

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

Michael P. Keane and Robert M. Sauer\*

September 22, 2003

## **Abstract**

This paper develops a new simulation estimation algorithm that is particularly useful for estimating dynamic panel data models with unobserved endogenous state variables. The new approach can deal with the commonly encountered and widely discussed “initial conditions problem,” as well as the more general problem of missing state variables at any point during the sample period. Repeated sampling experiments on a dynamic panel data probit model with serially correlated errors indicate that the estimator has good small sample properties and is computationally practical for use with panels of the size that are likely to be encountered in practice.

JEL Nos: C15, C23, C25

Keywords: Initial Conditions, Missing Data, Simulation Estimation

---

\*Mailing addresses: Department of Economics, Yale University, Box 208264, New Haven, CT 06250-8264, USA (Michael.Keane@yale.edu) and Department of Economics, Hebrew University of Jerusalem, Mt. Scopus, Jerusalem, Israel (mssauer@mscc.huji.ac.il), respectively.

# 1 Introduction

The problem of unobserved endogenous state variables frequently arises 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 all of the 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 assess the performance of 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 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. Simulation of the likelihood in dynamic models typically requires *conditional* simulations (of choice probabilities conditional on past history). The advantage of our algorithm lies in the fact that unconditional simulation is often straightforward in contexts where conditional simulation is computationally infeasible (as is often the case when past history is not fully observed).

In this paper, we describe how the SML algorithm developed by KW 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 esti-

mator on a panel data probit model with a time-varying exogenous covariate, lagged endogenous variables and serially correlated errors. 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)). Specification of a panel data probit model also allows us to focus on and further develop the estimation technique. 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 and is computationally practical for use with panels of the size that are likely to be encountered in practice.

Computationally tractable solutions to the problem of unobserved endogenous state variables have been proposed before in the econometrics literature. Most notably, Heckman (1981a) illustrated how, in a dynamic probit setting, one could usefully employ the simplifying assumption of equilibrium in the dynamic process and derive an expression for the marginal probability of the initial state. This marginal probability could then be incorporated into the likelihood function for consistent estimation. A drawback of this method, however, is that equilibrium implies the process has been in operation far into the past and that the exogenous variables are generated by a stationary stochastic process. Time and age trends must, therefore, be excluded from the set of explanatory variables.<sup>1</sup>

Heckman (1981a) also considered the estimation of fixed effects models as a potentially attractive alternative solution to the initial conditions problem. Estimation of a fixed effects model obviates the need to incorporate pre-sample information and the need to commit to a particular mixing distribution for individual effects. However, in nonlinear models of fixed panel length, the inconsistency of the fixed effects estimator is transmitted to the structural parameters. Moreover, the effective sample size for estimating the structural parameters of the model is the subsample of individuals that

---

<sup>1</sup>Card and Sullivan (1988) adopted this method for studying the dynamic effects of training on re-employment probabilities.

change state. This latter fact can easily lead to a form of small sample selection bias. This approach is also difficult to implement if the number of cross-sectional units is large.<sup>2</sup>

As a better alternative to assuming equilibrium or estimating a fixed effects model, Heckman (1981*a*) suggested approximating the marginal probability of the initial state by a probit function which has as its argument as much presample information on the exogenous variables as is available. The error term in the initial state index function can also be left freely correlated with the errors in the index functions during the sample period. This latter estimation strategy was shown to perform better than the fixed effects probit model in Monte Carlo simulations. However, this approximation procedure can still produce biases in structural parameters that are relatively large in magnitude.<sup>3</sup>

In stark contrast to the attention given to the initial conditions problem in dynamic panel data models, alternative practical solutions to the parallel problem of missing data during the sample period have not yet been fully 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). Although Rubin (1976), Dempster, Laird and Rubin (1977) and Little and Rubin (1987) have extensively analyzed the EM algorithm as a general solution to missingness problems, the EM algorithm has important limitations and has not been used very much

---

<sup>2</sup>In the special cases of the linear probability model and the logit model, fixed effects can be eliminated and structural parameters consistently estimated. For analyses of the conditional logit model see Chamberlain (1984) and Honore and Kyriazidou (2000).

<sup>3</sup>Monte Carlo experiments in Heckman (1981*a*) and Hyslop (1999) often produce estimated parameters with biases of more than 10%. Heckman and Singer (1984) suggested a similar type of solution to the initial conditions problem in multiple spell duration models. See Ham and LaLonde (1996) for an application.

in practice.<sup>4</sup>

When faced with missing data problems, applied economists frequently resort to the simpler methods of case deletion and imputation. Case deletion, which in the context of panel data models usually takes the form of cutting the individual's history short, is a questionable solution. It can cause large amounts of information to be discarded, resulting in inefficient estimates. Case deletion can also introduce biases to the extent that completely observed histories differ systematically from censored histories. In other words, biases are introduced when the data is not missing completely at random, or the missingness process is nonignorable (see Little and Rubin (1987)). 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 (see Schafer (1997)). An additional problem is that standard errors of estimates from models with imputed data usually do not reflect the additional variability due to the imputations.

The SML technique that we analyze in this paper offers a systematic unified “solution” to both the initial conditions problem and the problem of missingness during the sample period. Our proposed solution does not involve the censoring of incomplete histories or the use of ad hoc imputation methods. A key feature of the estimation technique is that we use unconditional simulations of the model to form the likelihood. This is in contrast to the usual approach to construction of the likelihood in dynamic models which requires calculation of the initial state probability and the probabilities of events at each date  $t$  conditional on the state at the start of time  $t$ .

---

<sup>4</sup>Analogous to the EM algorithm 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 EM and data augmentation is that the distribution of a missing value conditional on all other information can be quite complex in dynamic models (see also Ruud (1991)).

The key assumption that is required to form the likelihood in dynamic models with 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 has been used in applied work (see, e.g., Poterba and Summers (1995)). Moreover, if misclassification of this type 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. In our approach, the investigator has a great deal of flexibility in terms of the details of the specification of the classification error process, as we illustrate by considering two alternative specifications.<sup>5</sup>

The rest of this paper is organized as follows. Section 2 describes the general dynamic panel data probit model used in the analysis. Section 3 develops two different models of classification error that can be incorporated into the estimation algorithm. We call the two different models of classification error unbiased and biased classification error, respectively. Section 4 describes the estimation algorithm. Section

---

<sup>5</sup>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.

5 presents Monte-Carlo test results, assuming unbiased classification error, for both a random effects model and an  $AR(1)$  error model. Section 6 tests the estimation procedure under the biased classification error scheme. Section 7 summarizes and concludes.

## 2 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

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

where  $d_{it}$  is the indicator function defined by

$$d_{it} = \begin{cases} 1 & \text{if } u_{it} \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

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

The error term  $\varepsilon_{it}$  in (1) is assumed to be serially correlated. Serial correlation in the error term implies that lagged choices are endogenous. The source of serial correlation could be time-invariant individual effects, i.e.,

$$\varepsilon_{it} = \mu_i + \eta_{it} \quad (3)$$

where  $\mu_i$  is i.i.d. with zero mean and variance  $\sigma_\mu^2$ . Alternatively, serial correlation could derive from an  $AR(1)$  process,

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

where  $\eta_{it}$  is i.i.d. with zero mean and variance  $\sigma_\eta^2$ .

Although the model outlined above may appear somewhat specific, it is important to emphasize that the estimation procedure can accommodate a wide range of covariate specifications and distributions of the error term. In KW and Sauer (2003), the estimation technique is employed in a multinomial setting in which the error term is decomposed into a nonparametric individual effect and a multivariate normal i.i.d. disturbance.

### 3 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 model of misclassification that we consider is characterized by four classification error rates

$$\begin{aligned} \pi_{11t} &= \Pr(d_{it}^* = 1 \mid d_{it} = 1) \\ \pi_{01t} &= \Pr(d_{it}^* = 1 \mid d_{it} = 0) \\ \pi_{00t} &= 1 - \pi_{01t} \\ \pi_{10t} &= 1 - \pi_{11t} \end{aligned} \quad (5)$$

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

The investigator has a great deal of leeway in terms of how to specify the classification error rates  $\pi_{11t}$  and  $\pi_{01t}$ . In our Monte Carlo analysis of the estimation algorithm we will only 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 more often misreport than workers who do not change jobs. Hausman et. al. (1998) find evidence of this type of misclassification in the PSID and the Current Population Survey.

Covariate-dependent misclassification can 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 will focus on fairly simple specifications of the classification error process. A key distinction is whether classification error is biased or unbiased. We consider these cases in turn.

### 3.1 Unbiased Classification Error

The assumption that classification error is unbiased imposes a very simple structure on the classification error rates in (5). 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  $\Pr(d_{it}^* = 1) = \Pr(d_{it} = 1)$ . We find the assumption of classification error quite 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 (5) are linear in the true choice probability. To see this, note that by definition,

$$\Pr(d_{it}^* = 1) = \Pr(d_{it}^* = 1 \mid d_{it} = 1) \Pr(d_{it} = 1) + \Pr(d_{it}^* = 1 \mid d_{it} = 0) \Pr(d_{it} = 0) \quad (6)$$

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 choices in order to conserve on notation.

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

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

then these expressions can be substituted into (6) 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,  $\Pr(d_{it}^* = 1 \mid d_{it} = 1)$  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. In estimation,  $E$  is treated as a free parameter.

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

$$\begin{aligned} \pi_{11t} &= E + (1 - E) \Pr(d_{it} = 1) \\ \pi_{01t} &= (1 - E) \Pr(d_{it} = 1). \end{aligned} \quad (8)$$

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.

