Identifying the Correct Number of Classes in a Growth Mixture Model

Davood Tofighi

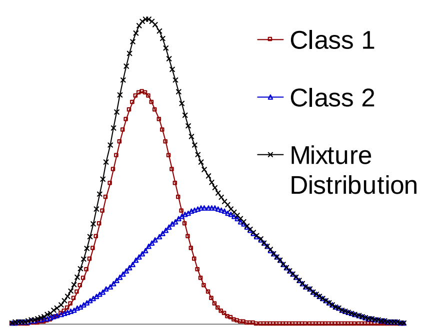
Craig Enders

Arizona State University

CILVR Conference, May 18, 2006

Mixture Modeling

- **Heterogeneity exists such that the data are comprised of two or more latent classes with different distributions**



$$f(\mathbf{y}) = \sum_{k=1}^{K} \pi_k \phi_k \left(\mathbf{y}; \mu_k, \Sigma_k\right)$$

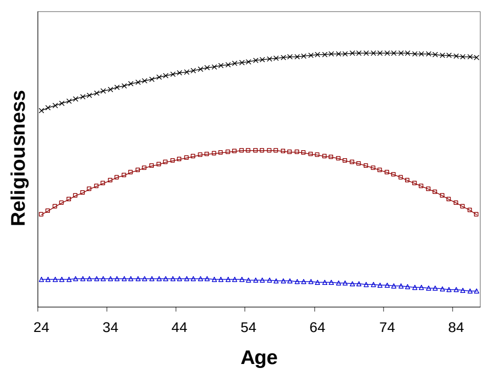
CILVR Conference, May 18, 2006

Growth Mixture Modeling (GMM)

- *k* latent classes with different growth trajectories and variance components
- Class membership may be related to covariates and distal outcomes
- GMM is analogous to a multiple-group growth model, but group membership is unobserved

Religiousness Example

- McCullough, Enders, Brion (2006)
- Three classes of religious development



Enumerating Latent Classes

- How many trajectory classes are there?
- Information-based criteria
  - BIC, AIC, etc.
- Likelihood ratio tests
  - Lo, Mendell, Rubin (2001)
- Goodness of fit tests
  - Tests based on model-implied skewness and kurtosis (Muthén, 2003)

Bayesian Information Criterion

- Based on the log likelihood and penalty terms related to model complexity

$$\text{BIC} = -2LL + p\ln(N)$$

- The sample-size adjusted BIC (SABIC) replaces *N* with (*N* + 2) / 24

Akaike Information Criterion

- Similar idea as the BIC …

$$\text{AIC} = -2LL + 2p$$

- The consistent AIC (CAIC) is

$$\text{CAIC} = -2LL + p(\ln[N] + 1)$$

- A sample-size adjusted CAIC (SACAIC) replaces *N* with (*N* + 2) / 24

CILVR Conference, May 18, 2006

Likelihood Ratio Tests

- Likelihood ratio tests can be used to compare a *k* versus *k* - 1 class model

$$\text{LRT} = -2(LL_{k-1} - LL_k)$$

- The LRT is not chi-square distributed
- Class probabilities for the nested *k* - 1 class model are at the boundary (zero)

CILVR Conference, May 18, 2006

Lo, Mendell, and Rubin (2001)

- Derived an appropriate reference distribution for the LRT, and an ad hoc adjustment to the test statistic
- LMR and adjusted LMR (ALMR)
- A small *p* value suggests that the *k* class model is favored over *k* - 1 classes

CILVR Conference, May 18, 2006

Bootstrapping The LRT

- The *k* and *k* - 1 class models are fit to a number of bootstrap samples
- A *p* value for the LRT is obtained from the empirical reference distribution of bootstrapped LRT values
- See Nylund, Asparouhov, and Muthén (2006) for detailed simulation results

CILVR Conference, May 18, 2006

Fit Tests Based On Multivariate Skewness And Kurtosis

- **Model-implied skewness and kurtosis from the *k* class model are compared to the sample moments (Muthén, 2003)**
- **Analogous to the GOF test in SEM**
- **A large *p* value indicates that the *k* class model accurately reproduces the higher-order moments**
- **Herein referred to as MST and MKT**

