# On some issues in trajectory modeling with finite mixture models

Jang SCHILTZ (University of Luxembourg)

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

June 30, 2015

LUXEMBOURG SCHOOL OF FINANCE

# Outline

# Outline

# 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, 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

# 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 $i$.

$\pi_j$ : 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 $i$.

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

$\pi_j$ : probability of a given subject to belong to group number $j$

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

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

# 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 $i$.

$\pi_j$ : probability of a given subject to belong to group number $j$

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

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

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

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, $P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$.

# 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.

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:

# 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

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

LUXEMBOURG SCHOOL OF FINANCE

# 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 :

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

# The case of a normal distribution (1)

Notations :

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

Then,

# The case of a normal distribution (1)

Notations :

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

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

# 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

LUXEMBOURG SCHOOL OF FINANCE

# An application example

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

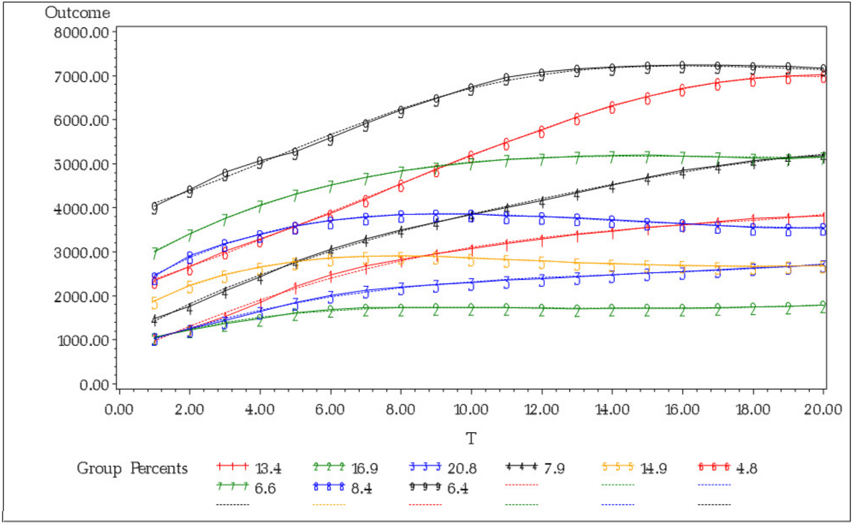
About 7 million salary lines corresponding to 718.054 workers.

Some sociological variables:

- gender (male, female)
- 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 |

# 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:

$$\pi_j(x_i) = \frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}, \tag{3}$$

where $\theta_j$ 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:

$$\pi_j(x_i) = \frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}, \tag{3}$$

where $\theta_j$ denotes the effect of $x_i$ on the probability of group membership.

$$L = \frac{1}{\sigma}\prod_{i=1}^{N}\sum_{j=1}^{r}\frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}\prod_{t=1}^{T}\phi\left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma}\right). \tag{4}$$

# Adding covariates to the trajectories (1)

# Adding covariates to the trajectories (1)

Let $z_1 \ldots z_M$ be covariates potentially influencing $Y$.

# Adding covariates to the trajectories (1)

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

We are then looking for trajectories

$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \alpha_1^j z_1 + ... + \alpha_M^j z_M + \varepsilon_{i_t}, \quad (5)$$

where $\varepsilon_{i_t} \sim \mathcal{N}(0, \sigma)$, $\sigma$ 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 $z_1 ... z_M$ be covariates potentially influencing $Y$.

