
Introduction to latent variable models

Lecture 4

Francesco Bartolucci

Department of Economics, Finance and Statistics

University of Perugia, IT

`bart@stat.unipg.it`

Outline

- Details on recursions and path prediction
- Interesting extensions: multivariate outcomes, individual covariates, multilevel data
- A multivariate approach with covariates in the measurement model with application to labour market
- A multivariate approach with covariates in the measurement model with application to criminal data and evaluation of nursing home performance

Details on the recursions

- In order to *implement the EM algorithm* it is necessary to compute

$$p(u_{t-1}, u_t | \mathbf{y}) = \frac{\sum_{\mathbf{u}^{(t-1,t)}} p(\mathbf{u}) p(\mathbf{y} | \mathbf{u})}{p(\mathbf{y})},$$

with $t = 2, \dots, T$, $u_{t-1}, u_t = 1, \dots, k$

- ▷ $\sum_{\mathbf{u}^{(t-1,t)}}$ is extended to all the possible configurations of the vector $\mathbf{u}^{(t-1,t)} = (u_1, \dots, u_{t-2}, u_{t+1}, \dots, u_T)$

- A *backward recursion* is exploited to compute the above posterior probabilities avoiding the sum
- By a suitable *marginalization* of $p(u_{t-1}, u_t | \mathbf{y})$ we obtain $p(u_t | \mathbf{y})$

Path prediction

- The posterior probabilities $p(u_t|\mathbf{y})$ may be used to *assign a latent state* to a subject at a given time occasion; the state assigned to subject i at occasion t is

$$\hat{u}_{it} : p(U_t = \hat{u}_{it}|\mathbf{y}) = \max_{u_t} p(u_t|\mathbf{y})$$

- More sophisticated is the problem of *path prediction*, i.e. finding the most likely sequence $\hat{\mathbf{u}} = (\hat{u}_{i1}, \dots, \hat{u}_{iT})$ for subject i :

$$\hat{\mathbf{u}}_i : p(U_1 = \hat{u}_{i1}, \dots, U_T = \hat{u}_{iT}|\mathbf{y}) = \max_{\mathbf{u}} p(\mathbf{u}|\mathbf{y})$$

- For this aim we need an *iterative algorithm* due to Viterbi (1967, *IEEE*) and further elaborated by Juan and Rabiner (1991, *Technometrics*)

Extensions of the LM model

- In real applications we may observe more response variables $\mathbf{Y}_t = (Y_{1t}, \dots, Y_{Jt})$ at each time occasion and then we need a *multivariate version of the model*
- In multivariate LM models the *local independence* assumption is usually extended by assuming that
 - ▷ the vectors \mathbf{Y}_t are conditionally independent given the latent process
 - ▷ the elements Y_{jt} of each vector \mathbf{Y}_t are conditionally independent given the corresponding latent variable U_t
- The *interpretation* is still based on considering each U_t as the only explanatory variable of the behavior of a subject at occasion t

- To include individual covariates, the LM model may be extended in two ways:
 1. *covariates in the measurement model*: the latent variables are used to account for the unobserved heterogeneity
 2. *covariates in the distribution of the latent process*: when the main interest is on a latent variable which is measured through the observable response variables (e.g. health status)
- *Multilevel versions* of the LM model are necessary when subjects are collected into clusters (e.g. nursing homes); in many applications it is interesting to study the cluster effect on the distribution of the latent variables (e.g. how the nursing home affects the health status)

Covariates in the measurement model: the extension of Bartolucci & Farcomeni (2009, *Jasa*)

- A *multivariate extension* of the LM model is proposed which is based on:
 - ▷ a *multivariate link function* to parametrize the conditional distribution of the vector of response variables
 - ▷ *covariates and lagged responses* may be included in the measurement model
- The resulting model also represents a generalization of *dynamic logit model* for binary panel data
- The model may be used in the presence of categorical response variables with *two or more categories*, possibly ordered

Multivariate link function

- *Basic notation:*
 - ▷ J : number of response variables observed at each occasion
 - ▷ y_{ijt} : categorical response variable j for subject i at occasion t
 - ▷ l_j : number of categories of y_{ijt} , indexed from 0 to $l_j - 1$
 - ▷ $\mathbf{y}_{it} = \{y_{ijt}, j = 1, \dots, J\}$: vector of response variables for subject i at occasion t
 - ▷ $\boldsymbol{\alpha}_{it}$: vector of time-varying subject-specific effects
- The *conditional distribution* of \mathbf{y}_{it} given $\boldsymbol{\alpha}_{it}$, \mathbf{x}_{it} and $\mathbf{y}_{i,t-1}$ is parametrized by marginal (with respect to other response variables) logits and log-odds ratios which may be of *local*, *global* or *continuation* type

- *Generalized logits* may be of type (with $z = 1, \dots, l_j - 1$):

▷ *local*:
$$\log \frac{p(y_{ijt} = z | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{ijt} = z - 1 | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}$$

▷ *global*:
$$\log \frac{p(y_{ijt} \geq z | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{ijt} < z | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}$$

▷ *continuation*:
$$\log \frac{p(y_{ijt} \geq z | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}{p(y_{ijt} = z - 1 | \boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})}$$
.

