Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence using Alternative Approaches

Kenneth Y. Chay
Department of Economics
UC-Berkeley

and

Dean Hyslop
Department of Economics
UCLA

October 1998

We thank Colin Cameron, David Card, Paul Devereux, Bo Honoré, Hilary Hoynes, Guido Imbens, Aviv Nevo, and James Powell for helpful suggestions and participants of the UC-Berkeley Labor Lunch and the 1998 Camp Econometrics. Bo Honoré generously provided the programs for one of the estimators used in this paper.

Identification and Estimation of Dynamic Binary Response Panel Data Models: Empirical Evidence using Alternative Approaches

ABSTRACT

This study compares empirical estimates from different specifications of the dynamic binary response panel data model that allows for both unobserved heterogeneity and first-order state dependence. We focus on the role of the unobservable initial conditions and individual effects in consistent estimation of structural state dependence in dynamic discrete processes. First, we examine alternative "random effects" approaches in which the conditional distributions of both the unobserved heterogeneity and the initial conditions are specified. Second, we consider a "fixed effects" conditional logit approach that does not require specification of either conditional distribution. Finally, we use the dynamic linear probability regression model to gauge the validity of the identifying assumptions underlying the nonlinear models.

The various models are used to estimate both female labor force participation and welfare participation equations using panel data from the Survey of Income and Program Participation (SIPP) and the Panel Study of Income Dynamics (PSID). As expected, assuming that the initial conditions are exogenous generates inflated estimates of the degree of state dependence. "Reasonable" assumptions on the initial conditions lead to reduced estimates of the state dependence that are relatively stable across specifications. The estimated effects of the other covariates are much more sensitive to the initial conditions restrictions. The evidence suggests that methods that inappropriately account for the effects of feedback result in understated estimates of the short-run and long-run impact of family structure changes on female welfare and labor force participation. We find that about fifty percent of the overall persistence in welfare participation is attributable to structural state dependence, while this figure is less than forty percent for female labor force participation. In addition, the less restrictive random effects estimators appear to be robust and precise, and linear probability approaches to the dynamic model, which are easier to implement, provide an attractive alternative.

Kenneth Chay
Department of Economics
UC-Berkeley
549 Evans Hall
Berkeley, CA 94720-3880
kenchay@econ.berkeley.edu

Dean Hyslop
Department of Economics
UCLA
405 Hilgard Ave
Los Angeles, CA 90095-1477
dhyslop@paua.sscnet.ucla.edu

Introduction

Understanding the dynamic processes underlying discrete economic phenomena is a topic of considerable interest to a wide range of researchers. There appears to be systematic commonality in the observed dynamics of several of these discrete processes, including welfare, labor force, and union participation; consumer purchase decisions; firm entry and exit decisions, import and export decisions, and decisions concerning capital equipment adoption and replacement; and the incidence of external debt crises in developing countries. In particular, all of these phenomena exhibit substantial serial persistence over time. In this study, we provide an empirical evaluation of different approaches to the problem of separately identifying and estimating the factors underlying this persistence.

In addition to the effects of observed covariates, there are two potential explanations for this serial dependence in discrete outcomes that have been emphasized in the literature (see Heckman 1981 a and c). On one hand, persistence may be the result of "true" or "structural" state dependence in which current participation directly affects the behavioral responses of individuals and, therefore, an individual's propensity to participate in the future. Operationally, in a dynamic discrete panel data model, the lagged outcome would be included in the model as a determinant of the current outcome. On the other hand, observed persistence can also result from permanent unobserved heterogeneity across individuals, in that individuals have different underlying propensities to experience an outcome in all periods. In this case, current participation does not structurally affect the future propensity to participate, but rather this source of serial correlation can be viewed as "spurious" state dependence.

In each of the examples mentioned above, distinguishing between these two sources of persistence is extremely important. For example, the debate over the impact of welfare programs hinges crucially on the ability to credibly estimate the degree of true state dependence. First, state dependence is often interpreted as a behavioral parameter that measures the addictive effects of welfare programs, also known as the "welfare trap".[1] If welfare has "narcotic" incentive effects, then changing welfare program parameters,

---

[1] A statement made by President Clinton summarizes the public perception that "welfare traps" are the primary source of serial dependence in welfare participation; "A long time ago I concluded that the current welfare system undermines the basic values of work, responsibility and family, trapping generation after generation in dependency." (*Time*, August 12, 1996) State dependence is variously labeled as "hysteresis" in the international trade context and

such as benefits levels, can reduce the average length of welfare spells. If most participation is due to permanent characteristics, then changing the nature of welfare programs will have little real effect. In addition, accounting for true lag adjustment in participation is necessary for obtaining consistent estimates of the long-run impact of changing welfare benefits and of the short-run and long-run relationship between participation and other variables such as fertility. Finally, since the outcomes examined are discrete, the existence of true state dependence suggests that small shocks to the process underlying the participation decision could have discontinuous, lasting effects.[2]

The ideal analysis of the relationship between current participation and past participation would involve a controlled experiment in which the researcher randomly assigns program participation across individuals and then observes differences in participation sequences in subsequent periods. Random assignment ensures that past participation would be unrelated to unobserved heterogeneity and that any observed serial dependence in participation could be attributed to state dependence. In the absence of this ideal, we use panel data on the participation histories of individuals to estimate econometric models which attempt to control for confounding determinants of serial dependence such as unobserved heterogeneity.

The analysis is complicated by two factors. First, the typical panel data set contains many individuals but only a small number of time periods. If the initial conditions of the process are unobservable, as is usually the case, and there exists both heterogeneity and state dependence, then misspecifying how the initial conditions are propagated through subsequent periods will lead to inconsistent estimates of the model parameters, particularly of the amount of structural state dependence. This "initial conditions" bias is inversely related to the length of the panel. Second, the outcome of interest is a nonlinear transformation of a continuous latent process. The literature on accounting for the role of initial conditions in the linear dynamic panel data model is well-developed. However, controlling for both unobserved heterogeneity and the unobservable initial conditions is more difficult in the nonlinear dynamic

---

the "narcotic" effects of arbitration in the context of labor dispute resolution. Becker, Grossman, and Murphy (1994) interpret state dependence as cigarette addiction.

[2] Examples of studies that examine welfare dynamics include Bane and Ellwood (1983), Plant (1984), Blank (1989), Gritz and MaCurdy (1992), Hoynes and MaCurdy (1994), and Gottschalk and Moffitt (1994). See Moffitt (1992) for a comprehensive review.

model (Heckman 1981b). Surprisingly, there is little empirical evidence on the relative performance of different estimators of the nonlinear model that attempt to purge the initial conditions bias.[3]

This study empirically examines several different approaches to estimating the dynamic binary response panel data model that allows for both unobserved heterogeneity and first-order state dependence. First, we analyze "random effects" models in which the distributions of both the unobserved individual effects and the initial conditions conditional on the observed covariates are completely specified. We estimate random effects models using both the "probit" and "logit" assumptions on the distributions of the errors, and three alternative assumptions on the initial conditions. Second, we estimate the dynamic "fixed effects" conditional logit model, recently developed by Honoré and Kyriazidou (1997), which does not require specification of the conditional distributions of either the unobserved heterogeneity or the initial conditions. Finally, we use the dynamic linear probability regression model to gauge the credibility of the identifying assumptions underlying the nonlinear models.

We evaluate the ability of these approaches to differentiate the competing sources of dependence in two discrete processes of interest to labor economists, female labor force participation and welfare participation. Not surprisingly, we find that the naïve assumption that the initial conditions are exogenous leads to inflated estimates of the degree of state dependence. Less restrictive assumptions on the initial conditions result in reduced estimates of the state dependence that are relatively stable across the various specifications. The estimated effects of the other covariates are much more sensitive to the initial conditions restrictions. The evidence suggests that methods that inappropriately account for the effects of feedback result in understated estimates of the short-run and long-run impact of family structure changes on female welfare and labor force participation. We find that about fifty percent of the overall persistence in welfare participation is attributable to structural state dependence, while this figure is less than forty percent for female labor force participation. In addition, the less restrictive random effects estimators appear to be robust and precise, and linear probability approaches to the dynamic model, which are more flexible and easier to implement, provide an attractive alternative.

---

[3] Interestingly, most empirical studies of welfare dynamics use duration models of welfare spells. Almost all of them ignore the problem of initial conditions and unobserved heterogeneity.

The Dynamic Binary Response Panel Data Model

The econometric literature on nonlinear panel data models, such as the binary response model, that allow solely for unobserved heterogeneity is well-established (see Chamberlain 1980, 1984, and 1992, Hsiao 1986, Manski 1987, and Honoré 1992). However, it is well-known that identification and estimation of nonlinear models that allow for both individual-specific effects and state dependence is much more tenuous, especially for the typical case of a panel data set with many individuals but a small number of time periods. One of the fundamental issues in estimating these types of models is credibly accounting for the effect of the unobserved initial conditions of the dynamic process on the estimated parameters (Heckman 1981b), also known as initial conditions bias.

Several methods for purging initial conditions bias in the linear dynamic regression model have been developed (e.g., Anderson and Hsiao 1981, Holtz-Eakin, Newey, and Rosen 1988, Arellano and Bond 1991, and Arellano 1995). In addition, there has been renewed interest among theoretical econometricians in developing estimators of nonlinear models that allow for both individual effects and state dependence (Honoré 1993, Arellano and Carrasco 1996, Hahn 1997, Honoré and Kyriazidou 1997, and An and Liu 1997). However, little is known about how well these methods do in practice. The purpose of this study is to perform an exploratory analysis of these approaches with labor market data.

The basic dynamic binary response panel data model that allows for unobserved heterogeneity and first-order state dependence has the following form:

(1a) $y_{it} = 1(y_{it}^{*} > 0) = 1(X_{it}'\beta + \gamma y_{it-1} + \alpha_i + u_{it} > 0); \; i = 1,\ldots,N; \; t = 1,\ldots,T,$

(1b) $P(y_{i0} = 1|X_i, \alpha_i) = p_0(X_i, \alpha_i)$

$P(y_{it} = 1|X_i, \alpha_i, y_{io},\ldots, y_{it-1}) = F(X_{it}'\beta + \gamma y_{it-1} + \alpha_i),$

where y_{it} is the participation outcome for individual i in period t, $1(\bullet)$ is an indicator function which is equal to one if the enclosed statement is true and zero otherwise, y_{it}^{*} is the latent process that guides the participation decision, X_{it} is the vector of exogenous determinants of participation, $X_i \equiv (X_{i1},\ldots,X_{iT})$, α_i is an unobserved individual-specific effect, u_{it} is the transitory error which is assumed to be i.i.d. over time with a distribution function $F(\bullet)$, and β and γ are the parameters of interest. In the context of welfare (and female labor force participation), γ is of particular interest since it represents structural state dependence in participation (i.e., the "welfare trap"). α_i, on the other hand, is the source of spurious state dependence attributable to permanent unobserved heterogeneity in earnings potential and tastes for leisure across individuals.

Identification of the parameters of the model described by equations (1a) and (1b) requires several assumptions. First, the model assumes that the observable determinants of participation (X_{it} and y_{it-1}) are "jointly" exogenous conditional on the individual effects. It should be noted, however, that the model does allow for feedback from the lagged discrete outcome into current values of the X's. Second, the functional form of the conditional mean of the latent process, y^*_{it}, must be correctly specified (e.g., a linear regression function with only first-order state dependence). Third, the transitory errors, u_{it}, are assumed to be serially uncorrelated over time and drawn from a stationary error distribution.[4] Finally, in order to close the model, assumptions are made on the functional form of the distributions of the unboservables. For example, assuming that u_{it} has a normal or logistic distribution implies the "probit" and "logit" models for F(•), respectively.

We examine both "fixed effects" and "random effects" approaches to specifying and estimating this model. Honoré and Kyriazidou (1997) have developed a fixed effects conditional logit approach to the dynamic binary response model, which extends Chamberlain's (1985) model to allow for exogenous covariates. In this approach, no assumptions are required on the conditional distributions of the individual effects or the initial conditions, which are allowed to have arbitrary relationships with the explanatory variables. Consequently, both the individual effects and the initial conditions can be "absorbed" with the proper conditioning statement.

In a random effects approach, the distribution of the unobserved individual effects, conditional on the observed covariates, is specified. Since the incidental parameters are not conditioned out in this case, the relationship between the (sample) initial conditions of the dynamic stochastic process and the unobserved heterogeneity must also be specified. We examine three alternative approaches to the initial

---

[4] In our previous empirical applications (Hyslop 1998 and Chay 1995), we found that first-order state dependence fit observed labor force participation sequences for women and men relatively well. In our contexts, we also allowed the errors, u_{it}, to be serially correlated. This study also examines models which allow for serially correlated errors. Consistent with the bulk of the previous econometric and empirical literature, we abstract from this case for now.

conditions “problem” that have commonly been adopted by researchers: 1) assume that the initial conditions (or pre-sample history) of the process are exogenous; 2) assume that the process is in equilibrium at the beginning of the sample period; and 3) use a reduced-form approximation to the initial sample observation which allows for a flexible relationship between the initial conditions, the unobserved heterogeneity, and the observed covariates. We show that models that assume that the initial conditions are exogenous are effectively ignoring serial dependence attributable to unobserved heterogeneity and, therefore, lead to upwardly biased estimates of the structural state dependence.

Fixed Effects Approaches

The initial conditions problem is particularly difficult in nonlinear models such as binary response. Under certain assumptions on the distribution of the transitory errors, one can sometimes avoid making stochastic restrictions on the individual effects by treating them as “fixed effects” and conditioning them out. In such cases, the individual effects are allowed to have an arbitrary correlation with the observed covariates, X_i. We illustrate this for the binary logit and linear models for when the researcher observes the individual for 2, 3, and 4 periods. We show that fixed effects approaches which absorb the incidental parameters but do not account for the role of the initial conditions lead to downwardly biased estimates of the state dependence, particularly in panels of a short and fixed length.

In the absence of state dependence (γ=0), it is well-known that if $T\geq 2$ and the transitory errors are i.i.d. over time with a logistic distribution, then conditioning on the sufficiency class, $\sum_{t=1}^{T} y_{it}$, absorbs the incidental parameters and leads to consistent estimation of β (e.g., Chamberlain 1980). In other words, the number of times an individual is observed participating is a sufficient statistic for the individual effects. It is worth noting that conditioning on the sufficiency class is analogous to the linear regression model based on deviations from individual means.[5] A model with both unobserved heterogeneity and first-order state dependence is not identified when individuals are observed for only two periods.

