Chapter 36

Classification Error Adjustments for Female Labour Force Transitions Using a Latent Markov Chain with Random Effects

Keith Humphreys

University of Southampton, UK.

1. Introduction

Because of the increasing labour force participation rate of mothers over the last two decades in the U.S. (Blau and Robins, 1991) and elsewhere, female labour force participation has become an important issue in the economic analysis of households. There is a vast literature concerned with the impact of economic factors (such as child care costs and wage rates) on female labour force participation (see Lehrer, 1992 and Ribar, 1992, for example). This paper studies the „transition“ of women between categories of the same variable measured through (discrete) time. Typically in such analyses misreporting of employment status is assumed not to occur even though it can be serious. For a review of the literature on misreporting of employment status see Holt et al. (1991). The potential bias introduced into the estimation of transitions by the presence of classification error is generally acknowledged as being of greater magnitude than that which is likely to occur in the measurement of cross-sectional proportions (Skinner and Torelli, 1993).

A number of classification error models for discrete time longitudinal discrete data exist in the literature. These usually describe individuals changing latent/true states between different points in time according to a Markov process and observed states are assumed to occur according to some model concerning classification error. For convenience, researchers have frequently assumed „independent classification error“ (ICE); see Shockey (1988) and Meyer (1988). Specifically it is assumed that (a) there is (conditional) independence between observed states at each time point, given the true states and (b), that the current observed state is independent of future/past true states given the current true state.

Van de Pol and Langeheine (1995) describe one approach to testing (b) where observed states are specified as being dependent on both the current and previous true state. Assumption (a) can be violated in a number of ways. Kristiansson (1984), for example, suggests that (in the context of labour force surveys) one may expect some „carry over“ effect from classifications made at successive time points if the classifications are made by the same interviewer. Assumption (b) is violated if individuals with particular latent state histories are more or less prone to classification error than others. The further assumption of population homogeneity is frequently assumed. This is strong and under this assumption classification errors are unlikely to be independent. Heterogeneity in transition probabilities will occur, in the context of labour force data if, for example, transitions between states of employment are dependent on a variety of demographic and socio-demographic factors. This seems likely. Heterogeneity in classification errors can occur if, as Singh and Lemaitre

(1989) suggest, some individuals are consistently more or less reliable than others due to some unobserved (and quite likely unobservable) factor.

In this paper an approach is developed for studying binary events which allows for heterogeneity between sample units by extending the latent Markov chain (which assumes ICE) by postulating regressions for the latent/true state membership probabilities and classification error probabilities on time varying covariate vectors and unobserved random components. Potential applications of these extensions are illustrated by the examples given above of how heterogeneity can occur in true state transitions and classification errors.

The family of Mixed Markov Latent Class models represents an alternative approach to allowing for heterogeneity in classification error and in latent/true state membership. These models can be fitted using PANMARK (van de Pol et. al., 1991). They allow for the simultaneous analysis of several subpopulations which can be either latent or be defined according to observed characteristics of the sample units. They are, however, restrictive since subpopulation membership is assumed to be time constant and classification errors, as well as transitions, are assumed independent within subpopulations. The more time points that are studied, the more important it is likely to be that the time varying nature of the covariates be accounted for. To introduce dependence on a time varying multi-category (H categories, say) covariate H^T subpopulations could be set up (where the model is fitted for T time points) but this is clumsy and too heavily parameterised for reliable inferences.

2. Extending the latent Markov chain to incorporate heterogeneity using observed time dependent covariates and unobserved random effects.

Consider a longitudinal study involving the observation of sample units $S = \{1,\ldots,N\}$over $T \geq 2$ consecutive time points. Let index i label the individual sample units $(i \in S)$ and index $t(= 1,\ldots,T)$ label the time points of the survey. Further, let Y_{it} denote the true state of unit i at time t, and y_{it} denote the measured state of unit i at time t, for a two state categorical variable (y_{it}, Y_{it} = 0 or 1; i = 1,…,N; t = 1,…,T). Let F represent some mapping function and $\mathbf{x_{it}} = (x_{it1},\ldots,x_{itR})'$ define an R dimensional covariate vector measured at occasion $t(= 1,\ldots,T)$ on sample unit i. The following notation is used.

