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

Michael P. Keane*and Robert M. Sauer†

June 3, 2007

Abstract

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 panel data probit models with serially correlated errors indicate that the estimator has good small sample properties. We also apply the estimator to a model of female labor force participation decisions using PSID data with serious missing data problems.

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

Keywords: Initial Conditions, Missing Data, Simulation, Female Labor Force Participation Decisions

---

*University of Technology Sydney and Arizona State University (michael.keane@uts.edu.au)

†University of Southampton (rms@soton.ac.uk)

1 Introduction

The problem of unobserved endogenous state variables arises frequently in the estimation of dynamic discrete choice panel data models. The problem is present whenever there are unobserved initial conditions, i.e., the history of the choice process begins prior to the first period of observed data. The problem is also present whenever panel data sets do not contain complete information on all choices for every individual within the sample period. Consistent estimation in either of these cases 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, 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) in order to estimate the parameters of a discrete choice dynamic programming problem with both unobserved initial conditions and missing choice data 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 function 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 performing unconditional simulations of data from a model is often straightforward in contexts where performing conditional simulations would be extremely difficult. Simulation of the likelihood in dynamic models typically requires conditional simulation (of choice probabilities conditional on past history), but when past history is not fully observed, conditional simulation is often computationally infeasible. For example, both the GHK and MCMC algorithms require simulation conditional on a draw from the distribution of past error terms. When the econometrician does not observe the entire past history, the draws required to implement these conditional simulation techniques can be very hard to obtain (see Geweke and Keane (2001)). This was the reason that KW could not use GHK.

In this study, we first describe how the SML algorithm developed by KW, which only requires unconditional simulations, can be extended to a number of cases beyond the specific discrete choice dynamic programming problem that they considered. In particular, we assess the performance of the estimator on simple panel data probit models with a time-varying exogenous covariate, lagged endogenous variables and serially correlated errors. Specification of panel data probit models, rather than discrete choice dynamic programs, allows us to focus on and further develop the estimation technique. Note also that the panel data probit model has been a leading case in past discussions of dynamic panel data models with unobserved initial conditions (see Heckman (1981a)). The results of a series of repeated sampling experiments on the dynamic probit model show that the SML estimator with the new algorithm has good small sample properties.

We also apply the algorithm to dynamic probit models of female labor force participation using PSID data corresponding to calendar years 1994-2003. A serious missing data problem naturally arises in these panel data because respondents are not interviewed in the years 1998, 2000 and 2002. We use the algorithm to test for the endogeneity of children and nonlabor income in dynamic models of labor force participation that include correlated random effects, $AR(1)$ serially correlated errors, and 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. Despite the rich error structure in the model, we reject the null hypothesis that fertility and nonlabor income are exogenous. This result holds in Markov models as well as in models which include a more general process for the effect of lagged endogenous state variables on current utility.

The rest of this paper is organized as follows. Section 2 reviews the literature on different approaches to dealing with the problem of unobserved endogenous state variables, and places our algorithm in context. Section 3 describes the general 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 the estimation algorithm in detail. Section 6 presents Monte-Carlo test results under our first model of classification error (unbiased classification error). Section 7 tests the estimation procedure under our second model of classification error (biased classification error). Section 8 applies the algorithm to a model of female labor force participation. Section 9 summarizes and concludes.

2 Background

Computationally tractable solutions to the initial conditions problem, a particular case of the more general problem of unobserved endogenous state variables, have been proposed before in the econometrics literature. Most notably, Heckman (1981*a*) illustrated how in non-linear dynamic discrete choice settings one could employ the assumption of equilibrium in the dynamic process to derive an expression for the marginal probability of the initial state. Given that equilibrium is a problematic assumption in this context, Heckman (1981*a*) also considered the estimation of fixed effects models, but concluded that it is best to approximate the marginal probability of the initial state by a probit function which has an error term in the initial state index function that is left freely correlated with the errors in the index functions during the sample period. This approximate solution performed better than the fixed effects probit model, but still produced biases of more than 10% in repeated sampling experiments.

More recently, Wooldridge (2003) re-considered Heckman's approximation procedure and proposed an alternative method for handling the initial conditions problem in dynamic, nonlinear random effects models. Wooldridge (2003) suggests conditioning the distribution of unobserved heterogeneity on the initial choice and the observed history of strictly exogenous explanatory variables. This approach is potentially computationally simpler than Heckman's approximation procedure, however, it has not yet been subject to Monte Carlo tests, nor has it been as widely used in empirical applications.

In contrast to the attention given to the initial conditions problem in dynamic panel data models, alternative solutions to the parallel problem of missing data during the sample period have not yet been extensively explored. Missingness problems frequently arise in data sets used by applied economists. For example, serious missing data problems exist in data sets such as the National Longitudinal Study of Youth (NLSY) and the Panel Study of Income Dynamics (PSID).

Perhaps the most widely known proposed solution to the problem of missing data during the sample period is the EM algorithm (see Dempster, Laird and Rubin (1977)). However, in EM it is often difficult to compute the conditional distribution required for the E (expectation) step of the algorithm (see Ruud (1991)). EM has also not been used very much in econometric applications. Another potential solution is the Gibbs-sampling data-augmentation algorithm. Geweke and Keane (2000) used this approach to deal with unobserved initial conditions and missing data in dynamic earnings models. The problem with data augmentation, as with EM, is that the distribution of a missing value conditional on all other information can be quite complex in dynamic models. Also, MCMC techniques can exhibit instability when trying to impute stochastic terms associated with large number of missing outcomes (a problem noted by Geweke and Keane (2000) when they needed to integrate over long pre-sample histories).