---

[5] In the normal linear model, the joint maximum likelihood (ML) estimator based on explicitly estimating individual dummies, the conditional ML estimator based on deviations from individual means, and the “marginal” ML estimator based on a correlated random effects model in which the individual effects are projected onto the X_i all coincide. Chamberlain (1985) notes that this provides an interpretation for why the joint ML estimator is consistent

When individuals are observed for three periods, the binary response model that allows for both unobserved heterogeneity and state dependence cannot be consistently identified. This fact is most transparent when considering the linear regression model based on first-differencing the data to eliminate the individual effects:

(2) $y^*_{i2} - y^*_{i1} = (X_{i2} - X_{i1})'\beta + \gamma(y^*_{i1} - y^*_{i0}) + (u_{i2} - u_{i1})$.

Due to the negative correlation between the first-differenced lagged dependent variable and the first-differenced error term, estimates of the state dependence will be severely attenuated. The bias in the fixed effects estimator of state dependence based on deviations from individual means is of O(1/T) and will be large in short panels. For example, if the joint distribution of the errors is stationary, then the estimator based on equation (2) converges to $(\gamma - 1)/2$ asymptotically.

This attenuation bias arises from the relationship between the initial conditions and the incidental parameters. In an attempt to absorb the incidental parameters, the data is first-differenced or deviated from individual means, which effectively absorbs the initial conditions as well. As a result, fixed effects estimators of the type in equation (2) attribute too much of the serial dependence to unobserved heterogeneity and understate the amount of state dependence by ignoring the role of initial conditions. In general, any approach which overstates (understates) the importance of unobserved heterogeneity will understate (overstate) the degree of structural state dependence. This problem is even more intractable in models based on nonlinear transformations of y^*_{it}, such as the binary response model. For example, Heckman (1981b) finds in monte carlo simulations that the attenuation bias in the autoregressive probit that jointly estimates the fixed effects is extremely large even when T=8. The problem of incidental parameters in nonlinear models with short panels exacerbates the initial conditions problem.

It is noteworthy that the linear dynamic regression model is identified in the three period case if the transitory errors are serially uncorrelated. Due to the state dependence, the second lag of the outcome (y^*_{i0}) is related to the first-difference of the lagged outcome ($y^*_{i1} - y^*_{i0}$), unless the process follows a unit

---

in the linear model despite the incidental parameters problem that the number of parameters is increasing with the number of individuals.

root. Therefore, y^*_{i0} can be used as an instrument for $(y^*_{i1} - y^*_{i0})$ in equation (2) since it is orthogonal to $(u_{i2}-u_{i1})$. This instrumental variables approach corrects the bias induced from purging the fixed effects.

If we observe individuals for at least four periods, then it is possible to identify the logit binary response model that allows for both sources of dependence. First, consider the model with no exogenous regressors ($\beta=0$). Chamberlain (1985) shows that if the transitory errors have an i.i.d. logistic distribution, then there exists a conditioning statement that will absorb both the individual effects and initial conditions. In particular, the model is:

(3a) $P(y_{i0} = 1|\alpha_i) = p_0(\alpha_i)$

$$P(y_{it} = 1|\alpha_i, y_{i0},\ldots, y_{it-1}) = \frac{\exp(\gamma y_{it-1} + \alpha_i)}{1+\exp(\gamma y_{it-1} + \alpha_i)} \qquad t=1,\ldots,3,$$

and it follows that:

(3b) $$P(y_{i0} = d_{i0}, y_{i1} = 1, y_{i2} = 0, y_{i3} = d_{i3}|\alpha_i, y_{i1} + y_{i2} = 1) = \frac{\exp[\gamma(d_{i0} - d_{i3})]}{1+\exp[\gamma(d_{i0} - d_{i3})]},$$

where $d_{i0}, d_{i3} \in \{0,1\}$, and the relationship between the initial observation and the unobserved heterogeneity is left unspecified. Here, consistent identification of γ is based on the fact that the joint probability in equation (3b) is independent of α_i. The fixed effects are eliminated by "fixing" the first and final period binary outcomes and conditioning on individuals who had changes in participation in the two intervening periods. Differences in the observed probabilities of the binary sequences ($y_{i0} = d_{i0}$, $y_{i1} = 1$, $y_{i2} = 0$, $y_{i3} = d_{i3}$) and ($y_{i0} = d_{i0}$, $y_{i1} = 0$, $y_{i2} = 1$, $y_{i3} = d_{i3}$) lead to inferences on the amount of path dependence in the process.

Honoré and Kyriazidou (1997) consider identification of the dynamic logit in the presence of additional explanatory variables that are strictly exogenous. They develop a conditional logit fixed effects estimator that is consistent and asymptotically normal provided that individuals are observed for at least four periods and the errors are i.i.d. logsitic. In this case, equation (1b) becomes:

(4) $P(y_{i0} = 1|X_i, \alpha_i) = p_0(X_i, \alpha_i)$

$$P(y_{it} = 1|X_i, \alpha_i, y_{i0},\ldots, y_{it-1}) = \frac{\exp(X_{it}\beta + \gamma y_{it-1} + \alpha_i)}{1+\exp(X_{it}\beta + \gamma y_{it-1} + \alpha_i)} \qquad t=1,\ldots,3,$$

where again the model is left unspecified in the initial period. Focusing on the same sequences as above, define events A and B as follows:

$A \equiv \{y_{i0} = d_{i0}, y_{i1} = 0, y_{i2} = 1, y_{i3} = d_{i3}\}$ and $B \equiv \{y_{i0} = d_{i0}, y_{i1} = 1, y_{i2} = 0, y_{i3} = d_{i3}\}$.

Here, conditioning only on the sufficiency class, $\sum_{t=1}^{T} y_{it}$, will not eliminate the individual effects.[6]

However, Honoré and Kyriazidou show that the individual effects can be eliminated and the parameters are identifiable if $X_{i2} = X_{i3}$. In this case, the conditional probabilities

$$\text{(5a)} \quad P(A|X_i, \alpha_i, A \cup B, X_{i2} = X_{i3}) = \frac{1}{1+\exp((X_{i1}-X_{i2})\beta+\gamma(d_{io}-d_{i3}))} \quad \text{and}$$

$$\text{(5b)} \quad P(B|X_i, \alpha_i, A \cup B, X_{i2} = X_{i3}) = \frac{\exp((X_{i1}-X_{i2})\beta+\gamma(d_{io}-d_{i3}))}{1+\exp((X_{i1}-X_{i2})\beta+\gamma(d_{io}-d_{i3}))}$$

do not depend on the individual effects. Identification of γ comes from differences in the observed frequencies of sequences that are identical except for the "path changes" that occur in the "intervening" periods among individuals whose exogenous characteristics are stationary in the final two periods. β is identified from changes in the exogenous variables in the two middle periods for these same individuals.

Based on this insight, one can derive an estimator of the logit model which puts increasing weight on those observations for which X_{i2} is "close" to X_{i3} (and asymptotically uses only observations where $X_{i2} = X_{i3}$). Honoré and Kyriazidou propose the following estimator for the four period case:

$$\text{(6)} \quad (\hat{\beta}, \hat{\gamma}) = \underset{\beta,\gamma}{\operatorname{argmax}} \sum_{i=1}^{N} 1(y_{i1}+y_{i2}=1) K\left(\frac{X_{i2}-X_{i3}}{\sigma_N}\right) \ln\left(\frac{\exp[(X_{i1}-X_{i2})'\beta+\gamma(y_{i0}-y_{i3})]^{y_{i1}}}{1+\exp[(X_{i1}-X_{i2})'\beta+\gamma(y_{i0}-y_{i3})]}\right),$$

where K(•) is a kernel weighting function which gives more weight to observations with "differences" that are smaller in magnitude, and σ_N is a bandwidth that goes to zero as N increases. They note that if $P(X_{i2}=X_{i3})>0$ (e.g., discrete covariates or controlled experiments) and if $(X_{i1}-X_{i2})$ has variation conditional on $(X_{i2}=X_{i3})$, then the estimator in equation (6) will have "root-N" convergence. However, if the regressors are continuous and not in the researcher's control, then the resulting estimator, while consistent and asymptotically normal, will have a rate of convergence slower than $N^{-1/2}$. The convergence rate will be slower as the number of continuous exogenous variables increases.[7]

Thinking about the estimator in equation (6) as a "pairwise differencing" or "matching" estimator provides intuition on how the dynamic conditional logit accounts for unobserved heterogeneity

---

[6] Card and Sullivan (1988) show that the minimal sufficient statistic for the individual effects in equation (4) is the entire data vector of participation sequences for individual i.

[7] Hahn (1997) derives results that suggest that it is not possible to consistently estimate the parameters of the model at a $N^{-1/2}$ rate of convergence.

and the initial conditions problem when estimating the amount of true state dependence, γ. The estimator only uses the information provided by individuals whose paths are identical in every way except for the "changes" that occur to them in the "intervening" periods. The state dependence is estimated by comparing individuals within a sufficiency class with identical starting and ending points and covariates that did not change in the final two periods. For example, among individuals with $X_{i2}=X_{i3}$, the existence of structural state dependence implies that the sequence (0,0,1,1) is more likely to occur than (0,1,0,1), and the sequence (1,1,0,0) should be more prevalent than (1,0,1,0).

The conditioning statement in equations (5a) and (5b) effectively matches individuals on observables to potentially control for the unobservable confounding factors of individual effects and initial conditions. In effect, one can be more "judicious" about which comparisons will credibly identify the amount of structural state dependence due to the additional information in the data afforded by having at least four periods. Individuals can be matched based on more observables. In the absence of the ideal of random assignment, the estimator relies on the identifying assumption that lagged participation is conditonally orthogonal to unobservable determinants of current participation.[8]

If the conditioning statement is true, then the conditional logit estimator will provide unbiased estimates of the contribution of structural state dependence. To the extent that it is not true, the estimator will not fully account for the role of initial conditions and will underestimate the amount of state dependence in finite samples for the reasons discussed above. Requiring stationary observables in the final two periods controls for the feedback effects of the lagged outcome on current values of the covariates. For example, women who experience changes in family structure (e.g., additional children, marriage) are also more likely to change their labor force or welfare participation behavior. Using these individuals to estimate the model will tend to lead to an understatement of the amount of structural state dependence. In addition, the importance of initial conditions, and therefore the size of the

---

[8] See Rubin (1979) and Robinson (1988) for examples of matching estimators from the viewpoint of statisticians and econometricians, respectively. Using the language of the causal inference literature, matching relies on the availability of "pre-exposure" variables that lead to conditional independence of the treatment assignment. This assumption is often referred to as "strong ignorability" conditional on the covariates. Card and Sullivan (1988) use this idea to condition out heterogeneity and state dependence as confounding factors in evaluating the impact of training programs on movements in and out of employment.

understatement, will be increasing in the amount of true state dependence. In monte carlos, Honoré and Kyriazidou find evidence of a positive relationship between the attenuation bias in their estimator for the state dependence and the true underlying state dependence.[9]

Examining equation (6), the dynamic conditional logit estimator is similar to an estimator of the linear "changes" regression model, which uses only certain changes and weights them to account for the initial conditions. By comparison, the instrumental variables (IV) approach to the linear dynamic model uses all "changes" in the lagged dependent variable that are attributable to additional lags of the variables to identify the amount of state dependence. If it is assumed that only y^*_{it-1} is pre-determined in the regression function, then the lagged history of the outcome ($y^*_{it-2}, y^*_{it-3}, \ldots$) can be used as instruments for changes in y^*_{it-1}. If it is also assumed that the covariates are either strictly exogenous or pre-determined, then all future and past values of the covariates (X_i) or their lagged history (X_{it-1}, $X_{it-2}, \ldots$) can be used as additional instruments for y^*_{it-1}, respectively.[10] We examine this approach by applying IV estimation to the linear probability model. In the less restrictive case which assumes that only y^*_{it-1} is pre-determined, the IV estimator only uses changes in y^*_{it-1} attributable to "reversion to the mean" (y^*_{it-2}) for identification.

Random Effects Approaches

Although fixed effects approaches do not require parametric assumptions about the conditional distributions of the individual effects and initial conditions, equation (6) illustrates that a lot of information is being absorbed in order to "non-parametrically" condition out these unobservables. Consequently, fixed effects estimators may be sub-optimal since they "throw away" comparisons between individuals that may be informative about the truth. However, if the statistical relationship between the individual effects, initial conditions, and explanatory variables is correctly specified, then

---

[9] Another potential caveat with the Honoré and Kyriazidou estimator is that it requires errors that are i.i.d. over time. Although the other conditional logit estimators require similar assumptions, the stationarity assumption is restrictive. For example, the estimator cannot accommodate unrestricted time-effects. Below, we examine both the independence and stationarity assumptions in more detail using the probit and linear probability models.

[10] See Arellano and Bover (1995) for a lucid description of efficient instrumental variables estimation of the linear dynamic panel data model.

one can use "random effects" estimators to increase efficiency and still consistently identify the parameters in equation (1a). Also, in the dynamic binary response model, the fixed effects approach can only be used if the errors have an i.i.d. logistic distribution. Random effects estimators, on the other hand, can be used under a variety of specifications if the stochastic restrictions are appropriate.

We examine both the "logit" (u_{it} ~ i.i.d. logistic) and "probit" (u_{it} ~ normal) models of equation (1). For the logit model, we consider two types of mixture models for the unobserved heterogeneity. In one, the heterogeneity distribution is assumed to be $\alpha_i \sim$ i.i.d. $N(0, \sigma^2_\alpha)$. In the other, we assume that the random effects have a common discrete distribution with a finite number of mass points. Since the positions and probabilities of the mass points are parameters of the likelihood function, this method allows for a relatively unrestricted specification of the individual effects (see Heckman and Singer 1984). Both specifications assume that the individual effects are independent of the observed covariates.[11] The probit model is used to examine the sensitivity of the results to the assumption that the errors are i.i.d. logistic. Here, we use monte-carlo integration methods to estimate a model that allows for the normal heterogeneity distribution and first-order serial correlation in the transitory errors: $u_{it} = \rho u_{it-1} + \nu_{it}$, $\nu_{it} \sim$ i.i.d. $N(0, \sigma^2_\nu)$.

