A Finite Mixture Model with Trajectories Depending on Covariates

Jang SCHILTZ (University of Luxembourg)

joint work with
Jean-Daniel GUIGOU (University of Luxembourg),
& Bruno LOVAT (University of Lorraine)

June 11, 2014

Outline

Outline

Outline

Outline

LUXEMBOURG SCHOOL OF FINANCE

General description of Nagin's model

We have a collection of individual trajectories.

General description of Nagin's model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.

General description of Nagin's model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.

Hence, this model can be interpreted as functional fuzzy cluster analysis.

General description of Nagin's model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.

Hence, this model can be interpreted as functional fuzzy cluster analysis.

Finite mixture model (Daniel S. Nagin (Carnegie Mellon University))

General description of Nagin's model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.

Hence, this model can be interpreted as functional fuzzy cluster analysis.

Finite mixture model (Daniel S. Nagin (Carnegie Mellon University))

- mixture : population composed of a mixture of unobserved groups

LUXEMBOURG SCHOOL OF FINANCE

General description of Nagin's model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.

Hence, this model can be interpreted as functional fuzzy cluster analysis.

Finite mixture model (Daniel S. Nagin (Carnegie Mellon University))

- mixture : population composed of a mixture of unobserved groups
- finite : sums across a finite number of groups

The Likelihood Function (1)

Consider a population of size N and a variable of interest Y.

The Likelihood Function (1)

Consider a population of size N and a variable of interest Y.

Let $Y_i = y_{i_1}, y_{i_2}, ..., y_{i_T}$ be T measures of the variable, taken at times $t_1, ...t_T$ for subject number i.

The Likelihood Function (1)

Consider a population of size N and a variable of interest Y.

Let $Y_i = y_{i_1}, y_{i_2}, ..., y_{i_T}$ be T measures of the variable, taken at times $t_1, ...t_T$ for subject number i.

π_j : probability of a given subject to belong to group number j

LUXEMBOURG SCHOOL OF FINANCE

The Likelihood Function (1)

Consider a population of size N and a variable of interest Y.

Let $Y_i = y_{i_1}, y_{i_2}, ..., y_{i_T}$ be T measures of the variable, taken at times $t_1, ... t_T$ for subject number i.

π_j : probability of a given subject to belong to group number j

$$\Rightarrow \pi_j \text{ is the size of group } j.$$

The Likelihood Function (1)

Consider a population of size N and a variable of interest Y.

Let $Y_i = y_{i_1}, y_{i_2}, \ldots, y_{i_T}$ be T measures of the variable, taken at times $t_1, \ldots t_T$ for subject number i.

π_j : probability of a given subject to belong to group number j

$\Rightarrow \pi_j$ is the size of group j.

$$\Rightarrow P(Y_i) = \sum_{j=1}^{r} \pi_j P^j(Y_i), \tag{1}$$

LUXEMBOURG SCHOOL OF FINANCE

The Likelihood Function (1)

Consider a population of size N and a variable of interest Y.

Let $Y_i = y_{i_1}, y_{i_2}, ..., y_{i_T}$ be T measures of the variable, taken at times $t_1, ...t_T$ for subject number i.

π_j : probability of a given subject to belong to group number j

$$\Rightarrow \pi_j \text{ is the size of group } j.$$

$$\Rightarrow P(Y_i) = \sum_{j=1}^{r} \pi_j P^j(Y_i), \tag{1}$$

where $P^j(Y_i)$ is probability of Y_i if subject i belongs to group j.

The Likelihood Function (2)

Aim of the analysis: Find r groups of trajectories of a given kind (for instance polynomials of degree 4, $P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$.

The Likelihood Function (2)

Aim of the analysis: Find r groups of trajectories of a given kind (for instance polynomials of degree 4, $P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$.

We try to estimate a set of parameters $\Omega = \left\{ \beta_0^j, \beta_1^j, \beta_2^j, \beta_3^j, \beta_4^j, \pi_j \right\}$ which allow to maximize the probability of the measured data.

The Likelihood Function (2)

Aim of the analysis: Find r groups of trajectories of a given kind (for instance polynomials of degree 4, $P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$.

We try to estimate a set of parameters $\Omega = \left\{ \beta_0^j, \beta_1^j, \beta_2^j, \beta_3^j, \beta_4^j, \pi_j \right\}$ which allow to maximize the probability of the measured data.

