Dynamic sequential analysis of careers

Fulvia Pennoni

Department of Statistics and Quantitative Methods

University of Milano-Bicocca

http://www.statistica.unimib.it/utenti/pennoni/

Email: fulvia.pennoni@unimib.it

joint work with R. Piccarreta

Outline

- Introduction: the survey
- Target: transition from school to work using sequences related to activities experienced by individuals between the ages of 16 and 22
- Baseline covariates (measured at the age of 16) describing individuals' and their family's characteristics
- Inferential analysis of the relation between the entire careers and the baseline covariates based on the Latent Markov (LM) model
- Main results

Introduction

- The interest is to study life courses represented as sequences of states describing the activities experienced by young individuals

- The goal is to identify factors exposing young people to risk patterns i.e. careers dominated by unemployment, as well as factors *"which encourage young people to remain in full-time education after the compulsory stage"* (Amstrong, 1999)

- The survey data are about a representative sample of 712 young people living in Northern Ireland who became eligible to leave school for the first time in July 1993

- Young people in the education system in Northern Ireland were streaming by the ability at age 11, with the upper third of each cohort entering the grammar school

Introduction

- The survey is made of two waves: the first made in 1995, and the second in 1999
- The data at the second waves are related to the transitions from school to work to identify those sequences if any risk patterns dominated by unemployment
- The monthly activities have been collected from 1993 to 1999 when they were first eligible to leave compulsory education (at the age of 16, July 1993) and an additional state, higher education, was added, possibly experienced after the age of 18
- The sample has been stratified according to the choice made by individuals at the end of compulsory school (school, further education, vocational training, and unemployment)

Sequential histories

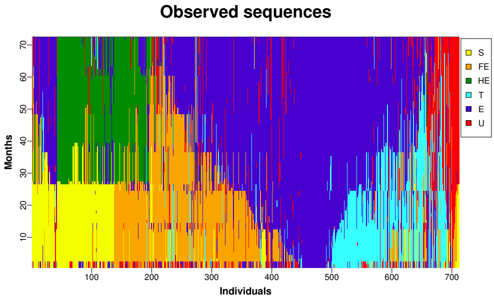
- The individual careers were built considering the monthly activities experienced by individuals during the $T = 72$ months:
  - Unemployment (U)
  - still being at School (S)
  - attending a Further Education college (FE)
  - Higher Education (HE)
  - being in Training (T)
  - Employed (E)

- According to the descriptive statistics the main transitions are those from school to higher education; from further education to employment and from training to employment

Sequential histories

- The observed patterns are reported in the figure, where individuals are arranged on the horizontal axis using a seriation procedure illustrated in Piccarreta and Lior (2010)

- The colours are related to the activity experienced in each month



Individual covariates

- The following set of (binary) background variables was collected at the age of 16 (in July 1993, first month of observation)

Covariate		%
Male		47.31
Catholic		43.99
Region	*Belfast*	11.25
	South	27.49
	SouthEast	10.61
	West	14.96
Grammar school		16.47
Father unemployed		14.96
High grades at school		33.25
Father manager		22.38
Cohabiting with both parents		57.67

- McVicar and Anyadike-Danes (2002) propose a multi-step approach: the dissimilarity between each pair of careers is measured using an alignment technique; then cluster analysis is applied to obtain groups of individuals with similar careers; lastly, a multinomial logistic regression is used to relate the probability of experiencing careers in each cluster to the explanatory variables

Individual covariates

- On the other hand, within our approach we suppose that the response variables indirectly measure a latent trait which may represent for example attitudes, aspiration which may evolve over time also according to experiences
- As argued by Amstrong (1999) attitudes and aspiration *"are likely to have an important influence on staying on over and above the observed covariates"*
- Therefore, we propose a LM model where the available covariates directly influence the latent part of the model in order to investigate their influence on these latent trait