### 3.2 Biased Classification Error

Any classification error scheme that doesn’t impose the linear relationships in (7) will, in general, lead to a biased classification error process in which  $\Pr(d_{it}^* = 1) \neq \Pr(d_{it} = 1)$ . The biased classification error scheme that we consider is characterized by the following index function,

$$l_{it} = \gamma_0 + \gamma_1 d_{it} + \gamma_2 d_{it-1}^* + \omega_{it} \quad (9)$$

where  $d_{it}^*$  denotes the reported choice and  $\omega_{it}$  is a stochastic term. If  $l_{it} > 0$  then  $d_{it}^* = 1$ , while  $d_{it}^* = 0$  otherwise. Notice that the specification in (9) allows the probability of reporting a particular choice to differ by the true choice, and also allows for persistence in misreporting. The greater in magnitude is  $\gamma_2$ , the more likely is persistent misreporting.

Assuming  $\omega_{it}$  is distributed logistically yields a tractable, nonlinear expression for the classification error rates,

$$\begin{aligned} \pi_{11t} &= \Pr(d_{it}^* = 1 \mid d_{it} = 1) = \frac{e^{\gamma_0 + \gamma_1 + \gamma_2 d_{it-1}^*}}{1 + e^{\gamma_0 + \gamma_1 + \gamma_2 d_{it-1}^*}} \\ \pi_{01t} &= \Pr(d_{it}^* = 1 \mid d_{it} = 0) = \frac{e^{\gamma_0 + \gamma_2 d_{it-1}^*}}{1 + e^{\gamma_0 + \gamma_2 d_{it-1}^*}}. \end{aligned} \quad (10)$$

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

1. Draw  $M$  times from the  $\varepsilon_{it}$  distribution for each individual  $i$  in every period  $t$  to form the sequence  $\left\{ \left\{ \{\varepsilon_{it}^m\}_{t=1}^T \right\}_{i=1}^N \right\}_{m=1}^M$ .
2. Given  $\left\{ \{x_{it}\}_{t=1}^T \right\}_{i=1}^N$  and the error sequence  $\left\{ \left\{ \{\varepsilon_{it}^m\}_{t=1}^T \right\}_{i=1}^N \right\}_{m=1}^M$ , construct  $M$  simulated choice histories for each individual  $i$   $\left\{ \left\{ \{d_{it}^m\}_{t=1}^T \right\}_{i=1}^N \right\}_{m=1}^M$  according to (1) and the decision rule (2).
3. Construct the classification error rates  $\left\{ \left\{ \{\hat{\pi}_{jkt}^m\}_{t=1}^T \right\}_{m=1}^M \right\}$  for each individual  $i$ . The procedure to do this depends on whether classification error is assumed to be unbiased or biased (see below.)
4. Form an unbiased simulator of the likelihood contribution for each individual  $i$  as:

$$\hat{P}(D_i^* \mid \theta, x_i) = \frac{1}{M} \sum_{m=1}^M \prod_{t=1}^T \left( \sum_{j=0}^1 \sum_{k=0}^1 \hat{\pi}_{jkt}^m I[d_{it}^m = j, d_{it}^* = k] \right)^{I(d_{it}^* \text{ observed})} \quad (11)$$

where  $\theta$  is the vector of model parameters.

In the case of unbiased classification error,  $\hat{\pi}_{jkt}^m$  depends on the true choice probability  $\Pr(d_{it} = 1)$  (see equation (8)). Therefore,  $\Pr(d_{it} = 1)$  must be simulated.

$\Pr(d_{it} = 1)$  can be computed by forming the unbiased simulator

$$\widehat{P}(d_{it} = 1 \mid H_{it}^m) = \frac{1}{M} \sum_{m=1}^M \Pr\left(\varepsilon_{it} \leq \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}^m \rho_\tau\right) \quad (12)$$

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  $t$ .<sup>6</sup>

The classification error rate for  $d_{it}^* = 1$  and  $d_{it}^m = 1$  in this case is,

$$\widehat{\pi}_{11t}^m = E + (1 - E) \widehat{P}(d_{it} = 1 \mid H_{it}^m). \quad (13)$$

In the case of biased measurement error,  $\widehat{\pi}_{jkt}^m$  depends on the reported choice in the previous period  $d_{i,t-1}^*$  (see equation (10)). If the reported choice in the previous period is missing,  $d_{i,t-1}^*$  must be simulated. The reported choice in the previous period is simulated according to (9) and is denoted as  $d_{i,t-1}^{*m}$ .

The classification error rate for  $d_{it}^* = 1$  and  $d_{it}^m = 1$  in this latter case is,

$$\widehat{\pi}_{11t}^m = \frac{e^{\gamma_0 + \gamma_1 + \gamma_2 d_{i,t-1}^{*(m)}}}{1 + e^{\gamma_0 + \gamma_1 + \gamma_2 d_{i,t-1}^{*(m)}}} \quad (14)$$

where  $d_{i,t-1}^{*(m)} = I(d_{i,t-1}^* \text{ observed}) d_{i,t-1}^* + (1 - I(d_{i,t-1}^* \text{ observed})) d_{i,t-1}^{*m}$ .

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  $I[d_{it}^m = j, d_{it}^* = k]$  “picks out” the appropriate classification error rate by comparing  $d_{it}^*$  to  $d_{it}^m$ . If  $d_{it}^*$  is unobserved, then the value of  $I(d_{it}^* \text{ observed})$  is zero, and there is no contribution to the likelihood (i.e., one simply enters one in the product) in period  $t$ .<sup>7</sup>

---

<sup>6</sup>In the case of a mean zero normal  $\varepsilon_{it}$ , the probability in the summation is  $\Phi(a)$  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  $\Phi$  is the standard normal c.d.f.

<sup>7</sup>An extension of the estimation procedure would be to specify the probability that the choice is not observed and incorporate this probability in the likelihood. This is necessary only if choices are not missing at random.

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 (11) builds the likelihood using unconditional simulations of the model. The simulation of conditional probabilities like  $P(d_{it} | H_{it})$  is completely avoided, circumventing the severe computational problems that typically arise if  $H_{it}$  is not fully observed.

Also note that the estimation procedure accommodates serial correlation in the error term through the draws on the  $\{\varepsilon_{it}\}_{t=1}^T$  histories and the resulting simulated choice history. These in turn determine the appropriate classification error rates that enter the likelihood. Similarly, the state space is updated according to previous simulated choices, rather than previous reported choices, which then determine current simulated choices. Thus, accommodating serial correlation does not require the use of a recursive simulator (such as GHK).

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  $\frac{M}{\sqrt{N}} \rightarrow \infty$  as  $N \rightarrow \infty$ . 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 simulate choice probabilities.

## 4.1 Missing Covariates and Initial Conditions

The estimation procedure 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,

$$x_{it} = \phi_2 x_{i,t-1} + \nu_{it} \quad (15)$$

where  $\nu_{it}$  is i.i.d. with zero mean and variance  $\sigma_v^2$ , and where  $x_{i0} = 0$ . If  $x_{it-1}$  is observed and  $x_{it}$  is missing, then the missing  $x_{it}$  is replaced by  $\hat{x}_{it}^m$  which equals  $\phi_2 x_{it-1}$  plus a draw from the  $\nu_{it}$  distribution. A new draw from the  $\nu_{it}$  distribution is taken for each simulated choice history  $m$ .

The likelihood contribution for each individual  $i$  in this case becomes

$$\hat{P}(D_i^*, x_i | \theta) = \frac{1}{M} \sum_{m=1}^M \prod_{t=1}^T f_m(x_{it})^{I(x_{it} \text{ observed})} \left( \sum_{j=0}^1 \sum_{k=0}^1 \hat{\pi}_{jkt}^m I[d_{it}^m = j, d_{it}^* = k] \right)^{I(d_{it}^* \text{ observed})} \quad (16)$$

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

Assuming  $\nu_{it}$  is distributed normally, the density of  $x_{it}$  according to draw sequence  $m$  is,

$$f_m(x_{it}) = \frac{1}{\sigma_v} \phi \left( \frac{x_{it} - \phi_2 \hat{x}_{it-1}^{(m)}}{\sigma_v} \right) \quad (17)$$

where  $\hat{x}_{it-1}^{(m)} = I(x_{i,t-1} \text{ observed}) x_{i,t-1} + (1 - I(x_{i,t-1} \text{ observed})) \hat{x}_{it-1}^m$  and  $\phi$  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).  $f_m(x_{it})$  affects the likelihood only when  $x_{it}$  is observed. The parameters  $\phi_2$  and  $\sigma_v$  now become part of the parameter vector  $\theta$ .

In the case of an initial conditions problem,  $t = 1$  is not the first period of observed data. Let  $t = \tilde{\tau}$  be the first period of observed data where  $\tilde{\tau} > 1$ .<sup>8</sup> 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  $\tilde{\tau}$ . If the  $x_{it}$ 's are also missing, the path of  $x_{it}$ 's must be simulated from  $t = 1$  until  $t = \tilde{\tau}$ .

---

<sup>8</sup>The first period of observed data could also be individual specific. In this case, there would be a different  $\tilde{\tau}_i$  for each individual.

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

$$\hat{P}(D_i^*, x_i | \theta) = \frac{1}{M} \sum_{m=1}^M \prod_{t=\tilde{\tau}}^T f_m(x_{it})^{I(x_{it} \text{ observed})} \left( \sum_{j=0}^1 \sum_{k=0}^1 \hat{\pi}_{jkt}^m I[d_{it}^m = j, d_{it}^* = k] \right)^{I(d_{it}^* \text{ observed})} \quad (18)$$

In (18), the first  $d_{it}^*$  is observed at  $t = \tilde{\tau}$ . In Heckman's approximation method, one would specify a distribution for  $d_{i\tilde{\tau}}^*$ . In our method, it is not necessary to construct a marginal distribution for the initial state. The distribution of the initial state in period  $\tilde{\tau}$  is implicitly determined by the simulated choice and covariate history from  $t = 1$  through  $t = \tilde{\tau} - 1$ .