An Artificial Data Example

- **An artificial data set (*N* = 1000) was generated from a population with three trajectory classes**
- **A sequence of models was fit (*k* = 1 to 4) to illustrate the class extraction process**

Analysis Results

	k = 1	*k* = 2	*k* = 3	*k* = 4
BIC	27333.77	26680.68	26623.59	26649.66
SABIC	27311.53	26633.03	26550.54	26560.73
LMR	N/A	$p < .001$	$p < .001$	$p = .09$
BLRT	N/A	$p < .001$	$p = 1.00$	$p = 1.00$
MST	$p < .001$	$p = .26$	$p = .82$	$p = .86$
MKT	$p < .001$	$p = .92$	$p = .49$	$p = .55$

CILVR Conference, May 18, 2006

More On The LRT

- Multiple *k* - 1 class models are possible
- For example, when testing a 3 class model, three different 2 class models could result
- M*plus* discards the first class when fitting the *k* - 1 model
- Starting values must be used to ensure that classes are ordered properly

CILVR Conference, May 18, 2006

Example

Correct ordering

- 1-class model
  - LL_1 = -13642.71
- 2-class model
  - LL_1 = -13642.71
  - LL_2 = -13288.53
- 3-class model
  - LL_2 = -13288.53
  - LL_3 = -13232.36

Incorrect ordering

- 1-class model
  - LL_1 = -13642.71
- 2-class model
  - LL_1 = -13642.71
  - LL_2 = -13288.53
- 3-class model
  - LL_2 = -13265.72
  - LL_3 = -13232.36

CILVR Conference, May 18, 2006

Purpose Of Study

- In the previous example, the fit indices did not agree on the number of classes
- Which index should be used to determine the number of classes?
- We designed a Monte Carlo simulation study to address this question

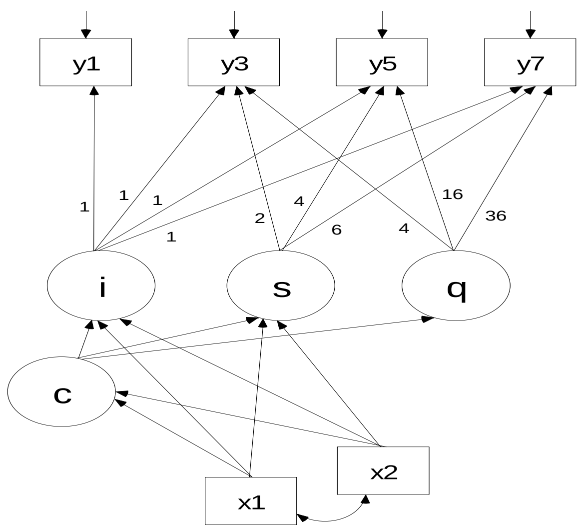
CILVR Conference, May 18, 2006

Simulation Procedure

- **Data were generated from a population with three trajectory classes**
- **A sequence of GMMs was fit (*k* = 2 to 4)**
- **Extraction was performed with and without covariates**
- **In what proportion of replications was the *k* = 3 class model recovered?**

