Implications of Classification Error for the Dynamics of Female Labor Supply

Michael P. Keane[*] and Robert M. Sauer[†]

December 31, 2007

Abstract

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.

Keywords: Female Labor Supply, Fertility, Discrete Choice, Classification Error, Simulated Maximum Likelihood

JEL Codes: J2,J6,C3,D1

---

[*]ARC Federation Fellow, University of Technology Sydney and Research Professor, Arizona State University (michael.keane@uts.edu.au)

[†]University of Southampton (rms@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)).[1]

Until recently, the consensus of the literature was that unobserved heterogeneity is an important source of persistence, 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.[2] But a recent Econometrica paper by Hyslop (1999) challenged these conclusions. Using recursive importance sampling techniques (see Keane (1994)) he was able to estimate a complex dynamic panel probit model of married women's labor supply that included a rich pattern of unobserved heterogeneity and true state dependence, as well as autoregressive (AR) errors.

Hyslop found that the equicorrelation assumption of the random effects model was soundly rejected. Allowing for autoregressive errors (the computationally diffi-

[1]Labor market and social policies can have very different effects depending on whether persistence is due to unobserved heterogeneity – i.e., relatively immutable differences across individuals in tastes for work, motivation, productivity, etc. – or due to state dependence – i.e., habit persistence, human capital accumulation, barriers to labor market entry (e.g., costs of job search), etc.. Also, whether fertility and nonlabor income may reasonably be treated as exogenous has important implications for the proper specification of labor supply functions and estimation of labor supply elasticities.

[2]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.

cult 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, rather surprisingly, he found that he could not reject that fertility and nonlabor income are exogenous to female labor supply decisions.

Here, we examine the sensitivity of Hyslop's results to a further elaboration of the panel probit model - the introduction of classification error in the dependent variable. As prior work has shown, mis-classification of employment status is important in micro data sets. Perhaps the best known evidence is provided by Poterba and Summers (1986). They find that, in the CPS, the probability an employed person falsely reports being unemployed or out-of-the-labor-force is 1.5%, while the probability an unemployed person falsely reports being employed is 4.0%.[3] Might Hyslop's results be sensitive to allowing for such mis-classification of employment status?

To address this issue, we nest Hyslop's (1999) dynamic panel probit model of married women's labor market participation decisions within a model of classification error in reported employment status. We first replicate Hyslop (1999)'s results, using PSID data on married women's labor market decisions between 1981 and 1987. We then show that inferences regarding exogeneity of fertility/non-labor income are indeed sensitive to classification error: allowing for mis-classification leads us to strongly reject the exogeneity hypothesis.

The intuition for the change in results is simple: If the data contain classification error, persistence in employment status is understated, and so is the importance of permanent unobserved heterogeneity. Allowing for classification error leads one to infer more persistence in "true" participation status, making unobserved heterogeneity

[3]These figures are derived from Poterba-Summers Table II. To obtain their results Poterba-Summers use the "CPS reconciliation data." In the reconciliation data, Census sends an interviewer to reinterview a household a week after its original interview. The interviewer determines if reports disagree and, in the event of a disagreement, attempts to determine true employment status.

more important. This increases the apparent magnitude of the covariance between individual effects and fertility/non-labor income as well.

Of course, any interpretation of data is conditional on ones assumptions, but to the extent one believes classification error in reported employment status is important in panel data, our results should move ones priors towards accepting the endogeneity of fertility and non-labor income. Thus, our results provide additional motivation for the importance of jointly modelling female labor supply and fertility, as in, e.g., Moffitt (1984), Hotz and Miller (1988), Keane and Wolpin (2006).

Our extension of Hyslop's model introduces a serious computational problem: with classification error, lagged true choices (and the true state of the agent more generally) become unobserved, making simulation of state contingent transition probabilities intractable. This makes the GHK approach to simulating the likelihood infeasible, as it relies on simulating transition probabilities (see Keane (1994)). Instead, following Keane and Wolpin (2001) and Keane and Sauer (2005) we show how to simulate the likelihood using only unconditional simulations.[4] As they describe, this is actually made possible by assuming classification error. However, as our paper is meant to be substantive rather than methodological, we refer the reader to those papers for details of the econometric methods.

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. Section 3 describes the PSID data used in the estimation. Section 4 presents the estimation results, while section 5 concludes.

