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

Michael P. Keane
Yale University

Robert M. Sauer
The Hebrew University of Jerusalem

June 2001
(Preliminary Version)

1 Introduction

This paper assesses the performance of a new simulated maximum likelihood (SML) estimator that is particularly useful for estimating dynamic panel data models with unobserved endogenous state variables and classification error in discrete choices. The novel estimation technique was recently introduced in Keane and Wolpin (2001*a*) in order to estimate the parameters of a discrete choice dynamic programming problem. The technique has since been applied, to estimate models of the same type, in Keane and Wolpin (2001*b*) and Sauer (2001). In this study, the performance of the estimator is assessed via repeated sampling experiments on a panel data probit model with a time-varying exogenous covariate, lagged endogenous variables and serially correlated errors. A simple panel data probit model is specified, rather than the more complex models to which the technique has previously been applied, in order to focus on the estimation method. The results of a series of repeated sampling experiments that allow for state-dependence, individual random effects and $AR(1)$ serially correlated errors show that the new SML estimator has good small sample properties and is computationally tractable for use with panels of the size that are likely to be encountered in practice.

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 high dimensional integrations that must be performed render the integrating out solution computationally impractical.

Several tractable solutions to the problem of unobserved initial conditions were considered in Heckman (1981a). Heckman illustrated how one could employ the simplifying assumption of equilibrium in the dynamic process at the start of the sample period 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.[1] However, assuming equilibrium at the start of the sample period is questionable when there are time or age trends in the exogenous covariates in the model. Heckman also noted that, as an alternative strategy, fixed effects could be estimated thus avoiding the initial conditions problem as well as specification of the mixing distribution of individual effects. However, in nonlinear models of fixed panel length, the inconsistency in the fixed effects estimates is transmitted to the structural parameters.[2] Moreover, repeated sampling experiments on the fixed effects probit model tend to produce estimates with considerable biases. As an alternative to assuming equilibrium or estimating fixed effects, Heckman suggested approximating the marginal probability of the initial state by a probit function where the error in the initial state index function is correlated with the errors in the index functions during the sample period.[3] This latter estimation strategy generally performs better than the fixed effects probit but still produces parameter estimates with disturbingly large biases.

The problem of unobserved endogenous state variables during the sample period is parallel to the problem of initial conditions. However, there is little discussion in

[1] Card and Sullivan (1988) adopt this strategy for studying the dynamic effects of training on re-employment probabilities.

[2] In the special case of the conditional logit model it is possible to consistently estimate fixed effects. For analyses of the conditional logit model see Chamberlain (1984) and Honore and Kyriazidou (2000).

[3] This approximation method was recently employed in Hyslop (1999). Heckman and Singer (1984) also suggested a similar type of solution to the initial conditions problem in multiple spell duration models. See Ham and LaLonde (1996) for an application.

the econometrics literature on alternative practical solutions to the challenges posed by "missingness".[4] The SML technique advocated in this paper not only provides a computationally practical alternative solution to the initial conditions problem but also simultaneously "solves" the problem of missingness during the sample period. The SML solution does not require approximating the marginal probability of the initial state or the marginal probability of the missing states during the sample period, but is rather based on matching simulated and reported choices and assuming a classification error process. The key computational advantage of the SML solution lies in the fact that the individual's likelihood contribution is not conditioned on the individual's state variables but is rather conditioned on the individual's simulated outcomes.

The classification error process that is incorporated into the estimation technique simply assumes that there is some probability that the reported choice is the true (simulated) choice and some probability that it is not. This type of classification error process for discrete choices has been previously analyzed. For example, Poterba and Summers (1995) find that the incorporation of empirical classification error rates into the likelihood function of a multinomial logit model substantially alters the conclusions on the effect of unemployment insurance benefits on the duration of unemployment.. More formally, Hausman, Abrevaya and Scott-Morton (1998) demonstrate how classical maximum likelihood estimation leads to biased and inconsistent parameter estimates when misclassification is present and not taken into account in estimation.[5]

The rest of this paper is organized as follows. Section II specifies the panel data

[4]Probit missingness has been analyzed in the statistical literature (Rubin (1976), Nordheim (1978)). Probit missingness is similar in spirit to the Heckman strategy for dealing with the initial conditions problem.

[5]Hausman, Abrevaya and Scott-Morton (1998) also note that a distributional assumption on the error term and a monotonicity condition are required for identification of classification error rates.

probit model and the classification error process in the dependent variable. Section III describes the estimation procedure. Section IV outlines the data generating process and presents Monte-Carlo test results for both a random effects model and an $AR(1)$ error model. Section V 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. u_{it} is assumed to take 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. Possible depreciation in the importance of past choices is captured through the weights ρ_τ.

The exogenous covariate x_{it} in (1) is allowed to be time-varying and stochastic. For example, x_{it} could follow an $AR(1)$ process,

$$x_{it} = \phi_1 x_{i,t-1} + \nu_{it} \quad (3)$$

where ν_{it} is i.i.d. with zero mean and variance σ_v^2.

The error term ε_{it} in (1) is also allowed to be serially correlated. For example, ε_{it} could contain time-invariant individual effects

$$\varepsilon_{it} = \mu_i + \eta_{it} \quad (4)$$

where μ_i is i.i.d. with zero mean and variance σ_μ^2. Alternatively, ε_{it} could follow the $AR(1)$ process

$$\varepsilon_{it} = \phi_2 \varepsilon_{i,t-1} + \eta_{it} \tag{5}$$

where η_{it} is i.i.d. with zero mean and variance σ_η^2.

The theoretical start of the process is, by definition, $d_{i0} = x_{i0} = 0$. However, the first time period, $t = 1$, may not be the first period of observed data. There may also be missing data during the sample period. There are thus four possible cases of missing data in any period t: *a)* d_{it} observed, x_{it} observed, *b)* d_{it} unobserved, x_{it} observed, *c)* d_{it} observed, x_{it} unobserved and *d)* d_{it} unobserved, x_{it} unobserved. An unobserved x_{it} in any period t can be handled by specifying and estimating the stochastic process generating the exogenous regressors. An unobserved d_{it} in any period t can be similarly handled by specifying and estimating a classification error process in the dependent variable.

The model of misclassification that we consider is defined by the following four classification error rates

$$\begin{aligned}
\pi_{11t} &= \Pr(d_{it} = 1 \text{ reported} \mid d_{it} = 1 \text{ occurred }) \\
\pi_{01t} &= \Pr(d_{it} = 1 \text{ reported} \mid d_{it} = 0 \text{ occurred }) \\
\pi_{00t} &= 1 - \pi_{01t} \\
\pi_{10t} &= 1 - \pi_{11t}
\end{aligned} \tag{6}$$

where π_{11t} is the probability that the first option is reported to be chosen ($d_{it} = 1$ reported) given that the first option is the true choice ($d_{it} = 1$ occurred); π_{01t} is the probability that the first option is reported to be chosen ($d_{it} = 1$ reported) given that the second option is the true choice ($d_{it} = 0$ occurred); and π_{00t} and π_{10t} are the corresponding conditional probabilities of reporting $d_{it} = 0$. The classification error rates are thus allowed to depend on the true choice but are otherwise assumed to be unconditional on the covariates in the model.