- Global and continuation logits are suitable for *ordinal variables*; local logits are commonly used with non-ordered categories
- *Marginal logits and log-odds ratios* are collected in the column vector $\boldsymbol{\eta}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ which has dimension:

$$\sum_j (l_j - 1) + \sum_{j_1 < r} \sum_{j_2 > j_1} (l_{j_1} - 1)(l_{j_2} - 1)$$

- The *vector of marginal effects* may be expressed as (Gloneck & McCullagh, 1995; Colombi & Forcina, 2001; Bartolucci *et al.*, 2007):

$$\boldsymbol{\eta}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) = \boldsymbol{C} \log[\boldsymbol{M} \boldsymbol{p}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})]$$

- ▷ $\boldsymbol{p}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$: probability vector for the conditional distribution of \boldsymbol{y}_{it} given $\boldsymbol{\alpha}_{it}$, \boldsymbol{x}_{it} and $\boldsymbol{y}_{i,t-1}$
 - ▷ \boldsymbol{C} : matrix of contrasts
 - ▷ \boldsymbol{M} : marginalization matrix
- We assume that all the *three and higher order log-linear interactions* for $\boldsymbol{p}(\boldsymbol{y}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ are equal to 0, so that the link function is a one-to-one transformation of this probability vector
 - A simple *Newton algorithm* may be used to obtain $\boldsymbol{p}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$ from $\boldsymbol{\eta}(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$

Parametrization of marginal effects

- We then *assume* that

$$\boldsymbol{\eta}_1(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) = \boldsymbol{\alpha}_{it} + \boldsymbol{X}_{it}\boldsymbol{\beta} + \boldsymbol{Y}_{it}\boldsymbol{\gamma}$$

$$\boldsymbol{\eta}_2(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1}) = \boldsymbol{\delta}$$

- ▷ $\boldsymbol{\eta}_1(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$: subvector containing marginal logits
- ▷ $\boldsymbol{\eta}_2(\boldsymbol{\alpha}_{it}, \boldsymbol{x}_{it}, \boldsymbol{y}_{i,t-1})$: subvector containing marginal log-odds ratios
- ▷ \boldsymbol{X}_{it} : design matrix defined on the basis of \boldsymbol{x}_{it} (e.g. $\boldsymbol{I} \otimes \boldsymbol{x}'_{it}$)
- ▷ \boldsymbol{Y}_{it} : design matrix defined on the basis of $\boldsymbol{y}_{i,t-1}$ (e.g. $\boldsymbol{I} \otimes \boldsymbol{y}'_{i,t-1}$)
- ▷ $\boldsymbol{\beta}$: vector of regression parameters for the covariates
- ▷ $\boldsymbol{\gamma}$: vector of parameters for the lagged responses (serial dependence)
- ▷ $\boldsymbol{\delta}$: vector of association parameters (contemporary association)

Latent Markov chain

- For each i , the random parameter vectors $\{\boldsymbol{\alpha}_{i1}, \dots, \boldsymbol{\alpha}_{iT}\}$ are assumed to follow a (unobservable) *first-order Markov chain* with
 - ▷ states ξ_u , $u = 1, \dots, k$
 - ▷ initial probabilities $\pi_u(\mathbf{y}_{i0})$, $u = 1, \dots, k$
 - ▷ transition probabilities $\pi_{v|u}$, $u, v = 1, \dots, k$
- *Dependence of the initial probabilities* on the initial observations in \mathbf{y}_{i0} is modeled on the basis of the parametrization

$$\boldsymbol{\psi}(\mathbf{y}_{i0}) = \mathbf{Y}_{i0}\boldsymbol{\tau}$$

- ▷ $\boldsymbol{\psi}(\mathbf{y}_{i0})$: column vector of logits $\log[\pi_u(\mathbf{y}_{i0})/\pi_1(\mathbf{y}_{i0})]$, $u = 2, \dots, k$
- ▷ \mathbf{Y}_{i0} : design matrix depending on \mathbf{y}_{i0} , typically $\mathbf{I} \otimes (1 \ \mathbf{y}'_{i0})$
- ▷ $\boldsymbol{\tau}$: vector of parameters

Maximum likelihood estimation

- Estimation is performed by *maximizing the log-likelihood*

$$\ell(\boldsymbol{\theta}) = \sum_i \log \left[\sum_{\boldsymbol{\alpha}_i} p(\boldsymbol{\alpha}_i) \prod_t p(\mathbf{y}_{it} | \boldsymbol{\alpha}_{it}, \mathbf{x}_{it}, \mathbf{y}_{i,t-1}) \right]$$

- $\ell(\boldsymbol{\theta})$ is maximized by an *EM algorithm* (Baum et al., 1970, Dempster et al., 1977) based on the *complete data log-likelihood*:

$$\begin{aligned} \ell^*(\boldsymbol{\theta}) = \sum_i \left\{ \sum_u w_{i1c} \log[\pi_u(\mathbf{y}_{i0})] + \sum_u \sum_v z_{iuv} \log(\pi_{v|u}) + \right. \\ \left. + \sum_t \sum_u w_{itu} \log[p(\mathbf{y}_{it} | \boldsymbol{\xi}_u, \mathbf{x}_{it}, \mathbf{y}_{i,t-1})] \right\} \end{aligned}$$

- ▷ w_{itu} : dummy variable equal to 1 if subject i is in latent state u at occasion t (i.e. $\boldsymbol{\alpha}_{it} = \boldsymbol{\xi}_u$) and to 0 otherwise
- ▷ $z_{iuv} = \sum_{t>1} w_{i,t-1,u} w_{itv}$: number of transitions from state u to v

EM algorithm

- The EM algorithm performs *two steps* until convergence in $\ell(\boldsymbol{\theta})$:
 - E:** compute the conditional expected value of the *complete data log-likelihood* given the current $\boldsymbol{\theta}$ and the observed data
 - M:** maximize this expected value with respect to $\boldsymbol{\theta}$
- Computing the *conditional expected value* of $\ell^*(\boldsymbol{\theta})$ is equivalent to computing the conditional expected value of w_{itu} and $w_{i,t-1,u}w_{itv}$. This is done by certain recursions taken from the literature on hidden Markov models (MacDonald & Zucchini, 1997)
- The parameters in $\boldsymbol{\theta}$ are *updated at the M-step* by simple iterative algorithms. An explicit formula is also available for updating parameters $\pi_{v|u}$

