Partnership dynamics for young American men: a latent Markov approach

Antonella Cuteri, Gabriella Cuteri
Department of Statistics
Sapienza - University of Rome
and Alessio Farcomeni
Department of Experimental Medicine
Sapienza - University of Rome

Abstract

In this paper we evaluate the impact of some economic, social and demographic factors on marital status evolution. We focus on a panel of American men aged between 18 and 35 years. We model marital status (married/cohabitant vs otherwise) through a recently introduced latent Markov approach. In doing so, we adjust for unobserved heterogeneity modeling the random intercept as a first-order homogeneous Markov chain. The state dependence effect is estimated simply by including the lagged response variable among the covariates. We find a strong state dependence for marital status, and a strong effect of employment status both in the current and previous year. It can be concluded that a stable employment and income source may favour stable union formation.

Keywords: Employment, Marital status, latent Markov model, Panel Study of Income Dynamics, State dependence

1 Introduction

At the beginning of adulthood every individual usually makes several different choices about his/her own life. Many important choices concern leaving the parent's house, dealing with relationships, leading a working life, etc. Couple membership roles may remain stable, or they may change by entry into marriage/cohabitation or dissolution of marriage/cohabitation. This kind of dynamics is well known to affect and be affected by economic well-being, employment status and health conditions. Many studies focus on economic well-being following a couple disruption, with particular attention to gender differences. There are two different points of view: part of existing literature supports that women experiencing couple breakdown undergo a worsening in the income level because of a lower participation to the labor market, custody of children and lack of state support (Gadalla, 2009; Jarvis and Jenkins, 1999; Manting and Bouman, 2006). On the other side, some authors assert that also men suffer economic consequences in terms of loss of income, due to alimony payments and rent payment for a new habitation (McManus and DiPrete, 2001).

Aassve *et al.* (2007) consider the effect of marital disruption on economic well-being using a propensity score matching technique combined with a difference-in-difference estimator and a range of different measures of economic well-being. They find a strong gender bias when using a conventional measure as income or poverty status which is smaller for other indices of well-being. Income is a fundamental element also in the formation process of stable unions: Aassve *et al.* (2002) proposed an econometric search-theoretic model for young Americans to examine how income determine decision to leave parental home and live alone or form a stable partnership.

Another interesting aspect is the relationship between employment status and couple formation/dissolution. A recently published study, using a Cox proportional hazard regression model, reports higher hazard of unemployment for separating men and women with respect to married/cohabiting people, with a worse situation for men (Covizzi, 2008).

In this study we focus on men, and try to explore marital disruption from a reverse perspective, investigating what are the factors that can lead to marital disruption. Our answers will be of course limited by the very limited amount of available information, but we will adopt a mixed-effects model in order to remove unobserved heterogeneity, partly adjust for selection bias, and hence clearly estimate the effects of the available covariates.

We use a recently proposed latent Markov model (Bartolucci and Farcomeni, 2009) for categorical panel data to study the evolution of marital status in a panel of men. Using a data set derived from the Panel Study of Income Dynamics (PSID), we want to establish the impact of a set of explanatory variables on union evolution for men aged between 18 and 35 years. We restrict our analysis on young subjects because their sentimental life is more likely to be turbulent, and to be affected by other factors (e.g., income, employment, education, etc.).

Bartolucci and Farcomeni (2009) propose a latent Markov model with covariates, extending work of Wiggins (1973) and Bartolucci (2006) among the others. The approach of Bartolucci and Farcomeni (2009) is developed for multivariate responses, and we specialize it here to the univariate response case. The proposed model assumes that, at each occasion, the categorical response variable depends on the covariates and on a subject-specific time-dependent intercept, which is assumed to be random. The random effects are assumed to be distributed as a categorical latent variable, which evolves over time according to a first-order homogeneous Markov chain. The subject-specific time-dependent intercept has been shown to be able to better take into account unobserved heterogeneity which could be time-varying, compared to more traditional mixed effects models. The fixed effects are directly modeled on the conditional distribution of the response variable, and unobserved heterogeneity is captured by the random effects. Furthermore, even after removing the effects of observed and unobserved heterogeneity, experiencing the event in the past may change the probability of observing the event in the future. In other words, the conditional probability of moving in or out of a state is not only a function of the covariates, but also a function of past experience (Hsiao, 2005). In order to directly measure the effect of having a stable relationship in the past on the probability of having a stable relationship in the present time, we introduce the lagged response as a further explanatory variable. This is the so called state dependence, see Heckman (1981). The final model can be seen as a transition model (Diggle *et al.*, 2002), which is particularly suitable for modelling persistent behaviours.