It is also assumed that the classification error process in (6) is unbiased. That is, the probability that a person is observed to choose an option is assumed to be equal to the true probability that the person chooses that option, or $\Pr(d_{it} = 1 \text{ reported}) = \Pr(d_{it} = 1 \text{ occurred})$. This assumption forces the structural parameters of the model to fit the conditional choice frequencies rather than allowing classification error to drive model fit.

The unbiasedness assumption is consistent with classification error rates that are linear in the true choice probability. To see this, let Y be the reported choice and let T be the true choice generated by the model. By definition,

$$\Pr(Y = 1) = \Pr(Y = 1 \mid T = 1)\Pr(T = 1) + \Pr(Y = 1 \mid T = 0)\Pr(T = 0) \qquad (7)$$

and by the assumption that $\Pr(Y = 1) = \Pr(T = 1)$, the classification error rates

$$\begin{aligned} \Pr(Y = 1 \mid T = 1) &= E + (1 - E)\Pr(T = 1) \\ \Pr(Y = 1 \mid T = 0) &= (1 - E)\Pr(T = 1) \end{aligned} \qquad (8)$$

can be substituted into (7) and shown to yield $\Pr(T = 1)$. Note that as the true choice probability approaches one, the classification error rate $\Pr(Y = 1 \mid T = 1)$ also approaches one, which must be the case to preserve unbiasedness. Further, as $\Pr(T = 1)$ approaches zero, $\Pr(Y = 1 \mid T = 1)$ approaches E. E can thus be interpreted as a "base" classification error rate. $\Pr(Y = 1 \mid T = 1)$ increases linearly above E as the true choice probability $\Pr(T = 1)$ grows. In estimation, E is treated as a free parameter, where $0 < E < 1$.

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

$$\begin{aligned} \pi_{11t} &= E + (1 - E)\Pr(d_{it} = 1 \text{ occurred}) \\ \pi_{01t} &= (1 - E)\Pr(d_{it} = 1 \text{ occurred}) \\ \pi_{00t} &= 1 - \pi_{01t} \\ \pi_{10t} &= 1 - \pi_{11t}. \end{aligned} \qquad (9)$$

These classification error rates will serve as the input to the individual's likelihood contribution.

3 Estimation

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 regressor 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, however, be missing choices in the data, let $I(d_{it}^* \neq NA)$ be an indicator function which equals one if d_{it}^* is observed, and zero otherwise, where NA denotes "not available." Under these conditions, estimation of the model requires constructing M simulated choices for each $\{x_{it}\}_{t=1}^T$ history as follows:

1. Draw M times from the ε_{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 choices for each individual i in every period t $\left\{\left\{\{d_{it}^m\}_{t=1}^T\right\}_{i=1}^N\right\}_{m=1}^M$ according to (1) and the decision rule (2).

3. In order to calculate the classification error rates in (9), compute the true choice probability $\Pr(d_{it} = 1 \text{ occurred})$ by using the unbiased frequency simulator

$$\widehat{P}_{it} = \widehat{P}\left(d_{it} = 1 \mid H_{it}^m\right) = \frac{1}{M}\sum_{m=1}^{M} \Pr\left(\varepsilon_{it} \leq \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}^m \rho_\tau\right) \tag{10}$$

   where $H_{it}^m = \left\{\{x_{i\tau}\}_{\tau=1}^t, \{d_{i\tau}^m\}_{\tau=1}^t\right\}$ is the history of the exogenous regressor and the simulated endogenous regressor through time t.[6]

4. Construct the classification error rates $\widehat{\pi}_{jkt}$ for each individual i using $\widehat{P}_{it}$ in place of $\Pr(d_{it} = 1 \text{ occurred})$.

[6] In the case of a standard normal ε_{it}, the probability in the summation is $\Phi(a)$ where $a = \left(\beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} d_{i\tau}^m \rho_\tau\right)$ and Φ is the standard normal c.d.f. Note that the c.d.f. yields a smooth simulator for $\Pr(d_{it} = 1 \text{ occurred})$.

5. Form the likelihood contribution for each individual i as:

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

where θ is the vector of model parameters.

The estimation procedure thus builds the likelihood contribution for each individual by averaging, over M simulated choice histories, the product of the appropriate classification error rates implied by simulated choice history $\{d_{it}^m\}_{t=1}^T$. The indicator function $I\left[d_{it}^m = j, d_{it}^* = k\right]$ "picks out" the appropriate classification error rate by comparing d_{it}^* to d_{it}^m whenever d_{it}^* is observed. If d_{it}^* is unobserved, then the value of $I\left(d_{it}^* \neq NA\right)$ is zero, and there is no contribution to the likelihood (or one enters the product) in period t. Note that each simulated choice history occurs with non-zero probability, reflecting the fact that any simulated choice history could be the true choice history when admitting classification error.

It is important to note that the estimation procedure accommodates serial correlation in the error term through the draws on the ε_{it} distribution and the resulting simulated choice history, which determines the appropriate classification error rates that enter the likelihood contribution. Similarly, the state space is updated according to previous simulated choices, rather than previous reported choices, which then determine current simulated choices. Thus, serial correlation does not require the use of a recursive simulator (such as GHK) and missing endogenous state variables do not necessitate integrating out over all possible choice sequences.

The estimation procedure needs to be only trivially modified to accommodate missing exogenous covariates and/or an initial conditions problem. In the case of missing exogenous covariates only, each missing x_{it} is simulated according to the assumed process generating the x_{it}'s. The parameters of this process are estimated along with the other parameters of the model. For example, if the x_{it}'s follow an $AR(1)$ process, x_{it-1} is observed and x_{it} is unobserved, then the missing x_{it} is simulated to be $\phi_2 x_{it-1}$ plus an i.i.d. random shock. ϕ_2.is the additional estimable parameter.[7]

In the case of an initial conditions problem, the path of the d_{it}'s and x_{it}'s are simulated in the same way as described above, i.e., from $d_{i0} = x_{i0} = 0$. However, classification error rates enter the likelihood contribution only once the first period of observed data is reached. As in the previous case, the parameters governing the stochastic process of the exogenous regressors are estimated. It is thus not necessary to specify a marginal probability for the unobserved initial state. The distribution of the initial conditions is implicitly determined by the prior history of simulated choices and exogenous covariates generated by the parameters of the model.