A random effects approach also requires the specification of the relation between the initial sample conditions and the unobserved heterogeneity and observables. We consider three different specifications of the initial conditions that vary in how restrictive they are. The most naive approach taken assumes that the initial conditions, y_{i0} (or the pre-sample history), are exogenous and can be ignored (e.g., Heckman 1978, 1981a and 1981c). This stochastic restriction is valid only if the disturbances that generate the process are serially independent, which is not the case in the presence of unobserved heterogeneity, or if the process begins at the start of the observed sample period (i.e., the first period of observation for every individual is the true initial period). Here,

$$(7) \qquad P(y_{i0}|X_i, \alpha_i) = P(y_{i0}),$$

---

[11] The fixed effects conditional logit estimator of Honoré and Kyrazidou allows for an arbitrary relation between the individual effects and the observables, as do the linear probability approaches examined below. We will eventually examine random effects models that allow for correlation between the observables and individual effects.

and the initial conditions and individual effects are assumed to be independent. Consequently, this conditional probability can be ignored when estimating the structural model. However, if unobserved heterogeneity is a determining factor in the initial conditions, then this approach will overstate the amount of state dependence in the process.

A second restriction on the initial conditions that has been used previously assumes that the dynamic stochastic process which generates the observed participation sequences is in equilibrium at the beginning of the sample period (Card and Sullivan 1988). Define $p_{i0}^{R} \equiv P(y_{i0} = 1|y_{i(-1)}=1, X_{i0}, \alpha_i)$ and $p_{i0}^{A} \equiv P(y_{i0} = 1|y_{i(-1)}=0, X_{i0}, \alpha_i)$ to be the "retention" and "accession" probabilities conditional on the initial period observables, respectively. Then the initial period steady-state participation probability is:

$$(8) \qquad p_{i0} \equiv P(y_{i0} = 1|X_i, \alpha_i) = p_{i0}\, p_{i0}^{R} + (1-p_{i0})\, p_{i0}^{A} = \frac{p_{i0}^{A}}{1+p_{i0}^{A}-p_{i0}^{R}},$$

which defines the probability of participation in the first observed period. We refer to this as the "equilibrium" initial conditions restriction. In the logit case:

$$p_{i0}^{R} = \frac{\exp(X_{i0}\beta+\gamma+\alpha_i)}{1+\exp(X_{i0}\beta+\gamma+\alpha_i)} \text{ and } p_{i0}^{A} = \frac{\exp(X_{i0}\beta+\alpha_i)}{1+\exp(X_{i0}\beta+\alpha_i)}.$$

This stationarity restriction is unlikely to hold when the observable covariates are time-varying and important determinants of participation.

Following Heckman (1981b), the third and least restrictive approach to the initial conditions problem that we examine uses a "reduced-form" approximation to the initial conditions:

$$(9) \qquad P(y_{i0} = 1|X_i, \alpha_i) = F(X_{i0}'\Pi_1 + \Pi_2\alpha_i).$$

In principle, this approach provides a particularly flexible characterization of the relation between the initial conditions and the individual effects when the "mass point" approach is used to model the heterogeneity distribution.[12] For the logit, the "reduced-form" initial conditions are:

$$P(y_{i0} = 1|X_i, \alpha_i) = \frac{\exp(X_{i0}\Pi_1+\Pi_2\alpha_i)}{1+\exp(X_{i0}\Pi_1+\Pi_2\alpha_i)}.$$

---

[12] Notice that we are using the decomposition of the joint distribution of $(y_{i0}, y_{i1}, \ldots, y_{iT})$ described in the Appendix which specifies $P(y_{i0}|X_i, \alpha_i)$. Equation (9) allows the initial state probabilities to vary across the M heterogeneity mass points $\alpha_1, \ldots, \alpha_M$. Alternatively, we could specify the conditional distribution of α_i given y_{i0}, $f(\alpha_i|y_{i0}, X_i)$, which allows for dependence between y_{i0} and α_i, while leaving the initial conditions of the process unrestricted.

Before proceeding, we note some potential caveats with our analysis. First, as the above discussion illustrates, there are an "infinite" number of ways to specify how the individual effects, initial conditions, and explanatory variables are jointly related. The approaches examined are diverse enough to provide a useful benchmark for the sensitivity of the estimated state dependence to different specifications of the model.[13] In a related point, all of our approaches rely on explicit specification of the shape of the distribution of the transitory errors, u_{it}. Honoré and Kyriazidou (1997) also develop an estimator of the model that is consistent if u_{it} is independent over time with a distribution function that is strictly increasing.[14] However, they infer that this semi-parametric estimator has a convergence rate slower than $n^{-1/3}$ and a non-normal asymptotic distribution.

Further, all of our approaches assume strict exogeneity of the explanatory variables in the mixture model. They will be consistent, however, as long as the feedback from the participation outcome into the X's can be summarized by first-order state dependence.[15] Our approaches also assume stationary and serially uncorrelated error terms. Finally, identification of the parameters in equation (1) requires that the conditional expectations function, in particular the lag length of the state dependence, is specified correctly. In the random effects estimators, the conditional expectations of the unobservables must be properly specified as well.[16]

Below, we use the dynamic linear probability regression model to examine the credibility of the identifying assumptions of the nonlinear models for two reasons. First, it is more flexible and easier to

---

[13] An and Liu (1997) use the method of "indirect inference" to allow for a relatively unrestricted conditional distribution for the initial conditions. Specifically, the "distance" between participation histories simulated under a maintained structural model and the actual participation histories is minimized using a "well-fitting" auxiliary model as the guiding metric.

[14] Their semi-parametric estimator adapts Manski's (1987) conditional maximum score estimator using a conditioning argument identical to the one used to identify the dynamic conditional logit.

[15] Since our focus is on estimating structural state dependence, we are less concerned with the exogeneity assumption on the X's. Arellano and Carrasco (1996) develop a random effects estimator for the dynamic binary choice model where the covariates are only assumed to be pre-determined. The linear probability model is used to examine this issue below.

[16] In the linear case, assumptions on the conditional expectations functions relating the individual effects and initial conditions to the covariates, such as linearity, are innocuous since the regression functions can be reinterpreted as linear minimum mean square error predictors. For example, Chamberlain (1980) shows that the correlated random effects estimator of the linear autoregressive panel data model with no covariates is identical to the instrumental variables estimator based on first-differences. In the nonlinear case, on the other hand, specifying the regression functions as being linear in the covariates when they are not leads to inconsistent estimates.

implement than the nonlinear models. Second, the methods for handling initial conditions in the linear model are well established. One potential drawback is that the linear probability model is not useful for forecasting exercises. In our study, however, this is not an issue since our interest is in the identification and estimation of the structural responses of individuals. Moreover, one could feed the estimates from the linear regression model into a dynamic probit or logit model to do forecasting.[17]

Data and Preliminary Analysis

Recently, a growing number of studies have applied the dynamic binary response panel data model to a wide array of topics.[18] However, there is little evidence on the relative performance of the different approaches that these studies have used. The objective of our study is to provide a systematic comparison of the random effects and fixed effects alternatives to the initial conditions problem and document whether the model is credibly and precisely identified when applied to two labor market processes: female labor force and welfare participation. Hopefully, this empirical benchmarking will provide guidance to researchers interested in estimating dynamic discrete-choice panel data models.

Our analysis uses samples from the two panel data sets predominantly used to examine these outcomes: the Survey of Income and Program Participation (SIPP) and the Panel Study of Income Dynamics (PSID). The SIPP extracts are used to examine both married women's labor force participation and female recipiency of Aid to Families with Dependent Children (AFDC) welfare benefits, measured at four-month intervals. The PSID extract pertains only to married women's annual labor force participation. A comparison of the SIPP and PSID labor force participation results provides information on the sensitivity of the estimates to the sampling interval between observations.

---

[17] One final point is noteworthy. Chamberlain (1993) shows that the parameters of a model with feedback (e.g., through state dependence) and multiplicative individual effects are often not identified, even in the linear case. As a result, this study assumes homogeneous effects of the variables of interest on participation probabilities.

[18] For example, the model has been applied to the following topics: evaluation of the impact of training programs (Card and Sullivan 1988, Ham and Lalonde 1996); the effects of rent control on housing quality (Moon and Stotsky 1993); the role of sunk costs in import and export decisions (Roberts and Tybout 1997); and persistence in the incidence of external debt crises in developing countries (Hajivassiliou and McFadden 1997). These studies used different approaches to the model, all of which are examined in this study.

The two SIPP longitudinal samples are derived from the 1990 panel, which contains 8 waves at four-month intervals and covers the previous 32 months. Although each wave collected month-specific information regarding the previous 4 months, we aggregate the monthly data in each wave to the 4-month level in order to mitigate problems such as "seam bias" and measurement error due to misreported month-to-month changes.[19] For the welfare participation analysis, the "at risk" population is defined to be all women who either received AFDC payments during or before the sample period, or whose average total family income during the sample period was less than the sample average family-specific poverty level. For the labor force participation analysis, we restrict the sample to be continuously married women whose husbands were labor force participants during each wave of the panel. In addition, the samples only include women who were 18-65 years old and could be matched across each of the 8 waves. The PSID sample consists of a seven year panel obtained from the 1986 panel, corresponding to waves 13-19, and covers annualized labor force participation over the period 1979-85. It only includes women who were continuously married, aged 18-60 in 1980, and had husbands who were labor force participants in each year.

Table 1 presents the frequencies associated with each of the possible participation sequences for the SIPP welfare participation sample (Panel A), the SIPP labor force participation sample (Panel B), and the PSID labor force participation sample (Panel C). In each panel, the sequences are sorted by “sufficiency classes” – that is, the total number of times an individual is observed to participate during the sample period. There are several patterns worth noting with respect to all three processes. First, each of the participation outcomes exhibits substantial serial persistence over time. The overwhelming majority of individuals either participate in all of the sample periods or never participate, effectively ruling out the possibility that the process underlying the sequences is independent over time. Interestingly, based on this metric, it appears that there is less serial persistence in PSID annualized labor force participation than in SIPP participation, probably due to the higher observation frequency in the SIPP.

Second, it appears that this serial dependence cannot be explained solely by unobserved heterogeneity. In particular, in the conditional logit framework and in the absence of true state dependence,

---

[19] See Marquis and Moore (1990) for a discussion of these issues in the SIPP.

conditioning on the sufficiency class absorbs the individual effects. This implies that if the only source of serial dependence arose from a permanent person-specific component, one would observe different numbers of women across each class but should also see approximately the same number of women in each sequence within a class (conditional on the exogenous regressors). However, when examining all three panels, it is clear that within sufficiency classes most women experience sequences in which participation periods and non-participation periods are "clumped" together. For example, in the class of sequences in which women participate a total of four times in welfare during the sample period (column 2 of Panel A), the most prevalent sequences are the ones in which an individual participates in either the first four consecutive periods or the last four consecutive periods.

Table 1 suggests that a model that allows for both unobserved heterogeneity and state dependence will fit the participation patterns better than models which allow for only one of these sources of persistence.[20] Our study gauges whether the examined estimators can credibly distinguish between these two sources of serial dependence in the data sets typically analyzed. The similarity between the welfare and labor force participation patterns underscores that another benefit of our study is that the findings may be generalizable to other discrete dynamic processes that are of interest to economists.

Next, we consider the relationships between observed individual characteristics and participation outcomes. Table 2 presents summary statistics for each of the three samples (column 1) and for various subsamples selected on the basis of the participation patterns observed in Table 1. The subsamples consist of those who participate in every period (column 2), never participate during the sample frame (column 3), have a single transition into participation (column 4) and out of participation (column 5), and have multiple transitions (column 6). Panel A summarizes the characteristics of the SIPP welfare sample. Women who always receive AFDC benefits are younger, less-educated, more likely to be black, less likely to be married, have more children, have lower earnings and income, and receive greater AFDC payments than other recipients. The converse of the above is true for women who never receive AFDC benefits. In addition, the

---

[20] Interestingly, Blank (1989) and Plant (1984), using different approaches than the ones we suggest, find little evidence of state dependence in program participation when examining data from the Seattle/Denver Income Maintenance Experiments.

two single transition subsamples have fairly similar characteristics, although those leaving welfare have higher earnings and income, on average. In fact, the earnings and income of the "leavers" are comparable to that of the non-recipient sample suggesting that earnings shocks lead to transitions out of welfare. Finally, women in the multiple transitions subsample are more likely to be married than the other sets of recipients suggesting that changes in family structure are another important determinant.

Panels B and C of Table 2 present the characteristics of the SIPP and PSID labor force participation samples. The relative characteristics across the subsamples are similar for the two extracts. Women who always participate are younger, better-educated, have fewer children, and have husbands with lower earnings than those who never participate. In both the SIPP and PSID samples, women who experience a single transition into the labor force are, on average, younger and have more children than women who have a single transition out. Using the additional detail on fertility available in the PSID, it seems that women who move into the labor force have fewer children aged 0-2, but more aged 3-5 and 6-17 than those leaving the labor force. This pattern is consistent with the notion that women leave the labor force to have children and return as the children age beyond infancy. The table suggests that there is a positive income effect of husbands' earnings on wives' non-market time, and a negative fertility effect of younger children on participation propensities.

To examine the relation between participation changes and changes in the covariates, Table 3 presents information on changes in characteristics associated with transitions out of participation (column 1), no change in participation (column 2), and transitions into participation (column 3) for the three extracts.[21] Panel A on SIPP welfare participation shows that transitions out of welfare are associated with a recent marriage, a decline in the number of children, and an increase in family income, while the converse is "strongly" true of transitions onto welfare. On the other hand, the covariates are very stationary during periods in which there are no changes in participation.

---

[21] The estimated standard errors presented in this table have not been corrected to account for the repeated observations in the panel. However, they should be approximately correct since only 10-20 percent of the samples in columns (1) and (3) involve repeat observations on the same individual.

For SIPP labor force participation, Panel B shows that transitions out of the labor force coincide with increases in the number of children and in husband's earnings, while movements into the labor force coincide with a decrease in the number of children. Based on PSID labor force participation in Panel C, it appears that the association between movements into and out of the labor force and children is largely driven by changes in the number of very young children, aged 0-2. In contrast to the SIPP sample, transitions into the labor force in the PSID coincide with a decline in husband's earnings, while there is little systematic evidence that movements out of the labor force correspond with an increase in spouse's earnings. Again, it appears that the observables are relatively stationary during periods with no transitions.