Classification error probabilities :

$$p\left(y_{it} = 1 \middle| Y_{it} = a\right) = P^t_{1|ai} = F^{-1}\left[\boldsymbol{\rho}^t_a \; \mathbf{x}_{it} + \varepsilon\left(\rho_a\right)_i\right] \qquad , a = 0,1; t = 1,\ldots,T,$$

$$p\left(y_{it} = 0 \middle| Y_{it} = a\right) = P^t_{0|ai} = 1 - P^t_{1|ai} \qquad , a = 0,1; t = 1,\ldots,T,$$

where $\boldsymbol{\rho}^1_a = \left(\rho^t_{a1},\ldots,\rho^t_{aR}\right) \qquad , a = 0,1; t = 1,\ldots,T.$

Latent state membership probabilities :

$$p\left(Y_{i1} = 1\right) = \Delta_{1|i} = F^{-1}\left[\boldsymbol{\delta}\, \mathbf{x}_{i1}\right] \quad \text{and} \quad p\left(Y_{i1} = 0\right) = \Delta_{0|i} = 1 - \Delta_{1|i}$$

$$p\left(Y_{it} = 1 \middle| Y_{it-1} = a\right) = M^t_{1|ai} = F^{-1}\left[\gamma^t_a\, \mathbf{x}_{it} + \varepsilon\left(\gamma_a\right)_i\right] \qquad , a = 0,1; t = 2,\ldots,T,$$

$$p\left(Y_{it}=0 \middle| Y_{it-1}=a\right)=M^{t}_{0|ai}=1-M^{t}_{1|ai} \qquad ,a=0,1;\ t=2,\ldots,T,$$

where $\delta=\left(\delta_1,\ldots,\delta_R\right)$ and $\boldsymbol{\gamma}^{t}_{a}=\left(\gamma^{t}_{a1},\ldots,\gamma^{t}_{aR}\right)$ $\qquad ,a=0,1;\ t=2,\ldots,T.$

Random effects are not included for the submodel for $\Delta_{a|i}\,(a=0,1)$ because it would not be possible to estimate a distribution since no repeated information is available on latent state membership at the first point in time. It is assumed that all heterogeneity in the probability of latent state membership at t = 1 can be captured by the observed covariates in $\mathbf{x}_{i1}$. It would only be possible to fit a random effect for this submodel if it were assumed that unmeasured determinants of latent state membership at t = 1 were the same as those of other probabilities, such as one of the latent state transitions (i.e., $\varepsilon(\gamma_0)_i$ or $\varepsilon(\gamma_1)_i$) and this seems unreasonable.

The motivation for specifying dependence of latent state membership and classification error on the random effects $\varepsilon(\gamma_a)_i$ and $\varepsilon(\rho_a)_i$ (a = 0,1) is described in the introduction. The random effect $\varepsilon(\rho_a)_i$ (a = 0,1) is introduced, for example, because it is possible that individuals have consistently different probabilities of being classified incorrectly and because these differences may not be perfectly predicted by the observed covariates. Each random effect is time constant and choices of distributional form are discussed in Section 2.1. If random effects $\varepsilon(\gamma_a)_i$ and $\varepsilon(\rho_a)_i$ (a = 0,1) are distributed with joint density $g(\boldsymbol{\varepsilon}_i)$, say, then the probability that any sample unit i(= 1,…,N) is measured with vector $\mathbf{y_i}=\mathbf{j}$ where $\mathbf{y}_i=(y_{i1},\ldots,y_{iT})$ and $\mathbf{j}=(j_1,\ldots,j_T)$, given $\mathbf{x}_{it}$,, is

$$p\left(\mathbf{y}_i=\mathbf{j}\middle|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}\right)=\int\sum_{a_1=0}^{1}\cdots\sum_{a_T=0}^{1}\left[\Delta_{a_1|i}\left(\prod_{t=2}^{T}M^{t}_{a_t|a_{t-1}}\right)\left(\prod_{t=1}^{T}P^{t}_{j_t|a_t}\right)\right]g\left(\boldsymbol{\varepsilon}_i\right)d\boldsymbol{\varepsilon}_i \qquad (1)$$

The integration is over the full range of each random effect distribution.