4 Monte-Carlo Tests

In this section the results of a series of Monte-Carlo tests of the SML estimator are presented. First, the data generating process is outlined and then the Monte-Carlo test results are discussed for both a random effects and an $AR(1)$ specification for the error term $\varepsilon_{it.}$. In each repeated sampling experiment, a vector of true model parameters is specified and used to create 50 Monte-Carlo data sets. Parameter estimates are then obtained for each data set. Each estimation uses a different seed for the random elements of the model that generate the simulated exogenous covariates and the simulated choices. However, the random draws are fixed over trial parameter iterations for a given data set. For each repeated sampling experiment, the true parameters, the mean, the median, the empirical standard deviations, the 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.

Given the true parameters of the model and $d_{i0} = x_{i0} = 0$, 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

[7] If $x_{i,t-1}$ is unobserved then the simulated $x_{i,t-1}$ is used to generate the unobserved x_{it}.

rates. The second stage consists of generating the sequence of true and reported choices together with a determination of when reported choices are missing. For ease of exposition, the case of missing choices only is considered. The two stages of the data generating process are outlined below.

4.1 Data Generating Process - Stage 1

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

$$P_{it} = P\left(\widetilde{d}_{it} = 1 \mid H_{it}^{m}\right) = \frac{1}{\widetilde{M}} \sum_{m=1}^{\widetilde{M}} \Pr\left(\varepsilon_{it} \leq \beta_0 + \beta_1 x_{it} + \sum_{\tau=0}^{t-1} \widetilde{d}_{i\tau}^{m} \rho_{\tau}\right) \qquad (14)$$

   where $H_{it}^{m} = \left\{ \{x_{i\tau}\}_{\tau=1}^{t}, \left\{ \widetilde{d}_{i\tau}^{m} \right\}_{\tau=1}^{t} \right\}$.
5. Construct the "true" classification error rates π_{jkt} for each individual i, according to (9), using P_{it} in place of $\Pr(d_{it} = 1 \text{ occurred})$.

4.2 Data Generating Process - Stage 2

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

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

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

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

6. Compare the uniform draws to the probability π^{NA} that d_{it}^{*} is missing in period t. That is, implement the following rule: if $\widetilde{U}_{it} < \pi^{NA}$ then $d_{it}^{*} = NA$, else $d_{it}^{*} \neq NA$.

Each data set thus consists of the panel $\{D_i^{*}, x_i\}_{i=1}^{N}$, where d_{it}^{*} may be not available in every period t. Generating missing exogenous covariates and/or missing choices that depend on the value of the exogenous covariates simply involves generalizing π^{NA}.

4.3 Random Effects Model

In the random effects model, the error term ε_{it} follows the components of variance structure in (4) with both μ_i and η_{it} assumed to be normally distributed. The exogenous covariate x_{it} is also assumed to be normally distributed and generated by the $AR(1)$ process in (3). The depreciation weights ρ_τ are specified to follow an exponential decay process, $\rho_\tau = \rho e^{-\alpha(t-\tau)}$. The parameter α captures the "speed" of depreciation in past choices. The vector of estimable parameters for this model is $\theta = \{\beta_0, \beta_1, \rho, \alpha, \sigma_\mu, E\}$ when there is no initial conditions problem and no missing exogenous covariates. The vector of estimable parameters is $\theta = \{\beta_0, \beta_1, \phi_1, \rho, \alpha, \sigma_\mu, E\}$ when there are missing exogenous covariates during the sample period or there is an initial conditions problem.[8]

The first set of three experiments sets the number of individuals N to 400 and the number of periods T to 10. It is assumed that there is no initial conditions problem and that there are no missing exogenous covariates during the sample period. The vector of true estimable parameters is set to $\theta = \{-.10, 1.00, 1.00, .50, .50, .75\}$. The process generating the exogenous covariate has an $AR(1)$ parameter ϕ_1 set to .25 and standard deviation σ_v set to .25. The standard deviation of the time varying component of ε_{it}, σ_η, is also set to .25. The frequency simulators, used both in the data generating process and in estimation, are formed with the number of draws $\widetilde{M}$ and M set to 500. The three experiments differ only in the proportion of missing choices in each period. Summary statistics, by time period, over the 50 datasets generated by this configuration of parameters are displayed in Table 1.

The first two columns in Table 1 show that there is an increasing proportion of individuals over time that are simulated to choose the first option, i.e., $d_{it} = 1$. The third and fourth columns display the mean and variance of $\beta' x = \beta_1 x_{it} + \rho \sum_{\tau=0}^{t-1} e^{-\alpha(t-\tau)} d_{i\tau}$ and the error term ε_{it}. The figures show that the mean of $\beta' x$ increases at a decreasing rate reflecting the increasing proportion of $d_{it} = 1$ the strong depreciation of past choices. Note also that the variances of $\beta' x$ and ε_{it} are roughly comparable.

The last two columns of Table 1 present the mean classification error rates for correct matches of simulated and reported choices. The mean classification error rate

[8] Identification conditions for this model (also known as a generalized Polya process) are discussed in Heckman (1981*b*). In short, the parameters generating serial correlation, as specified in (4) or (5) are identified off of transitions from $d_{it} = 0$. The structural state dependence parameters are identified off of transitions from $d_{it} = 1$.

for a correct match of $d_{it} = 1$ and $d^*_{it} = 1$, π_{11t}, is .865 in period 1 and increases over time to .964 in period 10. The mean classification error rate for a correct match of $d_{it} = 0$ and $d^*_{it} = 0$, π_{00t}, is .885 in period 1 and decreases over time to .786 in period 10. This pattern emerges since the proportion reporting $d^*_{it} = 1$ in each period must match the increasing true proportion, following the requirement of unbiased classification error. The high base classification error rate of $E = .75$ implies a low incidence of classification error and relatively large probabilities for correct matches. Decreases in E reduce the classification error rates for correct matches and increase the classification error rates for mismatches for a given proportion of simulated choices.[9]

The estimation results for the first set of three repeated sampling experiments are presented Table 2. The first panel of the table displays the results when there are no missing choices in the data. As the figures illustrate, the biases are quite small in magnitude. The biases in the estimates of ρ and σ_μ are the largest among the estimates, and the only statistically significant biases, but the magnitudes are only 3.4 percent and 2.8 percent, respectively. Note that the base classification error rate E is recovered with a bias of only 0.4 percent. The medians of the parameter estimates are also quite close to the means suggesting that the sampling distributions are symmetric.

The next two panels of Table 2 display the estimation results for 10 percent and 20 percent of the choices missing in each period, respectively. The biases in the parameter estimates retain the same pattern and remain negligible with no substantial differences in the empirical standard errors. The increased incidence of missing choices should not change the point estimates much since the missing choices do not substantially alter the reported choice frequencies.[10] The effect of a higher proportion of missing

[9]In estimation, the structural parameters of the model, and hence the simulated choice frequencies, adjust together with E to fit the reported choice frequencies while allowing for individual misreporting.

[10]This would be true even if choices were not missing at random since the estimation method fits

choices is thus only to reduce the effective sample size.[11]

The estimation results for the second set of three experiments are presented in Table 3. In all three experiments the number of individuals N is increased to 500. The first panel displays the results for 20 percent of the choices missing in each period. The biases in the parameter estimates remain negligible. The largest biases are again in the estimates of ρ and σ_μ, with biases of 3.17 percent and 3.18 percent, respectively. However, the increase in sample size and consequent decrease in the empirical standard errors produces higher t-statistics and more statistically significant biases.