## 4.2 Importance Sampling

The estimation procedure can also be modified to take advantage of importance sampling techniques that smooth the likelihood function and enable the use of standard gradient methods of optimization.<sup>9</sup> The non-smoothness of the simulated likelihood function arises because, holding the draw sequence  $\{\varepsilon_{it}^m\}_{t=1}^T$  fixed, a change in  $\theta$  can induce discrete changes in the  $\{d_{it}^m\}_{t=1}^T$  sequence. We smooth the likelihood by first constructing simulated choice histories  $\{d_{it}^m\}_{t=1}^T$  at an initial  $\theta = \theta_0$ . We then hold the  $\{d_{it}^m\}_{t=1}^T$  sequences fixed as we vary  $\theta$ . Each simulated choice sequence then has an associated importance sampling weight,  $W_m(\theta)$ , that varies with  $\theta$ . The basic idea of importance sampling is that, when we change  $\theta$ , sequences that are more (less) likely under the new  $\theta$  receive increased (reduced) weight. Thus we have

$$W_m(\theta) = \frac{P(d_{i1}^m, \dots, d_{iT}^m | \hat{\theta}, x_i)}{P(d_{i1}^m, \dots, d_{iT}^m | \theta_0, x_i)} \quad (19)$$

where the numerator is the joint probability that simulated choice history  $m$  occurs given the current vector of trial parameters  $\hat{\theta}$ . The denominator is the joint probability that simulated choice history  $m$  occurs given the initial vector of trial parameters  $\theta_0$ .

---

<sup>9</sup>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.



The joint probability of simulated choice history  $m$  in the dynamic probit model is simply  $\prod_{t=1}^T \Pr \left( \varepsilon_{it} \leq \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}^m \rho_\tau \right)$ .

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

$$\hat{P}(D_i^*, x_i | \theta) = \frac{1}{M} \sum_{m=1}^M W_m(\theta) \prod_{t=\tilde{\tau}}^T f_m(x_{it})^{I(x_{it} \text{ observed})} \left( \sum_{j=0}^1 \sum_{k=0}^1 \hat{\pi}_{jkt}^m I[d_{it}^m = j, d_{it}^* = k] \right)^{I(d_{it}^* \text{ observed})} \quad (20)$$

Note that (20) is just a special case of (18) with  $W_m = 1$  for each simulated choice history  $m$ .<sup>10</sup>

An important computational advantage of the re-weighting scheme over the implicit equal weighting scheme in (18) is that it requires simulated choice histories to be generated only once for each individual, with an initial vector of trial parameters  $\theta_0$ , as opposed to constructing simulated choice histories at each vector of trial parameters  $\hat{\theta}$ . KW used this smooth version of the algorithm to construct standard errors, but used the non-smooth version in estimation (using a simplex algorithm). Akerberg (2001) describes an analogous use of importance sampling and has a good discussion of how his approach differs from ours.

## 5 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 5.1, estimation results for a random effects specification are discussed. In subsection 5.2, 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

---

<sup>10</sup>The efficiency of importance sampling algorithms is often improved if weights are normalized to sum to one.

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

## 5.1 Random Effects Model

In the random effects model, the error term  $\varepsilon_{it}$  follows the components of variance structure in (3) with both  $\mu_i$  and  $\eta_{it}$  distributed normally. The exogenous covariate  $x_{it}$  is also assumed to be distributed normally and is generated by the  $AR(1)$  process in (15). The depreciation weights  $\rho_\tau$  are generated by an exponential decay process,  $\rho_\tau = \rho e^{-\alpha(t-\tau-1)}$ . The parameter  $\alpha$  captures the “speed” of depreciation in the effect of past choices. The vector of estimable parameters for this model is  $\theta = \{\beta_0, \beta_1, \phi_1, \sigma_v, \rho, \alpha, \sigma_\mu, E\}$ . In the special case of no initial conditions problem and no missing exogenous covariates,  $\phi_1$  and  $\sigma_v$  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  $\theta = \{-.10, 1.00, .25, 1.00, 1.00, .50, .80, .75\}$ . For reasons of identification, the variance of  $\varepsilon_{it}$  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  $\varepsilon_{it}$  ( $\sigma_\mu$  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  $\beta'x$  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  $\varepsilon_{it}$  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  $\varepsilon_{it}$  by the third period.

The Mean  $\pi_{11t}$  and Mean  $\pi_{00t}$  columns of Table 1 present the average probabilities of a correct classification. The average probability of a correct match of  $d_{it} = 1$  and  $d_{it}^* = 1$ ,  $\pi_{11t}$ , 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  $d_{it}^* = 0$ ,  $\pi_{00t}$ , is .887 in period 1 and decreases over time to .794 in period 10. This pattern emerges because  $\pi_{11t}$  is an increasing linear function of the proportion choosing  $d_{it} = 1$ , and  $\pi_{00t}$  is a decreasing linear function of the same proportion, as shown in (8). The slope of the linear functions is  $(1 - E)$ . The base classification error rate  $E$  is set to .75, implying that low probability events have a relatively high probability of being classified correctly.

### 5.1.1 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 throughout the study, 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  $\rho$  is statistically significant in all four panels, however, the magnitude of the bias never exceeds 5.1 percent. The biases in the estimates of  $\beta_1$  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. The increased incidence of missing choices does not change the point estimates much since the 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. 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. All the 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,  $\phi_1$  and  $\sigma_v$ , 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  $\rho$  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  $t = 1, \dots, 10$  but there are no missing choices or covariates from  $t = 11, \dots, 20$ .

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  $\beta_1, \rho, \sigma_v$  and  $\sigma_\mu$  are statistically significant. However, the magnitudes of the biases are negligible in magnitude. The maximum bias over these four parameter estimates 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,  $\phi_1$  and  $\sigma_v$ , are not estimated. In this case, the biases are generally substantial in magnitude. Note that the standard errors of the estimates

of  $\rho$  and  $\alpha$  increase dramatically and that  $\sigma_\mu$  is badly biased upwards. The incorrect treatment of the initial condition results in an overestimate of the importance of individual effects (inflated variance).<sup>11</sup>

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  $t = 1, \dots, 10$  but there are no missing choices or covariates from  $t = 11, \dots, 30$ . The observed choices in period  $t = 11, \dots, 20$  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  $t = 21, \dots, 30$ .  $u_{i21}$  in this method is given by:

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

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  $\rho$  and  $\alpha$  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

---

<sup>11</sup>The variance of the composite error term is restricted to be between zero and one. Since almost all of the estimates of  $\sigma_\mu$  are close to the upper boundary of one, the standard deviation over the fifty estimates is very small.

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  $\rho$  in this experiment is substantially biased downward and  $\sigma_\mu$  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  $\rho$  is substantial in magnitude (14%). As might be expected when treating the initial condition as exogenous, the estimate of  $\rho$  is biased upwards.<sup>12</sup>

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. In order to handle classification error, this method must be nested within our algorithm. The Heckman method specifies a different latent index function,  $u_{it}^H$ , in the first period of observed data. We specify the latent index at  $t = 10$  as

$$u_{it}^H = \gamma_0 + \gamma_1 x_{it} + \varepsilon_{it}^H \quad (21)$$

where the variance of  $\varepsilon_{it}^H$  is normalized to one and the correlation coefficient between  $\varepsilon_{it}^H$  and the individual effect  $\mu_i$  is  $\rho_{\mu\varepsilon^H}$ . As before, the likelihood function includes contributions from  $t = 10, \dots, 20$ . The parameters  $\gamma_0$ ,  $\gamma_1$  and  $\rho_{\mu\varepsilon^H}$  are estimated along with the other parameters of the model.

The estimation results show that nesting the Heckman method in our procedure

---

<sup>12</sup>In the  $AR(1)$  error model to be discussed below, treating the initial condition as exogenous produces a bias in the estimate of  $\rho$  which is considerably larger (23%).

works relatively well in the random effects model.  $\rho$  is over-estimated by only 6.4% on average. Although the biases are not substantial for Heckman’s approximate solution approach (except for the constant), simulation from the theoretical start of the process is clearly preferable as the parameter estimates are less biased and more precise.<sup>13</sup>

### 5.1.2 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 6 reports the results of three repeated sampling experiments that use the smooth SML algorithm and that are analogous to the repeated sampling experiments in Table 2 that use the non-smooth algorithm. The three experiments in Table 6 differ in the proportion of missing choices during the sample period. There are no missing exogenous covariates and no initial conditions problem.

Similar to the results in Table 2, the results in Table 6 illustrate that the bias in the estimate of  $\rho$  is statistically significant but the magnitude of the bias is negligible. The bias in the estimate of  $\rho$  never exceeds 4.8 percent. Note that the estimates of  $\rho$  and  $\alpha$  are less median biased than mean biased under the smooth algorithm. The maximum mean bias in the estimate of  $\alpha$  is 9.8% while the maximum median bias is 7.4%. In contrast to the estimates of  $\rho$  and  $\alpha$ , the other parameters of the model are more precisely estimated under the smooth algorithm. The upward bias in the

---

<sup>13</sup>The Heckman solution also produces more serious biases in the  $AR(1)$  error model. The Heckman estimate of  $\rho$  in the  $AR(1)$  error model is biased upward by 20%.



estimates of  $\rho$  and  $\alpha$ , the greater extent of mean bias as compared to median bias, and the relatively larger standard errors in estimates of these versus the other parameters are a consequence of several high valued outliers in the 50 repetitions.

It is 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 1.96 hours with a standard deviation of .89. The smooth version is 47% faster.

