A Finite Mixture Model with Trajectories Depending on Covariates

Jang SCHILTZ (University of Luxembourg)

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

June 11, 2014
1 The Basic Finite Mixture Model of Nagin
1. The Basic Finite Mixture Model of Nagin

2. Generalizations of the basic model
Outline

1 The Basic Finite Mixture Model of Nagin

2 Generalizations of the basic model

3 Our model
Outline

1. The Basic Finite Mixture Model of Nagin
2. Generalizations of the basic model
3. Our model
General description of Nagin’s model

We have a collection of individual trajectories.
General description of Nagin’s model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.
General description of Nagin’s model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.

Hence, this model can be interpreted as functional fuzzy cluster analysis.
General description of Nagin’s model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.

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

Finite mixture model (Daniel S. Nagin (Carnegie Mellon University))
General description of Nagin’s model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.

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

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

- mixture : population composed of a mixture of unobserved groups
General description of Nagin’s model

We have a collection of individual trajectories.

We try to divide the population into a number of homogenous sub-populations and to estimate, at the same time, a typical trajectory for each sub-population.

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

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

- mixture: population composed of a mixture of unobserved groups
- finite: sums across a finite number of groups
The Likelihood Function (1)

Consider a population of size $N$ and a variable of interest $Y$.
The Likelihood Function (1)

Consider a population of size $N$ and a variable of interest $Y$.

Let $Y_i = y_{i1}, y_{i2}, ..., y_{iT}$ be $T$ measures of the variable, taken at times $t_1, ... t_T$ for subject number $i$. 
Consider a population of size $N$ and a variable of interest $Y$.

Let $Y_i = y_{i1}, y_{i2}, \ldots, y_{iT}$ be $T$ measures of the variable, taken at times $t_1, \ldots t_T$ for subject number $i$.

$\pi_j$ : probability of a given subject to belong to group number $j$
The Likelihood Function (1)

Consider a population of size $N$ and a variable of interest $Y$.

Let $Y_i = y_{i1}, y_{i2}, ..., y_{iT}$ be $T$ measures of the variable, taken at times $t_1, ..., t_T$ for subject number $i$.

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

$\Rightarrow \pi_j$ is the size of group $j$. 
The Likelihood Function (1)

Consider a population of size $N$ and a variable of interest $Y$.

Let $Y_i = y_{i1}, y_{i2}, \ldots, y_{iT}$ be $T$ measures of the variable, taken at times $t_1, \ldots, t_T$ for subject number $i$.

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

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

$\Rightarrow P(Y_i) = \sum_{j=1}^{r} \pi_j P^j(Y_i), \quad (1)$
The Likelihood Function (1)

Consider a population of size \( N \) and a variable of interest \( Y \).

Let \( Y_i = y_{i1}, y_{i2}, \ldots, y_{iT} \) be \( T \) measures of the variable, taken at times \( t_1, \ldots, t_T \) for subject number \( i \).

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

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

\[ \Rightarrow P(Y_i) = \sum_{j=1}^{r} \pi_j P^j(Y_i), \quad (1) \]

where \( P^j(Y_i) \) is probability of \( Y_i \) if subject \( i \) belongs to group \( j \).
The Likelihood Function (2)

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

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

We try to estimate a set of parameters $\Omega = \{\beta_0^j, \beta_1^j, \beta_2^j, \beta_3^j, \beta_4^j, \pi_j\}$ which allow to maximize the probability of the measured data.
Aim of the analysis: Find $r$ groups of trajectories of a given kind (for instance polynomials of degree 4, $P(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4$).

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

Possible data distributions:
The Likelihood Function (2)

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

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

Possible data distributions:

- count data \( \Rightarrow \) Poisson distribution
The Likelihood Function (2)

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

We try to estimate a set of parameters $\Omega = \{\beta_{0j}, \beta_{1j}, \beta_{2j}, \beta_{3j}, \beta_{4j}, \pi_j\}$ which allow to maximize the probability of the measured data.

