. Thus, the smooth version is roughly twice as fast.

6.2 $AR(1)$ Error Model

In the $AR(1)$ error model, the error term ε_{it} follows the first-order serial correlation 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 same exponential decay process, $\rho_\tau = \rho e^{-\alpha(t-\tau-1)}$. The vector of estimable parameters is $\theta = \{\beta_0, \beta_1, \rho, \alpha, \phi_2, \sigma_v, \phi_1, E\}$.

Table 8 reports summary statistics, by time period, for a representative data set produced by the Polya model with $AR(1)$ errors. The data set 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 an $AR(1)$ error parameter of

.80 implies a considerable amount of serial correlation. As in the random effects model, the variance of ε_{it} is normalized to one and the frequency simulator that is used to compute true classification error rates has M set to 1000. A comparison of Tables 1 and 8 shows that the summary statistics produced by the $AR(1)$ error model are quite similar to the summary statistics produced by the random effects model.

6.2.1 Non-Smooth SML Algorithm

The order of repeated sampling experiments on the $AR(1)$ error 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 the results of increasing the incidence of missing choices and covariates during the sample period, assuming no initial conditions problem. As in the experiments on the random effects model, the bias in ρ is generally significant but negligible in magnitude, never exceeding 4.6%. The biases and standard errors of the parameter estimates are generally smaller in the $AR(1)$ error model than in the random effects model (compare Tables 3 and 9).

In Table 10, different solutions to the initial conditions problem are examined. 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 does in the random effects model. But the second panel, in which the initial conditions problem is ignored (i.e., just set $d_{i,10} = 0$), reveals serious biases. In particular, the $AR(1)$ parameter (ϕ_1) is substantially over-estimated (i.e., .92 vs. .80). The biases in the estimates of ρ and α are also very large. Since ρ is biased downward and α is biased upward, the estimates understate the importance of lagged choices.

The third panel shows results from treating the observed $d_{i,10}$ as exogenous. The magnitudes of the biases when using this approach are generally smaller in the $AR(1)$ model than in the random effects model. However, as in the random effects model, the estimates of ρ and α are biased upward, understating state dependence.

Table 11 examines different solutions to the initial conditions problem in the

Markov model with $AR(1)$ errors. As in the random effects model, simulating from the theoretical start of the process works well. Ignoring the initial conditions problem produces substantial biases that are similar in direction and magnitude to the random effects model (see Table 5). Treating the initial condition as exogenous (panel 3) or using the Heckman approximation method (panel 4) result in more serious biases in the $AR(1)$ error model than in the random effects model. In these latter two methods, the estimates of ρ are biased upward by 23% and 20%, respectively.[17]

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), the estimates have biases that are negligible in magnitude. Consistent with previously reported results for the random effects model, the $AR(1)$ model also converges much faster when using the smooth algorithm. For example, while the mean time to convergence over the 50 repetitions in the first panel of Table 9 (20% missing choices) was 3.07 hours with a standard deviation of .71, that over the 50 repetitions in the first panel of Table 12 was only 1.84 hours with a standard deviation of .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). The algorithm used to generate artificial data sets is described in Appendix B. Subsections 7.1 and 7.2, present results for Polya models with random effects and $AR(1)$ errors, respectively. In subsection 7.3, we present results for the Polya model with *both* random effects and $AR(1)$ errors.

[17] 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).

7.1 Random Effects Model

7.1.1 Non-Smooth SML Algorithm

The three panels of Table 13 report the results of using the non-smooth SML algorithm on Polya models with random effects and *biased* classification error. The vector of true structural parameters is the same as in the case of unbiased classification error. 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 in Table 13 differ in the true parameters of the classification error process, γ_0, γ_1 and γ_2.