[4] To the best of our knowledge, the few prior papers that have explicitly treated classification error in discrete choice models of labor supply consider 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 classification error 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 probit model.

2 A Panel Probit Model with Classification Error

We begin in Section 2.1 by presenting a model of married women's labor force participation *exactly* like that estimated by Hyslop (1999). Then, in Section 2.2, we extend it to allow for classification error.

2.1 The Basic Panel Probit Model - Hyslop (1999)

Consider the following specification for the 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., earnings of the husband), number of children in different age ranges, age, race, 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 (a key hypothesis which we will test).

In the simple static probit formulation, the coefficient γ is set to zero and u_{it} is assumed to be serially independent and normally distributed with zero mean and variance σ_u^2. The scale normalization is achieved by setting σ_u^2 equal to one.

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

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

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

Although α_i in (2) is usually assumed to be conditionally independent of X_{it}, it is possible to allow α_i to be correlated with Z_{it}, a vector that contains only the time varying elements of X_{it}.[5] This yields a correlated random effects model (CRE). The correlated random effects probit assumes that the individual effect takes the form,

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

where $\eta_i \sim N(0, \sigma_\eta^2)$ and is conditionally independent of Z_{it} and X_{it}. This implies that $\sigma_\eta^2 = Var(\alpha_i | Z_i)$, where $Z_i = (Z_{i0}, ..., Z_{iT})$, and that the variance of permanent unobserved heterogeneity is now $\sigma_\alpha^2 = Var\left(\sum_{t=0}^{T} Z_{it}'\delta_t\right) + \sigma_\eta^2$. In the CRE model, the δ_t's are estimated in addition to σ_η^2 and β. Thus, exogeneity of children and nonlabor income can be examined via hypothesis tests on δ_t.[6]

In order to see more clearly how the CRE model relaxes exogeneity, note that the conventional dynamic probit model assumes

$$P(h_{it} = 1 | X_i, h_{i,t-1}, \alpha_i) = P(h_{it} = 1 | X_{it}, h_{i,t-1}, \alpha_i) \tag{3a}$$

$$E(\alpha_i | X_{i1}, ..., X_{iT}) = E(\alpha_i) = 0. \tag{3b}$$

Together, these equations imply that, conditional on $(h_{i,t-1}, \alpha_i)$, only X_{it} helps predict h_{it} - i.e., leads and lags of X do not matter. Equation $(3a)$ is equivalent to $E(\varepsilon_{it} | X_{is}) = 0$ for all t and s, a type of strict exogeneity assumption we'll call SE-A. Together, $(3a)$ and $(3b)$ imply $E(u_{it} | X_{is}) = 0$ for all t and s, a stronger form of strict exogeneity we'll call SE-B. By dropping assumption $(3b)$, the CRE model relaxes SE-B while

[5] Letting a time invariant element of X_{it} shift α_i is equivalent to letting it shift $X_{it}'\beta$ by a constant.

[6] 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).

continuing to maintain SE-A.[7]

Next we relax the assumption that ε_{it} is serially uncorrelated. Such serial correlation could arise from persistence in shocks to tastes and/or productivity. Allowing ε_{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 σ_v^2, and conditionally independent of ε_{it-1}. We assume the process is stationary, so $\sigma_\varepsilon^2 = \frac{\sigma_v^2}{(1-\rho^2)}$.[8]

The scale normalization and independence assumption gives $\sigma_u^2 = \sigma_\eta^2 + \sigma_\varepsilon^2 = 1$, and variance stationarity in the $AR(1)$ process gives $\sigma_u^2 = \sigma_\eta^2 + \frac{\sigma_v^2}{(1-\rho^2)} = 1$. Thus, we can estimate ρ and σ_η^2, and "back out" σ_v^2 using the formula $\sigma_v^2 = (1-\rho^2)\left(1-\sigma_\eta^2\right)$.

Finally, in addition to estimating ρ and σ_η^2, we can allow for "true state dependence" by letting γ in (1) be non-zero. Thus, we decompose the persistence in observed choice behavior into that due to (i) permanent unobserved heterogeneity, (ii) first-order state dependence, and (iii) $AR(1)$ serial correlation.[9]

In dynamic probit models of the type specified in equations (1) through (4), it is

---

[7] Intuitively, the CRE model allows the unobserved individual effects α_i, which may capture tastes for work and/or latent skill endowments, to be correlated with fertility and non-labor income (in all periods). But it still maintains that current shocks to employment status ε_{it}, which may arise from transitory shocks to tastes and/or productivity, do not alter future fertility or non-labor income.

