IZA

IZA DP No. 2332

# Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior

Michael P. Keane
Robert M. Sauer

September 2006

# Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior


**Michael P. Keane**
*University of Technology Sydney*
*and Arizona State University*

**Robert M. Sauer**
*University of Southampton*
*and IZA Bonn*



Discussion Paper No. 2332
September 2006



IZA

P.O. Box 7240
53072 Bonn
Germany

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

IZA Discussion Paper No. 2332
September 2006

## ABSTRACT

# Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior

Two key issues in the literature on female labor supply are: (1) if persistence in employment status is due to unobserved heterogeneity or state dependence, and (2) if fertility is exogenous to labor supply. Until recently, the consensus was that unobserved heterogeneity is very important, and fertility is endogenous. But Hyslop (1999) challenged this. Using a dynamic panel probit model of female labor supply including heterogeneity and state dependence, he found that adding autoregressive errors led to a substantial diminution in the importance of heterogeneity. This, in turn, meant he could not reject that fertility is exogenous. Here, we extend Hyslop (1999) to allow classification error in employment status, using an estimation procedure developed by Keane and Wolpin (2001) and Keane and Sauer (2005). We find that a fairly small amount of classification error is enough to overturn Hyslop's conclusions, leading to overwhelming rejection of the hypothesis of exogenous fertility.

JEL Classification: J2, J6, C3, D1

Keywords: female labor supply, fertility, discrete choice, classification error, simulated maximum likelihood

Corresponding author:

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

# 1 Introduction

For many years, two key issues have played a major role in the literature on female labor supply. One is the attempt to distinguish true state dependence from unobserved heterogeneity as potential explanations for the substantial observed persistence in work decisions (see, e.g., Heckman and Willis (1977), Nakamura and Nakamura (1985), and Eckstein and Wolpin (1989)). The second is the attempt to determine whether children and nonlabor income can reasonably be viewed as exogenous to female labor supply (see, e.g., Chamberlain (1984), Rosenzweig and Schultz (1985), Mroz (1987) and Jakubson (1988)).

Distinguishing state dependence and unobserved heterogeneity can have important implications for policy makers choosing between different labor market and social policies. If persistence is due to unobserved heterogeneity – i.e., relatively immutable differences across individuals in tastes for work, motivation, productivity, etc. – then such policies may have very different effects than if persistence is due to state dependence – i.e., habit persistence, human capital accumulation while working (or depreciation when not), barriers to labor market entry (e.g., costs of job search), etc.. And decisions about whether fertility and nonlabor income may be treated as exogenous have important implications for the proper specification of labor supply functions and estimation of labor supply elasticities.

Until recently, the consensus of the literature was that unobserved heterogeneity is crucially important, and that fertility is endogenous (i.e., women with greater unobserved preferences for work and/or greater unobserved skill endowments tend to have fewer children).[1] But a recent paper by Hyslop (1999) challenged these conclusions. Using recursive importance sampling techniques (see Keane (1994)) he was able to estimate a dynamic panel probit model of female labor supply that included a rich

[1] For instance, Chamberlain (1984) estimated probit models for married womens' labor force participation, and Jakubson (1988) estimated panel Tobit models for married womens' hours, and they both overwhelmingly rejected exogeneity of children.

pattern of unobserved heterogeneity and true state dependence, as well as autoregressive errors. His rather surprising finding was that allowing for autoregressive errors (the computationally difficult part of the exercise) led to a substantial diminution in the apparent importance of permanent unobserved heterogeneity. This, in turn, led to diminution in the importance of correlation between unobserved heterogeneity and children/nonlabor income for labor supply behavior. Hence, he could not reject that fertility and nonlabor income are exogenous to female labor supply decisions.

In this paper, we contribute further to the literature on the determinants of female labor supply behavior by considering classification error. Specifically, we nest a dynamic panel probit model of married women's labor market participation decisions (like that of Hyslop (1999)) within a model of classification error in reported employment status. This introduces a serious computational problem: with classification error, the lagged true choice (and the true state of the agent more generally) becomes unobserved, making simulation of state contingent transition probabilities intractable. Instead, following Keane and Wolpin (2001) and Keane and Sauer (2005) we show how to simulate the likelihood using only unconditional simulations.[2]

Using PSID data on married women's labor market decisions between 1981 and 1987, we first replicate Hyslop (1999)'s results. We then show that even a relatively small amount of classification error can substantially alter conclusions regarding the importance of unobserved heterogeneity. Spurious transitions due to classification error exaggerate the extent of dynamism in the labor market. Consequently, correct-

[2] To the best of our knowledge, the few prior papers that have explicitly treated classification error in discrete choice models of labor supply have considered only static models. For example, Poterba and Summers (1995) show how the relationship between unemployment benefits and labor market participation is substantially altered when the likelihood function of the static multinomial logit model is generalized to take empirical classification error rates into account. Hausman, Abrevaya and Scott-Morton (1998) show how the estimated determinants of job transitions are affected when classification error rates are estimated jointly with the behavioral parameters of a static binary probit model.

ing for classification error greatly increases the estimated importance of permanent unobserved heterogeneity. It also increases the importance of correlation between unobserved heterogeneity and the number of children/level of nonlabor income for female labor supply behavior. Crucially, after controlling for classification error, we can strongly reject the hypothesis that fertility and nonlabor income are exogenous.

The results of this study suggest that researchers estimating dynamic discrete choice models should be careful to check the robustness of results to possible misclassification of the dependent variable. They also provide additional motivation for why it is important to jointly model female labor supply and fertility, as in, e.g., Moffitt (1984), Hotz and Miller (1988), and Keane and Wolpin (2006).

The rest of the paper is organized as follows: In section 2, we specify a dynamic probit model of female labor force participation decisions and nest it within a model of misclassification. In section 3, we outline the simulated maximum likelihood (SML) algorithm that we use to estimate the model. Section 4 describes the PSID data used in the estimation. Section 5 presents the estimation results, while section 6 concludes.

# 2 A Dynamic Panel Data Probit Model with Errors in Classification

## 2.1 Standard Panel Probit Models

Consider the following specification for a married woman's labor market participation decision rule,

$$h_{it} = 1\left(X_{it}'\beta + \gamma h_{it-1} + u_{it} > 0\right), \; i = 1, ..., N, \; t = 0, ..., T \tag{1}$$

where $h_{it}$ denotes the labor market participation choice of woman $i$ at time $t$. $h_{it}$ is equal to one when the expression in parentheses is true, and is equal to zero otherwise. $X_{it}$ is a vector of covariates for woman $i$ in year $t$ that includes measures of nonlabor income (e.g., permanent and transitory annual earnings of the husband), number of children in different age ranges, woman $i$'s age, race, and education, and time dummies. $h_{it-1}$ is woman $i$'s participation outcome in the previous period and $u_{it}$ is an error term. The decision rule is "reduced form" in the sense that we have substituted out for the wage as a function of $X_{it}$ and $h_{it-1}$, and the $X_{it}$ are assumed exogenous under the null (a key hypothesis which we will test).

In the simple static probit formulation, the coefficient $\gamma$ is set to zero and $u_{it}$ is assumed to be serially independent and normally distributed with zero mean and variance $\sigma_u^2$. Normalization for scale is satisfied by setting $\sigma_u^2$ equal to one.

In the static random effects (RE) version of the model, $u_{it}$ is decomposed into two components,

$$u_{it} = \alpha_i + \varepsilon_{it} \tag{2}$$

where $\alpha_i$ is a time-invariant individual effect that is distributed normally with zero mean and variance $\sigma_\alpha^2$. The individual effect $\alpha_i$ generates serial correlation in $u_{it}$. The transitory error component, $\varepsilon_{it}$, is assumed to be serially uncorrelated, conditionally independent of $\alpha_i$, and distributed normally with zero mean and variance $\sigma_\varepsilon^2$. Because $\sigma_u^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2$ and we normalize $\sigma_u^2 = 1$, only $\sigma_\alpha^2$ is directly estimated. Since $\alpha_i$ is meant to capture unobserved preference, motivation and productivity characteristics of woman $i$ that do not change over time, $\sigma_\alpha^2$ is the variance of permanent unobserved heterogeneity.

Although $\alpha_i$ in (2) is usually assumed to be conditionally independent of $X_{it}$, it is possible to allow $\alpha_i$ to be correlated with $Z_{it}$, a vector that contains only the time varying elements of $X_{it}$, e.g., transitory nonlabor income and the number of children in different age ranges.[3] This yields a correlated random effects model (CRE). The correlated random effects probit assumes that the individual effect takes the form,

$$\alpha_i = \sum_{t=0}^{T} Z_{it}'\delta_t + \eta_i \tag{3}$$

[3] Only the time-varying elements of $X_{it}$ can be included in $Z_{it}$ because letting a time invariant element of $X_{it}$ shift $\alpha_i$ is equivalent to letting it shift $X_{it}'\beta$ by a constant amount.

where $\eta_i$ is normally distributed with zero mean and variance $\sigma^2_\eta$. $\eta_i$ is assumed to be conditionally independent of $Z_{it}$ (and $X_{it}$). This implies that $\sigma^2_\eta = Var\,(\alpha_i|Z_i)$, where $Z_i = (Z_{i0}, ..., Z_{iT})$, and that the variance of permanent unobserved heterogeneity is now $\sigma^2_\alpha = Var\left(\sum_{t=0}^{T} Z'_{it}\delta_t\right) + \sigma^2_\eta$. In the correlated random effects model, the $\delta_t$'s are estimated in addition to $\sigma^2_\eta$ and $\beta$. Thus, the exogeneity of children in the household can be directly examined via hypothesis tests on $\delta_t$.[4]

The error term $u_{it}$ can be given a more complex structure than in (2) by relaxing the assumption that $\varepsilon_{it}$ is serially uncorrelated. Serial correlation in $\varepsilon_{it}$ could arise, for example, if data on accepted wages are not exploited in estimation and there is persistence in unobserved wage offers, given that we have substituted out for the wage.[5] Allowing $\varepsilon_{it}$ to follow an $AR(1)$ process we have,

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + v_{it} \tag{4}$$

where $v_{it}$ is normally distributed with zero mean and variance $\sigma^2_v$, and conditionally independent of $\varepsilon_{it-1}$. We assume the process is stationary, so $\sigma^2_\varepsilon = \frac{\sigma^2_v}{(1-\rho^2)}$.[6]

Because of the normalization $\sigma^2_u = 1$, $\sigma^2_v$ is not separately identified. However, the $AR(1)$ serial correlation coefficient $\rho$ can be estimated in addition to $\sigma^2_\eta$, the $\delta_t$'s and $\beta$. More specifically, from our scale normalization and independence assumptions, we have $\sigma^2_u = \sigma^2_\eta + \sigma^2_\varepsilon = 1$, and assuming variance stationarity in the $AR(1)$ process gives $\sigma^2_u = \sigma^2_\eta + \frac{\sigma^2_v}{(1-\rho^2)} = 1$. It then follows that $\sigma^2_v = (1-\rho^2)\left(1-\sigma^2_\eta\right)$. Thus, an estimate of $\sigma^2_v$ can be backed out from the estimates of $\rho$ and $\sigma^2_\eta$.