- Standard errors are computed on the basis of the *numerical derivative* of the score vector of the complete data log-likelihood
- A Viterbi algorithm has been implemented for *path prediction* on the basis of the covariates and the observed response configuration of a subject
- A deep *simulation study* has been carried out to assess the performance of the: (i) maximum likelihood estimator, (ii) method to compute standard errors and (iii) selection criteria for k :

$$AIC = -2\ell(\hat{\boldsymbol{\theta}}) + 2\#\text{param.}$$

$$BIC = -2\ell(\hat{\boldsymbol{\theta}}) + \log(n)\#\text{param.}$$

- The *ML estimator* showed a reduced bias and the *method to compute standard errors* showed good performance
- The good performance are maintained when data are generated from a version of the model in which the latent Markov chain is replaced by an *AR(1) process*
- *BIC has better performance than AIC*, since the latter tends to overestimate the number of latent states

An application (PSID data)

- Dataset concerning $n = 1446$ women followed from 1987 to 1993
- Two binary response variables:
 - ▷ *fertility*: equal to 1 if the woman had given birth to a child
 - ▷ *employment*: equal to 1 if the woman was employed
- Eight covariates (beyond a dummy variable for each year):
 - ▷ *race*: dummy variable equal to 1 for a black woman
 - ▷ *age* in 1986
 - ▷ *education* (in year of schooling)
 - ▷ *child 1-2*: number of children in the family between 1 and 2 years
 - ▷ *child 3-5*, *child 6-13*, *child 14-*
 - ▷ *income of the husband* (in dollars)

- The *conditional distribution* of \mathbf{y}_{it} (containing the two response variables) is modelled by two marginal logits and one log-odds ratio:

$$\log \frac{p(y_{i1t} = 1 | \boldsymbol{\alpha}_{it}, \mathbf{x}_{it}, \mathbf{y}_{i,t-1})}{p(y_{i1t} = 0 | \boldsymbol{\alpha}_{it}, \mathbf{x}_{it}, \mathbf{y}_{i,t-1})} = \alpha_{i1t} + \mathbf{x}'_{it} \boldsymbol{\beta}_1 + \mathbf{y}'_{i,t-1} \boldsymbol{\gamma}_1$$

$$\log \frac{p(y_{i2t} = 1 | \boldsymbol{\alpha}_{it}, \mathbf{x}_{it}, \mathbf{y}_{i,t-1})}{p(y_{i2t} = 0 | \boldsymbol{\alpha}_{it}, \mathbf{x}_{it}, \mathbf{y}_{i,t-1})} = \alpha_{i2t} + \mathbf{x}'_{it} \boldsymbol{\beta}_2 + \mathbf{y}'_{i,t-1} \boldsymbol{\gamma}_2$$

$$\begin{aligned} \log \frac{p(y_{i1t} = 1, y_{i2t} = 1 | \boldsymbol{\alpha}_{it}, \mathbf{x}_{it}, \mathbf{y}_{i,t-1})}{p(y_{i1t} = 1, y_{i2t} = 0 | \boldsymbol{\alpha}_{it}, \mathbf{x}_{it}, \mathbf{y}_{i,t-1})} + \\ - \log \frac{p(y_{i1t} = 0, y_{i2t} = 0 | \boldsymbol{\alpha}_{it}, \mathbf{x}_{it}, \mathbf{y}_{i,t-1})}{p(y_{i1t} = 0, y_{i2t} = 1 | \boldsymbol{\alpha}_{it}, \mathbf{x}_{it}, \mathbf{y}_{i,t-1})} = \delta \end{aligned}$$

- The process $\{\boldsymbol{\alpha}_{i1}, \dots, \boldsymbol{\alpha}_{iT}\}$, with $\boldsymbol{\alpha}_{it} = (\alpha_{i1t}, \alpha_{i2t})'$, follows an *homogenous Markov chain* with initial probabilities depending on \mathbf{y}_{i0}

- The model is estimated for an increasing number of latent states (k) from 1 to 5; AIC and BIC lead us to choose $k = 3$:

	k				
	1	2	3	4	5
log-lik.	-6219.0	-6050.0	-6011.5	-6004.7	-5993.6
# par.	37	44	53	64	77
AIC	12512	12188	12129	12137	12141
BIC	12707	12420	12409	12475	12548

- The same selection procedure has been considered for a latent class version of the proposed model (not so clear choice)

	k				
	1	2	3	4	5
log-lik.	-6219.0	-6064.3	-6031.7	-6025.1	-6022.7
# par.	37	42	47	52	57
AIC	12512	12213	12157	12154	12159
BIC	12707	12434	12405	12429	12460

- Estimates of the regression parameters ($k = 3$):

Effect	logit fertility	logit employment	log-odds ratio
intercept (average of support points)	-2.310	2.285	-1.238
race	-0.238	0.175	-
age	-0.219	0.060	-
age ² /100	-1.117	-0.110	-
education	0.158	0.085	-
child 1-2	0.180	-0.113	-
child 3-5	-0.382	-0.168	-
child 6-13	-0.613	0.027	-
child 14-	-0.891	0.071	-
income of the husband/1000	0.002	-0.011	-
lagged fertility	-1.476	-0.726	-
lagged employment	0.321	0.970	-

