Methodological challenges to the analysis of population-based trajectory profiles

Frauke Kreuter
Bengt Muthén

CILVR, 2006

Recent debate in Criminology

- How to model criminal trajectories?
  - Continuously varied growth?
  - Growth variation captured through trajectory classes?
- How many trajectory classes are needed to capture the variation in growth?
  - What are reasonable indicators to make this decision?
  - How to compare non-nested models?
- How to interpret trajectory classes?
  - Do classes approximate an unknown distribution or
  - do they show conceptually distinct groups?
  - What indicators can be used to support this decision process?

2

Model comparison

- Growth Curve Model
  (regular hierarchical linear model)
- Growth Mixture Model
- Non Parametric Growth Mixture Model
- Latent Class Growth Analysis
  (Group based trajectory models)
- What difference does it make
  - Substantive interpretation
  - Sensitivity towards outliers
  - Predictive power

3

Example – Criminal behavior

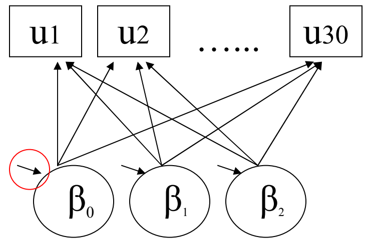
- "Cambridge study" (Farrington/West 1990)
- N = 411 (403)
- Age 10 to 40
- Number of convictions each year
- 60 percent never convicted
- In any given year 98.5% to 88.8% zero
- Biannual 97.1% to 83.2% zero

4

Representation of GCM

$$\ln(\lambda_{it}) = \beta_{0i} + \beta_{1i}x_t + \beta_{2i}x_t^{\,2}$$

$$\beta_{0i} = \alpha_0 + \zeta_{0i}$$



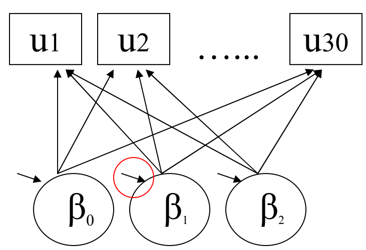
5

Representation of GCM

$$\ln(\lambda_{it}) = \beta_{0i} + \beta_{1i}x_t + \beta_{2i}x_t^{\,2}$$

$$\beta_{0i} = \alpha_0 + \zeta_{0i}$$

$$\beta_{1i} = \alpha_1 + \zeta_{1i}$$

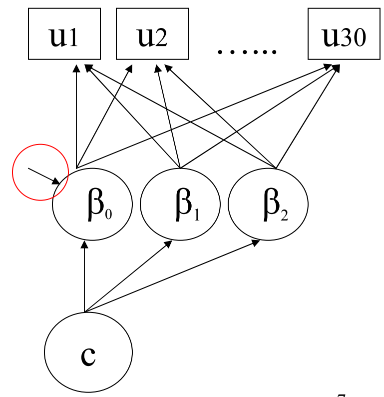


6

Representation of GMM

$$\ln(\lambda_{it|c_i=k}) = \beta_{0ki} + \beta_{1k} x_t + \beta_{2k} x_t^{\,2}$$

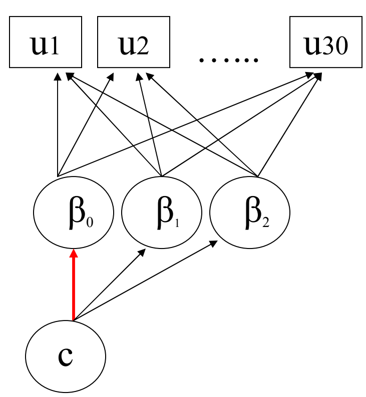
$$\beta_{0ki} = \alpha_{0k} + \zeta_{0ki}$$