Due to these computational difficulties in solving the missing data problem, applied economists frequently resort to the simpler methods of case deletion and imputation. Case deletion, which usually takes the form of cutting the individual's history short, can cause large amounts of information to be discarded, resulting in inefficient estimates. It can also introduce biases to the extent that completely observed histories differ systematically from censored histories. Imputation of missing values by ad hoc methods is no less problematic. Imputing averages tends to bias estimated variances and covariances toward zero while imputing predicted values from regression models tends to bias correlations away from zero. An additional problem is that standard errors of estimates from models with imputed data usually do not reflect the added variability due to the imputations.

In contrast to the previous literature, the SML estimation algorithm that we propose in this paper 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 imputation of missing values, and it is computationally simple. It is computationally simple because it does not require calculation of the initial state probability and the probabilities of events at each date t conditional on the state at the start of time t, which is the usual approach to construction of the likelihood in dynamic models. In our algorithm, unconditional simulations of the model are used to form the likelihood.

The key assumption that is required in order to form the likelihood in dynamic models using only unconditional simulations is that reported choices are measured with error. Assuming classification error in reported choices avoids the need to condition on past history, and avoids the usual problem in frequency simulation whereby an impractically large number of simulations is necessary to compute choice probabilities. Furthermore, the assumption that choices are measured with error is certainly valid in the vast majority of data sets that economists use.

The classification error process that we incorporate into the model simply specifies some probability that the reported choice is the true choice and some probability that it is not. Classification error of this type is frequently present in data sets with discrete outcomes and is popular in applied work (see, e.g., Poterba and Summers (1995) and Flinn (1997)). Moreover, 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)). 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.[1]

In our approach, the investigator has a great deal of flexibility in terms of the details of the classification error process. All that is required is that one can obtain a tractable expression for the probability of observed choices conditional on true choices. One can also specify the classification error process so that it is possible to estimate the extent of classification error in the data. We illustrate this flexibility of the algorithm by considering two different models of classification error in the repeated sampling 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 always unobserved to the researcher but the individual is assumed to choose the option which gives greatest utility. We will consider applications of our SML approach to models of the general form

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

[1] Hausman et. al. (1998) also demonstrate that a distributional assumption on the error term and a monotonicity condition are necessary for separate identification of structural parameters and classification error rates. The dynamic probit models that we consider meet these identification conditions.

where x_{it} is a strictly exogenous covariate[2] and d_{it} is the indicator function defined by

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

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 process.[3] Depreciation in the importance of past choices is captured through the weights ρ_τ. The theoretical start of the process in the dynamic probit model is, by definition, $d_{i0} = 0$.

The error term ε_{it} in (1) is assumed to be serially correlated. Serial correlation in the error term implies that 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

[2] Generalization to endogenous x_{it} is straightforward but requires that one specify the x_{it} process, as would be true in any ML approach.

[3] 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).

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 outlined above may appear somewhat restrictive, it should be noted that the estimation procedure can easily accommodate a wide range of alternative covariate specifications and distributions of the error term. For example, in KW a variant of the algorithm is employed in a multinomial choice setting with an error term that is decomposed into a nonparametric individual effect and a multivariate normal disturbance that is contemporaneously correlated across choices.

While we only consider the 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 do 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 that it is possible to 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

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

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

The investigator has a great deal of leeway in terms of how to further specify the classification error rates π_{11t} and π_{01t}. In our Monte Carlo analysis of the estimation algorithm we will consider cases in which the classification error rates are dependent on the true choice, but are otherwise unconditional on the covariates in the model. Classification error rates would depend on the true value of the dependent variable if, for example, workers who change jobs mis-report more often than workers who do not change jobs. Hausman et. al. (1998) find evidence of this type of misclassification in the PSID and the CPS. In a similar vein, Flinn (1997) finds that the mis-reporting of dismissals in the NLSY is an increasing function of the true dismissal state.

Covariate-dependent misclassification could also be easily incorporated into the classification error model. However, we note that if the measurement error process were made 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 does not necessarily provide guidance for specification of the model of misclassification. For both these reasons, we focus on fairly simple specifications of the classification error process. In what follows, we consider two different 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 classification error rates 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 the person 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 classification error rates in (6) are linear in the true choice probability. To see this, note that by definition,

$$\Pr(d_{it}^* = 1) = \Pr(d_{it}^* = 1 \mid d_{it} = 1)\Pr(d_{it} = 1) + \Pr(d_{it}^* = 1 \mid d_{it} = 0)\Pr(d_{it} = 0) \tag{7}$$

where, in writing $\Pr(d_{it}^* = 1)$ and $\Pr(d_{it} = 1)$, we have suppressed the obvious dependence of these probabilities on x_{it} and lagged true choices in order to conserve on notation.

If we write the classification error rates as the following linear functions of $\Pr(d_{it} = 1)$,

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

then these expressions can be substituted into (7) and shown 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,

$\Pr\left(d_{it}^{*}=1 \mid d_{it}=1\right)$ approaches E. E can thus be interpreted as a "base" classification error 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. E is treated as a free parameter, thus allowing for estimation of the extent of classification error.

In terms of the original notation, the classification error rates can be written as

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

Note the great parsimony that unbiasedness imposes on the classification error process (i.e., it depends on the single parameter E.) However, one could certainly generalize this specification by letting the base classification error rate E depend on covariates. In that case, one obtains unbiasedness conditional on covariates.

Note also that this model of unbiased classification error is similar to the "flexible" model of classification error considered in Hausman et. al. (1998). In both classification error schemes, the probability of the reported choice is increasing in the index function determining the true choice. The monotonicity condition for identification of classification error rates is thus satisfied. This is also true for the model of biased classification error that we consider below.

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 $\Pr\left(d_{it}^{*}=1\right) \neq \Pr\left(d_{it}=1\right)$. 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 a tractable, nonlinear expression for the classification error rates,

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

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).

