Group-based trajectory modeling

An application to economics

Jang SCHILTZ (University of Luxembourg)
joint work with Jean-Daniel GUIGOU (University of Luxembourg)
& Bruno LOVAT (University Nancy II)

August 24, 2009

Outline

Outline

Outline

Outline

Outline

Outline

Peter Molenaar (Luxembourg 2009):

Standard approach to statistical analysis in psychology:

- Analysis of **inter**-individual variation (variation **between** subjects in a population of subjects; individual differences)

- Strong assumption of **homogeneity** in (sub-)populations

- Aimed at generalization to the state of affairs at the **population level**

- **Implicit** assumption of applicability of results at the individual level of **intra**-individual variation

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.

By P.C.M Molenaar's definition, 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.

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, done 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, done 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, done at times $t_1, ...t_T$ for subject number i.

$P(Y_i)$ denotes the probability of Y_i

- count data $\Rightarrow$ Poisson distribution

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, done at times $t_1, ...t_T$ for subject number i.

$P(Y_i)$ denotes the probability of Y_i

- count data ⇒ Poisson distribution
- binary data ⇒ Binary logit distribution

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, done 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

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, done 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

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

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

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

Finite mixture model (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

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

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 censored normal distribution (1)

Y^*: latent variable measured by Y.

The case of a censored normal distribution (1)

Y^*: latent variable measured by Y.

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

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

The case of a censored normal distribution (1)

Y^*: latent variable measured by Y.

$$y^*_{i_t} = \beta^j_0 + \beta^j_1 Age_{i_t} + \beta^j_2 Age^2_{i_t} + \beta^j_3 Age^3_{i_t} + \beta^j_4 Age^4_{i_t} + \varepsilon_{i_t}, \quad (4)$$

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

Hence,

$$\begin{aligned} y_{i_t} = S_{min} &\quad \text{si} \quad y^*_{i_t} < S_{min}, \\ y_{i_t} = y^*_{i_t} &\quad \text{si} \quad S_{min} \leq y^*_{i_t} \leq S_{max}, \\ y_{i_t} = S_{max} &\quad \text{si} \quad y^*_{i_t} > S_{max}, \end{aligned}$$

where S_{min} and S_{max} dennote the minimum and maximum of the censored normal distribution.

The case of a censored normal distribution (2)

Notations :

The case of a censored normal distribution (2)

Notations :

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

The case of a censored normal distribution (2)

Notations :

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

The case of a censored normal distribution (2)

Notations :

- $\beta^j x_{i_t} = \beta^j_0 + \beta^j_1 Age_{i_t} + \beta^j_2 Age^2_{i_t} + \beta^j_3 Age^3_{i_t} + \beta^j_4 Age^4_{i_t}$.
- ϕ: density of standard centered normal law.
- Φ: cumulative distribution function of standard centered normal law.

The case of a censored normal distribution (2)

Notations :

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

Hence,

The case of a censored normal distribution (2)

Notations :

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

Hence,

$$p^j(y_{i_t} = S_{min}) = \Phi\left(\frac{S_{min} - \beta^j x_{i_t}}{\sigma}\right), \qquad (5)$$

The case of a censored normal distribution (2)

Notations :

- $\beta^j x_{i_t} = \beta^j_0 + \beta^j_1 Age_{i_t} + \beta^j_2 Age^2_{i_t} + \beta^j_3 Age^3_{i_t} + \beta^j_4 Age^4_{i_t}$.
- ϕ: density of standard centered normal law.
- Φ: cumulative distribution function of standard centered normal law.

Hence,

$$p^j(y_{i_t} = S_{min}) = \Phi\left(\frac{S_{min} - \beta^j x_{i_t}}{\sigma}\right), \tag{5}$$

$$p^j(y_{i_t}) = \frac{1}{\sigma}\phi\left(\frac{y_{i_t} - \beta^j x_{it}}{\sigma}\right) \quad \text{pour} \quad S_{min} \leq y_{it} \leq S_{max}, \tag{6}$$

The case of a censored normal distribution (2)

Notations :

- $\beta^j x_{i_t} = \beta^j_0 + \beta^j_1 Age_{i_t} + \beta^j_2 Age^2_{i_t} + \beta^j_3 Age^3_{i_t} + \beta^j_4 Age^4_{i_t}$.
- ϕ: density of standard centered normal law.
- Φ: cumulative distribution function of standard centered normal law.

Hence,

$$p^j(y_{i_t} = S_{min}) = \Phi\left(\frac{S_{min} - \beta^j x_{i_t}}{\sigma}\right), \tag{5}$$

$$p^j(y_{i_t}) = \frac{1}{\sigma}\phi\left(\frac{y_{i_t} - \beta^j x_{it}}{\sigma}\right) \quad \text{pour} \quad S_{min} \leq y_{it} \leq S_{max}, \tag{6}$$

$$p^j(y_{i_t} = S_{max}) = 1 - \Phi\left(\frac{S_{max} - \beta^j x_{i_t}}{\sigma}\right). \tag{7}$$

The case of a censored normal distribution (3)

If all the measures are in the interval $[S_{min}, S_{max}]$, we get

The case of a censored normal distribution (3)

If all the measures are in the interval $[S_{min}, S_{max}]$, we get

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

The case of a censored normal distribution (3)

If all the measures are in the interval $[S_{min}, S_{max}]$, we get

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

It is too complicated to get closed-forms equations

The case of a censored normal distribution (3)

If all the measures are in the interval $[S_{min}, S_{max}]$, we get

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

It is too complicated to get closed-forms equations

$\Rightarrow$ quasi-Newton procedure maximum research routine

The case of a censored normal distribution (3)

If all the measures are in the interval $[S_{min}, S_{max}]$, we get

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

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).

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

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

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}}, \tag{9}$$

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

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}}, \tag{9}$$

Finally,

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}} \prod_{t=1}^{T} \phi\left(\frac{y_{i_t} - \beta^j x_{i_t}}{\sigma}\right). \tag{10}$$

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:

$$\mathrm{BIC} = \log(L) - 0,5k\log(N), \tag{11}$$

where k denotes the number of parameters in the model.

Model Selection

Bayesian Information Criterion:

$$\text{BIC} = \log(L) - 0,5k\log(N), \tag{11}$$

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

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \quad (12)$$

Posterior Group-Membership Probabilities

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

Bayes's theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{12}$$

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

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \qquad (12)$$

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.