LM model with covariates

- The LM model with covariates affecting the latent Markov model (Bartolucci, Farcomeni, Pennoni 2013) has the following notation:
  - n: sample size
  - T: number of time occasions
  - Y_{it} categorical response variable for subject i at occasion t where $i = 1, \ldots, n, \;\; t = 1, \ldots, T,$
  - $\boldsymbol{Y}_t$: vector of response variables for each occasion

- For every i, the response variables in $\boldsymbol{Y}_i$ are conditionally independent given the *latent process* $V = (V_1, \ldots, V_T)$: *local independence* assumption

- The latent process V follows a *first-order homogeneous Markov chain* with discrete state space

Measurement model parameters

- The parameters of the *measurement model* (conditional distribution of each $\boldsymbol{Y}_t$ given $\boldsymbol{V}$) are:

$$\phi_{y|v} = f_{Y_t|V_t}(y|v),$$

for $t = 1, \ldots, T, v = 1, \ldots, k$

- The *individual covariates* $\boldsymbol{X}$ (time constant or varying) enter the model for the latent process

Inclusion of the covariates in the latent model

- The parameters of the *model for the latent process* (distribution of V_t) are:
  - *initial probabilities*

$$\pi_{v|\boldsymbol{x}} = f_{V_t|\boldsymbol{X}}(v|\bar{v}), \quad v = 1, \ldots, k$$

  - *transition probabilities* (may depend on time):

$$\pi_{v|\bar{v}\boldsymbol{x}} = f_{V_t|V_{(t-1)},\boldsymbol{X}}(v|\bar{v}, \boldsymbol{x}), \quad t = 2, \ldots, T, \; v, \bar{v} = 1, \ldots, k$$

  where:

  $\boldsymbol{x}$ denotes a realization of $\boldsymbol{X}$
  v denotes a realization of V_t
  $\bar{v}$ denotes a realization of $V_{(t-1)}$

Inclusion of the covariates in the latent model

- Inclusion of covariates is allowed by the adoption of suitable *link functions* for the initial and the transition probabilities

- We use *the formulation* which assumes the multinomial logit model for the initial probabilities

$$\log \frac{p(V_1 = v|\boldsymbol{X} = \boldsymbol{x})}{p(V_1 = 1|\boldsymbol{X} = \boldsymbol{x})} = \log \frac{\pi_{v|\boldsymbol{x}}}{\pi_{1|\boldsymbol{x}}} = \beta_{0v} + \boldsymbol{x}'\boldsymbol{\beta}_{1v},$$

- and the following for the transition probabilities

$$\log \frac{p(V_t = v|V_{(t-1)} = \bar{v}, \boldsymbol{X} = \boldsymbol{x})}{p(V_t = 1|V_{(t-1)} = \bar{v}, \boldsymbol{X} = \boldsymbol{x})} = \log \frac{\pi_{v|\bar{v}\boldsymbol{x}}}{\pi_{v|\bar{v}\boldsymbol{x}}} = \gamma_{0\bar{v}v} + \boldsymbol{x}'\boldsymbol{\gamma}_{1\bar{v}v},$$

where $v \geq 2$, for the first and the second equation

Maximum likelihood estimation

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

$$\ell(\boldsymbol{\theta}) = \sum_i \log \left[f_{\tilde{\boldsymbol{Y}}|\tilde{\boldsymbol{X}}}(\tilde{\boldsymbol{y}}_i|\tilde{\boldsymbol{x}}_i) \right]$$

  - $\tilde{\boldsymbol{x}}_i$ and $\tilde{\boldsymbol{y}}_i$ vectors observed data $i = 1, \ldots, n$
  - $\boldsymbol{\theta}$: vector of all model parameters

- $\ell(\boldsymbol{\theta})$ is efficiently computed by using backward *recursions* also employed in the hidden Markov literature

- Likelihood maximization is performed by an EM algorithm (Baum *et al.*, 1970, Dempster *et al.*, 1977) based on the *complete data log-likelihood* ($\ell^*(\boldsymbol{\theta})$), i.e. the log-likelihood that we could compute if we knew the latent state of each subject at every occasion

