

The copyright to this article is held by the Econometric Society, <http://www.econometricsociety.org/>. It may be downloaded, printed and reproduced only for personal or classroom use. Absolutely no downloading or copying may be done for, or on behalf of, any for-profit commercial firm or for other commercial purpose without the explicit permission of the Econometric Society. For this purpose, contact the Editorial Office of the Econometric Society at econometrica@econometricsociety.org

CLASSIFICATION ERROR IN DYNAMIC DISCRETE CHOICE MODELS: IMPLICATIONS FOR FEMALE LABOR SUPPLY BEHAVIOR

BY MICHAEL P. KEANE AND ROBERT M. SAUER¹

Two key issues in the literature on female labor supply are (i) whether persistence in employment status is due to unobserved heterogeneity or state dependence, and (ii) whether fertility is exogenous to labor supply. Until recently, the consensus was that unobserved heterogeneity is very important and fertility is endogenous. 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.

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)).² The second is whether children and nonlabor income can reasonably be viewed as exogenous to labor supply (see, e.g., Chamberlain (1984), Rosenzweig and Schultz (1985), Mroz (1987), and Jakubson (1988)).

Until recently, the consensus of the literature was that unobserved heterogeneity is an important source of persistence and that fertility is endogenous, that is, women with greater unobserved tastes for work or greater unobserved skill endowments tend to have fewer children.³ Hyslop (1999) challenged these conclusions. Using recursive importance sampling techniques

¹This research is supported in part by the Australian Research Council, through a grant to Michael Keane (ARC Grants FF0561843 and DP0774247), and the Economic and Social Research Council of the United Kingdom, through a grant to Robert Sauer (ESRC Grant RES-000-22-1529).

²Labor market 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, etc.).

³For instance, Chamberlain (1984) estimated probit models for married women's labor force participation, and Jakubson (1988) estimated panel tobit models for married women's hours, and they both overwhelmingly rejected exogeneity of children.

(see Keane (1994)), he estimated a complex panel probit model of married women's labor supply that included a rich pattern of unobserved heterogeneity, true state dependence, and autoregressive errors.

Hyslop found that the equicorrelation assumption of the random effects model was soundly rejected. Allowing for autoregressive errors led to a substantial diminution in the apparent importance of permanent unobserved heterogeneity. This, in turn, reduced the importance of correlation between unobserved heterogeneity and children/nonlabor income for labor supply behavior. Hence, rather surprisingly, he could not reject that fertility and nonlabor income are exogenous to female labor supply.

Here, we examine the potential sensitivity of Hyslop's results to classification error in employment status. Prior work has shown that misclassification of work status is important in micro data sets. Perhaps the best known evidence was provided by Poterba and Summers (1986). In the Current Population Survey (CPS), they found the probability an employed person falsely reports being unemployed or out of the labor force is 1.5%, while the probability an unemployed person reports being employed is 4.0%.⁴ Might Hyslop's results be sensitive to allowing for such misclassification of employment status?

To address this issue, we nest Hyslop's (1999) panel probit model of married women's work decisions within a model of classification error in reported employment status. We first replicate Hyslop's (1999) results, using the Panel Study of Income Dynamics (PSID) data on married women's work decisions between 1981 and 1987. We then show that inferences regarding exogeneity of fertility/nonlabor income are indeed sensitive to classification error: allowing for misclassification 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" employment status, making unobserved heterogeneity more important. This increases the apparent magnitude of the covariance between individual effects and fertility/nonlabor income as well.

Thus, to the extent one believes classification error in reported employment status is important in panel data, our results should move one's priors toward accepting the endogeneity of fertility and nonlabor income. This, in turn, provides additional motivation for the importance of jointly modelling female labor supply and fertility, as in, for example, Moffitt (1984), Hotz and Miller (1988), and Keane and Wolpin (2006).

⁴These figures are derived from Poterba and Summers' (1986) Table II. To obtain their results Poterba and 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.

Introducing classification error in the panel probit model creates a serious computational problem: lagged true choices are 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 simulate the likelihood using unconditional simulations. As our focus is on substantive results, we refer the reader to those papers for the econometric methods.