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 the $\{x_{it}\}_{t=1}^T$ history 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 $I\left(d^*_{it}\ \textit{observed}\right)$ be an indicator function which equals one if d^*_{it} is observed, and zero otherwise. Under these conditions, simulation of the likelihood function requires constructing M simulated choice histories for each $\{x_{it}\}_{t=1}^T$ history as follows:

1. For each individual i, draw M sequences of errors from the joint distribution of $(\varepsilon_{i1}, ..., \varepsilon_{iT})$ to form $\left\{\left\{\{\varepsilon^m_{it}\}_{t=1}^T\right\}_{i=1}^N\right\}_{m=1}^M$.
2. Given $\left\{\{x_{it}\}_{t=1}^T\right\}_{i=1}^N$ and the error sequences $\left\{\left\{\{\varepsilon^m_{it}\}_{t=1}^T\right\}_{i=1}^N\right\}_{m=1}^M$, 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 classification error rates $\left\{\{\widehat{\pi}_{jkt}^m\}_{t=1}^T\right\}_{m=1}^M$ 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 an unbiased 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.

<u>Step $(3a)$</u>:

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 computed by forming 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.[4]

Then $\widehat{\pi}_{11t}^m = \Pr\left(d_{it}^* = 1 \mid d_{it}^m = 1\right)$, and $\widehat{\pi}_{01t}^m = \Pr\left(d_{it}^* = 1 \mid d_{it}^m = 0\right)$ are, respectively,

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

[4] When ε_{it} is distributed normally with mean zero and variance σ_ε^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.

Step (3*b*):

In the special case of 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 the reported choice in the previous period is missing, $d_{i,t-1}^*$ must be simulated. The reported choice in the previous period can be easily simulated according to (10). The simulated $d_{i,t-1}^*$ is denoted as $d_{i,t-1}^{*m}$. 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 $\widehat{\pi}_{11t}^m = \Pr\left(d_{it}^* = 1 \mid d_{it}^m = 1\right)$ and $\widehat{\pi}_{01t}^m = \Pr\left(d_{it}^* = 1 \mid d_{it}^m = 0\right)$ are, respectively,

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

The estimation procedure described in steps (1) through (4) builds the likelihood contribution for each individual by averaging, over M simulated choice histories, the product of the appropriate classification error rates implied by the simulated choice history $\{d_{it}^m\}_{t=1}^T$ and the observed choice history $\{d_{it}^*\}_{t=1}^T$. In step (4) the indicator function $I\left[d_{it}^m = j, d_{it}^* = k\right]$ "picks out" the appropriate classification error rate by comparing d_{it}^* to d_{it}^m. If d_{it}^* is unobserved, then the value of $I\left(d_{it}^* \; observed\right)$ is zero, and there is no contribution to the likelihood (i.e., one simply enters one in the product) in period t.[5]

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

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

the model. The simulation of conditional probabilities like $P(d_{it} \mid H_{it})$ is completely avoided, circumventing the severe computational problems that typically arise if H_{it} is not fully observed. In the unconditional approach, the state space is updated according to previous simulated choices, rather than previous reported choices, which then determine current simulated choices.

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

5.1 Missing Covariates and Initial Conditions

The estimation procedure described above needs to be only slightly modified in order 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 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

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

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

where $f_m\left(x_{it}\right)$ 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,

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

where $\widehat{x}_{it-1}^{(m)} = I\left(x_{i,t-1}\ observed\right) x_{it-1} + \left(1 - I\left(x_{i,t-1}\ observed\right)\right)\widehat{x}_{it-1}^{m}$ and ϕ is the standard normal p.d.f.. Note that in the period in which x_{it} is missing, the density does not affect the likelihood (or one enters the product). $f_m\left(x_{it}\right)$ affects 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 $t = \widetilde{\tau}$.[6]

The likelihood contribution for each individual i in this case takes the form

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

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

In (19), the first d_{it}^* is observed at $t = \widetilde{\tau}$, which is the only difference between (17) and (19). In Heckman's approximation method, one would specify a distribution for

[6] 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 a non-stationarity), one could simulate outcomes accordingly from the two models.

$d^*_{i\widetilde{\tau}}$. 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 economic applications, the process has a natural start date (e.g., age 16 for decisions to stay in school or enter the labor force). In other applications, 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, and still handle the problem of missingness during the sample period. Hybrid approaches such as these will be explicitly considered below.

5.2 Importance Sampling

The estimation procedure can also be easily modified to take advantage of importance sampling techniques that smooth the likelihood function and enable the use of standard gradient methods of optimization.[7] The non-smoothness of the simulated likelihood function 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. We smooth the likelihood by first constructing simulated choice histories $\{d_{it}^m(\theta_0)\}_{t=1}^T$ at an initial θ_0. 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, we have

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

[7] The non-smooth version of the estimation algorithm considered until now necessitates the use of non-gradient methods of optimization such as the downhill simplex method.

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. The joint probability of simulated choice history m in (20) is

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

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. That is, the joint density of simulated choice history m in (22) is

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

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

The likelihood contribution for each individual i in the smooth version of the algorithm is

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

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

Note that (19) is just a special case of (24) with $W_m = 1$ for each simulated choice history m.[8]

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 Classification Error

In this section, Monte-Carlo tests of the SML estimator with unbiased classification error are reported. The algorithm used to generate artificial data sets with unbiased classification error is described in Appendix A. In subsection 6.1, estimation results for a random effects specification are discussed. In subsection 6.2, we discuss the estimation results for an $AR(1)$ specification for the error term. 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 elements of the model. 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 seed for the random elements of the model that generate the M unconditional simulations for each individual in the sample. For each repeated sampling experiment, the true parameters, the mean, the median, the empirical standard deviations, the root mean

[8] The efficiency of importance sampling algorithms is often improved if weights are normalized to sum to one.

square error of the estimates, and the t-statistics for the statistical significance of the biases, based on the empirical standard deviations, are reported.[9]

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 $\theta = \{\beta_0, \beta_1, \rho, \alpha, \phi_2, \sigma_v, \sigma_\mu, E\}$. 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 (1981b).