Use for Model diagnostics (2)

Diagnostic 1: Average Posterior Probability of Assignment

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

Use for Model diagnostics (2)

Diagnostic 1: Average Posterior Probability of Assignment

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

Diagonostic 2: Odds of Correct Classification

$$OCC_j = \frac{AvePP_j/1 - AvePP_j}{\hat{\pi}_j/1 - \hat{\pi}_j}. \tag{13}$$

Use for Model diagnostics (2)

Diagnostic 1: Average Posterior Probability of Assignment

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

Diagonostic 2: Odds of Correct Classification

$$OCC_j = \frac{AvePP_j/1 - AvePP_j}{\hat{\pi}_j/1 - \hat{\pi}_j}. \tag{13}$$

OCC_j should be greater than 5 for all groups.

Use for Model diagnostics (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.

Use for Model diagnostics (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 π_j should be small.

Outline

The data

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

The data

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

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

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)

The data

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

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

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

Dataset transformations

Mathematica programming

Dataset transformations

Mathematica programming

- 1 row per year $\longrightarrow$ 1 row per worker

Dataset transformations

Mathematica programming

- 1 row per year $\rightarrow$ 1 row per worker
- selection of the period we are interested in

Dataset transformations

Mathematica programming

- 1 row per year $\longrightarrow$ 1 row per worker
- selection of the period we are interested in
- taking out the years without years up to a maximum of five years

Dataset transformations

Mathematica programming

- 1 row per year $\longrightarrow$ 1 row per worker
- selection of the period we are interested in
- taking out the years without years up to a maximum of five years
- selection of the people who worked the number of years we are interested in

Dataset transformations

Mathematica programming

- 1 row per year $\rightarrow$ 1 row per worker
- selection of the period we are interested in
- taking out the years without years up to a maximum of five years
- selection of the people who worked the number of years we are interested in

Transformations in SPSS

Dataset transformations

Mathematica programming

- 1 row per year $\rightarrow$ 1 row per worker
- selection of the period we are interested in
- taking out the years without years up to a maximum of five years
- selection of the people who worked the number of years we are interested in

Transformations in SPSS

- elimination of all the workers who had monthly salaries above 15.000

Dataset transformations

Mathematica programming

- 1 row per year $\rightarrow$ 1 row per worker
- selection of the period we are interested in
- taking out the years without years up to a maximum of five years
- selection of the people who worked the number of years we are interested in

Transformations in SPSS

- elimination of all the workers who had monthly salaries above 15.000
- transforming all the salaries above 7.577 to 7.577

Dataset transformations

Mathematica programming

- 1 row per year $\rightarrow$ 1 row per worker
- selection of the period we are interested in
- taking out the years without years up to a maximum of five years
- selection of the people who worked the number of years we are interested in

Transformations in SPSS

- elimination of all the workers who had monthly salaries above 15.000
- transforming all the salaries above 7.577 to 7.577
- creation of the time variables necessary for the Proc Traj procedure

Proc Traj procedure

Selection of the time period for macroeconomic reasons

Proc Traj procedure

Selection of the time period for macroeconomic reasons (Crisis in the steel industry and emergence of the financial market place of Luxembourg)

Proc Traj procedure

Selection of the time period for macroeconomic reasons (Crisis in the steel industry and emergence of the financial market place of Luxembourg)

20 years of work for workers beginning their carrier between 1982 and 1987

Proc Traj procedure

Selection of the time period for macroeconomic reasons (Crisis in the steel industry and emergence of the financial market place of Luxembourg)

20 years of work for workers beginning their carrier between 1982 and 1987

Proc Traj Macro:

Proc Traj procedure

Selection of the time period for macroeconomic reasons (Crisis in the steel industry and emergence of the financial market place of Luxembourg)

20 years of work for workers beginning their carrier between 1982 and 1987

Proc Traj Macro:

```
DATA TEST;
      INPUT ID O1-O20 T1-T20;
      CARDS;
data
RUN;
```

Proc Traj procedure

Selection of the time period for macroeconomic reasons (Crisis in the steel industry and emergence of the financial market place of Luxembourg)

20 years of work for workers beginning their carrier between 1982 and 1987

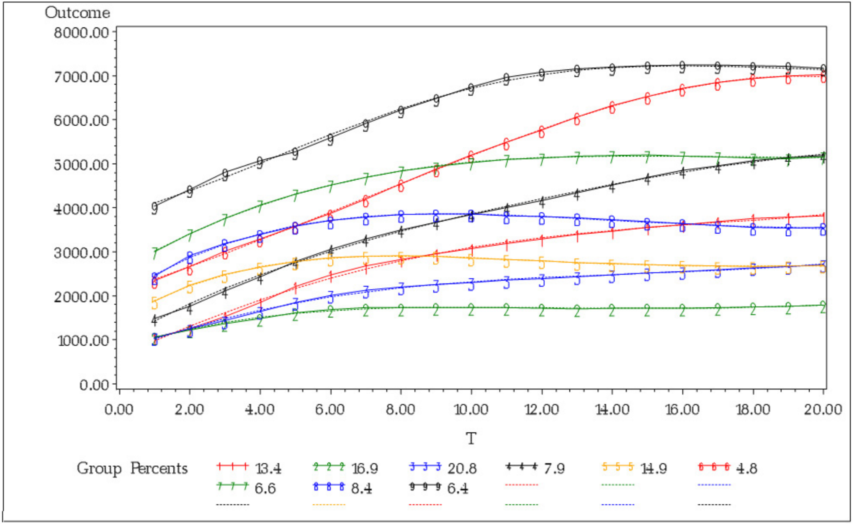
Proc Traj Macro:

```
DATA TEST;
      INPUT ID O1-O20 T1-T20;
      CARDS;
data
RUN;

PROC TRAJ DATA=TEST OUTPLOT=OP OUTSTAT=OS OUT=OF
OUTEST=OE ITDETAIL;
      ID ID; VAR O1-O20; INDEP T1-T20;
      MODEL CNORM; MAX 8000; NGROUPS 6; ORDER 4 4 4 4 4 4;
RUN;
```

Results for 9 groups (1)

Results for 9 groups (1)



Results for 9 groups (2)