We are then looking for trajectories

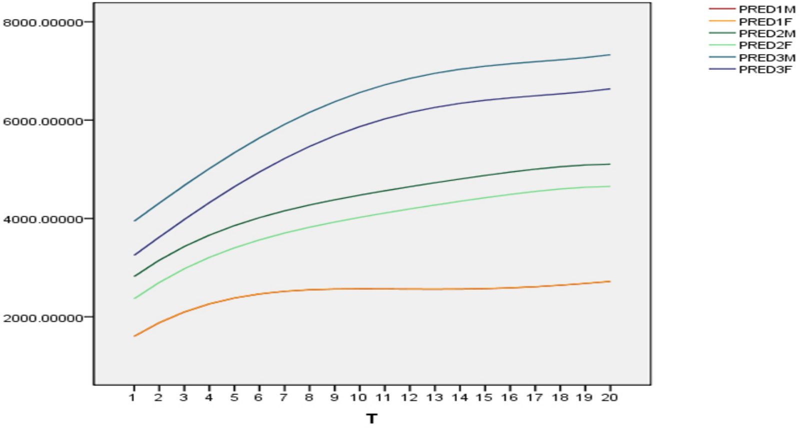
$$y_{i_t} = \beta_0^j + \beta_1^j Age_{i_t} + \beta_2^j Age_{i_t}^2 + \beta_3^j Age_{i_t}^3 + \beta_4^j Age_{i_t}^4 + \alpha_1^j z_1 + ... + \alpha_M^j z_M + \varepsilon_{i_t}, \quad (5)$$

where $\varepsilon_{i_t} \sim \mathcal{N}(0, \sigma)$, $\sigma$ 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 $x_1 ... x_M$ and $z_{i_1}, ..., z_{i_T}$ be covariates potentially influencing $Y$.

# Our model

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

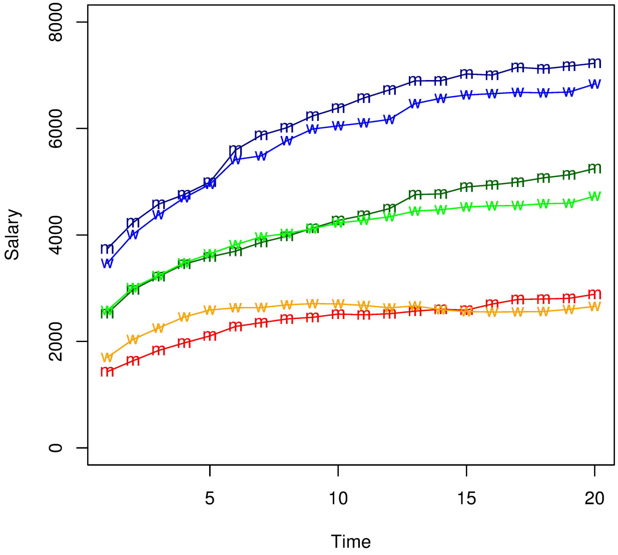
We propose the following model:

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

where $\varepsilon_{i_t} \sim \mathcal{N}(0, \sigma^j)$, $\sigma^j$ being the standard deviation, constant in group $j$.

LUXEMBOURG SCHOOL OF FINANCE

# Men versus women



# Statistical Properties

# Statistical Properties

The model's estimated parameters are the result of maximum likelihood estimation. As such, they are consistent and asymptotically normally distributed.

# Statistical Properties

The model's estimated parameters are the result of maximum likelihood estimation. As such, they are consistent and asymptotically normally distributed.

Confidence intervals of level $\alpha$ for the parameters $\beta_k^j$:

# Statistical Properties

The model's estimated parameters are the result of maximum likelihood estimation. As such, they are consistent and asymptotically normally distributed.

Confidence intervals of level $\alpha$ for the parameters $\beta_k^j$:

$$CI_\alpha(\beta_k^j) = \left[\hat{\beta}_k^j - t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j); \hat{\beta}_k^j + t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j)\right]. \quad (6)$$

# Statistical Properties

The model's estimated parameters are the result of maximum likelihood estimation. As such, they are consistent and asymptotically normally distributed.

Confidence intervals of level $\alpha$ for the parameters $\beta_k^j$:

$$CI_\alpha(\beta_k^j) = \left[\hat{\beta}_k^j - t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j); \hat{\beta}_k^j + t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j)\right]. \tag{6}$$

Confidence intervals of level $\alpha$ for the disturbance factor $\sigma_j$:

# Statistical Properties

The model's estimated parameters are the result of maximum likelihood estimation. As such, they are consistent and asymptotically normally distributed.