- Negative state dependence for fertility, positive state dependence for employment and negative association between the response variables

- Estimated support points for each latent state

Latent state	Support points	
	Fertility	Employment
1	-1.349	-5.358
2	-1.858	-1.066
3	-2.505	2.205

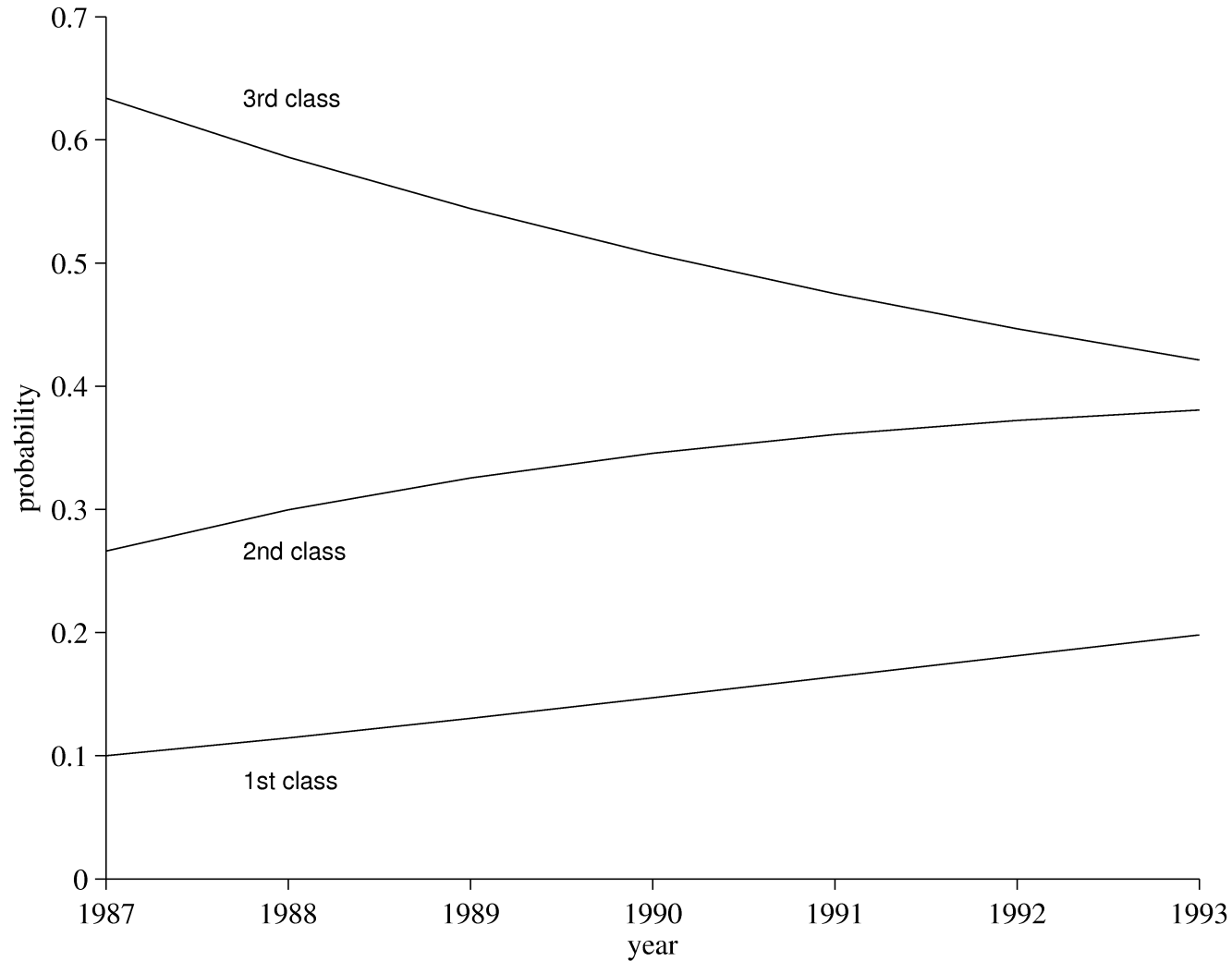
- Interpretation of the latent classes:
 - ▷ the 1st class is that of women with the highest propensity for fertility and the lowest for employment
 - ▷ the opposite for the 3rd class
 - ▷ the 2nd class has intermediate levels of propensity for both fertility and employment

- Estimated initial probability vector and transition probability matrix (averaged over all the subjects in the sample)

$$\hat{\boldsymbol{\pi}} = \begin{pmatrix} 0.010 \\ 0.266 \\ 0.634 \end{pmatrix}, \quad \hat{\mathbf{\Pi}} = \begin{pmatrix} 0.947 & 0.050 & 0.003 \\ 0.068 & 0.888 & 0.044 \\ 0.003 & 0.092 & 0.906 \end{pmatrix}$$

- The hypothesis that the transition matrix is diagonal must be rejected with a likelihood ratio statistic equal to 32.079 (p -value < 0.001)
- The assumption that the parameters in α_{it} for the unobserved heterogeneity are time-constant is restrictive; under this hypothesis the estimates of the association parameters are considerably different

- Estimated average probability of each latent state at every time occasion



- Estimates of the (contemporary, serial) association parameters for different values of k

	Effect	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
fertility	lag fert.	-1.476**	-1.469**	-1.482**	-1.452**	-1.499**
	lag empl.	-0.163	0.212	0.444**	0.443**	0.427**
employment	lag fert.	-0.478**	-0.733**	-0.704**	-0.654**	-0.747**
	lag empl.	2.949**	1.571**	1.008**	1.079**	0.746**
association	intercept	-1.213**	-1.286**	-1.130**	-1.651**	-1.173**