Table 1 reports summary statistics, by time period and over individuals, for a representative data set produced by the random effects model. The data set is generated with the number of individuals N set to 500, the number of periods T set to 10, no missing choices or missing exogenous covariates, and the vector of true parameters set at $\theta = \{-.10, 1.00, 1.00, .50, .25, .50, .80, .75\}$. For reasons of identification, the variance of ε_{it} is normalized to one, so that $\sigma_\mu^2 + \sigma_\eta^2 = 1$. The normalization implies that the individual effect accounts for 64 percent of the variance in ε_{it} (σ_μ is set to .80).

The Mean d_{it} column in Table 1 shows that there is an increasing proportion of individuals over time that 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

[9] 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.

$\beta' x$ column displays the mean and variance of $\beta' x = \beta_1 x_{it} + \rho \sum_{\tau=0}^{t-1} e^{-\alpha(t-\tau-1)} d_{i\tau}$ and the Mean ε_{it} column displays the mean and variance of the composite error term. The figures show that the mean of $\beta' x$ increases at a decreasing rate reflecting 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. 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 over time 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 missing choices during the sample period. The four panels correspond to data generating processes (DGPs) with no missing choices, 20% missing choices, 40% missing choices 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 the biases are negligible in magnitude. The bias in the estimate of ρ is statistically significant 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 in the parameter estimates 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 (1981a)). Note that the model in the first panel of Table 2, with no missing choices and no initial conditions problem, is difficult to estimate by classical maximum likelihood. Conditional choice probabilities are hard to construct when only lagged reported choices are known and not lagged lagged true choices.

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 not noticeably change the results. Lowering M to 500 also does not change the results, but is 61% faster. The mean time to convergence over the 50 repetitions in the second panel of Table 2 (20% missing choices and $M = 1000$) is 3.73 hours with a standard deviation of .92. The mean time to convergence with 20% missing choices and $M = 500$ is 1.46 hours with a standard deviation of .34. These experiments were run on a desktop computer containing two 1.0 GHz processors and 0.5 GHz RAM.

Table 3 reports the results of three repeated sampling experiments for a modified DGP where the exogenous covariate is missing for the same observations in which the choice is missing. The three panels display the estimation results for 20%, 40%

and 60% missing choices and covariates in each period, respectively. With missing choices and covariates, the parameters of the exogenous covariate process, ϕ_1 and σ_v, are estimated along with the other parameters of the model. 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 parameter estimates 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, but with likelihood contributions from periods $t = 11$ to $t = 20$ only. The biases in the estimates of $\beta_1, \rho, \sigma_\nu$ and σ_μ are statistically significant. However, the magnitudes of the biases are negligible in magnitude. The maximum bias over these four parameters is only 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 $d_{i,10} = 0$. Since there are no missing covariates in this experiment, the parameters of the exogenous covariate process, ϕ_1 and σ_ν, are not estimated. In this case, the biases are generally substantial in magnitude. Note that the standard errors of the estimates of ρ and α increase dramatically and that σ_μ is badly biased upwards. The incorrect treatment of the initial condition results in an overestimate of the importance of individual effects (inflated variance).[10]

[10] 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.

The third panel of Table 4 reports the results of handling the initial conditions problem by constructing a proxy for the initial value of the $\sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau$ term using the observed data. The number of periods in this experiment is increased to $T = 30$. 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 $t = 11, ..., 30$. 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 results in the previous panel, in which the initial conditions problem was ignored, the standard errors of the estimates of ρ and α increase dramatically and the incorrect treatment of the initial condition leads to upward bias in the estimated variance of the random effect. Also, the estimate of the base classification error rate 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 $\beta' x = \beta_1 x_{it} + \rho d_{it-1}$. The first panel of Table 5 reports the results of handling the initial conditions problem by simulating from $t = 0$ and including likelihood contributions from periods $t = 10$ to $t = 20$. 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. In the Markov model, 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 the results of constructing the initial condition by substituting the observed choice in period 10 into the utility function in period 11 (i.e., treating the choice at $t = 10$ as exogenous.) The biases produced in this method are generally less severe than ignoring the initial conditions problem but the bias in the estimate of ρ is substantial in magnitude (14%). As might be expected when treating the initial condition as exogenous, the estimate of ρ is biased upwards.[11]

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. The Heckman 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 coefficient between ε_{it}^H and the individual effect μ_i is $\rho_{\mu\epsilon^H}$. As before, the likelihood function includes contributions from $t = 10, ..., 20$. The parameters γ_0, γ_1 and $\rho_{\mu\epsilon^H}$ are estimated along with the other parameters of the model. We still use our algorithm to accommodate classification error and form the likelihood using only unconditional simulations from $t = 10, ..., 20$. In effect, we are nesting Heckman's procedure for handling the initial period within our algorithm.

The estimation 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 the 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

[11] In the $AR(1)$ error model to be discussed below, treating the initial condition as exogenous produces a bias in the estimate of ρ which is considerably larger (23%).

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. On the other hand, Wooldridge's method yields a better estimate of σ_μ than does Heckman's method.[12]

In order to illustrate that our algorithm can be useful even when there is no classification error at all in the data, Table 6 reports the results of three repeated sampling experiments when there is no initial conditions problem and no classification error in the DGP. The three panels display the estimation results for the full model (i.e., the generalized Polya Process) with 20%, 40% and 60% missing choices and covariates in each period, respectively. The results indicate negligible biases that never exceed 5%. The only effect of incorporating classification error in the algorithm when there no classification error in the DGP is to increase somewhat the standard

[12] 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.

errors (compare Tables 3 and 6). Note that the mean estimate of E tends towards the upper bound of one.