7

Representation of GMM - NP

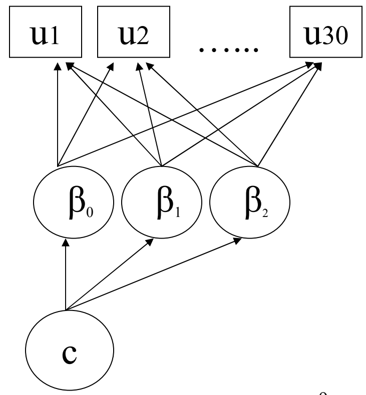
$$\ln(\lambda_{it|c_i=k}) = \beta_{0k} + \beta_{1k} x_t + \beta_{2k} x_t^{\,2}$$



8

Representation of LCGA

$$\ln(\lambda_{it|c_i=k}) = \beta_{0k} + \beta_{1k}x_t + \beta_{2k}x_t^{\ 2}$$



9

General features of all models

- Outcome treated as ZIP (Lambert 1992)
- Models estimated using random starting values
- Maximum likelihood estimation via Mplus
- Number of classes assessed using Bootstrap Likelihood Ratio Test (BLRT) (available now in Mplus V4)

10

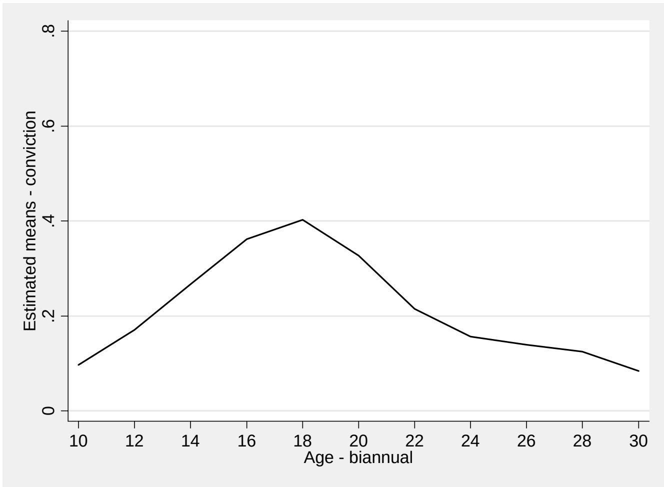
Modeling strategy

Growth Curve Model - Results

Model	RE	Log Likelihood	#P	BIC
Growth zip	i	-1481.3	7	3004.7
Growth zip	i s	-1469.6	9	2993.2
Growth zip	i s q	-1465.7	12	3003.5

12

Overall trajectory for "best" GCM



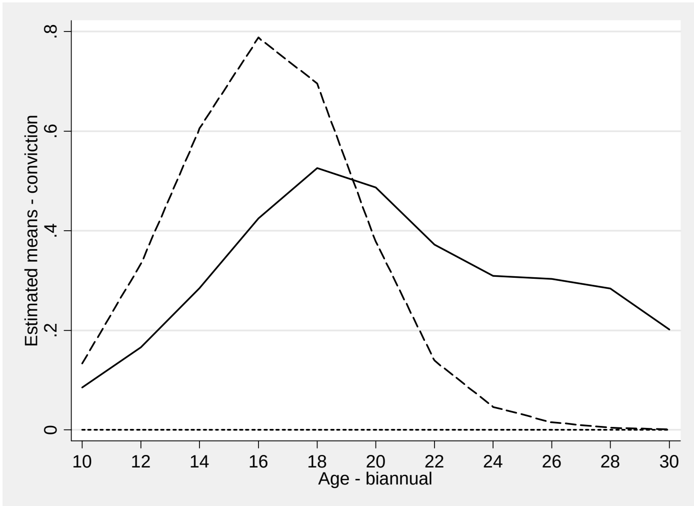
13

Growth Mixture Model - Results

Model	*Class*	*RE*	*Log Likelihood*	*#P*	*BIC*
Growth zip		i s	-1469.6	9	2993.2
GMM zip	1+0	i	-1473.3	8	2994.5
GMM zip	1+0	i s	-1461.8	10	2983.7
GMM zip	2+0	i	-1454.7	12	2981.5
GMM zip	3+0	i	-1450.7	16	2997.3