The first panel specifies values of γ_0, γ_1 and γ_2 that produce a relatively low level of bias in classifications. The parameters in the second panel generate an intermediate level of bias and the parameters in the third panel imply a relatively large bias. The conditional probabilities $\pi_{11t} = \Pr\left(d^*_{it} = 1 \mid d_{it} = 1\right)$ and $\pi_{01t} = \Pr\left(d^*_{it} = 1 \mid d_{it} = 0\right)$ are $(.97, .18)$, $(.95, .27)$ and $(.95, .50)$, in the first, second and third panels, respectively.

The results reveal relatively few statistically significant biases. Only the estimates of ρ and σ_v are consistently biased, but the magnitudes of these biases are negligible, never exceeding 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.[18] 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.

[18] The algorithm with a high extent of classification bias, and 20% missing choices and covariates, converges in similar time to the corresponding specification with unbiased classification error. For example, the time to convergence *per parameter* is .54 hours in the former case and .57 hours in the latter. The *overall* time to convergence for the unbiased and biased classification error models cannot be directly compared because they have a different number of parameters.

7.1.2 Smooth SML Algorithm

Table 14 reports the results of estimating 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 the choices and covariates are missing in each period, there is no initial conditions problem and there is a relatively low extent of true classification error bias. The results reveal slightly larger biases and standard errors than when using the non-smooth algorithm (compare Tables 13 panel 1 and 14). However, the biases remain small. The largest biases are in the estimates of ρ and α, which are biased 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, but for 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.

7.2.2 Smooth SML Algorithm (Importance Sampling)

Table 16 reports a similar experiment to that in Table 14 except with $AR(1)$ errors rather than random effects. The biases are once again negligible in magnitude and noticeably smaller than in the random effects specification. The estimates of ρ and α are biased by only 2.1% and 2.3%, respectively.

7.3 Random Effects and $AR(1)$ Errors

Finally, we consider a model with *both* random effects and $AR(1)$ errors. Here, the error term ε_{it} follows the error process in (5). The true σ_μ is set to .80 while the $AR(1)$ parameter ϕ_1 is set to $\phi_1 = .40$. 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. The smooth algorithm produces biases in the estimated parameters that 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 $\{\rho, \alpha, \sigma_\mu, \phi_1\}$.[19]

8 Application to Female Labor Force Participation

In this section, we use our algorithm to estimate dynamic probit models of married women's labor force participation, 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. We use our estimates to test for endogeneity of fertility and nonlabor income (following Hyslop (1999)).

[19]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 random Census subsample of families and nonrandom Survey of Economic Opportunities. 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.[20] Even in the seven years when labor force participation is reported, it is likely to be measured with error.[21]

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 labor

[20] 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).

[21] For example, Poterba and Summers (1986) used the so called "CPS reconciliation data" to assess the extent of classification error in reported employment status in the CPS. In the reconciliation data, Census sends an interviewer to reinterview a household a week after its original interview. The interviewer determines if reports disagree and, in the event if a disagreement, attempts to determine true employment status. The figures in Poterba and Summers Table II imply that 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%. Unfortunately, there is little direct evidence on classification error in the PSID itself, because the PSID validation study, analysed in Bound et al. (1994) only covered a sample of respondents who worked for a single large firm. As all participants were employed, these data cannot be used to assess the probability of falsely reported employment when ones true state is unemployed. However, Bound et al. report that between 29% and 37% of the variance in log hours is noise – see Table 3 panel B. This is suggestive that classification error in employment status is likely to also be important.

force participation rate is .82, while average annual husband's earnings (the proxy for nonlabor income) is \$46,000. The fertility variables are the number of children aged 0-2, 3-5 and 6-17. The last three variables in the table, also used as covariates, are age, the highest level of education attained 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 2004 panel (e.g., the fertility variables). In implementing our estimation procedure, we assume an $AR(1)$ process for the missing time-varying covariate (nonlabor income).

8.2 The Model

The models we fit to married womens' labor force participation decisions are

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

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 and $AR(1)$ errors. We also assume a classification error process where the probability of reporting a particular labor force participation state depends on the true participation status as well as lagged reported status (to allow for the possibility of persistence in misreporting one's state).