The second panel of Table 3 displays the results for 20 percent of the choices and 20 percent of the exogenous covariates missing in each period. The exogenous covariates were coded as missing whenever the choice was determined to be unobserved. In this case, the $AR(1)$ parameter of the process generating the exogenous covariates ϕ_1 is estimated along with the other parameters of the model. The figures show that the biases remain small and are not substantially different in magnitude from the previous panel. The bias in ϕ_1, however, is relatively large at 9.8 percent.

The third panel of Table 3 displays the results for 20 percent of the choices missing, 20 percent of the exogenous covariates missing and higher variances for the distribution of the individual effects and the distribution of the exogenous covariate. Specifically, σ_μ is increased from .50 to 1.00 and σ_v is increased from .25 to .75. The biases in the parameter estimates are not substantially different from the previous panel. The decrease in the statistical significance of the biases is due to the increase in the empirical standard errors of the estimates that result from the higher variances.

The third and fourth sets of three experiments are identical in form to the first and second sets, respectively, but have an E set to .50. The decrease in E from .75

conditional choice frequencies.

[11]Increasing the proportion of missing choices beyond 35 percent (not shown) begins to increase the empirical standard errors.

to .50 increases the incidence of classification error in reported choices. For example, the minimum π_{10} (in period 10) is doubled from .061 to 122. The fifth and sixth set of experiments are also of the same form as the previous two sets but have an E set to .25. An E of .25 further increases the incidence of classification error. The minimum π_{10} in this latter case is .183. The estimation results from these experiments are presented in Tables 4 through 7. The biases in the parameter estimates are generally small and similar in magnitude to those obtained with E set at .75. The effects of the increased incidence of classification error is to increase the empirical standard errors of the estimates and reduce the statistical significance of the biases. These latter effects are most noticeable when comparing the results for $E = .75$ to the results for $E = .25$.

The final set of experiments on the random effects model introduces an initial conditions problem together with 20 percent of the choices and exogenous covariates missing during the sample period. The number of individuals N is 500 and the total number of time periods T is 20. The simulated choices and covariates in the estimation procedure are generated starting from $t = 1$ but the first period of reported data is assumed to be $t = 11$. The first panel displays the estimation results for E set at .75. The figures show that the biases in the parameter estimates are not substantially affected by the introduction of an initial conditions problem. The magnitudes and patterns in the biases are similar to the corresponding experiment with no initial conditions problem. The second and third panels displays the results of the same experiment with E set at .50 and E set at .25, respectively. The effects of increasing the incidence of classification error is, again, to increase the empirical standard errors and reduce the statistical significance of the biases. This last set of three experiments confirm that the estimation method is quite robust to the simultaneous introduction of an initial conditions problem and missing unobserved endogenous state variables during the sample period.

4.4 $AR(1)$ Error Model

In the $AR(1)$ error model, the error term ε_{it} follows the process in (5) with η_{it} assumed to be normally distributed. As in the random effects model of the previous subsection, the exogenous covariate x_{it} is normally distributed and generated by the $AR(1)$ process in (3) and the depreciation weights ρ_τ follow the exponential decay process, $\rho_\tau = \rho e^{-\alpha(t-\tau)}$. The vector of estimable parameters for the $AR(1)$ error model is thus $\theta = \{\beta_0, \beta_1, \rho, \alpha, \phi_2, E\}$ when there is no initial conditions problem and no missing exogenous covariates. The vector of estimable parameters is $\theta = \{\beta_0, \beta_1, \phi_1, \rho, \alpha, \phi_2, E\}$ when there is an initial conditions problem or there are missing exogenous covariates during the sample period.

The order of the experiments for the $AR(1)$ error model is the same as in the random effects model. The first six experiments have the vector of true estimable parameters set at $\theta = \{-.10, 1.00, 1.00, .50, .50, .75\}$. The second and third sets of six experiments increase the incidence of classification error by replacing $E = .75$ with $E = .50$ and $E = .25$, respectively. As in the random effects model, the parameters ϕ_1, σ_v and σ_η are set to .25 and $\widetilde{M}$ and M are each set to 500. The summary statistics generated by the parameters in the $AR(1)$ error model are quite similar to those presented in Table 1.

The estimation results for the first six experiments with $E = .75$ are presented Tables 9 and 10. The results in Table 9 with $N = 400$ and an increasing incidence of missing choices show that the biases are negligible in magnitude. The largest bias, 3.1 percent, is in the estimate of ρ. Although obviously small in magnitude, the biases in ρ (and β_1) are statistically significant in all three panels. The effect of increasing N to 500 is displayed in the first panel of Table 10. The biases remain small in magnitude. The bias in the estimate of ρ increases slightly, to 3.4 percent and the bias in E becomes marginally significant due to the decrease in the empirical standard error. The second panel of the table displays the effect of introducing missing exogenous covariates. As in the random effects model, the estimation method has difficulty estimating the $AR(1)$ parameter of the process generating the exogenous covariates. Increasing the variance of the error term σ_η to .50 reduces the statistical significance of bias in ϕ_1, as its standard error increases,.and also increases the bias in ϕ_2. Nonetheless, in both of these latter cases, the structural parameter estimates are recovered with negligible biases.

Tables 11 through 14 display the results of the repeated sampling experiments with higher incidences of classification error, $E = .50$ and $E = .25$. The same pattern of results emerge as in the case of $E = .25$. The biases in the structural parameter estimates are negligible while the bias in the $AR(1)$ parameter generating the exogenous covariate is relatively larger. In general, the statistical significance of the bias decreases due to the higher standard errors that result from a higher degree of classification error. The final set of three experiments, which introduce an initial conditions problem at different levels of classification error, are reported in Table 15. The general conclusions are unchanged. The estimation method is also robust in the $AR(1)$ error model to the simultaneous introduction of missing endogenous state variables before and during the sample period.

5 Conclusion

This paper assessed the performance of a new, computationally practical SML estimator for dynamic panel data models with unobserved endogenous state variables and classification error in discrete choices. The estimation technique studied in this paper simultaneously "solves" the problem of initial conditions and the problem of missingness during the sample period by matching simulated and reported choices and assuming a classification error process.

A series of repeated sampling experiments on a panel data probit model with a time-varying exogenous covariate, lagged endogenous variables and serially correlated errors showed that the SML estimator has good small sample properties and is computationally tractable. Specifically, repeated sampling experiments were performed assuming different proportions of missing choices, exogenous covariates and classification error on both a random effects model and an $AR(1)$ error model. The results of the experiments, for both error specifications, indicate that the estimation method recovers the true structural parameters of the model with negligible biases. In the case of missing exogenous covariates, the estimation method performs less well in recovering the parameters that generate these covariates. However, in most applications, accurately recovering the parameters of this latter process will not be of central interest.