---

[4]The CRE model was first suggested by Chamberlain (1982) and first used by Chamberlain (1984) to test exogeneity of children to married womens' labor supply (i.e., employment status).

[5]The majority of non-linear discrete choice labor supply studies do not exploit accepted wage data in estimation, as in Heckman (1981), Hyslop (1999) and Magnac (2000). Eckstein and Wolpin (1989) is an exception.

[6]Note that we assume stationarity because Hyslop (1999) did so, and we want our results to differ from his only due to inclusion of classification error.

In addition to estimating $\rho$ and $\sigma_\eta^2$, $\gamma$ in (1) can be allowed to be non-zero. This permits the researcher to measure the relative importance of (i) permanent unobserved heterogeneity, (ii) first-order state dependence, and (iii) $AR(1)$ serial correlation as sources of persistence in observed choice behavior.

In dynamic probit models of the type specified in equations (1) through (4), it is well-known that if the $h_{it}$ process is not observed from its start, simply treating the first observed $h_{i,t-1}$ as exogenous can severely bias the parameter estimates. Several different corrections for this initial conditions problem have been developed. However, the Heckman approximate solution is the correction that is most often used.[7] The Heckman approximation takes the form,

$$\begin{aligned} h_{it} &= 1\left(X_{it}'\beta + \gamma h_{it-1} + u_{it} > 0\right), \, t \geq 1 \\ h_{i0} &= 1\left(X_{i0}'\beta_0 + u_{i0} > 0\right) \\ \rho_t &= corr\left(u_{i0}, u_{it}\right), \, t \geq 1, \end{aligned} \tag{5}$$

where $t = 0$ denotes the first period of observed data (not the start of the $h_{it}$ process). $u_{i0}$ is assumed to be distributed normally with zero mean and variance $\sigma_0^2$. Consistent with our normalization for scale $\sigma_u^2 = 1$, $t \geq 1$, we also normalize $\sigma_0^2$ to one. $\rho_t$ is the correlation coefficient between the error in the first period of observed data, $t = 0$, and the error in period $t$, $t \geq 1$.

Adopting the restriction that the $\rho_t$'s are equal implies that only one correlation coefficient, denoted by $\rho_0$, needs to be estimated. Notice that $\rho_0$ is also the covariance between $u_{i0}$ and the individual effect $\alpha_i$. To see this, consider the Choleski decomposition of $\Omega$, the variance-covariance matrix of $u_{i0}$ and $\alpha_i$,

$$\Omega = Var\begin{pmatrix} u_{i0} \\ \alpha_i \end{pmatrix} = E\begin{bmatrix} u_{i0}^2 & u_{i0}\alpha_i \\ \alpha_i u_{i0} & \alpha_i^2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_0 \\ \rho_0 & \sigma_\alpha^2 \end{bmatrix} = AA' \tag{6}$$

[7] Again, we choose this method for comparability with Hyslop (1999). See Heckman (1981) for more details on various solutions available. For a recently proposed alternative, see Wooldridge (2005).

where

$$A = \begin{bmatrix} 1 & 0 \\ \rho_0 & \sqrt{\sigma_\alpha^2 - \rho_0^2} \end{bmatrix} \tag{7}$$

is the Choleski factor of $\Omega$. Using $A$ to express $u_{i0}$ and $\alpha_i$ as functions of independent standard normal deviates, we have

$$\begin{pmatrix} u_{i0} \\ \alpha_i \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ \rho_0 & \sqrt{\sigma_\alpha^2 - \rho_0^2} \end{bmatrix} \begin{pmatrix} \xi_{i1} \\ \xi_{i2} \end{pmatrix} \tag{8}$$

where $\xi_{ir} \sim i.i.d.\ N(0,1)$, $r = 1, 2$. This implies that

$$\begin{aligned} u_{i0} &= \xi_{i1} \\ u_{it} &= \alpha_i + v_{it} = \rho_0 \xi_{i1} + \left(\sqrt{\sigma_\alpha^2 - \rho_0^2}\right) \xi_{i2} + v_{it}, \ t = 1, ..., T \end{aligned} \tag{9}$$

and $cov(u_{i0}, u_{it}) = E(u_{i0} u_{it}) = E\left(\xi_{i1}\left(\rho_0 \xi_{i1} + \left(\sqrt{\sigma_\alpha^2 - \rho_0^2}\right)\xi_{i2} + v_{it}\right)\right) = \rho_0 E\left(\xi_{i1}^2\right) + \left(\sqrt{\sigma_\alpha^2 - \rho_0^2}\right) E(\xi_{i1}\xi_{i2}) + E(\xi_{i1} v_{it}) = \rho_0$. Since the variances of $u_{i0}$ and $u_{it}$ are normalized to one, the correlation coefficient is equal to the covariance, or $corr(u_{i0}, u_{it}) = cov(u_{i0}, \alpha_i) = \rho_0$.

## 2.2 Incorporating Classification Error

Our contribution to the dynamic probit framework in equations (1) through (5) is to generalize it further by nesting it within a model of classification error in reported choices. Let $h_{it}^*$ denote the reported choice in the data, in contrast to $h_{it}$ which is the true choice generated by the decision rule. Then, consider the following index function,

$$l_{it} = \gamma_0 + \gamma_1 h_{it} + \gamma_2 h_{it-1}^* + \omega_{it} \tag{10}$$

where $l_{it} > 0$ implies $h_{it}^* = 1$, while $h_{it}^* = 0$ otherwise. In our model of classification error, we allow $h_{it}^*$ to be a function of $h_{it}$, as well as $h_{it-1}^*$, conditional on $h_{it}$. The latter is meant to capture possible persistence in misclassification. Persistent misclassification could be responsible for some of the persistence in reported choices, in addition to that generated through state dependence, unobserved heterogeneity and $AR(1)$ errors. The error term $\omega_{it}$ is simply assumed to be independent of $u_{it}$, conditional on $h_{it}$, and distributed logistically.[8] We let $\pi_{jk}$ denote the probability that a true $j$ is recorded as a $k$, where $j, k = 0, 1$.

The intuition for identification of the classification error rate parameters $\{\pi_{jk}\}$ in the static case is quite simple, as discussed by Hausman, Abrevaya and Scott-Morton (1998) (HAS). If we set $\gamma_2 = 0$ then we just have their "Model I." In their notation, we have:

$$\begin{aligned} \pi_{01} &= \frac{e^{\gamma_0}}{1+e^{\gamma_0}} = \alpha_0 \\ \pi_{10} &= \frac{1}{1+e^{\gamma_0+\gamma_1}} = \alpha_1. \end{aligned} \tag{11}$$

The first expression is $\Pr(h^*_{it} = 1 | h_{it} = 0)$, the probability a true 0 is misclassified as a 1, and the second term is $\Pr(h^*_{it} = 0 | h_{it} = 1)$, the probability a true 1 is misclassified as a 0. Notice that:

$$\begin{aligned} E\left[h^*_{it} | X'_{it}\beta\right] &= \alpha_0 \left[1 - F\left(X'_{it}\beta\right)\right] + (1-\alpha_1) F\left(X'_{it}\beta\right) \\ &= \alpha_0 + (1 - \alpha_0 - \alpha_1) F\left(X'_{it}\beta\right), \end{aligned} \tag{12}$$

where $F(\cdot)$ is the normal cdf. HAS point out that (besides the usual condition that $E(X'X)$ exists and is of full rank) identification of this model requires only that $\alpha_0 + \alpha_1 < 1$, which means the probability of an observed 1 is increasing in $F(X'_{it}\beta)$, the probability of a true 1, which in turn is increasing in $X'_{it}\beta$. This condition means classification error can't be so severe that people mis-report their state more often than not, which is certainly a mild requirement. HAS also note that extreme values of $X'_{it}\beta$ convey substantial information about $\alpha_0$ and $\alpha_1$, since, no matter how large

[8] The classification error specification in (10) has been shown to perform quite well in repeated sampling experiments on dynamic probit models, using our estimation procedure to be described below (see Keane and Sauer (2005)), in the sense that the parameters of the process can be recovered with precision, along with the parameters of the "true" process (1) – (5).

is $X_{it}'\beta$, the probability of an observed 1 cannot exceed $1 - \alpha_1$. Similarly, no matter how small is $X_{it}'\beta$, the probability of an observed 0 cannot exceed $1 - \alpha_0$.

HAS also consider identification of the parameter vector $\beta$ under more general models of classification error. They show that semi-parametric identification of $\beta$ (up to scale) in the static discrete choice model with classification error requires only (i) index sufficiency and (ii) that the (observed) choice probability be monotonically increasing in the latent index $X_{it}'\beta$.[9] Note that this monotonicity condition holds in "Model I," provided that $\alpha_0 + \alpha_1 < 1$. However, HAS note that one cannot identify the marginal effects of the $X_{it}$'s on the choice probabilities, or the measurement error rates, without imposing a parametric structure.