Finally, we examine the characteristics of individuals with participation sequences that may be informative for identifying the amount of structural state dependence in the context of the dynamic conditional logit model. Recall that the Honoré and Kyriazidou approach conditions out the unobserved heterogeneity by comparing 4-period participation sequences which have identical initial and final states but different transitions between the end states. Table 4 presents the sample characteristics for the sequences (0,1,0,1) versus (0,0,1,1) and (1,1,0,1) versus (1,0,1,1) for each of the three processes. The first important point is that for all three outcomes the relative frequency of y_t=1 conditional on (y_{t-3},y_{t-2},y_{t-1})=(0,0,1) is substantially greater than the one conditional on (y_{t-3},y_{t-2},y_{t-1})=(0,1,0); and the relative frequency of y_t=1 conditional on (y_{t-3},y_{t-2},y_{t-1})=(1,0,1) is substantially greater than that conditional on (y_{t-3},y_{t-2},y_{t-1})=(1,1,0). In a model with no covariates, these gaps suggest the strong presence of first-order structural state dependence in each of the processes. It also appears that there is more state dependence in welfare participation than in labor force participation and in the SIPP than in the PSID. Again, the latter observation may be attributable to the smaller sampling intervals in the SIPP.

Second, Honoré and Kyriazidou's estimator also relies on the covariates being approximately constant in the final two periods for identification. To address this issue, Table 4 also presents changes in the observables in the final two periods for each sequence. To varying degrees, the table entries suggest that the lives of women with sequences that may be informative about the structural state dependence are not stationary, especially with respect to number of dependent children and children aged 0-2. Consequently, the Honoré and Kyriazidou conditional logit estimator may exhibit some small sample bias. More generally, transitions into and out of welfare and labor force participation are associated with changes in family structure (marital status and number and ages of children) and in economic status (family income and husband's earnings). Therefore, accounting for any feedback relations is an important issue in trying to consistently estimate the amount of true state dependence and the effects of the other observables.

Empirical Results

Random Effects Models

Our analysis begins with the estimation results from the various random effects specifications of equation (1) for each of the three labor market samples. The alternative random effects models are based on different stochastic restrictions on the distributions of the transitory errors (u_{it}), random effects (α_i), and the initial conditions (y_{i0}), as discussed above. First, we consider models based on the assumption that u_{it} is i.i.d. logistically distributed for the two alternative specifications of α_i (i.i.d. $N(0, \sigma^2_\alpha)$ and a discrete mass point distribution) and the three different specifications of y_{i0} (exogenous, in equilibrium, and approximated using a flexible reduced-form). The Appendix provides the details for the likelihood function based on normal random effects and logistic errors. The normal random effects specification is convenient since it allows for estimation of the model using Gaussian Quadrature procedures (Butler and Moffitt 1982). For the mass point random effects, the number of mass points (k) is selected using a likelihood ratio statistic to gauge the adequacy of (k) versus (k+1) mass points and on the basis of model identification (i.e., convergence). The number of mass points used in our analysis varies between 3 and 4. Different numbers of mass points led to similar results.

Table 5 presents the results from the "logit" specification for SIPP welfare participation (Panel A) and SIPP and PSID labor force participation (Panels B and C, respectively). The upper and lower halves of each panel contain the estimates of models which do not include and do include covariates, respectively. The first, middle, and final set of columns of each panel pertain to the "exogenous", "equilibrium", and "reduced-form" specifications of the initial conditions, respectively.

For SIPP welfare participation in Panel A, the first column shows that the estimated state dependence in welfare participation is quite large when the initial conditions are treated as exogenous and is insensitive to including covariates. Column (2) suggests that including covariates and treating the initial conditions as being in equilibrium has little effect on the estimated state dependence. In addition, the estimation results for these three models implies that there is no unobserved heterogeneity in individual propensities to participate. However, the model with no covariates and equilibrium initial conditions leads to substantially reduced estimates of the state dependence and a finding of substantial unobserved heterogeneity. An important point is that the "unrestricted" estimates of the variance of the random effect were negative, leading us to restrict the random effects variance to be zero. Also, the non-parametric mass point models were not identified. These findings suggest that the nonlinear likelihood model in these cases may not be well-specified nor converging to the consistent optimum.

The next two columns of Panel A seem to show a substantial reduction in the estimated state dependence when a flexible reduced-form is used to approximate the initial conditions. Also, both the normal heterogeneity model (column 3) and the mass point heterogeneity model (column 4) appear to be "identified". The estimated state dependence is insensitive to which heterogeneity distribution is assumed and to whether covariates are included or not, suggesting that normal heterogeneity is a robust restriction and that the random effects can be approximated well by a small number of mass points. The estimated variance of the random effect in column (3) implies that unobserved heterogeneity accounts for about 60 percent of the total error variance. The estimated state dependence is about 30 percent lower than the estimates in columns (1) and (2), suggesting that the other initial conditions assumptions lead to overstated estimates of the structural state dependence.

The estimated coefficients on the covariates are very sensitive to the initial conditions restrictions, with the estimated effects of demographics (race and education) and family structure (marital status and number of kids) being much larger for the "reduced-form" initial conditions. This suggests that the unobservable initial conditions vary substantially across observable characteristics. Given the changes in the estimated state dependence across specifications, the estimated "long-run" effects of the covariates do not change as much as the "short-run" effects.[22]

---

[22] In the linear model, the long-run effects of the X's is equal to $\beta/(1-\gamma)$. When state dependence is not allowed for ($\gamma=0$), the short-run and long-run impacts of the covariates are identical (β).

The results for SIPP labor force participation in Panel B show that, similar to the results for SIPP welfare participation, specifying exogenous initial conditions (column 1) leads to a large estimated state dependence that is insensitive to the inclusion of covariates. The estimates in this case again imply no unobserved heterogeneity. This may be due to misspecification of the likelihood model since the unrestricted estimates of the heterogeneity variance were negative. In the next two columns, specifying the initial conditions as being in equilibrium substantially reduces the estimated state dependence and leads to a sizable estimated heterogeneity variance. The estimated state dependence is stable under different specifications of the mixing distribution and when covariates are not included. Columns (4) and (5) show that the reduced-form initial conditions approximation leads to almost identical estimates of the state dependence and random effects variance.

However, the initial conditions assumption does matter for the estimated effects of the covariates. With the exception of husband's current earnings, the estimated coefficients are substantially larger in magnitude in the reduced-form initial conditions model than in the equilibrium conditions model. Given that the estimated state dependence is stable, both the estimated short-run and long-run effects of the covariates are greater in the "reduced-form" model. This suggests that assuming stationarity in the initial period inappropriately accounts for feedback between participation and the covariates leading to estimated covariate effects that are biased down.

The PSID labor force participation results in Panel C are qualitatively similar to the SIPP labor force participation results. The only substantive differences are that there is generally less estimated state dependence across all specifications, and the model assuming exogenous initial conditions does lead to a non-zero, but very small, estimated heterogeneity variance. Again, the exogenous initial conditions specification leads to the largest estimated state dependence. The equilibrium initial conditions and reduced-form initial conditions models result in reduced estimates of the state dependence which are nearly identical to each other regardless of the heterogeneity specification or the inclusion of covariates. The estimated effects of the covariates are again very sensitive to the specification of the initial conditions. The magnitude of the coefficient estimates is much greater when the flexible reduced-form specification is used, with the only exception being the coefficient on husband's current earnings.

Interestingly, having very young kids, aged 0-2, has a bigger negative effect on female labor supply than having older kids.

To examine the robustness of the results to violations of the assumption that the transitory errors, u_{it}, are independent over time, we next consider random effects "probit" models in which we allow u_{it} to have a stationary first-order autoregressive AR(1) process. Here, we assume that $u_{it} = \rho u_{it-1} + \nu_{it}$, ν_{it} ~ i.i.d. $N(0, \sigma_\nu^2)$, α_i ~ i.i.d. $N(0, \sigma_\alpha^2)$, and we restrict $\sigma_\alpha^2 + \sigma_\nu^2/(1-\rho^2) = 1$ for identification.[23] The log-likelihood of this model is a function of T-variate integrals that are computationally intractable when using standard numerical approximation methods. Consequently, we estimate the model for all three specifications of the initial conditions using maximum simulated likelihood (MSL) methods to simulate the sample log-likelihood function using monte carlo draws.[24] The probit specification also allows us to examine the sensitivity of the estimates to alternative assumptions on the distribution of the transitory errors.

Table 6 presents the results of the probit specifications for SIPP welfare participation (Panel A), SIPP labor force participation (Panel B), and PSID labor force participation (Panel C). The general pattern of the estimates are similar to those obtained from the logit specifications in Table 5. In all cases, the estimated state dependence is substantially greater when the initial conditions are assumed to be exogenous than when a reduced-form approximation is used. This is particularly true of the two SIPP processes. With the exception of SIPP welfare participation, the equilibrium and reduced-form specifications of the initial conditions lead to broadly similar estimates of the state dependence, although the differences for SIPP labor force participation appear slightly greater than in Table 5. Also, the estimated state dependence is relatively insensitive to the inclusion of covariates. Interestingly, it now seems that there is a similar amount of state dependence in SIPP and PSID labor force participation.

Again, the smallest estimates of the heterogeneity variance occur when assuming exogenous initial conditions in all three samples and when assuming equilibrium initial conditions and allowing for

---

[23] As opposed to the linear model, the probit model can distinguish first-order state dependence from first-order serial correlation (Heckman 1978). In fact, state dependence can be distinguished from even more complicated error correlation structures since even the most general multivariate probit cannot generate a Markov chain. However, this result is attributable to the multivariate probit functional form (Chamberlain 1985).

[24] See Gourieroux and Monfort (1993) and Hyslop (1998) for more details on monte carlo integration methods in the context of nonlinear dynamic panel data models. The notes to Table 6 describe our implementation of the approach.

additional covariates in SIPP welfare participation. For the reduced-form initial conditions, the estimated fraction of the total error variance attributable to unobserved heterogeneity is almost identical to before for all three samples and is largest for SIPP labor force participation. As expected, the estimates of the state dependence have an inverse relationship with the estimates of the random effects variance across the initial conditions specifications and the three samples.

The estimated effects of the covariates appear to be more similar in the equilibrium and reduced-form initial conditions specifications than before, especially in the case of SIPP welfare participation. However, for the two labor force participation processes, the coefficient estimates are still larger in magnitude in the reduced-form model than in the equilibrium model, with the only exception being husband’s current earnings. Again, having children aged 0-2 has a much bigger negative effect on female labor supply than having older kids.

Finally, except for the exogenous initial conditions model in the PSID labor force participation sample, the estimated AR(1) serial correlation coefficient is always negative and statistically significant, with a magnitude of about 0.1-0.2. Although somewhat counterintuitive, the models estimated are already “absorbing” the contribution of permanent heterogeneity, unobserved initial conditions, and state dependence to positive serial correlation over time. Taken literally, this result implies that the estimates of the state dependence in the i.i.d. logit specifications are biased. One can approximately compare the logit and probit coefficients by multiplying the logit coefficients by the normalization factor ($3^{1/2}/\pi$).[25] Focusing on the reduced-form initial conditions model with normal heterogeneity, it appears that the probit estimates of the state dependence are identical to the logit estimates for PSID labor force participation and slightly smaller for the two SIPP processes. Although some of these differences may be due to the different distributional assumptions on the errors, one reason the estimates are reasonably similar is that the estimated negative serial correlation is relatively small in magnitude.

The random effects results can be summarized as follows. All specifications find evidence of substantial positive state dependence for each of the three binary response processes. The flexible reduced-

---

[25] See page 23 of Maddala (1983). Amemiya (1981) suggests a normalization factor equal to 0.625.

form initial conditions models usually result in the smallest estimated state dependence, and the exogenous initial conditions models lead to apparent overstatements of the behavioral dependence. With the exception of SIPP welfare participation, the state dependence estimates in the "equilibrium" and reduced-form initial conditions specifications are similar. In addition, the estimated state dependence is relatively insensitive to the inclusion of covariates, the specification used for the heterogeneity distribution, and the assumptions made on the distribution and serial correlation structure of the transitory errors. There appears to be more state dependence in welfare participation than in labor force participation.

The estimated effects of the covariates, other than husband's current earnings, are much more sensitive to the specification of the initial conditions, especially in the "logit" model. This implies that initial conditions do matter and that accounting for differences in unobserved initial conditions (or pre-sample histories) across individuals is important for obtaining consistent estimates of the effects of the covariates on dynamics in female welfare and labor force participation. The estimated short-run and long-run effects of family structure (marital status, number of kids) are larger in magnitude for the reduced-form initial conditions model than for the other two specifications. The fact that the estimated effect of husband's current earnings is stable across the initial conditions specifications lends credibility to this interpretation. One might not expect the initial conditions to vary by husband's current earnings, since it is a transitory variable conditional on husband's average earnings over the period. Since the flexible reduced-form approximation to the initial conditions is the least restrictive of our approaches, it probably provides the most reliable model estimates that account for the importance of feedback effects.

Fixed Effects Models

Next, we consider the results from estimation of the fixed effects conditional logit models suggested by Chamberlain (1985) and Honoré and Kyriazidou (1997). Under the assumptions of their models, these approaches non-parametrically condition out both the fixed effects and the initial conditions. Therefore, they provide a useful benchmark for gauging the robustness of the more parametric logit estimators in Table 5. However, since the fixed effects approaches absorb a substantial amount of variation in the data, some of which may be informative, correctly specified random effects estimators will be more efficient. Table 7 presents the dynamic conditional logit estimates of the state dependence for SIPP welfare participation (row 1) and SIPP labor force participation (row 2). Column 1 contains the Chamberlain estimator results when no covariates are included, while column 2 has the estimates from the Honoré and Kyriazidou approach with exogenous covariates.[26]

In a model that assumes no exogenous covariates, the conditional logit estimator leads to estimated state dependence that is almost identical to the random effects estimates based on the reduced-form approach to the initial conditions in Table 5 for both SIPP welfare and labor force participation. This strongly suggests that the reduced-form initial conditions approach is robust and that "reasonable" approaches to the initial conditions problem, reduced-form for SIPP welfare participation and equilibrium and reduced-form for SIPP labor force participation, effectively account for the impact of unobserved individual effects and initial conditions on the estimated state dependence. Interestingly, the sampling errors are only slightly larger for the conditional logit approach suggesting that its efficiency loss is low relative to the more parametric random effects estimator.