[8] The stationarity assumption may be controversial. We assume stationarity because Hyslop (1999) did so, and we want our results to differ from his only due to inclusion of classification error.

[9] As discussed by Wooldridge (2005), what distinguishes true state dependence (γ) from serial correlation (due either to random effects or an $AR(1)$ error component) in (1) is whether or not there is a causal effect of lagged X's on current choices. If the observed persistence in choices is generated entirely by serially correlated errors (i.e., $\gamma = 0$), then lagged X_{it}'s do not help to predict the current choice, conditional on the current X_{it}. Of course, this assertion rules out any *direct* effect of lagged X on the current choice. More generally, it is well known one cannot disentangle true state dependence from serial correlation without some parametric assumptions (see Chamberlain (1984) for discussion).

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. Heckman (1981) proposed an approximate solution to this initial conditions problem that proceeds as follows:[10] Assume that:

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

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 1 (to normalize for scale). ρ_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 ρ_t's are equal implies that only one correlation coefficient, denoted by ρ_0, needs to be estimated. It can easily be derived that ρ_0 is also the covariance between u_{i0} and the individual effect α_i. (The working paper version of the paper, Keane and Sauer (2006), provides a derivation).

2.2 Incorporating Classification Error

We generalize the dynamic probit framework in equations (1) through (5) 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. We let π_{jk} denote the probability that a true j is recorded as a k, where $j, k = 0, 1$, and assume these classification rates are determined by a logit model with the index function,

$$l_{it} = \gamma_0 + \gamma_1 h_{it} + \gamma_2 h_{it-1}^* + \omega_{it} \tag{6}$$

where $l_{it} > 0$ implies $h_{it}^* = 1$, while $h_{it}^* = 0$ otherwise.

[10] Again, we choose this method for comparability with Hyslop (1999). See Heckman (1981) and Wooldridge (2005) for details on various alternative solutions.

Naturally, we allow h_{it}^* to be a function of h_{it}, as the probability of a reported "1" should be greater when the person is actually employed.[11] We also include the lagged reported choice $h_{i,t-1}^*$ to accommodate persistence in misreporting. The error term ω_{it} is distributed logistically and independent of u_{it}, conditional on h_{it}, and $h_{i,t-1}^*$.[12]

Putting equations (1) through (5) and (6) together, we arrive at the following dynamic panel data probit model of female labor force participation decisions with classification error in reported choices,

$$\begin{aligned} h_{it} &= 1\left(X_{it}'\beta + \gamma h_{it-1} + u_{it} > 0\right) \\ u_{it} &= \alpha_i + \varepsilon_{it} \\ \alpha_i &= \sum_{t=0}^{T} Z_{it}'\delta_t + \eta_i \\ \varepsilon_{it} &= \rho\varepsilon_{it-1} + v_{it} \\ h_{i0} &= 1\left(X_{i0}'\beta_0 + u_{i0} > 0\right) \\ \rho_0 &= corr\left(u_{i0}, u_{it}\right) \\ l_{it} &= \gamma_0 + \gamma_1 h_{it} + \gamma_2 h_{it-1}^* + \omega_{it}, \end{aligned} \tag{7}$$

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

[11] Indeed, Hausman, Abrevaya and Scott-Morton (1998) (HAS) note that the key condition for identification of measurement error rates in parametric discrete choice models is that the probability of a reported "1" be increasing in the probability of a true "1" (and similarly for "0"). In our notation this requires that $\pi_{01} + \pi_{10} < 1$. This condition, which in our model is equivalent to $\gamma_1 > 0$, means classification error can't be so severe that people mis-report their state more often than not, which is certainly a mild requirement.

[12] The classification error specification in (6) has been shown to perform quite well in repeated sampling experiments on dynamic probit models using our estimation procedure (see Keane and Sauer (2005)). The parameters of the misclassification process can be recovered with precision, along with the parameters of the "true" process (1) – (5).