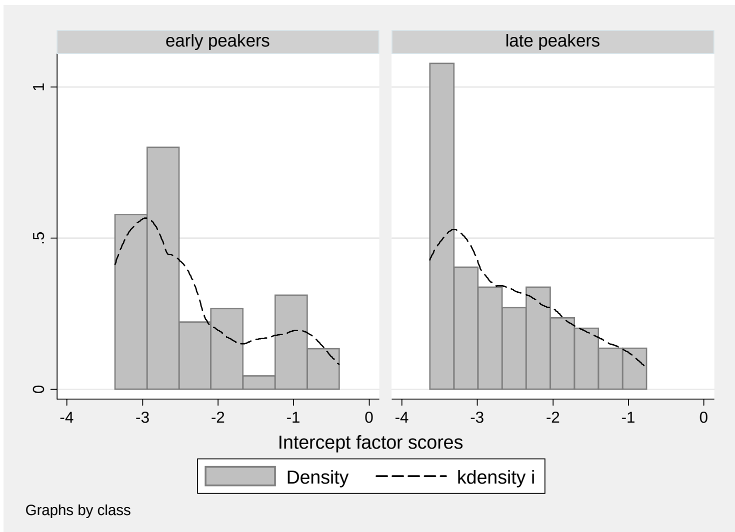
14

Growth trajectories for “best” GMM



15

Distribution of intercept factor scores



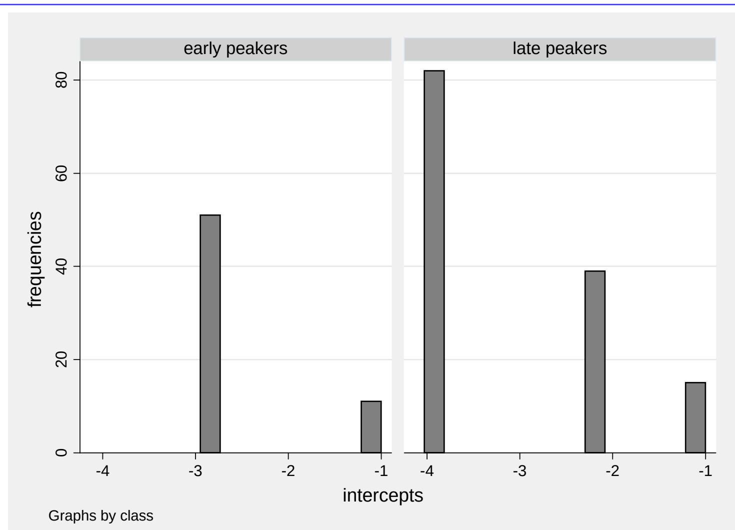
16

GMM non parametric - Results

Model	Class	Log Likelihood	#P	BIC
GMM-np zip	2(3)+0	-1444.5	16	2985
GMM-np zip	2(2+3)+0	-1444.4	15	2978.8
GMM-np zip	2(2+2)+0	-1457.7	13	2993