Similar to the previous findings, the Honoré and Kyriazidou estimation results in column 2 show that the estimates of the state dependence are insensitive to the inclusion of covariates.[27] We have not yet calculated standard errors for these estimates. Although the estimates of the state dependence were relatively insensitive to the choice of bandwidth, the coefficient estimates on the other covariates were much more sensitive. As a result, we do not present the estimated effects of the other covariates in this draft. Equation (6) shows that these coefficient estimates are based on weighted pairwise comparisons of within changes in the covariates conditional on their stationarity in other periods. It is possible that there is little variation in (X_{i1}-X_{i2}) conditional on (X_{i2}=X_{i3}) – that is, little identifying information remains after absorbing so many between and within comparisons. On the other hand, Tables 1 and 4 show that there

---

[26] Only the time varying variables in Table 5 are included as covariates. Pages 13-14 of Honoré and Kyriazidou (1997) show how to form the likelihood for the dynamic conditional logit model when there are more than 4 time periods. Due to time constraints, we have not yet applied the fixed effects logit esimator to the PSID labor force participation data. This is not a substantive issue since the results in Tables 5 and 6 suggest that the results for the PSID will be similar to the SIPP results. Also, since it is generally true that distinguishing between state dependence and unobserved heterogeneity becomes more difficult as the overall dependence in the process increases, the results for the SIPP may be particularly interesting. Regardless, we will examine the PSID data in the near future.

[27] The notes to the table describe our implementation of this estimator.

is variation in (y_{i0}-y_{i3}) conditional on sufficiency class and the same starting and ending points. The estimated state dependence is slightly lower than before for both processes. However, any potential attenuation bias resulting from not appropriately accounting for the feedback of the lagged outcome into current values of the covariates appears to be inconsequential. We will examine these issues more thoroughly in the revision, including obtaining estimates of the sampling errors.

Linear Probability Models

The results in Tables 5, 6, and 7 suggest that about 55-65 percent of the overall persistence in SIPP welfare participation can be attributed to first-order state dependence, while this figure is 40-55 percent for SIPP labor force participation. These estimates are consistent only if the regression function, the distribution of u_{it}, and the relationship between the unobservables and the covariates are correctly specified. For example, the dynamic probit results suggest that estimates of the state dependence based on the assumption that the transitory errors are independent over time may be slightly overstated.

Here, we use the dynamic linear probability regression model to generate additional estimates of the state dependence and to gauge the credibility of the identifying assumptions underlying the dynamic logit and probit estimators. Tables 8 and 9 present the linear probability results for SIPP welfare and labor force participation, respectively.[28] Panel A of both tables presents the least squares estimates when the data are treated as pooled cross-sections and when individual dummies are included in the regression to absorb the fixed effects. Panel B of the tables presents the various estimates based on the instrumental variables approaches applied to the first-differenced data that were discussed above. As opposed to the earlier results, all of the specifications here include unrestricted time-effects.

Several points in Panel A of both tables are noteworthy. First, the overall amount of first-order serial dependence is 0.87 in SIPP welfare participation and 0.85 in SIPP labor force participation. Column (2) of the panel suggests that almost none of this dependence can be explained by the observable

---

[28] See footnote 24. We will examine PSID labor force participation in the revision. Again, the SIPP may be especially interesting since it has greater overall serial dependence than the PSID. Arellano and Carrasco (1996) document an overall first-order autocorrelation of 0.65 in PSID female labor force participation.

covariates. Controlling for the dynamics embodied in first-order state dependence appears to substantially reduce the estimated coefficients on the covariates for both processes. However, the estimated "long-run" effects of the covariates in columns (1) and (2) are almost identical. Column (3) shows that using fixed effects to absorb any unobserved heterogeneity in the non-dynamic model slightly reduces the magnitude of the coefficients on the time-varying family structure variables when compared to the pooled cross-sectional model in column (1).[29]

In column (4), the least-squares fixed effects estimates imply that only about 42% and 34% of the overall persistence in welfare and labor force participation can be attributed to state dependence once the fixed effects are accounted for. However, as discussed above, these estimates will understate the importance of state dependence due to the attenuation bias that arises from assigning too much of the dependence to heterogeneity by not allowing for the contribution of the unobserved initial conditions. Recall that this bias is declining in the number of periods of observation, and that our SIPP samples follow individuals for 8 periods. Again, it appears that the estimated long-run impacts of the covariates are similar when one accounts for dynamics in the process.

Columns (5) and (6) of Panel A allow for richer dynamics. In Column (5) there is some slight evidence of negative second-order state dependence in both processes, although the magnitude of the estimates is small. If the true underlying model has positive first-order state dependence and negative AR(1) serial correlation in the transitory errors, then a model that allows for second-order state dependence but specifies independent errors will lead to a negative coefficient estimate on the 2-period lagged outcome.[30] The model in column (6) allows for two leads and two lags of the time-varying covariates to enter the regression. Although the estimates are small, it appears that the one-period lead of the variables helps predict current participation, especially marital status in the female labor force participation equation. Therefore, these variables may not be strictly exogenous since there appears to be

---

[29] Comparing the estimates from the pooled cross-sectional logit and the fixed-effects conditional logit without state dependence described earlier leads to a nearly identical finding.

[30] Recall that the dynamic probit estimates implied a small and negative AR(1) component for the errors. To examine this possibility in the context of the linear probability model, we examined the sample autocovariance structure of the residuals from the stacked cross-sectional linear probability regression that includes a lagged dependent variable. The results implied a negative AR(1) component for both SIPP processes.

a feedback relationship between the participation outcome and the covariates even conditional on the first-order state dependence. Since the coefficients on the lags of the covariates are generally small and statistically insignificant, with the exception of the first lag of marital status in welfare participation, it may be appropriate to assume that the variables are pre-determined.[31] It is important to note that if the covariates are pre-determined and not strictly exogenous, then their coefficient estimates will also suffer from attenuation biases in the fixed-effects regression when the role of initial conditions is ignored.

Panel B of both tables presents the results from applying two-stage least squares to the first-differenced data using the lag levels of the explanatory variables as instruments for their first-differences. As discussed earlier, this instrumental variables approach purges the attenuation biases that arise in fixed-effects estimation. In column (1), $(y_{it-1}-y_{it-2})$ is instrumented with y_{it-2}. The estimates of the state dependence imply that the estimates in column (4) of Panel A have attenuation biases of about 30% for both SIPP welfare and labor force participation, which is to be expected given the calculations provided in Nickell (1981). In column (2) it appears that using both y_{it-2} and y_{it-3} as instruments for $(y_{it-1}-y_{it-2})$ leads to estimates of the state dependence that are reduced by about 15-20%, suggesting the presence of some negative serial correlation in the transitory errors.[32] Consistent with the probit and logit results, it appears that failing to account for negative serial correlation in the errors leads to slightly overstated estimates of the state dependence. It is easy to show that in a model with no covariates, failing to account for a negative AR(1) correlation coefficient of about -0.15 to -0.20 (from column (3) of Table 6) will lead to overstatements of the size found.[33]

---

[31] Define $X_i^T=(X_{i1}, \ldots, X_{iT})$ and $X_i^t=(X_{i1}, \ldots, X_{it})$, then the strict exogeneity condition on the X's is $E(u_{it}|X_i^T)=0$. Another formulation of the strict exogeneity condition is $E(y_{it}|X_i^T, \alpha_i, y_{it-1})=E(y_{it}|X_i^t, \alpha_i, y_{it-1})$ – that is, leads of the X's do not enter the contemporaneous participation equation conditional on their history, the individual effects and first-order state dependence. The assumption that the X's are pre-determined can be formalized as $E(u_{it}|X_i^t)=0$ and allows for dynamic feedback from y to X of an unspecified form. See Arellano (1995) for a summary of these issues.

[32] When only y_{it-3} is used as an instrument for $(y_{it-1}-y_{it-2})$, the estimated state dependence is 0.65 and 0.57 for SIPP welfare and labor force participation, respectively. The over-identification statistic that tests the similarity of the IV coefficients for the two instruments has a p-value of about 1% for welfare participation and is much smaller for labor force participation. The first-stage relationship between the instruments and the lagged change in participation is very strong, and limited information maximum likelihood (LIML) IV estimation leads to identical results.

[33] In a model with no covariates, the upward bias in the estimated state dependence based on instrumenting $(y_{it-1}-y_{it-2})$ with y_{it-2} when there exists an AR(1) correlation coefficient, ρ, is of order $(\rho-1)\rho$.

In column (3) of Panel B, there is almost no evidence of second-order state dependence in either participation process, suggesting that first-order state dependence is a good approximation. In column (4), we examine the sensitivity of the results to the strict exogeneity assumption on the covariates. In particular, under the assumption that the explanatory variables are predetermined, we use X_{it-1} and X_{it-2} as instruments for $(X_{it}-X_{it-1})$. The estimates of the state dependence are virtually unchanged. Although the sampling errors increase substantially, the estimated effects of the family structure variables systematically increase in magnitude, particularly the estimated impact of number of children on female labor force participation.[34] On one hand, the change in the estimates may be attributable to measurement error in the family structure variables. However, given the size of the change in the estimated effect of kids in Table 9, it is more likely that instrumenting for the family structure variables more adequately accounts for the feedback from participation histories into fertility decisions.[35] These results are consistent with the finding that the method used to account for the initial conditions in the dynamic logit and probit models has a substantial impact on the estimated effects of the covariates. Recall that the less restrictive reduced-form initial conditions approximation led to the largest estimated effects. In both cases, the estimated short-run and long-run effects of the family structure variables on participation increase substantially when less restrictive approaches to the feedback problem are used.[36]

Column (5) of the tables examines the sensitivity of the results to potential non-stationarities in the errors. In particular, we include unrestricted state-time effects in the specifications to account for such factors as state-level economic shocks (labor demand shocks, changes in state unemployment and wage rates) and changes in state policies (e.g., changes in the level and coverage of public-assistance benefits). The estimates remain unchanged suggesting that non-stationarity in the error distribution is not a major source of bias in the conditional logit approaches examined above. Interestingly, for most of the

---

[34] Again the LIML results for both processes are almost identical to the 2SLS results that are presented.

[35] In the context of evaluating the impact of training programs on earnings, Ashenfelter and Card (1985) provide an example where the feedback relationship between individual earnings and participation in the training program may be complicated. See Chamberlain (1993) for the econometric identification problems that arise in this situation.

[36] Arellano and Carrasco (1996) find a similar result when examining the relationship between number of kids and female labor force participation using PSID data. Rosenzweig and Wolpin (1980) conclude that the impact of exogenous changes in fertility on female labor force participation is larger than the observed association between fertility and participation.

specifications, the effect of husband's current earnings on female labor force participation is negligible. In addition, the large sampling errors in the instrumental variables estimates which use lags of the X's as instruments reinforces the possibility that the Honoré and Kyriazidou dynamic conditional logit estimator will result in imprecise estimates of the coefficients on the X's.

Conclusion

This study compared empirical estimates obtained from different specifications of the dynamic binary response panel data model that allows for both unobserved heterogeneity and first-order state dependence. In nonlinear dynamic models, inappropriately accounting for the role of initial conditions and unobserved heterogeneity can lead to severe biases in estimates of the structural state dependence in short panels. We examined both random effects and fixed effects approaches to the model. In addition, the dynamic linear probability regression model was used to gauge the validity of the identifying assumptions underlying the nonlinear models. The various models were used to estimate both female labor force participation and welfare participation equations using panel data from the SIPP and PSID. Hopefully, empirical benchmarking of this type proves useful to researchers interested in estimating dynamic discrete-choice panel data models.

The linear probability regression results imply that about 50 percent of the overall persistence in SIPP welfare participation can be attributed to structural state dependence, while this figure is about 40 percent for SIPP labor force participation. This finding is nearly identical to the results from the dynamic probit that uses the reduced-form approximation to the initial conditions and allows for first-order serial correlation in the transitory errors. It appears that properly specified random effects estimators of the dynamic binary response panel data model can be both robust and precise. In addition, the linear probability regression model provides an attractive alternative to nonlinear likelihood models due to its simplicity and flexibility. The estimated effects of the other determinants of participation are very sensitive to the specification of the initial conditions. Methods that inappropriately account for the potential effects of feedback appear to understate both the short-run and long-run impact of family structure changes on female welfare and labor force participation. This provides evidence that labor supply plans do affect fertility decisions over the life-cycle (see Browning 1992).

Due to the lack of experimental data with a rich time-series, this study has examined non-experimental methods which use the longitudinal structure of the observational data to control for potential confounding factors. Future research which uses experiments and quasi-experiments to examine this question would serve a two-fold purpose. In addition to leading to credible estimates of the structural state dependence in discrete processes, it would provide a guiding metric for determining which non-experimental approaches perform well. In a related point, reliable administrative data that contain many individual observed over a long period at frequent intervals would facilitate an analysis of different approaches to forming the "right comparisons" for identifying the nonlinear dynamic model (e.g., random effects versus fixed effects approaches, discrete panel data models versus duration models).

Appendix: Random Effects Model

Here, we describe the details of the random effects model specification. For close comparability with the conditional logit estimator of Honoré and Kyriazidou (1997), we maintain the assumption that the idiosyncratic error, u_{it}, is logistically distributed. In addition, we assume that the random effects, α_i, are normally distributed with variance σ_α^2. Let $y_i = (y_{i0}, y_{i1}, \ldots, y_{iT})$. Then, given these assumptions:

(a1) $$P(y_i|X_i) = \int_{-\infty}^{+\infty} P(y_{i0} \mid X_i, \alpha) \prod_{t=1}^{T} P(y_{it} \mid X_i, y_{it-1}, \alpha)\, f(\alpha)\, d\alpha,$$

where $P(y_{it} \mid X_i, y_{it-1}, \alpha) = \dfrac{\exp\left(y_{it}\,(X_{it}\beta + \gamma\, y_{it-1} + \alpha)\right)}{1 + \exp\left(X_{it}\beta + \gamma\, y_{it-1} + \alpha\right)}$ and $f(\alpha) = \dfrac{1}{\sqrt{(2\pi)}\,\sigma_\alpha} \exp\left(\dfrac{-\alpha^2}{2\sigma_\alpha^2}\right)$.

It can easily be shown that (a1) is of the form:

(a2) $$P(y_i|X_i) = \int_{-\infty}^{+\infty} e^{-\tilde{\alpha}^2}\, h(\tilde{\alpha})\, d\tilde{\alpha},$$

where $h(\tilde{\alpha}) = \dfrac{1}{\sqrt{\pi}} P(y_{i0} \mid X_i, \tilde{\alpha}) \displaystyle\prod_{t=1}^{T} \dfrac{\exp\left(y_{it}\,(X_{it}\beta + \gamma\, y_{it-1} + \sqrt{2}\sigma_\alpha \tilde{\alpha})\right)}{1 + \exp\left(X_{it}\beta + \gamma\, y_{it-1} + \sqrt{2}\sigma_\alpha \tilde{\alpha}\right)}$ and $\tilde{\alpha} = \dfrac{\alpha}{\sqrt{2}\sigma_\alpha}$.