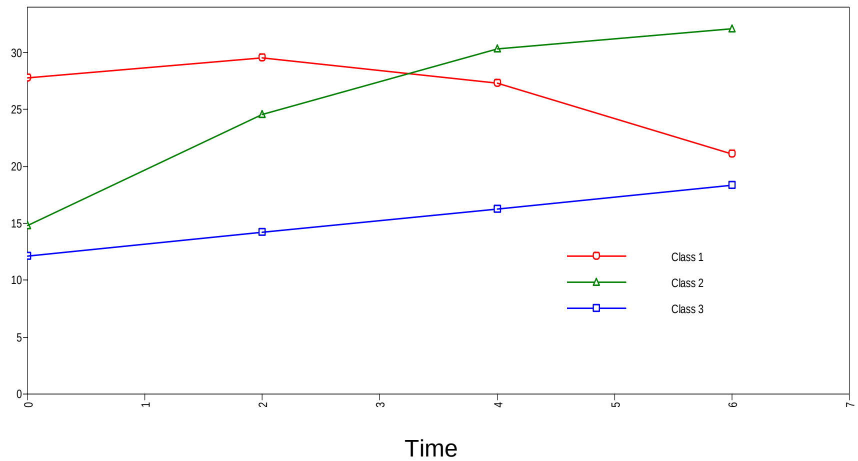
CILVR Conference, May 18, 2006

Data Generation Model



CILVR Conference, May 18, 2006

Population Trajectory Classes



CILVR Conference, May 18, 2006

Manipulated Variables

- **Number of repeated measures**
  - ***t* = 4, 7**
- **Sample size**
  - ***N* = 400, 700, 1000, 2000**
- **Mixing proportions**
  - **20%, 33%, 47% and 7%, 36%, 57%**
- **Within-class normality**
  - ***S* = 0, *K* = 0 and *S* = 1, *K* = 1**
- **Class separation**
  - **“High” and “Low”**

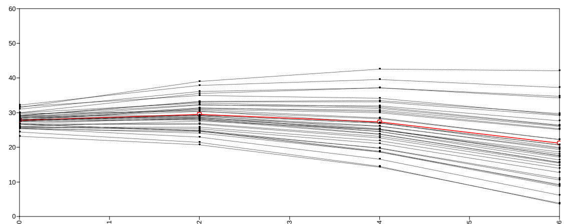
CILVR Conference, May 18, 2006

Manipulating Class Separation

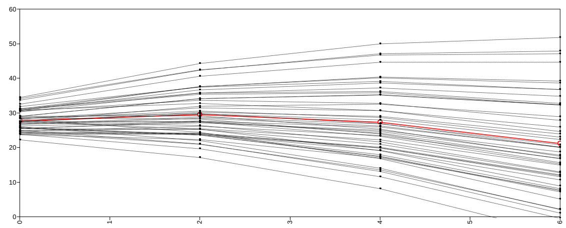
- **Definition of separation is subjective, but based on previous experience (e.g., McCullough, Enders, & Brion, 2005)**
- **Within-class variance components were increased in magnitude to create the low separation condition**
- **Mean growth trajectories did not change**