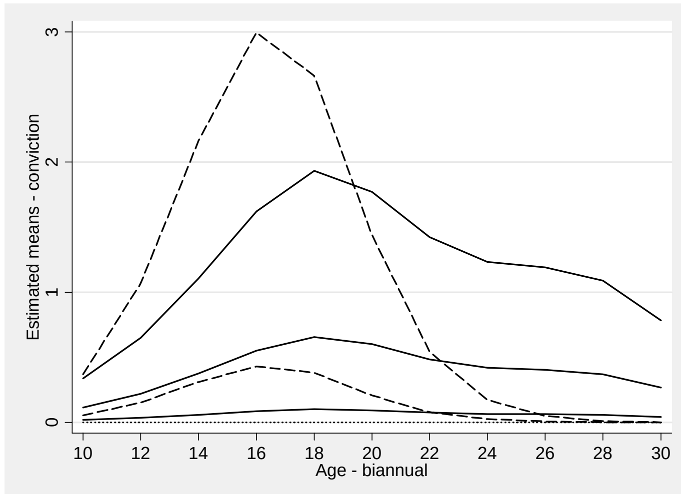
17

Support points for GMM-NP



18

Growth trajectories for "best" GMM-NP



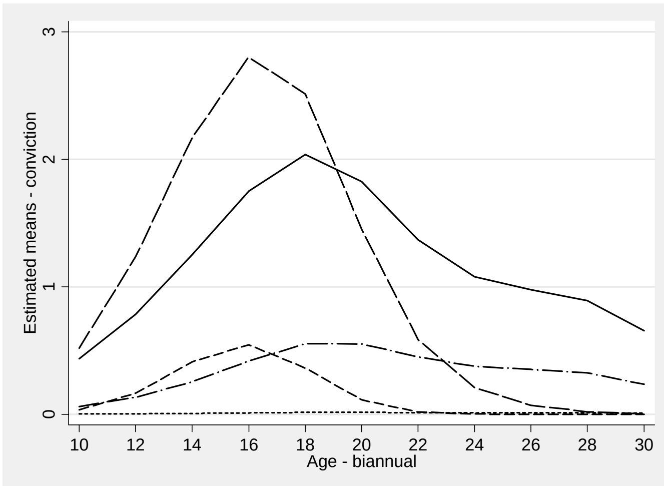
19

LCGA - Results

Model	*Class*	*Log Likelihood*	*#P*	*BIC*
GMM-np zip	2(2+3)+0	-1444.4	15	2978.8
LCGA zip	3	-1463.7	14	3011.6
LCGA zip	4	-1450.0	18	3008.0
LCGA zip	5	-1441.0	22	3014.0
LCGA zip	6	-1435.2	26	3026.4

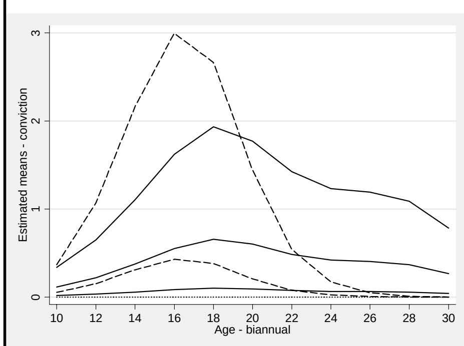
20

Growth trajectories for “best” LCGA

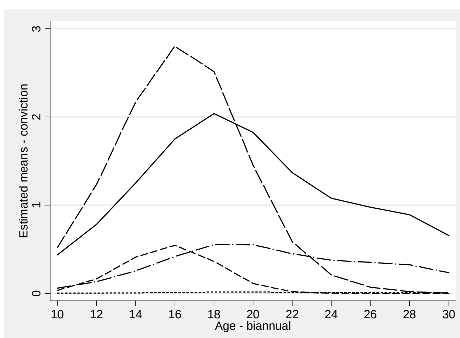


21

GMM-NP and LCGA



GMM - NP



LCGA

22

Model comparison

Favorite models

Model	*Class (re)*	*Log Likelihood*	*#P*	*BIC*
GMM zip	2+0 (i)	-1454.7	12	2981.5
GMM-np zip	2(2+3)+0	-1444.4	15	2978.8
LCGA zip	5	-1441.0	22	3014.0