- Estimates of the (contemporary, serial) association parameters for different values of k under the latent class version of the model

	Effect	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
fertility	lag fert.	-1.476**	-1.459**	-1.462**	-1.503**	-1.575**
	lag empl.	-0.163	-0.018	0.118	0.034	0.005
employment	lag fert.	-0.478**	-0.681**	-0.617**	-0.677**	-0.680**
	lag empl.	2.949**	2.061**	1.791**	1.751**	1.753**
association	intercept	-1.213**	-1.302**	-1.227**	-1.300**	-1.325**

- As an application of the Viterbi algorithm we consider a:
 - ▶ white woman in the sample who was 27 years old in 1986
 - ▶ with 12 years of education and no children in the same year
 - ▶ having a husband with income between 10,000\$ and 21,000\$ in the period of interest
 - ▶ no children in 1987 and 1993
 - ▶ having job position in 1987 and 1988 and continuously from 1991 to 1993, i.e. $\mathbf{y} = (1, 1, 0, 0, 1, 1, 1)$
- The predicted sequence of latent states is 3, 3, 2, 2, 2, 2, 2
- Overall, it results that 78.5% of the women started and persisted in the same latent state for the entire period, whereas for the 21.5% of the women we had one or more transitions between states

Covariates in the model for the latent process (1/2)

- The simplest case is when there is one *individual covariate* X which is categorical (e.g. gender)
- In this case we have to *estimate the initial and transition probabilities* which are specific for each level of X :

$$\pi_{u|x} = p(U_1 = u|x), \quad u = 1, \dots, k,$$

$$\pi_{v|u,x} = p(U_{t-1} = u, U_t = v|x), \quad u, v = 1, \dots, k, \quad t = 2, \dots, T$$

- For the analysis of a dataset concerning crimes of young people, Bartolucci *et al.* (2009, *JRSS-A*) considered a model which relies on a *non-homogenous Markov chain*, i.e. time-specific trans. probabilities:

$$\pi_{v|u,x}^{(t)} = p(U_{t-1} = u, U_t = v|x), \quad u, v = 1, \dots, k, \quad t = 2, \dots, T$$

Application to Criminal dataset

- Application based on the *England and Wales Offenders Index dataset*, which has the following features:
 - ▷ a court based record of the criminal histories of all offenders in England and Wales from 1963 to the current days
 - ▷ data concerns $n = 11,402$ individuals (9,232 males, 2,170 females) paroled from the cohort of those born in 1953, and followed through to 1993
 - ▷ Offences are combined into $J = 10$ major categories described in the Offenders Index Codebook (1998)
 - ▷ As in Francis *et al.* (2004) we have defined $T = 6$ time windows or age strips: 10-15, 16-20, 21-25, 26-30, 31-35, 36-40
 - ▷ the only covariate is gender (1=male, 2=female)

- The ten offence groups

<i>j</i>	Offence group
1	<i>Violence against the person</i>
2	<i>Sexual offences</i>
3	<i>Burglary</i>
4	<i>Robbery</i>
5	<i>Theft and Handling stolen goods</i>
6	<i>Fraud and Forgery</i>
7	<i>Criminal damage</i>
8	<i>Drug offences</i>
9	<i>Motoring offences</i>
10	<i>Other offences</i>

- Results of a preliminary fitting of a LM model with:
 - ▷ non-homogeneous Markov chain
 - ▷ different initial and transition probabilities for males and females
 - ▷ success probabilities of the first class constrained to zero

k	max. log-lik	#param.	BIC
1	-166,294	10	332,696
2	-114,297	32	228,941
3	-110,072	84	221,058
4	-108,694	156	219,084
5	-107,871	248	218,439
6	-107,399	360	218,713

- Results of fitting a constrained model with:
 - ▷ $k = 5$ latent states
 - ▷ common transition matrices for males and females
 - ▷ use of two transition matrices: the first until occasion \bar{t} and the second after occasion \bar{t} (partial homogeneity of the Markov chain)

\bar{t}	$\hat{\ell}_{\bar{t}}$	$r_{\bar{t}}$	$BIC_{\bar{t}}$
2	-108,470	88	217,896
3	-108,533	88	218,023
4	-108,541	88	218,039
5	-108,619	88	218,195

- The selected model is with $\bar{t} = 2$ (common transition probabilities after the second occasion)

- Estimated conditional probabilities of conviction for any offence group and any latent class

Offence group		Latent class				
		1	2	3	4	5
1	<i>Violence against the person</i>	0.000	0.003	0.158	0.018	0.227
2	<i>Sexual offences</i>	0.000	0.003	<i>0.029</i>	0.003	0.026
3	<i>Burglary</i>	0.000	0.032	0.006	0.016	0.487
4	<i>Robbery</i>	0.000	0.000	0.005	0.002	<i>0.039</i>
5	<i>Theft and Handling stolen goods</i>	0.000	0.096	0.067	0.546	0.777
6	<i>Fraud and Forgery</i>	0.000	0.000	0.019	0.130	0.149
7	<i>Criminal Damage</i>	0.000	0.016	0.091	0.010	0.233
8	<i>Drug offences</i>	0.000	0.000	0.075	0.016	<i>0.099</i>
9	<i>Motoring offences</i>	0.000	0.000	0.005	0.003	<i>0.044</i>
10	<i>Other offences</i>	0.000	0.000	0.060	0.039	0.347