The organization of paper is as follows. In next section we describe the data and in Section 3 the proposed model. In Section 3.1 we outline an EM type algorithm for fitting the proposed model, which is based on opportune forward-backward recursions. In Section 4 we show the results of fitting the proposed model on the PSID data, and conclude with a brief discussion about the implications of our findings in Section 5.

2 Data

The data we analyze is a subset of the Panel Study of Income Dynamics (PSID), which is a longitudinal survey of a representative sample of U.S. individuals (men, women, and children) and the families in which they reside. The Panel Study of Income Dynamics is primarily sponsored by the National Science Foundation the National Institute of Aging and the National Institute of Child Health and Human Development, and is conducted by the University of Michigan. This database is freely accessible from the website `http://psidonline.isr.umich.edu`, to which we refer for details. The central focus of the data is on income sources and amounts, employment and family composition changes but there are also additional topics (e.g., health, education).

Our response and few covariates of interest were found to be recorded continuously for years from 1990 to 1997. When needed, we carried recoded categorical covariates. We used a closed panel of $n = 1259$ men who were followed for the entire period.

Our dataset is based on young males who were aged between 18 and 35 years in 1991. These subjects are described in terms of different variables measured annually from 1990 to 1997. Hence we have observations for $T = 7$ years, starting in 1991; plus the response at time zero which is used to model the initial probabilities of the latent variables (see below).

The response variable is the marital status, that is a binary variable indicating whether the subject is married or cohabitant in the given year. In our panel we had at most one event for each year, so that the response is well defined. No subject had a marital disruption and (at least officially) found a new spouse within the same year.

The covariates we will consider include: age, total number of children, employment (dummy variable equal to 1 for a man who is employed in a certain year), education (in years of schooling), medical coverage (dummy variable equal to 1 for a man who is covered in a certain year), plus a dummy for each year.

We will evaluate effects of explanatory variables on the response and the effect of the lagged response, which is an indicator of pair stability or instability.

3 The proposed model

Let Y_{it} and X_{it} denote the binary response variable and the corresponding vector of strictly exogeneous covariates, respectively, for the i-th individual at time t, with $i = 1, \dots, n$ and $t = 1, \dots, T$. We also have observed Y_{i0}.

We model the conditional probability of an event for the i-th subject at time t through the following logit reparameterization:

$$\log \frac{p(y_{it} = 1 | x_{it}, \alpha_{it}, y_{i,t-1})}{p(y_{it} = 0 | x_{it}, \alpha_{it}, y_{i,t-1})} = \alpha_{it} + \beta' x_{it} + \gamma y_{i,t-1}. \tag{1}$$

The structural parameters are β, a vector of regression parameters for the exogeneous covariates, and the state dependence γ which directly measures the effect of experiencing the event at the previous time on the response. Finally, α_{it} summarizes the effects of unobserved covariates at time t. These parameters cannot be assumed to be fixed effects, for otherwise we would have the incidental parameter problem. Hence, we assume α_{it} is a random effect.

We assume α_{it} is a time-homogeneous Markov chain with state space $\{\xi_c; c = 1, \dots, k\}$ and in which the initial and transition probabilities are independent from the covariate vector x_{it}. The initial probability vector is allowed to depend on the outcome at time zero through the following logit reparameterization:

$$\log \frac{p(\alpha_{i1} = \xi_c | y_{i0})}{p(\alpha_{i1} = \xi_1 | y_{i0})} = \phi y_{i0} \quad c = 2, \dots, k; i = 1, \dots, n. \tag{2}$$

We denote with Π the transition probability matrix whose generic element is given by

$$\pi_{cd} = p(\alpha_{it} = \xi_d | \alpha_{i,t-1} = \xi_c) \quad c, d = 1, \dots, k; t = 2, \dots, T$$

We let the initial probabilities depend on the outcome at time zero to take into account the initial conditions problem (Heckman, 1981). The initial conditions problem arises since the first available observation may be correlated with the random parameters. This correlation is due to the fact that even this observation is generated from a distribution depending on observable and unobservable covariates which also affect the distribution of the outcome. It can be shown that modelling α_{i1} as a function of y_{i0} overcomes the initial condition problem and removes some of the selection bias. For further details see Hsiao (2005, Sec. 7.5.2).