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

1^{st} group

13.4 % of the population

1^{st} group

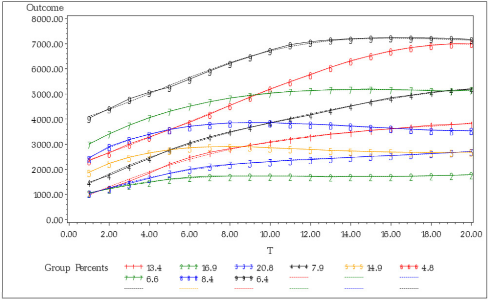
13.4 % of the population

$$P(x) = 590 + 388t - 14t^2 + 0.009t^4$$

1^{st} group

13.4 % of the population

$$P(x) = 590 + 388t - 14t^2 + 0.009t^4$$



1^{st} group

Age_initial

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	13,00	2	,1	,1	,1
	15,00	318	10,6	10,7	10,7
	16,00	540	18,1	18,1	28,8
	17,00	556	18,6	18,6	47,5
	18,00	494	16,5	16,6	64,0
	19,00	348	11,7	11,7	75,7
	20,00	187	6,3	6,3	82,0
	21,00	152	5,1	5,1	87,1
	22,00	90	3,0	3,0	90,1
	23,00	74	2,5	2,5	92,6
	24,00	37	1,2	1,2	93,8
	25,00	42	1,4	1,4	95,2
	26,00	27	,9	,9	96,1
	27,00	18	,6	,6	96,7
	28,00	16	,5	,5	97,3
	29,00	18	,6	,6	97,9
	30,00	9	,3	,3	98,2
	31,00	10	,3	,3	98,5
	32,00	11	,4	,4	98,9
	33,00	5	,2	,2	99,0
	34,00	2	,1	,1	99,1
	35,00	3	,1	,1	99,2
	36,00	6	,2	,2	99,4
	37,00	5	,2	,2	99,6
	38,00	3	,1	,1	99,7
	39,00	2	,1	,1	99,7
	40,00	3	,1	,1	99,8
	41,00	3	,1	,1	99,9
	43,00	1	,0	,0	100,0
	46,00	1	,0	,0	100,0
	Total	2983	99,9	100,0	
Missing	System	3	,1		
Total		2986	100,0		

1^{st} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	1213	57,8	57,8	57,8
	employé privé	887	42,2	42,2	100,0
	Total	2100	100,0	100,0	

1^{st} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	1213	57,8	57,8	57,8
	employé privé	887	42,2	42,2	100,0
	Total	2100	100,0	100,0	

Women:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	43	4,9	4,9	4,9
	employé privé	840	95,1	95,1	100,0
	Total	883	100,0	100,0	

2^{nd} group

16.9 % of the population

2^{nd} group

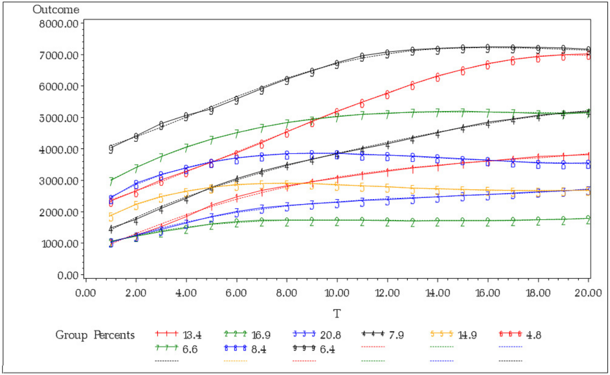
16.9 % of the population

$$P(x) = 785 + 278t - 28t^2 + 1.18t^3 - 0.016t^4$$

2^{nd} group

16.9 % of the population

$$P(x) = 785 + 278t - 28t^2 + 1.18t^3 - 0.016t^4$$



2^{nd} group

Age_initial

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15,00	588	15,7	15,7	15,7
	16,00	864	23,1	23,1	38,8
	17,00	489	13,1	13,1	51,9
	18,00	316	8,4	8,4	60,3
	19,00	277	7,4	7,4	67,8
	20,00	209	5,6	5,6	73,3
	21,00	148	4,0	4,0	77,3
	22,00	111	3,0	3,0	80,3
	23,00	107	2,9	2,9	83,1
	24,00	81	2,2	2,2	85,3
	25,00	68	1,8	1,8	87,1
	26,00	59	1,6	1,6	88,7
	27,00	57	1,5	1,5	90,2
	28,00	51	1,4	1,4	91,6
	29,00	41	1,1	1,1	92,7
	30,00	31	,8	,8	93,5
	31,00	32	,9	,9	94,4
	32,00	32	,9	,9	95,2
	33,00	23	,6	,6	95,8
	34,00	20	,5	,5	96,4
	35,00	22	,6	,6	97,0
	36,00	23	,6	,6	97,6
	37,00	24	,6	,6	98,2
	38,00	14	,4	,4	98,6
	39,00	14	,4	,4	99,0
	40,00	15	,4	,4	99,4
	41,00	10	,3	,3	99,6
	42,00	5	,1	,1	99,8
	43,00	3	,1	,1	99,8
	44,00	2	,1	,1	99,9
	45,00	4	,1	,1	100,0
	Total	3740	99,9	100,0	
Missing	System	2	,1		
Total		3742	100,0		

2^{nd} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	994	82,8	82,8	82,8
	employé privé	206	17,2	17,2	100,0
	Total	1200	100,0	100,0	

2^{nd} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	994	82,8	82,8	82,8
	employé privé	206	17,2	17,2	100,0
	Total	1200	100,0	100,0	

Women:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	1453	57,2	57,2	57,2
	employé privé	1087	42,8	42,8	100,0
	Total	2540	100,0	100,0	