Possible data distributions:

- count data $\Rightarrow$ Poisson distribution
- binary data $\Rightarrow$ Binary logit distribution
The Likelihood Function (2)

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

We try to estimate a set of parameters \( \Omega = \{ \beta_{0j}, \beta_{1j}, \beta_{2j}, \beta_{3j}, \beta_{4j}, \pi_j \} \) which allow to maximize the probability of the measured data.

Possible data distributions:

- count data \( \Rightarrow \) Poisson distribution
- binary data \( \Rightarrow \) Binary logit distribution
- censored data \( \Rightarrow \) Censored normal distribution
The case of a normal distribution (1)

Notations :

\[ \beta_j t_i = \beta_{j0} + \beta_{j1} \text{Age}_{i} + \beta_{j2} \text{Age}_{2i} + \beta_{j3} \text{Age}_{3i} + \beta_{j4} \text{Age}_{4i}. \]

\( \phi \): density of standard centered normal law.

Then, 

\[ L = \sigma_N \prod_{i=1}^r \sum_{j=1}^{\pi_j} T_t \prod_{t=1}^{\phi(y_{it} - \beta_{jt})}. \]

(2)

It is too complicated to get closed-forms equations.
The case of a normal distribution (1)

Notations:

- $\beta^j_{it} = \beta^j_0 + \beta^j_1 \text{Age}_{it} + \beta^j_2 \text{Age}_{it}^2 + \beta^j_3 \text{Age}_{it}^3 + \beta^j_4 \text{Age}_{it}^4$.
The case of a normal distribution (1)

Notations:
- \( \beta^j t_{it} = \beta^j_0 + \beta^j_1 \text{Age}_{it} + \beta^j_2 \text{Age}_{it}^2 + \beta^j_3 \text{Age}_{it}^3 + \beta^j_4 \text{Age}_{it}^4 \).
- \( \phi \): density of standard centered normal law.

Then,
\[ L = 1 / \sigma N \prod_{i=1}^r \sum_{j=1}^{\pi_j} T \prod_{t=1} \phi \left( y_{it} - \beta^j t_{it} / \sigma \right). \]
The case of a normal distribution (1)

Notations :
- $\beta^j_{t_{it}} = \beta^j_0 + \beta^j_1 \text{Age}_{it} + \beta^j_2 \text{Age}_{it}^2 + \beta^j_3 \text{Age}_{it}^3 + \beta^j_4 \text{Age}_{it}^4$. 
- $\phi$: density of standard centered normal law.

Then,
The case of a normal distribution (1)

Notations:

- \( \beta^j_{it} = \beta^j_0 + \beta^j_1 \text{Age}_{it} + \beta^j_2 \text{Age}_{it}^2 + \beta^j_3 \text{Age}_{it}^3 + \beta^j_4 \text{Age}_{it}^4 \).
- \( \phi \): density of standard centered normal law.

Then,

\[
L = \frac{1}{\sigma} \prod_{i=1}^N \sum_{j=1}^r \pi_j \prod_{t=1}^T \phi \left( \frac{y_{it} - \beta^j_{it}}{\sigma} \right). \tag{2}
\]
The case of a normal distribution (1)

Notations :

- \( \beta^j t_{it} = \beta^j_0 + \beta^j_1 \text{Age}_{it} + \beta^j_2 \text{Age}_{it}^2 + \beta^j_3 \text{Age}_{it}^3 + \beta^j_4 \text{Age}_{it}^4 \).
- \( \phi \): density of standard centered normal law.

Then,

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

It is too complicated to get closed-forms equations.
Available Software

SAS-based Proc Traj procedure
By Bobby L. Jones (Carnegie Mellon University).
Uses a quasi-Newton procedure maximum research routine.
Since the likelihood is neither convex nor a contraction, there are issues with local maxima.

R-package crimCV
By Jason D. Nielsen (Carleton University Ottawa).
Just implements a zero-inflation Poisson model.
Available Software

SAS-based Proc Traj procedure
By Bobby L. Jones (Carnegie Mellon University).
Available Software

**SAS-based Proc Traj procedure**

By Bobby L. Jones (Carnegie Mellon University).

Uses a quasi-Newton procedure maximum research routine.
Available Software

**SAS-based Proc Traj procedure**

By Bobby L. Jones (Carnegie Mellon University).