The distribution of the latent process is then given by

$$p(\alpha_{i,\leq T}) = p(\alpha_{i1}|y_{i0}) \prod_{t>1} p(\alpha_{it}|\alpha_{i,t-1})$$

where $\alpha_{i,\leq T} = (\alpha_{is}, s = 1, \ldots, T)$.

The model we propose is a special case of the class of latent Markov models introduced in Bartolucci and Farcomeni (2009). There are different alternative models for categorical panel data (see Hsiao, 2005). In this paper we prefer the proposed model because it explicitely takes into account the possibility of time-varying unobserved heterogeneity, and does not assume a Gaussian distribution for the random effects. Many models for categorical panel data assume a time-constant random effect (i.e., use $\alpha_{it} = \alpha_i$), which would be restrictive for the data at hand, as we prove below. One common approach to overcome this issue is to use an AR(1) process to model evolution in time of the Normal random effects. We prefer using a latent mixture because it has been shown to approximate well the latent distribution even when it is Normal (Bartolucci and Farcomeni, 2009), and will obviously provide much less bias when the latent distribution is far from being Normal. Further, there are different computational advantages in using a latent Markov process, for instance because there is no need for numerical integration.

3.1 Model fit

Indicating with θ the vectorized model parameters, which is composed by β, γ, ϕ and the off-diagonal elements of the matrix Π, the log-likelihood function can be expressed as:

$$\begin{aligned} l(\theta) &= \sum_{i=1}^{n} \log p(y_{i,\leq T}|x_{i,\leq T}) = \\ &= \sum_{i=1}^{n} \log[\sum_{\alpha_{i1}} \cdots \sum_{\alpha_{iT}} p(\alpha_{i,\leq T}) \prod_t p(y_{it}|\alpha_{it}, x_{it}, y_{i,t-1})] \end{aligned} \tag{3}$$

The log-likelihood involves a telecopic sum and it can obviously not be computed directly unless k and T are very small.

The solution proposed by Bartolucci and Farcomeni (2009) consists in using the EM algorithm (Dempster *et al.*, 1977) and ad-hoc recursions adapted from the hidden Markov literature.

First, we need to derive the complete data log-likelihood, i.e. the log-likelihood that we could compute if we knew the value of the latent process for each subject i and each time t. Let w_{itc} denote a dummy variable equal to 1 if $\alpha_{it} = \xi_c$ and to 0 otherwise.

Let $z_{icd} = \sum_{t>1} w_{i,t-1,c} w_{itd}$ be the number of transitions from state c to state d for the i-th subject. The complete data log-likelihood can be expressed as

$$l^*(\theta) = l_1^*(\beta, \gamma) + l_2^*(\phi) + l_3^*(\Pi) \tag{4}$$

where

$$l_1^*(\beta, \gamma) = \sum_i \sum_c \sum_t w_{itc} \log p(y_{it}|\alpha_{it} = \xi_c, x_{it}, y_{i,t-1}) \tag{5}$$

$$l_2^*(\phi) = \sum_i \sum_c w_{i1c} \log p(\alpha_{i1} = \xi_c | y_{i0}) \tag{6}$$

$$l_3^*(\Pi) = \sum_i \sum_c \sum_d z_{icd} \log \pi_{cd}. \tag{7}$$

The EM algorithm alternates the following steps until convergence.

- The E-step consists of computing the conditional expected value of $l^*(\theta)$, given the observed data and the current value of the parameters, which we denote with $\widetilde{\theta}$. This reduces to computing the conditional expected value of w_{itc} and z_{icd}

  $$w_{itc}(\widetilde{\theta}) = p(\alpha_{it} = \xi_c | x_{i,\le T}, y_{i,\le T})$$

  $$z_{icd}(\widetilde{\theta}) = \sum_{t>1} p(\alpha_{i,t-1} = \xi_c \alpha_{i,t} = \xi_d | x_{i,\le T}, y_{i,\le T})$$

  and substituting it in expression (4).

- The M-step updates the parameter estimates by maximizing separately the three addends of the expected value above with respect to the components of θ.

More formally, the E-step involves forward and backward recursions adapted from the hidden Markov literature (MacDonald and Zucchini, 1997), which are summarized in matrix notation in Appendix A. In Appendix A we also show how to compute (3).