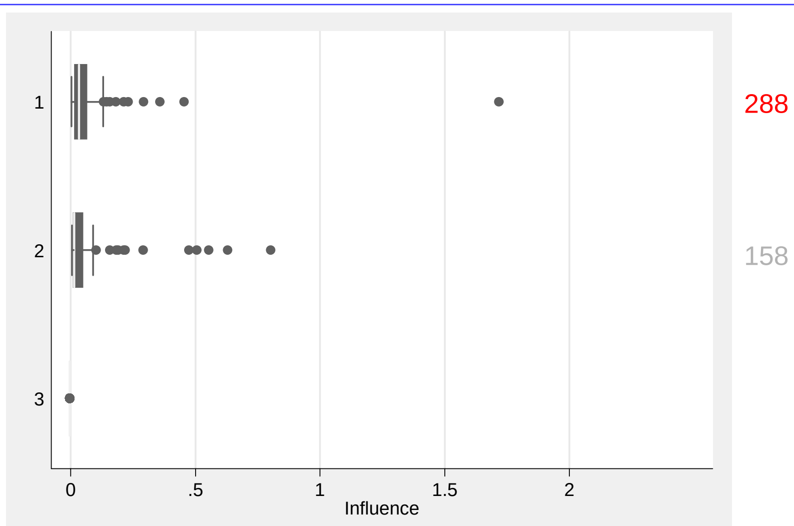
24

Substantive summary

- All models show substantive amount of boys in zero class
- All models pick up two substantive themes:
  - early peak and desistance
  - late peak and continuation
- Three classes (one zero and two substantive) seems to be all that is needed to fit the data.
- Variation around the substantive classes can be modeled non-parametrically.