Table 7 reports the results of three repeated sampling experiments that use the smooth SML algorithm and that are analogous to the repeated sampling experiments in Table 3. In these experiments, the exogenous covariate is missing for the same observation in which a choice is missing. In the first panel, with 20% missing choices and covariates in each period, the biases are not statistically significant and are negligible in magnitude. Increasing the proportion of missing choices and covariates to 40% increases the biases in the estimates of  $\rho$  and  $\alpha$ , but they remain negligible in magnitude. Further increasing the proportion of missing choices and covariates to 60% produces more serious biases in the estimates of  $\rho$  and  $\alpha$ . The direction of the biases in these latter parameter estimates and the large relative increase in their standard errors are suggestive of an identification problem when there is a large amount of missing information in the data. In general, when covariates are missing there is less information that can be used to “impute” missing choices.<sup>14</sup>

In summary, for the random effects model, the performance of the smooth algo-

---

<sup>14</sup>Although  $\rho$  and  $\alpha$  capture different aspects of the importance of past choices, the under-estimate of  $\rho$  generally decreases the impact of past choices while the under-estimate of  $\alpha$  generally increases the impact. Note that Little and Rubin (1987) also report identification problems in the EM algorithm when there is a large extent of missing information in the data.

rithm seems slightly inferior to that of the non-smooth algorithm. While the smooth algorithm is faster, it has the disadvantage that as  $\hat{\theta}$  departs from  $\theta_0$ , the importance sampling weights  $W_m(\theta)$  for some sequences  $m$  can get small, implying that relatively few sequences are doing all the “work” in simulating the choice probabilities. This is a well known problem with importance sampling algorithms.<sup>15</sup>

## 5.2 $AR(1)$ Error Model

In the  $AR(1)$  error model, the error term  $\varepsilon_{it}$  follows the first-order serial correlation process in (4) with  $\eta_{it}$  assumed to be distributed normally. As in the random effects model, the exogenous covariate  $x_{it}$  is also distributed normally and is generated by the  $AR(1)$  process in (15). The depreciation weights  $\rho_\tau$  follow the same exponential decay process,  $\rho_\tau = \rho e^{-\alpha(t-\tau-1)}$ . The vector of estimable parameters is  $\theta = \{\beta_0, \beta_1, \rho, \alpha, \phi_1, \sigma_v, \phi_2, E\}$ .

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  $\theta = \{-.10, 1.00, 1.00, .50, .25, 1.00, .80, .75\}$ . 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  $\varepsilon_{it}$  is normalized to one and the frequency simulator that is used to compute true classification error rates has  $\tilde{M}$  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.

---

<sup>15</sup>The conclusions from the experiments on the random effects model as well as the  $AR(1)$  error model are not sensitive to the extent of 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.

### 5.2.1 Non-Smooth SML Algorithm

The order of repeated sampling experiments on the  $AR(1)$  error model is the same as the order of experiments on the random effects model. Thus, Tables 9 – 12 are analogous to Tables 2 – 5. The four panels of Table 9 report the results of increasing the incidence of missing choices during the sample period with no initial conditions problem and no missing covariates. As in the corresponding experiments on the random effects model, the bias in  $\rho$  is significant in all four panels but negligible in magnitude, never exceeding 5.5%. The standard errors of the parameter estimates are generally smaller in the  $AR(1)$  error model. Negligible biases and relatively smaller standard errors in comparison to the random effects model also result when adding missing exogenous covariates (compare Tables 3 and 10).

In Table 11, different solutions to the initial conditions problem in the  $AR(1)$  error model are examined. The first panel shows that simulating choices from the theoretical start of the process works well. 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  $\rho$  and  $\alpha$  are also very large. Since  $\rho$  is biased downward and  $\alpha$  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  $\rho$  and  $\alpha$  are biased upward.

Table 12 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. As mentioned earlier, 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  $\rho$  are biased upward by 23% and 20%, respectively.

### 5.2.2 The Smooth SML Algorithm (Importance Sampling)

Table 13 reports the results of estimating the  $AR(1)$  error model with no missing exogenous covariates and no initial conditions problem using the smooth SML algorithm. Similar to the corresponding results in the random effects model, the bias in the estimate of  $\rho$  is statistically significant but never exceeds 5.3 percent. The estimates of  $\alpha$  and  $\rho$  are less median biased than mean biased and have relatively larger standard errors than the other parameter estimates, as was found in the random effects model. The maximum mean bias in the estimate of  $\alpha$  is 7.8% while the maximum median bias is only 3.6%.

The  $AR(1)$  error model converges much faster when using the smooth algorithm. The mean time to convergence over the 50 repetitions in the second 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 13 (20% missing choices) is 1.90 hours with a standard deviation of .78. The  $AR(1)$  error model converges slightly faster than the random effects model.

Table 14 reports the results of adding missing exogenous covariates. As in the random effects model, with 20% missing choices and covariates in each period, the biases are negligible in magnitude. The biases in the estimates of  $\rho$  and  $\alpha$  increase but remain negligible in magnitude when the proportion of missing choices and covariates is increased to 40%. As in the random effects model, increasing the proportion of missing choices and covariates to 60% produces rather serious biases in the estimates of  $\rho$  and  $\alpha$ . But, these biases are not quite as severe as in the corresponding experiment on the random effects model.

## 6 Monte-Carlo Tests - Biased Classification Error

In this section, Monte-Carlo tests of the SML estimator with biased classification error, as specified in (10), are performed. The algorithm used to generate artificial data sets with biased classification error is described in Appendix B. In subsection 6.1, estimation results for the random effects model are discussed and in subsection 6.2 we discuss estimation results for the  $AR(1)$  error model.

### 6.1 Random Effects Model

#### 6.1.1 Non-Smooth SML Algorithm

The three panels of Table 15 report the results of estimating the random effects model with 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 15 differ in the true parameters of the classification error process,  $\gamma_0$ ,  $\gamma_1$  and  $\gamma_2$ .

The first panel specifies values of the parameters that produce a relatively low level of classification error bias. The parameters in the second panel produce an intermediate level of bias and the parameters in the third panel produce a relatively high extent of 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  $\rho$  and  $\sigma_v$  are consistently significantly biased. However, the magnitudes of these biases are negligible. The biases rarely exceed 3 percent. Note that increases in the extent of biased classification error 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 is also faster than 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.<sup>16</sup>

### 6.1.2 Smooth SML Algorithm

The three panels of Table 16 report the results of estimating three different random effects models with biased classification error using the smooth SML algorithm. In all three panels, 20% of the choices are missing in each period, there is no initial conditions problem and there is a relatively low extent of true classification error bias, as in panel 1 of Table 15. In the first panel, the results of estimating a first-order Markov process with no missing exogenous covariates are reported. There are no statistically significant biases in the estimates of the structural parameters and the magnitudes of the biases are extremely small. There are significant biases in the estimates of the classification error process parameters. However, the magnitudes of the biases are negligible.

In the second panel, the results of estimating the Polya model with no missing covariates are reported. The same pattern as in the first panel emerges. There are no statistically significant biases in the estimates of the structural parameters, and the biases in the estimates of the parameters of the classification error process are negligible in magnitude. In the third panel, the results of estimating the Polya model with missing covariates are reported. The general conclusions are unchanged. As in the case of unbiased classification using the smooth SML algorithm, increasing the extent of missing choices and covariates increases the magnitude of the biases.

---

<sup>16</sup>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.

## 6.2 $AR(1)$ Error Model

### 6.2.1 Non-Smooth SML Algorithm

The three panels in Table 17 repeat the series of repeated sampling experiments in Table 15 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.

### 6.2.2 Smooth SML Algorithm

The three panels in Table 18 repeat the series of repeated sampling experiments in Table 16 with an  $AR(1)$  specification for the error term. The general conclusions from the corresponding experiments on the random effects model carry over to  $AR(1)$  errors. The biases are negligible in magnitude and mostly not statistically significant.

## 7 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 systematic unified approach to the initial conditions problem and the problem of missing data during the sample period. The parallel problems of unobserved initial conditions and missing data during the sample period are overcome by forming the likelihood with unconditional simulations of the model from the theoretical start of the process. In order to make it feasible to simulate the likelihood using unconditional simulations, it is necessary to assume a classification error process in discrete choices. The assumption that reported choices are misclassified, however, is a reasonable one in almost all empirical applications in economics. The estimation technique can 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 estimation method 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 both unbiased and biased classification error processes. The estimator was shown to have good small sample properties. Under the non-smooth version of the SML estimation algorithm, we found that biases are negligible in magnitude even for exceedingly high amounts of missing information in the data. Under the faster, smooth version of the SML estimation algorithm, biases were also negligible in magnitude except for cases in which there is a very high frequency of missing choices and missing exogenous covariates.

It was also shown that the estimation technique can be combined with Heckman's (1981a) approximate solution to the initial conditions problem. That is, one can specify an approximation to the marginal distribution of the initial condition, but use our approach to integrate over missing information during the sample period. Such a hybrid approach may be appealing when there is no natural starting point to the process. One caveat, however, is that the Heckman approximate solution method worked much less well in our experiments with the  $AR(1)$  error model.

Interestingly, our SML algorithm seems to perform better (in terms of consistently producing negligible bias) for models with biased 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.

Future research will examine the small sample properties of the estimation technique in more complex settings. For example, observed continuous outcomes, such as wages, can be incorporated into estimation by specifying measurement error densities that enter the likelihood. The algorithm can also be extended to a multinomial choice context, and to contexts where construction of the latent indices in the choice model requires solution of a dynamic program. It is in this latter context that particular versions of the estimation technique have been previously applied. The estimation method can also be further developed to handle cases in which the missing data are not missing at random.

## References

- [1] Akerberg, D. (2001), “A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation,” unpublished manuscript.
- [2] Card D. and D. Sullivan (1988), “Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment,” *Econometrica*, 56, 497-530.
- [3] Chamberlain, G. (1984), “Panel Data,” Chapter 22 in Handbook of Econometrics, Vol. 2, ed. by Z. Griliches and M.D. Intrilligator. Amsterdam: Elsevier Science Publishers B.V.
- [4] 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.
- [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] Ham, J.C., and R.J. LaLonde (1996), “The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence From Experimental Data on Training”, *Econometrica*, 64, 175-205.
- [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] Heckman, J., and B. Singer (1984), “Econometric Duration Analysis,” *Journal of Econometrics*, 24, 63-132.
- [11] Honore, B.E., and E. Kyriazidou (2000), “Panel Data Discrete Choice Models With Lagged Dependent Variables,” *Econometrica*, 68, 839-874.
- [12] Hyslop, D.R. (1999), “State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women,” *Econometrica*, 67, 1255-1294.
- [13] 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.
- [14] 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.
- [15] Little, R.J.A., and D.B. Rubin (1987), *Statistical Analysis with Missing Data*, New York: John Wiley & Sons.
- [16] Pakes A. and D. Pollard (1989), “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 57, 1027-1057.
- [17] 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.
- [18] Rubin, D.B. (1976) , “Inference and Missing Data,” *Biometrika*, 63, 581-592.