For the M-step, we separately maximize $l_1^*(\beta, \gamma)$, $l_2^*(\phi)$ and $l_3^*(\Pi)$. The first two are maximized using standard iterative algorithm of Newton-Raphson type for logit models. A closed form solution can instead be obtained for $\widehat{\Pi} = \arg\max l_3^*(\Pi)$, whose generic element is in fact given by

$$\widehat{\pi}_{cd} = \frac{\sum_i z_{icd}(\widetilde{\theta})}{\sum_{ic} z_{icd(\widetilde{\theta})}}; \quad c, d = 1, \ldots, k.$$

Standard errors for the maximum likelihood estimators are obtained considering the observed information matrix which is computed as minus the numerical derivative of the score vector used in the M-step.

The EM algorithm is guaranteed to converge to a local maximum for the likelihood. Since the likelihood in latent variable models can be multimodal, we try different starting values for the parameters in order to be more likely to end up with the global maximum. For more details see Bartolucci and Farcomeni (2009).

4 Data Analysis

In analyzing the dataset, the most interesting scientific question concerns the direct effect of employment on marital status. Also of interest are the strength of the state dependence effect and how the response depends on the covariates. The proposed approach allows us to separate these effects from the effect of the unobserved heterogeneity, which is modeled by a latent process. The latent process is time-varying, allowing us to model unobserved effects which may change over time.

We began with a descriptive analysis of our data set. We recall that we considered $n = 1259$ men aged between 18 and 35 years in 1991 with a median and mean age of 29 and 28.5 years respectively (Figure 1(a)). The number of children varies between 0 and 8 with modal and median value equal to 2 (Figure 1(b)). For each year of the panel, we report in Table 1 proportions for binary variables and the boxplot for education in Figure 2.

	91	92	93	94	95	96	97
% married	0.63	0.65	0.69	0.70	0.72	0.73	0.73
% employed	0.86	0.87	0.90	0.90	0.91	0.90	0.91
% medical covered	0.02	0.02	0.02	0.02	0.02	0.02	0.01

Table 1: Descriptive statistics for binary variables per year.

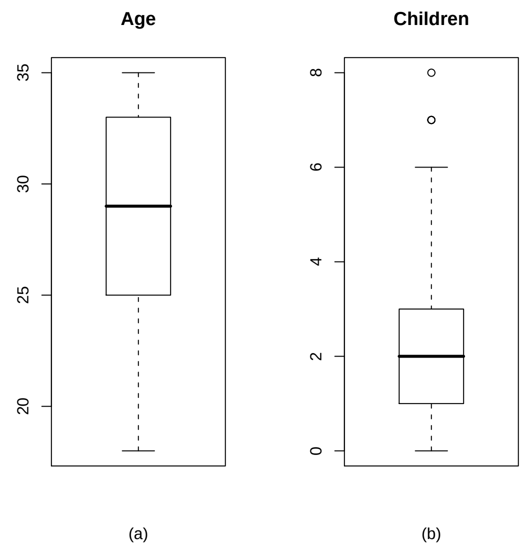

Figure 1: Boxplot of age (a) and number of children (b).

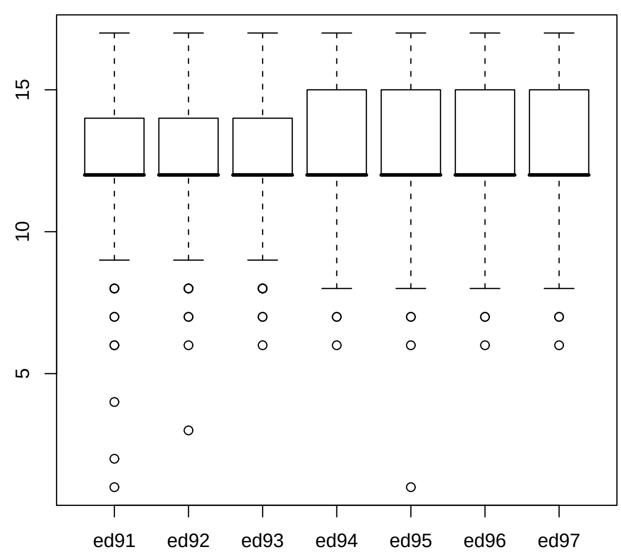

Figure 2: Boxplot of education in each occasion.