To fit the model using maximum likelihood estimation the incomplete data loglikelihood $\sum_{i=1}^{N}\ln\left(p(\mathbf{y}_i|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT})\right)$ is maximised with respect to the parameter set described above. The „predicted latent/true counts“ for each vector of possible responses, $\mathbf{a}^*=(a^*_1,\ldots,a^*_T)'$ say $(a^*_t=0,1;\ t=1,\ldots,T)$ are calculated by evaluating the following posterior probabilities at the maximum likelihood parameter estimates and summing these across all sample units (i = 1,…,N).

$$p\left(\mathbf{Y}_i=\mathbf{a}^*\middle|\,\mathbf{x}_i,\mathbf{z}_{i1},\ldots,\mathbf{z}_{iT}\right)=\frac{\int\Delta_{a^*_1|i}\left(\prod_{t=2}^{T}M^{t}_{a^*_t|a^*_{t-1}i}\right)\left(\prod_{t=1}^{T}P^{t}_{y_{it}|a^*_t i}\right)g\left(\boldsymbol{\varepsilon}_i\right)d\left(\boldsymbol{\varepsilon}_i\right)}{\int\sum_{a_1=0}^{1}\cdots\sum_{a_T=0}^{1}\Delta_{a_1|i}\left(\prod_{t=2}^{T}M^{t}_{a_t|a_{t-1}i}\right)\left(\prod_{t=1}^{T}P^{t}_{y_{it}|a_t i}\right)g\left(\boldsymbol{\varepsilon}_i\right)d\left(\boldsymbol{\varepsilon}_i\right)}$$

The „predicted observed“ counts $\mathbf{y}_i=\mathbf{j}$ are calculated by evaluating and summing across all sample units expression (1) at the maximum likelihood parameter estimates.

2.1 Distributional assumptions for the random effects and evaluation of expression (1)

Under common choices of distribution for the random effect and link function (Gaussian and logistic respectively, for example), evaluating the integral in (1) requires numerical evaluation techniques (Gaussian quadrature, say) and this will be computationally intensive. There is one approach in the literature however, due to Conoway (1990), which can be applied and is not computationally intensive. In the context of a single random effects model for (repeated) binary responses (assumed to be measured without error) Conoway describes how, under particular parametric formulations for the random effect, ε_i, say, the probability of observing any vector of responses can be evaluated analytically if the linear predictor (which includes the random effect) is linked to the probability of an event occurring through the log-log function. One of the distributions considered by Conoway for the random effect is the log-Gamma. If ε_i~logGamma(η,λ) then it has density,

$$f_{\varepsilon_i}\left(\varepsilon_i|\eta,\lambda\right)=\frac{\lambda^{\eta}\exp\left(\eta\varepsilon_i\right)\exp\left(-\lambda\exp(\varepsilon_i)\right)}{\Gamma(\eta)}\quad;\quad\varepsilon_i,\eta,\lambda\succ 0$$

Lawless (1982) describes properties of this distribution, which is negatively skewed, with skewness decreasing as η increases. λ essentially represents a location parameter. The log-Gamma produces a wide variety of shapes for the random effects distribution, including some that closely resemble normal distributions. Conoway considers the following random effects binary response log-log link regression model.

$$p(Y_{it}=1)=\exp\left(-\exp\left(\boldsymbol{\beta}_t'\mathbf{x}_{it}+\varepsilon_i\right)\right)$$

