A new model selection criterion for finite mixture models

Jang SCHILTZ (University of Luxembourg)

joint work with Cédric NOEL(University of Lorraine)

62nd ISI World Statistics Congress 2019
August 20, 2019

LUXEMBOURG SCHOOL OF FINANCE

Outline

Outline

Outline

Outline

Outline

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.

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.

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

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

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

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

LUXEMBOURG SCHOOL OF FINANCE

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:

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

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

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

LUXEMBOURG SCHOOL OF FINANCE

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

Notations :

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

Then,

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \pi_j \prod_{t=1}^{T} \phi\left(\frac{y_{i_t} - \beta^j t}{\sigma}\right). \tag{3}$$

It is too complicated to get closed-forms solutions.

Outline

LUXEMBOURG SCHOOL OF FINANCE

Predictors of trajectory group membership

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

Multinomial logit model:

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

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

$$L = \frac{1}{\sigma} \prod_{i=1}^{N} \sum_{j=1}^{r} \frac{e^{x_i\theta_j}}{\sum_{k=1}^{r} e^{x_i\theta_k}} \prod_{t=1}^{T} \phi\left(\frac{y_{i_t} - \beta^j t}{\sigma}\right). \tag{5}$$

LUXEMBOURG SCHOOL OF FINANCE

Adding covariates to the trajectories

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

We are then looking for trajectories

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

where $\varepsilon_{i_t} \sim \mathcal{N}(0, \sigma)$, σ 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.

An application example

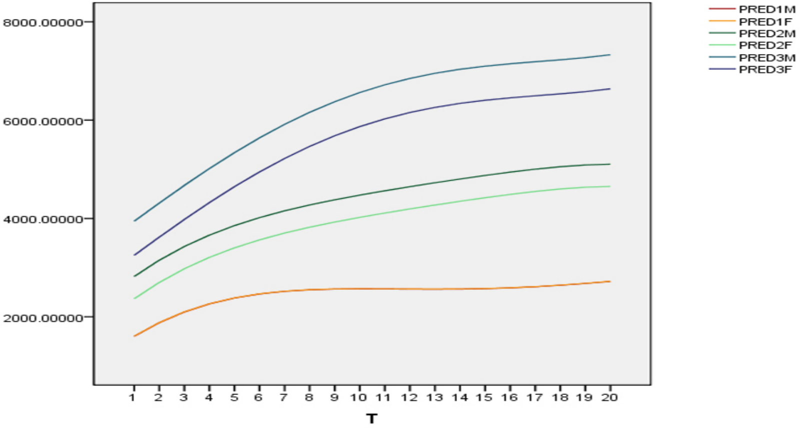
The data : Salaries of workers in the private sector in Luxembourg from 1987 to 2006.

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 (2)



Outline

Our model

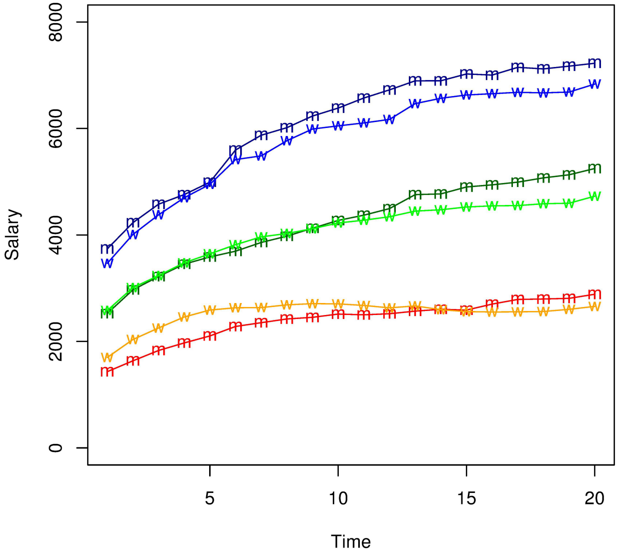
Let $x_1 ... x_M$ and z_t be covariates potentially influencing Y.

We propose the following model:

$$y_{i_t} = \left(\beta_0^j + \sum_{l=1}^{M} \alpha_{0l}^j x_{i_l} + \gamma_0^j z_{i_t}\right) + \left(\beta_1^j + \sum_{l=1}^{M} \alpha_{1l}^j x_{i_l} + \gamma_1^j z_{i_t}\right) t$$
$$+ \left(\beta_2^j + \sum_{l=1}^{M} \alpha_{2l}^j x_{i_l} + \gamma_2^j z_{i_t}\right) t^2 + \left(\beta_3^j + \sum_{l=1}^{M} \alpha_{3l}^j x_{i_l} + \gamma_3^j z_{i_t}\right) t^3$$
$$+ \left(\beta_4^j + \sum_{l=1}^{M} \alpha_{4l}^j x_{i_l} + \gamma_4^j z_{i_t}\right) t^4 + \varepsilon_{i_t}^j,$$

where $\varepsilon_{i_t} \sim \mathcal{N}(0, \sigma^j)$, σ^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 α for the parameters β_k^j:

$$CI_\alpha(\beta_k^j) = \left[\hat{\beta}_k^j - t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j); \hat{\beta}_k^j + t_{1-\alpha/2;N-(2+M)s} ASE(\hat{\beta}_k^j)\right]. \tag{7}$$