3^{rd} group

20.8 % of the population

3^{rd} group

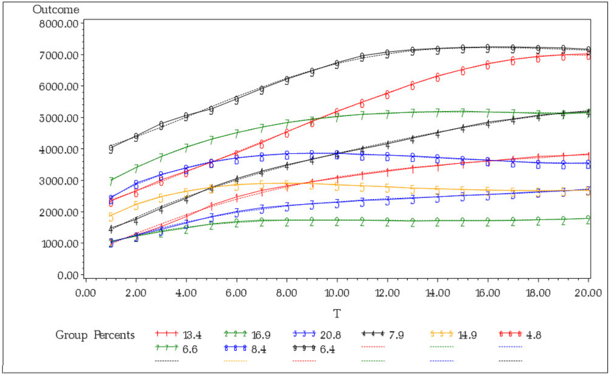
20.8 % of the population

$P(x) = 709 + 318t - 21.5t^2 + 0.62t^3$

3^{rd} group

20.8 % of the population

$$P(x) = 709 + 318t - 21.5t^2 + 0.62t^3$$



3^{rd} group

Age_initial

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10,00	1	,0	,0	,0
	13,00	2	,0	,0	,1
	14,00	3	,1	,1	,1
	15,00	635	13,7	13,7	13,9
	16,00	1045	22,6	22,6	36,5
	17,00	700	15,1	15,1	51,6
	18,00	542	11,7	11,7	63,3
	19,00	358	7,7	7,7	71,1
	20,00	288	6,2	6,2	77,3
	21,00	195	4,2	4,2	81,5
	22,00	168	3,6	3,6	85,2
	23,00	140	3,0	3,0	88,2
	24,00	98	2,1	2,1	90,3
	25,00	81	1,8	1,8	92,1
	26,00	68	1,5	1,5	93,5
	27,00	42	,9	,9	94,4
	28,00	47	1,0	1,0	95,5
	29,00	30	,6	,6	96,1
	30,00	30	,6	,6	96,8
	31,00	27	,6	,6	97,3
	32,00	14	,3	,3	97,6
	33,00	16	,3	,3	98,0
	34,00	13	,3	,3	98,3
	35,00	14	,3	,3	98,6
	36,00	16	,3	,3	98,9
	37,00	11	,2	,2	99,2
	38,00	9	,2	,2	99,4
	39,00	10	,2	,2	99,6
	41,00	8	,2	,2	99,7
	42,00	4	,1	,1	99,8
	43,00	3	,1	,1	99,9
	44,00	3	,1	,1	100,0
	45,00	1	,0	,0	100,0
	48,00	1	,0	,0	100,0
	Total	4623	100,0	100,0	
Missing	System	1	,0		
Total		4624	100,0		

3^{rd} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	2452	78,2	78,2	78,2
	employé privé	684	21,8	21,8	100,0
	Total	3136	100,0	100,0	

3^{rd} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	2452	78,2	78,2	78,2
	employé privé	684	21,8	21,8	100,0
	Total	3136	100,0	100,0	

Women:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	539	36,2	36,2	36,2
	employé privé	948	63,8	63,8	100,0
	Total	1487	100,0	100,0	

4^{th} group

7.9 % of the population

4^{th} group

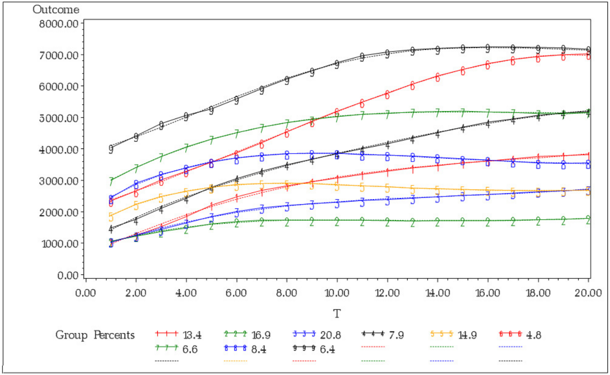
7.9 % of the population

$$P(x) = 976 + 474t - 29.6t^2 - 0.029t^4$$

4^{th} group

7.9 % of the population

$$P(x) = 976 + 474t - 29.6t^2 - 0.029t^4$$



4^{th} group

Age_initial

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15,00	40	2,3	2,3	2,3
	16,00	94	5,4	5,4	7,7
	17,00	221	12,7	12,7	20,4
	18,00	303	17,4	17,4	37,7
	19,00	303	17,4	17,4	55,1
	20,00	220	12,6	12,6	67,7
	21,00	160	9,2	9,2	76,9
	22,00	114	6,5	6,5	83,4
	23,00	66	3,8	3,8	87,2
	24,00	39	2,2	2,2	89,4
	25,00	38	2,2	2,2	91,6
	26,00	27	1,5	1,5	93,2
	27,00	31	1,8	1,8	95,0
	28,00	22	1,3	1,3	96,2
	29,00	16	,9	,9	97,1
	30,00	11	,6	,6	97,8
	31,00	12	,7	,7	98,5
	32,00	5	,3	,3	98,7
	33,00	2	,1	,1	98,9
	34,00	4	,2	,2	99,1
	35,00	5	,3	,3	99,4
	36,00	1	,1	,1	99,4
	37,00	2	,1	,1	99,5
	39,00	1	,1	,1	99,6
	41,00	1	,1	,1	99,7
	42,00	1	,1	,1	99,7
	43,00	2	,1	,1	99,8
	46,00	2	,1	,1	99,9
	49,00	1	,1	,1	100,0
	Total	1744	99,9	100,0	
Missing	System	2	,1		
Total		1746	100,0		

4^{th} group

Résidence et nationalité

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	résident de nationalité luxembourgeoise	1211	69,4	69,4	69,4
	résident étranger	260	14,9	14,9	84,3
	frontalier	273	15,6	15,7	100,0
	Total	1744	99,9	100,0	
Missing	System	2	,1		
Total		1746	100,0		

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	117	6,7	6,7	6,7
	employé privé	1627	93,2	93,3	100,0
	Total	1744	99,9	100,0	
Missing	System	2	,1		
Total		1746	100,0		

4^{th} group

Résidence et nationalité

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	résident de nationalité luxembourgeoise	1211	69,4	69,4	69,4
	résident étranger	260	14,9	14,9	84,3
	frontalier	273	15,6	15,7	100,0
	Total	1744	99,9	100,0	
Missing	System	2	,1		
Total		1746	100,0		

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	117	6,7	6,7	6,7
	employé privé	1627	93,2	93,3	100,0
	Total	1744	99,9	100,0	
Missing	System	2	,1		
Total		1746	100,0		