The rest of the paper is organized as follows: Section 2 presents our panel probit model with classification error. Section 3 describes the PSID data used in the estimation. Section 4 presents the estimation results, while Section 5 concludes. Simulations are available in the online supplement (Keane and Sauer (2009))

2. A PANEL PROBIT MODEL WITH CLASSIFICATION ERROR

In Section 2.1 we present a model of married women's labor supply decisions *exactly* like that in Hyslop (1999). 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 work decision rule:

$$(1) \quad h_{it} = 1(X'_{it}\beta + \gamma h_{it-1} + u_{it} > 0) \quad (i = 1, \dots, N, t = 0, \dots, T),$$

where h_{it} indicates whether woman i works at time t . She chooses to work if and only if (iff) the expression in parentheses is true. X_{it} is a vector of covariates including nonlabor income, number of children in different age ranges, age, race, education, and time dummies. h_{it-1} is lagged employment status and u_{it} is an error term. The decision rule is in "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 to be exogenous (a key hypothesis which we will test).

In the simple static probit model, 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 1.

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

$$(2) \quad u_{it} = \alpha_i + \varepsilon_{it},$$

where α_i is a time-invariant individual effect distributed $N(0, \sigma_\alpha^2)$. This individual effect, which captures unobserved time-invariant taste and productivity characteristics of woman i , generates serial correlation in u_{it} . The transitory error component, ε_{it} , is serially uncorrelated, conditionally independent of α_i ,

and distributed $N(0, \sigma_\varepsilon^2)$. Because $\sigma_u^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2$ and we normalize $\sigma_u^2 = 1$, only σ_α^2 is directly estimated.

Although α_i in (2) is usually assumed to be conditionally independent of X_{it} , one can 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), where

$$(3) \quad \alpha_i = \sum_{t=0}^T Z'_{it} \delta_t + \eta_i.$$

Here $\eta_i \sim N(0, \sigma_\eta^2)$ and is conditionally independent of Z_{it} and X_{it} . The variance of permanent heterogeneity is now $\sigma_\alpha^2 = \text{Var}(\sum_{t=0}^T Z'_{it} \delta_t) + \sigma_\eta^2$. 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

To see how the CRE model relaxes exogeneity, note that the basic panel probit assumes

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

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

These equations imply that, conditional on $(h_{i,t-1}, \alpha_i)$, only X_{it} helps predict h_{it} ; that is, 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 we will 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 will call SE-B. By dropping assumption (3b), the CRE model relaxes SE-B while maintaining SE-A.⁷

Next we allow ε_{it} to be serially correlated. This could arise from persistence in shocks to tastes and/or productivity. Letting ε_{it} follow an AR(1) process, we have

$$(4) \quad \varepsilon_{it} = \rho \varepsilon_{it-1} + v_{it},$$

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

⁶The CRE model was first suggested by Chamberlain (1982) and first used by him (Chamberlain (1984)) to test exogeneity of children to married women's labor supply (i.e., employment status).

⁷Intuitively, 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 nonlabor 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 nonlabor income.

where v_{it} is normally distributed with zero mean and variance σ_v^2 , and is conditionally independent of ε_{it-1} . We assume the process is stationary, so $\sigma_\varepsilon^2 = (\sigma_v^2)/(1 - \rho^2)$.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 $\sigma_u^2 = \sigma_\eta^2 + \sigma_v^2/(1 - \rho^2) = 1$. Thus, we can estimate ρ and σ_η^2 , and “back out” σ_v^2 using the formula $\sigma_v^2 = (1 - \rho^2)(1 - \sigma_\eta^2)$.

Finally, in addition to estimating ρ and σ_η^2 , we can allow for “true state dependence” by letting γ in (1) be nonzero. 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 like (1)–(4), it is well known that if the h_{it} process is not observed from its start, simply treating the first observed h as exogenous can lead to severe bias. Heckman (1981) proposed an approximate solution to this initial conditions problem where the first observed h is determined by a probit model:¹⁰

$$\begin{aligned}
 (5) \quad h_{it} &= 1(X'_{it}\beta + \gamma h_{it-1} + u_{it} > 0) && (t \geq 1), \\
 h_{i0} &= 1(X'_{i0}\beta_0 + u_{i0} > 0), \\
 \rho_t &= \text{corr}(u_{i0}, u_{it}) && (t \geq 1).
 \end{aligned}$$

Here $t = 0$ is the first period of observed data (not the start of the h_{it} process) and h_{i0} is the first observed h . The error in the first period probit for h_{i0} , denoted u_{i0} , is assumed to be $N(0, 1)$. ρ_t is the correlation between u_{i0} and the errors u_{it} for $t \geq 1$.