where $\boldsymbol{\beta}_t(t = 1,\ldots,T)$ is a P dimensional vector of regression coefficients and ε_i is assumed independent of the covariate vector $\mathbf{x}_{it}$. Responses are assumed independent given the covariate vector at each time point and ε_i. Let $\pi_i(\mathbf{a})$ define the unconditional probability of observing the vector $\mathbf{Y}_i = \mathbf{a}$ where $\mathbf{Y}_i = (Y_{i1},\ldots,Y_{iT})$ and $\mathbf{a} = (a_1,\ldots,a_T)$ (a_t = 0,1; t = 1,…,T). Let $\boldsymbol{\pi}_i$ be a vector of length 2^T containing the probabilities $\pi_1(\mathbf{a})$ with the first element corresponding to $\mathbf{a}$ = (1,1,…,1) and with the left element being fastest to change. The last element of $\pi_1(\mathbf{a})$ corresponds to $\mathbf{a}$ = (0,0,…,0). Now let T^* define a subset of time points from the set {1,2,…,T} and $\pi_i^*(T^*)$ be the joint probability that Y_{it} = for all t εT^*. If $\boldsymbol{\pi}_i^*$ is a vector of length 2^T containing $\pi_i^*(T^*)$ with the subsets of T^* in the order $\varnothing$,{1},{2},{1,2},{3},{1,3},{2,3},…,{1,2,3,…,T} then $\boldsymbol{\pi}_i=\mathbf{A}^{-1}\boldsymbol{\pi}_i^*$, where $\mathbf{A}$ is the $2^T\times 2^T$ design matrix of ones and zeros whose columns correspond to the effects in the saturated model for a 2^T factorial design and $\varnothing$ denotes the empty set. For ε_i~logGamma(η,λ) Conoway shows that $\pi_i^*(T^*)$ takes the form of the moment generating function for a Gamma distributed variate. Specifically,

$$\pi_i^*\left(T^*\right)=\left[\frac{\lambda}{\lambda+\sum_{t\in T^*}\exp\left(\boldsymbol{\beta}_t'\mathbf{x}_{it}\right)}\right]^{\eta}$$

Without extending Conoway's technique for random effects from multivariate distributions it is, in order to apply it to evaluate (1), necessary to make the additional assumption that random effects are mutually uncorrelated. Let $g_{\gamma_a}(\varepsilon(\gamma_a)_i)$ with parameters

$\eta(\gamma_a)$ and $\lambda(\gamma_a)$ and $g_{\rho_a}(\varepsilon(\rho_a)_i)$ with parameters $\eta(\rho_a)$ and $\lambda(\rho_a)$ denote the densities of $\varepsilon(\gamma_a)_i$ and $\varepsilon(\rho_a)_i$ (a = 0,1) respectively. Under this assumption (after inter-changing integrals with summations) (1) can be re-written as a summation across 2^T elements.

$$p\left(\mathbf{y}_i = \mathbf{j} \middle| \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}\right) = \sum_{a_1=0}^{1} \cdots \sum_{a_T=0}^{1} \Delta_{a_1|i} I_1(\mathbf{a})_i I_2(\mathbf{a}, \mathbf{j})_i$$

where

$$I_1(\mathbf{a})_i = p\left(Y_{i2} = a_2, \ldots, Y_{iT} = a_T \middle| Y_{i1} = a_1, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}\right) =$$
$$\int \prod_{t=2}^{T} \left(M^t_{a_t|a_{t-1}}\right)^{(1-a_{t-1})} g_{\gamma_0}\left(\varepsilon(\gamma_0)_i \middle| \eta(\gamma_0)\lambda(\gamma_0)\right) d\varepsilon(\gamma_0)_i \times \int \prod_{t=2}^{T} \left(M^t_{a_t|a_{t-1}}\right)^{a_{t-1}} g_{\gamma_1}\left(\varepsilon(\gamma_1)_i \middle| \eta(\gamma_1), \lambda(\gamma_1)\right) d\varepsilon(\gamma_1)_i$$

and

$$I_2(\mathbf{a}, \mathbf{j})_i = p\left(\mathbf{y}_i = \mathbf{j} \middle| \mathbf{Y}_i = \mathbf{a}, \mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}\right) =$$
$$\int \prod_{t=1}^{T} \left(P^t_{j_t|a_t}\right)^{(1-a_t)} g_{\rho_0}\left(\varepsilon(\rho_0)_i \middle| \eta(\rho_0), \lambda(\rho_0)\right) d\varepsilon(\rho_0)_i \times \int \prod_{t=1}^{T} \left(P^t_{j_t|a_t}\right)^{a_t} g_{\rho_1}\left(\varepsilon(\rho_1)_i \middle| \eta(\rho_1), \lambda(\rho_1)\right) d\varepsilon(\rho_1)_i \quad .$$