5^{th} group

14.9 % of the population

5^{th} group

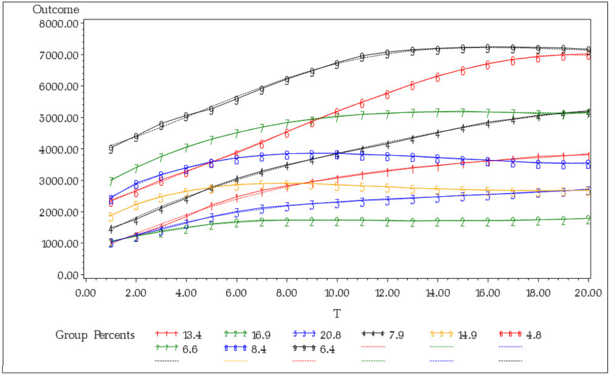
14.9 % of the population

$$P(x) = 1452 + 490t - 29.6t^2 + 1.38t^3 - 0.028t^4$$

5^{th} group

14.9 % of the population

$$P(x) = 1452 + 490t - 29.6t^2 + 1.38t^3 - 0.028t^4$$



5^{th} group

Age_initial

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15,00	13	,4	,4	,4
	16,00	65	2,0	2,0	2,4
	17,00	147	4,5	4,5	6,8
	18,00	276	8,4	8,4	15,2
	19,00	306	9,3	9,3	24,5
	20,00	358	10,9	10,9	35,4
	21,00	332	10,1	10,1	45,5
	22,00	264	8,0	8,0	53,5
	23,00	217	6,6	6,6	60,1
	24,00	172	5,2	5,2	65,3
	25,00	158	4,8	4,8	70,1
	26,00	161	4,9	4,9	75,0
	27,00	130	3,9	4,0	79,0
	28,00	100	3,0	3,0	82,0
	29,00	94	2,9	2,9	84,9
	30,00	69	2,1	2,1	87,0
	31,00	65	2,0	2,0	88,9
	32,00	68	2,1	2,1	91,0
	33,00	61	1,9	1,9	92,9
	34,00	48	1,5	1,5	94,3
	35,00	56	1,7	1,7	96,0
	36,00	36	1,1	1,1	97,1
	37,00	23	,7	,7	97,8
	38,00	21	,6	,6	98,5
	39,00	19	,6	,6	99,0
	40,00	8	,2	,2	99,3
	41,00	8	,2	,2	99,5
	42,00	6	,2	,2	99,7
	43,00	5	,2	,2	99,8
	44,00	5	,2	,2	100,0
	Total	3291	99,9	100,0	
Missing	System	2	,1		
Total		3293	100,0		

5^{th} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	1910	81,0	81,0	81,0
	employé privé	449	19,0	19,0	100,0
	Total	2359	100,0	100,0	

5^{th} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	1910	81,0	81,0	81,0
	employé privé	449	19,0	19,0	100,0
	Total	2359	100,0	100,0	

Women:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	71	7,6	7,6	7,6
	employé privé	861	92,4	92,4	100,0
	Total	932	100,0	100,0	

6^{th} group

4.8 % of the population

6^{th} group

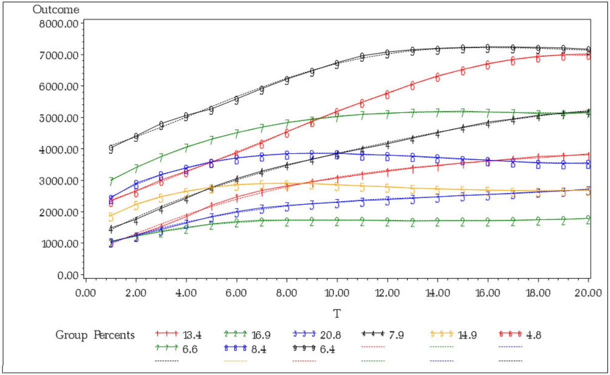
4.8 % of the population

$P(x) = 2089 - 0.017t^4$

6^{th} group

4.8 % of the population

$$P(x) = 2089 - 0.017t^4$$



6^{th} group

Age_initial

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15,00	1	,1	,1	,1
	16,00	8	,8	,8	,9
	17,00	38	3,6	3,6	4,5
	18,00	56	5,3	5,3	9,8
	19,00	103	9,7	9,8	19,5
	20,00	140	13,2	13,3	32,8
	21,00	175	16,5	16,6	49,4
	22,00	103	9,7	9,8	59,1
	23,00	91	8,6	8,6	67,8
	24,00	66	6,2	6,3	74,0
	25,00	64	6,0	6,1	80,1
	26,00	52	4,9	4,9	85,0
	27,00	37	3,5	3,5	88,5
	28,00	36	3,4	3,4	91,9
	29,00	18	1,7	1,7	93,6
	30,00	12	1,1	1,1	94,8
	31,00	12	1,1	1,1	95,9
	32,00	8	,8	,8	96,7
	33,00	7	,7	,7	97,3
	34,00	6	,6	,6	97,9
	35,00	5	,5	,5	98,4
	36,00	4	,4	,4	98,8
	37,00	4	,4	,4	99,1
	38,00	1	,1	,1	99,2
	41,00	3	,3	,3	99,5
	43,00	3	,3	,3	99,8
	45,00	2	,2	,2	100,0
	Total	1055	99,2	100,0	
Missing	System	8	,8		
Total		1063	100,0		

6^{th} group

Résidence et nationalité

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	résident de nationalité luxembourgeoise	651	61,2	61,7	61,7
	résident étranger	184	17,3	17,4	79,1
	frontalier	220	20,7	20,9	100,0
	Total	1055	99,2	100,0	
Missing	System	8	,8		
Total		1063	100,0		

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	3	,3	,3	,3
	employé privé	1052	99,0	99,7	100,0
	Total	1055	99,2	100,0	
Missing	System	8	,8		
Total		1063	100,0		

7^{th} group

6.6 % of the population

7^{th} group

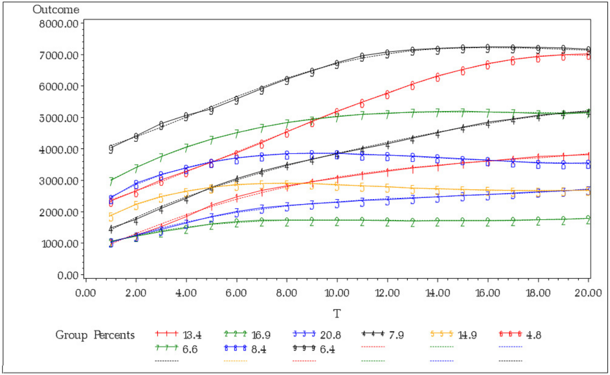
6.6 % of the population

$P(x) = 2556 + 484t - 29.9t^2 + 0.66t^3$

7^{th} group

6.6 % of the population

$$P(x) = 2556 + 484t - 29.9t^2 + 0.66t^3$$



7^{th} group

Age_initial

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	16,00	1	,1	,1	,1
	17,00	8	,5	,5	,6
	18,00	34	2,3	2,3	3,0
	19,00	116	7,9	8,0	10,9
	20,00	168	11,5	11,5	22,5
	21,00	180	12,3	12,4	34,8
	22,00	191	13,1	13,1	47,9
	23,00	164	11,2	11,3	59,2
	24,00	115	7,9	7,9	67,1
	25,00	101	6,9	6,9	74,0
	26,00	76	5,2	5,2	79,3
	27,00	44	3,0	3,0	82,3
	28,00	37	2,5	2,5	84,8
	29,00	29	2,0	2,0	86,8
	30,00	23	1,6	1,6	88,4
	31,00	16	1,1	1,1	89,5
	32,00	25	1,7	1,7	91,2
	33,00	17	1,2	1,2	92,4
	34,00	16	1,1	1,1	93,5
	35,00	20	1,4	1,4	94,8
	36,00	20	1,4	1,4	96,2
	37,00	11	,8	,8	97,0
	38,00	13	,9	,9	97,9
	39,00	9	,6	,6	98,5
	40,00	9	,6	,6	99,1
	41,00	4	,3	,3	99,4
	42,00	4	,3	,3	99,7
	43,00	2	,1	,1	99,8
	44,00	2	,1	,1	99,9
	45,00	1	,1	,1	100,0
	Total	1456	99,6	100,0	
Missing	System	6	,4		
Total		1462	100,0		

