An R-package for finite mixture models

Jang SCHILTZ (University of Luxembourg)

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

December 7, 2013

LUXEMBOURG SCHOOL OF FINANCE

An R-package for finite mixture models

Jang SCHILTZ (University of Luxembourg)

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

December 7, 2013

Outline

1. The Basic Finite Mixture Model of Nagin

Outline

1. The Basic Finite Mixture Model of Nagin
2. Generalizations of the basic model

Outline

1. The Basic Finite Mixture Model of Nagin
2. Generalizations of the basic model
3. Muthén's model

Outline

Outline

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 a mean 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 a mean trajectory for each sub-population.

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

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.

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.

$P(Y_i)$ denotes the probability of Y_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.

$P(Y_i)$ denotes the probability of Y_i

- count data $\Rightarrow$ Poisson distribution

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.

$P(Y_i)$ denotes the probability of Y_i

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

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.

$P(Y_i)$ denotes the probability of Y_i

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

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.

$P(Y_i)$ denotes the probability of Y_i

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

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

LUXEMBOURG SCHOOL OF FINANCE

The Likelihood Function (2)

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

The Likelihood Function (2)

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

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

LUXEMBOURG SCHOOL OF FINANCE

The Likelihood Function (2)

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

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

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)

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

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

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.

$P^j(Y_i)$: probability of Y_i if subject i belongs to group j

The Likelihood Function (2)

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

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

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.

$P^j(Y_i)$: probability of Y_i if subject i belongs to 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 (2)

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

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

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.

$P^j(Y_i)$: probability of Y_i if subject i belongs to group j

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

<u>Finite mixture model</u> (Daniel S. Nagin (Carnegie Mellon University))

LUXEMBOURG SCHOOL OF FINANCE

The Likelihood Function (2)

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

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

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.

$P^j(Y_i)$: probability of Y_i if subject i belongs to group j

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

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

- finite : sums across a finite number of groups

The Likelihood Function (2)

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

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

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.

$P^j(Y_i)$: probability of Y_i if subject i belongs to group j

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

<u>Finite mixture model</u> (Daniel S. Nagin (Carnegie Mellon University))

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

The Likelihood Function (3)

Hypothesis: In a given group, conditional independence is assumed for the sequential realizations of the elements of Y_i !!!

The Likelihood Function (3)

Hypothesis: In a given group, conditional independence is assumed for the sequential realizations of the elements of Y_i !!!

$$\Rightarrow P^j(Y_i) = \prod_{t=1}^{T} p^j(y_{i_t}), \tag{2}$$

where $p^j(y_{i_t})$ denotes the probability of y_{i_t} given membership in group j.

The Likelihood Function (3)

Hypothesis: In a given group, conditional independence is assumed for the sequential realizations of the elements of Y_i !!!

$$\Rightarrow P^j(Y_i) = \prod_{t=1}^{T} p^j(y_{i_t}), \qquad (2)$$

where $p^j(y_{i_t})$ denotes the probability of y_{i_t} given membership in group j.

Likelihood of the estimator:

The Likelihood Function (3)

Hypothesis: In a given group, conditional independence is assumed for the sequential realizations of the elements of Y_i !!!

$$\Rightarrow P^j(Y_i) = \prod_{t=1}^{T} p^j(y_{i_t}), \tag{2}$$

where $p^j(y_{i_t})$ denotes the probability of y_{i_t} given membership in group j.

Likelihood of the estimator:

$$L = \prod_{i=1}^{N} P(Y_i)$$

The Likelihood Function (3)

Hypothesis: In a given group, conditional independence is assumed for the sequential realizations of the elements of Y_i !!!

$$\Rightarrow P^j(Y_i) = \prod_{t=1}^{T} p^j(y_{i_t}), \tag{2}$$

where $p^j(y_{i_t})$ denotes the probability of y_{i_t} given membership in group j.

Likelihood of the estimator:

$$L = \prod_{i=1}^{N} P(Y_i) = \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} p^j(y_{i_t}). \tag{3}$$

The case of a normal distribution (1)

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \varepsilon_{i_t}, \tag{4}$$

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

The case of a normal distribution (1)

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

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

Notations :

LUXEMBOURG SCHOOL OF FINANCE

The case of a normal distribution (1)

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \varepsilon_{i_t}, \qquad (4)$$

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

Notations :

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

LUXEMBOURG SCHOOL OF FINANCE

The case of a normal distribution (1)

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \varepsilon_{i_t}, \tag{4}$$

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

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)

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \varepsilon_{i_t}, \qquad (4)$$

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

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.