Putting equations (1) through (5) and (10) together, we arrive at the following dynamic panel data probit model of female labor force participation decisions with classification error in reported choices,

$$\begin{aligned} h_{it} &= 1\left(X_{it}'\beta + \gamma h_{it-1} + u_{it} > 0\right) \\ u_{it} &= \alpha_i + \varepsilon_{it} \\ \alpha_i &= \sum_{t=0}^{T} Z_{it}'\delta_t + \eta_i \\ \varepsilon_{it} &= \rho\varepsilon_{it-1} + v_{it} \\ h_{i0} &= 1\left(X_{i0}'\beta_0 + u_{i0} > 0\right) \\ \rho_0 &= corr\left(u_{i0}, u_{it}\right) \\ l_{it} &= \gamma_0 + \gamma_1 h_{it} + \gamma_2 h_{it-1}^* + \omega_{it}, \end{aligned} \tag{13}$$

for $i = 1, ..., N$ and $t = 0, ..., T$. The full vector of estimable parameters is $\theta = \left\{\beta, \gamma, \delta, \sigma_\eta^2, \rho, \beta_0, \rho_0, \gamma_0, \gamma_1, \gamma_2\right\}$.[10]

[9] Intuitively, we can find loci of $X_{it}$ values such that the choice probability is constant, implying that the latent index is constant. We can then infer $\beta$ from the movements of the elements of $X_{it}$ within those loci.

[10] Placing various restrictions on the parameters in $\theta$ yields a range of simpler probit models.

## 2.3 Identification

It is important to understand how the parameters of model (13) are identified. We begin by discussing how one can separately identify state dependence ($\gamma$) from serial correlation in the errors (due either to random effects or an $AR(1)$ error component) in the outcome equation (1). The key point is that, if the observed persistence in choices is generated entirely by serially correlated errors, so that $\gamma = 0$ in (1), then lagged $X_{it}$'s do not help to predict the current choice, conditional on the current $X_{it}$. That is:

$$E\left(h_{it}|X_{it}, X_{i,t-1}, ..., X_{i0}\right) = E\left(h_{it}|X_{it}\right). \tag{14}$$

However, if true state dependence is present (i.e., $\gamma \neq 0$), then lagged $X$'s do help to predict the current choice, even conditional on the current $X$. Thus, the presence of a causal effect of lagged $X$'s on current choices is a distinguishing feature of discrete choice models with true state dependence (see Erdem (1998) and Wooldridge (2005). As these authors note, this assertion rules out any *direct* effect of lagged X on current choice).[11]

Of course, as is well known, one cannot disentangle true state dependence from various sources of serial correlation in the errors without some parametric assumptions (see Chamberlain (1984) for discussion). To see this, notice that even if $\gamma = 0$, serial correlation in the errors will imply that:

$$E\left(h_{it}|X_{it}, X_{i,t-1}, h_{i,t-1}\right) \neq E\left(h_{it}|X_{it}\right). \tag{15}$$

That is, $X_{i,t-1}$ and $h_{i,t-1}$ will still help to predict $h_{it}$ because they provide signals of $u_{i,t-1}$ (which is correlated with $u_{it}$).[12] Thus, unless one correctly models the serial

[11]Note that the situation is very different for linear models. In a linear model, if $u_t = X_t'\beta + \varepsilon_t$ where $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$ then we can always write that $u_t = \rho u_{t-1} + X_t'\beta + X_{t-1}'(\rho\beta) + v_t$, so that serial correlation in $\varepsilon_t$ and an effect of lagged X on $u_t$ are observationally equivalent.

[12]For example, if $X_{i,t-1}$ is such that $h_{i,t-1} = 1$ is unlikely, but we nevertheless observe $h_{i,t-1} = 1$, it implies that $u_{i,t-1}$ had a large positive value (despite the fact that $u_{i,t-1}$ is independent of $X_{i,t-1}$).

correlation in the errors, lagged choice will tend to be spuriously significant in the equation for $h_{it}$. Conversely, incorrect specification of how past choices affect current choice will, if true state dependence is present, lead to incorrect inferences about serial correlation.

Thus, assuming a random effects plus first-order Markov structure ($RE + AR(1)$), as in (1), (2) and (4), will lead to a particular decomposition of the sources of persistence into those due to each of these components, but different assumptions may lead to different conclusions.[13] Given this, one should obviously check that one's substantive results – in the present case, conclusions about exogeneity of fertility and nonlabor income to female labor supply – are robust to various alternative structures of serial correlation and state dependence. This was a key point of Hyslop (1999), who examined robustness to several alternative specifications – specifically: $RE$ alone, $RE + AR(1)$, and $RE + AR(1)$+first order state dependence ($SD(1)$). This is a key point of our work as well. We push further in this direction, by asking whether results are robust to allowing for classification error in the outcome variable, both with and without persistence.

The only really new identification issue that arises in our work is how one can distinguish persistence in classification errors (i.e., $\gamma_2 > 0$) from true state dependence (i.e., $\gamma > 0$) or from $AR(1)$ errors. If the true model has either true state dependence ($\gamma > 0$) or $AR(1)$ errors ($\rho > 0$), and we omit this in our empirical specification, we would expect our model to "sop up" this mis-specification by setting $\gamma_2 > 0$. This is because $h^*_{i,t-1}$ is correlated with both $h_{i,t-1}$ and $u_{i,t-1}$. We will see this very clearly in our empirical results below (i.e., the importance of $\gamma_2$ drops substantially when $AR(1)$ errors are included).

Now, to understand how the parameters $\gamma$ and $\gamma_2$ can be distinguished, take first

[13] For example, if state dependence was actually higher than first-order, estimating a first-order structure might lead one to exaggerate the importance of the random effects.

the case where there is no serial correlation in the errors. Then consider the object:

$$E\left(h_{it}^{*}|X_{it},h_{i,t-1}^{*},X_{i,t-1}\right) \tag{16}$$

If $X_{i,t-1}$ matters, it implies there is true state dependence (i.e., persistence in observed outcomes is not due to persistence in classification error alone). The point is that $h_{i,t-1}^{*}$ measures $h_{i,t-1}$ only with error, so additional information is gained by conditioning on $X_{i,t-1}$.[14] Conversely, if there is persistence in classification error but *no* true state dependence, we should have:

$$E\left(h_{it}^{*}|X_{it},h_{i,t-1}^{*},X_{i,t-1}\right)=E\left(h_{it}^{*}|X_{it},h_{i,t-1}^{*}\right). \tag{17}$$

That is, in a first-order Markov model, the lagged state is only a sufficient statistic for lagged inputs if the lagged state is measured *without* error.[15]

Now, if there is serial correlation, the situation is not so simple. Whether or not there is true state dependence depends on whether the above condition (17) holds after integrating out the correlated errors. Hence, our results will depend on the assumed parametric form of error distributions and serial correlation. Of course, this is no different from the situation that arises in trying to distinguish various sources of persistence in models *without* correlated classification error, as our earlier discussion emphasized.

# 3 The Estimation Algorithm

To motivate our estimation procedure, consider first the model of equation (13) without classification error. In panel probit models with $AR(1)$ errors, the order of integration required to form the probability of an observed choice history, and hence

[14]That is, $X_{i,t-1}$ is correlated with $h_{i,t-1}$, even conditional on $h_{i,t-1}^{*}$.

[15]As was the case with equation (14), this assertion rules out any direct effect of lagged X on current choice.

the likelihood function, is $T$, the number of time periods. As we'll see in the next section, we use the same data as Hyslop (1999), where $T = 7$. Numerical evaluation of such high dimensional integrals as many times as would be necessary for maximization of the log-likelihood (i.e., thousands of times) is not feasible. Hence, Hyslop (1999) adopted the GHK recursive importance sampling algorithm developed in Keane (1994) to simulate the likelihood function. We will refer to this estimation procedure as SML-GHK.

Hyslop's use of SML-GHK allowed him to extend the correlated random effects approach of Chamberlain (1984) and Jakubson (1988) to include dynamics (i.e., $AR(1)$ errors and first order state dependence). Chamberlain (1984) and Jakubson (1988) do not use ML because (since they use 4 waves of the PSID) with an unrestricted covariance matrix they get 4 dimensional integrals. This led both of these authors to use a minimum distance technique, invented by Chamberlain (1982), in which they estimate a separate probit (or Tobit) for each year, and then back out what the implied coefficient estimates would have been had all the years been estimated jointly. But this technique does not allow for $AR(1)$ errors or state dependence. Hyslop was able to include dynamics and estimate by ML because he simulated the likelihood.

However, once classification error is introduced, it is no longer feasible to use the SML-GHK algorithm to form the likelihood. GHK works by breaking up the probability of a choice history into a string of transition probabilities, and simulating each transition probability along the string. This becomes infeasible when, due to classification error in endogenous variables, the true state of an agent at each point in time is unobserved. However, while introduction of classification error makes use of the GHK recursive algorithm infeasible, Keane and Wolpin (2001) pointed out that it makes *unconditional* simulation of the probabilities of choice histories feasible.

Thus, we estimate the dynamic probit model in equation (13) using SML combined with the unconditional simulation procedure developed in Keane and Wolpin (2001). They originally applied this procedure to estimation of the structural parameters of complex dynamic programming problems, but Keane and Sauer (2005) show that the procedure is also useful for estimating a range of simpler dynamic panel data models.

For purposes of illustrating the SML algorithm, denote the observed data by $\{h_i^*, X_i\}_{i=1}^N$ where $h_i^* = \{h_{it}^*\}_{t=0}^T$ is the history of reported choices for woman $i$ and $X_i = \{X_{it}\}_{t=0}^T$ is the history of the vector of covariates. Simulation of the likelihood function requires constructing $M$ simulated choice histories for each $\{X_{it}\}_{t=0}^T$ history as follows:

1. For each woman $i$, where $i = 1, ..., N$, draw $M$ sequences of errors from the joint distribution of $(u_{i0}, ..., u_{iT})$ to form $\left\{\left\{\{u_{it}^m\}_{t=0}^T\right\}_{i=1}^N\right\}_{m=1}^M$. This entails forming the error sequences $\left\{\{\xi_{i1}^m, \xi_{i2}^m\}_{i=1}^N\right\}_{m=1}^M$ and $\left\{\left\{\{v_{it}^m\}_{t=1}^T\right\}_{i=1}^N\right\}_{m=1}^M$.