- [19] Ruud, P.A. (1991), “Extensions of Estimation Methods Using the EM Algorithm,” *Journal of Econometrics*, 49, 305-341.
- [20] Sauer, R.M. (2003), “Educational Financing and Lifetime Earnings,” *Review of Economic Studies*, forthcoming.
- [21] Schafer, J.L. (1997), *Analysis of Incomplete Multivariate Data*, London: Chapman & Hall.

## 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  $N$  times from the  $x_{it}$  distribution in every period  $t$  to construct the sequence  $\left\{ \left\{ x_{it} \right\}_{t=1}^T \right\}_{i=1}^N$ .
2. Draw  $\widetilde{M}$  times from the  $\varepsilon_{it}$  distribution for each individual  $i$  in every period  $t$  to form the sequence  $\left\{ \left\{ \left\{ \widetilde{\varepsilon}_{it}^m \right\}_{t=1}^T \right\}_{i=1}^N \right\}_{m=1}^{\widetilde{M}}$ . Note that  $\widetilde{M}$  will generally differ from the number of simulated choice histories  $M$  generated for each individual in estimation.
3. Given  $\left\{ \left\{ x_{it} \right\}_{t=1}^T \right\}_{i=1}^N$  and the error sequence  $\left\{ \left\{ \left\{ \widetilde{\varepsilon}_{it}^m \right\}_{t=1}^T \right\}_{i=1}^N \right\}_{m=1}^{\widetilde{M}}$ , 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  $H_{it}^m = \left\{ \{x_{i\tau}\}_{\tau=1}^t, \{\tilde{d}_{i\tau}^m\}_{\tau=1}^t \right\}$ .

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

## Stage 2

1. Draw  $N$  times from the  $\varepsilon_{it}$  distribution for each individual  $i$  in every period  $t$  to form the sequence  $\left\{ \{\varepsilon_{it}\}_{t=1}^T \right\}_{i=1}^N$ .
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  $\left\{ \{\tilde{U}_{it}\}_{t=1}^T \right\}_{i=1}^N$ .
6. Compare the uniform draws to the probability  $\pi^{obs}$  that  $d_{it}^*$  is missing in period  $t$ . That is, implement the following rule: if  $\tilde{U}_{it} < \pi^{obs}$  then  $I(d_{it}^* \text{ observed}) = 1$ , else  $I(d_{it}^* \text{ observed}) = 0$ .

Note that step 6 does not specify  $\pi^{obs}$  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  $\pi^{obs}$  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  $N$  times from the  $x_{it}$  distribution in every period  $t$  to construct the sequence  $\left\{ \left\{ x_{it} \right\}_{t=1}^T \right\}_{i=1}^N$ .
2. Draw  $N$  times from the  $\varepsilon_{it}$  distribution for each individual  $i$  in every period  $t$  to form the sequence  $\left\{ \left\{ \varepsilon_{it} \right\}_{t=1}^T \right\}_{i=1}^N$ .
3. Given  $\left\{ \left\{ x_{it} \right\}_{t=1}^T \right\}_{i=1}^N$  and  $\left\{ \left\{ \varepsilon_{it} \right\}_{t=1}^T \right\}_{i=1}^N$ , construct  $N$  true choices  $\left\{ \left\{ d_{it} \right\}_{t=1}^T \right\}_{i=1}^N$  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  $\left\{ \left\{ U_{it} \right\}_{t=1}^T \right\}_{i=1}^N$ .
5. Construct  $N$  reported choices  $\left\{ \left\{ d_{it}^* \right\}_{i=1}^N \right\}_{t=1}^T$  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  $\pi_{jkt}$  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  $\left\{ \left\{ \tilde{U}_{it} \right\}_{t=1}^T \right\}_{i=1}^N$ .
7. Implement the following rule: if  $\tilde{U}_{it} < \pi^{obs}$  then  $I(d_{it}^* \text{ observed}) = 1$ , else  $I(d_{it}^* \text{ observed}) = 0$ .



Table 1

Summary Statistics  
 Representative Data Set  
 Random Effects Model  
 Unbiased Classification Error

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

Note:  $d_{it}$  is the true choice,  $d_{it}^*$  is the reported choice,  $\pi_{11t}$  and  $\pi_{00t}$  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:

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

Table 2

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

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
No Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0975	-.0950	.0427	.0427	.42
$\beta_1$	1.0000	1.0171	1.0196	.0552	.0578	2.20
$\rho$	1.0000	1.0463	1.0462	.0513	.0691	6.38
$\alpha$	.5000	.4912	.4926	.0499	.0506	-1.22
$\sigma_\mu$	.8000	.8062	.8009	.0269	.0276	1.62
$E$	.7500	.7408	.7417	.0162	.0186	-3.99
20% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0995	-.1017	.0428	.0428	.08
$\beta_1$	1.0000	1.0114	1.0199	.0611	.0622	1.32
$\rho$	1.0000	1.0450	1.0356	.0528	.0694	6.04
$\alpha$	.5000	.4864	.4985	.0719	.0731	-1.34
$\sigma_\mu$	.8000	.8095	.8066	.0259	.0275	2.59
$E$	.7500	.7409	.7399	.0184	.0206	-3.50
40% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1025	-.1001	.0530	.0530	-.33
$\beta_1$	1.0000	1.0183	1.0265	.0612	.0648	2.09
$\rho$	1.0000	1.0505	1.0425	.0524	.0728	6.81
$\alpha$	.5000	.4887	.4882	.0633	.0643	-1.26
$\sigma_\mu$	.8000	.8047	.7989	.0339	.0343	.98
$E$	.7500	.7437	.7412	.0231	.0239	-1.94
60% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1070	-.1052	.0596	.0600	-.82
$\beta_1$	1.0000	1.0147	1.0161	.0860	.0872	1.21
$\rho$	1.0000	1.0485	1.0562	.0603	.0773	5.68
$\alpha$	.5000	.4970	.4982	.0817	.0817	-.26
$\sigma_\mu$	.8000	.8016	.8012	.0486	.0487	.23
$E$	.7500	.7477	.7426	.0287	.0288	-.55

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

Table 3

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

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
20% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1051	-.1023	.0436	.0439	-.83
$\beta_1$	1.0000	1.0167	1.0191	.0611	.0634	1.92
$\rho$	1.0000	1.0479	1.0446	.0444	.0653	7.63
$\alpha$	.5000	.4977	.5031	.0656	.0657	-.24
$\phi_2$	.2500	.2520	.2505	.0176	.0177	.80
$\sigma_\nu$	.5000	.5015	.5016	.0057	.0059	1.86
$\sigma_\mu$	.8000	.8056	.8017	.0287	.0292	1.38
$E$	.7500	.7428	.7430	.0172	.0187	-2.95
40% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1087	-.1099	.0539	.0546	-1.15
$\beta_1$	1.0000	1.0141	1.0233	.0678	.0692	1.48
$\rho$	1.0000	1.0458	1.0374	.0636	.0784	5.10
$\alpha$	.5000	.4953	.4949	.0600	.0602	.56
$\phi_2$	.2500	.2521	.2546	.0253	.0254	.59
$\sigma_\nu$	.5000	.5012	.5012	.0069	.0070	1.21
$\sigma_\mu$	.8000	.8046	.8063	.0347	.0350	.94
$E$	.7500	.7474	.7416	.0245	.0246	-.74
60% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0997	-.1116	.0542	.0543	.05
$\beta_1$	1.0000	1.034	1.0258	.0894	.0924	1.85
$\rho$	1.0000	1.0401	1.0512	.0682	.0791	4.15
$\alpha$	.5000	.4957	.4973	.0721	.0722	-.42
$\phi_2$	.2500	.2507	.2498	.0372	.0373	.13
$\sigma_\nu$	.5000	.5011	.5017	.0089	.0090	.88
$\sigma_\mu$	.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.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta} - \beta)}{\text{Std}(\hat{\beta})} \right)$ . The model is the same as in Table 1.

Table 4

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

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
Simulate from start of process with $d_{i0} = 0$ ( $t = 11, \dots, 20$ )						
$\beta_0$	-.1000	-.1001	-.1022	.0295	.0295	-.02
$\beta_1$	1.0000	1.0286	1.0337	.0454	.0537	4.46
$\rho$	1.0000	1.0298	1.0253	.0324	.0440	6.51
$\alpha$	.5000	.5044	.5004	.0320	.0323	.98
$\phi_2$	.2500	.2501	.2526	.0135	.0135	.05
$\sigma_\nu$	.5000	.5015	.5025	.0042	.4985	2.56
$\sigma_\mu$	.8000	.8130	.8145	.0245	.0277	3.74
$E$	.7500	.7450	.7410	.0193	.0199	-1.82
Assume process starts with $d_{i,10} = 0$ ( $t = 11, \dots, 20$ )						
$\beta_0$	-.1000	.9367	.9513	.0543	1.0381	135.05
$\beta_1$	1.0000	.2966	.2844	.0938	.7096	-53.01
$\rho$	1.0000	.9543	.9333	.3278	.3310	-.99
$\alpha$	.5000	.4187	.3995	.2957	.3067	-1.94
$\sigma_\mu$	.8000	.9905	.9923	.0090	.1907	149.11
$E$	.7500	.7144	.7125	.0230	.0424	-10.96
Use reported data from $t = 11, \dots, 20$ to proxy for initial condition at $t = 21$ ( $t = 11, \dots, 30$ )						
$\beta_0$	-.1000	-.5239	-.4859	.3039	.5216	-9.86
$\beta_1$	1.0000	.4742	.4671	.1788	.5553	-20.80
$\rho$	1.0000	1.0522	1.1064	.3076	.3120	1.20
$\alpha$	.5000	.5839	.6139	.2299	.2448	2.58
$\sigma_\mu$	.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.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta} - \beta)}{\text{Std}(\hat{\beta})} \right)$ . The model is the same as in Table 1.