3. A study of transitions in female labour force participation

In this section the movement of a sample of women from the Panel Study of Income Dynamics (PSID), who are aged between 16 and 49 over all of the five years, 1983 to 1987, of the survey, between the states „in the labour force“(0) and „keeping house“(1) is studied. The PSID includes two subsamples, an equal probability sample and an unequal probability sample which oversamples low income households (Hill, 1992). To avoid the complication of weighting, only sample units from the equal probability sample are used. For this analyses any units that have missing values on any of the covariates have been excluded. This results in a sample of 1726 women. Table 1 displays the contingency table for their reported states of labour force participation.

response vector	*observed freq.*	*response vector*	*observed freq.*	*response vector*	*observed freq.*	*response vector*	*observed freq.*
0 0 0 0 0	948	0 1 0 0 0	29	1 0 0 0 0	76	1 1 0 0 0	46
0 0 0 0 1	35	0 1 0 0 1	6	1 0 0 0 1	11	1 1 0 0 1	14
0 0 0 1 0	24	0 1 0 1 0	7	1 0 0 1 0	4	1 1 0 1 0	10
0 0 0 1 1	24	0 1 0 1 1	12	1 0 0 1 1	8	1 1 0 1 1	17
0 0 1 0 0	28	0 1 1 0 0	15	1 0 1 0 0	14	1 1 1 0 0	32
0 0 1 0 1	5	0 1 1 0 1	5	1 0 1 0 1	6	1 1 1 0 1	16
0 0 1 1 0	6	0 1 1 1 0	6	1 0 1 1 0	5	1 1 1 1 0	40
0 0 1 1 1	30	0 1 1 1 1	35	1 0 1 1 1	10	1 1 1 1 1	202

Table 1: Contingency table of observed U.S. PSID. female labour force participation data, 1983-87.

All models fitted in this paper assume time constant processes underlying transitions between latent states and classification errors. Given that female labour participation has been rising in the U.S. (Blau and Robins, 1991), this may not be completely realistic, but for simplicity time trends are assumed to negligible. First some models for contingency tables - (latent) mixtures of (homogeneous) latent Markov chains - are fitted (Table 2). The latent class model, with a single 2 state latent variable, assumes time constant membership of latent states for sample units, but that the observed states differ according to a (time constant) process of classification error. Likelihood Ratio (LR) values suggest that this model and the single population Markov chain provide poor explanations of the observed 2^5 table of flows.

	InL	*LR* χ^2	*Pearson* χ^2	*df*
Static factor with two latent classes	-3607.48	366.38	490.06	28
Single Markov chain	-3590.73	332.88	551.25	28
Single latent Markov chain	-3464.18	79.77	87.88	26
Two latent Markov chains	-3432.73	16.86	16.49	20

Table 2: A mixed Markov latent class analysis of PSID female labour force participation data, 1983 to 1987.

A significant improvement in fit is encountered when a model incorporating classification errors is fitted (row 3, Table 2), but it is not until population heterogeneity is incorporated, by fitting a latent mixture of two latent Markov chains (row 4, Table 2), that a model that provides a „good fit“ (a non-significant difference between the observed and predicted observed tables, at a 5 percent level of significance) to the observed table of flows is found.

notation	*name*	*covariate details(measured at t)*
x_{it1}	intercept	1 for all sample units.
x_{it2}	baby I	1 if woman has 1 child aged under 6, 0 otherwise
x_{it3}	baby II	1 if woman has 2+ children aged under 6, 0 otherwise.
x_{it4}	child	1 if woman has 1+ children aged 6 or over, 0 otherwise.
x_{it5}	age	{age-me(87)/sd(87)}, where me(87) and sd(87) are sample means and standard deviations of ages in 1987.
x_{it6}	age squared	x_{it5} squared.
x_{it7}	race	0 if white, 1 if non-white.
x_{it8}	education	1 if reached 12th grade, 0 otherwise.
x_{it9}	partner	1 if living with spouse or cohabiting, 0 otherwise.
x_{it10}	city	1 if largest city in county has population 100,000 or more, 0 otherwise

Table 3: Covariates considered as determinants of latent state membership and classification error probabilities.