2. Given $\left\{\{X_{it}\}_{t=0}^T\right\}_{i=1}^N$ and the error sequences $\left\{\left\{\{u_{it}^m\}_{t=0}^T\right\}_{i=1}^N\right\}_{m=1}^M$, construct $M$ simulated choice histories for each woman $i$ $\left\{\left\{\{h_{it}^m\}_{t=0}^T\right\}_{i=1}^N\right\}_{m=1}^M$ according to the true choice model in $(1) - (5)$.

3. Construct the classification error rates $\left\{\{\pi_{jkt}^m\}_{t=0}^T\right\}_{m=1}^M$ for each woman $i$, according to the model of misclassification in (10), where $j$ denotes the simulated choice and $k$ denotes the reported choice.

4. Form an unbiased simulator of the likelihood contribution for each woman $i$ as:
$$\widehat{P}\left(h_i^* \mid \theta, X_i\right) = \frac{1}{M}\sum_{m=1}^{M}\prod_{t=0}^{T}\left(\sum_{j=0}^{1}\sum_{k=0}^{1}\pi_{jkt}^m I\left[h_{it}^m = j, h_{it}^* = k\right]\right) \tag{18}$$
where $\theta$ is the vector of model parameters.

Given the model of misclassification in (10), there are four possible classification error rates which can enter steps (3) − (4) of the algorithm,

$$\begin{aligned}
\pi_{11t} &= \Pr\left(h^*_{it} = 1 \mid h_{it} = 1, h^*_{it-1}\right) = \frac{e^{\gamma_0+\gamma_1+\gamma_2 h^*_{it-1}}}{1+e^{\gamma_0+\gamma_1+\gamma_2 h^*_{it-1}}} \\
\pi_{01t} &= \Pr\left(h^*_{it} = 1 \mid h_{it} = 0, h^*_{it-1}\right) = \frac{e^{\gamma_0+\gamma_2 h^*_{it-1}}}{1+e^{\gamma_0+\gamma_2 h^*_{it-1}}} \\
\pi_{10t} &= \Pr\left(h^*_{it} = 0 \mid h_{it} = 1, h^*_{it-1}\right) = 1-\pi_{11t} = \frac{1}{1+e^{\gamma_0+\gamma_1+\gamma_2 h^*_{it-1}}} \\
\pi_{00t} &= \Pr\left(h^*_{it} = 0 \mid h_{it} = 0, h^*_{it-1}\right) = 1-\pi_{01t} = \frac{1}{1+e^{\gamma_0+\gamma_2 h^*_{it-1}}}
\end{aligned} \tag{19}$$

where $\pi_{11t}$ denotes the probability that a one is correctly classified as a one in time $t$, and $\pi_{01t}$ is the probability that a zero is misclassified as a one in time $t$. $\pi_{10t}$ and $\pi_{00t}$ are the corresponding conditional probabilities for reporting a zero. Note that only two classification error rates can be estimated due to the adding up constraint.

In step (4), the likelihood contribution for each woman $i$ is built up by averaging, over $M$ simulated choice histories, the product of the appropriate classification error rates in (19) implied by the simulated choice history $\{h^m_{it}\}^T_{t=0}$ and the observed choice history $\{h^*_{it}\}^T_{t=0}$. The index function $I\left[h^m_{it} = j, h^*_{it} = k\right]$ "picks out" the appropriate classification error rate depending on the values of $h^*_{it}$ and $h^m_{it}$. Generating $M$ simulated choice histories serves to integrate out the true choice probability from the likelihood contribution so that only classification error rates appear in (18). Consistency requires that $M$ and $N$ grow large.[16]

A drawback of the estimation procedure described above is that it does not produce a smooth simulated likelihood function. Holding the draw sequences $\left\{\{\xi^m_{i1}, \xi^m_{i2}\}^N_{i=1}\right\}^M_{m=1}$ and $\left\{\left\{\{v^m_{it}\}^T_{t=1}\right\}^N_{i=1}\right\}^M_{m=1}$ fixed, a change in $\theta$ can induce discrete changes in the $\{h^m_{it}\}^T_{t=0}$ sequence. To partially offset the problem of non-smoothness, one can simply use non-gradient methods of optimization such as the downhill simplex method. But

[16] The small sample properties of the estimator, and generalization of the estimation algorithm for handling missing endogenous state variables during the sample period, can be found in Keane and Sauer (2005). The asymptotic properties of the estimator are the same as in Pakes and Pollard (1989) and Lee (1992).

these are typically very slow to converge. Instead, we consider a modification to the estimation procedure that takes advantage of importance sampling techniques that smooth the likelihood function and enable the use of standard gradient methods of optimization.

We smooth the likelihood by first constructing simulated choice histories $\{h_{it}^m(\theta_0)\}_{t=0}^T$ at an initial $\theta_0$. These simulated choice histories are generated from latent variable sequences $\{U_{it}^m(\theta_0)\}_{t=0}^T$, where $U_{it}^m(\theta_0) = X_{it}'\beta^0 + \gamma^0 h_{it-1}^m + u_{it}^m$. We then hold the $\{h_{it}^m(\theta_0)\}_{t=0}^T$ and $\{U_{it}^m(\theta_0)\}_{t=0}^T$ sequences fixed as we vary $\theta$. Each simulated choice sequence then receives an importance sampling weight, $W_m(\theta)$, that varies with $\theta$, and takes the form,

$$W_m(\theta) = \frac{\prod_{t=0}^{T} \frac{1}{\sigma_v}\phi\left(\frac{U_{it}^m(\theta_0) - X_{it}'\beta - \gamma h_{it-1}^m - \alpha_i^m}{\sigma_v}\right)}{\prod_{t=0}^{T} \frac{1}{\sigma_v^0}\phi\left(\frac{U_{it}^m(\theta_0) - X_{it}'\beta^0 - \gamma^0 h_{it-1}^m - \alpha_i^m}{\sigma_v^0}\right)} \tag{20}$$

where $\sigma_v = \sqrt{\left(1-\sigma_\eta^2\right)\left(1-\rho^2\right)}$ and $\phi$ is the standard normal probability density function. The numerator is the product of standardized $U_{it}^m(\theta_0)$ densities, given the current vector of trial parameters $\theta$, and the denominator is the product of standardized $U_{it}^m(\theta_0)$ densities at the initial vector of trial parameters $\theta_0$. Thus, when $\theta$ changes, sequences that are more (less) likely under the new $\theta$ receive increased (reduced) weight.

The likelihood contribution for each woman $i$ in the smooth case takes the form,

$$\widehat{P}\left(h_i^* \mid \theta, X_i\right) = \frac{1}{M}\sum_{m=1}^{M} W_m(\theta) \prod_{t=0}^{T}\left(\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jkt}^m I\left[h_{it}^m = j, h_{it}^* = k\right]\right). \tag{21}$$

Note that (18) is just a special case of (21) with $W_m = 1$. The simulated likelihood function that results from multiplying likelihood contributions of the form in (21) can also be used to construct standard errors for the estimates obtained by maximizing the likelihood function with contributions defined by (18).[17]

[17]See Keane (1994), Keane and Wolpin (1997), Keane and Wolpin (2001), Bajari, Hong and Ryan

# 4 Data

The data used in estimation are drawn from the 1986 panel of the Panel Study of Income Dynamics (PSID), including both the random Census subsample of families and nonrandom Survey of Economic Opportunities. The sample corresponds to the seven calendar years 1979-85 and includes only women who are between the ages of 18 and 60 in 1980, who are continuously married during the period and who have husbands that are labor force participants in each year. The sample is exactly the same as the one which was used in Hyslop (1999), who graciously provided us with his data set.

We chose to use the same data as in Hyslop (1999) in order to facilitate a comparison of estimation results with and without classification error taken into account. Note that a married woman is classified in these data as a labor force participant if she reports positive annual hours worked and positive annual earnings. Possible measurement error in the annual hours of work and annual earnings variables could be a source of classification error in participation status.

Table 1 presents selected means and standard errors in the estimation sample. The labor market participation rate, which is calculated by computing a participation rate over 7 years for each woman, and then averaging this rate over the $N = 1812$ women in the sample, is .70. Calculating the participation rate over the 1812 women in each year separately reveals a u-shaped pattern over the sample period. The participation rate in 1979 is .71, falls monotonically to .68 in 1982, and subsequently rises to a peak of .73 in 1985. The additional variables displayed in the table, which help explain the level and movements in participation rates, are a woman's nonlabor income, number of children in different age ranges, age, education and race.

Nonlabor income for each woman $i$ in the sample is proxied by her husband's

(2004), and Fernandez-Villaverde and Rubio-Ramirez (2004) for other applications of importance sampling in econometrics.

earnings in year $t$ ($y_{it}$). The sample mean of 29,590 (in constant 1987 dollars) is computed by first averaging annual earnings for each husband between 1979 and 1985, and then averaging over the sample size of 1812. As in Hyslop (1999), the natural logarithm of husband's average earnings over the sample period $y_{mp} = \ln(\frac{1}{N}\sum_i y_{it})$ is used as a proxy for a woman's permanent nonlabor income. A woman's transitory nonlabor income is proxied by $y_{mt} = \ln(y_{it}) - y_{mp}$. $y_{mp}$ and $y_{mt}$ enter as separate covariates in estimation.

The influence of children on female participation rates is captured by defining three different variables, the number of children aged 0-2 years, the number of children aged 3-5 years and the number of children aged 6-17 years. The means of these latter variables are also computed by averaging their values over time for each individual (between 1978 and 1985), and then averaging over the 1812 women in the sample.[18] The number of children aged 0-2 years lagged one year also appears as a separate covariate in estimation (see Hyslop (1999) for discussion). The last three variables in the table, which are also used as covariates in estimation, are age, the highest level of education attained over the sample period (which is then held constant from 1979 to 1985), and race (which equals one if black).