Confidence intervals of level $\alpha$ for the parameters $\beta_k^j$:

$$CI_\alpha(\beta_k^j) = \left[\hat{\beta}_k^j - t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j); \hat{\beta}_k^j + t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j)\right]. \tag{6}$$

Confidence intervals of level $\alpha$ for the disturbance factor $\sigma_j$:

$$CI_\alpha(\sigma_j) = \left[\sqrt{\frac{(N-(2+M)s-1)\hat{\sigma}_j{}^2}{\chi^2_{1-\alpha/2;N-(2+M)s-1}}}; \sqrt{\frac{(N-(2+M)s-1)\hat{\sigma}_j{}^2}{\chi^2_{\alpha/2;N-(2+M)s-1}}}\right]. \tag{7}$$

LUXEMBOURG SCHOOL OF FINANCE

# Outline

# An application example

# An application example

**The data : second dataset** Salaries of workers in the private sector in Luxembourg from 1987 to 2006.

# An application example

**The data : second dataset** Salaries of workers in the private sector in Luxembourg from 1987 to 2006.

About 1.3 million salary lines corresponding to 85.049 workers.

# An application example

**The data : second dataset** Salaries of workers in the private sector in Luxembourg from 1987 to 2006.

About 1.3 million salary lines corresponding to 85.049 workers.

Some sociological variables:

- gender (male, female)

# An application example

**The data : second dataset** Salaries of workers in the private sector in Luxembourg from 1987 to 2006.

About 1.3 million 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 workers in the private sector in Luxembourg from 1987 to 2006.

About 1.3 million salary lines corresponding to 85.049 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship
- working sector

# An application example

**The data : second dataset** Salaries of workers in the private sector in Luxembourg from 1987 to 2006.

About 1.3 million salary lines corresponding to 85.049 workers.

Some sociological variables:

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

# An application example

**The data : second dataset** Salaries of workers in the private sector in Luxembourg from 1987 to 2006.

About 1.3 million salary lines corresponding to 85.049 workers.

Some sociological variables:

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

# An application example

**The data : second dataset** Salaries of workers in the private sector in Luxembourg from 1987 to 2006.

About 1.3 million salary lines corresponding to 85.049 workers.

Some sociological variables:

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

# If group membership does not depend on the covariates

# If group membership does not depend on the covariates

We analyze if the fact to be either a Luxembourg resident or a commuter has an influence on the salary.

# If group membership does not depend on the covariates

We analyze if the fact to be either a Luxembourg resident or a commuter has an influence on the salary.

Unlike the case of the gender, this covariate does not distinguish trajectory membership.

# If group membership does not depend on the covariates

We analyze if the fact to be either a Luxembourg resident or a commuter has an influence on the salary.

Unlike the case of the gender, this covariate does not distinguish trajectory membership.

That means that Nagin's model does not provide different trajectories for residents and non residents.

# If group membership does not depend on the covariates

We analyze if the fact to be either a Luxembourg resident or a commuter has an influence on the salary.

Unlike the case of the gender, this covariate does not distinguish trajectory membership.

That means that Nagin's model does not provide different trajectories for residents and non residents.

In our model, this is not true, but the trajectories for the two groups remain of course very close.

# If group membership does not depend on the covariates

We analyze if the fact to be either a Luxembourg resident or a commuter has an influence on the salary.

Unlike the case of the gender, this covariate does not distinguish trajectory membership.

That means that Nagin's model does not provide different trajectories for residents and non residents.

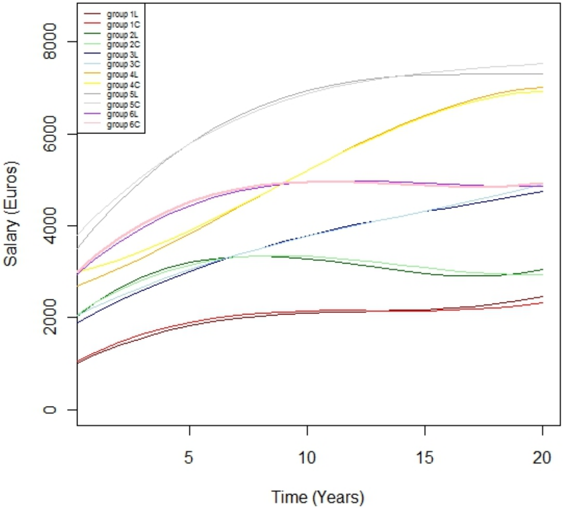
In our model, this is not true, but the trajectories for the two groups remain of course very close.