Hyslop (1999) adopted the restriction that the ρ_t 's are equal. Let ρ_0 denote their common value. In this case ρ_0 is also the covariance between u_{i0} and the individual effect α_i . (See Keane and Sauer (2006) for a derivation.)

⁸The 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.

⁹As discussed by Wooldridge (2005), what distinguishes true state dependence (γ) from serial correlation (i.e., random effects or AR(1) errors) 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).

¹⁰Again, we choose this method for comparability with Hyslop (1999). See Heckman (1981), Wooldridge (2005), and Keane and Wolpin (1997) for details on various alternative solutions.

2.2. *Incorporating Classification Error*

We generalize the panel probit model in (1)–(5) by nesting it within a model of classification error. Let h_{it}^* denote the reported choice, while h_{it} is the true choice. Let π_{jk} denote the probability 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

$$(6) \quad l_{it} = \gamma_0 + \gamma_1 h_{it} + \gamma_2 h_{it-1}^* + \omega_{it},$$

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

Naturally, we allow h_{it}^* to be a function of h_{it} , as the probability of a reported “1” should be greater if the person is actually employed.¹¹ We also include the lagged reported choice $h_{i,t-1}^*$ to capture 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

Combining (1)–(6), we arrive at the following panel data probit model of female labor supply decisions with classification error in reported choices:

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

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

¹¹Hausman, Abrevaya, and Scott-Morton (1998) noted 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 $\pi_{01} + \pi_{10} < 1$, which in our model is equivalent to $\gamma_1 > 0$. Thus, classification error can not be so severe that people misreport their state more often than not—certainly a mild requirement.

¹²Keane and Sauer (2005) showed that the classification error scheme in (6) performs quite well in repeated sampling experiments on panel probit models using our estimation procedure (i.e., both the parameters of (6) and the “true” process (1)–(5) are recovered with precision).

3. DATA

We use the same data as Hyslop (1999), who graciously gave us his data set. While in some cases we might have made different decisions in defining covariates or constructing the sample, it is essential the data be identical to facilitate replication. The data are from the 1986 Panel Study of Income Dynamics (PSID), including both the random core sample of families and nonrandom Survey of Economic Opportunity. The sample period is 1979–1985. We include women aged 18–60 in 1980, who were continuously married during the period and whose husbands were employed each year. This gives $N = 1812$ women and 12,684 person/year observations. A woman is classified as employed if she reports positive annual hours worked and positive earnings.

Table I reports means and standard deviations of variables in the analysis sample. The average employment rate over the whole sample is .70. Covariates used to predict employment are nonlabor income, number of children in three different age ranges (0–2, 3–5, and 6–17), age, education, and race (equal to 1 if black).¹³

As in Hyslop (1999), the log of husband's average earnings over the sample period $y_{ip} = \ln(\frac{1}{T} \sum_t y_{it}^m)$ is used as a proxy for permanent nonlabor income. Transitory nonlabor income is proxied by $y_{it} = \ln(y_{it}^m) - y_{ip}$. y_{ip} and y_{it} enter as separate covariates in estimation. The number of children aged 0–2 years lagged 1 year also appears as a covariate (see Hyslop (1999) for discussion).

The degree of persistence in employment status is very strong. $P(h_{it} = 1|h_{i,t-1} = 1)$ is 91%, while $P(h_{it} = 0|h_{i,t-1} = 0)$ is 78%. There is also an important asymmetry in transition rates. $P(h_{it} = 1|h_{i,t-1} = 1, h_{i,t-2} = 0)$ is 72.2%, while $P(h_{it} = 1|h_{i,t-1} = 0, h_{i,t-2} = 1)$ is only 40.3%. This implies 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 employment 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 II–IV present SML estimates of the CRE model in (7). In addition to the variables reported, all models control for race, education, a quadratic in age, unrestricted year effects, and the lagged value of number of children aged 0–2.