At the beginning of the period around 63% of men where cohabitant or married. Since we follow a closed panel of young men, as could be expected at the end of the period this percentage has increased. Same tendency is found for employment. On the contrary, percentages of men with a medical coverage and median education remain constant over time. Globally, the men that at the begin of the panel were and remained not married are 222 while the married men who never separated are 643; the remaining subjects have at least one status transition.

We have examined bivariate cross-sectional relationships between the response variable and the time-dependent covariates. In Table 2 we report the contingency tables for the marital status and employment for each year. The Pearson chi-squared test on each table indicates strong association between the variables in each year, with p-values all of the order of magnitude of 10^{-10} or lower.

We then applied the proposed model to the PSID data described above. In order to choose the number of states of the latent process we repeatedly fit the model for different values of k and then chose the model minimizing the Akaike Information Criterion (AIC). AIC (Akaike, 1973) is defined as $-2l(\widehat{\theta}) + 2g$, where g denotes the number of non-redundant parameters.

Table 3 reports the AIC values for $k = 1, \ldots, 4$ together with the corresponding maximum log-likelihood and number of parameters. Based on these results, we chose a final model with $k = 3$ latent states.

The estimated effects (besides a dummy for each year) for the model with three latent states are reported in Table 4. Medical coverage is the only covariate whose effect is not significant on marital status for the data at hand, besides the year dummies. The other considered covariates all have a significant positive effect on marital status.

As could be expected, subjects elder at entrance in the panel, the probability of being married increases. Elder subjects still tend to favor stable relationships.

We think it is important to underline the other effects. First of all, working increases the probability of being married, with a high log-odds of about 1.17. Note that unobserved heterogeneity and the effects of the other covariates (like age and education) is removed from this log-odds, which then leads to conclude that promoting stable jobs may actually have a positive effect on helping men form stable couples.

The probability of being married increases also with the number of chil-

	Employment					
Marital	1990			1991		
status	no	yes	Total	no	yes	Total
0	138	364	502	115	355	470
1	36	721	757	60	729	789
Total	174	1085	1259	175	1084	1259
Marital	1992			1993		
status	no	yes	Total	no	yes	Total
0	109	331	440	75	317	392
1	54	765	819	47	820	867
Total	163	1096	1259	122	1137	1259
Marital	1994			1995		
status	no	yes	Total	no	yes	Total
0	70	303	373	68	284	352
1	50	836	886	49	858	907
Total	120	1139	1259	117	1142	1259
Marital	1996			1997		
status	no	yes	Total	no	yes	Total
0	68	268	336	61	284	345
1	53	870	923	53	861	914
Total	121	1138	1259	114	1145	1259

Table 2: Contingency tables for marital status and employment per year. Marital status equals one for married or cohabitant subjects, zero otherwise.

dren, which can be deemed as a proxy for tightness of the bond (and probably also for religious beliefs, since in USA couples with strong religious beliefs tend to have an high number of children *and*, as recommended by many religions, do not separate easily). Also a large number of years of education tends to increase the probability of being married. Learned men tend to marry later, and later marriages tend to last longer.

Finally, and most importantly, there is a very high log-odds of around 4.18 for the lagged response variable; indicating a very strong state dependence even after removing the effects of observed and unobserved covariates.

In Table 5 we report the estimates for the three support points and the estimated transition probabilities. The estimated initial probability vector, averaged over all subjects in the sample, is $(0.1835, 0.3008, 0.5157)$.

From a comparison of the three support points in Table 5, we can observe that the first latent state corresponds to men with extremely low propensity of forming a stable union whereas the third latent state corresponds to men with the highest propensity. The second latent state is intermediate, but summarizing a propensity much closer to the third rather than to the first state.

From estimated transition probability matrix we can see that for states 2 and 3 diagonal elements are close to 1, denoting a strong persistence in the state. For individuals in the first state, instead, there is a moderate probability of transition to the second state.

The likelihood ratio statistic for the hypothesis that the transition matrix is diagonal is equal to 46.864, which, on the basis of the approach of Bartolucci (2006), leads us to reject the null hypothesis. This result indicate that a time-constant subject-specific intercept α_i, which corresponds to a latent class model (i.e., a diagonal transition matrix), would be restrictive for the data at hand.