Note that the model in (29) is more general than those considered in the repeated sampling experiments because we allow for correlated random effects (the last equation 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. Then 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 that they are uncorrelated with the individual random effects.[22]

In estimating (29), we take the theoretical start of the process, $t = 0$, to be age 16. Of course, most women in the sample are not observed at age 16. To deal with this initial conditions problem, we simulate participation and nonlabor income from age 16 onward. We estimated the model using the smooth algorithm with the importance sampling weights defined in (22). The number of simulated choices for each individual in each time period, M, is set to 250.[23]

8.3 Estimation Results

8.3.1 The Markov Model

Table 19 displays results for four different Markov versions of the 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

[22]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 error terms ξ_{it}.

[23]Setting M=1000, as in the repeated sampling experiments, is not computationally practical for N=1310. Thus, we performed additional repeated sampling experiments with N=1310 and M=250. Biases remain negligible, although standard errors are higher. Biases also remain negligible with changes in N and T.

measured effects of nonlabor income, fertility, age, race and education, and 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 participation. Permanent unobserved heterogeneity is also important in explaining persistence in participation. The estimate $\widehat{\sigma}_\mu = .89$ implies that 79% 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.

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 (participation) when the true state is zero (nonparticipation) increases from 29.9% to 67.7% if participation is reported in the previous period. Similarly, when participation is reported in the previous period, the probability of mis-reporting nonparticipation when the true state is participation 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.[24]

Column (2) reports results for the correlated random effects version of the model. Allowing for correlated random effects produces qualitatively similar point estimates and standard errors 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

[24]The Pearson chi-squared statistic is calculated by computing the frequency of actual and predicted sequences of participation over the seven years of observed choices in the ten-year panel. In order to avoid small cell problems, the number of cells is reduced from 128 (2^7) to 48 by combining "similar" cells. This is the same procedure that Hyslop (1999) employs to evaluate goodness-of-fit.

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. The results show that $AR(1)$ serial correlation is an important component of persistence in female labor force participation, in addition to random effects and first-order state dependence. In Column (3), the point estimate is $\widehat{\phi}_1 = .608$, and it is precisely estimated. However, estimates of other parameters remain qualitatively similar, and the model exhibits only a modest improvement in fit when $AR(1)$ serial correlation is introduced.

Column (4) reports the estimation results for the correlated random effects Markov model in (29), with random effects and $AR(1)$ errors, and where the individual effect is allowed to be correlated with nonlabor income and fertility. Allowing for correlated random effects produces qualitatively similar results to those obtained in Column (3). However, log-likelihood improves by 31 points leading to clear rejection of the null hypothesis $\delta_t = 0$ (i.e., the likelihood ratio χ^2 has a p-value of .0002). 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 displays estimation results for the four different versions of the Polya model that correspond to the four versions of the Markov model in Table 19. If we look at the most general model, 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 $AR(1)$ serial correlation 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 participation is an important determinant of current participation, 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).[25] 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, 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

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

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 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. The estimator was shown to have good small sample properties. Under both the non-smooth and smooth versions of the algorithm, we found that biases are negligible in magnitude even for high amounts of missing information in the data.

The new SML estimation algorithm can also be combined with either Heckman's (1981a) or Wooldridge's (2003) approximate solution to the initial conditions problem. Such a hybrid approach may be appealing when there is no natural starting point to the choice process, and missing data is a problem 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). Overall, it is preferable to simulate choices from the theoretical start of the process if it can be determined.

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 female labor force participation 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 participation outcomes 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.

References

[1] Ackerberg, D. (2001), "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation," unpublished manuscript.

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

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

[4] Chamberlain, G. (1984), "Panel Data," Chapter 22 in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M.D. Intrilligator. Amsterdam: Elsevier Science Publishers B.V.

[5] Dempster, A.P., N.M. Laird and D.B. Rubin (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, B 39, 1-38.