Possible data distributions:

LUXEMBOURG SCHOOL OF FINANCE

The Likelihood Function (2)

Aim of the analysis: Find r groups of trajectories of a given kind (for instance polynomials of degree 4, $P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$.

We try to estimate a set of parameters $\Omega = \left\{ \beta_0^j, \beta_1^j, \beta_2^j, \beta_3^j, \beta_4^j, \pi_j \right\}$ which allow to maximize the probability of the measured data.

Possible data distributions:

- count data $\Rightarrow$ Poisson distribution

The Likelihood Function (2)

Aim of the analysis: Find r groups of trajectories of a given kind (for instance polynomials of degree 4, $P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$.

We try to estimate a set of parameters $\Omega = \left\{ \beta_0^j, \beta_1^j, \beta_2^j, \beta_3^j, \beta_4^j, \pi_j \right\}$ which allow to maximize the probability of the measured data.

Possible data distributions:

- count data $\Rightarrow$ Poisson distribution
- binary data $\Rightarrow$ Binary logit distribution

The Likelihood Function (2)

Aim of the analysis: Find r groups of trajectories of a given kind (for instance polynomials of degree 4, $P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$.

We try to estimate a set of parameters $\Omega = \left\{\beta_0^j, \beta_1^j, \beta_2^j, \beta_3^j, \beta_4^j, \pi_j\right\}$ which allow to maximize the probability of the measured data.

Possible data distributions:

- count data $\Rightarrow$ Poisson distribution
- binary data $\Rightarrow$ Binary logit distribution
- censored data $\Rightarrow$ Censored normal distribution

The case of a normal distribution (1)

Notations :

The case of a normal distribution (1)

Notations :

- $\beta^j t_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4$.

The case of a normal distribution (1)

Notations :

- $\beta^j t_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4$.
- ϕ: density of standard centered normal law.

The case of a normal distribution (1)

Notations :

- $\beta^j t_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4.$
- ϕ: density of standard centered normal law.

Then,

The case of a normal distribution (1)

Notations :

- $\beta^j t_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4$.
- ϕ: density of standard centered normal law.

Then,

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi\left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma} \right). \tag{2}$$

The case of a normal distribution (1)

Notations :

- $\beta^j t_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4$.
- ϕ: density of standard centered normal law.

Then,

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi \left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma} \right). \tag{2}$$

It is too complicated to get closed-forms equations.

Available Software

Available Software

SAS-based Proc Traj procedure

By Bobby L. Jones (Carnegie Mellon University).

Available Software

SAS-based Proc Traj procedure

By Bobby L. Jones (Carnegie Mellon University).

Uses a quasi-Newton procedure maximum research routine.

Available Software

SAS-based Proc Traj procedure

By Bobby L. Jones (Carnegie Mellon University).

Uses a quasi-Newton procedure maximum research routine.

Since the likelyhood is nor convex, nor a contraction, there are issues with local maxima.

Available Software

SAS-based Proc Traj procedure

By Bobby L. Jones (Carnegie Mellon University).

Uses a quasi-Newton procedure maximum research routine.

Since the likelyhood is nor convex, nor a contraction, there are issues with local maxima.

R-package crimCV

By Jason D. Nielsen (Carleton University Ottawa).

Available Software

SAS-based Proc Traj procedure

By Bobby L. Jones (Carnegie Mellon University).

Uses a quasi-Newton procedure maximum research routine.

Since the likelyhood is nor convex, nor a contraction, there are issues with local maxima.

R-package crimCV

By Jason D. Nielsen (Carleton University Ottawa).

Just implements a zero-inflation Poission model.

Model Selection (1)

Bayesian Information Criterion:

$$\text{BIC} = \log(L) - 0,5k\log(N), \tag{3}$$

where k denotes the number of parameters in the model.

Rule:

The bigger the BIC, the better the model!

LUXEMBOURG SCHOOL OF FINANCE

Model Selection (2)

Leave-one-out Cross-Validation Apporach:

$$CVE = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left|y_{i_t} - \hat{y}_{i_t}^{[-i]}\right|. \tag{4}$$

Rule:

The smaller the CVE, the better the model!

Posterior Group-Membership Probabilities

Posterior probability of individual i's membership in group j : $P(j/Y_i)$.

Bayes's theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{5}$$

Bigger groups have on average larger probability estimates.

To be classified into a small group, an individual really needs to be strongly consistent with it.

An application example

An application example

The data : first dataset Salaries of workers in the private sector in Luxembourg from 1940 to 2006.