Latent class description		Latent class description	
1	<i>Non offenders</i>	4	<i>Theft and fraud offenders</i>
2	<i>Incidental offenders</i>	5	<i>High frequency and varied offenders</i>
3	<i>Violent offenders</i>		

- Estimated initial probabilities for males and females

	$\hat{\pi}_{u 1}$	$\hat{\pi}_{u 2}$
1	0.496	0.963
2	0.472	0.020
3	0.000	0.000
4	0.000	0.016
5	0.033	0.000

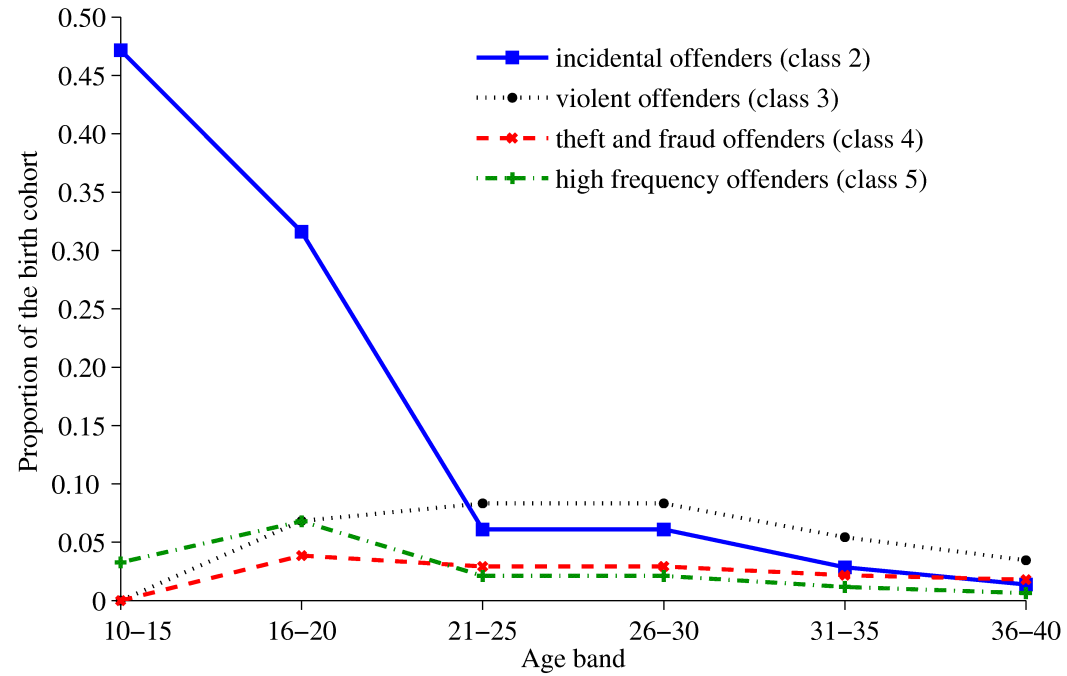
- Estimated transition probabilities from age band 10-15 to age band 16-20 for both males and females

u	v				
	1	2	3	4	5
1	0.960	0.009	0.003	0.028	0.000
2	0.068	0.648	0.140	0.053	0.092
3	1.000	0.000	0.000	0.000	0.000
4	0.846	0.000	0.005	0.149	0.000
5	0.040	0.175	0.033	0.000	0.753

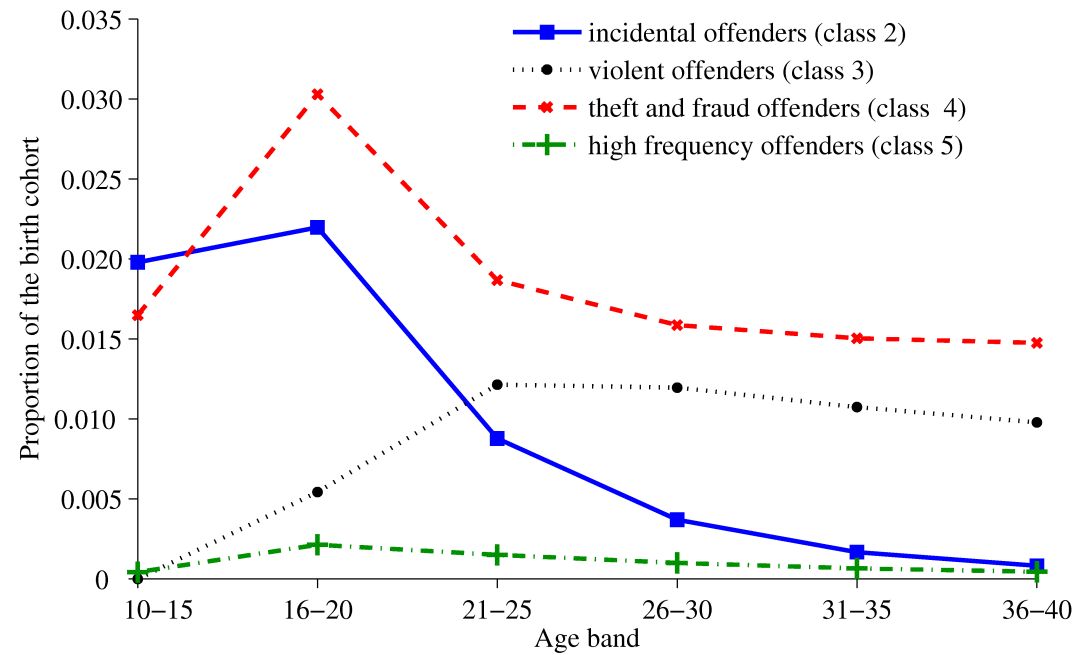
- Estimated transition probabilities from one age band to another for male and female offenders over 16

u	v				
	1	2	3	4	5
1	0.993	0.000	0.005	0.012	0.000
2	0.327	0.371	0.206	0.078	0.018
3	0.583	0.000	0.395	0.015	0.007
4	0.809	0.001	0.008	0.180	0.002
5	0.000	0.276	0.217	0.039	0.468

males



females



Covariates in the model for the latent process (2/2)