CILVR Conference, May 18, 2006

Class 1: High Versus Low Separation



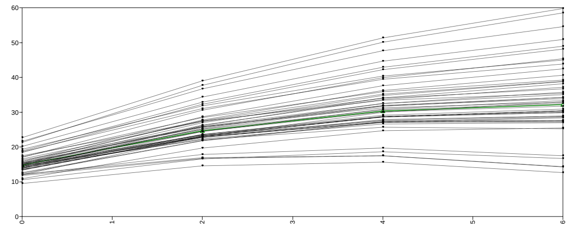
High Separation



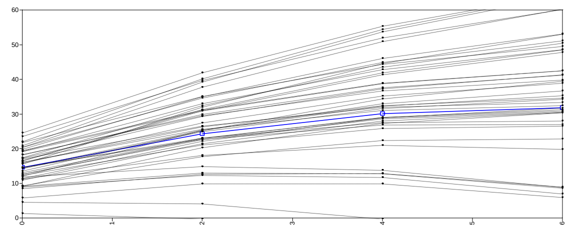
Low Separation

CILVR Conference, May 18, 2006

Class 2: High Versus Low Separation



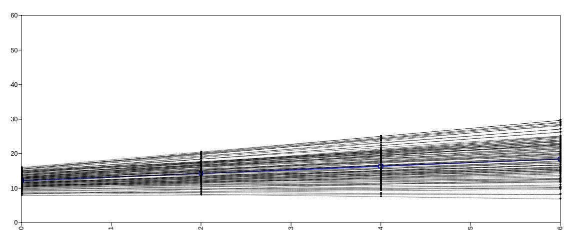
High Separation



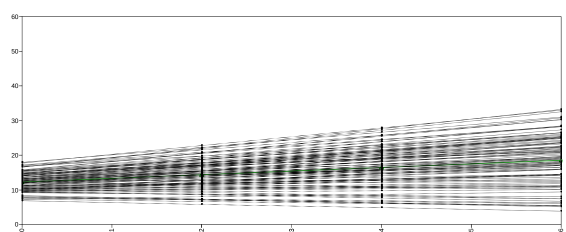
Low Separation

CILVR Conference, May 18, 2006

Class 3: High Versus Low Separation



High Separation



Low Separation

CILVR Conference, May 18, 2006

Average Class Probabilities

Low Separation

	1	2	3
1	.85	.03	.12
2	.02	.81	.18
3	.07	.13	.81

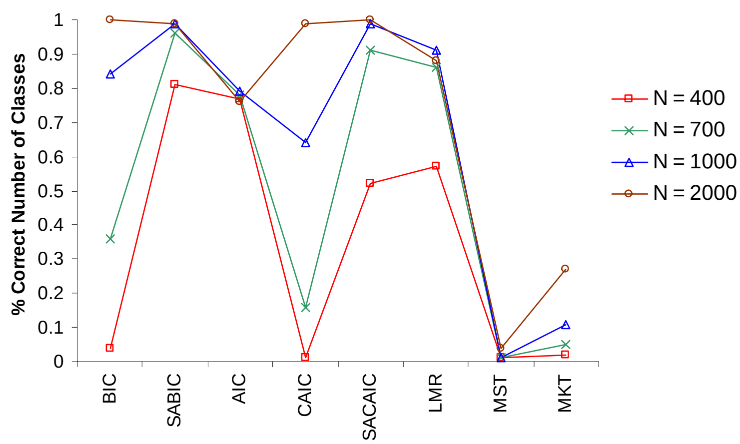
High Separation

	1	2	3
1	.91	.08	.01
2	.05	.89	.08
3	.01	.11	.89

CILVR Conference, May 18, 2006

Sample Size (No Covariates)

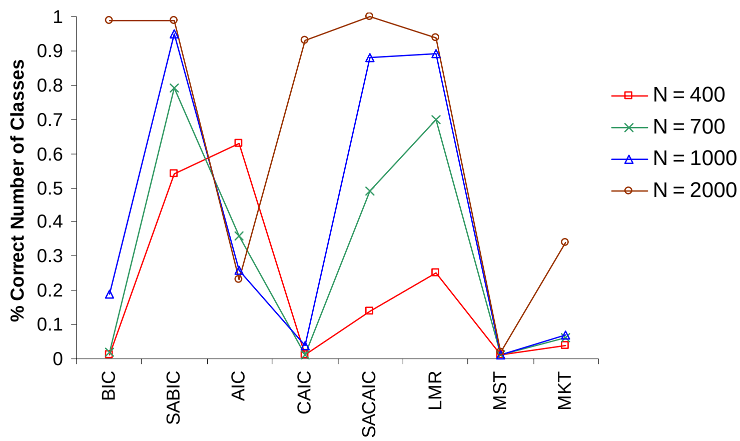
Normality, High Separation, *t* = 4, Proportions = 20:33:47



CILVR Conference, May 18, 2006

Sample Size (With Covariates)

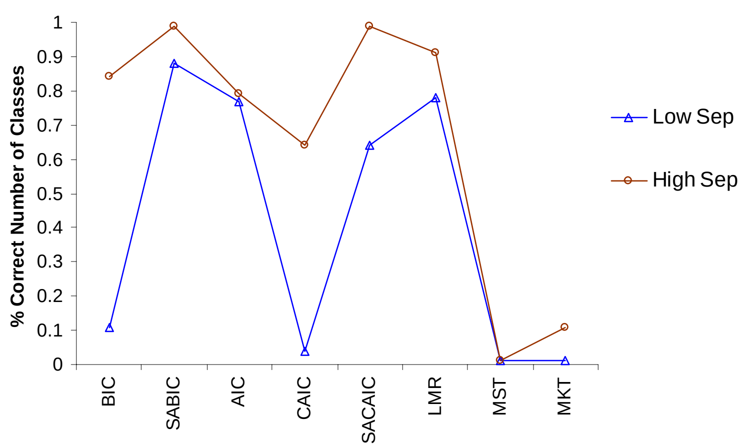
Normality, High Separation, t = 4, Proportions = 20:33:47



