Robustness of groups and trajectories in Nagin's finite mixture model

Jang SCHILTZ (University of Luxembourg)

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

June 7, 2012

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 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.

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

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

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

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

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

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

<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

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

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

- finite : sums across a finite number of groups

LUXEMBOURG SCHOOL OF FINANCE

The Likelihood Function (2)

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

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

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

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

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

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 case of a censored normal distribution

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

The case of a censored normal distribution

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

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

The case of a censored normal distribution

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

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

It is too complicated to get closed-forms equations

The case of a censored normal distribution

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

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

It is too complicated to get closed-forms equations

$\Rightarrow$ quasi-Newton procedure maximum research routine

LUXEMBOURG SCHOOL OF FINANCE

The case of a censored normal distribution

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

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

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

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

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}}, \qquad (3)$$

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

$$\pi_j = \frac{e^{\theta_j}}{\sum_{j=1}^{r} e^{\theta_j}}, \tag{3}$$

Finally,

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

Model Selection

Bayesian Information Criterion:

Model Selection

Bayesian Information Criterion:

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

where k denotes the number of parameters in the model.

Model Selection

Bayesian Information Criterion:

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

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

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{6}$$

Posterior Group-Membership Probabilities

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

Bayes's theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{6}$$

Bigger groups have on average larger probability estimates.

Posterior Group-Membership Probabilities

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

Bayes's theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{6}$$

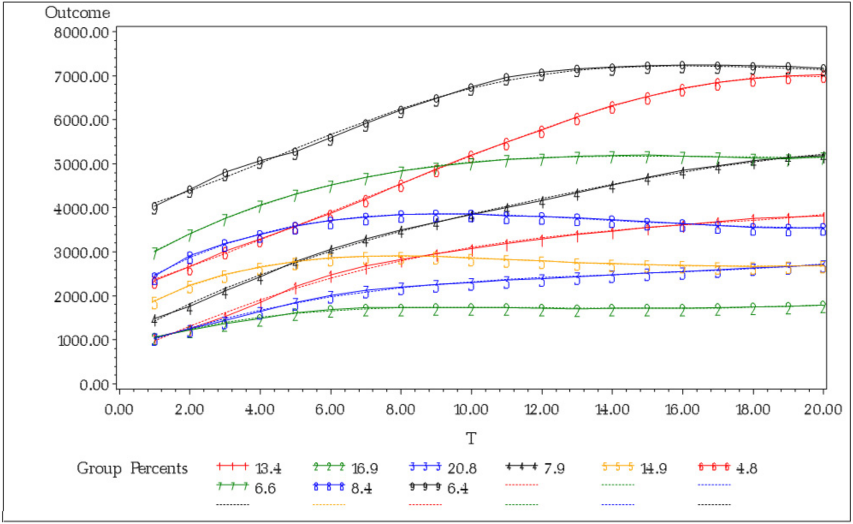
Bigger groups have on average larger probability estimates.

