Model selection in generalized finite mixture models

Jang SCHILTZ (University of Luxembourg)

July 11, 2016
Outline

1 Nagin’s Finite Mixture Model
Outline

1. Nagin’s Finite Mixture Model
2. Generalizations of Nagin’s model
Outline

1. Nagin’s Finite Mixture Model
2. Generalizations of Nagin’s model
3. Our model
Outline

1. Nagin’s Finite Mixture Model
2. Generalizations of Nagin’s model
3. Our model
4. Model Selection
Outline

1. Nagin’s Finite Mixture Model
2. Generalizations of Nagin’s model
3. Our model
4. Model Selection
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
Consider a population of size $N$ and a variable of interest $Y$. 

The Likelihood Function (1) 

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 $j$. 

$P_j(Y_i) = \sum_{j=1}^{\pi} \pi_j P_j(Y_i)$, (1) 

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

Jang SCHILTZ (University of Luxembourg) Model selection in generalized finite mixture 

July 11, 2016 5 / 29
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 $Y_i = y_{i1}, y_{i2}, \ldots, y_{iT}$ 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 $j$. 
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 $j$.

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

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

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

Statistical Model:

\[
y_{it} = \beta_0^j + \beta_1^j t + \beta_2^j t^2 + \beta_3^j t^3 + \beta_4^j t^4 + \varepsilon_{it},
\]

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

We try to estimate a set of parameters \( \Omega = \{ \beta_0^j, \beta_1^j, \beta_2^j, \beta_3^j, \beta_4^j, \pi_j, \sigma \} \) 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
Possible data distributions

- count data $\Rightarrow$ Poisson distribution
Possible data distributions

- count data $\Rightarrow$ Poisson distribution
- binary data $\Rightarrow$ Binary logit distribution
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:

\[ \beta_j t = \beta_{j0} + \beta_{j1} t + \beta_{j2} t^2 + \beta_{j3} t^3 + \beta_{j4} t^4. \]

\[ \phi: \text{density of standard centered normal law.} \]

Then,

\[ L = \frac{1}{\sigma N} \prod_{i=1}^{r} \sum_{j=1}^{\pi_j} T \prod_{t=1}^{\phi(y_i t - \beta_j t \sigma)}. \]

(3)

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

Notations:

- $\beta^j t = \beta^j_0 + \beta^j_1 t + \beta^j_2 t^2 + \beta^j_3 t^3 + \beta^j_4 t^4$. 

$\phi$: density of standard centered normal law.

Then,

$$L = \frac{1}{\sigma_N} \prod_{i=1}^{r} \sum_{j=1}^{\pi_j} T \prod_{t=1}^{\phi(y_i t - \beta^j t \sigma)}.$$ 

(3)

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

Notations:

- \( \beta^j t = \beta^j_0 + \beta^j_1 t + \beta^j_2 t^2 + \beta^j_3 t^3 + \beta^j_4 t^4 \).
- \( \phi \): density of standard centered normal law.
The case of a normal distribution (1)

Notations :
- \( \beta^i t = \beta^i_0 + \beta^i_1 t + \beta^i_2 t^2 + \beta^i_3 t^3 + \beta^i_4 t^4 \).
- \( \phi \): density of standard centered normal law.

Then,
The case of a normal distribution (1)

Notations:

- $\beta^j t = \beta^j_0 + \beta^j_1 t + \beta^j_2 t^2 + \beta^j_3 t^3 + \beta^j_4 t^4$.
- $\phi$: density of standard centered normal law.

Then,

$$L = \frac{1}{\sigma} \prod_i^{N} \sum_j^r \pi_j \prod_{t=1}^T \phi \left( \frac{y_{it} - \beta^j t}{\sigma} \right).$$ (3)
The case of a normal distribution (1)

Notations:
- $\beta^j t = \beta^j_0 + \beta^j_1 t + \beta^j_2 t^2 + \beta^j_3 t^3 + \beta^j_4 t^4$.
- $\phi$: density of standard centered normal law.