6.1.2 The Smooth SML Algorithm (Importance Sampling)

The smooth version of the estimation algorithm, differs from the non-smooth version in that the former requires simulated choice histories to be generated only once for each individual in the sample, at the initial vector of trial parameters. The smooth version enables the use of standard gradient methods of optimization as opposed to generally more time consuming non-gradient methods of optimization such as the downhill simplex method. Thus, the smooth version of the algorithm should be faster to converge. 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), and that 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. Table 7 reveals a few statistically significant biases, but the biases are trivial.

It is also important to note that there is a large difference in mean time to convergence between the smooth and non-smooth algorithms in these experiments. 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. The mean time to convergence over the 50 repetitions in the first panel of Table 6 (20% missing choices) is 1.94 hours with a standard deviation of .97. The smooth version is 47% faster.

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 and over individuals, for a representative data set produced by the $AR(1)$ error model. The data set is generated with the number of individuals N set to 500, the number of periods T set to 10, no missing choices or missing covariates, and the vector of true parameters set at $\theta = \{-.10, 1.00, 1.00, .50, .25, .50, .80, .75\}$. 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 $\widetilde{M}$ 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 the order of experiments on 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 $AR(1)$ error model as it does in the random effects model. The second panel, in which the initial conditions problem is ignored, reveals serious biases. As in the corresponding experiment on the random effects model, in which the standard deviation of the individual effect is substantially biased upward, the $AR(1)$ error parameter is substantially over-estimated. 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 when using observed data to form a proxy for the initial value of the $\sum_{\tau=0}^{t-1} d_{i\tau}\rho_\tau$ term. The magnitudes of the biases when using this approach are generally smaller in the $AR(1)$ error model than in the random effects model. However, as in the random effects model, the estimates of ρ and α are biased upward.

Table 11 examines different solutions to the initial conditions problem in the Markov model. 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. Treating the initial condition as exogenous (panel 3) and 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.[13]

6.2.2 The Smooth SML Algorithm (Importance Sampling)

Table 12 reports the results of estimating the $AR(1)$ error model with missing exogenous covariates and no initial conditions problem using the smooth SML algorithm with the weights in (22). As in the random effects model, with 20% missing choices

[13] The Wooldridge approach is not estimated in the $AR(1)$ case because it was developed specifically for a random effects model. The Wooldridge method works by forming an expectation of the random effect conditional on the initial condition and the information on the exogenous covariates.

and covariates in each period, the biases are negligible in magnitude. Consistent with previously reported results for the random effects model, the $AR(1)$ error model also converges much faster when using the smooth algorithm. The mean time to convergence over the 50 repetitions in the first panel of Table 9 (20% missing choices) is 3.07 hours with a standard deviation of .71. The mean time to convergence over the 50 repetitions in the first panel of Table 12 (20% missing choices) is 1.84 hours with a standard deviation of .72. The $AR(1)$ error model converges slightly faster than the random effects model.

7 Monte-Carlo Tests - Biased Classification Error

In this section, Monte-Carlo tests of the SML estimator with biased classification error, as specified in (11), are performed. The algorithm used to generate artificial data sets with biased classification error is described in Appendix B. In subsections 7.1 and 7.2, estimation results for the Polya process random effects and $AR(1)$ error models are discussed, respectively. In subsection 7.3, we consider the estimation results for the Polya process model with combined random effects and $AR(1)$ errors.

7.1 Random Effects Model

7.1.1 Non-Smooth SML Algorithm

The three panels of Table 13 report the results of estimating with random effects and biased classification error using the non-smooth SML algorithm. 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 classification error bias. 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 classification error rates $\pi_{11t} = \Pr(d^*_{it} = 1 \mid d_{it} = 1)$ and $\pi_{01t} = \Pr(d^*_{it} = 1 \mid d_{it} = 0)$ are $(.97, .18)$, $(.95, .27)$ and $(.95, .50)$, in the first, second and third panels, respectively.

The estimation results indicate relatively few statistically significant biases. Only the estimates of ρ and σ_v are consistently biased. However, the magnitudes of these biases are negligible, never exceeding 3 percent. Note that increases in classification error bias leads to larger empirical standard errors. The more classification error bias, the less efficient are the estimates.

In general, the algorithm seems to perform very well for the DGPs with biased classification error, both in terms of uncovering the structural parameters and in terms of uncovering the parameters of the classification error process. The algorithm with a high extent of classification error bias and 20% missing choices and covariates converges in similar time to the corresponding specification with unbiased classification error. The time to convergence per parameter is .54 hours in the former case and .57 hours in the latter.[14]

7.1.2 Smooth SML Algorithm

Table 14 reports the results of estimating the random effects 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.

[14]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.

7.2 $AR(1)$ Error Model

7.2.1 Non-Smooth SML Algorithm

The three panels in Table 15 repeat the series of repeated sampling experiments in Table 13 with an $AR(1)$ specification for the error term rather than a random effects specification. The results in all three panels 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 the results of estimating the same experiment as in Table 14 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 + $AR(1)$ Error Model

In the random effects + $AR(1)$ error model, the error term ε_{it} follows the error process in (5). As in the random effects only model, the true σ_μ is set to .80. However, the $AR(1)$ error parameter ϕ_1 is lowered from its value in the $AR(1)$ errors only model from $\phi_1 = .80$, to $\phi_1 = .40$.

7.3.1 Smooth SML Algorithm (Importance Sampling)

Table 17 reports the results of estimating the combined error model using the smooth algorithm with the weights in (22). 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 negligible 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 the corresponding biases are 2.1% and 2.3% (see Table 16). Thus, the combined error model is not only more general; it also yields smaller biases and standard errors than estimating with only random effects.