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

LUXEMBOURG SCHOOL OF FINANCE

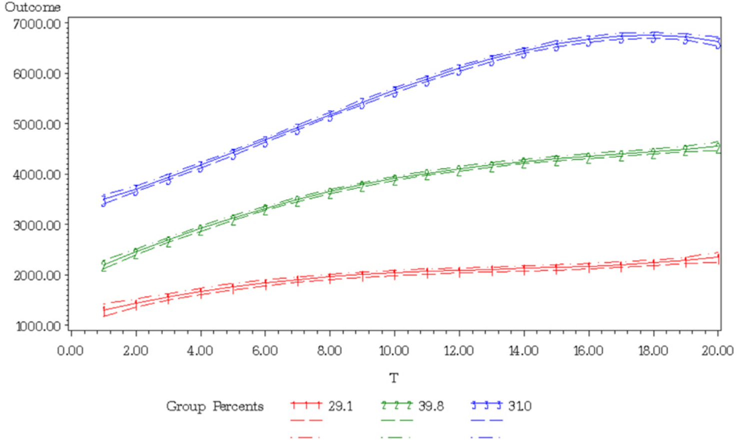
Application: Salary trajectories

Application: Salary trajectories

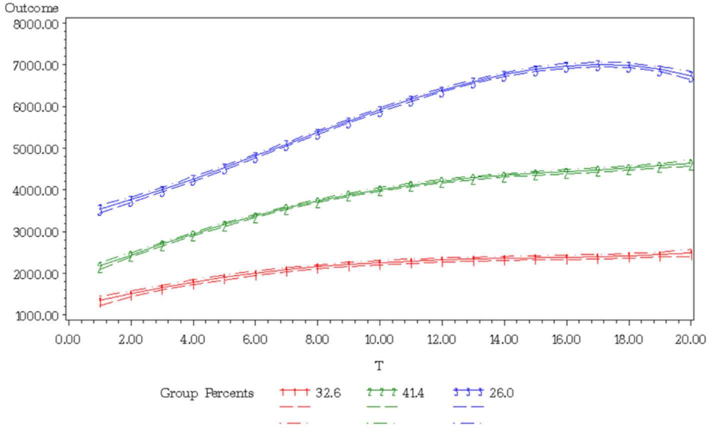


Outline

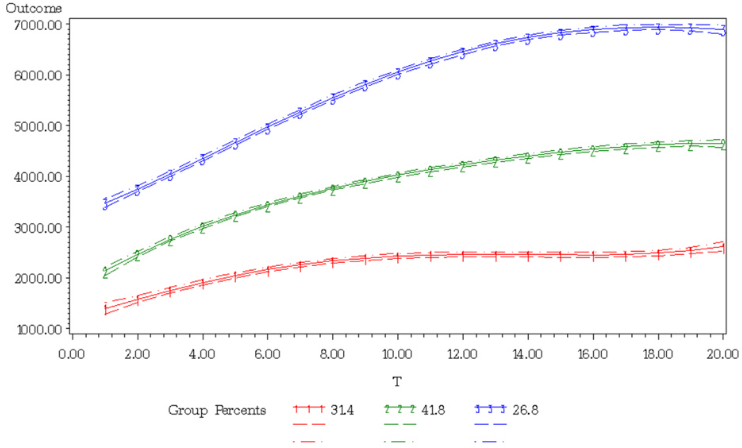
Result for 3 groups : workers beginning their career in 1982



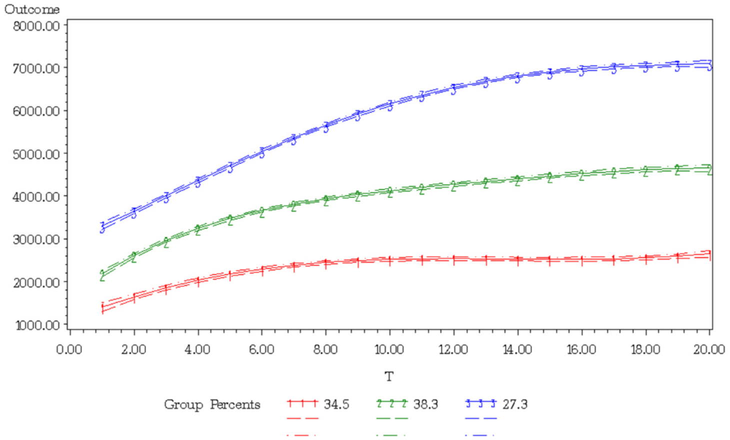
Result for 3 groups : workers beginning their career in 1983



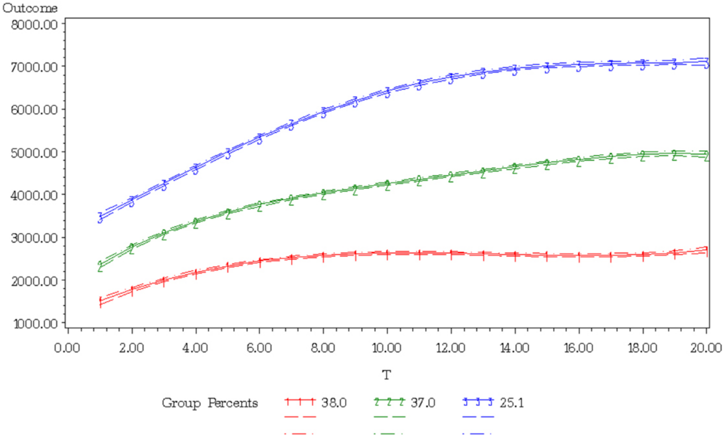
Result for 3 groups : workers beginning their career in 1984



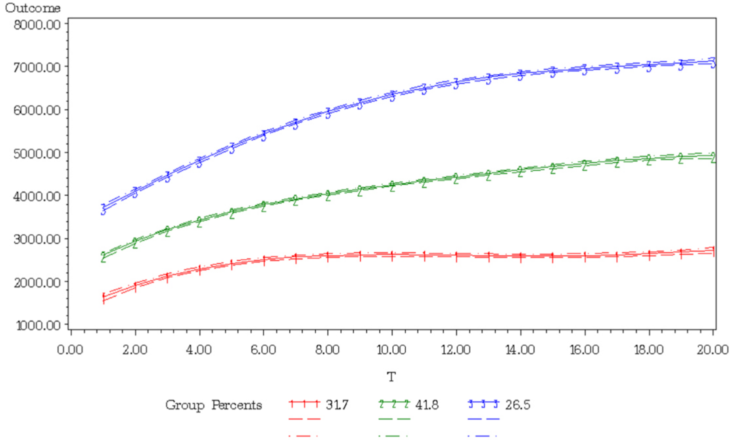
Result for 3 groups : workers beginning their career in 1985