Then,

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

(3)

It is too complicated to get closed-forms equations.
An application example

The data:
Salaries of workers in the private sector in Luxembourg from 1987 to 2006.
About 1.3 million salary lines corresponding to 85,049 workers.
Some sociological variables:
gender (male, female)
nationality and residentship
working sector
year of birth
year of birth of children
age in the first year of professional activity
An application example

The data: Salaries of workers in the private sector in Luxembourg from 1987 to 2006.
An application example

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

About 1.3 million salary lines corresponding to 85,049 workers.
An application example

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

About 1.3 million salary lines corresponding to 85,049 workers.

Some sociological variables:

- gender (male, female)
An application example

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

About 1.3 million salary lines corresponding to 85,049 workers.

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

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

About 1.3 million salary lines corresponding to 85,049 workers.

Some sociological variables:
- gender (male, female)
- nationality and residentship
- working sector
An application example

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

About 1.3 million salary lines corresponding to 85,049 workers.

Some sociological variables:
- gender (male, female)
- nationality and residentship
- working sector
- year of birth
An application example

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

About 1.3 million salary lines corresponding to 85,049 workers.

Some sociological variables:
- gender (male, female)
- nationality and residentship
- working sector
- year of birth
- year of birth of children
An application example

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

About 1.3 million salary lines corresponding to 85,049 workers.

Some sociological variables:
- gender (male, female)
- nationality and residentship
- working sector
- year of birth
- year of birth of children
- age in the first year of professional activity
Result for 9 groups (dataset 1)
Result for 9 groups (dataset 1)
Outline

1. Nagin’s Finite Mixture Model
2. Generalizations of Nagin’s model
3. Our model
4. Model Selection
Predictors of trajectory group membership

\[
\pi_j(x_i) = e^{x_i \theta_j \sum_k e^{x_i \theta_k}},
\]

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

\[
L = \sigma N \prod_i \sum_j e^{x_i \theta_j \sum_k e^{x_i \theta_k}} T \prod_t \phi(y_{it} - \beta_j t \sigma).
\]
Predictors of trajectory group membership

$x$: vector of variables potentially associated with group membership (measured before $t_1$).
Predictors of trajectory group membership

$x$ : 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.
Predictors of trajectory group membership

$x$: vector of variables potentially associated with group membership (measured before $t_1$).

Multinomial logit model:

$$
\pi_j(x_i) = \frac{e^{x_i \theta_j}}{r \sum_{k=1}^{r} e^{x_i \theta_k}}, \quad (4)
$$

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

$$
L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \frac{e^{x_i \theta_j}}{r \sum_{k=1}^{r} e^{x_i \theta_k}} \prod_{t=1}^{T} \phi \left( \frac{y_{it} - \beta^j t}{\sigma} \right). \quad (5)
$$
Adding covariates to the trajectories (1)

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

We are then looking for trajectories $y_{it} = \beta_{j0} + \beta_{j1}t + \beta_{j2}t^2 + \beta_{j3}t^3 + \beta_{j4}t^4 + \alpha_{j1}z_1 + \ldots + \alpha_{jM}z_M + \epsilon_{it}$, \hspace{1cm} (6)

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

$$y_{it} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^4 + \alpha_1 z_1 + \ldots + \alpha_M z_M + \epsilon_{it}, \quad (6)$$

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

We are then looking for trajectories

$$y_{it} = \beta^j_0 + \beta^j_1 t + \beta^j_2 t^2 + \beta^j_3 t^3 + \beta^j_4 t^4 + \alpha^j_1 z_1 + ... + \alpha^j_M z_M + \varepsilon_{it},$$

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

We are then looking for trajectories

$$y_{it} = \beta_0^j + \beta_1^j t + \beta_2^j t^2 + \beta_3^j t^3 + \beta_4^j t^4 + \alpha_1^j z_1 + \ldots + \alpha_M^j z_M + \varepsilon_{it}, \quad (6)$$

where $\varepsilon_{it} \sim N(0, \sigma)$, $\sigma$ 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 (2)
Adding covariates to the trajectories (2)
Outline

1. Nagin’s Finite Mixture Model
2. Generalizations of Nagin’s model
3. Our model
4. Model Selection
Our model

Let \( x_1, \ldots, x_M \) and \( z_t \) be covariates potentially influencing \( Y \).