In order to get a sense of the correlation between participation rates and the presence of young children in the household, Figure 1 presents the results of two nonparametric regressions. The curve labelled "Participation Rates" displays the results of locally weighted regressions of a woman's labor market participation rate, calculated over 1979 to 1985 for each individual as in Table 1, on a woman's average number of children aged 0-5 between 1978 and 1985. The figure shows that the

[18] There is substantial over-time variation in the number of children in different age ranges and transitory nonlabor income. The over-time standard deviations and their standard errors (in parentheses) are .159 (.005), .182 (.005), .375 (.001), and .149 (.008), for the three fertility variables (in ascending age order) and transitory nonlabor income, respectively. Significant variation in these variables is important for the CRE estimator.

estimated mean participation rate declines quite sharply with increases in the average number of young children in the household. The curve labelled "Participation Rate Residuals" displays nonparametric regression results using the residuals from a prior OLS regression as the dependent variable. The prior linear regression has a women's average labor market participation rate between 1979 and 1985 as the dependent variable and permanent nonlabor income, age in 1979, age squared in 1979, education and race as covariates. The estimated mean participation rate, after controlling for these additional covariates, similarly declines sharply with increases in the average number of young children.

In addition to being influenced by the presence of young children, female participation rates often display a high degree of underlying persistence. The extent of persistence in employment states in the sample is displayed in Table 2, which computes transitions from participation at time $t-1$ to participation at time $t$, as well as transitions from participation at times $t-2$ *and* $t-1$ to participation at time $t$. The top panel of the table shows an extraordinarily high degree of persistence. The probability of participation at $t$ given participation at $t-1$ is 91%. The persistence in nonparticipation is also high, but not quite as great: 78 percent of nonparticipants at time $t-1$ remain nonparticipants at time $t$. The rate of transition from nonparticipation to participation (.22) is, therefore, 2 and 1/2 times the rate of transition from participation to nonparticipation (.09).

The bottom panel of the table illustrates an important asymmetry in transition rates. The transition rate from nonparticipation at $t-2$ *and* participation at $t-1$ to participation at $t$ (.722) is considerably bigger than the transition rate from participation at $t-2$ *and* nonparticipation at $t-1$ to participation at $t$ (.403). This implies that the error structure is not only random effects (equicorrelation). There is also some type of short run persistence, like first-order serial correlation or first-order state dependence.

The transition patterns displayed in Table 2 are critical for identification of the relative importance of permanent unobserved heterogeneity, AR(1) serial correlation and first-order state-dependence. But, if a non-negligible number of these transitions are spurious, due to misclassification of participation status, there may be a substantial effect on estimates of the relative importance of these factors, as well as on conclusions regarding the endogeneity of nonlabor income and fertility in a correlated random effects model.

# 5 Estimation Results

Tables 3-5 present selected SML estimates of different versions of the general model in (13). In addition to the reported parameter estimates, all specifications control for the number of children aged 0-2 in the previous year, race, maximum years of education, a quadratic in age, and unrestricted year effects. The tables also report the results of likelihood ratio tests for the endogeneity of fertility and nonlabor income.

## 5.1 Random Effects

### 5.1.1 Uncorrelated RE Model

Column (1) of Table 3 reports estimates of a RE model with no AR(1) serial correlation, no first-order state dependence and no correction for classification error (No CE). The estimates were obtained by Hyslop (1999) using the SML-GHK algorithm.[19] The estimated coefficients in Column (1) show that the negative effect of permanent nonlabor income on labor market participation is relatively stronger than the negative effect of transitory nonlabor income. The estimated coefficients on the fertility variables indicate that younger children in the household have a larger depressing effect

[19] Note that these estimates could have been obtained without simulation (e.g., using a numerical method like quadrature.) The reason to use SML here is so that differences with $AR(1)$ models reported in Tables 4-5 don't arise due to simulation per se.

on the probability of participation than do older children. The estimate of $Var(\eta_i)$ implies that 75.9% of the overall error variance is due to permanent unobserved heterogeneity.[20] The nonlabor income and fertility effects as well as the variance of unobserved heterogeneity are precisely estimated.

Column (2) presents the same selection of estimated coefficients after correcting for classification error with the SML algorithm described in Section 3. The model of classification error assumes that there is no persistence in misclassification (No Persistent CE), which is equivalent to imposing the restriction $\gamma_2 = 0$ in (13). Allowing for CE does not produce substantial changes in the coefficients of the covariates.

Importantly, however, note that the estimated variance of the individual effect in Column (2) is considerably larger than in Column (1). The point estimate of the variance increases by 22% (to 93.8%). This implies that permanent unobserved heterogeneity accounts for 93.8% of the overall error variance, as opposed to 76% with no correction for classification error. This large increase in the importance of permanent unobserved heterogeneity suggests that misclassification exaggerates the frequency of transitions between labor market states. This is consistent with Poterba and Summers (1995) where it is found that misclassification exaggerates the flow out of unemployment.

Additional evidence on the presence of misclassification is provided by the estimates of $\gamma_0$ and $\gamma_1$ in Column (2). Using $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$ to calculate the classification error rates in (19), the probability of reporting participation, when the true state is nonparticipation ($\widehat{\pi}_{01}$) is .082. The probability of reporting nonparticipation, when the true state is participation ($\widehat{\pi}_{10}$) is .010. These classification error rates are not large, but they are significantly different from zero.[21] Comparing the log-likelihoods

[20] The proportion of the overall error variance $\sigma_u^2$ due to permanent unobserved heterogeneity is $\frac{\sigma_\eta^2}{\sigma_u^2} = \frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_\varepsilon^2} = \sigma_\eta^2$, following the normalization for scale, $\sigma_u^2 = 1$.

[21] The estimates of $\widehat{\pi}_{10}$ and $\widehat{\pi}_{01}$ that we obtain seem relatively modest. For example, they are much smaller than those found in HAS for job changes (Table 7, Column (6)), where $\widehat{\pi}_{10} = .289$

in Columns (1) and (2) by a likelihood ratio test produces a chi-squared test statistic, with two degrees of freedom, equal to 10.52 and a p-value of .005. If there was no significant classification error, the likelihood ratio test would have revealed this, and our SML algorithm would have produced statistically indistinguishable estimates to those obtained in Column (1) using SML-GHK.

Thus, the results in Column (2) show that even a fairly "small" amount of classification error in the data (i.e., error rates of 8% or less) can lead to serious biases in estimation. In particular, classification error can lead to a severe attenuation bias in the importance of permanent unobserved heterogeneity in female labor supply behavior.

Column (3) presents SML estimates of the random effects model allowing for persistence in misclassification (Persistent CE). The Persistent CE model relaxes the restriction that $\gamma_2 = 0$ in (13). Estimating this model of classification error also produces a much larger variance of the individual effect, as in Column (2). The estimate of $Var(\eta_i)$ in Column (3) implies that permanent unobserved heterogeneity accounts for 94.3% of the overall error variance.

Does allowing for persistence in classification error make a difference? The estimates of the classification error parameters $\gamma_0$, $\gamma_1$ and $\gamma_2$ in Column (3) imply that the probability of reporting participation, when the true state is nonparticipation *and nonparticipation is reported in the previous period*, is .080. The probability of reporting nonparticipation, when the true state is participation *and participation is reported in the previous period*, is .004. These are similar estimated classification error rates to those obtained in the No Persistent CE model where classification error rates are not a function of reported labor market status in the previous period.

---

and $\widehat{\pi}_{01} = .202$. Poterba and Summers (1995) (Table I, Panel B) also report higher classification error rates, ranging between 5 and 28 percent for entry/exit from unemployment. Card (1995) and Freeman (1984) find classification error rates in reported union status that vary between 2 and 3 percent.

However, in the Persistent CE model, the estimated classification error rates significantly change when a different participation status from the current true one is reported in the previous period. That is, the probability of reporting participation, when the true state is nonparticipation, *but participation is reported in the previous period*, is .520. The probability of reporting nonparticipation, when the true state is participation, *but nonparticipation is reported in the previous period*, is .053. The substantial increases in the probability of reporting the wrong labor market state, when that same labor market state is reported in the previous period, suggest that persistent misclassification may be an important source of recorded persistence in female labor force participation data.

Note also that there is a dramatic improvement in the log-likelihood when persistent misclassification is introduced into the model. However, as we will see below, there is also evidence that, to a great extent, including lagged choice in the classification error process proxies for the lagged choice variable and/or $AR(1)$ errors in the participation equation (which are omitted in Table 3).

#### 5.1.2 Correlated RE Model

The last three columns of Table 3, Columns (4)-(6), consider the correlated RE version of the model with No CE, No Persistent CE and Persistent CE respectively. Adding CE to the CRE model produces only modest changes in the estimated effects of fertility and nonlabor income, just as when we added classification error to the RE model. There is also a similar increase in $\widehat{\sigma}^2_\eta$. The main advantage of the CRE model is that it allows one to test the null hypothesis that the individual effect is uncorrelated with fertility and nonlabor income. At the bottom of Columns (4)-(6) we report four separate hypothesis tests - for exogeneity of children in three age ranges and exogeneity of nonlabor income. The chi-squared test statistics and p-values indicate that one can clearly reject the hypothesis that fertility and nonlabor income are exogenous covariates, regardless of whether classification error is included in the model.

Note, however, that the chi-squared test statistics produced by both the No Persistent CE and Persistent CE models are generally much larger than the test statistics produced by the No CE model. This occurs because the CE models generate much larger variances of the individual effect. The increased $\sigma_\alpha^2$ makes it easier to detect correlations between the individual effect and fertility and nonlabor income. Note that $\widehat{\sigma}_\alpha^2$ is bigger in the CRE models with classification error because both $Z_{it}'\widehat{\delta}_t$ is more important and $\widehat{\sigma}_\eta^2$ is much larger (recall that $\sigma_\alpha^2 = Var(\sum_{t=0,T} Z_{it}'\delta_t) + \sigma_\eta^2$).