Result for 3 groups : workers beginning their career in 1986



Result for 3 groups : workers beginning their career in 1987



Previous work

- Sampson R.J., Laub J.H. and Eggleston E.P. 2004. On the Robustness and Validity of Groups. *Journal of Quantitative Criminology* **20-1** p.37-42.
- Nagin D.S. and Tremblay R.E. 2005. Developmental trajectory groups: Fact or a useful statistical fiction? *Journal of Criminology* **43-4** p.873-904.
- Sampson R.J. and Laub J.H. 2005. Seductions of method: Rejoinder to Nagin and Tremblay's "Developmental trajectory groups: fact or fiction?. *Journal of Criminology* **43-4** p.905-913.

The statistical shape analysis approach

The statistical shape analysis approach

Comparing the geometrical figure of the trajectories

The statistical shape analysis approach

Comparing the geometrical figure of the trajectories

$\longrightarrow$ statistical shape analyis:

The statistical shape analysis approach

Comparing the geometrical figure of the trajectories

$\longrightarrow$ statistical shape analyis:

Compute the mean shape of the different results.

The statistical shape analysis approach

Comparing the geometrical figure of the trajectories

$\longrightarrow$ statistical shape analyis:

Compute the mean shape of the different results.

Use Ziezold's test for every set of trajectories to see if it is significantly different from the mean set of trajectories.

The statistical shape analysis approach

Comparing the geometrical figure of the trajectories

$\longrightarrow$ statistical shape analyis:

Compute the mean shape of the different results.

Use Ziezold's test for every set of trajectories to see if it is significantly different from the mean set of trajectories.

Remark:

This apporach is just useful to compare a whole set of models.

The mean shape

To compare the standardized and centered sets of landmarks, we need to define the mean shape of all the objects and a distance function which allows us to evaluate how "near" every object is from this mean shape.

The mean shape

To compare the standardized and centered sets of landmarks, we need to define the mean shape of all the objects and a distance function which allows us to evaluate how "near" every object is from this mean shape.

The term "mean" is here used in the sense of Fréchet (1948).

The mean shape

To compare the standardized and centered sets of landmarks, we need to define the mean shape of all the objects and a distance function which allows us to evaluate how "near" every object is from this mean shape.

The term "mean" is here used in the sense of Fréchet (1948).

If X demotes a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with values in a metric space (Ξ, d), an element $m \in \Xi$ is called a mean of $x_1, x_2, ..., x_k \in \Xi$ if

$$\sum_{j=1}^{k} d(x_j, m)^2 = \inf_{\alpha \in \Xi} \sum_{j=1}^{k} d(x_j, \alpha)^2. \tag{7}$$

The mean shape

To compare the standardized and centered sets of landmarks, we need to define the mean shape of all the objects and a distance function which allows us to evaluate how "near" every object is from this mean shape.

The term "mean" is here used in the sense of Fréchet (1948).

If X demotes a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with values in a metric space (Ξ, d), an element $m \in \Xi$ is called a mean of $x_1, x_2, ..., x_k \in \Xi$ if

$$\sum_{j=1}^{k} d(x_j, m)^2 = \inf_{\alpha \in \Xi} \sum_{j=1}^{k} d(x_j, \alpha)^2. \tag{7}$$

That means that the mean shape is defined as the shape with the smallest variance of all shapes in a group of objects.

Ziezold's test

We consider to subsets A and B of the sample of size n and $N - n$ respectively.

Ziezold's test

We consider to subsets A and B of the sample of size n and $N - n$ respectively.

The subset A is a realization of a distribution P and the subset B is an independent realization of a distribution Q.

Ziezold's test

We consider to subsets A and B of the sample of size n and $N - n$ respectively.

The subset A is a realization of a distribution P and the subset B is an independent realization of a distribution Q.

The test hypotheses are:

$$\begin{array}{ll} \textit{Hypothesis:} & H_0 : P = Q \\ \textit{Alternative:} & H_1 : P \neq Q \end{array}$$

Ziezold's test (2)

1. Computing the mean shape m_0 of subset A.

Ziezold's test (2)

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

Ziezold's test (2)

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

$$u_0 = \sum_{j=1}^{n} card\big(b_k : d(b_k, m_0) < d(a_j, m_0)\big).$$

Ziezold's test (2)

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

$$u_0 = \sum_{j=1}^{n} card\big(b_k : d(b_k, m_0) < d(a_j, m_0)\big).$$

3. Determination of all the possibilities of dividing the set into two subset with the same proportion.

Ziezold's test (2)

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

$$u_0 = \sum_{j=1}^{n} card\big(b_k : d(b_k, m_0) < d(a_j, m_0)\big).$$

3. Determination of all the possibilities of dividing the set into two subset with the same proportion.
4. Comparing the u_0-value to all possible u-values. Computing the rank (small u-value mean a small rank).