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 $\ell^*(\boldsymbol{\theta})$ at the *E-step* is equivalent to computing the *posterior probabilities* of each latent state at every occasion

- The parameters in $\boldsymbol{\theta}$ are updated at the *M-step* by simple iterative algorithms (explicit formulae are occasionally available)

- The model is estimated by using the `R` library named `LMest`[1] (Bartolucci, Pandofi, Pennoni, 2017)

[1] V2.4, available from `http://CRAN.R-project.org/package=LMest`

Path prediction

- By using an adapted version of the *Viterbi* (1967) algorithm a forward recursion computes the joint conditional probability of the latent states given the responses and the covariates $\hat{f}_{\boldsymbol{V}|\tilde{\boldsymbol{X}},\tilde{\boldsymbol{Y}}}(\boldsymbol{v}|\tilde{\boldsymbol{x}},\tilde{\boldsymbol{y}})$ according to the ML estimates $\hat{\boldsymbol{\theta}}$

- The optimal global state is found by means of computations and maximisations of the joint posterior probabilities

- the function `est_lm_cov_latent` of the package LMest is employed to estimate the model parameters

LM likelihood parameter estimates

- We applied the function `search.model.LM` to select the value of k on the basis of the observed data
- When the model selection is performed according to the Bayesian Information Criterion $k = 5$ *latent states* are selected
- We show the maximum log-likelihood at convergence, the number of parameters, and AIC and BIC indices for models from M_1 to M_6

Model	$\hat{\ell}$	#par	*BIC*	*AIC*
M_1	-79949.82	5	159932.47	159909.63
M_2	-50437.18	46	101176.50	100966.37
M_3	-33479.81	111	67688.69	67181.63
M_4	-22671.54	200	46656.69	45743.08
M_5	-16049.34	313	34154.48	32724.67
M_6	-15725.62	450	34406.88	32351.25

Results

- According to the *estimated conditional probabilities* the *latent states may be characterized as*

  "School $\rightarrow V_S$"
  "Employment $\rightarrow V_E$"
  "Further Education $\rightarrow V_{FE}$"
  "Jobless or Training $\rightarrow V_{J\&T}$"
  "Higher Education $\rightarrow V_{HE}$"

- $V_{U\&T}$ refers to an alternation between unemployment and firm training, consistently with the observed data structure

- The estimates of the β parameters allow evaluating the impact of covariates on the initial probabilities, that is the probability of starting with a state different from the reference latent state, here V_S

Results

- For the sake of illustration, we report below the estimates of the γ's coefficients measuring the impact of the baseline covariates on the probability of transitioning from $V_{U\&T}$ to the other latent states rather than to the reference category V_S

$\hat{\gamma}_{1v_{U\&T}v}$	V_E	V_{FE}	$V_{U\&T}$	V_{HE}
Intercept	-9.09^{**}	-1.08	-3.91^{**}	13.74^{**}
Male	-0.37	0.08	-0.23	-0.14
Catholic	0.39	-0.36^{**}	$-0.55^{\dagger}$	-0.47
Belfast	0.29	$-0.30^{\dagger}$	-0.42	-8.58
South	-1.43^{*}	-0.23	0.23	-1.00
SouthEast	-0.03	-0.24	-0.17	-8.06
West	0.18	-0.27	0.13	0.71
Grammar school	1.09^{**}	0.09	0.16	$1.51^{\dagger}$
Father unemployed	0.14	$-0.38^{\dagger}$	-0.07	-7.75
High grades at school	1.09^{**}	0.11	0.88^{**}	1.93^{*}
Father manager	$0.69^{\dagger}$	-0.08	0.20	0.85
Cohabiting with both parents	0.60	-0.06	-0.20	-0.25

Results

- The following are the estimates of the γ's coefficients measuring the impact of the baseline covariates on the probability of transitioning from V_S to the other latent states