Hence,

The case of a normal distribution (1)

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \varepsilon_{i_t}, \qquad (4)$$

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

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.

Hence,

$$p^j(y_{i_t}) = \frac{1}{\sigma}\phi\left(\frac{y_{i_t} - \beta^j t_{it}}{\sigma}.\right) \qquad (5)$$

LUXEMBOURG SCHOOL OF FINANCE

The case of a normal distribution (2)

So we get

The case of a normal distribution (2)

So we get

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi \left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma} \right). \qquad (6)$$

The case of a normal distribution (2)

So we get

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi \left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma} \right). \tag{6}$$

The estimations of π_j must be in $[0, 1]$.

The case of a normal distribution (2)

So we get

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi \left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma} \right). \tag{6}$$

The estimations of π_j must be in $[0, 1]$.

It is difficult to force this constraint in model estimation.

A computational trick

Instead, we estimate the real parameters θ_j such that

A computational trick

Instead, we estimate the real parameters θ_j such that

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}}, \tag{7}$$

LUXEMBOURG SCHOOL OF FINANCE

A computational trick

Instead, we estimate the real parameters θ_j such that

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}}, \tag{7}$$

Finally,

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}} \prod_{t=1}^{T} \phi \left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma} \right). \tag{8}$$

A computational trick

Instead, we estimate the real parameters θ_j such that

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}}, \tag{7}$$

Finally,

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}} \prod_{t=1}^{T} \phi\left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma}\right). \tag{8}$$

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).

LUXEMBOURG SCHOOL OF FINANCE

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).

LUXEMBOURG SCHOOL OF FINANCE

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.

Future Software

Future Software

R-package FMM

By Jang Schiltz & Mounir Shal (University of Luxembourg).

Future Software

R-package FMM

By Jang Schiltz & Mounir Shal (University of Luxembourg).

Uses the EM Algortihm.

Future Software

R-package FMM

By Jang Schiltz & Mounir Shal (University of Luxembourg).

Uses the EM Algortihm.

Allows the estimation of a generalised version of Nagin's model, as well as Muthen's model.

Future Software

R-package FMM

By Jang Schiltz & Mounir Shal (University of Luxembourg).

Uses the EM Algortihm.

Allows the estimation of a generalised version of Nagin's model, as well as Muthen's model.

Will take us probably another year before completion.

LUXEMBOURG SCHOOL OF FINANCE

Model Selection (1)

Model Selection (1)

Bayesian Information Criterion:

$$\mathrm{BIC} = \log(L) - 0,5k\log(N), \tag{9}$$

where k denotes the number of parameters in the model.

Model Selection (1)

Bayesian Information Criterion:

$$\mathrm{BIC} = \log(L) - 0,5k\log(N), \tag{9}$$

where k denotes the number of parameters in the model.

Rule:

The bigger the BIC, the better the model!

Model Selection (2)

Leave-one-out Cross-Validation Apporach:

Model Selection (2)

Leave-one-out Cross-Validation Apporach:

$$CVE = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \left| y_{i_t} - \hat{y}_{i_t}^{[-i]} \right|. \tag{10}$$

Model Selection (2)

Leave-one-out Cross-Validation Apporach:

$$CVE = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left|y_{i_t} - \hat{y}_{i_t}^{[-i]}\right|. \tag{10}$$

Rule:

The smaller the CVE, the better the model!

Their "proof" that CVE is better than BIC

Their "proof " that CVE is better than BIC

TO1:

ngr	llike	AIC	BIC	CVE
1	−13967.63	27945.26	27982.26	1.0902792
2	−11929.40	23880.81	23962.22	0.9128347
3	−11424.68	22883.37	23009.18	0.9592355
4	−11191.28	22428.55	22598.77	0.9052791
5	−11016.19	22090.37	22304.99	0.8535441
6	−10886.30	21842.61	22101.63	0.8334242
7	−10805.59	21693.18	21996.60	0.8261734
8	−10732.58	21559.16	21906.99	**0.8123785**
9	−10684.54	**21475.08**	**21867.31**	0.8240060

Posterior Group-Membership Probabilities

Posterior Group-Membership Probabilities

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

Posterior Group-Membership Probabilities

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

Bayes's theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \qquad (11)$$

Posterior Group-Membership Probabilities

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

Bayes's theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{11}$$

Bigger groups have on average larger probability estimates.

Posterior Group-Membership Probabilities

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

Bayes's theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{11}$$

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.

LUXEMBOURG SCHOOL OF FINANCE

Model Fit (1)

Model Fit (1)

Diagnostic 1: Average Posterior Probability of Assignment

AvePP should be at least $0,7$ for all groups.