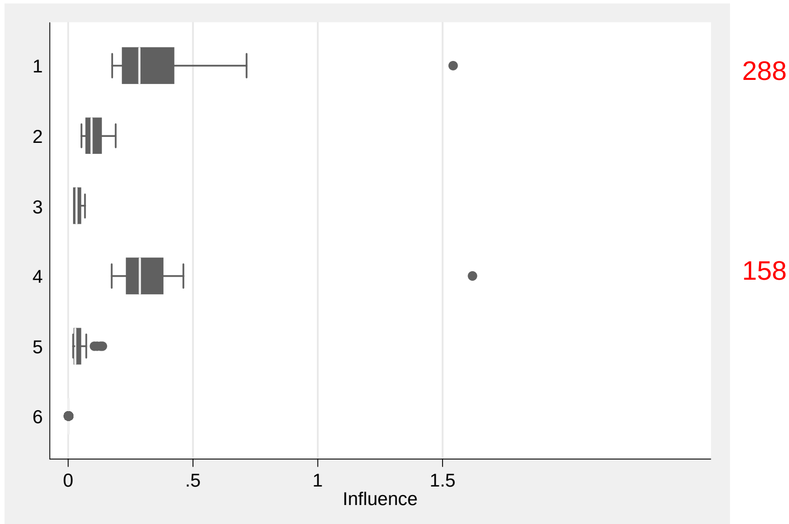
25

Influential cases – GMM



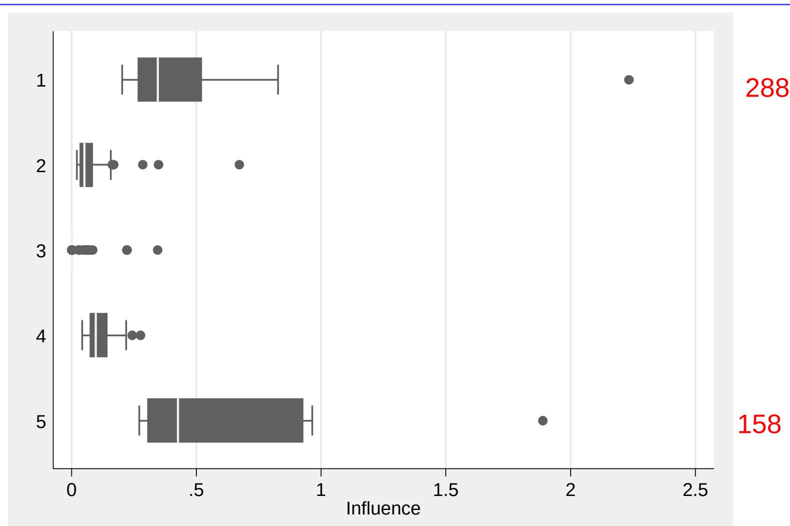
26

Influential cases
GMM (2(3)+0) NP



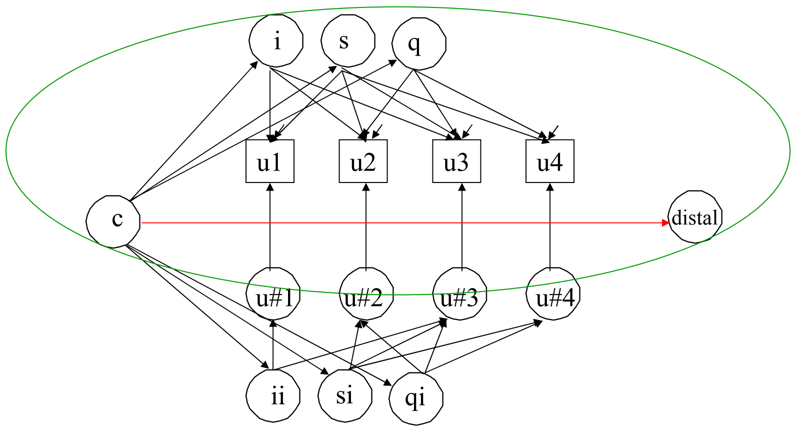
27

Influential cases – LCGA



28

Predictive power



29

Adding distal outcome

Model	*Class*	*Log Likelihood*	*#P*	*BIC*
GMM zip	2+0	-1477.5	15	3044.9
GMM zip NP	2(2+3)+0	-1462.0	21	3049.9
LCGA zip	5	-1458.1	27	3078.1

30

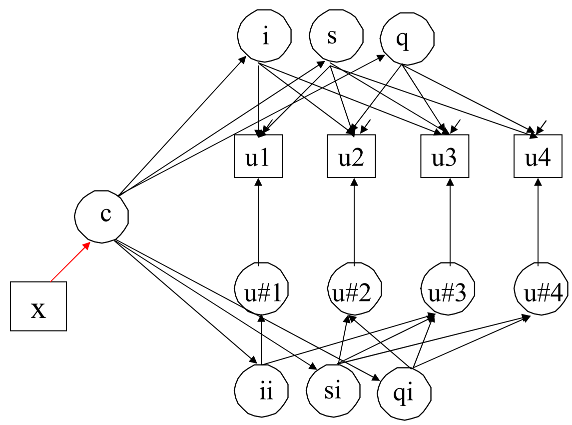
Summary of model examination

- Removing influential cases does not change model results.
- Adding distal outcome shows non-zero probability for the late-peaking class.
- In GMM-NP and LCGA non-zero probability for one of the late-peaking themes.
- Normality assumption of GMM is harmless

31

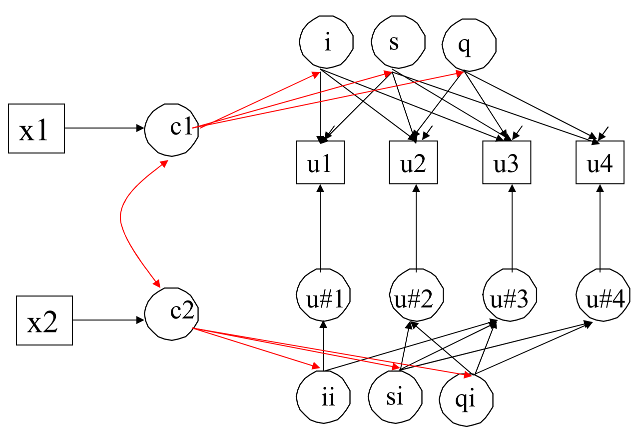
Model extensions

Model extensions



33

Model extensions cont'd



34