7^{th} group

Résidence et nationalité

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	résident de nationalité luxembourgeoise	632	43,2	43,4	43,4
	résident étranger	273	18,7	18,8	62,2
	frontalier	551	37,7	37,8	100,0
	Total	1456	99,6	100,0	
Missing	System	6	,4		
Total		1462	100,0		

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	19	1,3	1,3	1,3
	employé privé	1437	98,3	98,7	100,0
	Total	1456	99,6	100,0	
Missing	System	6	,4		
Total		1462	100,0		

8^{th} group

8.4 % of the population

8^{th} group

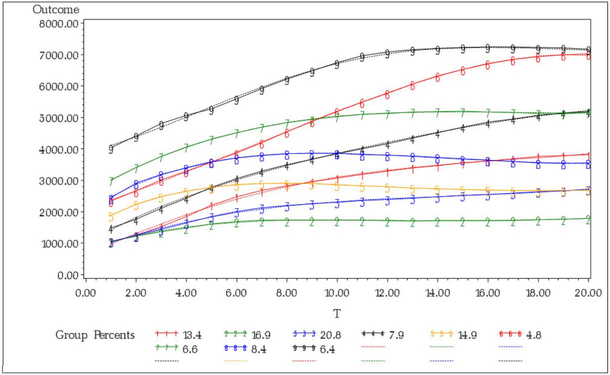
8.4 % of the population

$$P(x) = 1987 + 537t - 52.7t^2 + 2.06t^3 - 0.028t^4$$

8^{th} group

8.4 % of the population

$$P(x) = 1987 + 537t - 52.7t^2 + 2.06t^3 - 0.028t^4$$



8^{th} group

Age_initial

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	15,00	3	,2	,2	,2
	16,00	17	,9	,9	1,1
	17,00	58	3,1	3,1	4,2
	18,00	123	6,6	6,6	10,8
	19,00	204	10,9	10,9	21,7
	20,00	210	11,2	11,3	33,0
	21,00	225	12,0	12,1	45,0
	22,00	187	10,0	10,0	55,0
	23,00	134	7,2	7,2	62,2
	24,00	107	5,7	5,7	68,0
	25,00	95	5,1	5,1	73,0
	26,00	65	3,5	3,5	76,5
	27,00	56	3,0	3,0	79,5
	28,00	60	3,2	3,2	82,7
	29,00	49	2,6	2,6	85,4
	30,00	34	1,8	1,8	87,2
	31,00	24	1,3	1,3	88,5
	32,00	33	1,8	1,8	90,2
	33,00	38	2,0	2,0	92,3
	34,00	18	1,0	1,0	93,2
	35,00	19	1,0	1,0	94,3
	36,00	24	1,3	1,3	95,6
	37,00	17	,9	,9	96,5
	38,00	14	,7	,8	97,2
	39,00	13	,7	,7	97,9
	40,00	13	,7	,7	98,6
	41,00	7	,4	,4	99,0
	42,00	8	,4	,4	99,4
	43,00	1	,1	,1	99,5
	44,00	8	,4	,4	99,9
	46,00	2	,1	,1	100,0
	Total	1866	99,8	100,0	
Missing	System	3	,2		
Total		1869	100,0		

8^{th} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	594	52,6	52,6	52,6
	employé privé	535	47,4	47,4	100,0
	Total	1129	100,0	100,0	

8^{th} group

Men:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	594	52,6	52,6	52,6
	employé privé	535	47,4	47,4	100,0
	Total	1129	100,0	100,0	

Women:

Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	6	,8	,8	,8
	employé privé	731	99,2	99,2	100,0
	Total	737	100,0	100,0	

9^{th} group

6.4 % of the population

9^{th} group

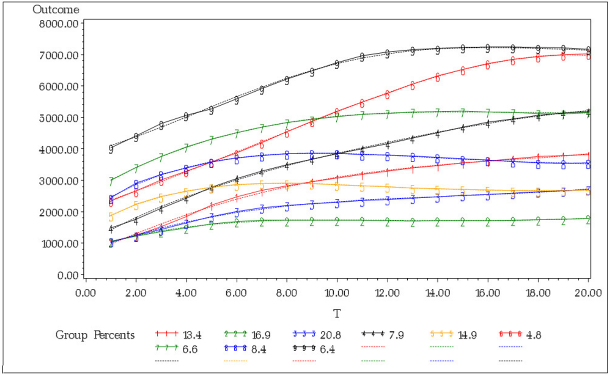
6.4 % of the population

$P(x) = 3873 + 206t + 30t2 - 2.89t^3 + 0.06t^4$

9^{th} group

6.4 % of the population

$$P(x) = 3873 + 206t + 30t2 - 2.89t^3 + 0.06t^4$$



9th group

Age_initial

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	17,00	1	,1	,1	,1
	18,00	4	,3	,3	,4
	19,00	19	1,3	1,4	1,7
	20,00	35	2,5	2,5	4,2
	21,00	68	4,8	4,8	9,0
	22,00	107	7,5	7,6	16,6
	23,00	123	8,7	8,7	25,4
	24,00	167	11,8	11,9	37,2
	25,00	150	10,6	10,7	47,9
	26,00	107	7,5	7,6	55,5
	27,00	92	6,5	6,5	62,0
	28,00	79	5,6	5,6	67,7
	29,00	65	4,6	4,6	72,3
	30,00	54	3,8	3,8	76,1
	31,00	48	3,4	3,4	79,5
	32,00	41	2,9	2,9	82,4
	33,00	33	2,3	2,3	84,8
	34,00	28	2,0	2,0	86,8
	35,00	38	2,7	2,7	89,5
	36,00	25	1,8	1,8	91,3
	37,00	24	1,7	1,7	93,0
	38,00	18	1,3	1,3	94,2
	39,00	19	1,3	1,4	95,6
	40,00	19	1,3	1,4	96,9
	41,00	12	,8	,9	97,8
	42,00	10	,7	,7	98,5
	43,00	7	,5	,5	99,0
	44,00	7	,5	,5	99,5
	45,00	4	,3	,3	99,8
	46,00	3	,2	,2	100,0
	Total	1407	99,2	100,0	
Missing	System	11	,8		
Total		1418	100,0		