PANMARK permits the use of multiple starting values. This option was used with 100 sample sets of starting values to ensure that the „best“ fitting two latent Markov chains model was selected. Note that restricting one of the latent Markov chains so that it's members stay in either of the latent states over all time points does not provide an equally adequate fit and fitting an extra chain does not provide a significant improvement of fit either. It is not obvious why the two latent Markov chains model (hereafter referred to as model I) provides a good fit. Perhaps the most sensible explanation would be related to whether or not women have young children, but the analyses that follow demonstrate that this explanation is too simplistic. Furthermore a good fit at the aggregate level may not imply good prediction (of latent flows) at an individual level (see Humphreys, 1994).

Next some of the heterogenous population models described in Section 2 are fitted. A number of covariates (Table 3) that are usually found to be related to female labour force participation (see Blau and Robins, 1991) are included into the analysis. Only a subset of these covariates x_{it5}, x_{it6}, x_{it7}, and x_{it8} are considered as possible determinants of classification error probabilities.

Time constant classification errors and latent state transitions are assumed (ie. the t superscripts are dropped). Covariates are incorporated using a stepwise selection procedure with LR tests with 5 percent significance levels as the base line for the inclusion or exclusion of the covariates („model II“), and then in addition the random effect terms („model III“) are stepwise selected. Goodness of fit is assessed primarily on the optimised value of the objective/log-likelihood function. Since models I,II and III are nested within the single latent Markov chain, this model is used as a baseline (Table 4).

Model	*# parameters*	*lnL*	*LR*	*Pearson* χ^2
Single latent Markov chain	5	-3464.2	*	80.0
I	11	-3432.7	62.9 (6)	16.5
II	19	-3205.8	518.8 (14)	78.3
III	22	-3170.3	587.9 (17)	28.7

Table 4: A summary of models I, II and III. The LR-statistic equals 2 times change in log-likelihood from latent Markov chain (df in parentheses).

For models I and III, Table 5 shows the observed number of women in the sample falling into the eight possible combinations of the two states over the three years 1983 to 1985 (the five years' flows have been aggregated over the years 1986, 1987 to simplify presentation), along with the „predicted observed“ flows and „predicted true“ flows obtained under the models. To alleviate concerns of parameter identification several sets of starting values were tried for the parameters in models II and III. In all cases estimates converged to the same solution.

response vector, t = 1,2,3.	observed frequency	predicted observed frequency		predicted true/latent frequency	
		I	III	I	III
0 0 0	1031	1044.4	1029.7	1097.0	901.1
0 0 1	69	67.3	72.5	35.2	48.6
0 1 0	54	52.8	51.4	5.5	12.4
0 1 1	61	63.8	67.6	35.4	58.2
1 0 0	99	103.6	98.8	78.7	81.5
1 0 1	35	40.0	40.6	4.5	12.6
1 1 0	87	76.7	76.3	75.5	78.0
1 1 1	290	292.6	289.0	394.3	533.7

Table 5: Predicted latent and predicted observed frequencies (t=1,2,3) under models I and III.

Models I and III produce very different predictions for the latent/true flows (Table 5), but without estimating the variance associated with the predicted latent flows under each of the models it is difficult to make reliable inferences. Humphreys (1994) describes a simulation study based on this data which provides such estimates and demonstrates that there is enough accuracy to conclude that the predictions for the latent flows are very sensitive to the model specification. The primary purpose of that simulation study is to see how closely the true flows can be predicted. A number of samples are randomly generated from model III (with the parameter values in Table 6) and the means and standard deviations of the observed flows and predicted true flows (under the assumed true model III with re-estimated coefficients) are obtained. By treating the population as infinite, the standard deviations of the estimates of the observed and predicted true flows under model III, in Table 5, can be approximated by the standard deviations of the flows recorded in the simulation study. In most cases the increase in the standard deviation from the observed table to the predicted true frequencies is more than offset by the predicted bias in the observed flow (the mean square error of the predicted true flow is less than that for the observed flow).