CILVR Conference, May 18, 2006

Class Separation (No Covariates)

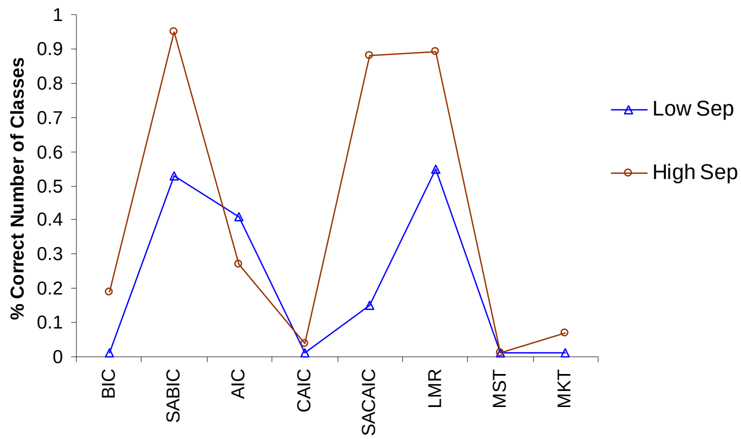
Normality, N = 1000, t = 4, Proportions = 20:33:47



CILVR Conference, May 18, 2006

Class Separation (Covariates)

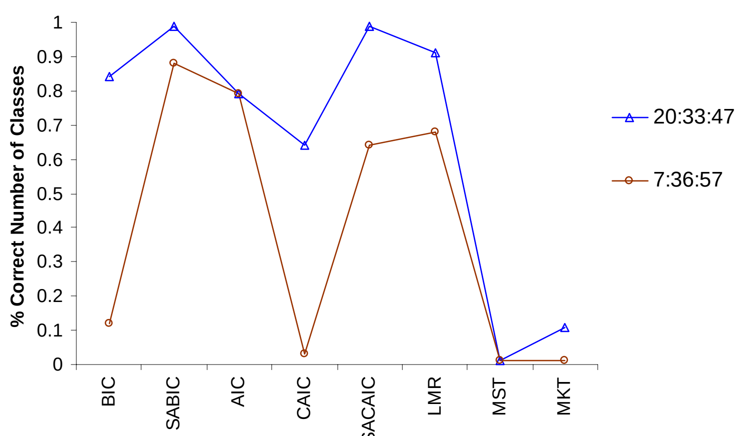
Normality, *N* = 1000, *t* = 4, Proportions = 20:33:47



CILVR Conference, May 18, 2006

Mixing Proportion (No Covariates)

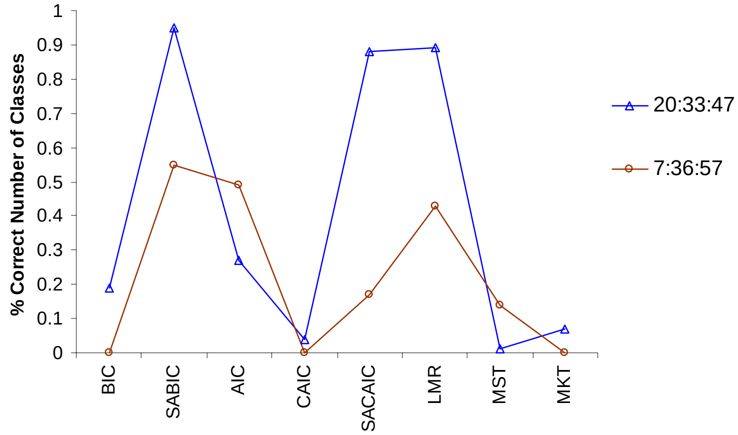
Normality, *N* = 1000, *t* = 4, High Separation



CILVR Conference, May 18, 2006

Mixing Proportion (Covariates)