We complete our analysis by showing the estimated average probability for each latent state at each time occasion in Figure 3. We can note that the third state is the mode in each year, but that its probability decreases over time, in favour of the second state whose probability increases. Also for the first state we record a decrease in probability. The consequence is that the proportion of men with very low propensity of forming a stable union tends to decrease over time.

	$k = 1$	$k = 2$	$k = 3$	$k = 4$
log-lik	-1799.9	-1787.8	-1767.4	-1762.9
# par	7	12	19	28
AIC	3613.8	3599.7	3572.7	3581.8

Table 3: Maximum log-likelihood, number of parameters and AIC for different values of k, the number of latent states.

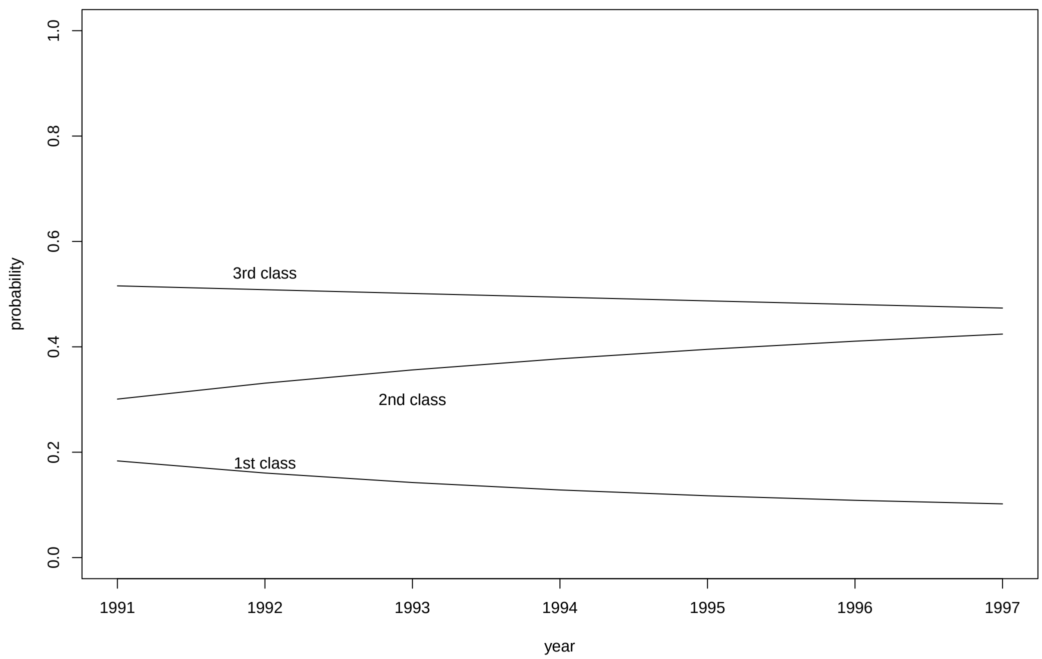

Figure 3: Estimated average probability for each latent state.

4.1 Modeling the state dependence for employment

We repeated the analysis considering as further explanatory variable the lagged employment, retaining unchanged the number of latent states ($k = 3$). The new model has an AIC value equal to 3561.0, which is the lowest observed. We also investigated the possible presence of interactions between covariates and the lagged response variable, finding no significant relationships.

The estimated effects for this model are reported in Table 6. The log-odds estimates are fairly stable, with the only exception of the log-odds for employment. When one considers the effect of the lagged employment, the current employment effect decreases from 1.17 to 0.68. The state dependence for employment is estimated as 0.83, and both effects are significant. Hence, employment status influences the likelihoods of avoiding marital disruption both on the present and on the following year.

For completeness, we report also estimates for the three support points and the transition probabilities, which are quite stable too (Table 7).

5 Discussion

This work is concerned with evaluating the impact of some economic, social and demographic factors on the marital status dynamics which are present in a panel of American young men. We use a latent Markov model in which the random intercept evolves according to a Markov chain and allows us to consider time-varying unobserved heterogeneity. Selection bias is tackled in a unified modeling framework.

We have shown that time-constant subject-specific intercepts α_i would

	estimate	standard error	p-value
age	0.0365	0.0177	0.0392
# children	0.6399	0.1390	<0.0001
employment	1.1688	0.2203	<0.0001
education	0.1154	0.0409	0.0048
medical coverage	-0.4584	0.4462	0.3043
lag marital status	4.1845	0.2363	<0.0001

Table 4: Estimated effects for the model with 3 latent states.