8 Application to Female Labor Force Participation Decisions

In this section, we apply the smooth algorithm, with the importance sampling weights defined in (22), to a model of female labor force participation, using PSID data corresponding to the calendar years 1994-2003. Because respondents are not interviewed every year during the sample period (in 1998, 2000 and 2002), there are both missing choices (missing endogenous state variables) and missing covariates, in addition to an initial conditions problem. As in the repeated sampling experiments, we estimate dynamic probit models with these data and imbed an $AR(1)$ process for the missing time-varying covariate (nonlabor income). We use the estimates of the models to test for the endogeneity of fertility and nonlabor income in married women's labor supply behavior.

8.1 The Data

The data used in estimation are drawn from the 2004 panel of the Panel Study of Income Dynamics (PSID), including both the random Census subsample of families and nonrandom Survey of Economic Opportunities. The sample corresponds to the ten calendar years 1994-2003. Restricting the sample to these years produces a panel with serious missing data problems and a panel length which is the same as in the repeated sampling experiments. The missing data problem arises because the PSID switched from an annual to a biannual survey after the 1997 wave was completed.

Hence, PSID families were not interviewed in 1998, 2000, and 2002.[15]

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

Table 18 presents selected means, standard deviations and standard errors in the estimation sample. The labor market participation rate of .82 is calculated by computing a participation rate over the seven non-missing years for each woman, and then averaging this rate over the 1310 women in the sample. Average annual husband's earnings (the proxy for nonlabor income) is calculated in a similar way. The fertility variables appearing in the table, used as covariates in estimation, are the number of children aged 0-2 years, the number of children aged 3-5 years and the number of children aged 6-17 years. The means of these latter variables are computed by averaging over all ten sample years. The last three variables in the table, which are also used as covariates in estimation, are age, the highest level of education attained over the sample period (which is then held constant from 1994-2003), and race (which equals one if black). All covariates except nonlabor income are available for the full ten years because they are either not time-varying (age, education, race) or easily re-constructed from information in the 2004 panel (the fertility variables).

[15] 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).

8.2 The Model

The Markov and Polya models we use to explain married woman i's labor force participation decision in year t are

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

where y_{it} is the husband's annual earnings in year t, and X_{it} is a vector containing the fertility, race and education covariates, in addition to year effects. The error structure for both the Markov and Polya models is random effects with $AR(1)$ serially correlated transitory 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.

Note that the model in (29) expands upon the models considered in the repeated sampling experiments because we allow for correlated random effects (the last equation in (29)). Estimating with correlated random effects enables one to test for the endogeneity of fertility and nonlabor income by formulating the null hypothesis H_0: $\delta = 0$. W_{it} in the correlated random effects equation contains $\ln(y_{it})$ and the three fertility variables in time t.

The initial conditions problem that arises in the dynamic models in (29) is solved by simulating participation outcomes and nonlabor income realizations from the theoretical start of the process, $t = 0$, which we take to be age 16. The models are estimated 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 250.[16]

8.3 Estimation Results

8.3.1 Markov Model

Table 19 displays the estimation results for four different versions of the Markov model. Column (1) reports point estimates and asymptotic standard errors for a restricted version of the Markov model in (29), where there are random effects but no $AR(1)$ serial correlation in transitory errors nor correlated random effects (i.e., $\phi_1 = 0$ and $\delta = 0$). The results show precisely measured effects of nonlabor income, fertility, age, race and education. The signs of the coefficients and the relative magnitudes of the fertility effects are all in the expected directions.

The estimate $\widehat{\rho} = 2.31$ in Column (1) implies strong positive state dependence in female participation decisions. The estimate of σ_μ suggests that permanent unobserved heterogeneity is also important in explaining persistence in participation. $\widehat{\sigma}_\mu = .89$ implies that 89% of the total error variance is due to the individual effect. Note that the estimated $AR(1)$ coefficient in the nonlabor income process, $\widehat{\phi}_2 = .999$, implies that the nonlabor income process is a random walk. The unit root is perhaps not surprising given that nonlabor income must be simulated a large number of periods for many women (back to age 16). The mean age in the sample is 37.

The estimates of the classification error process, γ_0, γ_1 and γ_2, indicate that classification error in reported choices is not negligible, and that reported choices in the previous period have a significant influence on misclassification rates in the current

[16] Setting M=1000, as in the repeated sampling experiments, is computationally practical for N=500, but not for N=1310. Additional repeated sampling experiments with $N = 1310$ and M=250 were performed. Biases remain negligible, although standard errors are higher. Biases also remain negligible with changes in N and T.

period. 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 a zero (nonparticipation) when the true state is one (participation) falls from 7.3% to 1.6%.

Column (2) reports the estimation results for the correlated random effects version of the model, without $AR(1)$ serial correlation in transitory errors. Allowing for correlated random effects produces qualitatively similar point estimates and standard errors to those obtained in Column (1). However, there is a large improvement in the log-likelihood which leads to the rejection of the null hypothesis $\delta = 0$. The likelihood ratio chi-square statistic is 44.58 with 27 degrees of freedom. It has a p-value of .0243. Thus, we find clear evidence that fertility and nonlabor income are not exogenous within a random effects probit model containing a first-order Markov process in past choices.

Column (3) reports the estimated parameters and standard errors when the error structure is random effects + $AR(1)$ serially correlated transitory errors, but $\delta = 0$. The additional parameter that is estimated in this specification, relative to Column (1), is ϕ_1. The results show positive $AR(1)$ serial correlation. $AR(1)$ serial correlation is an important component of persistence in female labor force participation, in addition to individual effects and first-order state dependence. The point estimate $\widehat{\phi}_1 = .608$ is precisely estimated. However, the Pearson chi-square statistic does not reveal a substantial improvement in fit with the introduction of $AR(1)$ serial correlation.[17]

---

[17]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. This is the

Column (4) reports the estimation results for the full Markov model in (29), where there are random effects, $AR(1)$ serially correlated errors, and the individual effect is allowed to be correlated with nonlabor income and fertility outcomes. Allowing for correlated random effects produces qualitatively similar results to those obtained in Column (3). However, there is a large improvement in the log-likelihood which once again leads to rejection of the null hypothesis $\delta = 0$. The likelihood ratio chi-square statistic is 62.08 with a p-value of .0002. Note that the fit of the model improves more substantially when estimating with correlated random effects and $AR(1)$ serial correlation. The Pearson chi-squared statistic falls from 58.32 with a p-value of .1245, to 56.40 with a p-value of .1637.