3 Data

We use the same data as Hyslop (1999), who graciously provided us with his data set. While in some instance we might have made different decisions in constructing the sample or defining the covariates, it is essential that the data be identical to facilitate replication. The data are from the 1986 Panel Study of Income Dynamics (PSID), including both the random Census subsample of families and nonrandom Survey of Economic Opportunities. The sample period covers the seven 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. This gives a sample size of $N = 1812$ women and $12,684$ person/year observations. A woman is classified as a labor force participant if she reports positive annual hours worked and positive annual earnings.

Table 1 presents selected means and standard errors in the estimation sample. The average labor market participation rate over the whole sample is .70. The additional variables displayed in the table, which are used as covariates to predict participation, are a woman's nonlabor income, her number of children in three different age ranges (0-2, 3-5 and 6-17), and her age, education and race (equal to one if black).[13]

Nonlabor income for each woman i in the sample is proxied by her husband's earnings in year t (y_{it}) (in 1987$). As in Hyslop (1999), the log of husband's average earnings over the sample period $y_{mp} = \ln(\frac{1}{T}\sum_t y_{it})$ is used as a proxy for permanent nonlabor income. 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 number of children aged 0-2 years lagged one year also appears as a covariate (see Hyslop (1999) for discussion).

The degree of persistence in participation is very strong. The probability of par-

[13] There is substantial over-time variation in numbers of children and transitory nonlabor income. The over-time standard deviations of the three fertility variables (in ascending age order) and transitory nonlabor income are .159, .182, .375, and .149, respectively. Significant variation in these variables is important for the CRE estimator.

ticipation at t given participation at $t-1$ is 91%, while the probability of nonparticipation given nonparticipation at $t-1$ is 78%. The is also 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 larger than the transition rate from participation at $t-2$ *and* nonparticipation at $t-1$ to participation at t (.403). This implies that the model is not simply RE, but that there is also some type of short run persistence (e.g., AR(1) errors and/or state dependence).

Transition patterns are critical for identifying the relative importance of random effects, AR(1) errors and first-order state-dependence. But, if some 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 the CRE model.

4 Estimation Results

Tables 2-4 present selected SML estimates of the correlated random effects (CRE) model in (7). 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.

4.1 Basic CRE Model

Column (1) of Table 2 reports estimates of the CRE model with no $AR(1)$ serial correlation, no first-order state dependence ($SD(1)$) and no correction for classification error (No CE). The estimates were obtained by Hyslop (1999) using the SML-GHK algorithm.[14] The parameter estimates are as expected: the negative effect of "per-

[14] Note that these RE models could have been estimated 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 columns 3-4 don't arise due to simulation per se.

manent" nonlabor income on labor market participation is stronger than that of transitory nonlabor income, and younger children have a larger negative effect on participation than older children. The estimate of $Var(\eta_i)$ implies that 80.4% of the overall error variance is due to permanent unobserved heterogeneity.[15]

The bottom four rows of the table report likelihood ratio tests for exogeneity of children in the three age ranges (0-2, 3-5 and 6-17) and nonlabor income (i.e., tests of H(0): $\delta_t = 0$ for $t = 0, .., T$). The p-values indicate the null hypothesis that children and nonlabor income are exogenous is clearly rejected.

Column (2) presents estimates of the exact same model except that we use our SML algorithm, based on unconditional simulation, instead of SML-GHK.[16] We also fix the level of classification error to near zero, i.e., $\pi_{01} = \pi_{10} = .0025$. The purpose of this exercise is to verify that any difference between our results and those of Hyslop (1999) that we may find later is due to introduction of classification error, not due to use of a different simulation method. Comparing Columns (1) and (2), we see that our results and those of Hyslop are essentially identical - the alternative estimation method makes almost no difference.

Next, we introduce classification error. In column (3), we present estimates of the CRE model assuming no persistence in misclassification (No Persistent CE) - i.e., $\gamma_2 = 0$ in (7). In column (4) we estimate the general model that allows persistence. Allowing for classification error (of either type) does not produce substantial changes in the coefficients of the covariates. It does, however, substantially increase the estimate of the fraction of the overall error variance due to permanent unobserved heterogeneity from about 80% to about 94%. This large increase in the importance of permanent unobserved heterogeneity suggests that misclassification exaggerates the frequency of transitions between labor market states. Given the increased importance

[15] The proportion of the overall error variance σ_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, $\sigma_u^2 = 1$.