We propose the following model:

\[
y_{it} = \left( \beta_{j0} + \sum_{l=1}^{M} \alpha_{j0} l x_{il} + \gamma_{j0} z_{it} \right) + \left( \beta_{j1} + \sum_{l=1}^{M} \alpha_{j1} l x_{il} + \gamma_{j1} z_{it} \right) t + \left( \beta_{j2} + \sum_{l=1}^{M} \alpha_{j2} l x_{il} + \gamma_{j2} z_{it} \right) t^2 + \left( \beta_{j3} + \sum_{l=1}^{M} \alpha_{j3} l x_{il} + \gamma_{j3} z_{it} \right) t^3 + \left( \beta_{j4} + \sum_{l=1}^{M} \alpha_{j4} l x_{il} + \gamma_{j4} z_{it} \right) t^4 + \varepsilon_{ji} t,
\]

where \( \varepsilon_{ji} \sim N(0, \sigma_j) \), \( \sigma_j \) being the standard deviation, constant in group \( j \).
Our model

Let $x_1...x_M$ and $z_t$ be covariates potentially influencing $Y$. 
Our model

Let $x_1...x_M$ and $z_t$ be covariates potentially influencing $Y$.

We propose the following model:

$$ y_{it} = \left( \beta^j_0 + \sum_{l=1}^{M} \alpha^j_{0l} x_{il} + \gamma^j_0 z_{it} \right) + \left( \beta^j_1 + \sum_{l=1}^{M} \alpha^j_{1l} x_{il} + \gamma^j_1 z_{it} \right) t $$

$$ + \left( \beta^j_2 + \sum_{l=1}^{M} \alpha^j_{2l} x_{il} + \gamma^j_2 z_{it} \right) t^2 + \left( \beta^j_3 + \sum_{l=1}^{M} \alpha^j_{3l} x_{il} + \gamma^j_3 z_{it} \right) t^3 $$

$$ + \left( \beta^j_4 + \sum_{l=1}^{M} \alpha^j_{4l} x_{il} + \gamma^j_4 z_{it} \right) t^4 + \varepsilon^j_{it}, $$

where $\varepsilon^j_{it} \sim N(0, \sigma^j)$, $\sigma^j$ being the standard deviation, constant in group $j$. 
Men versus women
Statistical Properties

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

Confidence intervals of level $\alpha$ for the parameters $\beta_{jk}$:

$$CI_{\alpha}(\beta_{jk}) = \left[ \hat{\beta}_{jk} - t_{1-\alpha/2; N-(2+M)s_{\text{ASE}}(\hat{\beta}_{jk})}; \hat{\beta}_{jk} + t_{1-\alpha/2; N-(2+M)s_{\text{ASE}}(\hat{\beta}_{jk})} \right].$$

(7)

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

$$CI_{\alpha}(\sigma_j) = \left[ \left( N-(2+M)s_{-1} \right) \chi^2_{1-\alpha/2; N-(2+M)s_{-1}} \sigma_j^2; \left( N-(2+M)s_{-1} \right) \chi^2_{\alpha/2; N-(2+M)s_{-1}} \sigma_j^2 \right].$$

(8)
Statistical Properties

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

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

Confidence intervals of level $\alpha$ for the parameters $\beta^j_k$:

$$
CI_{\alpha}(\beta^j_k) = \left[ \hat{\beta}^j_k - t_{1-\alpha/2; N-(2+M)s_{\text{ASE}}(\hat{\beta}^j_k)}; N-(2+M)s_{\text{ASE}}(\hat{\beta}^j_k) + t_{1-\alpha/2; N-(2+M)s_{\text{ASE}}(\hat{\beta}^j_k)} \right].
$$

(7)

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

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

(8)
Statistical Properties

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

Confidence intervals of level $\alpha$ for the parameters $\beta^j_k$:

$$CI_\alpha(\beta^j_k) = \left[ \hat{\beta}^j_k - t_{1-\alpha/2;N-(2+M)s}ASE(\hat{\beta}^j_k); \hat{\beta}^j_k + t_{1-\alpha/2;N-(2+M)s}ASE(\hat{\beta}^j_k) \right]. \quad (7)$$
Statistical Properties

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

Confidence intervals of level $\alpha$ for the parameters $\beta^j_k$:

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

(7)

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

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

(8)
The model’s estimated parameters are the result of maximum likelihood estimation. As such, they are consistent and asymptotically normally distributed.