8.3.2 Polya Model

Table 20 displays estimation results for four different versions of the Polya model, corresponding to the four versions of the Markov model in Table 19. Comparing Column (1) of Table 20 with Column (1) of Table 19 shows that the Polya model produces stronger effects of nonlabor income and age but a weaker effect of race. The fertility effects, the estimated variance of the random effect, and the estimates of the $AR(1)$ process for nonlabor income are similar in the two models.

The Polya process estimates, $\widehat{\rho}$ and $\widehat{\alpha}$, imply that past participation choices are important determinants of current decisions but the influence of past choices falls quickly over time. For example, when $d_{i,t-1} = 1$, u_{it} increases by .6363 (the point estimate of ρ), holding all else constant. This is in contrast to an increase in u_{it} of 2.3148 in the Markov model. In the Markov model $d_{i,t-2} = 1$ has no effect on u_{it}, while in the Polya model, $d_{i,t-2} = 1$ increases u_{it} by .0959. Moving further into the past, when $d_{i,t-3} = 1$, increases u_{it} by increases .0145, and when $d_{i,t-4} = 1$, u_{it} increases by only .0022.

The estimates of the classification error parameters in Column (1) imply that

same procedure that Hyslop (1999) employs to evaluate goodness-of-fit.

the Polya model produces lower classification error rates than the Markov model. The estimated misclassification 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 $d^*_{i,t-1} = 1$. Note that the Polya model yields are large improvement in the log-likelihood (105.51 points) compared to the Markov model, with only one additional parameter (α). The Polya model also fits the data considerably better than does the Markov model. The Pearson chi-squared statistic in the Polya model is 54.62 with a p-value of .2075 (compared to 59.52 with a p-value of .1024 in the Markov model).

Column (2) reports the results of the correlated random effects Polya model. Note that the null hypothesis of exogenous fertility and nonlabor income is once again rejected. The likelihood ratio test statistic is 46.42 with a p-value of .0158. Column (3) adds $AR(1)$ serially correlated transitory errors to the random effects error structure in Column (1). The estimated $AR(1)$ coefficient is smaller than in the corresponding Markov specification (.4606 compared to .6084) but is still substantially positive and precisely estimated. Column (4) reports the results of the full Polya specification. The results are qualitatively similar to those in Column (3). Importantly, the null hypothesis of exogenous fertility and nonlabor income is still rejected and the fit of the model improves substantially.

Despite the rich error structure, we find that in both Markov and Polya models for the influence of past participation choices, one can reject the null hypothesis of exogenous fertility and nonlabor income. This is in contrast to findings in Hyslop (1999). The discrepancy in results is mainly due to the fact that we correct for classification error within our SML estimation algorithm (see Keane and Sauer (2006) for more details).

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 simulations 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 and MCMC 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 also be determined in estimation.

The SML estimation algorithm was tested via a series of repeated sampling experiments on a general 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 easily combined with Heckman's (1981*a*) approximate solution and Wooldridge's (2003) alternative solution to the initial conditions problem. Such a hybrid approach may be appealing when there is no natural starting point to the choice process, but it is necessary to integrate over missing information during the sample period. Heckman's approximate solution method was found to work better than Wooldridge's approach in our experiments with a random effects model. Heckman's approximate solution 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 when the start of the process can be determined.

Interestingly, our SML algorithm seems to perform better (in terms of consistently producing negligible bias) for models with biased classification error than for models with unbiased classification error. In order to impose the constraint that classification error be unbiased, one must specify that classification error rates are functions of true choice probabilities. This means that classification error rates must themselves be simulated. This induces some additional noise and computation time into the likelihood simulation. In contrast, with biased classification error, one can specify that classification error rates are closed form functions of true choices (and perhaps also lagged observed choices and covariates). Our algorithm already requires that true choice histories be simulated, so once this has been done, no additional simulation is necessary to form the classification error rates. This saves computation time and avoids one component of simulation error.

We also apply the algorithm to dynamic probit models of female labor force participation using PSID data corresponding to calendar years 1994-2003. A serious missing data problem naturally arises in these data because respondents are not interviewed in the years 1998, 2000 and 2002. We solve the initial conditions problem that arises in this context 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 using the smooth version of the algorithm and assuming persistent classification error. The empirical application shows that nonlabor income and fertility remain endogenous even in dynamic probit models with rich error structures.

Future research will examine the small sample properties of the estimation technique in more complex dynamic probit settings. 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, and 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] 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.

[3] 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.

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

[5] 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.

[6] 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.

[7] 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.

[8] 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).

[9] 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).

[10] 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.

[11] 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.

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

[13] 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.

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

[15] 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

Given a vector of true parameters of the model and 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. The first stage consists of generating the exogenous covariates and computing the "true" classification error rates. The second stage consists of generating the sequence of true choices and misclassified choices, using the true classification rates computed in the first stage. The second stage also determines if a choice is missing from the data. 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 "true" 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 missing exogenous covariates as well as data that is missing at random or missing not at random 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 $I\left(d^*_{it}\ observed\right) = 1$, else $I\left(d^*_{it}\ observed\right) = 0$.

Table 1

Summary Statistics
Representative Data Set
Polya Model
Random Effects
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
Polya Model
Random Effects
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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$. The model is the same as in Table 1.