[16] Hyslop used 40 draws to implement GHK, while we use M=1500 simulated choice histories.

of the random effects, it is not surprising that the chi-squared test statistics for the hypotheses of exogenous fertility and nonlabor income increase substantially, leading to even stronger rejections of the null hypothesis of exogeneity.[17]

The estimates of γ_0 and γ_1 in column (3) can be used to calculate the classification error rates implied by the model. The probability of reporting participation, when the true state is nonparticipation ($\widehat{\pi}_{01}$) is .081. The probability of reporting nonparticipation, when the true state is participation ($\widehat{\pi}_{10}$) is .010. These classification error rates can be compared to the analogous rates of 4% and 1.5% obtained by Poterba and Summers for the CPS. The overall (i.e., unconditional) error rate implied by our model is only 1.8%. Thus, we see that even a fairly "small" amount of classification error can lead to a substantial attenuation bias in the importance of permanent unobserved heterogeneity.

Comparing the log-likelihoods in Columns (2) and (3) by a likelihood ratio test produces a chi-squared test statistic, with two degrees of freedom, equal to 19.96 and a p-value of .000. Thus, introducing classification error leads to a significant improvement in fit.

The estimate γ_1 in column (4) implies there is considerable persistence in misreporting. However, we will reserve further discussion of this point until we get to the models with AR(1) errors. The reason is that, as we shall see, in models with only random effects the parameter γ_1 tends to "sop up" omitted serial correlation in the time varying error component.

4.2 CRE with AR(1) Errors

Table 3 reports estimates of the same sequence of models as in Table 2, except that we now allow for $AR(1)$ serial correlation in the transitory error. Columns (1) and

[17] The increased σ_α^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 larger (recall that $\sigma_\alpha^2 = Var(\sum_{t=0,T} Z'_{it}\delta_t) + \sigma_\eta^2$).

(2) reproduce the rather dramatic finding from Hyslop (1999). Specifically, at the bottom of the table we see that with the introduction of $AR(1)$ errors we can no longer reject the null hypothesis that fertility and non-labor income are exogenous at any conventional level of significance.

The introduction of $AR(1)$ errors has a very modest impact on the estimated coefficients of nonlabor income and fertility. But, the importance of the individual effect is considerably reduced - i.e., it drops from 80% of the overall error variance in column (1) of Table 2 to 55% in column (1) of Table 3. The estimated $AR(1)$ coefficient ($\widehat{\rho}$) is .696 and is precisely estimated. Relaxing the restriction that $\rho = 0$ results in a 225 point improvement in the log-likelihood (compare columns (1) in Tables 2 and 3). Thus, $AR(1)$ serial correlation appears to be an important component of the persistence in reported labor market states. This replicates Hyslop's other main result: that the equicorrelation assumption is soundly rejected.

Table 3 columns (3)-(4) introduce classification error. Here we see our main result. When classification is introduced, the fraction of variance accounted for by random effects increases from about 55% to 83%. And the hypotheses of exogenous fertility and non-labor income are soundly rejected. This is true regardless of whether we allow for persistence in classification error.

Note that this change in the exogeneity test results 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 easier to detect (and more important as a determinant of labor supply behavior).

The $AR(1)$ parameter $\widehat{\rho}$ also increases (slightly) when we introduce classification error, from .70 in column (1) to .75 in columns (3)-(4). Thus, Hyslop's other main finding - that equicorrelation is strongly rejected - is still supported when we add classification error.[18]

[18]Note also that introduction of $AR(1)$ errors into (either) classification error model reduces the

The estimates of γ_0 and γ_1 in Column (3) imply that $\widehat{\pi}_{01}$ is .066 and $\widehat{\pi}_{10}$ is .005. These error rates are again comparable to the figures of 4% and 1.5% obtained by Poterba and Summers (1986). The overall (i.e., unconditional) rate of misclassification implied by our model is 1.3%. A likelihood ratio test for the joint significance of γ_0 and γ_1 produces a χ^2 statistic with two degrees of freedom equal to 31.8, implying a p-value of .000.