- With *continuous covariates*, a suitable parametrization of the initial and transition probabilities must be adopted
- The simplest parametrization is based on *multinomial logits*:

$$\log \frac{\pi_{iu}}{\pi_{i1}} = \mathbf{x}'_{i1} \boldsymbol{\beta}_u, \quad u = 2, \dots, k$$

$$\log \frac{\pi_{iv|u}^{(t)}}{\pi_{iu|u}^{(t)}} = \mathbf{x}'_{it} \boldsymbol{\gamma}_{uv}, \quad u, v = 2, \dots, k, u \neq v$$

- For the analysis of dataset concerning the *health condition of sample of elderly people* hosted in nursing homes, Bartolucci *et al.* (2009, AOAS) exploited a model based on a parametrization which constrains the latent states to be ordinal

Application on the dataset about elderly people

- Data about 1,093 elderly people, admitted in 2003 to 11 nursing homes in Umbria, who responded to 9 items about their health status:

Item	%
1 [CC1] Does the patient show problems in recalling what recently happened (5 minutes)?	72.6
2 [CC2] Does the patient show problems in making decisions regarding tasks of daily life?	64.2
3 [CC3] Does the patient have problems in being understood?	43.9
4 [ADL1] Does the patient need support in moving to/from lying position, turning side to side and positioning body while in bed?	54.4
5 [ADL2] Does the patient need support in moving to/from bed, chair, wheelchair and standing position?	59.0
6 [ADL3] Does the patient need support for eating?	28.7
7 [ADL4] Does the patient need support for using the toilet room?	63.5
8 [SC1] Does the patient show presence of pressure ulcers?	15.4
9 [SC2] Does the patient show presence of other ulcers?	23.1

- Responding 1 to each item is sign of bad health conditions, whereas 0 is sign of good health conditions
- The available covariates are: (i) gender (0 = male, 1 = female), (ii) 11 dummies for the nursing homes, and (iii) age
- Some summary statistics referred to patients and nursing homes:

Variables	min	mean	max
n.patients	55.00	99.36	177.00
% males	20.30	33.58	40.70
age	32.00	80.69	102.00
#occasions	1.00	4.67	20.00

- The panel is unbalanced since the number of interview occasions is not the same for all subjects (T_i = number of occasions for subject i)

Model assumptions

- *Local independence* (for multivariate model), so that the variables in $\mathbf{Y}_{it} = (Y_{i1t}, \dots, Y_{iJt})$ are conditionally independent given the corresponding latent variable U_{it}
- About the *conditional distribution* of Y_{ijt} given U_{it} :

$$p(Y_{ijt} = 1 | U_{it} = u, \mathbf{x}_{it}) = \lambda_{ju}, \quad i = 1, \dots, n, \quad t = 1, \dots, T_i$$

We also assume monotonicity:

$$\lambda_{j1} \leq \lambda_{j2} \leq \dots \leq \lambda_{jk}, \quad j = 1, \dots, J,$$

so that the latent states result ordered (they are decreasing ordered in terms of goodness of health condition)

- *Initial probabilities* of the latent process are modeled via global logits

$$\log \frac{\pi_{i1} + \cdots + \pi_{iu}}{\pi_{i,u+1} + \cdots + \pi_{ik}} = \alpha_u + \mathbf{x}'_{i1} \boldsymbol{\beta}, \quad i = 1, \dots, n, \quad u = 1, \dots, k-1,$$

which takes into account the ordinal nature of the states as in the *proportional odds model* of McCullagh (1980)

- *Transition probability matrix* is assumed to be tridiagonal, e.g. with $k = 5$ states

$$\mathbf{\Pi}_i^{(t)} = \begin{pmatrix} \pi_{i1|1}^{(t)} & \pi_{i2|1}^{(t)} & 0 & 0 & 0 \\ \pi_{i1|2}^{(t)} & \pi_{i2|2}^{(t)} & \pi_{i3|2}^{(t)} & 0 & 0 \\ 0 & \pi_{i2|3}^{(t)} & \pi_{i3|3}^{(t)} & \pi_{i4|3}^{(t)} & 0 \\ 0 & 0 & \pi_{i3|4}^{(t)} & \pi_{i4|4}^{(t)} & \pi_{i5|4}^{(t)} \\ 0 & 0 & 0 & \pi_{i4|5}^{(t)} & \pi_{i5|5}^{(t)} \end{pmatrix}.$$

- A tridiagonal transition matrix has an advantage in terms of parsimony of the model; we also assume:

$$\log \frac{\pi_{i,u-1|u}^{(t)}}{\pi_{iu|u}^{(t)}} = \mathbf{x}'_{it} \gamma_{u1} \quad \text{logit for the probability of improving}$$

$$\log \frac{\pi_{i,u+1|u}^{(t)}}{\pi_{iu|u}^{(t)}} = \mathbf{x}'_{it} \gamma_{u2} \quad \text{logit for the probability of worsening}$$

- A constrained version of the model assumes:

$$\gamma_{u1} = \gamma_1, \quad \gamma_{u2} = \gamma_2, \quad u = 1, \dots, k$$

since the vectors of covariates \mathbf{x}_{it} include dummies for the nursing homes, the corresponding coefficients may be used for ranking