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

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

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

[9] Geweke, J. and M.P. Keane (2001), "Computationally Intensive Methods for Integration in Econometrics," in *Handbook of Econometrics,* Volume V, eds., J.J. Heckman and E. Leamer, Elsevier Science B.V., pp. 3463-3568.

[10] Gould, E. (2007), "Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium," *Review of Economics and Statistics*, 74, 477-506.

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

[12] Heckman, J. (1981*a*), "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 (MIT Press, Cambridge, MA).

[13] Heckman, J. (1981*b*), "Statistical Models for Discrete Panel Data," in C. Manski and D. McFadden, eds., The Structural Analysis of Discrete Data (MIT Press, Cambridge, MA).

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

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

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

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

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

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

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

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

[22] Wooldridge, J.M. (2003), "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics*, forthcoming.

Appendix A

Data Generating Process - Unbiased Classification Error

Defining the initial conditions of the model as $d_{i0} = x_{i0} = 0$, each data set in the repeated sampling experiments is constructed in two stages. In the first stage we generate the exogenous covariates and compute the classification error rates. In the second stage we generate the sequences of true and reported choices (using the error rates computed in the first stage). The second stage also determines if a choice is missing. The two stages of the data generating process are as follows:

Stage 1

1. Draw N sequences from the joint distribution of $(x_{i1}, ..., x_{iT})$ to form $\left\{ \{x_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$.

2. Draw $\widetilde{M}$ times from the joint distribution of $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ to form $\left\{ \left\{ \{\widetilde{\varepsilon}_{it}^{m}\}_{t=1}^{T} \right\}_{i=1}^{N} \right\}_{m=1}^{\widetilde{M}}$. Note that $\widetilde{M}$ will generally differ from the number of simulated choice histories M generated for each individual in estimation.

3. Given $\left\{ \{x_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$ and the error sequence $\left\{ \left\{ \{\widetilde{\varepsilon}_{it}^{m}\}_{t=1}^{T} \right\}_{i=1}^{N} \right\}_{m=1}^{\widetilde{M}}$, construct $\widetilde{M}$ simulated choices for each individual i in every period t $\left\{ \left\{ \left\{ \widetilde{d}_{it}^{m} \right\}_{t=1}^{T} \right\}_{i=1}^{N} \right\}_{m=1}^{\widetilde{M}}$ according to (1) and the decision rule (2).

4. Form the frequency simulator $\widehat{P}\left(\widetilde{d}_{it} = 1 \mid H_{it}^{m} \right) =$ $\frac{1}{\widetilde{M}} \sum_{m=1}^{\widetilde{M}} \Pr\left(\varepsilon_{it} \leq \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} \widetilde{d}_{i\tau}^{m} \rho_\tau \right)$ where $H_{it}^{m} = \left\{ \{x_{i\tau}\}_{\tau=1}^{t}, \left\{ \widetilde{d}_{i\tau}^{m} \right\}_{\tau=1}^{t} \right\}$.

5. Construct the classification error rates π_{jkt} for each individual i, according to (8), using $\widehat{P}$ in place of $\Pr(d_{it} = 1)$.

Stage 2

1. Draw N sequences of errors from the joint distribution of $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ to form $\left\{ \{\varepsilon_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$.

2. Given the $\left\{ \{x_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$ sequence generated in the first stage, and the error sequence $\left\{ \{\varepsilon_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$, construct N true choices $\left\{ \{d_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$ according to (1) and the decision rule (2).

3. In order to construct the sequence of reported choices, draw T times for each individual i from a uniform random number generator to obtain the sequence $\left\{ \{U_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$.

4. Compare the uniform random draws to the classification error rates to determine if choices are correctly reported. That is, construct N reported choices $\left\{ \{d_{it}^{*}\}_{i=1}^{N} \right\}_{t=1}^{T}$ by implementing the following rule: if $d_{it} = 1$ and $U_{it} < \pi_{11t}$ then $d_{it}^{*} = 1$, else $d_{it}^{*} = 0$. Similarly, if $d_{it} = 0$ and $U_{it} < \pi_{00t}$ then $d_{it}^{*} = 0$, else $d_{it}^{*} = 1$.