Column (4) presents estimates allowing for persistence in misclassification (Persistent CE). The estimate of γ_2 implies substantial persistence. For instance, the probability of reporting participation, when the true state is nonparticipation *and nonparticipation is reported in the previous period*, is .064. But if participation was reported in the previous period, this error rate rises to .250. Similarly, the probability of reporting nonparticipation, when the true state is participation *and participation is reported in the previous period*, is only .003. But if nonparticipation was reported in the previous period, this error rate rises to .015. The substantial increases in the probability of reporting the wrong labor market state, when that same state was reported in the previous period, suggest that persistent misclassification may be an important source of recorded persistence in female labor force participation data.

Comparing Column (4) of Tables 2 and 3, we see that introduction of the $AR(1)$ error component leads to a drop in the estimated persistence in misclassification. $\widehat{\gamma}_2$ falls from 2.61 in column (4) of Table 2 to 1.58 in column (4) of Table 3.

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 column (4), relaxing the restriction that $\gamma_2 = 0$ results in a relatively large improvement in the log-likelihood of 13 points. Note, however, that this is much smaller than the 206 point improvement we saw in column (4) of Table 2 when an

---

fraction of variance due to heterogeneity from about 94% to 83% (compare columns (3)-(4) in Table 2 to columns (3)-(4) in Table 3).

$AR(1)$ 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.[19]

4.3 CRE with AR(1) Errors and SD(1)

Table 4 reports results for more general CRE models which allow for both $AR(1)$ serial correlation and first order state dependence ($SD(1)$). As in Hyslop (1999), 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 model without classification error from Hyslop (1999). The coefficient on lagged participation is a strong 1.042 and is precisely estimated. The inclusion of lagged participation in the model improves the log-likelihood by 20 points over the $CRE{+}AR(1)$ model, and reduces the variance of the individual effect from 55% to 49%. The estimate of the $AR(1)$ serial correlation coefficient $\widehat{\rho}$ falls dramatically from .696 to $-.213$.

Despite these differences, Hyslop's main result from Table (3) remains unchanged: In the $CRE{+}AR(1){+}SD(1)$ model, the hypothesis of exogeneity of fertility and nonlabor income cannot be rejected at any conventional level of significance.

Columns (3)-(4) report estimates of models that include classification error. These models produce substantial improvements in the log-likelihood: 32 points with no persistence, and an additional 26 points when persistence in classification error is allowed. They also produce very different estimates of the importance of random effects, AR(1) errors and state dependence. In each case, the first order state dependence coefficient falls to about .73 (as opposed to 1.04 in column (1)), and the fraction of the error variance due to random effects increases to .78 (as opposed to .49). The $AR(1)$ coefficient is about .62 to .65 (as opposed to $-.21$).

[19] The intuition for how the parameters ρ and γ_2 are distinguished is similar to that for how serial correlation and state dependence are distinguished. Specially, if $\gamma_2 > 0$ it implies that lagged X's help to predict current choices, while $\rho > 0$ does not have this implication.

Thus, failure to account for classification error produces substantial attenuation biases in the importance of unobserved heterogeneity and $AR(1)$ serial correlation, and an upward bias in extent of first order state dependence.[20] 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 ($\widehat{\pi}_{01} = .064$ and $\widehat{\pi}_{10} = .015$) are similar in magnitude to those obtained in earlier specifications and remain statistically significant. They are also quite close to the analogous rates calculated by Poterba and Summers for the CPS (i.e., 4% and 1.5%, respectively). The overall error rate implied by our model is 1.8%. Also, the estimated degree of persistence in misclassification is only slightly smaller than in the RE+AR(1) model (compare γ_2 in Column (4) of Tables 3 and 4).[21]

Finally, the classification error models in columns (3)-(4) again reject overwhelmingly the hypotheses of exogenous fertility and non-labor income. The difference in results from column (4) is again a direct result of the greater estimated variance of the random effect in models that accommodate classification error.

[20] 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.

[21] The intuition for how one can distinguish true state dependence $\gamma > 0$ from persistence in misclassification $\gamma_2 > 0$ is as follows. If there is persistence in classification error but *no* true state dependence, we should have:

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

However, in a first-order Markov model, the lagged state is only a sufficient statistic for lagged inputs if it is measured *without* error. Thus, if true state dependence is also present (in addition to persistent misreporting) then lagged X's will help to predict current current choices even conditional on the lagged (measured) choice.

