A generalised finite mixture model

Jang SCHILTZ (University of Luxembourg)

November 27, 2013

LUXEMBOURG SCHOOL OF FINANCE

Outline

Outline

Outline

1. Nagin's Finite Mixture Model

2. Generalization of the basic model

3. The Luxemburgish salary trajectories

LUXEMBOURG SCHOOL OF FINANCE

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 subpopulations and to estimate a mean trajectory for each subpopulation.

General description of Nagin's model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous subpopulations and to estimate a mean trajectory for each subpopulation.

This is still an inter-individual model, but unlike other classical models such as standard growth curve models, it allows the existence of subpolulations with completely different behaviors.

LUXEMBOURG SCHOOL OF FINANCE

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.

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

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

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

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

$\Rightarrow \pi_j$ 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

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

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

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

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

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

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

Finite mixture model (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 !!!

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

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)

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

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 :

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 :

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

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)

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

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)

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

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,

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

The case of a normal distribution (2)

So we get

The case of a normal distribution (2)

So we get

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

The case of a normal distribution (2)

So we get

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

It is too complicated to get closed-forms equations

The case of a normal distribution (2)

So we get

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

It is too complicated to get closed-forms equations

$\Rightarrow$ quasi-Newton procedure maximum research routine

The case of a normal distribution (2)

So we get

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

It is too complicated to get closed-forms equations

$\Rightarrow$ quasi-Newton procedure maximum research routine

Software:

SAS-based Proc Traj procedure
by Bobby L. Jones (Carnegie Mellon University).

LUXEMBOURG SCHOOL OF FINANCE

A computational trick

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

A computational trick

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

It is difficult to force this constraint in model estimation.

A computational trick

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

It is difficult to force this constraint in model estimation.

Instead, we estimate the real parameters θ_j such that

LUXEMBOURG
SCHOOL OF FINANCE

A computational trick

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

It is difficult to force this constraint in model estimation.

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

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

It is difficult to force this constraint in model estimation.

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,

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

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.

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.

Model Selection

Bayesian Information Criterion:

Model Selection

Bayesian Information Criterion:

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

where k denotes the number of parameters in the model.

Model Selection

Bayesian Information Criterion:

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

where k denotes the number of parameters in the model.

Rule:

The bigger the BIC, the better the model!

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

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{10}$$

Posterior Group-Membership Probabilities

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

Bayes's theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{10}$$

Bigger groups have on average larger probability estimates.

LUXEMBOURG SCHOOL OF FINANCE

Posterior Group-Membership Probabilities

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

Bayes's theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{10}$$

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

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:

$$\pi_j(x_i) = \frac{e^{x_i \theta_j}}{\sum_{k=1}^{r} e^{x_i \theta_k}}, \tag{11}$$

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:

$$\pi_j(x_i) = \frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}, \tag{11}$$

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

$$L = \frac{1}{\sigma}\prod_{i=1}^{N}\sum_{j=1}^{r}\frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}}\prod_{t=1}^{T}\phi\left(\frac{y_{i_t} - \beta^j t_{i_t}}{\sigma}\right). \tag{12}$$

LUXEMBOURG SCHOOL OF FINANCE

Group membership probabilities

Group membership probabilities

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

Group membership probabilities

The Wald test which indicates whether any number of ocefficients 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

Adding covariates to the trajectories

$$y_{i_t} = \beta_0^j + \beta_1^j t + \beta_2^j t^2 + \beta_3^j t^3 + \beta_4^j t^4 + \alpha_1^j z_{1t} + ... + \alpha_L^j z_{Lt} + \varepsilon_{i_t}, \quad (13)$$

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

Adding covariates to the trajectories

$$y_{i_t} = \beta_0^j + \beta_1^j t + \beta_2^j t^2 + \beta_3^j t^3 + \beta_4^j t^4 + \alpha_1^j z_{1t} + \ldots + \alpha_L^j z_{Lt} + \varepsilon_{i_t}, \quad (13)$$

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

Outline

The data : first dataset

Salaries of workers in the private sector in Luxembourg from 1940 to 2006.

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.

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)

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)

LUXEMBOURG SCHOOL OF FINANCE

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)

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

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

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.

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.

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)

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

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

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

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

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

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.

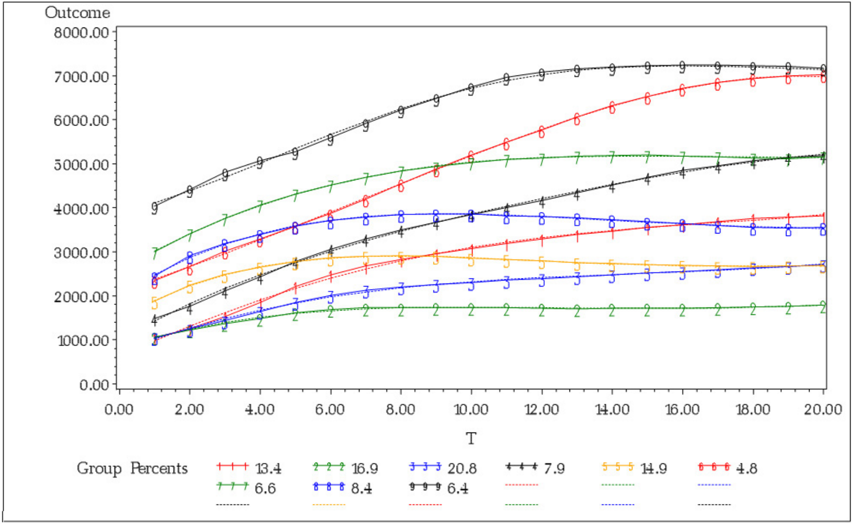
1.303.010 salary lines corresponding to 85.049 workers.

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

Result for 9 groups (dataset 1)

Result for 9 groups (dataset 1)



Adding covariates to the trajectories (dataset 2)

Adding covariates to the trajectories (dataset 2)