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 the estimation method by specifying measurement error densities that enter the likelihood contribution. The performance of the method should also be examined in a multinomial probit framework both with and without the nested solution of a dynamic program. Lastly, the estimation method can be adapted to take advantage of importance sampling re-weighting techniques that can substantially reduce the computational burden.

References

[1] Card D. and Daniel Sullivan (1988) , "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, 56, 497-530.

[2] 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.

[3] Ham, J.C., and Robert J. LaLonde (1996), "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence From Experimental Data on Training", *Econometrica*, 64, 175-205.

[4] 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.

[5] 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).

[6] 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).

[7] Heckman, J., and B. Singer (1984), "Econometric Duration Analysis," *Journal of Econometrics*, 24, 63-132.

[8] Honore, B.E., and Ekaterini Kyriazidou (2000), "Panel Data Discrete Choice Models With Lagged Dependent Variables," *Econometrica*, 68, 839-874.

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

[10] Keane, M.P., and K.I. Wolpin, (2001*a*), "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," *International Economic Review*, forthcoming.

[11] Keane, M.P., and K.I. Wolpin, (2001*b*), "Estimating the Effect of Welfare on the Education, Employment, Fertility and Marriage Decisions of Women", unpublished manuscript.

[12] Nordheim, E.V. (1978), "Inference from Nonrandomly Missing Data," unpublished Ph.D. dissertation, University of Minnesota.

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

[14] Rubin, D.B. (1976) "Inference and Missing Data," *Biometrika*, 63, 581-592.

[15] Sauer, R.M. (2001), "Educational Financing and Lifetime Earnings," unpublished manuscript.

Table 1

Random Effects Model
Summary Statistics

t	d_{it}	$\beta' x$	ε_{it}	π_{11t}	π_{00t}
1	.4336	.0010 (.0027)	.0027 (.2525)	.8654	.8846
2	.5703	.2661 (.1690)	.0028 (.3123)	.8885	.8615
3	.6702	.5096 (.2654)	-.0032 (.3251)	.9076	.8424
4	.7402	.7146 (.3243)	.0014 (.3349)	.9225	.8275
5	.7961	.8826 (.3427)	-.0040 (.3382)	.9344	.8156
6	.8371	1.0169 (.3375)	.0023 (.3372)	.9435	.8065
7	.8665	1.1247 (.3127)	-.0060 (.3383)	.9507	.7993
8	.8930	1.2078 (.2847)	-.0026 (.3358)	.9563	.7937
9	.9105	1.2751 (.2577)	-.0033 (.3408)	.9607	.7893
10	.9245	1.3264 (.2291)	.0043 (.3414)	.9640	.7860

Note: d_{it} is the simulated choice, d^*_{it} is the reported choice, $\beta' x = \beta_1 x_{it} + \rho \sum_{\tau=0}^{t-1} e^{-\alpha(t-\tau)} d_{i\tau}$, ε_{it} is the error term and π_{11} and π_{00} are the classification error rates for correct classifications. Standard deviations are in parentheses.

Table 2

Repeated Sampling Experiment
Random Effects Model
$(R = 50,\ N = 400,\ T = 10,\ E = .75)$

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
No Missing Choices						
β_0	-.1000	-.0934	-.0955	.0276	.0284	1.69
β_1	1.0000	1.0073	1.0068	.0611	.0616	.85
ρ	1.0000	1.0335	1.0348	.0563	.0655	4.21
α	.5000	.5023	.4971	.0519	.0520	.31
σ_μ	.5000	.5138	.5107	.0357	.0383	2.73
E	.7500	.7467	.7428	.0158	.0161	-1.49
10% Missing Choices						
β_0	-.1000	-.0970	-.1021	.0298	.0300	.72
β_1	1.0000	1.0130	1.0239	.0667	.0680	1.38
ρ	1.0000	1.0394	1.0418	.0589	.0709	4.73
α	.5000	.4970	.5057	.0638	.0638	-.33
σ_μ	.5000	.5141	.5105	.0450	.0471	2.21
E	.7500	.7464	.7449	.0160	.0164	-1.59
20% Missing Choices						
β_0	-.1000	-.0977	-.994	.0269	.0270	.60
β_1	1.0000	1.0162	1.0235	.0673	.0692	1.70
ρ	1.0000	1.0271	1.0229	.0530	.0595	3.62
α	.5000	.5023	.5023	.0474	.0475	.34
σ_μ	.5000	.5113	.5092	.0295	.0316	2.71
E	.7500	.7484	.7488	.0168	.0169	-.67

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 1.85 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 3

Repeated Sampling Experiment
Random Effects Model
($R = 50,\ N = 500,\ T = 10,\ E = .75$)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
20% Missing Choices						
β_0	-.1000	-.1071	-.1092	.0245	.0256	-2.06
β_1	1.0000	1.0154	1.0174	.0476	.0500	2.29
ρ	1.0000	1.0317	1.0239	.0537	.0623	4.18
α	.5000	.5001	.5013	.0515	.0515	.02
σ_μ	.5000	.5159	.5200	.0460	.0487	2.44
E	.7500	.7457	.7440	.0147	.0153	-2.09
20% Missing Choices and X's						
β_0	-.1000	-.1101	-.1075	.0271	.0290	-.264
β_1	1.0000	1.0240	1.0177	.0549	.0599	3.09
ρ	1.0000	1.0292	1.0304	.0459	.0544	4.50
α	.5000	.5009	.5116	.0524	.0524	.12
ϕ_1	.2500	.2787	.2741	.0404	.0496	5.02
σ_μ	.5000	.5114	.5153	.0475	.0489	1.70
E	.7500	.7465	.7456	.0136	.0140	-1.80
20% Missing Choices and X's (Higher Variances)						
β_0	-.1000	-.0989	-.0925	.0499	.0500	.15
β_1	1.0000	1.0195	1.0117	.0356	.0406	3.87
ρ	1.0000	1.0070	1.0041	.0362	.0369	1.37
α	.5000	.5055	.5012	.0238	.0244	1.64
ϕ_1	.2500	.2747	.2770	.0467	.0528	3.74
σ_μ	1.0000	1.0112	1.0056	.0421	.0435	1.89
E	.7500	.7497	.7492	.0131	.0132	-.15

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 2.21 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 4

Repeated Sampling Experiment
Random Effects Model
$(R = 50,\ N = 400,\ T = 10,\ E = .50)$