Given an assumption on the initial conditions of the process -- i.e., $P(y_{i0}|X_i,\alpha)$ -- (a2) can be estimated using Gaussian Quadrature procedures (see Butler and Moffit 1982).

References

Amemiya, Takeshi (1981), "Qualitative Response Models: A Survey," *Journal of Economic Literature*, Vol. 19, pp. 483-536.

An, Mark Yuying, and Ming Liu (1997), "Using Indirect Inference to Solve the Initial Conditions Problem," unpublished manuscript, Department of Economics, Duke University.

Anderson, T. W., and Cheng Hsiao (1981), "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, Vol. 76, pp. 598-606.

Arellano, Manuel (1995), "Panel Data Models with Predetermined Instruments," unpublished manuscript, CEMFI, Madrid.

Arellano, Manuel and Stephen Bond (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, Vol. 58, pp. 277-97.

Arellano, Manuel and Olympia Bover (1995), "Another Look at the Instrumental Variable Estimation of Error-Components Models," *Journal of Econometrics*, Vol. 68, pp. 29-51.

Arellano, Manuel and Raquel Carrasco (1996), "Binary Choice Panel Data Models with Predetermined Variables," Working Paper No. 9618, CEMFI.

Ashenfelter, Orley and David Card (1985), "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs," *Review of Economics and Statistics*, pp. 648-660.

Bane, Mary Jo and David T. Ellwood (1983), "The Dynamics of Dependence: The Routes to Self Sufficiency," Prepared for the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation. Cambridge, MA: Urban Systems Research and Engineering, Inc.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy (1994), "An Empirical Analysis of Cigarette Addiction," *American Economics Review*, Vol. 84, No. 3, pp. 396-418.

Blank, Rebecca (1989), "Analyzing the Length of Welfare Spells," *Journal of Public Economics*, Vol. 39, pp. 245-73.

Browning, Martin (1992), "Children and Household Economic Behavior," *Journal of Economic Literature*, Vol. 30, pp. 1434-1475.

Butler, J.S., and Robert Moffitt (1982), "A Computationally Efficient Quadrature Procedure for the One Factor Multinomial Probit Model," *Econometrica*, Vol. 50, pp. 761-764.

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

Chamberlain, Gary (1980), "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, Vol. 47, pp. 225-38.

Chamberlain, Gary (1984), "Panel Data," in Handbook of Econometrics, Vol. II, edited by Zvi Griliches and Michael Intriligator. Amsterdam: North Holland.

Chamberlain, Gary (1985), “Heterogeneity, Omitted Variables Bias, and Duration Dependence,” in Longitudinal Analysis of Labor Market Data, edited by James Heckman and Burton Singer. Cambridge University Press.

Chamberlain, Gary (1992), “Binary Response Models for Panel Data: Identification and Information,” unpublished manuscript, Department of Economics, Harvard University.

Chamberlain, Gary (1993), “Feedback in Panel Data Models,” unpublished manuscript, Department of Economics, Harvard University.

Chay, Kenneth Y. (1995), “Evaluating the Impact of the 1964 Civil Rights Act on the Economics Status of Black Men Using Censored Longitudinal Earnings Data,” unpublished manuscript, Department of Economics, Princeton University.

Gottschalk, Peter and Robert A. Moffitt (1994), “Welfare Dependence: Concepts, Measures, and Trends,” *American Economic Review*, Vol. 84, No. 2, pp. 38-42.

Gourieroux, Christian and Alain Monfort (1993), “Simulation-based Inference: A Survey with Special Reference to Panel Data Models,” *Journal of Econometrics*, Vol. 59, pp. 5-33.

Gritz, Mark R., and Thomas MaCurdy (1992), “Patterns of Welfare Utilization and Multiple Program Participation Among Young Women,” Working Papers in Economics, Stanford University.

Hahn, Jinyong (1997), “Information Bound of the Dynamic Panel Logit Model with Fixed Effects,” unpublished manuscript, Department of Economics, University of Pennsylvania.

Hajivassiliou, Vassilis and Daniel McFadden (1997), “The Method of Simulated Scores with Application to Models of External Debt Crises,” *Econometrica*, forthcoming.

Ham, John C., and Robert J. Lalonde (1996), “The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training,” *Econometrica*, Vol. 64, No. 1, pp. 175-205.

Heckman, James J. (1978), “Simple Statistical Models for Discrete Panel Data Developed and Applied to Test the Hypothesis of True State Dependence Against the Hypothesis of Spurious State Dependence,” Annales de l'INSEE, No. 30-31, pp. 227-269.

Heckman, James J. (1981a), “Statistical Models for Discrete Panel Data,” Chapter 3 in Manski, Charles and Daniel McFadden (eds.), Structural Analysis of Discrete Data, MIT Press, Cambridge, MA.

Heckman, James J. (1981b), “The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process,” Chapter 4 in Manski, Charles and Daniel McFadden (eds.), Structural Analysis of Discrete Data, MIT Press, Cambridge, MA.

Heckman, James J. (1981c), “Heterogeneity and State Dependence,” in Rosen, Sherwin (ed.), Studies in Labor Markets, University of Chicago Press.

Heckman, James J., and Burton Singer (1984), “A Method for Minimizing the Impact of Distributional Assumption in Econometric Models for Duration Data,” *Econometrica*, Vol. 52, pp. 271-320.

Holtz-Eakin, D., W. Newey, and H.S. Rosen (1988), "Estimating Vector Autoregressions with Panel Data," *Econometrica*, Vol. 56, pp. 1371-95.

Honoré, Bo E. (1992), "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, Vol. 60, pp. 533-565.

Honoré, Bo E. (1993), "Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables," *Journal of Econometrics*, Vol. 59, pp. 35-61.

Honoré, Bo E. and Ekaterini Kyriazidou (1997), "Panel Data Discrete Choice Models with Lagged Dependent Variables," unpublished manuscript, Department of Economics, Princeton University.

Hoynes, Hilary (1996), "Local Labor Markets and Welfare Spells: Do Demand Conditions Matter?" NBER Working Paper Number 5643.

Hoynes, Hilary and Thomas MaCurdy (1994), "Has the Decline in Benefits Shortened Welfare Spells?" *American Economic Review*, Vol. 84, No. 2, pp. 43-48.

Hsiao, Cheng (1986), Analysis of Panel Data. Cambridge University Press.

Hyslop, Dean R. (1998), "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Participation Behavior: Monte Carlo Results and Empirical Evidence for Married Women," unpublished manuscript, UCLA.

Maddala, G. S., (1983), Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press.

Manski, Charles (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, Vol. 55, pp. 357-362.

Marquis, Kent H. and Jeffrey C. Moore (1990), "Measurement Errors in the Survey of Income and Program Participation (SIPP) Program Reports," *1990 Annual Research Conference Proceedings*, Washington, D.C.: U.S. Bureau of the Census, 1990.

Moffitt, Robert (1992), "Incentive Effects of the U.S. Welfare System: A Review," *Journal of Economic Literature*, Vol. 15, pp. 1-61.

Moon, C.-G. and J.G. Stotsky (1993), "The Effect of Rent Control on Housing Quality Change: A Longitudinal Analysis," *Journal of Political Economy*, Vol. 101, pp. 1114-1148.

Nickell, Stephen (1981), "Biases in Dynamic Models with Fixed Effects," *Econometrica*, Vol. 49, pp. 1417-1426.

Plant, Mark W. (1984), "An Empirical Analysis of Welfare Dependence," *American Economic Review*, Vol. 74, No. 4, pp. 673-684.

Roberts, Mark J. and Tames R. Tybout (1997), "The Decision to Export in Columbia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, Vol. 87, pp. 545-564.

Robinson, Peter (1988), “Root-N-Consistent Semiparametric Regression,” *Econometrica*, Vol. 56, pp. 931-954.

Rosenzweig, Mark R. and Kenneth I. Wolpin, (1980) “Life-Cycle Labor Supply and Fertility: Causal Inferences From Household Models,” *Journal of Political Economy*, Vol. 88, pp. 328-348.

Rubin, Donald (1979), “Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies,” *Journal of the American Statistical Association*, Vol. 74, pp. 318-328.

Table 1: Participation Sequence Frequencies

Panel A: SIPP Four-Monthly Welfare Participation

Seq	Freq	Seq	Freq	Seq	Freq
00000000	1076	00000111	27	00011111	16
		00001011	2	00101111	2
00000001	27	00001101	1	01100111	3
00000010	3	00001110	3	01111100	2
00000100	10	00010110	2	10011101	1
00001000	10	00011001	1	11000111	1
00010000	14	00011100	6	11100011	4
00100000	12	00100011	1	11101100	1
01000000	10	00101100	1	11110011	3
10000000	19	00111000	5	11110001	1
	105	01001100	1	11111000	15
		01100100	1		49
00000011	16	01110000	1		
00000110	6	10000011	1	00111111	29
00001001	3	10101000	1	01101111	2
00001010	1	11000001	2	01111110	1
00001100	9	11000100	1	11001111	1
00010100	1	11100000	20	11100111	1
00011000	5		77	11110110	2
00100010	1			11111001	2
00110000	5	00001111	28	11111100	16
01000001	1	00010111	2		54
01010000	1	00011011	1		
01100000	1	00011101	1	01111111	34
10000100	1	00011110	4	10111111	6
10010000	2	00101101	1	11101111	3
10100000	1	00111001	1	11110111	2
11000000	29	00111100	3	11111011	3
	83	01101100	1	11111101	1
		01111000	2	11111110	12
		10000111	2		61
		10111000	1		
		11100100	2	11111111	364
		11100001	1		
		11110000	15		
			65		

Note: A "1" in the t^{th} position of a sequence indicates participation in the t^{th} wave; a "0" indicates non-participation.

Table 1: Participation Sequence Frequencies

Panel B: SIPP Four-Monthly Labor Force Participation

Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq
00000000	1162	00000111	39	00001111	36	00011111	40	00111111	66
		00001011	3	00010111	2	00101111	1	01011111	3
00000001	31	00001101	1	00011011	1	00110111	2	01101111	2
00000010	13	00001110	6	00011101	3	00111011	1	01110111	4
00000100	12	00010011	1	00011110	8	00111110	7	01111011	4
00001000	13	00011001	1	00100111	2	01011101	1	01111101	3
00010000	17	00011100	4	00101011	1	01100111	1	01111110	8
00100000	13	00100011	1	00101101	1	01101101	4	10011111	17
01000000	17	00110001	1	00111001	1	01110011	1	10110111	1
10000000	58	00110010	1	00111010	1	01110110	1	10111101	1
	174	00110100	1	00111100	4	01111001	1	10111110	4
		00111000	4	01000111	1	01111100	3	11001111	9
00000011	44	01000011	4	01001101	1	10001111	8	11011011	1
00000110	7	01000101	1	01001110	1	10010111	1	11011101	1
00001001	3	01000110	1	01100101	1	10011011	2	11011110	4
00001010	1	01001001	1	01100110	2	10011110	4	11100111	19
00001100	7	01001010	1	01101100	1	10100111	2	11101011	4
00010001	1	01001100	1	01101001	1	10101011	2	11101101	2
00010010	2	01010010	1	01110100	1	10101101	1	11101110	6
00010100	1	01010100	1	01111000	5	10110011	3	11110011	10
00100001	2	01100010	2	10000111	6	10110101	1	11110101	4
00100100	3	01101000	1	10010011	1	10111100	3	11110110	3
00101000	1	01110000	3	10011100	1	11000111	7	11111010	4
00110000	5	10000011	10	10100110	1	11001011	1	11111001	15
01000001	5	10000101	1	10101100	1	11001110	1	11111100	49
01010000	2	10000110	2	10110100	3	11010111	1		244
01100000	10	10010001	2	10111000	2	11010101	2		
10000001	5	10010010	3	11000011	9	11010110	1	01111111	73
10000010	6	10010100	1	11000101	2	11011001	1	10111111	30
10000100	2	10100001	1	11000110	3	11011010	1	11011111	38
10010000	3	10100100	1	11001010	1	11011100	4	11101111	19
10100000	4	10110000	2	11010011	1	11100011	6	11110111	25
11000000	76	11000001	4	11011000	2	11100101	2	11111011	32
	190	11000010	1	11100010	3	11100110	5	11111101	24
		11000100	2	11100001	7	11101001	2	11111110	80
		11001000	4	11100100	6	11101100	3		321
		11010000	2	11101000	3	11110001	6		
		11100000	38	11110000	55	11110010	6	11111111	3035
			154		181	11110100	2		
						11111000	61		
							202		

Note: A "1" in the t^th position of a sequence indicates participation in the t^th wave; a "0" indicates non-participation.

Table 1: Participation Sequence Frequencies

Panel C: PSID Annual Labor Force Participation

Seq	Freq	Seq	Freq	Seq	Freq	Seq	Freq
0000000	192	0000111	22	0001111	22	0011111	34
		0001011	5	0010111	1	0101111	6
0000001	27	0001101	2	0011011	2	0110111	7
0000010	12	0001110	2	0011101	1	0111011	1
0000100	5	0010011	3	0011110	6	0111101	3
0001000	9	0010101	1	0100111	8	0111110	4
0010000	10	0010110	2	0101011	3	1001111	13
0100000	11	0011001	1	0101101	0	1010111	5
1000000	36	0011010	0	0101110	1	1011011	4
	110	0011100	0	0110011	1	1011101	3
		0100011	4	0110101	1	1011110	0
0000011	26	0100101	0	0110110	1	1100111	17
0000101	4	0100110	0	0111001	1	1101011	7
0000110	5	0101001	0	0111010	0	1101101	0
0001001	2	0101010	0	0111100	5	1101110	2
0001010	1	0101100	1	1000111	16	1110011	9
0001100	4	0110001	0	1001011	1	1110101	5
0010001	1	0110010	1	1001101	1	1110110	4
0010010	1	0110100	2	1001110	1	1111001	8
0010100	0	0111000	3	1010011	2	1111010	8
0011000	3	1000011	12	1010101	1	1111100	19
0100001	1	1000101	0	1010110	0		159
0100010	1	1000110	2	1011001	1		
0100100	1	1001001	0	1011010	2	0111111	45
0101000	1	1001010	0	1011100	2	1011111	28
0110000	6	1001100	0	1100011	11	1101111	14
1000001	3	1010001	0	1100101	0	1110111	17
1000010	1	1010010	0	1100110	0	1111011	15
1000100	3	1010100	2	1101001	0	1111101	9
1001000	2	1011000	3	1101010	3	1111110	28
1010000	4	1100001	4	1101100	3		156
1100000	28	1100010	5	1110001	4		
	98	1100100	1	1110010	2	1111111	873
		1101000	4	1110100	4		
		1110000	21	1111000	14		
			103		121		

Note: A "1" in the tth position of a sequence indicates participation in the tth wave; a "0" indicates non-participation.