Model Fit (1)

Diagnostic 1: Average Posterior Probability of Assignment

AvePP should be at least $0,7$ for all groups.

Diagonostic 2: Odds of Correct Classification

$$OCC_j = \frac{AvePP_j/1 - AvePP_j}{\hat{\pi}_j/1 - \hat{\pi}_j}. \tag{12}$$

Model Fit (1)

Diagnostic 1: Average Posterior Probability of Assignment

AvePP should be at least $0,7$ for all groups.

Diagonostic 2: Odds of Correct Classification

$$OCC_j = \frac{AvePP_j/1 - AvePP_j}{\hat{\pi}_j/1 - \hat{\pi}_j}. \tag{12}$$

OCC_j should be greater than 5 for all groups.

LUXEMBOURG SCHOOL OF FINANCE

Model Fit (2)

Diagonostic 3: Comparing $\hat{\pi}_j$ to the Proportion of the Sample Assigned to Group j

The ratio of the two should be close to 1.

Model Fit (2)

Diagonostic 3: Comparing $\hat{\pi}_j$ to the Proportion of the Sample Assigned to Group j

The ratio of the two should be close to 1.

Diagonostic 4: Confidence Intervals for Group Membership Probabilities

The confidence intervals for group membership probabilities estimates should be narrow, i.e. standard deviation of $\hat{\pi}_j$ should be small.

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.

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.

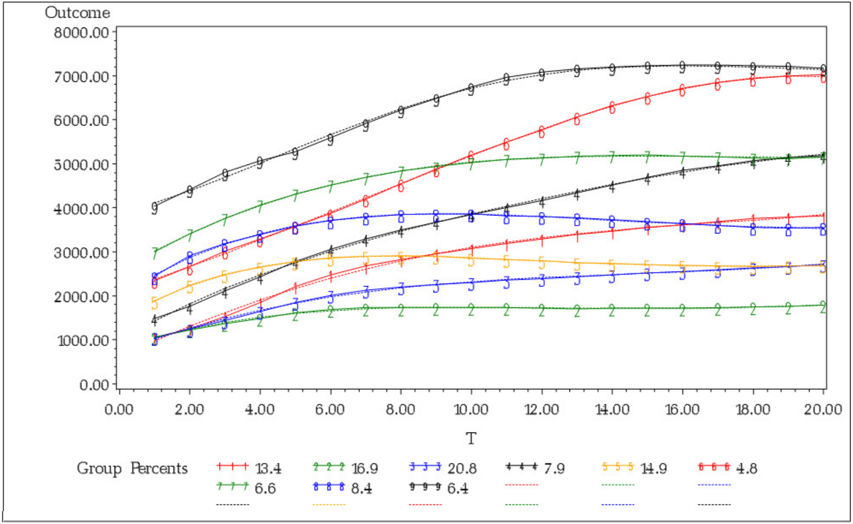
About 7 million salary lines corresponding to 718.054 workers.

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:

$$\pi_j(x_i) = \frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}, \tag{13}$$

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:

$$\pi_j(x_i) = \frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}, \tag{13}$$

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

$$L = \frac{1}{\sigma}\prod_{i=1}^{N}\sum_{j=1}^{r}\frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}\prod_{t=1}^{T}\phi\left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma}\right). \tag{14}$$

Group membership probabilities

Group membership probabilities

The Wald test which indicates whether any number of coefficients is significally different, allows the statistical testing of the predictors.

Group membership probabilities

The Wald test which indicates whether any number of coefficients is significally different, allows the statistical testing of the predictors.

Confidence intervals for the probabilities of group membership can be computed by a parametric bootstrap technique.

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

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \alpha_1^j z_1 + ... + \alpha_L^j z_L + \varepsilon_{i_t}, \quad (15)$$

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.

LUXEMBOURG SCHOOL OF FINANCE

Adding covariates to the trajectories (1)

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

We are then looking for trajectories

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \alpha_1^j z_1 + ... + \alpha_L^j z_L + \varepsilon_{i_t}, \quad (15)$$

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 estimation of parameters α_l^j is not implemented in proc traj procedure; it is just possible to plot the impact of the covariates.

Adding covariates to the trajectories (1)

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

We are then looking for trajectories

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \alpha_1^j z_1 + ... + \alpha_L^j z_L + \varepsilon_{i_t}, \quad (15)$$

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 estimation of parameters α_l^j is not implemented in proc traj procedure; it is just possible to plot the impact of the covariates.

Moreover, the influence of the covariates is limited to the intercept of the trajectory.