Confidence intervals of level $\alpha$ for the parameters $\beta^j_k$:

$$ CI_\alpha(\beta^j_k) = \left[ \hat{\beta}^j_k - t_{1-\alpha/2;N-(2+M)s} \text{ASE}(\hat{\beta}^j_k), \hat{\beta}^j_k + t_{1-\alpha/2;N-(2+M)s} \text{ASE}(\hat{\beta}^j_k) \right]. $$

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

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

We analyze the influence of the consumer price index (CPI) on the salary. CPI and time have a correlation of 0.995. Hence a model like

$$S_t = (\beta_{j0} + \gamma_{j0} z_t) + (\beta_{j1} + \gamma_{j1} z_t) t + (\beta_{j2} + \gamma_{j2} z_t) t^2 + (\beta_{j3} + \gamma_{j3} z_t) t^3,$$

(9)

where $S$ denotes the salary and $z_t$ is Luxembourg's CPI in year $t$ of the study, makes no sense. Because of obvious multicolinearity problems, almost none of the parameters would be significant. Therefore, we simplify the model and calibrate

$$S_t = (\beta_{j0} + \gamma_{j0} z_t) + \gamma_{j1} z_t t + \gamma_{j2} z_t t^2 + \gamma_{j3} z_t t^3.$$

(10)
Attention to multicolinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary.
Attention to multicollinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary. CPI and time have a correlation of 0.995.
Attention to multicolinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary. CPI and time have a correlation of 0.995. Hence a model like

\[ S_{it} = (\beta_0^i + \gamma_0^i z_t) + (\beta_1^i + \gamma_1^i z_t) t + (\beta_2^i + \gamma_2^i z_t) t^2 + (\beta_3^i + \gamma_3^i z_t) t^3, \quad (9) \]

where \( S \) denotes the salary and \( z_t \) is Luxembourg’s CPI in year \( t \) of the study, makes no sense.
Attention to multicolinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary. CPI and time have a correlation of 0.995. Hence a model like

$$S_{it} = (\beta_0^i + \gamma_0^i z_t) + (\beta_1^i + \gamma_1^i z_t) t + (\beta_2^i + \gamma_2^i z_t) t^2 + (\beta_3^i + \gamma_3^i z_t) t^3,$$  \hspace{1cm} (9)

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

Because of obvious multicolinearity problems, almost none of the parameters would be significant.
Attention to multicolinearity issues!

We analyze the influence of the consumer price index (CPI) on the salary. CPI and time have a correlation of 0.995. Hence a model like

$$S_{it} = (\beta_0^j + \gamma_0^j z_t) + (\beta_1^j + \gamma_1^j z_t) t + (\beta_2^j + \gamma_2^j z_t) t^2 + (\beta_3^j + \gamma_3^j z_t) t^3,$$  \hspace{1cm} (9)

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

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

Therefore, we simplify the model and calibrate

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

(10)
### Results for group 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>321.381</td>
<td>1189.430</td>
<td>-2213.502 2856.093</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>1689.492</td>
<td>277.834</td>
<td>-4.232 7.611</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.400</td>
<td>0.120</td>
<td>0.143 0.656</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.034</td>
<td>0.007</td>
<td>-0.049 -0.019</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.0008</td>
<td>0.0002</td>
<td>0.0005 0.0013</td>
</tr>
</tbody>
</table>

### Results for group 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>7688.158</td>
<td>951.103</td>
<td>5660.197 9714.832</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>-13.095</td>
<td>2.222</td>
<td>-17.822 -8.350</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.260</td>
<td>0.096</td>
<td>1.055 1.465</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.097</td>
<td>0.006</td>
<td>-0.109 -0.085</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.0025</td>
<td>0.0002</td>
<td>0.0022 0.0028</td>
</tr>
</tbody>
</table>

### Results for group 3

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>682.638</td>
<td>196.327</td>
<td>141.924 1101.045</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>-11.367</td>
<td>4.586</td>
<td>-21.135 -1.586</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.983</td>
<td>0.199</td>
<td>0.559 1.406</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.048</td>
<td>0.012</td>
<td>-0.073 -0.023</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.0010</td>
<td>0.0003</td>
<td>0.0003 0.0017</td>
</tr>
</tbody>
</table>
### Results for group 4