Comparing models II and III to the homogeneous single latent Markov chain, it can be seen that including the covariates and random effects improves very significantly the fit at the individual level (column 4 Table 4). To compare the fit at an aggregate level the Pearson χ^2 values are also included. Although conventionally this statistic would only be used for the aggregate level models, this information does give an indication that model II does not provide a sizeable improvement in fit at the aggregate level over the single latent Markov chain. This is not surprising. In fact because model II contains covariates it might have actually been expected that the aggregate level fit would be worse. Inclusion of random effects in model III provides a sizeable improvement of fit at the individual level over model II. (Since $\varepsilon(\rho_0)_i$ does not significantly contribute to a reduction in the value of the objective function it is not included).

Because model I is not nested within models II and III classical test procedures for comparing the fit of the models (LR tests) do not apply. Under the Akaike Information Criterion (AIC), which is used to discriminate between non-nested models, model III is most preferred (models I, II, III have AIC values 6887.5, 6447.6, 6374.6 respectively).

δ_1 (interc.)	0.795	δ_9 (partner)	-0.985	γ_{15}(age)	-0.271	$\eta(\gamma_0)$	1.494
δ_2 (baby I)	-0.889	γ_{02} (baby I)	-1.192	ρ_{01} (interc.)	1.547	$\lambda(\gamma_0)$	0.052
δ_3 (baby II)	-2.532	γ_{03} (baby II)	-2.035	ρ_{05} (age)	-0.083	$\eta(\gamma_1)$	1.346
δ_4 (child)	-0.637	γ_{04} (age)	0.547	ρ_{17} (race)	0.658	$\lambda(\gamma_1)$	8.855
δ_6 (age²)	-0.067	γ_{05}(educ.)	0.992			$\eta(\rho_1)$	0.280
δ_8 (educ.)	0.831	γ_{06} (partner)	-1.669			$\eta(\rho_1)$	0.460

Table 6: Parameter estimates for the heterogenous population latent Markov chain model III

The signs of the estimated coefficients under model III conform with what one would expect them to be (for example, the probability of participating in the labour force is smaller when women have children and is increased when women have more education; see Table 6). In the models where random effects are included intercepts are not included since they are redundant.

The estimated distributions of $\varepsilon(\rho_1)_i$ and $p(y_{it} = 1|Y_{it} = 1)$ are heavily (negatively) skewed. For example, the mean and median values of $p(y_{it} = 1|Y_{it} = 1)$ for the 'white' subpopulation are 0.70 and 0.87 respectively. The estimated distribution of the random effect $\varepsilon(\rho_1)_i$ has mean -2.99 and median - 2.03. These values are obtained by generating a very large number of values from the estimated distribution. The means have been checked against theoretical values (Lawless, 1982). In general, women „keeping house“ appear to be quite likely to report their status incorrectly (possibly because „in the labour force“ may be perceived to be a more desirable response), but some more than others. The severe misreporting estimated for state 1 results in the sizeable upward adjustment to flow 111 in Table 5.

The estimated distributions of the random effects associated with the latent state transitions are less skewed ($\varepsilon(\gamma_0)_i$) has mean 2.99 and median 3.12 and $\varepsilon(\gamma_1)_i$ has mean –2.30 and median –2.15; for women in the baseline category of each of the covariates included in the submodels for $p(Y_{it} = 1|Y_{it-1} = a)$, a = 0,1, the respective probabilities are 0.01, 0.00 and 0.87, 0.89 respectively). The estimated transitions out of the state „in the labour force“ are very small for women without children.

4. Discussion

For the U.S. female labour force participation flows data a mixed Markov latent class model that fits well was found but there seems no convincing substantive argument for the existence of two latent subpopulations. Intuitively a more plausible formulation is one which extends the single latent Markov chain by postulating regressions for both the classification error and true state transition probabilities (which incorporate observed covariates and unobserved random effects). This model, which estimates large differences in the classification error probabilities across sample units, provides a better fit at an individual level, but a slightly poorer fit at the aggregate level.