Uses a quasi-Newton procedure maximum research routine. Since the likelihood is nor convex, nor a contraction, there are issues with local maxima.

R-package crimCV

By Jason D. Nielsen (Carleton University Ottawa).

Just implements a zero-inflation Poission model.

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

**SAS-based Proc Traj procedure**

By Bobby L. Jones (Carnegie Mellon University).

Uses a quasi-Newton procedure maximum research routine.

Since the likelihood is nor convex, nor a contraction, there are issues with local maxima.

**R-package crimCV**

By Jason D. Nielsen (Carleton University Ottawa).
### Available Software

<table>
<thead>
<tr>
<th>SAS-based Proc Traj procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Bobby L. Jones (Carnegie Mellon University).</td>
</tr>
<tr>
<td>Uses a quasi-Newton procedure maximum research routine.</td>
</tr>
<tr>
<td>Since the likelihood is nor convex, nor a contraction, there are issues with local maxima.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-package crimCV</th>
</tr>
</thead>
<tbody>
<tr>
<td>By Jason D. Nielsen (Carleton University Ottawa).</td>
</tr>
<tr>
<td>Just implements a zero-inflation Poisson model.</td>
</tr>
</tbody>
</table>
Bayesian Information Criterion:

$$\text{BIC} = \log(L) - 0.5k \log(N),$$  \hspace{1cm} (3)

where $k$ denotes the number of parameters in the model.

**Rule:**

The bigger the BIC, the better the model!
Model Selection (2)

Leave-one-out Cross-Validation Approach:

\[ CVE = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} |y_{it} - \hat{y}_{it}^{[-i]}| \]. (4)

Rule:

The smaller the CVE, the better the model!
Posterior Group-Membership Probabilities

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

Bayes’s theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}.$$  \hspace{2cm} (5)

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.
An application example

The data: first dataset
Salaries of workers in the private sector in Luxembourg from 1940 to 2006.
About 7 million salary lines corresponding to 718,054 workers.

Some sociological variables:
- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)
- working status (white collar worker, blue collar worker)
- year of birth
- age in the first year of professional activity

Jang SCHILTZ (University of Luxembourg) joint work with Jean-Daniel GUIGOU (University of Luxembourg), & Bruno LOVAT (University of Lorraine)
An application example

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

- 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
An application example

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

About 7 million salary lines corresponding to 718,054 workers.
An application example

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

About 7 million salary lines corresponding to 718,054 workers.

Some sociological variables:
- gender (male, female)
An application example

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

About 7 million salary lines corresponding to 718,054 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)
An application example

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

About 7 million salary lines corresponding to 718,054 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)
- working status (white collar worker, blue collar worker)
An application example

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

About 7 million salary lines corresponding to 718,054 workers.

Some sociological variables:
- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)
- working status (white collar worker, blue collar worker)
- year of birth
An application example

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

About 7 million salary lines corresponding to 718,054 workers.

Some sociological variables:

- gender (male, female)
- nationality and residentship (luxemburgish residents, foreign residents, foreign non residents)
- working status (white collar worker, blue collar worker)
- year of birth
- age in the first year of professional activity
Result for 9 groups (dataset 1)
Result for 9 groups (dataset 1)
### 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. The Basic Finite Mixture Model of Nagin
2. Generalizations of the basic model
3. Our model
Predictors of trajectory group membership

\[ \pi_j(x_i) = e^{x_i \theta_j} \sum_{k=1}^{r} e^{x_i \theta_k}, \quad (6) \]

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

\[ L = \sigma \prod_{t=1}^{T} \phi(y_{it} - \beta_{jt}) \cdot \prod_{i=1}^{N} \prod_{j=1}^{r} e^{x_i \theta_j} \sum_{k=1}^{r} e^{x_i \theta_k} \cdot \prod_{t=1}^{T} \phi(y_{it} - \beta_{jt}) \cdot \prod_{i=1}^{N} \prod_{j=1}^{r} e^{x_i \theta_j} \sum_{k=1}^{r} e^{x_i \theta_k}, \quad (7) \]
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:

$$\pi_j(x_i) = \frac{e^{x_i \theta_j}}{\sum_{k=1}^{r} e^{x_i \theta_k}}, \quad (6)$$

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

\( x_i \): vector of variables potentially associated with group membership (measured before \( t_1 \)).