An application example

The data : first dataset Salaries of workers in the private sector in Luxembourg from 1940 to 2006.

About 7 million salary lines corresponding to 718.054 workers.

LUXEMBOURG SCHOOL OF FINANCE

An application example

The data : first dataset Salaries of workers in the private sector in Luxembourg from 1940 to 2006.

About 7 million salary lines corresponding to 718.054 workers.

Some sociological variables:

- gender (male, female)

An application example

The data : first dataset Salaries of workers in the private sector in Luxembourg from 1940 to 2006.

About 7 million salary lines corresponding to 718.054 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)

An application example

The data : first dataset Salaries of workers in the private sector in Luxembourg from 1940 to 2006.

About 7 million salary lines corresponding to 718.054 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)
- working status (white collar worker, blue collar worker)

An application example

The data : first dataset Salaries of workers in the private sector in Luxembourg from 1940 to 2006.

About 7 million salary lines corresponding to 718.054 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)
- working status (white collar worker, blue collar worker)
- year of birth

An application example

The data : first dataset Salaries of workers in the private sector in Luxembourg from 1940 to 2006.

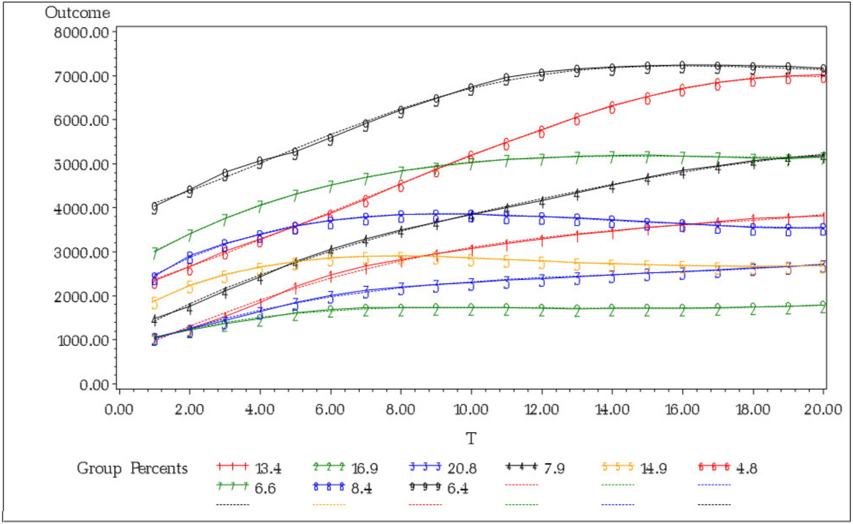
About 7 million salary lines corresponding to 718.054 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)
- working status (white collar worker, blue collar worker)
- year of birth
- age in the first year of professional activity

Result for 9 groups (dataset 1)

Result for 9 groups (dataset 1)



Results for 9 groups (dataset 1)

Maximum Likelihood Estimates
Model: Censored Normal (CNORM)

Group	Parameter	Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
1	Intercept	589.03067	18.46813	31.894	0.0000
	Linear	387.72145	11.31617	34.263	0.0000
	Quadratic	-14.36621	2.12997	-6.745	0.0000
	Cubic	-0.01563	0.15109	-0.103	0.9176
	Quartic	0.00856	0.00358	2.395	0.0166
2	Intercept	784.79156	15.75939	49.798	0.0000
	Linear	277.63602	9.78078	28.386	0.0000
	Quadratic	-28.36731	1.83236	-15.481	0.0000
	Cubic	1.17739	0.12972	9.076	0.0000
	Quartic	-0.01635	0.00307	-5.330	0.0000
3	Intercept	709.28728	15.90545	44.594	0.0000
	Linear	318.88029	8.97949	35.512	0.0000
	Quadratic	-21.54540	1.69611	-12.703	0.0000
	Cubic	0.62010	0.12002	5.167	0.0000
	Quartic	-0.00440	0.00284	-1.554	0.1203

Outline

Predictors of trajectory group membership

Predictors of trajectory group membership

x_i : vector of variables potentially associated with group membership (measured before t_1).

Predictors of trajectory group membership

x_i : vector of variables potentially associated with group membership (measured before t_1).

Multinomial logit model:

$$\pi_j(x_i) = \frac{e^{x_i \theta_j}}{\sum_{k=1}^{r} e^{x_i \theta_k}}, \tag{6}$$

where θ_j denotes the effect of x_i on the probability of group membership.