5. In order to determine if a reported choice is missing, draw T times for each individual i from a uniform random number generator to obtain the sequence $\left\{ \left\{ \widetilde{U}_{it} \right\}_{t=1}^{T} \right\}_{i=1}^{N}$.

6. Compare the uniform draws to the probability π^{obs} that d_{it}^{*} is missing in period t. That is, implement the following rule: if $\widetilde{U}_{it} < \pi^{obs}$ then $I\left(d_{it}^{*}\ observed\right) = 1$, else $I\left(d_{it}^{*}\ observed\right) = 0$.

Note that step 6 does not specify π^{obs} as a function of the exogenous covariates or the observed choices. The data are thus missing completely at random. Generating an initial conditions problem and/or non-randomly missing covariates simply involves modifying π^{obs} accordingly.

Appendix B

Data Generating Process - Biased Classification Error

The data generating process in the case of biased classification error follows the same general rules as in the case of unbiased classification error. The only difference is that the data generating process can be accomplished in one stage rather than two. True choice probabilities do not need to be simulated. The procedure is as follows:

1. Draw N sequences from the joint distribution of $(x_{i1}, ..., x_{iT})$ to form $\left\{ \{x_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$.

2. Draw N sequences of errors from the joint distribution of $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ to form $\left\{ \{\varepsilon_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$.

3. Given $\left\{ \{x_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$ and $\left\{ \{\varepsilon_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$, construct N true choices $\left\{ \{d_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$ according to (1) and the decision rule (2).

4. Draw T times for each individual i from a uniform random number generator to obtain the sequence $\left\{ \{U_{it}\}_{t=1}^{T} \right\}_{i=1}^{N}$.

5. Construct N reported choices $\left\{ \{d_{it}^{*}\}_{i=1}^{N} \right\}_{t=1}^{T}$ by implementing the following rule: if $d_{it} = 1$ and $U_{it} < \pi_{11t}$ then $d_{it}^{*} = 1$, else $d_{it}^{*} = 0$. Similarly, if $d_{it} = 0$ and $U_{it} < \pi_{00t}$ then $d_{it}^{*} = 0$, else $d_{it}^{*} = 1$. The "true" classification error rates π_{jkt} are obtained directly from (11). It is assumed that $d_{i0}^{*} = d_{i0} = 0$.

6. Draw T times for each individual i from a uniform random number generator to obtain the sequence $\left\{ \left\{ \widetilde{U}_{it} \right\}_{t=1}^{T} \right\}_{i=1}^{N}$.

7. Implement the following rule: if $\widetilde{U}_{it} < \pi^{obs}$ then $I\left(d_{it}^{*}\ \mathit{observed}\right) = 1$, else $I\left(d_{it}^{*}\ \mathit{observed}\right) = 0$.

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:

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

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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$. 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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$. 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 (.2701)	-.0330 (1.0164)	.8622	.8878	500
2	.5740	.5700	.4709 (.5272)	-.0220 (1.0525)	.8935	.8565	500
3	.6340	.6280	.8778 (.8917)	-.0146 (.9698)	.9128	.8372	500
4	.6940	.6800	1.1514 (1.1668)	-.0055 (.8593)	.9265	.8235	500
5	.7380	.7420	1.3771 (1.2028)	.0504 (.8507)	.9367	.8133	500
6	.7700	.7840	1.5895 (1.2453)	.0311 (.8962)	.9454	.8046	500
7	.8000	.7960	1.7679 (1.1408)	.0392 (.9582)	.9537	.7963	500
8	.8360	.8620	1.8576 (1.1427)	.0142 (.9893)	.9588	.7912	500
9	.8480	.8260	1.9912 (1.1048)	.0086 (1.0212)	.9640	.7860	500
10	.8600	.8720	2.0187 (.9955)	.0233 (.9182)	.9677	.7823	500

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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. 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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$. 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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. 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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$. 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).