5 Conclusion

Estimating the relative importance of state dependence and permanent unobserved heterogeneity, and the influence of children and nonlabor income, have long been important topics in the literature on female labor supply. Hyslop (1999) contributed to this literature by estimating dynamic probit models of married women's labor force participation decisions, using PSID data from 1979 to 1985. His innovation was to relax the equicorrelation assumption of the common random effects model by allowing for an AR(1) error component. He obtained two main findings: (i) the AR(1) error component is important, and when it is included the importance of random effects is substantially reduced, and (ii) once the AR(1) error component is included the hypothesis that fertility and husband's income are exogenous - in the sense of being uncorrelated with the random effects - cannot be rejected.

We extend Hyslop's model by nesting it within a model of classification error in reported employment status. Our estimates imply that the extent of classification error in the data is rather modest - i.e., employment status is misclassified in about 1.3% to 1.8% of cases on average. The extent of classification error that we estimate for the PSID is in the ballpark of estimates obtained by Poterba and Summers for employment status in the CPS, which gives face validity to our results.

Crucially, we find that even these modest levels of classification error (i.e., 1.3% to 1.8%) are sufficient to cause models that fail to account for it to substantially understate the important of individual random effects. This is obviously due to the spurious transitions created by mis-classification. After correcting for classification error, we obtain a large increase in the estimated variance of the random effects. As a result, correlation between the random effects and fertility/nonlabor income becomes easier to detect, and we soundly reject the hypothesis that fertility and nonlabor income are exogenous. This is in sharp contrast to main result (ii) in Hyslop (1999). Our results suggest that researchers estimating dynamic discrete choice models should be careful to consider the possible impact of misclassification on their results.

References

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

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

[3] 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.

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

[5] Heckman, J. 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.

[6] 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.

[7] 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.

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

[9] 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.

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

[11] 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 available at SSRN: http://ssrn.com/abstract=448240.

[12] Keane, M.P., and R.M.Sauer (2006), "Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior," IZA Discussion Paper No. 2332.

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

[14] 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.

[15] 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," PIER Working Paper No. 06-004.

[16] 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.

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

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

[19] 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.

[20] 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.

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

[22] Poterba J.M., and L.H. Summers (1986), "Reporting Errors and Labor Market Dynamics," *Econometrica*, 54, 1319-1338.

[23] 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.

[24] 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.

[25] 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).

Table 2
Correlated Random Effects Probit Models of Participation
(SML Estimates)

	Hyslop	Keane and Sauer		
	No CE	No CE	No Persistent CE	Persistent CE
	(1)	(2)	(3)	(4)
y_{mp}	-.341	-.336	-.400	-.375
	(.05)	(.05)	(.04)	(.04)
y_{mt}	-.099	-.103	-.127	-.172
	(.03)	(.03)	(.02)	(.03)
$\#Kids0\text{-}2_t$	-.300	-.305	-.290	-.388
	(.03)	(.03)	(.04)	(.05)
$\#Kids3\text{-}5_t$	-.247	-.245	-.265	-.271
	(.03)	(.03)	(.03)	(.04)
$\#Kids6\text{-}17_t$	-.084	-.083	-.090	-.087
	(.03)	(.03)	(.02)	(.03)
$Var\ (\eta_i)$	.804	.829	.938	.943
	(.02)	(.04)	(.07)	(.10)
γ_0	-	-	-2.427	-2.386
			(.09)	(.11)
γ_1	-	-	6.996	5.056
			(.21)	(.19)
γ_2	-	-	-	2.611
				(.11)
Log-Likelihood	-4888.38	-4887.75	-4878.27	-4672.62
N	1812	1812	1812	1812
$\delta_{\#Kids0-2}= 0$	32.36(.00)**	35.31(.00)**	52.14(.00)**	57.34(.00)**
$\delta_{\#Kids3-5}= 0$	12.77(.12)	13.02(.11)	49.04(.00)**	61.04(.00)**
$\delta_{\#Kids6-17}= 0$	21.74(.01)**	23.01(.00)**	49.50(.00)**	61.19(.00)**
$\delta_{ymt}= 0$	48.50(.00)**	48.71(.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 γ's are the classification error parameters. * indicates significance at the 1% level and ** indicates significance at the 5% level.