Predictors of trajectory group membership

x_i : vector of variables potentially associated with group membership (measured before t_1).

Multinomial logit model:

$$\pi_j(x_i) = \frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}, \tag{6}$$

where θ_j denotes the effect of x_i on the probability of group membership.

$$L = \frac{1}{\sigma}\prod_{i=1}^{N}\sum_{j=1}^{r}\frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}\prod_{t=1}^{T}\phi\left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma}\right). \tag{7}$$

Adding covariates to the trajectories (1)

Adding covariates to the trajectories (1)

Let $z_1 ... z_L$ be covariates potentially influencing Y.

Adding covariates to the trajectories (1)

Let $z_1 ... z_L$ be covariates potentially influencing Y.

We are then looking for trajectories

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \alpha_1^j z_1 + ... + \alpha_L^j z_L + \varepsilon_{i_t}, \quad (8)$$

where $\varepsilon_{i_t} \sim \mathcal{N}(0, \sigma)$, σ being a constant standard deviation and z_l are covariates that may depend or not upon time t.

Adding covariates to the trajectories (1)

Let $z_1 ... z_L$ be covariates potentially influencing Y.

We are then looking for trajectories

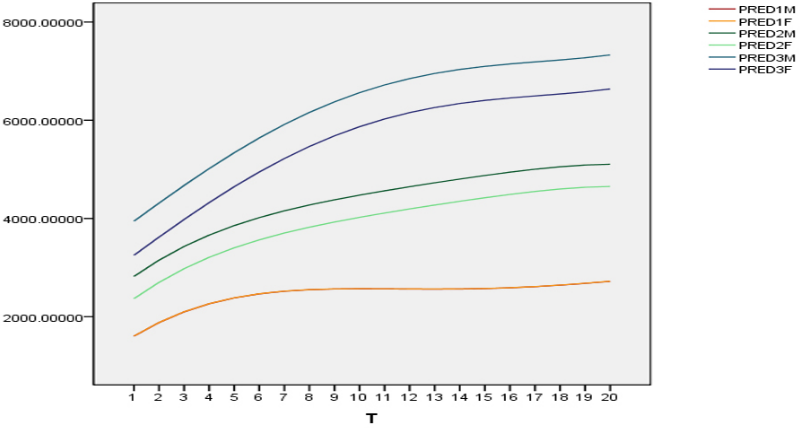
$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \alpha_1^j z_1 + ... + \alpha_L^j z_L + \varepsilon_{i_t}, \quad (8)$$

where $\varepsilon_{i_t} \sim \mathcal{N}(0, \sigma)$, σ being a constant standard deviation and z_l are covariates that may depend or not upon time t.

Unfortunately the influence of the covariates in this model is limited to the intercept of the trajectory.

Adding covariates to the trajectories (2)

Adding covariates to the trajectories (2)



Outline

Our model

Our model

Let $x_1 ... x_L$ and $z_{i_1}, ..., z_{i_T}$ be covariates potentially influencing Y.

Our model

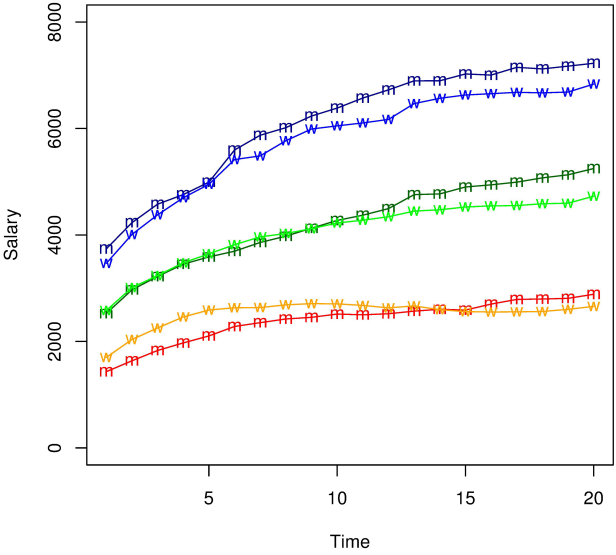
Let $x_1 \ldots x_L$ and $z_{i_1}, \ldots, z_{i_T}$ be covariates potentially influencing Y.