Multinomial logit model:

\[
\pi_j(x_i) = \frac{e^{x_i \theta_j}}{\sum_{k=1}^{r} e^{x_i \theta_k}},
\]

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

\[
L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \frac{e^{x_i \theta_j}}{e^{x_i \theta_k}} \prod_{t=1}^{T} \phi \left( \frac{y_{it} - \beta_j t_{it}}{\sigma} \right).
\]
Adding covariates to the trajectories (1)
Adding covariates to the trajectories (1)

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

$$y_{it} = \beta_{j0} + \beta_{j1} \text{Age}_{it} + \beta_{j2} \text{Age}_{2it} + \beta_{j3} \text{Age}_{3it} + \beta_{j4} \text{Age}_{4it} + \alpha_{j1} z_{1i} + \ldots + \alpha_{jL} z_{Li} + \epsilon_{it},$$

where $\epsilon_{it} \sim N(0, \sigma)$, $\sigma$ being a constant standard deviation and $z_l$ are covariates that may depend or not upon time $t$. 

Unfortunately the influence of the covariates in this model is limited to the intercept of the trajectory.
Adding covariates to the trajectories (1)

Let \( z_1 \ldots z_L \) be covariates potentially influencing \( Y \).

We are then looking for trajectories

\[
y_{it} = \beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2 + \beta_3^j \text{Age}_{it}^3 + \beta_4^j \text{Age}_{it}^4 + \alpha_1^j z_1 + \ldots + \alpha_L^j z_L + \varepsilon_{it}, \tag{8}
\]

where \( \varepsilon_{it} \sim \mathcal{N}(0, \sigma) \), \( \sigma \) being a constant standard deviation and \( z_l \) are covariates that may depend or not upon time \( t \).
Adding covariates to the trajectories (1)

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

We are then looking for trajectories

$$y_{it} = \beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2 + \beta_3^j \text{Age}_{it}^3 + \beta_4^j \text{Age}_{it}^4 + \alpha_1^j z_1 + \ldots + \alpha_L^j z_L + \epsilon_{it}, \quad (8)$$

where $\epsilon_{it} \sim \mathcal{N}(0, \sigma)$, $\sigma$ being a constant standard deviation and $z_l$ are covariates that may depend or not upon time $t$.

Unfortunately the influence of the covariates in this model is limited to the intercept of the trajectory.
Adding covariates to the trajectories (2)
Adding covariates to the trajectories (2)
Outline

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

Let $x_1, \ldots, x_L$ and $z_{i1}, \ldots, z_{iT}$ be covariates potentially influencing $Y$. We propose the following model:

$$y_{it} = \left( \beta_{j0} + \sum_{l=1}^{L} \alpha_{j0l} x_l + \gamma_{j0} z_{it} \right) + \left( \beta_{j1} + \sum_{l=1}^{L} \alpha_{j1l} x_l + \gamma_{j1} z_{it} \right) \text{Age}_{iit} + \left( \beta_{j2} + \sum_{l=1}^{L} \alpha_{j2l} x_l + \gamma_{j2} z_{it} \right) \text{Age}_{2iit} + \left( \beta_{j3} + \sum_{l=1}^{L} \alpha_{j3l} x_l + \gamma_{j3} z_{it} \right) \text{Age}_{3iit} + \left( \beta_{j4} + \sum_{l=1}^{L} \alpha_{j4l} x_l + \gamma_{j4} z_{it} \right) \text{Age}_{4iit} + \epsilon_{it},$$

where $\epsilon_{it} \sim N(0, \sigma^2)$, $\sigma$ being a constant standard deviation.
Our model

Let $x_1 ... x_L$ and $z_{i1}, ..., z_{iT}$ be covariates potentially influencing $Y$. 
Our model

Let $x_1 \ldots x_L$ and $z_{i1}, \ldots, z_{iT}$ be covariates potentially influencing $Y$.