Table 3
Correlated Random Effects Probit Models of Participation with AR(1) Errors
(SML Estimates)

	Hyslop	Keane and Sauer		
	No CE	No CE	No Persistent CE	Persistent CE
	(1)	(2)	(3)	(4)
y_{mp}	-.332	-.327	-.345	-.345
	(.05)	(.04)	(.00)	(.00)
y_{mt}	-.097	-.108	-.112	-.085
	(.03)	(.03)	(.01)	(.01)
$\#Kids0\text{-}2_t$	-.272	-.251	-.306	-.307
	(.03)	(.03)	(.02)	(.02)
$\#Kids3\text{-}5_t$	-.234	-.219	-.265	-.269
	(.03)	(.02)	(.01)	(.01)
$\#Kids6\text{-}17_t$	-.077	-.083	-.079	.083
	(.02)	(.02)	(.01)	(.01)
$Var\,(\eta_i)$	.546	.582	.830	.831
	(.04)	(.03)	(.03)	(.04)
ρ	.696	.710	.746	.748
	(.04)	(.05)	(.00)	(.00)
γ_0	-	-	-2.650	-2.675
			(.12)	(.13)
γ_1	-	-	7.909	6.837
			(.35)	(.85)
γ_2	-	-	-	1.576
				(.19)
Log-Likelihood	-4663.71	-4662.55	-4646.65	-4633.67
N	1812	1812	1812	1812
$\delta_{\#Kids0-2}= 0$	9.65(.29)	10.27(.25)	36.05(.00)**	37.31(.00)**
$\delta_{\#Kids3-5}= 0$	9.37(.31)	10.39(.24)	43.80(.00)**	35.17(.00)**
$\delta_{\#Kids6-17}= 0$	8.04(.43)	9.44(.31)	52.44(.00)**	34.53(.00)**
$\delta_{ymt}= 0$	8.22(.22)	8.91(.18)	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 γ's are the classification error parameters. ρ is the AR(1) serial correlation coefficient. * indicates significance at the 1% level and ** indicates significance at the 5% level.

Table 4
Correlated Random Effects Probit Models of Participation with AR(1) Errors and First-Order State Dependence
(SML Estimates)

	Hyslop	Keane and Sauer		
	No CE	No CE	No Persistent CE	Persistent CE
	(1)	(2)	(3)	(4)
y_{mp}	-.285	-.291	-.362	-.451
	(.05)	(.05)	(.01)	(.01)
y_{mt}	-.140	-.137	-.134	-.186
	(.04)	(.05)	(.03)	(.03)
$\#Kids0\text{-}2_t$	-.252	-.254	-.322	-.420
	(.05)	(.05)	(.05)	(.05)
$\#Kids3\text{-}5_t$	-.135	-.131	-.158	-.171
	(.05)	(.04)	(.03)	(.03)
$\#Kids6\text{-}17_t$	-.054	-.053	-.072	-.110
	(.04)	(.04)	(.02)	(.03)
$Var\,(\eta_i)$	.485	.519	.781	.787
	(.04)	(.06)	(.09)	(.11)
ρ	-.213	-.141	.619	.649
	(.04)	(.03)	(.03)	(.03)
h_{t-1}	1.042	1.031	.733	.726
	(.09)	(.07)	(.03)	(.04)
$Corr\,(u_{i0}, u_{it})$	.494	.561	.835	.853
	(.03)	(.09)	(.18)	(.21)
γ_0	-	-	-2.684	-2.252
			(.09)	(.08)
γ_1	-	-	6.842	5.427
			(.14)	(.21)
γ_2	-	-	-	1.335
				(.17)
Log-Likelihood	-4643.52	-4641.62	-4609.70	-4583.94
N	1812	1812	1812	1812
$\delta_{\#Kids0-2}= 0$	3.39(.91)	6.02(.65)	39.80(.00)**	36.91(.00)**
$\delta_{\#Kids3-5}= 0$	3.84(.87)	6.78(.56)	35.90(.00)**	32.25(.00)**
$\delta_{\#Kids6-17}= 0$	3.34(.91)	6.89(.55)	32.97(.00)**	31.19(.00)**
$\delta_{ymt}= 0$	2.92(.82)	5.92(.43)	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 γ's are the classification error parameters. ρ 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.