## 5.2 Random Effects with AR(1) Errors

### 5.2.1 Uncorrelated RE Model

Table 4 reports estimates of the same sequence of RE models as in Table 3. But the models in Table 4 are more general in that they allow for $AR(1)$ serial correlation in the transitory error. That is, the restriction that $\rho = 0$ in (13) is relaxed. Column (1) reports the estimates of the No CE random effects model obtained by Hyslop (1999). Introduction of $AR(1)$ serially correlated errors has a very modest impact on the estimated effects of nonlabor income and fertility. But, the importance of the individual effect is considerably reduced. The variance of the individual effect drops to 55.9% of the overall error variance, compared to 75.9% without $AR(1)$ serial correlation. The estimated $AR(1)$ coefficient ($\widehat{\rho}$) is .687 and is precisely estimated. Thus, $AR(1)$ serial correlation appears to be an important component of the persistence in reported labor market states. Relaxing the restriction that $\rho = 0$ results in a large improvement in the log-likelihood (i.e., by 235 points, comparing Column (1) in Table 3 to Column (1) in Table 4).

Column (2) reports the corresponding No Persistent CE results. Once again, introducing classification error produces a dramatic increase in the variance of the random effect. Permanent unobserved heterogeneity accounts for 83% of the error variance in Column (2), as opposed to only 55.9% in Column (1). Note that the increase in the variance of the random effect is *not* accompanied by a decrease in the strength of the $AR(1)$ serial correlation coefficient. Indeed, $\widehat{\rho}$ slightly increases from .687 in Column (1) to .748 in Column (2). The introduction of $AR(1)$ serial correlation into the No Persistent CE model reduces the fraction of variance due to heterogeneity from 93.8% (see Table 3, Column (2)) to 83% (Table 4, Column (2)).

The estimates of $\gamma_0$ and $\gamma_1$ in Column (2) mean that $\widehat{\pi}_{01}$ is .060 and $\widehat{\pi}_{10}$ is .006. These imply similar estimated classification error rates to those obtained in Column (2) of Table 3. $\widehat{\pi}_{01}$ and $\widehat{\pi}_{10}$ are significantly different from zero. A likelihood ratio test for their joint significance produces a chi-squared statistic, with two degrees of freedom, equal to 9.62 with a p-value of .008. Again, the introduction of a "small" amount of classification error (in this case, a 6% error rate or less) leads to a large increase in the estimate of $Var(\eta_i)$.

Column (3) reports the Persistent CE estimation results. The point estimates and standard errors of the fertility and nonlabor income effects are quite similar to those obtained in Column (2) in the No Persistent CE model. There is also little effect on the importance of permanent unobserved heterogeneity and extent of $AR(1)$ serial correlation after allowing for persistent misclassification. However, there is a noticeable change in the extent of persistence in misclassification. $\widehat{\gamma}_2$ falls to 1.56 in comparison to the estimate of 2.53 obtained in Column (3) of Table 3 without $AR(1)$ serial correlation in the model.

Thus, the strength of the persistence in misclassification is sensitive to the inclusion of $AR(1)$ serial correlation in the model, but both of these sources of dynamics are important in explaining the persistence in labor market states recorded in the data. In Table 4, relaxing the restriction that $\gamma_2 = 0$ results in a relatively large improvement in the log-likelihood of 17 points. Note, however, that this is much smaller than the 203 point improvement we saw in Table 3 when an $AR(1)$ error was not included. Thus, while still highly significant, persistence in classification error does not lead to nearly so great a likelihood improvement once another source of short run persistence ($AR(1)$ errors) is allowed for.

### 5.2.2 Correlated RE Model

Columns (4) – (6) report the correlated RE results for the No CE, No Persistent CE and Persistent CE models. There are no substantial changes in the nonlabor income and fertility effects, the point estimate of $\sigma_\eta^2$, the extent of AR(1) serial correlation, or the classification error rate parameters, in comparison to the corresponding results in Columns (1) – (3). However, comparing Column (4) with Columns (5) – (6), we see there is a crucial difference between the models with and without classification error, in terms of the tests for the endogeneity of fertility and nonlabor income.

As in Hyslop (1999), in the $RE + AR(1)$ model without classification error, the null hypothesis of the exogeneity of fertility and nonlabor income is *not* rejected. However, this surprising result turns out to be very sensitive to accounting for classification error. When we add CE, either with or without persistence, the test statistics dramatically increase in value and the null hypothesis of exogeneity is overwhelmingly rejected. Thus, the conclusion reached in Hyslop (1999), that richer error structures (i.e., $RE + AR(1)$ errors) can correct for the endogeneity of fertility and nonlabor income, is not robust to the inclusion of classification error in the model.[22]

Once again, the difference in the results is related to the severe attenuation bias in the variance of the individual effect when classification error is ignored. Note that $\widehat{\sigma}_\eta^2$ is 83% of the variance in Columns (5) and (6) that include CE, but only 55% in Column (4) where CE is not included. This is consistent with the overall importance of the random effect increasing when we account for measurement error. As the importance of the RE increases, the correlation between it and fertility/nonlabor income becomes

[22] In Table 3, with no AR(1) serial correlation in the model, the null hypothesis of the exogeneity of nonlabor income and fertility was rejected regardless of any correction for classification error. The null hypothesis continues to be rejected, after inclusion of AR(1) serial correlation, only when classification error is taken into account.

easier to detect (and more important as a determinant of labor supply behavior).

## 5.3 RE with AR(1) Errors and First-Order State Dependence

### 5.3.1 Uncorrelated RE Model

Table 5 reports the results of estimating more general RE models which allow for both AR(1) serial correlation and first order state dependence ($SD(1)$). The initial conditions problem that arises when $SD(1)$ is included in the model is dealt with by employing the Heckman approximate solution. Column (1) reports the No CE estimation results of this model from Hyslop (1999). The coefficient on lagged participation is a strong 1.063 and is precisely estimated. The inclusion of lagged participation in the model reduces the variance of the individual effect from 55.9% to 48.2% of the total error variance compared to Column (1) of Table 4. Note that the estimate of the $AR(1)$ serial correlation coefficient $\widehat{\rho}$ falls dramatically from .687 in Column (1) of Table 4 to -.219.[23]

The estimates of the No Persistent CE model in Column (2) are considerably different. In particular, the $AR(1)$ serial correlation coefficient falls by much less, the $SD(1)$ effect is more moderate and the variance of unobserved heterogeneity is larger. Specifically, $\widehat{\rho}$ remains positive and falls only to .589 (as opposed to -.219), the first order state dependence coefficient is .843 (as opposed to 1.063), and the variance of unobserved heterogeneity is .732 (as opposed to .479).

Thus, failure to account for classification error produces substantial attenuation biases in the importance of unobserved heterogeneity and $AR(1)$ serial correlation,

[23] It is worth recalling that what identifies the $AR(1)$ parameter vs. state dependence is whether lagged $X$'s help predict current choice in the reduced form of the model where one substitutes out for $h_{i,t-1}$.

and an upward bias in extent of first order state dependence.[24] The relative importance of permanent unobserved heterogeneity and first-order state dependence in explaining persistence in the data is thus quite sensitive to misclassification of labor market states. Note that the estimated classification error rates ($\widehat{\pi}_{01} = .073$ and $\widehat{\pi}_{10} = .015$) are similar in magnitude to those obtained in the corresponding specifications in Tables 3 and 4 and remain statistically significant.

The estimates of the Persistence CE model in Column (3) lead to similar general conclusions. Allowing for persistence in classification error further weakens the $SD(1)$ effect and slightly strengthens the importance of permanent unobserved heterogeneity and $AR(1)$ serial correlation. There is still substantial persistence in misclassification and the value of the log-likelihood increases by a relatively large amount (i.e., 17 points), when $\gamma_2$ is included.

### 5.3.2 Correlated RE Model

Columns (4)-(6) report the results of estimating the correlated RE version of the model with No CE, No Persistent CE and Persistent CE. There is, as in Table 4, a crucial difference in terms of the tests for the endogeneity of fertility and nonlabor income. The correlated RE model with AR(1) serial correlation and first order state dependence, but no classification error, fails to reject the null hypothesis of the exogeneity of fertility and nonlabor income. In contrast, the versions of the model that include classification error (with or without persistence) clearly indicate that fertility and nonlabor income are endogenous, regardless of the introduction of a richer error structure, or dynamics in the form of first order state dependence.

[24] The main parameter of the Heckman approximate solution to the initial conditions problem, $\widehat{\rho}_0 = \widehat{Corr}(u_{i0}, u_{it})$, also suffers from an attenuation bias.

# 6 Conclusion

Estimating the relative importance of state dependence and permanent unobserved heterogeneity, in addition to the influence of children and nonlabor income, has long been an important focus in panel data studies of female labor supply. In this paper, we contribute to the literature on female labor supply by nesting a dynamic panel data probit model of labor market participation decisions within a model of classification error in reported employment status. Relatively few papers have treated classification error in discrete choice models as a serious problem, and to the best of our knowledge, there are no previous studies that have attempted to correct for classification error in dynamic panel probit models with complex error structures.

Using PSID data on married women's labor market participation decisions between 1979 and 1985, we find that ignoring even a fairly small amount of classification error in the data (i.e., error rates of 8% or less) leads to severe attenuation biases in the variance of permanent unobserved heterogeneity and the extent of AR(1) serial correlation. It also leads to an overestimate of the effect of first order state dependence. Adjusting for classification error considerably reweights the relative importance of these factors that are crucial in explaining the persistence in female labor market participation decisions. Our model of classification error also reveals that persistent misclassification is an additional source of persistence in data on female labor force participation, above and beyond the persistence generated by permanent unobserved heterogeneity, $AR(1)$ serial correlation in transitory errors and first order state dependence.

After correcting for classification error, and obtaining a large increase in the estimated variance of permanent unobserved heterogeneity in dynamic correlated random effects versions of the model, we can reject the hypothesis that fertility outcomes and nonlabor income are exogenous covariates. This is in sharp contrast to previous findings where classification error is not taken into account (see Hyslop (1999)). This suggests that researchers estimating dynamic nonlinear discrete choice models should be careful to consider the possible impact of misclassification of the dependent variable on their results.