We propose the following model:

$$
y_{it} = \left( \beta^j_0 + \sum_{l=1}^{L} \alpha^j_{0l} x_l + \gamma^j_0 z_{it} \right) + \left( \beta^j_1 + \sum_{l=1}^{L} \alpha^j_{1l} x_l + \gamma^j_1 z_{it} \right) \text{Age}_{it}
$$

$$ + \left( \beta^j_2 + \sum_{l=1}^{L} \alpha^j_{2l} x_l + \gamma^j_2 z_{it} \right) \text{Age}^2_{it} + \left( \beta^j_3 + \sum_{l=1}^{L} \alpha^j_{3l} x_l + \gamma^j_3 z_{it} \right) \text{Age}^3_{it}
$$

$$ + \left( \beta^j_4 + \sum_{l=1}^{L} \alpha^j_{4l} x_l + \gamma^j_4 z_{it} \right) \text{Age}^4_{it} + \varepsilon_{it},$$

where $\varepsilon_{it} \sim \mathcal{N}(0, \sigma)$, $\sigma$ being a constant standard deviation.
Men versus women

![Chart showing salary trends over time for men and women.](image-url)
Statistical Properties

Both Nagin’s and our model can be written as

\[ L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_{j} \prod_{t=1}^{T} \phi \left( \frac{\text{observed data} - \text{modelled data}}{\sigma} \right) \]. (9)
Both Nagin’s and our model can be written as

\[
L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi \left( \frac{\text{observed data} - \text{modelled data}}{\sigma} \right). \tag{9}
\]

Nagin’s model:

\[
y_{it} = \beta_0^j + \beta_1^j \text{Age}_{it} + \beta_2^j \text{Age}_{it}^2 + \beta_3^j \text{Age}_{it}^3 + \beta_4^j \text{Age}_{it}^4 + \alpha_1^j z_1 + \ldots + \alpha_L^j z_L + \varepsilon_{it}, \tag{10}
\]

Our model:

\[
y_{it} = \left( \beta_0^j + \sum_{l=1}^{L} \alpha_{0l}^j x_l + \gamma_0^j z_{it} \right) + \left( \beta_1^j + \sum_{l=1}^{L} \alpha_{1l}^j x_l + \gamma_1^j z_{it} \right) \text{Age}_{it} + \\
\left( \beta_2^j + \sum_{l=1}^{L} \alpha_{2l}^j x_l + \gamma_2^j z_{it} \right) \text{Age}_{it}^2 + \left( \beta_3^j + \sum_{l=1}^{L} \alpha_{3l}^j x_l + \gamma_3^j z_{it} \right) \text{Age}_{it}^3 + \\
\left( \beta_4^j + \sum_{l=1}^{L} \alpha_{4l}^j x_l + \gamma_4^j z_{it} \right) \text{Age}_{it}^4 + \varepsilon_{it},
\]
Parameter estimation

A way of estimating our model with the existing software (if group membership does not depend on the covariates):

- Use the latest version of proc.traj to test if the covariates have indeed an influence on the trajectories.
- Apply proc.traj to the data without covariates to do the clustering and obtain the number of groups and the constitution of the groups.
- Use your favorite regression model software to get the trajectories separately for each group.
Parameter estimation

A way of estimating our model with the existing software (if group membership does not depend on the covariates):

- Use the latest version of proc.traj to test if the covariates have indeed an influence on the trajectories.
Parameter estimation

A way of estimating our model with the existing software (if group membership does not depend on the covariates):

- Use the latest version of proc.traj to test if the covariates have indeed an influence on the trajectories.
- Apply proc.traj to the data without covariates do the clustering and obtain the number of groups and the constitution of the groups.
A way of estimating our model with the existing software (if group membership does not depend on the covariates):

- Use the latest version of proc.traj to test if the covariates have indeed an influence on the trajectories.
- Apply proc.traj to the data without covariates do the clustering and obtain the number of groups and the constitution of the groups.
- Use your favorite regression model software to get the trajectories separately for each group.
Bibliography