Latent states	Support points	Transition probabilities		
1	-36.1800	0.8048	0.1952	0.0000
2	-5.6943	0.0196	0.9804	0.0000
3	-3.4444	0.0134	0.0006	0.9859

Table 5: Support points for each latent state and transition probability matrix.

	estimate	standard error	p-value
age	0.0320	0.0172	0.0628
# children	0.6714	0.1227	<0.0001
employment	0.6861	0.2464	0.0054
education	0.1122	0.0414	0.0067
medical coverage	-0.4736	0.4482	0.2907
lag employment	0.8345	0.2380	0.0004
lag marital status	4.1808	0.2450	<0.0001

Table 6: Estimated effects for the model with 3 latent states and the lagged employment as further explanatory variable.

have been restrictive for the data at hand. This finding issues a warning to the widely spread use of time-constant random effects in longitudinal modeling. In our opinion researchers should routinelly check this assumption and possibly overcome it with a latent Markov approach like the one we used, or an autoregressive structure in case of normality assumptions on the random effects.

The exceptionally high estimated state dependence for our response variable, after removal of observed and unobserved heterogeneity, denotes a low likelihood of marital disruption, in general, at least for the young men in the first years of a stable relationship. Men entering a stable relationship are likely to hold it. Further, the proportion of men with very low propensity of forming a stable union tends to decrease over time.

As highlighted in Aassve *et al.* (2002) for the probability of movements out of parental home for independent living or marriage, various sources of income affect differently the household formation: own income availability favour marriage or cohabitation while parental income availability delay

Latent states	Support points	Transition probabilities		
1	-37.150400	0.8087	0.1913	0.0001
2	-5.8347	0.0214	0.9786	0.0000
3	-3.5745	0.0134	0.0004	0.9862

Table 7: Support points and transition probability matrix for the model with lagged employment.

them. We agree with their findings and indirectly confirm them. Possibility of easily accessing labour market and obtaining stable employment is a fundamental tool to command own income resources. In our opinion, indeed, the most important finding relates to the effect of employment. Even if we partly removed unobserved heterogeneity and selection bias, we cannot claim a causal effect of employment on marital status. Nevertheless, the very strong association detected suggests that, at least partly, a causal relation could be present. Further, one should note that while political decisions can have direct effects on the labour market, they may not have a direct effect on marital status. In this sense, an intervention boosting stability of employment may be likely to have a beneficial effect also on helping young men establish stable relationships.

Acknowledgements: This work is part of a project of the Ph.D. program in Methodological Statistics of the Department of Statistics, Sapienza - University of Rome. Authors are grateful to Prof. Giovanna Jona Lasinio for organization and support.

References

Aassve A., Betti G., Mazzuco S. and Mencarini L. (2007) Marital Disruption and Economic Wellbeing: A comparative analysis. *Journal of Royal Statistical Society Series A*, **170**, 781-799.

Aassve A., Burgess S., Chesher A. and Propper C. (2002) Transitions from home to marriage of young americans. *Journal of Applied Econometrics*, **17**, 1-23.

Aassve A., Burgess S., Propper C. and Dickson M. (2006) Employment, family union and childbearing decisions in Great Britain. *Journal of Royal Statistical Society Series A*, **169**, 781-804

Akaike, H. (1973) *Information theory as an extension of the maximum likelihood principle*, In *Second International symposium on information theory*, Petrov, B. N. and Csaki F., eds.; Akademiai Kiado, Budapest, 267-281.

Bartolucci F. (2006) Likelihood inference for a class of latent markov models under linear hypotheses on the transition probabilities. *Journal of the Royal Statistical Society, Series B*, **68**, 155-178.

Bartolucci F. and Farcomeni A. (2009) A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, **104**, 816-831.

Covizzi I., (2008). Does Union Dissolution Lead to Unemployment? A Longitudinal Study of Health and Risk of Unemployment for Women and Men Undergoing Separation. *European Sociological Review*, **24(3)**, 347–361.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B*, **39**, 1-38.

Diggle P.J., Heagerty P., Liang K.-Y. and Zeger S.L. (2002). *Analysis of Longitudinal Data*. Oxford University Press, New York.