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$\widehat{Std(\beta)}$	$RMSE$	t-Stat
No Missing Choices						
β_0	-.1000	-.1010	-.1037	.0260	.0260	-.26
β_1	1.0000	1.0021	1.0072	.0715	.0715	.20
ρ	1.0000	1.0408	1.0340	.0558	.0691	5.17
α	.5000	.5045	.4960	.0538	.0540	.59
σ_μ	.5000	.5088	.5154	.0399	.0408	1.56
E	.5000	.4979	.4990	.0201	.0202	-.74
10% Missing Choices						
β_0	-.1000	-.0992	-.1008	.0251	.0252	.21
β_1	1.0000	1.0004	1.0136	.0847	.0847	.04
ρ	1.0000	1.0341	1.0331	.0478	.0587	5.05
α	.5000	.5067	.4996	.0702	.0705	.67
σ_μ	.5000	.5060	.5150	.0521	.0525	.81
E	.5000	.4970	.4962	.0198	.0200	-1.08
20% Missing Choices						
β_0	-.1000	-.1017	-.1037	.0243	.0243	-.50
β_1	1.0000	1.0237	1.0190	.1536	.1554	1.09
ρ	1.0000	1.0515	1.0371	.1214	.1319	3.00
α	.5000	.5018	.5038	.0559	.0550	.23
σ_μ	.5000	.5181	.4995	.0777	.0798	1.65
E	.5000	.4977	.4995	.0212	.0213	-.77

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $\widehat{Std(\beta)}$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t statistics are calculated as $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 1.36 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 5

Repeated Sampling Experiment
Random Effects Model
($R = 50,\ N = 500,\ T = 10,\ E = .50$)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
20% Missing Choices						
β_0	-.1000	-.1042	-.1041	.0264	.0267	-1.13
β_1	1.0000	1.0275	1.0136	.0629	.0686	3.10
ρ	1.0000	1.0210	1.0200	.0555	.0593	2.68
α	.5000	.5024	.5097	.0606	.0607	.28
σ_μ	.5000	.5196	.5205	.0476	.0515	2.91
E	.5000	.4975	.4953	.0197	.0198	-.89
20% Missing Choices and X's						
β_0	-.1000	-.1031	-.1021	.0291	.0293	-.75
β_1	1.0000	1.0221	1.0204	.0562	.0604	2.79
ρ	1.0000	1.0227	1.0151	.0620	.0660	2.59
α	.5000	.5078	.5095	.0490	.0496	1.13
ϕ_1	.2500	.2634	.2655	.1059	.1068	.90
σ_μ	.5000	.5154	.5153	.0482	.0506	2.26
E	.5000	.4992	.4969	.0187	.0187	-.30
20% Missing Choices and X's (Higher Variances)						
β_0	-.1000	-.0993	-.1015	.0538	.0538	.09
β_1	1.0000	1.0071	1.0035	.0715	.0718	.70
ρ	1.0000	1.0166	1.0153	.0580	.0603	2.03
α	.5000	.5151	.5144	.0436	.0462	2.45
ϕ_1	.2500	.2536	.2774	.1456	.1456	.18
σ_μ	1.0000	1.0052	.9993	.0650	.0652	.56
E	.5000	.5001	.4992	.0202	.0202	.04

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 2.18 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 6

Repeated Sampling Experiment
Random Effects Model
($R = 50,\ N = 400,\ T = 10,\ E = .25$)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$\widehat{Std(\beta)}$	$RMSE$	t-Stat
No Missing Choices						
β_0	-.1000	-.1012	-.1033	.0369	.0369	-.23
β_1	1.0000	1.0364	1.0151	.1517	.1560	1.70
ρ	1.0000	1.0307	1.0388	.0824	.0880	2.64
α	.5000	.4938	.4922	.0818	.0821	-.53
σ_μ	.5000	.5164	.5130	.0933	.0947	1.24
E	.2500	.2438	.2458	.0296	.0296	-.40
10% Missing Choices						
β_0	-.1000	-.1041	-.1029	.0342	.0345	-.85
β_1	1.0000	1.0199	1.0127	.0844	.0867	1.67
ρ	1.0000	1.0347	1.0282	.0682	.0765	3.60
α	.5000	.4956	.4878	.0547	.0549	-.56
σ_μ	.5000	.5062	.5159	.0511	.0514	.86
E	.2500	.2480	.2484	.0294	.0294	-.49
20% Missing Choices						
β_0	-.1000	-.1032	-.1035	.0388	.0389	-.59
β_1	1.0000	1.0219	1.0135	.1126	.1147	1.38
ρ	1.0000	1.0423	1.0336	.0703	.0820	4.25
α	.5000	.5031	.5006	.0786	.0787	.28
σ_μ	.5000	.5048	.5081	.0725	.0726	.47
E	.2500	.2483	.2467	.0338	.0339	-.35

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $\widehat{Std(\beta)}$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t statistics are calculated as $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 1.21 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 7

Repeated Sampling Experiment
Random Effects Model
($R = 50$, $N = 500$, $T = 10$, $E = .25$)

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
20% Missing Choices						
β_0	-.1000	-.1062	-.1097	.0290	.0296	-1.51
β_1	1.0000	1.0213	1.0309	.0932	.0956	1.62
ρ	1.0000	1.0329	1.0369	.0572	.0660	4.07
α	.5000	.5032	.4967	.0474	.0475	.48
σ_μ	.5000	.5163	.5222	.0578	.0601	2.00
E	.2500	.2430	.2361	.0320	.0328	-1.55
20% Missing Choices and X's						
β_0	-.1000	-.1076	-.1054	.0285	.0295	-1.89
β_1	1.0000	1.0266	1.0319	.0699	.0747	2.69
ρ	1.0000	1.0205	1.0132	.0597	.0631	2.43
α	.5000	.4964	.4929	.0438	.0440	-.59
ϕ_1	.2500	.2739	.2780	.0663	.0704	2.55
σ_μ	.5000	.5156	.5159	.0368	.0400	3.00
E	.2500	.2454	.2392	.0320	.0324	-1.01
20% Missing Choices and X's (Higher Variances)						
β_0	-.1000	-.1074	-.1061	.0570	.0575	-.92
β_1	1.0000	.9921	1.0039	.0925	.0928	-.61
ρ	1.0000	1.0231	1.0135	.1295	.1316	1.26
α	.5000	.5209	.5072	.1026	.1047	1.44
ϕ_1	.2500	.2937	.2895	.2306	.2347	1.34
σ_μ	1.0000	.9896	1.0063	.1249	.1254	-.59
E	.2500	.2535	.2570	.0316	.0318	.77

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 1.78 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 8

Repeated Sampling Experiment
Random Effects Model
($R = 50,\ N = 500, T = 20$)