LUXEMBOURG SCHOOL OF FINANCE

An application example

The data : second dataset Salaries of all workers in Luxembourg which began to work in Luxembourg between 1980 and 1990 at an age less than 30 years.

An application example

The data : second dataset Salaries of all workers in Luxembourg which began to work in Luxembourg between 1980 and 1990 at an age less than 30 years.

1.303.010 salary lines corresponding to 85.049 workers.

An application example

The data : second dataset Salaries of all workers in Luxembourg which began to work in Luxembourg between 1980 and 1990 at an age less than 30 years.

1.303.010 salary lines corresponding to 85.049 workers.

Some sociological variables:

- gender (male, female)

An application example

The data : second dataset Salaries of all workers in Luxembourg which began to work in Luxembourg between 1980 and 1990 at an age less than 30 years.

1.303.010 salary lines corresponding to 85.049 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship

An application example

The data : second dataset Salaries of all workers in Luxembourg which began to work in Luxembourg between 1980 and 1990 at an age less than 30 years.

1.303.010 salary lines corresponding to 85.049 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship
- sector of activity

LUXEMBOURG SCHOOL OF FINANCE

An application example

The data : second dataset Salaries of all workers in Luxembourg which began to work in Luxembourg between 1980 and 1990 at an age less than 30 years.

1.303.010 salary lines corresponding to 85.049 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship
- sector of activity
- year of birth

An application example

The data : second dataset Salaries of all workers in Luxembourg which began to work in Luxembourg between 1980 and 1990 at an age less than 30 years.

1.303.010 salary lines corresponding to 85.049 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship
- sector of activity
- year of birth
- age in the first year of professional activity

An application example

The data : second dataset Salaries of all workers in Luxembourg which began to work in Luxembourg between 1980 and 1990 at an age less than 30 years.

1.303.010 salary lines corresponding to 85.049 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship
- sector of activity
- year of birth
- age in the first year of professional activity
- marital status

An application example

The data : second dataset Salaries of all workers in Luxembourg which began to work in Luxembourg between 1980 and 1990 at an age less than 30 years.

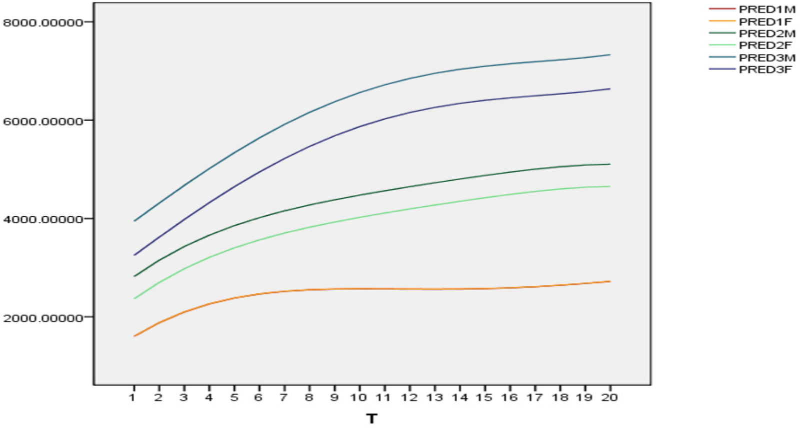
1.303.010 salary lines corresponding to 85.049 workers.

Some sociological variables:

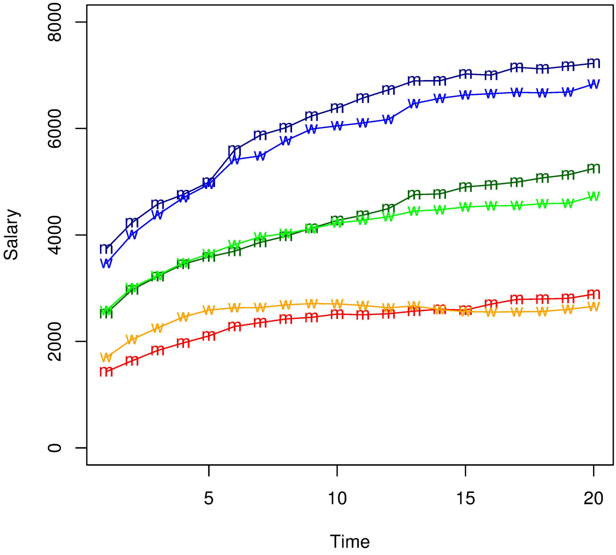
- gender (male, female)
- nationality and residentship
- sector of activity
- year of birth
- age in the first year of professional activity
- marital status
- year of birth of children

Adding covariates to the trajectories (3)

Adding covariates to the trajectories (3)



What's really going on



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.