Confidence intervals of level α for the disturbance factor σ_j:

$$CI_\alpha(\sigma_j) = \left[\sqrt{\frac{(N-(2+M)s-1)\hat{\sigma}_j{}^2}{\chi^2_{1-\alpha/2;N-(2+M)s-1}}}; \sqrt{\frac{(N-(2+M)s-1)\hat{\sigma}_j{}^2}{\chi^2_{\alpha/2;N-(2+M)s-1}}}\right]. \tag{8}$$

Results for group 1

Parameter	Estimate	Standard error	95% confidence Lower	intervals Upper
β_0	321.381	1189.430	-2213.502	2856.093
γ_0	1689.492	277.834	-4.232	7.611
γ_1	**0.400**	0.120	0.143	0.656
γ_2	**-0.034**	0.007	-0.049	-0.019
γ_3	**0.0008**	0.0002	0.0005	0.0013

Results for group 2

Parameter	Estimate	Standard error	95% confidence Lower	intervals Upper
β_0	**7688.158**	951.103	5660.197	9714.832
γ_0	**-13.095**	2.222	-17.822	-8.350
γ_1	**1.260**	0.096	1.055	1.465
γ_2	**-0.097**	0.006	-0.109	-0.085
γ_3	**0.0025**	0.0002	0.0022	0.0028

Results for group 3

Parameter	Estimate	Standard error	95% confidence Lower	intervals Upper
β_0	**682.638**	196.327	141.924	1101.045
γ_0	**-11.367**	4.586	-21.135	-1.586
γ_1	**0.983**	0.199	0.559	1.406
γ_2	**-0.048**	0.012	-0.073	-0.023
γ_3	**0.0010**	0.0003	0.0003	0.0017

Results for group 4

Parameter	Estimate	Standard error	95% confidence Lower	intervals Upper
β_0	**8473.081**	1859.349	4511.016	12434.892
γ_0	**-13.083**	4.342	-22.335	-3.825
γ_1	**0.927**	0.188	0.527	1.328
γ_2	-0.013	0.011	-0.036	0.010
γ_3	-0.0003	0.0003	-0.0009	0.0004

Results for group 5

Parameter	Estimate	Standard error	95% confidence Lower	intervals Upper
β_0	4798.276	3205.141	-2034.302	11630.238
γ_0	-2.846	7.488	-18.806	13.115
γ_1	**1.315**	0.324	0.0624	2.006
γ_2	**-0.081**	0.019	-0.122	-0.040
γ_3	**0.0016**	0.0005	0.0005	0.0027

Results for group 6

Parameter	Estimate	Standard error	95% confidence Lower	intervals Upper
β_0	**8332.439**	1139.127	5903.348	10759.713
γ_0	**-12.472**	2.661	-18.145	-6.800
γ_1	**1.378**	0.015	1.132	1.623
γ_2	**-0.094**	0.007	-0.108	-0.079
γ_3	**0.0022**	0.0002	0.0018	0.0026

LUXEMBOURG SCHOOL OF FINANCE

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

Model Selection (1)

Bayesian Information Criterion:

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

where k denotes the number of parameters in the model.

Rule:

The bigger the BIC, the better the model!

LUXEMBOURG SCHOOL OF FINANCE

Model Selection (2)

Leave-one-out Cross-Validation Approach:

$$CVE = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\left|y_{i_t} - \hat{y}_{i_t}^{[-i]}\right|. \qquad (10)$$

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

$$\Rightarrow P(j/Y_i) = \frac{P(Y_i/j)\hat{\pi}_j}{\sum_{j=1}^{r} P(Y_i/j)\hat{\pi}_j}. \tag{11}$$

Our Model Selection Criterion

We propose to choose the number of groups which explains the predicted group memberships best. Let's define the posterior probability criterion PPC by

$$PPC = \sum_{i=1}^{N} (\max_{j=1,\ldots r} P(j/Y_i)) \qquad (12)$$

Rule:

The bigger the PPC, 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.

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: Optimal mix of funded and unfunded pension systems: the case of Luxembourg. *Pensions* **17-4** p. 208-222.
- Schiltz, J. 2015: A generalization of Nagin's finite mixture model. In: Dependent data in social sciences research: Forms, issues, and methods of analysis' Mark Stemmler, Alexander von Eye & Wolfgang Wiedermann (Eds.). Springer 2015.