Table 5

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

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
Simulate from start of process with $d_{i0} = 0$ ( $t = 10, \dots, 20$ )						
$\beta_0$	-.1000	-.1127	-.1086	.0391	.0411	-2.30
$\beta_1$	1.0000	1.0379	1.0364	.0324	.0500	8.25
$\rho$	1.0000	1.0330	1.0319	.0386	.0508	6.04
$\phi_2$	.2500	.2496	.2511	.0136	.0136	-.19
$\sigma_\nu$	.5000	.5014	.5011	.0045	.4986	2.17
$\sigma_\mu$	.8000	.8137	.8133	.0294	.0324	3.29
$E$	.7500	.7293	.7294	.0150	.0256	-9.75
Assume process starts with $d_{i9} = 0$ ( $t = 10, \dots, 20$ )						
$\beta_0$	-.1000	.1598	.1594	.0775	.2712	23.70
$\beta_1$	1.0000	.9126	.9171	.0693	.1115	-8.92
$\rho$	1.0000	.6396	.6171	.1025	.3747	-24.87
$\sigma_\mu$	.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 $t = 11$ ( $t = 10, \dots, 20$ )						
$\beta_0$	-.1000	-.1882	-.1867	.0771	.1171	-8.09
$\beta_1$	1.0000	1.0328	1.0480	.0595	.0679	3.90
$\rho$	1.0000	1.1369	1.1465	.1024	.1710	9.45
$\sigma_\mu$	.8000	.7838	.7843	.0460	.0488	-2.49
$E$	.7500	.7240	.7262	.0233	.0349	-7.91
Use Heckman's approximation method to proxy for initial condition at $t = 11$ ( $t = 10, \dots, 20$ )						
$\beta_0$	-.1000	-.1721	-.1705	.0728	.1025	-7.01
$\beta_1$	1.0000	.9868	.9831	.0616	.0630	-1.52
$\rho$	1.0000	1.0637	1.0673	.1074	.1249	4.20
$\sigma_\mu$	.8000	.7735	.7767	.0472	.0542	-3.97
$E$	.7500	.7438	.7456	.0181	.0191	-2.44
$\gamma_0$		.3819	.3843	.0757		
$\gamma_1$		.6857	.6799	.1008		
$\rho_{\mu\epsilon^H}$		.6565	.6589	.0627		

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta} - \beta)}{\text{Std}(\hat{\beta})} \right)$ .

The Markov model replaces  $\sum_{\tau=0}^{t-1} d_{i\tau} \rho_\tau$  in Table 1 with  $\rho d_{i,t-1}$ .

Table 6

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

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
20% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1033	-.1034	.0158	.0161	-1.50
$\beta_1$	1.0000	1.0000	1.0001	.0093	.0093	.04
$\rho$	1.0000	1.0485	1.0301	.1049	.1155	3.27
$\alpha$	.5000	.5493	.5371	.0840	.0974	4.15
$\sigma_\mu$	.8000	.7993	.7990	.0033	.0034	-1.57
$E$	.7500	.7454	.7453	.0165	.0171	-1.97
40% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1014	-.1018	.0178	.0179	-.57
$\beta_1$	1.0000	1.0003	1.0000	.0107	.0107	.22
$\rho$	1.0000	1.0397	1.0224	.1192	.1257	2.35
$\alpha$	.5000	.5419	.5061	.0943	.1032	3.14
$\sigma_\mu$	.8000	.7998	.7995	.0038	.0038	-.43
$E$	.7500	.7492	.7476	.0215	.0215	-.28
60% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1011	-.0986	.0204	.0204	-.39
$\beta_1$	1.0000	1.0002	1.0011	.0121	.0121	.13
$\rho$	1.0000	1.0417	1.0138	.1277	.1343	2.31
$\alpha$	.5000	.5487	.5170	.1211	.1305	2.85
$\sigma_\mu$	.8000	.8000	.8001	.0046	.0046	.04
$E$	.7500	.7511	.7452	.0275	.0275	.29

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta} - \beta)}{\text{Std}(\hat{\beta})} \right)$ . The model is the same as in Table 1.

Table 7

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

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
20% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0948	-.0960	.0178	.0185	2.06
$\beta_1$	1.0000	.9986	.9989	.0106	.0107	-.92
$\rho$	1.0000	.9923	.9906	.1098	.1101	-.50
$\alpha$	.5000	.5154	.4954	.0953	.0966	1.15
$\phi_2$	.2500	.2508	.2535	.0145	.0145	.40
$\sigma_\nu$	.5000	.5014	.5015	.0055	.0057	1.77
$\sigma_\mu$	.8000	.8010	.8011	.0038	.0039	1.93
$E$	.7500	.7462	.7469	.0168	.0172	-1.61
40% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0867	-.0922	.0265	.0297	3.56
$\beta_1$	1.0000	.9970	.9959	.0153	.0155	-1.39
$\rho$	1.0000	.9455	.9416	.1201	.1319	-3.21
$\alpha$	.5000	.4752	.4757	.0805	.0842	-2.18
$\phi_2$	.2500	.2510	.2498	.0185	.0185	.38
$\sigma_\nu$	.5000	.5013	.5017	.0065	.0066	1.47
$\sigma_\mu$	.8000	.8031	.8019	.0059	.0067	3.69
$E$	.7500	.7514	.7504	.0215	.0216	.45
60% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0539	-.0567	.0355	.0582	9.19
$\beta_1$	1.0000	.9959	.9999	.0262	.0265	-1.10
$\rho$	1.0000	.7979	.7738	.1362	.2437	-10.50
$\alpha$	.5000	.3826	.3714	.0986	.1533	-8.42
$\phi_2$	.2500	.2630	.2636	.0286	.0314	3.21
$\sigma_\nu$	.5000	.5000	.5000	.0078	.0078	.03
$\sigma_\mu$	.8000	.8111	.8106	.0095	.0146	8.31
$E$	.7500	.7540	.7465	.0283	.0286	.99

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta} - \beta)}{\text{Std}(\hat{\beta})} \right)$ . The model is the same as in Table 1.

Table 8

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

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

Note:  $d_{it}$  is the true choice,  $d_{it}^*$  is the reported choice,  $\pi_{11t}$  and  $\pi_{00t}$  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  $\tilde{M}$  set to 1000. The model is:

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



Table 9

Repeated Sampling Experiments  
AR(1) Error Model  
Unbiased Classification Error  
(No Missing X's, No Initial Conditions Problem)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
No Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1008	-.0967	.0418	.0418	-.14
$\beta_1$	1.0000	1.0105	1.0101	.0492	.0503	1.50
$\rho$	1.0000	1.0357	1.0317	.0479	.0598	5.27
$\alpha$	.5000	.5058	.5088	.0371	.0375	1.11
$\phi_1$	.8000	.8002	.7943	.0266	.0266	.06
$E$	.7500	.7457	.7445	.0137	.0143	-2.23
20% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1020	-.1028	.0414	.0414	-.35
$\beta_1$	1.0000	1.0045	1.0136	.0552	.0554	.57
$\rho$	1.0000	1.0377	1.0356	.0408	.0556	6.53
$\alpha$	.5000	.5066	.5125	.0514	.0519	.91
$\phi_1$	.8000	.8006	.8016	.0257	.0257	.17
$E$	.7500	.7469	.7463	.0191	.0193	-1.14
40% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1034	-.1034	.0423	.0424	-.57
$\beta_1$	1.0000	.9906	1.0015	.0711	.0718	-.93
$\rho$	1.0000	1.0470	1.0361	.0540	.0716	6.15
$\alpha$	.5000	.5072	.5154	.0585	.0590	.87
$\phi_1$	.8000	.7974	.7984	.0331	.0333	-.56
$E$	.7500	.7488	.7470	.0244	.0245	-.36
60% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1056	-.1065	.0605	.0607	-.66
$\beta_1$	1.0000	.9904	.9850	.0800	.0806	-.85
$\rho$	1.0000	1.0553	1.0506	.0731	.0917	5.34
$\alpha$	.5000	.5054	.5064	.0560	.0562	.68
$\phi_1$	.8000	.7986	.8035	.0405	.0405	-.25
$E$	.7500	.7479	.7430	.0360	.0361	-.42

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  
AR(1) Error Model  
Unbiased Classification Error  
(Missing X's, No Initial Conditions Problem)

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
20% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1042	-.0981	.0391	.0394	-.76
$\beta_1$	1.0000	1.0021	1.0060	.0519	.0519	.29
$\rho$	1.0000	1.0444	1.0393	.0424	.0614	7.40
$\alpha$	.5000	.5057	.5058	.0423	.0428	1.12
$\phi_2$	.2500	.2521	.2486	.0181	.0183	.83
$\sigma_\nu$	.5000	.5018	.5024	.0057	.0060	2.21
$\phi_1$	.8000	.7996	.8003	.0264	.0264	-1.12
$E$	.7500	.7473	.7486	.0174	.0176	-1.08
40% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1052	-.1014	.0400	.0403	-.92
$\beta_1$	1.0000	1.0036	1.0011	.0566	.0567	.45
$\rho$	1.0000	1.0460	1.0400	.0446	.0640	7.30
$\alpha$	.5000	.5018	.5053	.0405	.0405	.32
$\phi_2$	.2500	.2522	.2531	.0261	.0262	.61
$\sigma_\nu$	.5000	.5019	.5026	.0067	.0070	1.98
$\phi_1$	.8000	.8002	.7989	.0301	.0301	.05
$E$	.7500	.7504	.7524	.0251	.0251	.12
60% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.1041	-.0996	.0524	.0526	-.55
$\beta_1$	1.0000	1.0003	1.0124	.0748	.0748	.03
$\rho$	1.0000	1.0433	1.0372	.0610	.0748	5.03
$\alpha$	.5000	.5047	.5077	.0621	.0623	.54
$\phi_2$	.2500	.2521	.2514	.0384	.0385	.39
$\sigma_\nu$	.5000	.5007	.5018	.0086	.0086	.61
$\phi_1$	.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.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta} - \beta)}{\text{Std}(\hat{\beta})} \right)$ . The model is the same as in Table 8.