Misreporting is estimated to be very mild for women „in the labour force" (just under 1% for women of mean age), but for women „keeping house" it is estimated to be severe. Horvath (1982) similarly estimates serious misreporting of employment status by women (in a twelve month retrospective report of employment status, the U.S. Work Experience Survey underestimates person- years unemployment by 32% for women, compared to 2% for men). It is however possible that the findings from the analysis presented here are related to an inadequate definition of labour market status. This is a problem that may be particularly serious for women. Gönül (1992), for example, argues that men who are not in the labour force form a reasonably homogeneous group in terms of transition to employment, but that for women there is a large distinction between those that do and do not search for employment. This may be the case for the state „keeping house". Similarly unobserved 'state heterogeneity' may be present in the state „in the labour force" (there may be clear distinctions in behaviours between the part-time and full-time employed for example). Extending the analysis to more than two states would be informative. Model based estimates of flows have been demonstrated to be very sensitive to model specification and this result highlights the need to reflect substantive theory, concerned with processes of classification error, by allowing for the types of complexities that have been considered in this paper.

Acknowledgements

The author would like to thank Chris Skinner and Frank Van de Pol for their comments on earlier versions of this paper. Research was supported by grant numbers R000 23 2522 and H519 25 5005 from the Economic and Social Research Council. The data were made available through the ESRC Data Archive by the Inter-University Consortium for Political and Social Research, Ann Arbor, Michigan. The data were originally collected by J.N. Morgan. Neither the original investigator, nor the Consortium, nor the Archive bear any responsibility for the analysis presented here.

References

Blau, D.M. and Robins, P.K. (1991). Child care demand and labor supply of young mothers. *Demography. 28*, 3, 333-351.

Conoway, M.R. (1990). A random effects model for binary data. *Biometrics, 46*, 317-328.

Gönül, F. (1992). New evidence on whether unemployment and out of the labour force are distinct states. *Journal of Human Resources. 27*, 329-361.

Hill, M.S. (1992). *The Panel study of Income Dynamics : A users guide*. Newbury Park. CA: Sage.

Holt, D., McDonald, J.W. and Skinner, C.J. (1991). The effect of measurement error on event history. pp.665-686 in P. Biemer et al (Eds.) Measurement Errors in Surveys. Wiley.

Horvath, F,W. (1982). Forgotten unemployment: Recall bias in retrospective data. *Monthly Labor Review, 105*, 3, 40-43.

Humphreys, K. (1994). Estimation of longitudinal flows in the presence of classification error. The use of models with latent classes and random effects. Centre for Survey Data Analysis, University of Southampton, *Technical Report No. 94-10*.

Kristiansson, K.E. (1984). Gross-flow estimates in the Swedish labour force surveys. *Statistical Journal of the U.N. Economic Commission for Europe, 2*, 357-367.

Lawless, J.F. (1982). *Statistical models and methods for lifetime data*. New York : Wiley.

Lehrer, E.L. (1992). The impact of children on married women's labor supply. *Journal of Human Resources. 27*, 422-444.

Meyer, B.D. (1988). Classification error models and labor market dynamics. *Journal of Business and Economic Statistics. 6*, 385-390.

Ribar, D.C. (1992). Child care and the labor supply of married women. *Journal of Human Resources. 27*, 134-165.

Shockey, J.W. (1988). Adjusting for response error in panel surveys: a latent class approach. *Sociological Methods and Research. 17*, 65-92.

Singh, A.C. and Lemaitre, G.E. (1989). On an ε-contamination model for multinomial response error with application to labour flows. *Proceedings of the Section on Survey Research methods*. American Statistical Association.

Skinner, C.J. and Torelli, N. (1993). Measurement error and the estimation of gross flows from longitudinal economic data. *Statistica. 53*, 3, 391-405.

Van de Pol, F.J.R. and Langeheine, R. (1995). Separating change and measurement error in Panel surveys (with an application to labour market data). To appear in P.P. Biemer (Ed.) *Survey Measurement and Process Quality*.

Van de Pol, F.J.R., Langeheine, R. and Jong, W. de (1991*). Panmark User manual. Panel analysis using Markov chains*. Version 2.2. Netherlands Central Bureau of Statistics.