¹³There is substantial variation over time in numbers of children and transitory nonlabor income. The standard deviations over time 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 identification of the δ_i in the CRE estimator.

TABLE I
 SAMPLE CHARACTERISTICS PSID WAVES 12–19 (1978–1985) ($N = 1812$)^a

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

^aMeans and standard errors (in parentheses) for 1812 continuously married women in the PSID between 1979 and 1985, aged 18–60 in 1980, and with a husband who has positive annual hours worked and positive wages each year during the sample period. Earnings are in thousands of 1987 dollars. Variable definitions and sample selection criteria are the same as those chosen by Hyslop (1999).

TABLE II
CORRELATED RANDOM EFFECTS PROBIT MODELS OF EMPLOYMENT STATUS
(SML ESTIMATES)^a

	Hyslop	Keane and Sauer		
	No CE 1	No CE 2	CE 3	CE 4
y_{ip}	-.341 (.05)	-.336 (.05)	-.400 (.04)	-.375 (.04)
y_{it}	-.099 (.03)	-.103 (.03)	-.127 (.02)	-.172 (.03)
#Kids0-2 _t	-.300 (.03)	-.305 (.03)	-.290 (.04)	-.388 (.05)
#Kids3-5 _t	-.247 (.03)	-.245 (.03)	-.265 (.03)	-.271 (.04)
#Kids6-17 _t	-.084 (.03)	-.083 (.03)	-.090 (.02)	-.087 (.03)
Var(η_i)	.804 (.02)	.829 (.04)	.938 (.07)	.943 (.10)
γ_0	—	—	-2.427 (.09)	-2.386 (.11)
γ_1	—	—	6.996 (.21)	5.056 (.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)**
$\delta_{y_{it}=0}$	48.50 (.00)**	48.71 (.00)**	50.08 (.00)**	62.60 (.00)**

^aAll specifications include number of children aged 0–2 years lagged 1 year, race, maximum years of education over the sample period, a quadratic in age, and unrestricted year effects. Nonlabor income is measured by y_{ip} and y_{it} , which denote husband’s permanent (sample average) and transitory (deviations from sample average) annual earnings, respectively. Var(η_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.

4.1. Basic CRE Model

Column 1 of Table II 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 “permanent” nonlabor income on work decisions is stronger

¹⁴This CRE model could have been estimated without simulation (e.g., using quadrature). We use SML so differences with the AR(1) models in columns 3 and 4 do not arise due to simulation per se.

TABLE III
CORRELATED RANDOM EFFECTS PROBIT MODELS OF EMPLOYMENT STATUS WITH AR(1)
ERRORS (SML ESTIMATES)^a

	Hyslop	Keane and Sauer		
	No CE 1	No CE 2	CE 3	CE 4
y_{ip}	-.332 (.05)	-.327 (.04)	-.345 (.00)	-.345 (.00)
y_{it}	-.097 (.03)	-.108 (.03)	-.112 (.01)	-.085 (.01)
#Kids0-2 _t	-.272 (.03)	-.251 (.03)	-.306 (.02)	-.307 (.02)
#Kids3-5 _t	-.234 (.03)	-.219 (.02)	-.265 (.01)	-.269 (.01)
#Kids6-17 _t	-.077 (.02)	-.083 (.02)	-.079 (.01)	-.083 (.01)
Var(η_i)	.546 (.04)	.582 (.03)	.830 (.03)	.831 (.04)
ρ	.696 (.04)	.710 (.05)	.746 (.02)	.748 (.03)
γ_0	—	—	-2.650 (.12)	-2.675 (.13)
γ_1	—	—	7.909 (.35)	6.837 (.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)**
$\delta_{y_{it}}=0$	8.22 (.22)	8.91 (.18)	53.84 (.00)**	40.45 (.00)**

^aAll specifications include number of children aged 0-2 years lagged 1 year, race, maximum years of education over the sample period, a quadratic in age, and unrestricted year effects. Nonlabor income is measured by y_{ip} and y_{it} , which denote husband's permanent (sample average) and transitory (deviations from sample average) annual earnings, respectively. Var(η_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.

than that of transitory nonlabor income, and young children have a larger negative effect on employment than older children. The estimate of Var(η_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

¹⁵The proportion of the overall error variance σ_u^2 due to permanent unobserved heterogeneity is $\sigma_\eta^2/\sigma_u^2 = \sigma_\eta^2/(\sigma_\eta^2 + \sigma_\varepsilon^2) = \sigma_\eta^2$, following the normalization for scale, $\sigma_u^2 = 1$.