We calibrated the model

$$S_{it} = (\beta_0^j + \alpha_0^j x_i) + (\beta_1^j + \alpha_1^j x_i)t + (\beta_2^j + \alpha_2^j x_i)t^2 + (\beta_3^j + \alpha_3^j x_i)t^3, \quad (8)$$

where $S$ denotes the salary and $x$ the country of residence variable (The Luxembourg resident are coded by 1 and the commuters by 0).

# Resident versus non resident workers



# Parameter estimation

A way of estimating our model with the existing software:

# Parameter estimation

A way of estimating our model with the existing software:

- 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:

- 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:

- 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.

# Attention to multicolinearity issues!

# Attention to multicolinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary.

# Attention to multicolinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary.

CPI and time have a correlation of 0.995.

# Attention to multicolinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary.

CPI and time have a correlation of 0.995.

Hence a model like

$$S_{it} = (\beta_0^j + \gamma_0^j z_{it}) + (\beta_1^j + \gamma_1^j z_{it})t + (\beta_2^j + \gamma_2^j z_{it})t^2 + (\beta_3^j + \gamma_3^j z_{it})t^3, \quad (9)$$

where $S$ denotes the salary and $z_t$ is Luxembourg's CPI in year $t$ of the study, makes no sense.

LUXEMBOURG SCHOOL OF FINANCE

# Attention to multicolinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary.

CPI and time have a correlation of 0.995.

Hence a model like

$$S_{it} = (\beta_0^j + \gamma_0^j z_{it}) + (\beta_1^j + \gamma_1^j z_{it})t + (\beta_2^j + \gamma_2^j z_{it})t^2 + (\beta_3^j + \gamma_3^j z_{it})t^3, \quad (9)$$

where $S$ denotes the salary and $z_t$ is Luxembourg's CPI in year $t$ of the study, makes no sense.

Because of obvious multicolinearity problems, almost none of the parameters would be significant.

# Attention to multicolinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary.

CPI and time have a correlation of 0.995.

Hence a model like

$$S_{it} = (\beta_0^j + \gamma_0^j z_{it}) + (\beta_1^j + \gamma_1^j z_{it})t + (\beta_2^j + \gamma_2^j z_{it})t^2 + (\beta_3^j + \gamma_3^j z_{it})t^3, \quad (9)$$

where $S$ denotes the salary and $z_t$ is Luxembourg's CPI in year $t$ of the study, makes no sense.

Because of obvious multicolinearity problems, almost none of the parameters would be significant.

Therefore, we simplify the model and calibrate

$$S_{it} = (\beta_0^j + \gamma_0^j z_{it}) + \gamma_1^j z_{it} t + \gamma_2^j z_{it} t^2 + \gamma_3^j z_{it} t^3. \quad (10)$$

# Attention to multicolinearity issues!

We observe a significant influence of the CPI for all six groups, which is not astonishing, since by law, the salaries are coupled with the CPI.

# Attention to multicolinearity issues!

We observe a significant influence of the CPI for all six groups, which is not astonishing, since by law, the salaries are coupled with the CPI.

For groups two, three and six all parameters are significant. The trajectories in groups one and five do not have any constant term, nor a linear dependency on the CPI but depend only on the interaction of CPI and time. Group four, finally, exhibits only linear behaviour with respect to CPI, as well as the interaction of CPI and time.

# Attention to multicolinearity issues!

We observe a significant influence of the CPI for all six groups, which is not astonishing, since by law, the salaries are coupled with the CPI.

For groups two, three and six all parameters are significant. The trajectories in groups one and five do not have any constant term, nor a linear dependency on the CPI but depend only on the interaction of CPI and time. Group four, finally, exhibits only linear behaviour with respect to CPI, as well as the interaction of CPI and time.

The disturbance terms for the six groups are $\sigma_1 = 41.49$, $\sigma_2 = 33.18$, $\sigma_3 = 68.48$, $\sigma_4 = 64.84$, $\sigma_5 = 111.83$ and $\sigma_6 = 39.74$

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.
- 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.