LUXEMBOURG SCHOOL OF FINANCE

Ziezold's test (2)

1. Computing the mean shape m_0 of subset A.
2. Computing the u-value

$$u_0 = \sum_{j=1}^{n} card\big(b_k : d(b_k, m_0) < d(a_j, m_0)\big).$$

3. Determination of all the possibilities of dividing the set into two subset with the same proportion.
4. Comparing the u_0-value to all possible u-values. Computing the rank (small u-value mean a small rank).
5. Calculate the p-value for H_0. $p_{r=i} = \frac{1}{\binom{N}{n}}$ for $i = 1, \ldots, \binom{N}{n}$, where r is the rank for which we assume a uniform distribution.

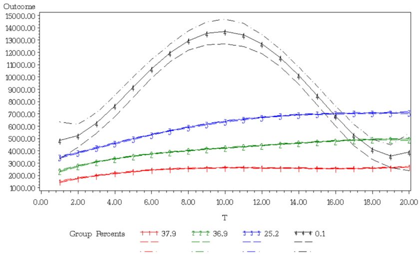
LUXEMBOURG SCHOOL OF FINANCE

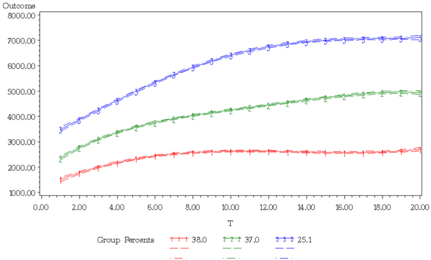
The statistical shape analysis approach

Are these sets of trajectories different?

The statistical shape analysis approach

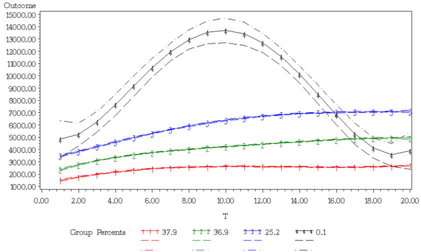
Are these sets of trajectories different?

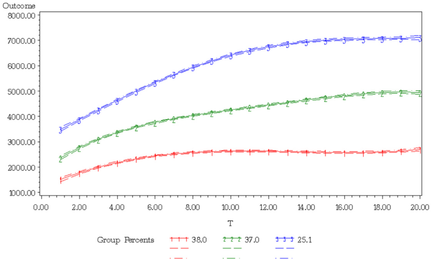




The statistical shape analysis approach

Are these sets of trajectories different?

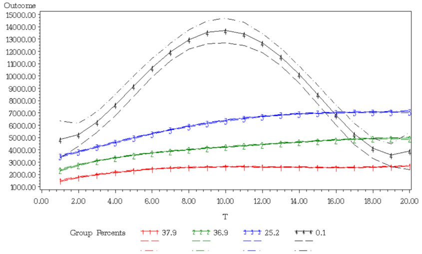


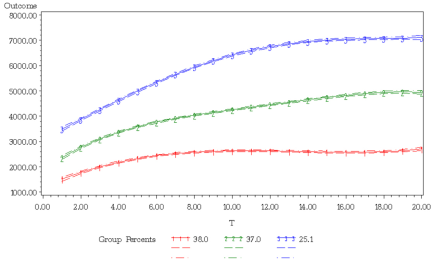


Shape Analysis says yes,

The statistical shape analysis approach

Are these sets of trajectories different?





Shape Analysis says yes, but are they really?

The statistical shape analysis approach

Alternative methodology

To avoid this kind of situation, one can take the estimated parameters of the model as landmarks and perform a statistical "shape" analysis on these.

The classical statistics approach

The classical statistics approach

Compare the estimated parameters:

The classical statistics approach

Compare the estimated parameters:

- Performing the Wald test to see if the parameters differ between two models.

The classical statistics approach

Compare the estimated parameters:

- Performing the Wald test to see if the parameters differ between two models.
- Compare the confidence intervals of the parameters and see if they have an intersection.

Functional Data Analysis Approach

Functional Data Analysis Approach

Compare the set of trajectories as functions:

Functional Data Analysis Approach

Compare the set of trajectories as functions:

Consider a metrical space on the continuous functions defined on the time interval of the trajectories and use tests on functional data to analyze the time stability of the results.

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: Analysis of the salary trajectories in Luxembourg : a finite mixture model approach. To appear.
- Giebel S. 2011: *Zur Anwendung der statistischen Formanalyse*. Phd Thesis, University of Luxembourg.
- Schiltz, J. 2012: Robustness of groups and trajectories in Nagin's finite mixture model. To appear.