TABLE IV
CORRELATED RANDOM EFFECTS PROBIT MODELS OF EMPLOYMENT STATUS WITH AR(1)
ERRORS AND FIRST-ORDER STATE DEPENDENCE (SML ESTIMATES)^a

	Hyslop		Keane and Sauer	
	No CE 1	No CE 2	CE 3	CE 4
y_{ip}	-.285 (.05)	-.291 (.05)	-.362 (.01)	-.451 (.01)
y_{it}	-.140 (.04)	-.137 (.05)	-.134 (.03)	-.186 (.03)
#Kids0-2 _t	-.252 (.05)	-.254 (.05)	-.322 (.05)	-.420 (.05)
#Kids3-5 _t	-.135 (.05)	-.131 (.04)	-.158 (.03)	-.171 (.03)
#Kids6-17 _t	-.054 (.04)	-.053 (.04)	-.072 (.02)	-.110 (.03)
Var(η_i)	.485 (.04)	.519 (.06)	.781 (.09)	.787 (.11)
ρ	-.213 (.04)	-.141 (.03)	.619 (.03)	.649 (.03)
h_{t-1}	1.042 (.09)	1.031 (.07)	.733 (.03)	.726 (.04)
Corr(u_{i0}, u_{it})	.494 (.03)	.561 (.09)	.835 (.18)	.853 (.21)
γ_0	—	—	-2.684 (.09)	-2.252 (.08)
γ_1	—	—	6.842 (.14)	5.427 (.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)**
$\delta_{y_{it}=0}$	2.92 (.82)	5.92 (.43)	47.70 (.00)**	38.20 (.00)**

^aAll specifications include number of children aged 0–2 years lagged 1 year, race, maximum years of education over the sample period, a quadratic in age, and unrestricted year effects. Nonlabor income is measured by y_{ip} and y_{it} , which denote husband’s permanent (sample average) and transitory (deviations from sample average) annual earnings, respectively. Var(η_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.

exogeneity of children in three age ranges (0–2, 3–5, and 6–17) and nonlabor income (i.e., tests of $H(0): \delta_t = 0 \forall t$). The null hypothesis that children and nonlabor income are exogenous is clearly rejected.

Column 2 presents estimates of the exact same model except 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, that is, $\pi_{01} = \pi_{10} = .0025$. 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 use of a different simulation method. Comparing columns 1 and 2, we see the results are essentially identical—the alternative estimation method makes almost no difference.

Next, we introduce classification error. Column 3 presents estimates of the model with no persistence in misclassification (i.e., $\gamma_2 = 0$ in (7)), while column 4 reports the model that allows persistence. Allowing for classification error (of either type) produces little change in the coefficients on covariates, but it increases the estimated fraction of the overall error variance due to unobserved heterogeneity from about 80% to 94%.¹⁷ This large increase in the importance of unobserved heterogeneity suggests that misclassification exaggerates the frequency of transitions between labor market states. Given the increased importance of the random effects, it is not surprising that the χ^2 statistics for the hypotheses of exogenous fertility and nonlabor income increase substantially, leading to even stronger rejections of exogeneity.¹⁸

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 working when the true state is not working ($\hat{\pi}_{01}$) is .081. Conversely, the probability of falsely reporting nonemployment ($\hat{\pi}_{10}$) 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.¹⁹ 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 substantial attenuation bias in the importance of unobserved heterogeneity.

The estimate of γ_2 in column 4 implies considerable persistence in misreporting. However, we 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

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

¹⁷Comparing columns 2 and 3 by a likelihood ratio test produces a $\chi^2(2)$ of 19.96 and a p-value of .000. Thus, introducing classification error also leads to a significant improvement in fit.

¹⁸The increased σ_α^2 makes it easier to detect correlations between the individual effect and fertility and nonlabor income. Note that $\hat{\sigma}_\alpha^2$ is bigger in the CRE models with classification error *both* because $Z'_{it}\hat{\delta}_i$ is more important and because $\hat{\sigma}_\eta^2$ is larger (recall that $\sigma_\alpha^2 = \text{Var}(\sum_{t=0,T} Z'_{it} \times \delta_i) + \sigma_\eta^2$).