In this work we have only considered models with first order state dependence in the main equation and the measurement error process. The relative importance of unobserved heterogeneity vs. state dependence, and the conclusions about the endogeneity of fertility outcomes and nonlabor income may also be sensitive to the assumption of first order state dependence. Therefore, it may be worthwhile to estimate similar models which allow for more complex forms of structural state dependence. Future work may also consider the robustness of these findings to different, and perhaps more general, models of classification error.

# References


[1] Bajari, P., Hong H., and S. Ryan (2004), "Identification and Estimation of Discrete Games of Incomplete Information," unpublished manuscript.

[2] Card, D. (1996), "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrics*, 64(4), pp. 957-979.

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

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

[5] Eckstein, Z. and K.I. Wolpin (1989), "Dynamic Labor Force Participation of Married Women and Endogenous Work Experience," *Review of Economic Studies*, 56, pp. 375-390.

[6] Erdem, T. (1998), "An Empirical Analysis of Umbrella Branding," *Journal of Marketing Research,* 35(3), pp. 339-351.

[7] Fernandez-Villaverde J., J.F. Rubio-Ramirez (2004), "Estimating Non-linear Dynamic Equilibrium Economies: A Likelihood Approach," unpublished manuscript.

[8] Freeman, R. (1984), "Longitudinal Analysis of the Effects of Trade Unions," *Journal of Labor Economics*, 2, pp. 1-26.

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

[10] Heckman, J. J. (1981), "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time–Discrete Data Stochastic Process and Some Monte Carlo Evidence," In Structural Analysis of Discrete Data with Econometric Applications, edited by Charles F. Manski and Daniel McFadden. Cambridge, Mass.: MIT Press.

[11] Heckman, J.J. and R. Willis (1977), "A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women," *Journal of Political Economy*, 85, pp. 27-58.

[12] Hotz, V.J. and R.A. Miller (1988), "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply," *Econometrica*, 56(1), pp. 91-118.

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

[14] Jakubson, G. (1988), "The Sensitivity of Labor Supply Parameter Estimates to Unobserved Individual Effects: Fixed- and Random-Effects Estimates in a Nonlinear Model Using Panel Data," *Journal of Labor Economics*, 6, pp. 302-329.

[15] Keane, M.P. (1994), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95-116.

[16] Keane, M.P., and R.M. Sauer (2005), "A Computationally Practical Simulation Estimation Algorithm for Dynamic Panel Data Models with Unobserved Endogenous State Variables," unpublished manuscript.

[17] Keane, M.P., and K.I. Wolpin (1997), "The Career Decisions of Young Men," *Journal of Political Economy*, 105(3), 473-522.

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

[19] Keane, M.P., and K.I. Wolpin (2006), "The Role of Labor and Marriage Markets, Preference Heterogeneity and the Welfare System in the Life Cycle Decisions of Black, Hispanic and White Women," unpublished manuscript.

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

[21] Magnac T. (2000), "Subsidized Training and Youth Unemployment: Distinguishing Unobserved Heterogeneity From State Dependence in Labour Market Histories," *Economic Journal*, 110, pp. 805-857.

[22] Moffitt R. (1984), "Life Cycles Profile of Labour Supply and Fertility," *Review of Economic Studies*, 51(2), pp. 263-278.

[23] Mroz T.A. (1987), "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55(4), pp. 765-799.

[24] Nakamura, A and M. Nakamura (1985), "Dynamic Models of the Labor Force Behavior of Married Women Which Can Be Estimated Using Limited Amounts of Information," *Journal of Econometrics*, 27, pp. 273-298.

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

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

[27] Rosenzweig, M.R and T.P. Schultz (1985), "The Demand for and Supply of Births: Fertility and its Life Cycle Consequences," *American Economic Review*, 75(5),pp. 992-1015.

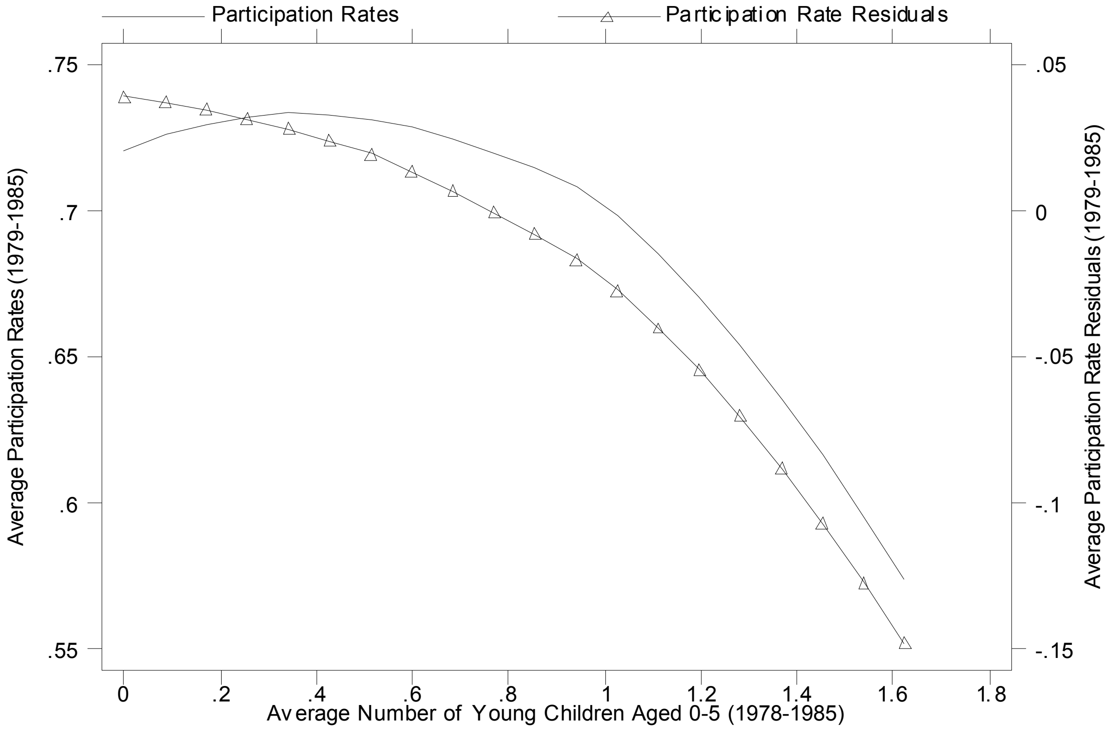
[28] Wooldridge, J.M. (2005), "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Effects", *Journal of Applied Econometrics*, 20(1), pp. 39-54.

Table 1
Sample Characteristics
PSID Waves 12-19 (1978-85)
(N=1812)

| | Mean (1) | Std. Dev. (2) |
|---|---|---|
| Participation | .705 | .362 |
| (avg. over 1979-1985) | (.008) | |
| Participation 1979 | .710 | .454 |
| | (.011) | |
| Participation 1980 | .694 | .461 |
| | (.011) | |
| Participation 1981 | .687 | .464 |
| | (.011) | |
| Participation 1982 | .682 | .466 |
| | (.011) | |
| Participation 1983 | .700 | .458 |
| | (.011) | |
| Participation 1984 | .733 | .442 |
| | (.010) | |
| Participation 1985 | .727 | .445 |
| | (.010) | |
| Husband's Annual Earnings | 29.59 | 19.97 |
| (avg. over 1979-1985) | (.47) | |
| No. Children aged 0-2 years | .249 | .313 |
| (avg. over 1978-1985) | (.007) | |
| No. Children aged 3-5 years | .296 | .338 |
| (avg. over 1978-1985) | (.008) | |
| No. Children aged 6-17 years | .989 | .948 |
| (avg. over 1978-1985) | (.022) | |
| Age | 34.34 | 9.77 |
| (1980) | (.02) | |
| Education | 12.90 | 2.33 |
| (maximum over 1979-1985) | (.05) | |
| Race | .216 | .412 |
| (1=Black) | (.010) | |

Note: Means and standard errors (in parentheses) for 1812 continuously married women in the PSID between 1979 and 1985, aged 18-60 in 1980, with positive annual earnings and hours worked each year for both partners in the married couple. Earnings are in thousands of 1987 dollars. Variable definitions and sample selection criteria are the same as those chosen by Hyslop (1999).

Figure 1
Labor Market Participation and Young Children




Note: The curve labelled "Participation Rates" displays the results of locally weighted (nonparameteric) regressions of a woman's labor market participation rate between 1979 and 1985 on the average number of children aged 0-5 between 1978 and 1985. The curve labelled "Participation Rate Residuals" displays analogous results using the residuals from a prior OLS regression as the dependent variable. The prior linear regression uses a women's labor market participation rate between 1979 and 1985 as the dependent variable and permanent non-labor income, age in 1979, age squared in 1979, education and race as covariates.

Table 2
Participation Transition Matrices

| | | Participate in t | | |
|---|---|---|---|---|
| | Participate in t-1 | 0 | 1 | Total |
| | 0 | .780 | .220 | 3249 |
| | 1 | .090 | .910 | 7623 |

| | | Participate in t | | |
|---|---|---|---|---|
| Participate in t-2 | Participate in t-1 | 0 | 1 | Total |
| 0 | 0 | .832 | .168 | 2134 |
| 1 | 0 | .597 | .403 | 590 |
| 0 | 1 | .278 | .722 | 632 |
| 1 | 1 | .063 | .937 | 5704 |

Note: The cells contain row percentages. The last column contains row totals (person-years) of the past participation profile.