$\hat{\gamma}_{1\bar{v}_S v}$	V_E	V_{FE}	$V_{U\&T}$	V_{HE}
Intercept	0.56	−1.07	−2.65	−8.43**
Male	−0.07	0.00	0.64†	0.06
Catholic	−0.01	−0.06	0.14	0.12
Belfast	−0.47	−0.98	0.83	−0.21
South	−0.28	−0.69†	−0.25	0.50
SouthEast	−0.70	0.43	−1.29	0.59
West	−0.72	−0.17	−0.29	0.51
Grammar school	−0.64†	−1.08**	−0.81†	0.37
Father unemployed	−0.05	0.23	0.58	−0.10
High grades at school	−0.61†	0.19	−1.91**	0.97**
Father manager	−0.61†	−0.55	0.26	0.26
Cohabiting with both parents	0.21	−0.03	0.29	−0.03

Results

- By averaging estimated initial probabilities conditioned to the covariates's levels we can make the following comparisons

- The initial probabilities broken down by father occupation (manager vs not manager)

Father manager: $\hat{\pi}_1$					Father not manager: $\hat{\pi}_1$				
V_S	V_E	V_{FE}	$V_{U\&T}$	V_{HE}	V_S	V_E	V_{FE}	$V_{U\&T}$	V_{HE}
0.21	0.23	0.14	0.42	0.00	0.18	0.25	0.14	0.43	0.00

- The only difference is the probability to stay at school in the first period July 1993

Results

- By averaging the estimated estimated transition probabilities conditioned to the covariates's levels we can make many comparisons
  - For example if we broken down by father occupation (manager vs not manager) (from row $\bar{v}$ to column v):

	Father manager: $\hat{\pi}_{v\|\bar{v}}$					Father not manager: $\hat{\pi}_{v\|\bar{v}}$				
	V_S	V_E	V_{FE}	$V_{U\&T}$	V_{HE}	V_S	V_E	V_{FE}	$V_{U\&T}$	V_{HE}
V_S	0.95	0.01	0.01	0.01	0.02	0.94	0.02	0.01	0.01	0.01
V_E	0.00	0.97	0.01	0.01	0.01	0.00	0.98	0.01	0.01	0.00
V_{FE}	0.00	0.02	0.96	0.01	0.01	0.00	0.03	0.95	0.02	0.00
$V_{U\&T}$	0.01	0.05	0.03	0.91	0.00	0.00	0.04	0.02	0.94	0.00
V_{HE}	0.00	0.02	0.07	0.00	0.98	0.00	0.01	0.08	0.00	0.89

  - There is quite a lot of persistency in the same state during the period of observation
  - We notice that the highest difference in probabilities (0.09) among the two matrices is related to further education for those having a father manager; the other differences are lower then 0.03

Results

- The initial probabilities broken down by catholic religion (catholic vs not catholic)

Catholic: $\hat{\pi}_1$					Non catholic: $\hat{\pi}_1$				
V_S	V_E	V_{FE}	$V_{U\&T}$	V_{HE}	V_S	V_E	V_{FE}	$V_{U\&T}$	V_{HE}
0.20	0.23	0.15	0.42	0.00	0.17	0.25	0.12	0.44	0.00

  - Catholic individuals show a higher probability to proceed with school and further education after the end of the compulsory school compared to non catholic individuals
  - They also have a lower probability to be employed compared to their counterpart

Results

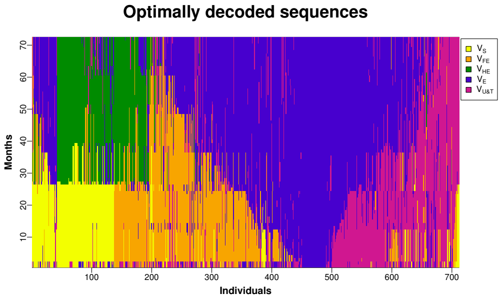
- By averaging the estimated estimated transition probabilities conditioned to the catholic religion