¹⁹The CPS asks about current employment while the PSID asks about annual employment, so the two measures are not strictly comparable. Our prior is that a current measure would tend to have *less* error (i.e., it is easier to say if you are employed today than if you were employed at all during the past year), so that CPS error rates would be lower than PSID error rates. Of course, this is only speculative. The point of our comparison is merely to show that our error rates are not implausibly high.

with only random effects the parameter γ_2 tends to “sop up” omitted serial correlation in ε_{it} .

4.2. CRE With AR(1) Errors

Table III reports estimates of the same sequence of models as in Table II, except now we allow for AR(1) serial correlation in the transitory error. Columns 1 and 2 reproduce the rather dramatic finding from Hyslop (1999). Specifically, with the introduction of AR(1) errors we can no longer reject the null hypothesis that fertility and nonlabor income are exogenous at any conventional level of significance.

Introducing AR(1) errors has little impact on the nonlabor income and fertility coefficients, but the importance of the individual effect is considerably reduced, dropping from 80% of the overall error variance to only 55% (compare column 1 of Tables II and III). The estimated AR(1) coefficient ($\hat{\rho}$) is .696 and precisely estimated, and including it leads to a 225 point improvement in the log-likelihood (compare column 1 in Tables II and III). Thus, AR(1) serial correlation appears to be an important source of persistence in reported labor market states. This replicates Hyslop’s other main result: that the equicorrelation assumption is soundly rejected.

Table III columns 3 and 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 nonlabor 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).

Also note that the AR(1) parameter actually increases (slightly) when we introduce classification error, from .70 in column 1 to .75 in columns 3 and 4. Thus, Hyslop’s other main finding—the rejection of equicorrelation—is still supported.²⁰

The estimates of γ_0 and γ_1 in column 3 imply that $\hat{\pi}_{01}$ is .066 and $\hat{\pi}_{10}$ 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 $\chi^2(2)$ statistic of 31.8, implying a p -value of .000.

²⁰Note also that introduction of AR(1) errors into (either) classification error model reduces the fraction of variance due to permanent unobserved heterogeneity from about 94% to 83% (compare columns 3 and 4 in Table II to columns 3 and 4 in Table III).

Column 4 presents estimates allowing for persistence in misclassification. The estimate of γ_2 implies substantial persistence. There is a substantial increase in the probability of falsely reporting a particular labor market state if that same state was reported in the previous period.²¹ This suggests that persistent misclassification may be an important source of recorded persistence in female employment.

Note, however, that the strength of persistence in misclassification is sensitive to the inclusion of AR(1) errors in the model. In Table III column 4, relaxing the restriction that $\gamma_2 = 0$ results in a rather large improvement in the log-likelihood of 13 points, but this is much smaller than the 206 point improvement we saw in Table II column 4 when we added persistent classification error to a model without AR(1) errors.²² Thus, while still significant, persistent classification error does not give nearly so great a likelihood improvement once we allow for AR(1) errors.²³

4.3. CRE With AR(1) Errors and SD(1)

Table IV 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), we deal with the initial conditions problem that arises when SD(1) is included in the model by using the Heckman approximate solution. Column 1 reports the model without classification error from Hyslop (1999). The coefficient on lagged employment is a strong 1.042 and is precisely estimated. Including lagged employment 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 $\hat{\rho}$ falls dramatically from .696 to $-.213$.

Column 2 reports estimates of the same model, except using our SML estimator. The estimates are little different from Hyslop's and his main finding is again replicated: 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 and 4 report estimates of models that include classification error. These models produce substantial improvements in the log-likelihood:

²¹For instance, the probability of reporting employment, when the true state is nonemployment and nonemployment is reported in the previous period, is .064, but if employment was reported in the previous period, this error rate rises to .250. Similarly, the probability of reporting nonemployment, when the true state is employment and employment is reported in the previous period, is only .003, but if nonemployment was reported in the previous period, this error rate rises to .015.

²²Also, comparing column 4 of Tables II and III, we see that including the AR(1) error component leads to a drop in the estimated persistence in misclassification (i.e., $\hat{\gamma}_2$ falls from 2.61 to 1.58).