Unobserved Initial Conditions
20% Missing Choices and X's

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
$E = .75$						
β_0	-.1000	-.0986	-.1044	.0303	.0303	.32
β_1	1.0000	1.0280	1.0286	.0599	.0661	3.31
ρ	1.0000	1.0260	1.0166	.0441	.0512	4.17
α	.5000	.4855	.4878	.0390	.0416	-2.63
ϕ_1	.2500	.2808	.2797	.0570	.0648	3.82
σ_μ	.5000	.5265	.5278	.0338	.0430	5.55
E	.7500	.7455	.7476	.0160	.0166	-1.99
$E = .50$						
β_0	-.1000	-.0982	-.1015	.0303	.0304	.41
β_1	1.0000	1.0223	1.0288	.0687	.0722	2.30
ρ	1.0000	1.0082	1.0058	.0411	.0420	1.41
α	.5000	.4993	.5019	.0271	.0271	-.17
ϕ_1	.2500	.2740	.2767	.0550	.0600	3.09
σ_μ	.5000	.5146	.5096	.0313	.0346	3.30
E	.5000	.5025	.5018	.0228	.0230	.78
$E = .25$						
β_0	-.1000	-.1018	-.1056	.0339	.0229	-.39
β_1	1.0000	1.0252	1.0251	.0658	.0704	2.71
ρ	1.0000	1.0160	1.0117	.0534	.0557	2.12
α	.5000	.5007	.4976	.0263	.0263	.18
ϕ_1	.2500	.2631	.2731	.0825	.0835	1.13
σ_μ	.5000	.5068	.5089	.0324	.0331	1.49
E	.2500	.2512	.2520	.0278	.0278	.31

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 5.22 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 9

Repeated Sampling Experiment
$AR(1)$ Error Model
$(R = 50,\ N = 400,\ T = 10,\ E = .75)$

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
No Missing Choices						
β_0	-.1000	-.0992	-.0974	.0173	.0173	.33
β_1	1.0000	1.0226	1.0196	.0555	.0599	2.88
ρ	1.0000	1.0248	1.0240	.0377	.0451	4.65
α	.5000	.5009	.5052	.0347	.0347	.19
ϕ_2	.5000	.5125	.5203	.0527	.0542	1.68
E	.7500	.7461	.7455	.0214	.0217	-1.31
10% Missing Choices						
β_0	-.1000	-.0987	-.0959	.0170	.0171	.52
β_1	1.0000	1.0197	1.0201	.0557	.0591	2.50
ρ	1.0000	1.0285	1.0276	.0408	.0498	4.94
α	.5000	.5023	.5064	.0365	.0366	.45
ϕ_2	.5000	.5140	.5080	.0581	.0597	1.70
E	.7500	.7442	.7430	.0226	.0233	-1.82
20% Missing Choices						
β_0	-.1000	-.0994	-.0979	.0169	.0169	.26
β_1	1.0000	1.0225	1.0172	.0646	.0684	2.47
ρ	1.0000	1.0280	1.0308	.0440	.0522	4.49
α	.5000	.5026	.5014	.0371	.0372	.49
ϕ_2	.5000	.5086	.5006	.0564	.0570	1.08
E	.7500	.7457	.7421	.0244	.0248	-1.25

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 1.91 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 10

Repeated Sampling Experiment
$AR(1)$ Error Model
$(R = 50,\ N = 500,\ T = 10,\ E = .75)$

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$\widehat{Std(\beta)}$	$RMSE$	t-Stat
20 % Missing Choices						
β_0	-.1000	-.1015	-.1024	.0182	.0182	-.58
β_1	1.0000	1.0212	1.0209	.0606	.0642	2.47
ρ	1.0000	1.0335	1.0383	.0451	.0562	5.26
α	.5000	.5057	.5059	.0301	.0306	1.35
ϕ_2	.5000	.5104	.5063	.0563	.0572	1.30
E	.7500	.7447	.7454	.0183	.0191	-2.07
20% Missing Choices and X's						
β_0	-.1000	-.1034	-.1041	.0166	.0169	-1.47
β_1	1.0000	1.0206	1.0269	.0618	.0652	2.36
ρ	1.0000	1.0220	1.0191	.0378	.0437	4.12
α	.5000	.5046	.5072	.0293	.0296	1.12
ϕ_1	.2500	.2906	.3003	.0690	.0801	4.17
ϕ_2	.5000	.5027	.4971	.0555	.0556	.35
E	.7500	.7492	.7493	.0204	.0204	-.29
20% Missing Choices and X's (Higher Variances)						
β_0	-.1000	-.1023	-.1005	.0222	.0223	-.73
β_1	1.0000	1.0307	1.0288	.0548	.0628	3.96
ρ	1.0000	1.0287	1.0240	.0299	.0414	6.81
α	.5000	.5009	.5016	.0230	.0230	.28
ϕ_1	.2500	.2630	.2668	.0517	.0533	1.78
ϕ_2	.5000	.5299	.5272	.0485	.0570	4.36
E	.7500	.7354	.7368	.0186	.0236	-5.56

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $\widehat{Std(\beta)}$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t statistics are calculated as $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 2.50 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 11

Repeated Sampling Experiment
$AR(1)$ Error Model
$(R = 50,\ N = 400,\ T = 10,\ E = .50)$

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
No Missing Choices						
β_0	-.1000	-.0977	-.0989	.0164	.0166	.97
β_1	1.0000	1.0233	1.0278	.0633	.0675	2.60
ρ	1.0000	1.0331	1.0293	.0419	.0534	5.57
α	.5000	.0509	.5067	.0224	.0242	2.84
ϕ_2	.5000	.5204	.5085	.0682	.0712	2.11
E	.5000	.4904	.4898	.0292	.0308	-2.32
10% Missing Choices						
β_0	-.1000	-.0987	-.1019	.0178	.0178	.53
β_1	1.0000	1.0182	1.0257	.0675	.0699	1.90
ρ	1.0000	1.0295	1.0264	.0477	.0561	4.37
α	.5000	.5048	.5088	.0281	.0285	1.22
ϕ_2	.5000	.5156	.5136	.0636	.0655	1.74
E	.5000	.4899	.4881	.0326	.0342	-2.19
20% Missing Choices						
β_0	-.1000	-.0993	-.0997	.0169	.0169	.31
β_1	1.0000	1.0238	1.0302	.0667	.0708	2.52
ρ	1.0000	1.0346	1.0337	.0457	.0573	5.35
α	.5000	.5075	.5134	.0299	.0308	1.77
ϕ_2	.5000	.5116	.4967	.0604	.0615	1.35
E	.5000	.4870	.4901	.0393	.0414	-2.34

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 1.82 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 12

Repeated Sampling Experiment
$AR(1)$ Error Model
$(R = 50,\ N = 500,\ T = 10,\ E = .50)$