	Catholic: $\hat{\pi}_{v\|\bar{v}}$					Non catholic: $\hat{\pi}_{v\|\bar{v}}$				
	V_S	V_E	V_{FE}	$V_{U\&T}$	V_{HE}	V_S	V_E	V_{FE}	$V_{U\&T}$	V_{HE}
V_S	0.94	0.02	0.01	0.02	0.01	0.94	0.02	0.02	0.01	0.01
V_E	0.00	0.98	0.01	0.01	0.00	0.00	0.98	0.01	0.01	0.00
V_{FE}	0.00	0.03	0.95	0.01	0.01	0.00	0.03	0.95	0.01	0.01
$V_{U\&T}$	0.01	0.03	0.01	0.94	0.00	0.01	0.05	0.02	0.92	0.00
V_{HE}	0.00	0.01	0.07	0.00	0.92	0.00	0.01	0.08	0.00	0.91

- The catholic have a higher probability of remaining in $U\&T$ (unemployment and training) compared to their counterpart

- They also show a lower probability to transit towards E (empoloyment) once in $U\&T$ compared to their counterpart

Predicted sequence of latent states



- To each individual a (vertical) bar is associated, with colours depending on the predicted activity in each month (the ordering of individuals on the horizontal axes is the same used for the original sequences)
- The predicted paths are coherent with the observed ones

Results

- On the basis of the estimates of the covariate effects on the initial probabilities and on the transition probabilities we conclude that under the selected model
  - To be male affect positively the probability of begin in the class with moderate or high employment with respect to the class of low work employment
  - Young people with high grades have a higher probability to change from low to high employment than subjects with low grades
  - Young people leaving in the Belfast have a higher probability to be employed than subjects leaving on the other areas
- The model is able to account for the longitudinal phenomena as well as to consider the unobserved cultural factors influencing individual choices

Main References

- Amstrong D. M. (1999). School performance and staying on: a micro analysis for Northen Irland. *The Manchester School*, **76**, 203-230.
- Bartolucci, F. and Farcomeni, A. and Pennoni, F. (2010). An overview of latent Markov models for longitudinal categorical data, *Technical report available at http://arxiv.org/abs/1003.2804*.
- Piccarreta, R., Lior, O. (2010). Exploring sequences: A graphical tool based upon Multidimensional Scaling, *JRSS Series A*,ă**173**,165-184
- Piccarreta, R. and Billari, F. (2007). Clustering work and family trajectories using a divisive algorithm. *Journal of the Royal Statistical Society, Series A*, **170**, 1061–1078.
- Dempster A. P., Laird, N. M. & Rubin, D. B. (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion), *JRSS-B*, **39**, 1-38.
- Mcvicar, D., Anyadike-Danes, M. (2002). Predicting successful and unsuccessful transitions from school to work by using sequence methods. *JRSSA*, **165**, 317–334.
- MacDonald, I. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series*. London: Chapman & Hall.
- Wiggins, L.M. (1973). *Panel Analysis: Latent probability models for attitude and behavious processes*, Amsterdam: Elsevier.

Main References

- Pennoni, F. (2014). *Issues on the estimation of latent variable and latent class models*. Scholars' Press, Saarbucken.
- Bartolucci, F., Farcomeni A., Pennoni, F. (2014). Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates, *with discussion*, *Test*, **23**, 433-465.
- Bartolucci, F., Farcomeni A., Pennoni F. (2014). Rejoinder on: Latent Markov Models: a review of a general framework for the analysis of longitudinal data with covariates, *Test*, **23**, 484-496.
- Bartolucci, F. and Farcomeni, A. and Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*, Chapman and Hall/CRC press, Boca Raton.
- Baum, L.E., Petrie, T., Soules, G. and Weiss, N. (1970), A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Annals of Mathematical Statistics*, **41**, 164-171.
- Bartolucci, F.; Pennoni, F. and Vittadini, G. (2011). Assessment of school performance through a multilevel latent Markov Rasch model, *Journal of Educational and Behavioural Statistics* , **36**, 491-522
- Welch, L. R. (2003). Hidden Markov Models and the Baum-Welch Algorithm, *IEEE Information Theory Society Newsletter*, **53**, 1-13.