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>8473.081</td>
<td>1859.349</td>
<td>4511.016 - 12434.892</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>-13.083</td>
<td>4.342</td>
<td>-22.335 - 3.825</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.927</td>
<td>0.188</td>
<td>0.527 - 1.328</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.013</td>
<td>0.011</td>
<td>-0.036 - 0.010</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>-0.0003</td>
<td>0.0003</td>
<td>-0.0009 - 0.0004</td>
</tr>
</tbody>
</table>

### Results for group 5

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>4798.276</td>
<td>3205.141</td>
<td>-2034.302 - 11630.238</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>-2.846</td>
<td>7.488</td>
<td>-18.806 - 13.115</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.315</td>
<td>0.324</td>
<td>0.0624 - 2.006</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.081</td>
<td>0.019</td>
<td>-0.122 - 0.040</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.0016</td>
<td>0.0005</td>
<td>0.0005 - 0.0027</td>
</tr>
</tbody>
</table>

### Results for group 6

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_0$</td>
<td>8332.439</td>
<td>1139.127</td>
<td>5903.348 - 10759.713</td>
</tr>
<tr>
<td>$\gamma_0$</td>
<td>-12.472</td>
<td>2.661</td>
<td>-18.145 - 6.800</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.378</td>
<td>0.015</td>
<td>1.132 - 1.623</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.094</td>
<td>0.007</td>
<td>-0.108 - 0.079</td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.0022</td>
<td>0.0002</td>
<td>0.0018 - 0.0026</td>
</tr>
</tbody>
</table>
Disturbance terms

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

1. Nagin’s Finite Mixture Model
2. Generalizations of Nagin’s model
3. Our model
4. Model Selection
Bayesian Information Criterion:

\[ \text{BIC} = \log(L) - \frac{0.5k \log(N)}{N}, \]  

where \( k \) denotes the number of parameters in the model.

Rule: The bigger the BIC, the better the model!
Bayesian Information Criterion:

\[ \text{BIC} = \log(L) - 0.5k \log(N), \quad (11) \]

where \( k \) denotes the number of parameters in the model.
Bayesian Information Criterion:

\[ \text{BIC} = \log(L) - 0.5k \log(N), \quad (11) \]

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:

\[
\text{CVE} = N \sum_{i=1}^{T} \left| y_{it} - \hat{y}_{it} \right|.
\] (12)

Rule: The smaller the CVE, the better the model!
Model Selection (2)

Leave-one-out Cross-Validation Approach:

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

(12)
Leave-one-out Cross-Validation Approach:

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

(12)

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

(13)

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.
Posterior Group-Membership Probabilities

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

Bayes’s theorem $\Rightarrow P(j/Y_i) = P(Y_i/j)^{\pi_j}r_j \sum_{j=1}^{J} P(Y_i/j)^{\pi_j}$. 

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.
Posterior Group-Membership Probabilities

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

Bayes’s theorem

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

(13)

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.
Posterior Group-Membership Probabilities

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

Bayes’s theorem

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{13}$$

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

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \quad (13)$$

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.
Our Model Selection Criterion

We propose to take the number of groups which maximizes the classification probabilities.

\[ SP = N \sum_{i=1}^{N} \log(\max_{j} P(j/Y_i)) \] (14)

Rule: The bigger the SP, the better the model!
Our Model Selection Criterion

We propose to take the number of groups which maximizes the classification probabilities.

\[
SP = \sum_{i=1}^{N} \log(\max_j P(j/Y_i))
\]  

(14)
Our Model Selection Criterion

We propose to take the number of groups which maximizes the classification probabilities.

\[
SP = \sum_{i=1}^{N} \log(\max_j P(j/Y_i))
\]  

(14)

Rule:
The bigger the SP, the better the model!
Advantages

- Computationally easy
- Does not depend on the number of parameters in the model. Hence there is no need for a correction term.
Advantages

- Computationally easy
Advantages

- Computationally easy
- Does not depend on the number of parameters in the model. Hence there is no need for a correction term.
Bibliography