Table 3

Repeated Sampling Experiments
Polya Model
Random Effects
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
Polya Model
Random Effects
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
Markov Model
Random Effects
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
Use reported data at $t = 10$ to proxy for initial condition at $t = 11$ $(t = 10, ..., 20)$						
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
Markov Model
Random Effects
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 $t = 11$ $(t = 10, ..., 20)$						
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 $(t = 10, ..., 20)$						
$\widetilde{\beta}_0$	-.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
Polya Model
Random Effects
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	-	.9249	.9290	.0566	-	-
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	-	.9180	.9230	.0496	-	-
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	-	.9366	.9341	.0698	-	-

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

Table 7

Repeated Sampling Experiments
Polya Model
Random Effects
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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$. The model is the same as in Table 1.

Table 8

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

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

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

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

Table 9

Repeated Sampling Experiments
Polya Model
$AR(1)$ 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
$AR(1)$ 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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$. The model is the same as in Table 8.

Table 11

Repeated Sampling Experiments
Markov Model
$AR(1)$ 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
Use reported data at $t = 10$ to proxy for initial condition at $t = 11$ $(t = 10, ..., 20)$						
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
$AR(1)$ 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
Polya Model
Random Effects
Biased Classification Error
(20% Missing Choices and X's, No Initial Conditions Problem)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$\widehat{Std(\beta)}$	$RMSE$	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. $\widehat{Std(\beta)}$ 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 14

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

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

Table 15

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

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

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

Table 16

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

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$\widehat{Std(\beta)}$	$RMSE$	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. $\widehat{Std(\beta)}$ 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
Random Effects + $AR(1)$ 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 $\sqrt{50}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}\left(\widehat{\beta}\right)}\right)$.
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 annual earnings and hours worked each non-missing year for both partners in the married couple. 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
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 (.0001)	-.1510 (.0001)	-.1697 (.0001)	-.1646 (.0000)
$\#kids0\text{-}2_t$	-.6433 (.0015)	-.5382 (.0026)	-.6659 (.0003)	-.4271 (.0005)
$\#kids3\text{-}5_t$	-.3342 (.0004)	-.3524 (.0006)	-.3650 (.0003)	-.3379 (.0004)
$\#kids6\text{-}17_t$	-.0845 (.0014)	-.0830 (.0023)	-.0808 (.0001)	0.0734 (.0004)
$age_t/10$	.6676 (.0033)	.5818 (.0040)	.6887 (.0002)	.6792 (.0002)
$age_t^2/100$	-.1438 (.0001)	-.1364 (.0000)	-.1525 (.0000)	-.1565 (.0000)
$race_i$	.5547 (.0007)	.5467 (.0007)	.4518 (.0003)	.4533 (.0003)
$education_i$	.0501 (.0001)	.0407 (.0008)	.0581 (.0000)	.0392 (.0000)
ρ	2.3148 (.0015)	2.3582 (.0015)	2.4047 (.0007)	2.5099 (.0007)
ϕ_2	.9993 (.0002)	.9993 (.0002)	.9992 (.0001)	.9993 (.0000)
σ_ν	.2719 (.0006)	.2718 (.0006)	.2758 (.0006)	.2755 (.0004)
σ_μ	.8947 (.0012)	.8949 (.0012)	.8877 (.0003)	.8905 (.0003)
γ_0	-.8535 (.0541)	-.9716 (.0500)	-0.8346 (.0425)	-.9454 (.0471)
γ_1	3.3974 (.0624)	3.4328 (.0639)	3.6335 (.0714)	3.5653 (.0685)
γ_2	1.5943 (.0924)	1.6178 (.0940)	1.7012 (.0937)	1.6734 (.0947)
ϕ_1	-	-	.6084 (.0004)	.6136 (.0006)
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:

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

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
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 (.0001)	-.3111 (.0001)	-.3066 (.0001)	-.3040 (.0001)
$\#kids0\text{-}2_t$	-.5964 (.0030)	-.6000 (.0028)	-.6495 (.0004)	-.6339 (.0004)
$\#kids3\text{-}5_t$	-.3648 (.0006)	-.3565 (.0007)	-.3325 (.0003)	-.3466 (.0003)
$\#kids6\text{-}17_t$	-.0145 (.0011)	-.0123 (.0019)	-.0211 (.0003)	-.0225 (.0011)
$age_t/10$	.7527 (.0031)	.5387 (.0040)	.7081 (.0001)	.7263 (.0001)
$age_t^2/100$	-.1310 (.0000)	-.1074 (.0001)	-.1262 (.0000)	-.1280 (.0000)
$race_i$	.3083 (.0006)	.2272 (.0005)	.2945 (.0002)	.2684 (.0002)
$education_i$	.0652 (.0000)	.0558 (.0000)	.0630 (.0000)	.0611 (.0000)
ρ	.6363 (.0005)	.7281 (.0008)	.6758 (.0003)	.6979 (.0003)
α	1.8924 (.0020)	1.9502 (.0019)	2.1278 (.0020)	2.1457 (.0011)
ϕ_2	.9994 (.0002)	.9994 (.0002)	.9994 (.0001)	.9994 (.0001)
σ_ν	.2743 (.0003)	.2736 (.0004)	.2742 (.0002)	.2736 (.0002)
σ_μ	.8949 (.0013)	.8970 (.0013)	.8952 (.0003)	.8960 (.0003)
γ_0	-1.1203 (.0482)	-.8962 (.0457)	-.9940 (.0457)	-.9404 (.0463)
γ_1	3.8880 (.0761)	3.6738 (.0769)	3.6809 (.0760)	3.7190 (.0781)
γ_2	1.6520 (.0985)	1.5320 (.0977)	1.5658 (.0974)	1.6096 (.0987)
ϕ_1	-	-	.4606 (.0004)	.4596 (.0005)
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:

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

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).