Table 3
Random Effects Probit Models of Participation
(SML Estimates)

| | RE | | | CRE | | |
|---|---|---|---|---|---|---|
| | No CE | No Persistent CE | Persistent CE | No CE | No Persistent CE | Persistent CE |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $y_{mp}$ | -.314 | -.362 | -.349 | -.341 | -.400 | -.375 |
| | (.05) | (.02) | (.02) | (.05) | (.04) | (.04) |
| $y_{mt}$ | -.106 | -.107 | -.141 | -.099 | -.127 | -.172 |
| | (.03) | (.02) | (.03) | (.03) | (.02) | (.03) |
| $\#Kids0\text{-}2_t$ | -.354 | -.309 | -.328 | -.300 | -.290 | -.388 |
| | (.03) | (.03) | (.04) | (.03) | (.04) | (.05) |
| $\#Kids3\text{-}5_t$ | -.293 | -.291 | -.270 | -.247 | -.265 | -.271 |
| | (.03) | (.02) | (.03) | (.03) | (.03) | (.04) |
| $\#Kids6\text{-}17_t$ | -.097 | -.103 | -.101 | -.084 | -.090 | -.087 |
| | (.02) | (.01) | (.02) | (.03) | (.02) | (.03) |
| $Var(\eta_i)$ | .759 | .938 | .943 | .804 | .938 | .943 |
| | (.01) | (.06) | (.08) | (.02) | (.07) | (.10) |
| $\gamma_0$ | - | -2.409 | -2.445 | - | -2.427 | -2.386 |
| | | (.09) | (.11) | | (.09) | (.11) |
| $\gamma_1$ | - | 6.991 | 5.326 | - | 6.996 | 5.056 |
| | | (.21) | (.21) | | (.21) | (.19) |
| $\gamma_2$ | - | - | 2.527 | - | - | 2.611 |
| | | | (.11) | | | (.11) |
| *Log-Likelihood* | -4916.05 | -4910.79 | -4707.84 | -4888.38 | -4878.27 | -4672.62 |
| $N$ | 1812 | 1812 | 1812 | 1812 | 1812 | 1812 |
| $\delta_{\#Kids0-2}=0$ | - | - | - | 32.36(.00)$^{**}$ | 52.14(.00)$^{**}$ | 57.34(.00)$^{**}$ |
| $\delta_{\#Kids3-5}=0$ | - | - | - | 12.77(.12) | 49.04(.00)$^{**}$ | 61.04(.00)$^{**}$ |
| $\delta_{\#Kids6-17}=0$ | - | - | - | 21.74(.01)$^{**}$ | 49.50(.00)$^{**}$ | 61.19(.00)$^{**}$ |
| $\delta_{ymt}=0$ | - | - | - | 48.50(.00)$^{**}$ | 50.08(.00)$^{**}$ | 62.60(.00)$^{**}$ |

Note: All specifications include number of children aged 0-2 years lagged one year, race, maximum years of education over the sample period, a quadratic in age, and unrestricted year effects. Non-labor income is measured by $y_{mp}$ and $y_{mt}$ which denote husband's permanent (sample average) and transitory (deviations from sample average) annual earnings, respectively. $Var(\eta_i)$ is the variance of permanent unobserved heterogeneity and the $\gamma$'s are the classification error parameters. $^{*}$ indicates significance at the 1% level and $^{**}$ indicates significance at the 5% level.

Table 4
Random Effects Probit Models of Participation with AR(1) Errors
(SML Estimates)

| | RE+AR(1) | | | CRE+AR(1) | | |
|---|---|---|---|---|---|---|
| | No CE (1) | No Persistent CE (2) | Persistent CE (3) | No CE (4) | No Persistent CE (5) | Persistent CE (6) |
| $y_{mp}$ | -.316 | -.346 | -.347 | -.332 | -.345 | -.345 |
| | (.05) | (.00) | (.00) | (.05) | (.00) | (.00) |
| $y_{mt}$ | -.097 | -.077 | -.070 | -.097 | -.112 | -.085 |
| | (.03) | (.01) | (.01) | (.03) | (.01) | (.01) |
| $\#Kids0\text{-}2_t$ | -.311 | -.305 | -.302 | -.272 | -.306 | -.307 |
| | (.03) | (.02) | (.02) | (.03) | (.02) | (.02) |
| $\#Kids3\text{-}5_t$ | -.270 | -.274 | -.273 | -.234 | -.265 | -.269 |
| | (.03) | (.01) | (.01) | (.03) | (.01) | (.01) |
| $\#Kids6\text{-}17_t$ | -.089 | -.077 | -.075 | -.077 | -.079 | .083 |
| | (.02) | (.00) | (.00) | (.02) | (.01) | (.01) |
| $Var\,(\eta_i)$ | .559 | .830 | .832 | .546 | .830 | .831 |
| | (.04) | (.03) | (.04) | (.04) | (.03) | (.04) |
| $\rho$ | .687 | .748 | .747 | .696 | .746 | .748 |
| | (.03) | (.00) | (.00) | (.04) | (.00) | (.00) |
| $\gamma_0$ | - | -2.676 | -2.426 | - | -2.650 | -2.675 |
| | | (.12) | (.13) | | (.12) | (.13) |
| $\gamma_1$ | - | 7.836 | 6.836 | - | 7.909 | 6.837 |
| | | (.33) | (.87) | | (.35) | (.85) |
| $\gamma_2$ | - | - | 1.560 | - | - | 1.576 |
| | | | (.18) | | | (.19) |
| *Log-Likelihood* | -4681.54 | -4676.73 | -4659.57 | -4663.71 | -4646.65 | -4633.67 |
| $N$ | 1812 | 1812 | 1812 | 1812 | 1812 | 1812 |
| $\delta_{\#Kids0-2}= 0$ | - | - | - | 9.65(.29) | 36.05(.00)** | 37.31(.00)** |
| $\delta_{\#Kids3-5}= 0$ | - | - | - | 9.37(.31) | 43.80(.00)** | 35.17(.00)** |
| $\delta_{\#Kids6-17}= 0$ | - | - | - | 8.04(.43) | 52.44(.00)** | 34.53(.00)** |
| $\delta_{ymt}= 0$ | - | - | - | 8.22(.22) | 53.84(.00)** | 40.45(.00)** |

Note: All specifications include number of children aged 0-2 years lagged one year, race, maximum years of education over the sample period, a quadratic in age, and unrestricted year effects. Non-labor income is measured by $y_{mp}$ and $y_{mt}$ which denote husband's permanent (sample average) and transitory (deviations from sample average) annual earnings, respectively. $Var\,(\eta_i)$ is the variance of permanent unobserved heterogeneity and the $\gamma$'s are the classification error parameters. $\rho$ is the AR(1) serial correlation coefficient. * indicates significance at the 1% level and ** indicates significance at the 5% level.

Table 5
Random Effects Probit Models of Participation with AR(1) Errors and First-Order State Dependence
(SML Estimates)

| | RE+AR(1)+SD(1) | | | CRE+AR(1)+SD(1) | | |
|---|---|---|---|---|---|---|
| | No CE (1) | No Persistent CE (2) | Persistent CE (3) | No CE (4) | No Persistent CE (5) | Persistent CE (6) |
| $y_{mp}$ | -.272 | -.394 | -.399 | -.285 | -.362 | -.451 |
| | (.05) | (.02) | (.01) | (.05) | (.01) | (.01) |
| $y_{mt}$ | -.140 | -.174 | -.202 | -.140 | -.134 | -.186 |
| | (.04) | (.03) | (.03) | (.04) | (.03) | (.03) |
| $\#Kids0\text{-}2_t$ | -.296 | -.347 | -.343 | -.252 | -.322 | -.420 |
| | (.04) | (.04) | (.03) | (.05) | (.05) | (.05) |
| $\#Kids3\text{-}5_t$ | -.174 | -.114 | -.138 | -.135 | -.158 | -.171 |
| | (.04) | (.03) | (.02) | (.05) | (.03) | (.03) |
| $\#Kids6\text{-}17_t$ | -.048 | -.025 | -.034 | -.054 | -.072 | -.110 |
| | (.02) | (.01) | (.01) | (.04) | (.02) | (.03) |
| $Var(\eta_i)$ | .479 | .732 | .740 | .485 | .781 | .787 |
| | (.04) | (.07) | (.09) | (.04) | (.09) | (.11) |
| $\rho$ | -.219 | .589 | .678 | -.213 | .619 | .649 |
| | (.04) | (.03) | (.02) | (.04) | (.03) | (.03) |
| $h_{t-1}$ | 1.063 | .843 | .769 | 1.042 | .733 | .726 |
| | (.09) | (.03) | (.03) | (.09) | (.03) | (.04) |
| $Corr(u_{i0}, u_{it})$ | .482 | .816 | .820 | .494 | .835 | .853 |
| | (.03) | (.13) | (.21) | (.03) | (.18) | (.21) |
| $\gamma_0$ | - | -2.541 | -2.488 | - | -2.684 | -2.252 |
| | | (.09) | (.09) | | (.09) | (.08) |
| $\gamma_1$ | - | 6.712 | 5.996 | - | 6.842 | 5.427 |
| | | (.21) | (.25) | | (.14) | (.21) |
| $\gamma_2$ | - | - | 1.100 | - | - | 1.335 |
| | | | (.22) | | | (.17) |
| *Log-Likelihood* | -4655.36 | -4650.17 | -4633.00 | -4643.52 | -4609.70 | -4583.94 |
| $N$ | 1812 | 1812 | 1812 | 1812 | 1812 | 1812 |
| $\delta_{\#Kids0-2}=0$ | - | - | - | 3.39(.91) | 39.80(.00)** | 36.91(.00)** |
| $\delta_{\#Kids3-5}=0$ | - | - | - | 3.84(.87) | 35.90(.00)** | 32.25(.00)** |
| $\delta_{\#Kids6-17}=0$ | - | - | - | 3.34(.91) | 32.97(.00)** | 31.19(.00)** |
| $\delta_{ymt}=0$ | - | - | - | 2.92(.82) | 47.70(.00)** | 38.20(.00)** |

Note: All specifications include number of children aged 0-2 years lagged one year, race, maximum years of education over the sample period, a quadratic in age, and unrestricted year effects. Non-labor income is measured by $y_{mp}$ and $y_{mt}$ which denote husband's permanent (sample average) and transitory (deviations from sample average) annual earnings, respectively. $Var(\eta_i)$ is the variance of permanent unobserved heterogeneity and the $\gamma$'s are the classification error parameters. $\rho$ is the AR(1) serial correlation coefficient and $h_{t-1}$ is lagged participation status. $Corr(u_{i0}, u_{it})$ is the error correlation relevant for the Heckman approximate solution to the initial conditions problem. * indicates significance at the 1% level and ** indicates significance at the 5% level.