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
20% Missing Choices						
β_0	-.1000	-.1001	-.1009	.0150	.0150	-.03
β_1	1.0000	1.0185	1.0215	.0774	.0796	1.69
ρ	1.0000	1.0298	1.0233	.0423	.0518	4.98
α	.5000	.5122	.5086	.0302	.0326	2.85
ϕ_2	.5000	.5034	.5040	.0607	.0607	.40
E	.5000	.4923	.4956	.0294	.0304	-1.85
20% Missing Choices and X's						
β_0	-.1000	-.1021	-.1001	.0151	.0152	-.99
β_1	1.0000	1.0258	1.0217	.0740	.0784	2.46
ρ	1.0000	1.0171	1.0127	.0397	.0433	3.07
α	.5000	.5079	.5097	.0318	.0328	1.76
ϕ_1	.2500	.2927	.2878	.1007	.1094	3.00
ϕ_2	.5000	.4925	.4960	.0627	.0631	-.85
E	.5000	.5018	.5017	.0261	.0261	.50
20% Missing Choices and X's (Higher Variances)						
β_0	-.1000	-.1006	-.1009	.0214	.0214	-.20
β_1	1.0000	1.0333	1.0411	.0763	.0763	3.43
ρ	1.0000	1.0159	1.0144	.0339	.0339	3.76
α	.5000	.5039	.5038	.0199	.0199	1.43
ϕ_1	.2500	.2719	.2771	.0540	.0540	3.14
ϕ_2	.5000	.5091	.5104	.0485	.0485	1.36
E	.5000	.4951	.4988	.0363	.0363	-.97

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 1.92 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 13

Repeated Sampling Experiment
$AR(1)$ Error Model
$(R = 50,\ N = 400,\ T = 10,\ E = .25)$

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
No Missing Choices						
β_0	-.1000	-.0984	-.0984	.0164	.0165	.70
β_1	1.0000	1.0145	1.0160	.0621	.0638	1.66
ρ	1.0000	1.0209	1.0179	.0411	.0461	3.59
α	.5000	.5054	.5069	.0262	.0267	1.47
ϕ_2	.5000	.5257	.5153	.0719	.0754	2.23
E	.2500	.2495	.2543	.0364	.0364	-.09
10% Missing Choices						
β_0	-.1000	-.0976	-.0956	.0173	.0174	1.00
β_1	1.0000	1.0189	1.0121	.0843	.0864	1.59
ρ	1.0000	1.0192	1.0177	.0482	.0519	2.82
α	.5000	.5016	.5043	.0334	.0334	.33
ϕ_2	.5000	.5337	.5281	.0940	.0998	2.54
E	.2500	.2467	.2475	.0417	.0418	-.56
20% Missing Choices						
β_0	-.1000	-.0983	-.0956	.0193	.0193	.63
β_1	1.0000	1.0225	1.0218	.0879	.0907	1.81
ρ	1.0000	1.0242	1.0215	.0642	.0686	2.67
α	.5000	.5041	.5078	.0286	.0288	1.01
ϕ_2	.5000	.5188	.5164	.0695	.0720	1.92
E	.2500	.2354	.2412	.0607	.0624	-1.70

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 1.59 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 14

Repeated Sampling Experiment
$AR(1)$ Error Model
$(R = 50,\ N = 500,\ T = 10,\ E = .25)$

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$Std(\widehat{\beta})$	$RMSE$	t-Stat
20% Missing Choices						
β_0	-.1000	-.0973	-.0960	.0159	.0162	1.22
β_1	1.0000	1.0064	1.0101	.0722	.0724	.63
ρ	1.0000	1.0258	1.0211	.0346	.0431	5.27
α	.5000	.5112	.5116	.0196	.0226	4.06
ϕ_2	.5000	.5094	.5213	.0716	.0722	.93
E	.2500	.2452	.2435	.0477	.0479	-.72
20% Missing Choices and X's						
β_0	-.1000	-.0992	-.0994	.0171	.0171	.33
β_1	1.0000	1.0155	1.0232	.0718	.0734	1.53
ρ	1.0000	1.0046	1.0060	.0449	.0451	.73
α	.5000	.5048	.5060	.0315	.0318	1.08
ϕ_1	.2500	.3227	.2937	.2476	.2581	2.08
ϕ_2	.5000	.4991	.5119	.1108	.1108	-.06
E	.2500	.2644	.2630	.0471	.0492	2.17
20% Missing Choices and X's (Higher Variances)						
β_0	-.1000	-.0952	-.0949	.0225	.0230	1.50
β_1	1.0000	1.0094	1.0065	.0826	.0831	.80
ρ	1.0000	1.0067	1.0050	.0302	.0309	1.58
α	.5000	.5118	.5097	.0224	.0253	3.74
ϕ_1	.2500	.2863	.2846	.1041	.1102	2.47
ϕ_2	.5000	.4917	.4969	.0681	.0686	-.86
E	.2500	.2595	.2571	.0540	.0548	1.24

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $Std(\widehat{\beta})$ 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 $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 1.58 hours on a Pentium III 933 MHz with 500 megabytes of memory.

Table 15

Repeated Sampling Experiment
$AR(1)$ Error Model
$(R = 50,\ N = 500, T = 20)$

Unobserved Initial Conditions
20% Missing Choices and X's

Parameter	True Value	Mean $\widehat{\beta}$	Median $\widehat{\beta}$	$\widehat{Std(\beta)}$	$RMSE$	t-Stat
$E = .75$						
β_0	-.1000	-.0920	-.0987	.0319	.0329	1.76
β_1	1.0000	1.0182	1.0128	.1051	.11066	1.22
ρ	1.0000	1.0152	1.0093	.0319	.0354	3.37
α	.5000	.5033	.4997	.0185	.0187	1.25
ϕ_1	.2500	.2816	.2768	.0719	.0786	3.10
ϕ_2	.5000	.5113	.5109	.0623	.0633	1.28
E	.7500	.7374	.7392	.0451	.0468	-1.97
$E = .50$						
β_0	-.1000	-.0881	-.0997	.0751	.0761	1.12
β_1	1.0000	1.0175	1.0144	.1146	.1159	1.08
ρ	1.0000	.9955	1.0027	.0675	.0677	-.47
α	.5000	.5034	.5057	.0144	.0148	1.69
ϕ_1	.2500	.2769	.2821	.1010	.1045	1.88
ϕ_2	.5000	.4779	.4970	.0841	.0869	-1.86
E	.5000	.5034	.5115	.0630	.0631	.38
$E = .25$						
β_0	-.1000	-.0932	-.0966	.0344	.0350	1.40
β_1	1.0000	.9801	.9900	.1301	.1316	-1.08
ρ	1.0000	1.0055	1.0131	.0394	.0398	.99
α	.5000	.5042	.5016	.0175	.0180	1.71
ϕ_1	.2500	.2932	.2796	.1016	.1104	3.00
ϕ_2	.5000	.4982	.5090	.0935	.0935	-.14
E	.2500	.2520	.2476	.0935	.0936	.15

Note: R is the number of replications. N is the number of individuals. T is the number of time periods. Mean $\widehat{\beta}$ and Median $\widehat{\beta}$ refer to the mean and the median, respectively, of the estimated parameters over all R data sets. $\widehat{Std(\beta)}$ and $RMSE$ refer to the sample standard deviation and the root mean square error, respectively, of the estimated parameters. The t statistics are calculated as $\sqrt{R}\left(\frac{\text{Mean}\widehat{\beta}-\beta}{\text{Std}(\widehat{\beta})}\right)$. The average CPU hours for one replication of the estimations in this table is 6.49 hours on a Pentium III 933 MHz with 500 megabytes of memory.