9^{th} group

Résidence et nationalité

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	résident de nationalité luxembourgeoise	468	33,0	33,3	33,3
	résident étranger	475	33,5	33,8	67,0
	frontalier	464	32,7	33,0	100,0
	Total	1407	99,2	100,0	
Missing	System	11	,8		
Total		1418	100,0		

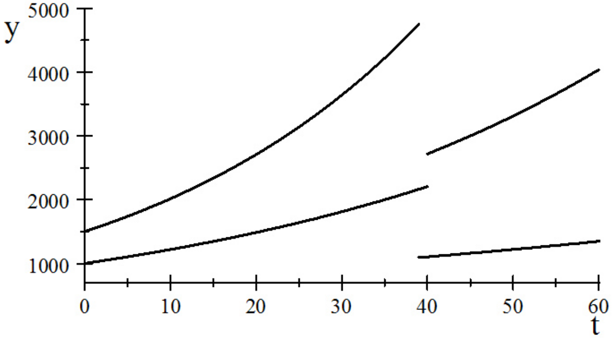
Classe d'employé

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ouvrier	1	,1	,1	,1
	employé privé	1406	99,2	99,9	100,0
	Total	1407	99,2	100,0	
Missing	System	11	,8		
Total		1418	100,0		

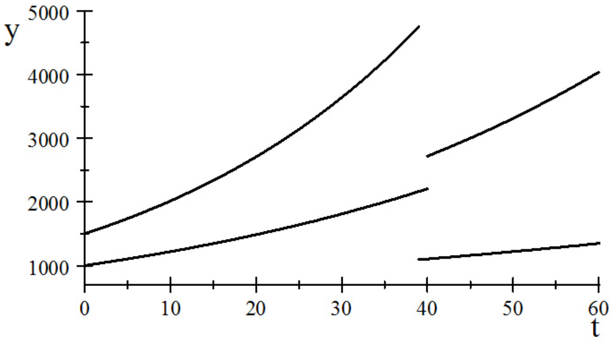
Outline

1. Nagin's Finite Mixture Model
2. The Luxemburgish salary trajectories
3. Description of the groups
4. Economic Modeling
5. Outlook

Dummy example

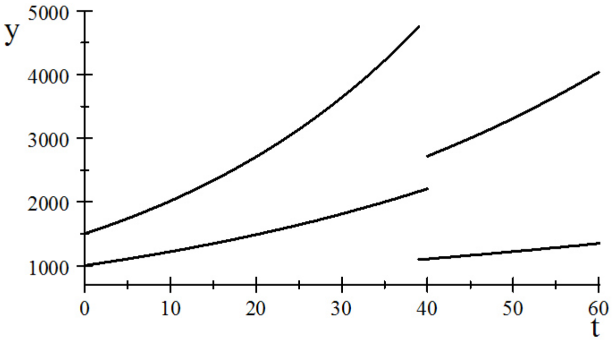


Dummy example



2 trajectories S^1 and S^2 with groupe size 60% and 40% of the population.

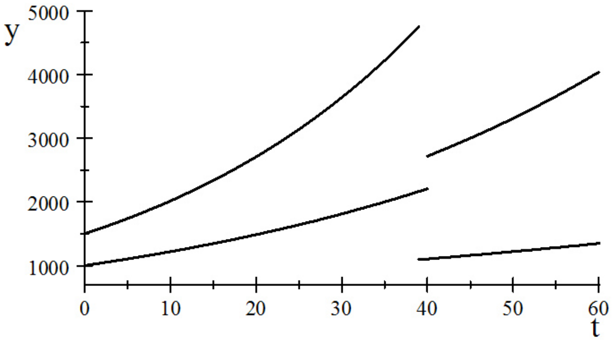
Dummy example



2 trajectories S^1 and S^2 with groupe size 60% and 40% of the population.

Length of the professional life: $T = 40$ years.

Dummy example



2 trajectories S^1 and S^2 with groupe size 60% and 40% of the population.

Length of the professional life: $T = 40$ years.

Additional life expectancy: $T^* = 20$ years.

Hypotheses

Salaries grow linearly, S^1 with a starting value of 1500 and a growth coefficient of 3 %, S^2 with a starting value of 1000 and a growth coefficient of 2 %.

Hypotheses

Salaries grow linearly, S^1 with a starting value of 1500 and a growth coefficient of 3 %, S^2 with a starting value of 1000 and a growth coefficient of 2 %.

Pensions grow also linearly, S^1 with a starting value of 2718 and a growth coefficient of 2%, S^2 with a starting value of 1104 and a growth coefficient of 1%.

Hypotheses

Salaries grow linearly, S^1 with a starting value of 1500 and a growth coefficient of 3 %, S^2 with a starting value of 1000 and a growth coefficient of 2 %.

Pensions grow also linearly, S^1 with a starting value of 2718 and a growth coefficient of 2%, S^2 with a starting value of 1104 and a growth coefficient of 1%.

Luxembourg adopts a repartition model, which means that the current pensions are paid with the tax incomes from the current workers. Each generation hence pays the pension for the generation before it.

Replacement rate in the repartition model

Replacement rate = first pension / last salary

Replacement rate in the repartition model

Replacement rate = first pension / last salary

For S^1, $t_{rep} = \frac{2718}{1500(1+0.03)^{39}} \simeq 57\%$.

Replacement rate in the repartition model

Replacement rate = first pension / last salary

For S^1, $t_{rep} = \frac{2718}{1500(1+0.03)^{39}} \simeq 57\%$.

For S^2, $t_{rep} = \frac{1104}{1000(1+0.02)^{39}} = 50\%$.

Replacement rate in the repartition model

Replacement rate = first pension / last salary

For S^1, $t_{rep} = \frac{2718}{1500(1+0.03)^{39}} \simeq 57\%$.

For S^2, $t_{rep} = \frac{1104}{1000(1+0.02)^{39}} = 50\%$.

A worker who's trajectory is S^1 with a probability of 75 % and S^2 with a probability of 25 % has a replacement rate of

$$t_{rep} = \frac{0.75 \times 2718 + 0.25 \times 1104}{0.75 \times 1500(1+0.03)^{39} + 0.25 \times 1000(1+0.02)^{39}} \simeq 56\%.$$

Coverage potential in a repartition & capitalization model

We want to know the sum a that we have to put every year in a saving account to get a desired replacement rate t_{aim}.