We propose the following model:

$$y_{i_t} = \left(\beta_0^j + \sum_{l=1}^{L} \alpha_{0l}^j x_l + \gamma_0^j z_{i_t}\right) + \left(\beta_1^j + \sum_{l=1}^{L} \alpha_{1l}^j x_l + \gamma_1^j z_{i_t}\right) Age_{i_t}$$
$$+ \left(\beta_2^j + \sum_{l=1}^{L} \alpha_{2l}^j x_l + \gamma_2^j z_{i_t}\right) Age_{i_t}^2 + \left(\beta_3^j + \sum_{l=1}^{L} \alpha_{3l}^j x_l + \gamma_3^j z_{i_t}\right) Age_{i_t}^3$$
$$+ \left(\beta_4^j + \sum_{l=1}^{L} \alpha_{4l}^j x_l + \gamma_4^j z_{i_t}\right) Age_{i_t}^4 + \varepsilon_{i_t},$$

where $\varepsilon_{i_t} \sim \mathcal{N}(0, \sigma)$, σ being a constant standard deviation.

Men versus women



Statistical Properties

Both Nagin's and our model can be written as

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi \left(\frac{\text{observed data - modelled data}}{\sigma} \right). \tag{9}$$

LUXEMBOURG SCHOOL OF FINANCE

Statistical Properties

Both Nagin's and our model can be written as

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi \left(\frac{\text{observed data - modelled data}}{\sigma} \right). \tag{9}$$

Nagin's model:

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \alpha_1^j z_1 + \ldots + \alpha_L^j z_L + \varepsilon_{i_t}, \tag{10}$$

Our model:

$$\begin{aligned} y_{i_t} = {} & \left(\beta_0^j + \sum_{l=1}^{L} \alpha_{0l}^j x_l + \gamma_0^j z_{i_t} \right) + \left(\beta_1^j + \sum_{l=1}^{L} \alpha_{1l}^j x_l + \gamma_1^j z_{i_t} \right) Age_{i_t} \\ & + \left(\beta_2^j + \sum_{l=1}^{L} \alpha_{2l}^j x_l + \gamma_2^j z_{i_t} \right) Age_{i_t}^2 + \left(\beta_3^j + \sum_{l=1}^{L} \alpha_{3l}^j x_l + \gamma_3^j z_{i_t} \right) Age_{i_t}^3 \\ & + \left(\beta_4^j + \sum_{l=1}^{L} \alpha_{4l}^j x_l + \gamma_4^j z_{i_t} \right) Age_{i_t}^4 + \varepsilon_{i_t}, \end{aligned}$$

Parameter estimation

A way of estimating our model with the existing software (if group membership does not depend on the covariates) :

Parameter estimation

A way of estimating our model with the existing software (if group membership does not depend on the covariates) :

- Use the latest version of proc.traj to test if the covariates have indeed an influence on the trajectories.

Parameter estimation

A way of estimating our model with the existing software (if group membership does not depend on the covariates) :

- Use the latest version of proc.traj to test if the covariates have indeed an influence on the trajectories.
- Apply proc.traj to the data without covariates do the clustering and obtain the number of groups and the constitution of the groups.

Parameter estimation

A way of estimating our model with the existing software (if group membership does not depend on the covariates) :

- Use the latest version of proc.traj to test if the covariates have indeed an influence on the trajectories.
- Apply proc.traj to the data without covariates do the clustering and obtain the number of groups and the constitution of the groups.
- Use your favorite regression model software to get the trajectories separately for each group.

LUXEMBOURG SCHOOL OF FINANCE

Bibliography

- Nagin, D.S. 2005: *Group-based Modeling of Development*. Cambridge, MA.: Harvard University Press.
- Jones, B. and Nagin D.S. 2007: Advances in Group-based Trajectory Modeling and a SAS Procedure for Estimating Them. *Sociological Research and Methods* **35** p.542-571.
- Nielsen, J.D. et al. 2012: Group-Based criminal trajectory analysis using cross-validation criteria. http://www.probability.ca/jeff/research.html.
- Muthén, B., Shedden, K. 1999: Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm. *Biometrics* **55** p.463-469.
- Guigou, J.D, Lovat, B. and Schiltz, J. 2012: Optimal mix of funded and unfunded pension systems: the case of Luxembourg. *Pensions* **17-4** p. 208-222.
- Schiltz, J. 2015: A generalization of Nagin's finite mixture model. In: Dependent data in social sciences research: Forms, issues, and methods of analysis' Mark Stemmler, Alexander von Eye & Wolfgang Wiedermann (Eds.). Springer 2015.