Gadalla T.M., (2009). Impact of Marital Dissolution on Men's and Women's Incomes: A Longitudinal Study. *Journal of Divorce and Remarriage*, **50**, 55–65.

Jarvis S. and Jenkins S.P., (1999) Marital splits and income changes: Evidence from the British Household Panel Survey. *Population Studies*, **53**, 237–254.

Heckman, J. J. (1981), *The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process*, In *Structural Analysis of Discrete Data*, McFadden D. L. and Manski C. A, eds.; MIT Press, Cambridge, MA.

Hsiao C. (2005). *Analysis of Panel Data*. Cambridge University Press, New York.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and other Models for Discrete-Valued Time Series*. Chapman and Hall, London.

Manting D. and Bouman A.M., (2006) Short- and Long-Term Economic Consequences of the Dissolution of Marital and Consensual Unions. The Example of the Netherlands*. *European Sociological Review* **22(4)**, 413–429.

McManus P.A. and DiPrete T.A. (2001) Losers and winners: the financial consequences of separation and divorce for men. *American Sociological Review*, **66**, 246-268.

Vermunt J.K., Langeheine R. and B??kenholt U. (1999) Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, **24**, 179-207.

Wiggins L.M. (1973) *Panel Analysis: Latent Probability Models for Attitude and Behavior Processes*. Amsterdam: Elsevier.

A Computing the posterior expected values of the latent indicators

In order to compute the posterior probabilities for the E-step of the EM algorithm, and the observed log-likelihood, we make use of the following recursions. We give the recursions in matrix notation for notational convenience and for ease of implementation.

First, we use a forward recursion which consists of computing, for $t = 1, \ldots, T$, the vector

$$q_{it}(y_{i1}, \ldots, y_{it}) = \begin{cases} \mathrm{diag}[\mathrm{u_{i1}}(\mathrm{y_{i1}})]\mathrm{p}(\alpha_{\mathrm{i1}} = \xi_{\mathrm{c}}|\mathrm{y_{i0}}) & \text{if } t = 1, \\ \mathrm{diag}[\mathrm{u_{it}}(\mathrm{y_{it}})]\Pi' \mathrm{q_{it}}(\mathrm{y_{i1}}, \ldots, \mathrm{y_{i,t-1}}) & \text{otherwise,} \end{cases}$$

where $u_{it}(y_{i1}, \ldots, y_{it})$ is a column vector with elements $p(y_{it}|\alpha_{it} = \xi_c, x_{it}, y_{i,t-1})$, $c = 1, \ldots, k$.

The observed log-likelihood (3) for the i-th subject is computed as the sum of the elements of the vector $q_{iT}(y_{i1}, \ldots, y_{iT})$.

After computing $q_{it}(y_{i1}, \dots, y_{it})$, $i = 1, \dots, n$; $t = 1, \dots, T$, consider the matrix $V_{it}(y_{i1}, \dots, y_{iT})$ with elements $p(\alpha_{i,t-1} = \xi_c, \alpha_{it} = \xi_d | x_{i1}, \dots, x_{iT}, y_{i0}, \dots, y_{iT})$, for $c, d = 1, \dots, k$. This matrix may be computed through the following recursion, for $t = 2, \dots, T$:

$$V_{it}(y_{i1}, \dots, y_{iT}) = \frac{\mathrm{diag[q_{i,t-1}(y_{i1}, \dots, y_{i,t-1})]\Pi diag[u_{it}(y_{i1}, \dots, y_{it})]diag[v_{it}(y_{it}, \dots, y_{iT})]}}{p(y_{i,\leq T} | x_{i,\leq T})},$$

where the denominator is the observed log-likelihood for the i-th subject and the vector $v_{it}(y_{it}, \dots, y_{iT})$ is computed through the backward recursion

$$v_{it}(y_{it}, \dots, y_{iT}) = \begin{cases} 1_k, & \text{if } t = T \\ \mathrm{\Pi diag[u_{i,t+1}(y_{i1}, \dots, y_{i,t+1})]v_{i,t+1}(y_{i,t+1}, \dots, y_{iT})}, & \text{otherwise;} \end{cases}$$

where 1_k denotes a vector of ones of size k. The probabilities $w_{itc}(\widetilde{\theta}) = p(\alpha_{it} = \xi_c | x_{i,\leq T}, y_{i,\leq T})$ are then computed as suitable sums of the elements of the matrix $V_{it}(y_{i1}, \dots, y_{iT})$.