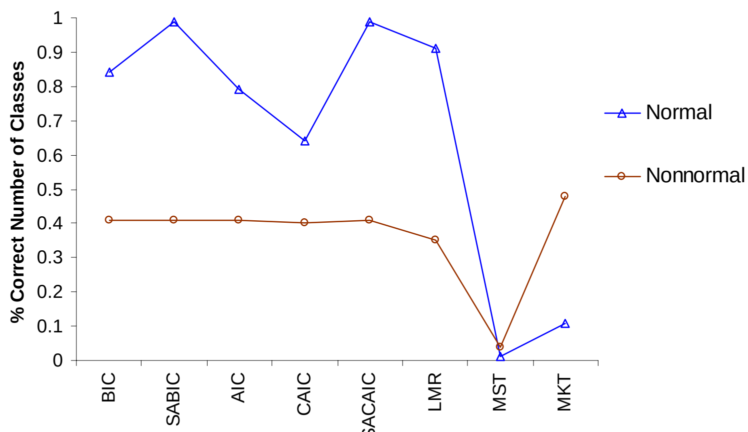
Normality, *N* = 1000, *t* = 4, High Separation



CILVR Conference, May 18, 2006

Normality (No Covariates)

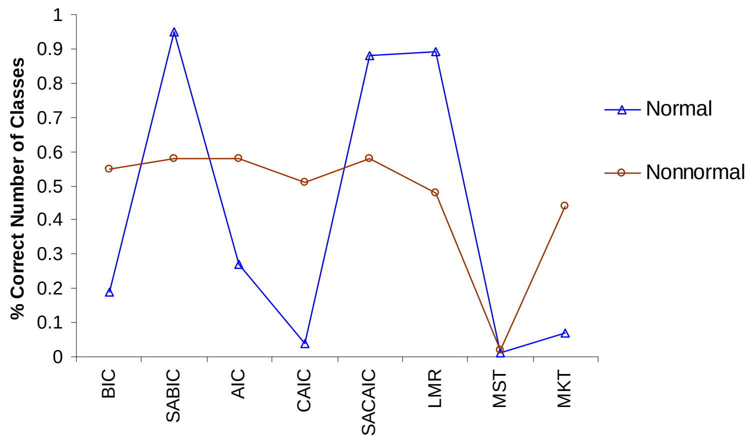
N = 1000, *t* = 4, Proportions = 20:33:47, High Separation



CILVR Conference, May 18, 2006

Normality (Covariates)

N = 1000, *t* = 4, Proportions = 20:33:47, High Separation



Information-Based Criteria

- **The SABIC very accurately detected the number of latent classes**
- **At small *N*s, it had a slight tendency to extract too few classes**
- **Note that the *k* class model was retained if the SABIC decreased by <u>any</u> amount**
- **Other information-based criteria performed poorly**

Likelihood Ratio Tests

- The LMR and ALMR performed well, but were somewhat less powerful than the SABIC
- LMR tended to extract too few classes at small *N*s, and too many classes at large *N*s

MSK and MKT

- MSK and MKT uniformly extracted too few classes
- The performance of these measures may be model-dependent
- MSK was slightly more accurate than the LMR in a pilot study with a slightly different set of mixture distributions

The Use Of Covariates

- The inclusion of covariates dramatically decreased power, and resulted in the extraction of too few classes
- e.g., At N = 400, the SABIC extracted two classes 39% of the time, compared to 14% when covariates were excluded
- The use of covariates should be avoided unless N is very large

CILVR Conference, May 18, 2006

Nonnormal Data

- Mild violations of within-class normality led to the extraction of too many classes
- None of the tests we studied was immune to this problem
- Is bootstrapping the LRT a solution?

CILVR Conference, May 18, 2006

Bootstrapping The LRT

- Preliminary results suggest that the bootstrap is less powerful than the LMR
- e.g., The k = 2 class model was correctly rejected about 75% and 10% of the time in the high and low separation conditions, respectively
- See Nylund, Asparouhov, and Muthén (2006) for detailed simulation results

How Likely Is Over-Extraction?

- Extracting too many classes frequently produced problematic solutions (e.g., negative variances, unstable solutions)
- Estimating class-specific variances probably prevents over-extraction
- Invoking constraints to attain convergence (e.g., fixing variances to zero) is likely a sign of over-extraction