²³The 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, $\gamma_2 > 0$ implies that lagged X 's help to predict current choices, while $\rho > 0$ does not have this implication.

32 points with no persistence in classification error and an additional 26 points when persistence is allowed. Compared to Hyslop (1999), they produce quite different estimates of the importance of random effects, AR(1) errors, and state dependence. The first-order state dependence coefficient falls to about .73 (compared to 1.04 in column 1), the fraction of the error variance due to random effects increases to .78 (compared to .49 in column 1), and the AR(1) coefficient increases to about .62 to .65 (compared to $-.21$). Particularly notable is the complete reversal in sign on the AR(1) coefficient, back to a more plausible positive value.

Thus, failure to account for classification error produces severe attenuation biases in the importance of unobserved heterogeneity and AR(1) serial correlation, and an upward bias in the extent of first-order state dependence.²⁴ 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 III and IV).²⁵

Finally, the classification error models in columns 3 and 4 again reject overwhelmingly the hypotheses of exogenous fertility and nonlabor income. The difference in results from columns 1 and 2 is again a direct result of the greater estimated variance of the random effect in models that accommodate classification error.

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 panel probit models of married women's employment decisions, using PSID data from 1979 to 1985. His innovation was to relax the equicorrelation assumption of the common

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

²⁵The 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

$$E(h_{it}^* | X_{it}, h_{i,t-1}^*, X_{i,t-1}) = E(h_{it}^* | X_{it}, h_{i,t-1}^*).$$

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 choices even conditional on the lagged (measured) choice.

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, that is, 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 ignore it to substantially understate the importance of individual random effects. This is obviously due to the spurious transitions created by misclassification. 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

- CHAMBERLAIN, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46. [978]
- (1984): "Panel Data," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: Elsevier Science, Chapter 22. [975,978,979]
- ECKSTEIN, Z., AND K. I. WOLPIN (1989): "Dynamic Labor Force Participation of Married Women and Endogenous Work Experience," *Review of Economic Studies*, 56, 375–390. [975]
- 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. [980]
- 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*, ed. by C. F. Manski and D. McFadden. Cambridge, MA: MIT Press. [979]
- 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, 27–58. [975]
- HOTZ, V. J., AND R. A. MILLER (1988): "An Empirical Analysis of Life Cycle Fertility and Female Labor Supply," *Econometrica*, 56, 91–118. [976]
- HYSLOP, D. R. (1999): "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women," *Econometrica*, 67, 1255–1294. [975-977,979, 981,982,986-990]

- 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, 302–329. [975]
- KEANE, M. P. (1994): "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116. [976,977]
- 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 <http://ssrn.com/abstract=448240>. [977,980]
- (2006): "Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior," Discussion Paper 2332, IZA. [979]
- (2009): "Supplement to 'Classification Error in Dynamic Discrete Choice Models: Implications for Female Labor Supply Behavior'," *Econometrica Supplemental Material*, 77, http://www.econometricsociety.org/ecta/Supmat/7642_simulations.zip. [977]
- KEANE, M. P., AND K. I. WOLPIN (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522. [979]
- (2001): "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," *International Economic Review*, 42, 1051–1103. [977]
- (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," Working Paper 06-004, PIER. [976]
- MOFFITT, R. (1984): "Life Cycles Profile of Labour Supply and Fertility," *Review of Economic Studies*, 51, 263–278. [976]
- MROZ, T. A. (1987): "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions," *Econometrica*, 55, 765–799. [975]
- 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, 273–298. [975]
- POTERBA, J. M., AND L. H. SUMMERS (1986): "Reporting Errors and Labor Market Dynamics," *Econometrica*, 54, 1319–1338. [976]
- 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, 992–1015. [975]
- 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, 39–54. [979]

University of Technology Sydney, City Campus, P.O. Box 123, Broadway, Sydney, NSW 2007, Australia and Arizona State University, Tempe, AZ 85287, U.S.A.; michael.keane@uts.edu.au

and

School of Economics, Finance and Management, University of Bristol, 8 Woodland Road, Bristol BS8 1TN, U.K.; robert.sauer@bristol.ac.uk.

Manuscript received December, 2007; final revision received November, 2008.