Table 11

Repeated Sampling Experiments  
AR(1) Error Model  
Unbiased Classification Error  
(No Missing Choices or X's, Initial Conditions Problem)

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
Simulate from start of process with $d_{i0} = 0$ ( $t = 11, \dots, 20$ )						
$\beta_0$	-.1000	-.0896	-.0925	.0265	.0285	2.77
$\beta_1$	1.0000	1.0224	1.0221	.0479	.0529	3.31
$\rho$	1.0000	1.0194	1.0148	.0298	.0356	4.60
$\alpha$	.5000	.5121	.5128	.0238	.0267	3.59
$\phi_2$	.2500	.2511	.2531	.0138	.0139	.56
$\sigma_\nu$	.5000	.5011	.5013	.0047	.0049	1.58
$\phi_1$	.8000	.8071	.8100	.0280	.0289	1.80
$E$	.7500	.7420	.7455	.0261	.0273	-2.16
Assume process starts with $d_{i,10} = 0$ ( $t = 11, \dots, 20$ )						
$\beta_0$	-.1000	.9503	.9682	.0605	1.0520	122.84
$\beta_1$	1.0000	.1699	.3883	.4544	.9463	-12.92
$\rho$	1.0000	.5849	.5266	.2792	.5003	-10.51
$\alpha$	.5000	.7102	.7385	.3180	.3812	4.67
$\phi_1$	.8000	.9221	.9259	.0316	.1261	27.33
$E$	.7500	.7656	.7485	.1323	.1332	.83
Use reported data from $t = 11, \dots, 20$ to proxy for initial condition at $t = 21$ ( $t = 11, \dots, 30$ )						
$\beta_0$	-.1000	-.0862	-.0812	.0617	.0632	1.58
$\beta_1$	1.0000	.9406	.9781	.0932	.1105	-4.50
$\rho$	1.0000	1.0445	1.0219	.0924	.1026	3.41
$\alpha$	.5000	.5908	.5674	.0737	.1170	8.72
$\phi_1$	.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.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta} - \beta)}{\text{Std}(\hat{\beta})} \right)$ . The model is the same as in Table 8.

Table 12

Repeated Sampling Experiments  
AR(1) Error Model  
Unbiased Classification Error  
First Order Markov Process  
(No Missing Choices or X's, Initial Conditions Problem)

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
Simulate from start of process with $d_{i0} = 0$ ( $t = 10, \dots, 20$ )						
$\beta_0$	-.1000	-.1171	-.1125	.0429	.0462	-2.81
$\beta_1$	1.0000	1.0185	1.0191	.0323	.0373	4.05
$\rho$	1.0000	1.0354	1.0316	.0465	.0585	5.38
$\phi_2$	.2500	.2511	.2509	.0139	.0140	.56
$\sigma_\nu$	.5000	.5013	.5016	.0050	.0052	1.89
$\phi_1$	.8000	.8081	.8077	.0266	.0278	2.15
$E$	.7500	.7401	.7403	.0126	.0160	-5.58
Assume process starts with $d_{i9} = 0$ ( $t = 10, \dots, 20$ )						
$\beta_0$	-.1000	.1895	.1797	.0547	.2946	37.43
$\beta_1$	1.0000	.8189	.8025	.0727	.1951	-17.63
$\rho$	1.0000	.5932	.5807	.1054	.4202	-27.29
$\phi_1$	.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 $t = 11$ ( $t = 10, \dots, 20$ )						
$\beta_0$	-.1000	-.2416	-.2501	.0492	.1500	-20.36
$\beta_1$	1.0000	1.0150	1.0239	.0430	.0456	2.46
$\rho$	1.0000	1.2330	1.2380	.0702	.2434	23.47
$\phi_1$	.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 $t = 11$ ( $t = 10, \dots, 20$ )						
$\beta_0$	-.1000	-.2181	-.2206	.0538	.1298	-15.54
$\beta_1$	1.0000	1.0333	1.0315	.0471	.0577	5.00
$\rho$	1.0000	1.1997	1.2129	.0604	.2086	23.37
$\phi_1$	.8000	.7727	.7746	.0316	.0418	-6.13
$E$	.7500	.7385	.7385	.0116	.0164	-7.00
$\gamma_0$		.4149	.4118	.0564		
$\gamma_1$		.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.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta} - \beta)}{\text{Std}(\hat{\beta})} \right)$ .

The Markov model replaces  $\sum_{\tau=0}^{t-1} d_{i\tau} \rho_\tau$  in Table 8 with  $\rho d_{i,t-1}$ .

Table 13

Repeated Sampling Experiments  
 AR(1) Error Model  
 Unbiased Classification Error  
 Smooth Algorithm  
 (No Missing X's, No Initial Conditions Problem)

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
20% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0988	-.0987	.0188	.0188	.47
$\beta_1$	1.0000	.9994	.9995	.0103	.0103	-.39
$\rho$	1.0000	1.0531	1.0358	.1631	.1715	2.30
$\alpha$	.5000	.5358	.5136	.0865	.0936	2.92
$\phi_1$	.8000	.8004	.8001	.0040	.0040	.72
$E$	.7500	.7488	.7477	.0178	.0178	-.46
40% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0965	-.0969	.0191	.0194	1.28
$\beta_1$	1.0000	.9993	1.0005	.0104	.0105	-.47
$\rho$	1.0000	1.0508	1.0286	.1398	.1488	2.57
$\alpha$	.5000	.5398	.5263	.0870	.0957	3.23
$\phi_1$	.8000	.8009	.8007	.0040	.0041	1.58
$E$	.7500	.7522	.7504	.0223	.0234	.70
60% Missing Choices ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0928	-.0912	.0228	.0239	2.24
$\beta_1$	1.0000	.9992	.9980	.0095	.0096	-.60
$\rho$	1.0000	1.0350	1.0013	.1607	.1644	1.54
$\alpha$	.5000	.5297	.5103	.0936	.0982	2.24
$\phi_1$	.8000	.8018	.8017	.0045	.0049	2.79
$E$	.7500	.7533	.7487	.0297	.0300	.77

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta}) - \beta}{\text{Std}(\hat{\beta})} \right)$ . The model is the same as in Table 8.

Table 14

Repeated Sampling Experiments  
 AR(1) Error Model  
 Unbiased Classification Error  
 Smooth Algorithm  
 (Missing X's, No Initial Conditions Problem)

Parameter	True Value	Mean $\hat{\beta}$	Median $\hat{\beta}$	$Std(\hat{\beta})$	$RMSE$	t-Stat
20% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0916	-.0910	.0169	.0189	3.48
$\beta_1$	1.0000	1.0001	1.0004	.0117	.0117	.03
$\rho$	1.0000	.9852	.9869	.1012	.1023	-1.03
$\alpha$	.5000	.5048	.4932	.0614	.0616	.55
$\phi_2$	.2500	.2516	.2535	.0154	.0155	.74
$\sigma_\nu$	.5000	.5015	.5025	.0055	.0057	1.90
$\phi_1$	.8000	.8016	.8017	.0037	.0040	3.11
$E$	.7500	.7495	.7485	.0174	.0174	-.19
40% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0861	-.0899	.0227	.0266	4.33
$\beta_1$	1.0000	.9935	.9929	.0171	.0183	-2.68
$\rho$	1.0000	.9454	.9176	.1317	.1425	-2.93
$\alpha$	.5000	.4937	.4705	.0947	.0949	-.47
$\phi_2$	.2500	.2564	.2561	.0211	.0221	2.13
$\sigma_\nu$	.5000	.5006	.5009	.0066	.0067	.69
$\phi_1$	.8000	.8030	.8020	.0047	.0056	4.52
$E$	.7500	.7530	.7522	.0216	.0218	.97
60% Missing Choices and X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0669	-.0652	.0337	.0472	6.93
$\beta_1$	1.0000	.9836	.9857	.0274	.0324	-4.50
$\rho$	1.0000	.8601	.8131	.1910	.2368	-5.18
$\alpha$	.5000	.4537	.4303	.1159	.1248	-2.82
$\phi_2$	.2500	.2702	.2665	.0254	.0324	5.61
$\sigma_\nu$	.5000	.4995	.5000	.0083	.0083	-.47
$\phi_1$	.8000	.8078	.8076	.0073	.0107	7.56
$E$	.7500	.7559	.7548	.0286	.0292	1.46

Note: The number of replications in each experiment is 50 and the number of individuals in the sample is 500.  $Std(\hat{\beta})$  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  $\sqrt{50} \left( \frac{\text{Mean}(\hat{\beta} - \beta)}{\text{Std}(\hat{\beta})} \right)$ . The model is the same as in Table 8.