Coverage potential in a repartition & capitalization model

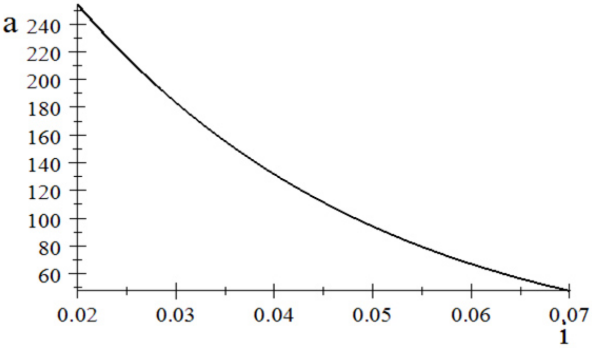
We want to know the sum *a* that we have to put every year in a saving account to get a desired replacement rate t_{aim}.

a of course depends on the account's interest rate *i*.

Coverage potential in a repartition & capitalization model

We want to know the sum *a* that we have to put every year in a saving account to get a desired replacement rate t_{aim}.

a of course depends on the account's interest rate *i*.



If $i \sim U(2\%; 7\%)$, *a* varies between 46 euros and 252 euros with a mean of 124 euros.

Capitalization effort coefficient

τ_2 = Sum of the salaries on the salary trajectory / sum of the pensions on the pension trajectory = 16.5 on average.

Capitalization effort coefficient

τ_2 = Sum of the salaries on the salary trajectory / sum of the pensions on the pension trajectory = 16.5 on average.

That means that you need 16.5 euros from the salary to get 1 euro by capitalization for the pension.

Capitalization effort coefficient

τ_2 = Sum of the salaries on the salary trajectory / sum of the pensions on the pension trajectory = 16.5 on average.

That means that you need 16.5 euros from the salary to get 1 euro by capitalization for the pension.

In fact, if $i \sim U(2\%; 7\%)$, τ_2 varies between 18 euros and 15 euros.

Capitalization effort coefficient

τ_2 = Sum of the salaries on the salary trajectory / sum of the pensions on the pension trajectory = 16.5 on average.

That means that you need 16.5 euros from the salary to get 1 euro by capitalization for the pension.

In fact, if $i \sim U(2\%; 7\%)$, τ_2 varies between 18 euros and 15 euros.

τ_2 depends on a, hence a not only allows to get the desired replacement rate, but a also serves to control the variability of the capitalization effort coefficient.

Capitalization effort coefficient

$\tau_2 =$ Sum of the salaries on the salary trajectory / sum of the pensions on the pension trajectory $= 16.5$ on average.

That means that you need 16.5 euros from the salary to get 1 euro by capitalization for the pension.

In fact, if $i \sim U(2\%; 7\%)$, τ_2 varies between 18 euros and 15 euros.

τ_2 depends on a, hence a not only allows to get the desired replacement rate, but a also serves to control the variability of the capitalization effort coefficient.

We need a compromise between a high replacement rate and a small capitalization effort coefficient.

Repartition effort coefficient

τ_1 = weighted mean of the salaries on the salary trajectory / weighted mean of the pensions in the repartition model on the pension trajectory = 2.7 on average.

Repartition effort coefficient

τ_1 = weighted mean of the salaries on the salary trajectory / weighted mean of the pensions in the repartition model on the pension trajectory = 2.7 on average.

That means that the active worker have to earn 2.7 euros to pay 1 euro of pesnion by repartition. pause

τ_1 depends on the demographic rate d.

Repartition effort coefficient

τ_1 = weighted mean of the salaries on the salary trajectory / weighted mean of the pensions in the repartition model on the pension trajectory = 2.7 on average.

That means that the active worker have to earn 2.7 euros to pay 1 euro of pesnion by repartition. pause

τ_1 depends on the demographic rate d. In fact, if $d \sim U(0\%; 5\%)$, τ_1 varies between 6.7 euros and 1.6 euros.

Repartition effort coefficient

τ_1 = weighted mean of the salaries on the salary trajectory / weighted mean of the pensions in the repartition model on the pension trajectory = 2.7 on average.

That means that the active worker have to earn 2.7 euros to pay 1 euro of pesnion by repartition. pause

τ_1 depends on the demographic rate d. In fact, if $d \sim U(0\%; 5\%)$, τ_1 varies between 6.7 euros and 1.6 euros.

	Market risk	Demographic risk
Repartition	Negligeable	Extreme
Capitalization	Extreme	Negligeable

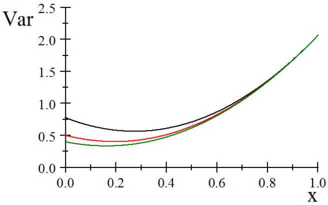
Global effort coefficient

$$\tau = x\tau_1 + (1 - x)\tau_2$$

is the number of euros necessary to pay 1 euro for the pension.

Here x euros come from repartition and $1 - x$ euros from capitalization.

We want to limit the risk of the hybrid system without reducing the pension and in the same time minimize the capitalization effort.



Aim and solution

Aim : volatility of τ = (volatility of τ_1)$/k$.

Aim and solution

Aim : volatility of τ = (volatility of τ_1)$/k$.

Solution:

$$x = \frac{1}{k^2}.$$

Outline

Further research

- Do the computations for the 9 groups of real trajectories instead of the dummy ones.

Further research

- Do the computations for the 9 groups of real trajectories instead of the dummy ones.
- Add macroeconomic covariates to the finite mixture model to get trajectories depending on the general economical situation.

Further research

- Do the computations for the 9 groups of real trajectories instead of the dummy ones.
- Add macroeconomic covariates to the finite mixture model to get trajectories depending on the general economical situation.
- Extend the Nagin model to exponential trajectories and Pareto distributions.

Further research

- Do the computations for the 9 groups of real trajectories instead of the dummy ones.
- Add macroeconomic covariates to the finite mixture model to get trajectories depending on the general economical situation.
- Extend the Nagin model to exponential trajectories and Pareto distributions.
- Get another decision criterion for the optimal number of groups in the model.

Further research

- Do the computations for the 9 groups of real trajectories instead of the dummy ones.
- Add macroeconomic covariates to the finite mixture model to get trajectories depending on the general economical situation.
- Extend the Nagin model to exponential trajectories and Pareto distributions.
- Get another decision criterion for the optimal number of groups in the model.
- Combining dynamic clustering and dynamic factor analysis.