Table 2: Summary Statistics

	Full Sample (1)	Always Participate (2)	Never Participate (3)	Single Transition		Multiple Transitions (6)
				Into Welfare (4)	Out of Welfare (5)	
A: SIPP Welfare Participation						
Age	34.05	30.51	37.08	29.05	31.30	30.24
	(0.26)	(0.45)	(0.37)	(0.74)	(0.79)	(0.74)
Black	0.31	0.43	0.25	0.32	0.37	0.37
	(0.01)	(0.03)	(0.01)	(0.04)	(0.04)	(0.03)
Education	11.26	10.92	11.32	11.31	11.66	11.25
	(0.06)	(0.12)	(0.09)	(0.17)	(0.19)	(0.17)
Married	0.29	0.05	0.39	0.24	0.19	0.30
	(0.01)	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)
Children<18	1.85	2.49	1.56	1.96	1.95	2.08
	(0.03)	(0.06)	(0.04)	(0.09)	(0.11)	(0.09)
AFDC Payments	313.54	370.11	0	274.81	274.58	267.30
	(5.95)	(9.05)		(11.64)	(13.15)	(13.13)
Earnings	341.02	45.66	440.47	235.53	435.31	379.16
	(12.78)	(10.29)	(19.59)	(30.53)	(47.95)	(33.98)
Income	567.83	475.21	590.28	511.13	654.38	613.30
	(13.15)	(14.45)	(20.16)	(32.89)	(52.57)	(42.64)
Family Income	1239.12	835.56	1276.17	1401.55	1316.48	1597.96
	(28.03)	(48.21)	(38.11)	(100.00)	(93.60)	(106.16)
Poverty Level ($)	1026.43	1093.44	981.04	1065.06	1039.65	1109.90
	(7.61)	(15.77)	(10.75)	(23.10)	(23.50)	(23.02)
Average Participation	0.30	1	0	0.53	0.45	0.37
	(0.01)			(0.02)	(0.02)	(0.02)
Sample Size	1934	364	1076	177	126	191

Note: Estimated standard errors are in parentheses. See text for sample selection criteria. AFDC payments, earnings and income figures are monthly (in 1990 $).

Table 2: Summary Statistics
(Continued)

	Full Sample (1)	Always Participate (2)	Never Participate (3)	Single Transition: Into Labor Force (4)	Single Transition: Out of Labor Force (5)	Multiple Transitions (6)
B: SIPP Labor Force Participation						
Age	39.04	38.57	42.40	36.41	38.09	37.36
	(0.13)	(0.16)	(0.33)	(0.52)	(0.55)	(0.37)
Black	0.07	0.07	0.05	0.05	0.07	0.08
	(0.003)	(0.005)	(0.01)	(0.01)	(0.01)	(0.01)
Education	13.21	13.66	12.37	12.87	13.05	12.90
	(0.04)	(0.05)	(0.09)	(0.14)	(0.12)	(0.10)
Children<18	1.24	1.13	1.30	1.72	1.19	1.40
	(0.02)	(0.02)	(0.04)	(0.07)	(0.06)	(0.05)
Husband's Earnings	2961.92	2880.18	3251.02	2885.25	2929.67	2893.65
	(21.45)	(27.44)	(54.54)	(78.65)	(81.07)	(60.16)
Average Participation	0.68	1	0	0.56	0.51	0.57
	(0.01)			(0.01)	(0.01)	(0.01)
Sample Size	5663	3035	1162	329	417	720
C: PSID Labor Force Participation						
Age	34.34	34.52	39.66	33.12	34.35	32.08
	(.23)	(.31)	(.81)	(.44)	(.89)	(.44)
Black	0.22	0.25	0.24	0.15	0.16	0.20
	(.01)	(.01)	(.03)	(.03)	(.03)	(.02)
Education	12.90	13.26	11.86	12.90	12.85	12.67
	(.05)	(.08)	(.17)	(.16)	(.21)	(.11)
Children 0-2	0.26	0.21	0.24	0.28	0.33	0.35
	(0.01)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)
Children 3-5	0.30	0.24	0.26	0.44	0.26	0.41
	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.02)
Children 6-17	0.98	0.95	0.95	1.26	0.61	1.05
	(0.02)	(0.03)	(0.07)	(0.07)	(0.07)	(0.05)
Husband's Earnings	29.59	27.90	35.17	33.64	31.46	28.22
	(.47)	(.64)	(1.93)	(1.97)	(1.56)	(.72)
Average Participation	0.70	1	0	0.55	0.46	0.57
	(.01)			(.02)	(.02)	(.01)
Sample Size	1812	873	192	176	146	425

Note: Estimated standard errors are in parentheses. See text for sample selection criteria. SIPP earnings and income figures are monthly (in 1990 $); PSID earnings are annual (in 1987 $1000).

Table 3: Summary Statistics by Transitions

Change in	Transition Out-of Participation (1)	No Change In Participation (2)	Transition Into Participation (3)
A: SIPP Welfare Participation			
Children<18	-0.043	-0.006	0.164
	(0.03)	(0.004)	(0.03)
Family Income	80.62	0.17	-187.06
	(51.2)	(8.5)	(49.8)
Married	0.023	0.001	-0.092
	(0.01)	(0.001)	(0.02)
Poverty Level	-3.06	-0.25	5.73
	(7.1)	(0.9)	(6.1)
Sample Size	342	12793	403
B: SIPP Labor-force Participation			
Children<18	0.021	-0.002	-0.025
	(0.01)	(0.001)	(0.01)
Husband's Monthly Earnings	79.73	-14.60	-4.59
(1990 $)	(37.4)	(6.1)	(37.9)
Sample Size	1351	37069	1221
C: PSID Labor-force Participation			
Children 0-2	0.053	-0.021	-0.078
	(0.02)	(0.004)	(0.02)
Children 3-5	0.034	0.000	0.024
	(0.02)	(0.004)	(0.02)
Children 6-17	0.044	0.013	0.077
	(0.02)	(0.005)	(0.02)
Husband's Annual Earnings	0.35	0.34	-1.08
(1981 $1000)	(0.39)	(0.12)	(0.38)
Sample Size	685	9471	716

Notes: Standard errors in parentheses, uncorrected for panel sample.

Table 4: Summary Statistics of Conditioning Sequences

	Conditioning Sequence (y_{t-3},y_{t-2},y_{t-1})			
	(0,1,0)	**(0,0,1)**	**(1,1,0)**	**(1,0,1)**
A: SIPP Welfare Participation				
Relative Frequency: y_t=1	0.159	0.715	0.109	0.842
Period (t-1) to t change in:				
Children<18	0.011	0.092	0.011	0.079
	(0.07)	(0.04)	(0.04)	(0.04)
Family Income	-166.03	137.15	30.27	59.92
	(113.84)	(59.47)	(63.42)	(175.55)
Married	0.023	-0.004	0.022	0
	(0.02)	(0.01)	(0.01)	
Sample size	88	260	184	38
B: SIPP Labor Force Participation				
Relative Frequency: y_t=1	0.247	0.732	0.334	0.787
Period (t-1) to t change in:				
Children<18	-0.009	-0.003	0.023	-0.036
	(0.02)	(0.01)	(0.01)	(0.01)
Husband's Earnings	-94.16	-27.90	-101.32	-76.77
	(94.79)	(49.47)	(47.76)	(77.94)
Sample size	219	594	734	305
C: PSID Labor Force Participation				
Relative Frequency: y_t=1	0.403	0.752	0.413	0.700
Period (t-1) to t change in:				
Children 0-2	-0.086	-0.085	0.010	-0.063
	(0.05)	(0.02)	(0.02)	(0.02)
Children 3-5	0.079	-0.036	0.000	-0.029
	(0.05)	(0.03)	(0.03)	(0.03)
Children 6-17	0.058	0.072	0.017	0.048
	(0.04)	(0.03)	(0.03)	(0.03)
Husband's Earnings	-0.194	0.785	0.339	0.799
	(0.79)	(0.82)	(0.67)	(0.55)
Sample size	139	306	303	207

Notes: Standard errors in parentheses.

Table 5: Random Effects Logit Models

	Initial Conditions treated as			
	Exogenous	In Equilibrium	Reduced Form	
	(1)	(2)	(3)	(4)
A: SIPP Welfare Participation				
No Covariates				
Participation$_{t-1}$	5.503	4.012	3.831	
	(0.10)	(0.16)	(0.12)	
Var(α_i)(a)	0	4.060	1.880	
		(0.73)	(0.17)	
Log-likelihood	-2840.89	-3890.62	-3875.88	
With Covariates				
Participation$_{t-1}$	5.137	5.159	3.693	3.662
	(0.10)	(0.10)	(0.15)	(0.15)
Var(α_i)(a)	0	0	4.940	-.---
			(0.81)	
Race (Black=1)	0.101	0.276	0.430	0.435
	(0.08)	(0.05)	(0.17)	(0.18)
Years of Education	-0.075	-0.064	-0.177	-0.182
	(0.01)	(0.01)	(0.03)	(0.03)
Family Poverty Level	0.199	0.312	0.437	0.475
	(0.20)	(0.11)	(0.35)	(0.35)
Age/10	-0.251	0.087	0.007	0.109
	(0.21)	(0.21)	(1.43)	(0.47)
Age^2/100	-0.012	0.055	-0.111	-0.136
	(0.03)	(0.03)	(0.20)	(0.07)
Married	-1.338	-1.269	-2.892	-2.987
	(0.11)	(0.08)	(0.25)	(0.23)
#Kids	0.223	0.179	0.455	0.432
	(0.04)	(0.03)	(0.09)	(0.08)
Log-likelihood	-2672.41	-3670.33	-3480.82	-3476.22

Notes: All models assume logistic errors. Quasi-MLE standard errors are in parentheses. In column (3) the random effects are assumed to be Normally distributed (the unrestricted variance in columns (1) and (2) was negative). In column (4) the random effects are assumed to have a mass point distribution with 4 mass points. The sample consists of N=1934 individuals, observed in each of T=8 periods.

(a) In the random effects models, the normalisation adopted is variance of the logistic distribution: i.e. Var(u_{it}) = $\pi^2/3 \approx 3.290$. This provides a way to evaluate the fraction of total error due to the random effect.

Table 5: Random Effects Logit Models
(Continued)

	Initial Conditions treated as				
	Exogenous	In Equilibrium		Reduced Form	
	(1)	(2)	(3)	(4)	(5)
B: SIPP Labor Force Participation					
No Covariates					
Participation$_{t-1}$	5.181	2.788		2.762	
	(0.04)	(0.08)		(0.08)	
Var(α_i)$^{(a)}$	0	14.585		16.212	
		(1.66)		(1.68)	
Log-likelihood	-9378.61	-12171.26		-12151.76	
With Covariates					
Participation$_{t-1}$	5.074	2.866	2.877	2.897	3.013
	(0.06)	(0.10)	(0.08)	(0.08)	(0.08)
Var(α_i)$^{(a)}$	0	10.938	-.---	10.016	-.--
		(1.47)		(0.71)	
Race (Black=1)	0.137	0.138	0.071	0.573	0.393
	(0.08)	(0.15)	(0.14)	(0.25)	(0.27)
Years of Education	0.069	0.066	0.065	0.273	0.230
	(0.01)	(0.01)	(0.01)	(0.08)	(0.02)
Husband's Avg Earnings	-0.245	-0.274	-0.275	-0.993	-0.835
	(0.04)	(0.07)	(0.06)	(0.11)	(0.10)
Husband's Current Earnings	-0.145	-0.156	-0.163	-0.142	-0.144
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Age/10	1.576	2.240	2.088	5.168	4.503
	(0.15)	(0.27)	(0.23)	(0.45)	(0.35)
Age^2/100	-0.212	-0.295	-0.278	-0.706	-0.629
	(0.02)	(0.03)	(0.03)	(0.05)	(0.04)
#Kids	-0.122	-0.122	-0.123	-0.505	-0.445
	(0.02)	(0.03)	(0.03)	(0.05)	(0.05)
Log-likelihood	-9246.52	-12095.68	-12083.11	-11826.08	-11841.58

Notes: All models assume logistic errors. Quasi-MLE standard errors are in parentheses. In columns (2) and (4) the random effects are assumed to be Normally distributed (the unrestricted variance in column (1) was negative). In columns (3) and (5) the random effects are assumed to have a mass point distribution with 3 and 4 mass points respectively. The sample consists of N=5663 individuals, observed in each of T=8 periods.

(a) In the random effects models, the normalisation adopted is variance of the logistic distribution: i.e. Var(u_{it}) = $\pi^2/3 \approx 3.290$. This provides a way to evaluate the fraction of total error due to the random effect.

Table 5: Random Effects Logit Models

(Continued)

	Initial Conditions treated as					
	Exogenous		**In Equilibrium**		**Reduced Form**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
C: PSID Labor Force Participation						
No Covariates						
Participation$_{t-1}$	3.463		2.121		2.024	
	(0.26)		(0.11)		(0.11)	
Var(α_i)(a)	0.264		3.553		4.701	
	(0.48)		(0.42)		(0.57)	
Log-likelihood	-4016.38		-4909.84		-4899.96	
With Covariates						
Participation$_{t-1}$	3.062	3.110	2.045	1.972	1.924	1.925
	(0.22)	(0.26)	(0.11)	(0.11)	(0.11)	(0.10)
Var(α_i)(a)	0.959	-.---	3.353	-.---	4.505	-.---
	(0.43)		(0.39)		(0.53)	
Race (Black=1)	0.232	0.169	0.259	0.156	0.417	0.398
	(0.11)	(0.10)	(0.14)	(0.12)	(0.17)	(0.20)
Years of Education	0.176	0.166	0.138	0.106	0.294	0.286
	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.09)
Husband's Avg Earnings	-0.483	-0.488	-0.431	-0.339	-0.823	-0.686
	(0.07)	(0.07)	(0.10)	(0.05)	(0.14)	(0.23)
Husband's Current Earnings	-0.301	-0.203	-0.321	-0.254	-0.365	-0.322
	(0.10)	(0.04)	(0.09)	(0.07)	(0.10)	(0.10)
Age/10	1.273	1.409	1.073	1.023	2.412	2.088
	(0.37)	(0.35)	(0.41)	(0.29)	(0.51)	(0.46)
Age^2/100	-0.190	-0.203	-0.162	-0.151	-0.344	-0.308
	(0.05)	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)
#Kids0-2	-0.665	-0.650	-0.719	-0.577	-0.867	-0.805
	(0.09)	(0.07)	(0.09)	(0.08)	(0.10)	(0.10)
#Kids3-5	-0.190	-0.148	-0.096	0.021	-0.346	-0.227
	(0.08)	(0.07)	(0.08)	(0.07)	(0.09)	(0.09)
#Kids6-17	-0.025	-0.028	0.098	0.103	-0.100	-0.048
	(0.04)	(0.04)	(0.05)	(0.03)	(0.06)	(0.08)
Log-likelihood	-3889.94	-3898.48	-4798.33	-4806.57	-4688.56	-4691.09

Notes: All models assume logistic errors. Quasi-MLE standard errors are in parentheses. In columns (1), (3) and (5) the random effects are assumed to be Normally distributed. In columns (2), (4) and (6) the random effects are assumed to have a mass point distribution with 3 mass points. The sample consists of N=1812 individuals, observed in each of T=7 years.