Table 15

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

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
Low Classification Error Bias ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0922	-.944	.0387	.0394	1.42
$\beta_1$	1.0000	1.0198	1.0131	.0531	.0567	2.63
$\rho$	1.0000	1.0144	1.0102	.0390	.0415	2.61
$\alpha$	.5000	.5031	.5104	.0489	.0490	.45
$\phi_2$	.2500	.2489	.2456	.0161	.0161	-.47
$\sigma_\nu$	.5000	.5018	.5018	.0050	.0053	2.47
$\sigma_\mu$	.8000	.8068	.8041	.0239	.0248	1.99
$\gamma_0$	-3.5000	-3.4867	-3.4762	.0580	.0595	1.62
$\gamma_1$	5.0000	4.9845	5.0033	.0728	.0744	-1.51
$\gamma_2$	2.0000	2.0161	2.0236	.0446	.0475	2.56
Medium Classification Error Bias ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0941	-.0988	.0425	.0429	.98
$\beta_1$	1.0000	1.0045	1.0119	.0608	.0609	.52
$\rho$	1.0000	1.0222	1.0232	.0465	.0515	3.37
$\alpha$	.5000	.5160	.5253	.0658	.0677	1.71
$\phi_2$	.2500	.2476	.2452	.0162	.0163	-1.04
$\sigma_\nu$	.5000	.5022	.5026	.0050	.0054	3.04
$\sigma_\mu$	.8000	.8049	.8041	.0272	.0276	1.29
$\gamma_0$	-3.0000	-2.9902	-2.9826	.0561	.0570	1.24
$\gamma_1$	4.0000	3.98	3.9951	.0776	.0787	-1.19
$\gamma_2$	2.0000	2.0104	2.0134	.0782	.0789	.94
High Classification Error Bias ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0988	-.0918	.0708	.0708	.12
$\beta_1$	1.0000	1.0145	1.0068	.0693	.0708	1.48
$\rho$	1.0000	1.0218	1.0228	.0791	.0820	1.94
$\alpha$	.5000	.5088	.5328	.0993	.0997	.63
$\phi_2$	.2500	.2484	.2460	.0164	.0165	-.70
$\sigma_\nu$	.5000	.5021	.5028	.0051	.2980	2.90
$\sigma_\mu$	.8000	.8023	.7999	.0406	.3050	.40
$\gamma_0$	-3.0000	-2.9918	-2.9983	.0638	.0643	.91
$\gamma_1$	3.0000	2.9842	2.9920	.0829	.0844	-1.34
$\gamma_2$	3.0000	3.0190	3.0371	.1018	.1036	-1.32

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

Table 16

Repeated Sampling Experiments  
 Random Effects Model  
 Biased Classification Error  
 Low Classification Error Bias  
 Smooth Algorithm  
 (20% Missing Choices, No Initial Conditions Problem)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
No Missing X's (Markov model) ( $t = 1, \dots, 10$ )						
$\beta_0$	-1.000	-.1006	-.0996	.0179	.0179	-.22
$\beta_1$	1.0000	.9985	.9990	.0087	.0088	-1.19
$\rho$	1.0000	.9972	.9915	.0561	.0561	-.35
$\sigma_\mu$	.8000	.7997	.7993	.0037	.0038	-.47
$\gamma_0$	-3.5000	-3.3219	-3.2916	.2840	.3352	4.43
$\gamma_1$	5.0000	4.7449	4.7237	.3041	.3969	-5.93
$\gamma_2$	2.0000	2.0537	2.0093	.1870	.1945	2.03
No Missing X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0968	-.0963	.0148	.0152	1.52
$\beta_1$	1.0000	.9982	1.0002	.0139	.0140	-.94
$\rho$	1.0000	.9902	.9904	.0863	.0868	-.80
$\alpha$	.5000	.5049	.4922	.0662	.0664	.53
$\sigma_\mu$	.8000	.8006	.8004	.0032	.0033	1.37
$\gamma_0$	-3.5000	-3.3357	-3.2874	.2398	.2907	4.84
$\gamma_1$	5.0000	4.7959	4.8140	.2442	.3183	-5.91
$\gamma_2$	2.0000	2.0410	2.0349	.1835	.1881	1.58
Missing X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0965	-.0957	.0159	.0163	1.54
$\beta_1$	1.0000	.9973	.9993	.0135	.0138	-1.43
$\rho$	1.0000	.9926	.9897	.0933	.0936	-.56
$\alpha$	.5000	.5082	.4982	.0698	.0731	.83
$\phi_2$	.2500	.2505	.2477	.0150	.0150	.23
$\sigma_\nu$	.5000	.5019	.5023	.0049	.0053	2.65
$\sigma_\mu$	.8000	.8007	.8007	.0034	.0035	1.41
$\gamma_0$	-3.5000	-3.3503	-3.3084	.2461	.2880	4.30
$\gamma_1$	5.0000	4.8156	4.8272	.2598	.3186	-5.02
$\gamma_2$	2.0000	2.0356	2.0363	.1779	.1815	1.41

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 in panels 2 and 3 are the same as in Table 1. The Markov model in panel 1 replaces  $\sum_{\tau=0}^{t-1} d_{i\tau} \rho_\tau$  in Table 1 with  $\rho d_{i,t-1}$ .



Table 17

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

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
Low Classification Error Bias ( $t = 1, \dots, 10$ )						
$\beta_0$	-1.000	-.1033	-.1039	.0406	.0407	.57
$\beta_1$	1.0000	1.0176	1.0114	.0649	.0673	1.91
$\rho$	1.0000	1.0322	1.0325	.0385	.0502	5.92
$\alpha$	.5000	.5017	.5050	.0461	.0461	.25
$\phi_2$	.2500	.2496	.2502	.0165	.0165	-.16
$\sigma_\nu$	.5000	.5018	.5023	.0049	.0052	2.62
$\phi_1$	.8000	.7987	.7961	.0264	.0265	-.35
$\gamma_0$	-3.5000	-3.4987	-3.4809	.0664	.0665	.14
$\gamma_1$	5.0000	4.9831	5.0056	.0697	.0717	-1.72
$\gamma_2$	2.0000	2.0265	2.0196	.0451	.0513	4.15
Medium Classification Error Bias ( $t = 1, \dots, 10$ )						
$\beta_0$	-1.000	-.0893	-.0982	.0525	.0536	1.44
$\beta_1$	1.0000	1.0075	1.0040	.0745	.0749	.71
$\rho$	1.0000	1.0283	1.0364	.0534	.0604	3.75
$\alpha$	.5000	.5162	.5101	.0540	.0563	2.12
$\phi_2$	.2500	.2478	.2469	.0163	.0164	-.94
$\sigma_\nu$	.5000	.5024	.5027	.0046	.0052	3.74
$\phi_1$	.8000	.8016	.8023	.0312	.0312	.35
$\gamma_0$	-3.0000	-3.0058	-3.0009	.0716	.0718	-.57
$\gamma_1$	4.0000	3.9802	3.9803	.0735	.0761	-1.90
$\gamma_2$	2.0000	2.0151	2.0227	.0659	.0676	1.62
High Classification Error Bias ( $t = 1, \dots, 10$ )						
$\beta_0$	-1.000	-.0926	-.0896	.0756	.0758	.69
$\beta_1$	1.0000	1.0135	1.0201	.0778	.0790	1.23
$\rho$	1.0000	1.0276	1.0255	.0682	.0735	2.86
$\alpha$	.5000	.5074	.5033	.0624	.0629	.83
$\phi_2$	.2500	.2476	.2446	.0152	.0153	-1.10
$\sigma_\nu$	.5000	.5019	.5030	.0051	.0055	2.62
$\phi_1$	.8000	.7980	.8046	.0386	.0387	-.36
$\gamma_0$	-3.0000	-3.0026	-2.9870	.0823	.0824	-.23
$\gamma_1$	3.0000	2.9899	2.9807	.0680	.0687	-1.04
$\gamma_2$	3.0000	3.0186	3.0185	.0693	.0717	1.90

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

Table 18

Repeated Sampling Experiments  
 AR(1) Error Model  
 Biased Classification Error  
 Low Classification Error Bias  
 Smooth Algorithm  
 (20% Missing Choices, No Initial Conditions Problem)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
No Missing X's (Markov model) ( $t = 1, \dots, 10$ )						
$\beta_0$	-1.000	-.0957	-.0974	.0135	.0142	2.26
$\beta_1$	1.0000	.9985	.9991	.0088	.0090	-1.19
$\rho$	1.0000	.9849	.9853	.0404	.0432	-2.65
$\phi_1$	.8000	.8009	.8008	.0031	.0032	2.12
$\gamma_0$	-3.5000	-3.3853	-3.3548	.3321	.3514	2.44
$\gamma_1$	5.0000	4.7582	4.7928	.3004	.3856	-5.69
$\gamma_2$	2.0000	2.1277	2.1208	.2004	.2376	4.51
No Missing X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0957	-.0963	.0187	.0192	1.64
$\beta_1$	1.0000	1.0003	1.0022	.0108	.0108	.17
$\rho$	1.0000	1.0021	.9954	.0993	.0993	.15
$\alpha$	.5000	.5048	.4999	.0450	.0453	.76
$\phi_1$	.8000	.8009	.8005	.0039	.0039	1.58
$\gamma_0$	-3.5000	-3.3875	-3.3778	.2774	.2993	2.87
$\gamma_1$	5.0000	4.8043	4.8417	.2679	.3318	-5.16
$\gamma_2$	2.0000	2.1189	2.1113	.1795	.2153	4.68
Missing X's ( $t = 1, \dots, 10$ )						
$\beta_0$	-.1000	-.0954	-.0958	.0185	.0191	1.75
$\beta_1$	1.0000	1.0003	1.0014	.0108	.0108	.20
$\rho$	1.0000	1.0017	.9906	.0983	.0983	.12
$\alpha$	.5000	.5046	.5012	.0443	.0445	.73
$\phi_2$	.2500	.2497	.2497	.0133	.0133	-.14
$\sigma_\nu$	.5000	.5019	.5024	.0047	.0050	2.83
$\phi_1$	.8000	.8009	.8004	.0038	.0039	1.63
$\gamma_0$	-3.5000	-3.3885	-3.3811	.2752	.2970	2.86
$\gamma_1$	5.0000	4.8058	4.8436	.2676	.3307	-5.13
$\gamma_2$	2.0000	2.1185	2.1083	.1785	.2143	4.69

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 in panels 2 and 3 are the same as in Table 8. The Markov model in panel 1 replaces  $\sum_{\tau=0}^{t-1} d_{i\tau} \rho_\tau$  in Table 8 with  $\rho d_{i,t-1}$ .