- The model may be seen as a multilevel LM model with fixed rather than random effects for the clusters

- Model selection based on BIC:

Model	k	#param.	$\hat{\ell}$	BIC
M1: unrestricted LM model	1	9	-27824	55769
	2	59	-18992	38397
	3	97	-17126	34931
	4	135	-15880	32705
	5	173	-15188	31586
	6	211	-14893	31262
	7	249	-14660	31063
	8	287	-14568	31143
M2: based on restrictions $\gamma_{u1} = \gamma_1$ and $\gamma_{u2} = \gamma_2$	7	109	-14868	30499
M3: M2 + no gender effect on initial prob.	7	108	-14870	30495
M4: M2 + no age effect on initial prob.	7	108	-14888	30531
M5: M2 + no nursing home effect on initial prob.	7	99	-14926	30544
M6: M2 + no gender effect on transition prob.	7	107	-14870	30490
M7: M2 + no age effect on transition prob.	7	107	-14870	30489
M8: M2 + no time effect on transition prob.	7	107	-14885	30518
M9: M2 + no nursing home effect on transition prob.	7	89	-14982	30587
M10: M2 + restrictions defining M3, M6, M7	7	104	-14875	30478

- Estimated conditional response probabilities given the latent states

Latent state (u)	Item (j)								
	1-CC1	2-CC2	3-CC3	4-ADL1	5-ADL2	6-ADL3	7-ADL4	8-SC1	9-SC2
state1	0.000	0.000	0.000	0.027	0.016	0.000	0.036	0.023	0.068
state2	0.747	0.000	0.000	0.027	0.016	0.011	0.089	0.026	0.068
state3	0.747	0.725	0.290	0.027	0.016	0.011	0.199	0.026	0.089
state4	0.747	0.725	0.290	0.824	0.950	0.273	0.968	0.117	0.224
state5	0.997	0.999	0.982	0.824	0.950	0.273	0.997	0.117	0.224
state6	0.997	0.999	0.982	1.000	1.000	0.927	0.997	0.117	0.250
state7	0.997	0.999	0.982	1.000	1.000	0.927	0.997	1.000	1.000

- Estimates of the regression parameters for the initial probabilities

	estimate	s.e.	<i>t</i> -stat.	<i>p</i> -value
intercept 1	-0.393	0.054	-7.306	0.000
intercept 2	-1.584	0.093	-16.960	0.000
intercept 3	-2.865	0.121	-23.765	0.000
intercept 4	-3.488	0.131	-26.589	0.000
intercept 5	-4.731	0.162	-29.188	0.000
age	0.040	0.006	7.208	0.000
dummy 1	-1.190	0.443	-2.685	0.007
dummy 2	-1.006	0.482	-2.084	0.037
dummy 3	-1.172	0.501	-2.338	0.019
dummy 4	-1.234	0.483	-2.554	0.011
dummy 5	-0.965	0.481	-2.007	0.045
dummy 6	-0.801	0.510	-1.572	0.116
dummy 7	-1.229	0.531	-2.315	0.021
dummy 8	-1.859	0.466	-3.989	0.000
dummy 9	-0.369	0.465	-0.794	0.427
dummy 10	-2.538	0.519	-4.893	0.000
dummy 11	-1.208	0.485	-2.490	0.013

- Averaged over all subjects we have the initial probability vector

$$\hat{\pi} = (0.133, 0.048, 0.217, 0.282, 0.111, 0.134, 0.073)'$$

- Estimates of the regression parameters for the transition probabilities

	Improvement effect (γ_1)				Worsening effect (γ_2)			
	estimate	s.e.	<i>t</i> -stat.	<i>p</i> -value	estimate	s.e.	<i>t</i> -stat.	<i>p</i> -value
time	-0.009	0.002	-5.121	0.000	-0.003	0.002	-2.020	0.043
dummy 1	-3.295	0.372	-8.869	0.000	-3.317	0.304	-10.903	0.000
dummy 2	-3.677	0.352	-10.447	0.000	-2.632	0.201	-13.076	0.000
dummy 3	-2.703	0.270	-10.004	0.000	-2.876	0.258	-11.161	0.000
dummy 4	-2.218	0.326	-6.794	0.000	-3.393	0.487	-6.964	0.000
dummy 5	-0.112	0.298	-0.378	0.705	-1.053	0.363	-2.902	0.004
dummy 6	-0.282	0.383	-0.735	0.462	-1.094	0.448	-2.445	0.014
dummy 7	0.025	0.346	0.074	0.941	-0.925	0.384	-2.410	0.016
dummy 8	-4.123	0.638	-6.463	0.000	-3.092	0.330	-9.378	0.000
dummy 9	-1.990	0.338	-5.883	0.000	-1.686	0.243	-6.942	0.000
dummy 10	-2.951	1.037	-2.846	0.004	-2.436	0.560	-4.346	0.000
dummy 11	-3.025	0.503	-6.018	0.000	-2.427	0.316	-7.675	0.000

- Scores computed as difference between improving and worsening scores for each nursing home:

h	Unidimensional score			
	estimate	s.e.	t -stat.	p -value
1	0.120	0.412	0.291	0.771
2	-0.947	0.379	-2.500	0.012
3	0.271	0.385	0.704	0.481
4	1.273	0.528	2.409	0.016
5	1.039	0.342	3.034	0.002
6	0.911	0.457	1.994	0.046
7	1.049	0.385	2.722	0.006
8	-0.933	0.577	-1.619	0.106
9	-0.206	0.341	-0.604	0.546
10	-0.417	1.078	-0.387	0.699
11	-0.500	0.529	-0.946	0.344