(a) The normalisation adopted is variance of the logistic distribution: i.e. Var(u_{it}) = $\pi^2/3 \approx 3.290$. This provides a way to evaluate the fraction of total error due to the random effect.

Table 6: Random Effects Probit Models

	Initial Conditions treated as		
	Exogenous (1)	**in Equilibrium (2)**	**Reduced Form (3)**
A: SIPP Welfare Participation			
No Covariates			
Participation$_{t-1}$	3.027	1.289	1.713
	(0.06)	(0.14)	(0.18)
Var(α_i)$^{(a)}$	0.119	0.664	0.514
	(0.03)	(0.04)	(0.07)
ρ	-0.169	-0.055	-0.169
	(0.05)	(0.05)	(0.04)
Simulated Log-likelihood	-2828.98	-3891.87	-3877.92
With Covariates			
Participation$_{t-1}$	2.739	3.117	1.621
	(0.07)	(0.05)	(0.13)
Var(α_i)$^{(a)}$	0.182	0	0.502
	(0.03)		(0.05)
ρ	-0.158	-0.222	-0.164
	(0.05)	(0.02)	(0.04)
Race (Black=1)	0.056	0.132	0.131
	(0.04)	(0.04)	(0.05)
Years of Education	-0.038	-0.054	-0.058
	(0.01)	(0.01)	(0.01)
Family Poverty Level	0.126	0.306	0.175
	(0.09)	(0.10)	(0.11)
Age/10	-0.132	-0.049	-0.038
	(0.11)	(0.10)	(0.14)
Age^2/100	-0.006	-0.027	0.030
	(0.01)	(0.01)	(0.02)
Married	-0.658	-0.718	-0.919
	(0.05)	(0.04)	(0.06)
#Kids	0.110	0.108	0.147
	(0.02)	(0.02)	(0.03)
Simulated Log-likelihood	-2665.35	-3447.51	-3481.61

Notes: The models assume Normally distributed random effects and stationary AR(1) errors with correlation ρ. Estimation by Maximum Simulated Likelihood methods, using the Smooth Recursive Conditioning Simulator with 20 replications. Quasi-MLE standard errors are in parentheses. The sample consists of N=1934 individuals, observed in each of T=8 periods.

[a] Fraction of total error variance attributable to the random effect.

Table 6: Random Effects Probit Models

(continued)

	Initial Conditions treated as		
	Exogenous (1)	in Equilibrium (2)	Reduced Form (3)
B: SIPP Labor Force Participation			
No Covariates			
Participation$_{t-1}$	2.842	0.898	1.154
	(0.04)	(0.05)	(0.07)
Var(α_i)$^{(a)}$	0.122	0.753	0.722
	(0.01)	(0.01)	(0.02)
ρ	-0.092	-0.108	-0.201
	(0.04)	(0.02)	(0.02)
Simulated Log-likelihood	-9340.23	-12208.70	-12166.81
With Covariates			
Participation$_{t-1}$	2.762	0.944	1.205
	(0.04)	(0.06)	(0.07)
Var(α_i)$^{(a)}$	0.137	0.727	0.633
	(0.02)	(0.02)	(0.03)
ρ	-0.095	-0.109	-0.201
	(0.04)	(0.03)	(0.02)
Race (Black=1)	0.064	0.053	0.133
	(0.04)	(0.04)	(0.06)
Years of Education	0.037	0.026	0.071
	(0.004)	(0.005)	(0.01)
Husband's Avg Earnings	-0.127	-0.097	-0.231
	(0.02)	(0.02)	(0.03)
Husband's Current Earnings	-0.069	-0.044	-0.045
	(0.03)	(0.02)	(0.02)
Age/10	0.826	0.772	1.372
	(0.07)	(0.09)	(0.11)
Age^2/100	-0.111	-0.103	-0.186
	(0.01)	(0.01)	(0.01)
#Kids	-0.065	-0.048	-0.128
	(0.01)	(0.01)	(0.01)
Simulated Log-likelihood	-9201.08	-12124.47	-11827.16

Notes: The models assume Normally distributed random effects and stationary AR(1) errors with correlation ρ. Estimation by Maximum Simulated Likelihood methods, using the Smooth Recursive Conditioning Simulator with 20 replications. Quasi-MLE standard errors are in parentheses. The sample consists of N=5663 individuals, observed in each of T=8 periods.

[a] Fraction of total error variance attributable to the random effect.

Table 6: Random Effects Probit Models
(continued)

	Initial Conditions treated as		
	Exogenous (1)	in Equilibrium (2)	Reduced Form (3)
C: PSID Labor Force Participation			
No Covariates			
$\text{Participation}_{t-1}$	1.751	1.247	1.226
	(0.07)	(0.11)	(0.11)
$\text{Var}(\alpha_i)^{(a)}$	0.301	0.436	0.428
	(0.05)	(0.04)	(0.04)
ρ	0.085	-0.238	-0.248
	(0.04)	(0.04)	(0.04)
Simulated Log-likelihood	-3954.59	-4910.52	-4905.93
With Covariates			
$\text{Participation}_{t-1}$	1.664	1.155	1.110
	(0.07)	(0.10)	(0.10)
$\text{Var}(\alpha_i)^{(a)}$	0.316	0.451	0.453
	(0.05)	(0.04)	(0.04)
ρ	0.081	-0.229	-0.230
	(0.04)	(0.04)	(0.05)
Race (Black=1)	0.103	0.120	0.135
	(0.04)	(0.05)	(0.06)
Years of	0.076	0.064	0.099
Education	(0.01)	(0.01)	(0.01)
Husband's	-0.207	-0.196	-0.283
Avg Earnings	(0.04)	(0.04)	(0.04)
Husband's	-0.126	-0.149	-0.137
Current Earnings	(0.04)	(0.04)	(0.04)
Age/10	0.671	0.704	0.918
	(0.15)	(0.16)	(0.17)
$\text{Age}^2/100$	-0.094	-0.094	-0.125
	(0.02)	(0.02)	(0.02)
#Kids0-2	-0.287	-0.279	-0.304
	(0.04)	(0.04)	(0.04)
#Kids3-5	-0.052	-0.012	-0.081
	(0.03)	(0.03)	(0.03)
#Kids6-17	-0.006	0.028	-0.022
	(0.02)	(0.02)	(0.02)
Simulated Log-likelihood	-3846.86	-4795.67	-4697.38

Notes: The models assume Normally distributed random effects and stationary AR(1) errors with correlation ρ. Estimation by Maximum Simulated Likelihood methods, using the Smooth Recursive Conditioning Simulator with 20 replications. Quasi-MLE standard errors are in parentheses. The sample consists of N=1812 individuals, observed in each of T=7 years.

(a) Fraction of total error variance attributable to the random effect.

Table 7: Fixed Effects Dynamic Conditional Logit Models (Very Preliminary)

	No Exogenous Covariates	**Exogenous Covariates**
SIPP AFDC Partic$_{t-1}$	**3.669** **(0.131)**	**3.501**
SIPP Labor Force Partic$_{t-1}$	**2.817** **(0.092)**	**2.758**

Notes: Estimated standard errors in parentheses. The dynamic conditional logit without covariates was estimated using the Chamberlain approach, while the model with covariates was estimated using the Honoré and Kyriazidou estimator. A normal kernel weighting function was used. The bandwidth was set to 3 for the AFDC participation model and 2 for the labor force participation model.

Table 8: Linear Probability Models for AFDC Participation from the SIPP

A. Least Squares Estimates

Overall 1st-Order Serial Correlation in Participation: 0.874 (0.004)

	Pooled Cross-Section		Fixed-Effects			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Participation(t-1)		0.836		0.370	0.323	0.125
		(0.005)		(0.008)	(0.010)	(0.013)
Participation(t-2)					-0.076	
					(0.010)	
Married(t)	-0.328	-0.061	-0.271	-0.192	-0.206	-0.165
	(0.008)	(0.005)	(0.013)	(0.013)	(0.015)	(0.024)
Married(t+1)						-0.102
						(0.025)
Married(t+2)						-0.026
						(0.025)
Married(t-1)						-0.056
						(0.023)
Married(t-2)						-0.005
						(0.022)
Kids(t)	0.095	0.016	0.052	0.035	0.029	0.042
	(0.004)	(0.002)	(0.006)	(0.006)	(0.006)	(0.009)
Kids(t+1)						0.017
						(0.009)
Kids(t+2)						-0.010
						(0.010)
Kids(t-1)						0.009
						(0.009)
Kids(t-2)						0.014
						(0.009)
Poverty(t)/100	-0.012	-0.001	-0.006	-0.002	0.002	-0.003
	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)
Black	0.060	0.006				
	(0.007)	(0.004)				
Education	-0.017	-0.003				
	(0.001)	(0.001)				

Notes: Estimated standard errors in parentheses. All specifications include age and age-squared and unrestricted time effects. Married is an indicator for marital status, and Kids is the number of children.

B. Instrumental Variables Estimation of First-Differences

	(1)	(2)	(3)	(4)	(5)
Participation(t-1)	0.546	0.458	0.485	0.436	0.435
	(0.045)	(0.039)	(0.060)	(0.035)	(0.035)
Participation(t-2)			0.026		
			(0.018)		
Married	-0.156	-0.143	-0.152	-0.223	-0.229
	(0.020)	(0.022)	(0.025)	(0.133)	(0.133)
Kids	0.030	0.032	0.034	0.068	0.066
	(0.008)	(0.008)	(0.009)	(0.029)	(0.029)
Poverty/100	-0.003	-0.003	-0.001	-0.011	-0.012
	(0.003)	(0.004)	(0.004)	(0.020)	(0.020)
State-time effects	No	No	No	No	Yes

Notes: Estimated standard errors in parentheses. Specifications (1)-(4) include unrestricted time effects.
Model (1) treats Participation(t-1) as an endogenous variables and instruments its first-difference with Participation(t-2).
Model (2) uses Participation(t-2) and Participation(t-3) as instruments for the first-difference in Participation(t-1).
Model (3) uses Participation(t-2) and Participation(t-3) as instruments for the first-difference in Participation(t-1) and Participation(t-3) and Participation(t-4) as instruments for the first-difference in Participation(t-2).
Model (4) treats all of the variables as being pre-determined and not strictly exogenous. Participation(t-2) and Participation(t-3) are instruments for the first-difference in Participation(t-1). Married(t-1) and Married(t-2) are instruments for Married; Kids(t-1) and Kids(t-2) are instruments for Kids; etc.
Model (5) is the same as Model (4) except it includes unrestricted state-time effects.

Table 9: Linear Probability Models for Labor Force Participation from the SIPP

A. Least Squares Estimates

Overall 1st-Order Serial Correlation in Participation: 0.854 (0.003)

	Pooled Cross-Section		**Fixed-Effects**			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Participation(t-1)		0.843 (0.003)		0.287 (0.005)	0.239 (0.006)	0.065 (0.008)
Participation(t-2)					-0.065 (0.006)	
Huband(t)	-0.006 (0.006)	-0.009 (0.003)	-0.007 (0.003)	-0.007 (0.003)	-0.007 (0.003)	-0.005 (0.005)
Husband(t+1)						0.010 (0.004)
Husband(t+2)						0.004 (0.004)
Husband(t-1)						-0.001 (0.004)
Husband(t-2)						-0.002 (0.004)
Kids(t)	-0.065 (0.002)	-0.007 (0.001)	-0.044 (0.004)	-0.028 (0.004)	-0.031 (0.005)	-0.036 (0.008)
Kids(t+1)						-0.015 (0.007)
Kids(t+2)						-0.003 (0.007)
Kids(t-1)						0.002 (0.008)
Kids(t-2)						-0.005 (0.008)
Black	0.088 (0.0085)	0.012 (0.005)				
Education	0.022 (0.001)	0.003 (0.0005)				

Notes: Estimated standard errors in parentheses. All specifications include age and age-squared and unrestricted time effects. Husband is the husband's current average earnings, and Kids is the number of children.

B. Instrumental Variables Estimation of First-Differences

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Participation(t-1)	0.413	0.329	0.320	0.331	0.329
	(0.024)	(0.019)	(0.032)	(0.019)	(0.019)
Participation(t-2)			0.022		
			(0.012)		
Husband	-0.007	-0.008	-0.007	-0.002	-0.002
	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)
Kids	-0.019	-0.021	-0.023	-0.146	-0.148
	(0.006)	(0.007)	(0.007)	(0.052)	(0.052)
State-time effects	No	No	No	No	Yes

Notes: Estimated standard errors in parentheses. Specifications (1)-(4) include unrestricted time effects.
Model (1) treats Participation(t-1) as an endogenous variables and instruments its first-difference with Participation(t-2).
Model (2) uses Participation(t-2) and Participation(t-3) as instruments for the first-difference in Participation(t-1).
Model (3) uses Participation(t-2) and Participation(t-3) as instruments for the first-difference in Participation(t-1) and Participation(t-3) and Participation(t-4) as instruments for the first-difference in Participation(t-2).
Model (4) treats all of the variables as being pre-determined and not strictly exogenous. Participation(t-2) and Participation(t-3) are instruments for the first-difference in Participation(t-1). Husband(t-1) and Husband(t-2) are instruments for Husband; Kids(t-1) and Kids(t-2) are instruments for Kids.
Model (5) is the same as Model (4) except it includes unrestricted state-time effects.