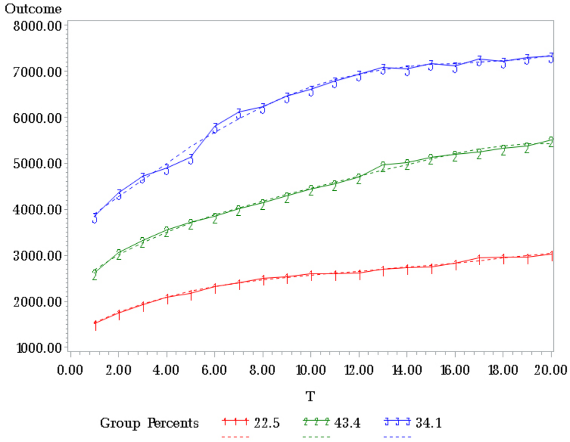
We propose the following model:

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

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

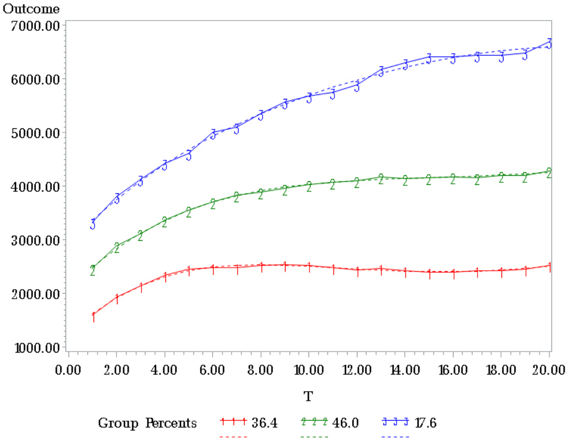
An alternative analysis (1)

Salary trajectories of the men



An alternative analysis (2)

Salary trajectories of the women



Outline

Muthén's model (1)

Muthén and Shedden (1999): Generalized growth curve model

Muthén's model (1)

Muthén and Shedden (1999): Generalized growth curve model

Elegant and technically demanding extension of the uncensored normal model.

Muthén's model (1)

Muthén and Shedden (1999): Generalized growth curve model

Elegant and technically demanding extension of the uncensored normal model.

Adds random effects to the parameters β^j that define a group's mean trajectory.

Muthén's model (1)

Muthén and Shedden (1999): Generalized growth curve model

Elegant and technically demanding extension of the uncensored normal model.

Adds random effects to the parameters β^j that define a group's mean trajectory.

Trajectories of individual group members can vary from the group trajectory.

Muthén's model (1)

Muthén and Shedden (1999): Generalized growth curve model

Elegant and technically demanding extension of the uncensored normal model.

Adds random effects to the parameters β^j that define a group's mean trajectory.

Trajectories of individual group members can vary from the group trajectory.

Software:

Mplus package by L.K. Muthén and B.O Muthén.

LUXEMBOURG SCHOOL OF FINANCE

Muthén's model (2)

Advantage of GGCM

Fewer groups are required to specify a satisfactory model.

Muthén's model (2)

Advantage of GGCM

Fewer groups are required to specify a satisfactory model.

Disadvantages of GGCM

1. Difficult to extend to other types of data.

Muthén's model (2)

Advantage of GGCM

Fewer groups are required to specify a satisfactory model.

Disadvantages of GGCM

1. Difficult to extend to other types of data.
2. Group cross-over effects.

Muthén's model (2)

Advantage of GGCM

Fewer groups are required to specify a satisfactory model.

Disadvantages of GGCM

1. Difficult to extend to other types of data.
2. Group cross-over effects.
3. Can create the illusion of non-existing groups.

Outline

LUXEMBOURG SCHOOL OF FINANCE

Outlook

Outlook

- Relationship between finite mixture models and hierarchical cluster analysis of functions.

LUXEMBOURG SCHOOL OF FINANCE

Outlook

- Relationship between finite mixture models and hierarchical cluster analysis of functions.
- Stability of the trajectories in finite mixture models.
  - Using classical statistics.
  - Using statistical shape analysis.
  - Using functional data analysis.

Outlook

- Relationship between finite mixture models and hierarchical cluster analysis of functions.

- Stability of the trajectories in finite mixture models.
  - Using classical statistics.
  - Using statistical shape analysis.
  - Using functional data analysis.

- Handling missing data.

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., Rosenthal, J.S., Sun, Y., Day, D.M., Bevc, I., Duchesne, T. 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.