Finite Mixture Growth Models: Problems and Opportunities

Daniel J. Bauer
North Carolina State University

Seminar Presentation
University of Michigan
October 3, 2003

Outline of Talk

- Theoretical Motivation
- Empirical Examples
- The Growth Mixture Model
- Problems and Opportunities
  - *Distributional Assumptions*
  - *Model Specification*
  - *Nonlinear Relationships*
- Conclusions

Theoretical Motivation

"To establish the development of pathology, an entire profile of developmental lines or pathways needs to be examined and compared to normal development for each line of functioning" (Loeber et al., 1993, p. 104).

“...there is still little that we can say with confidence about...why antisocial trajectories develop, why they broaden and deepen with development in some children yet taper off in others, and why they are so difficult to deflect once stabilized” (Richters & Cicchetti, 1993, p. 3).

“…temporary versus persistent antisocial persons constitute… two qualitatively distinct categories of individuals, each in need of its own distinct theoretical explanation.”

(Moffitt, 1993)

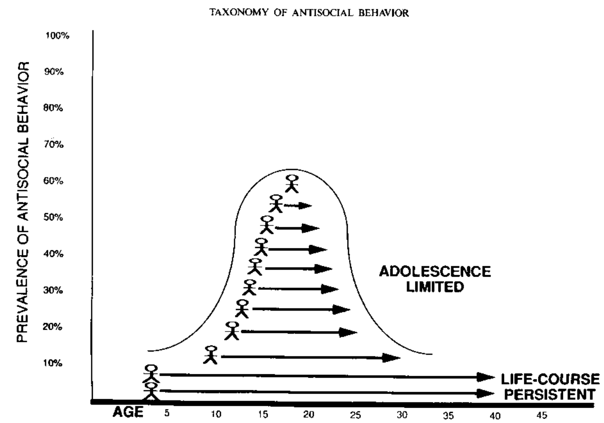
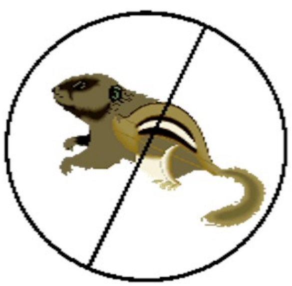


Figure 3. Hypothetical illustration of the changing prevalence of participation in antisocial behavior across the life course. (The solid line represents the known curve of crime over age. The arrows represent the duration of participation in antisocial behavior by individuals.)



“There are gophers,

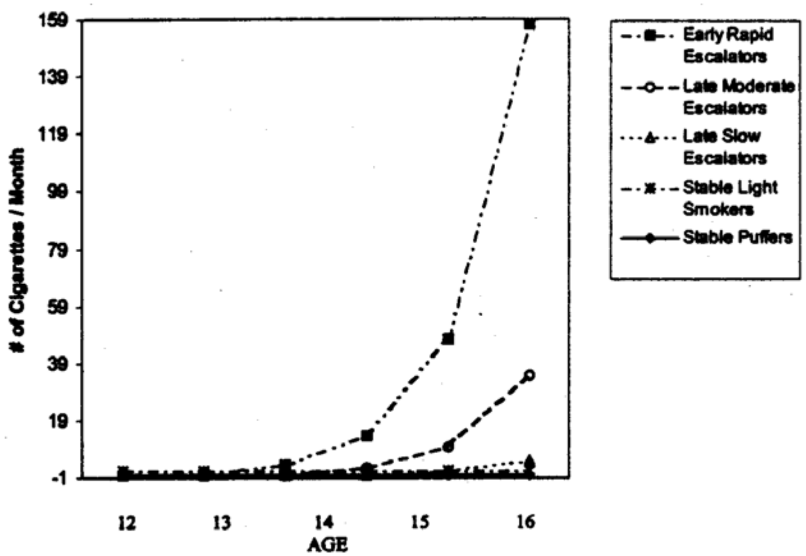
there are chipmunks,

but there are no gophmunks.”

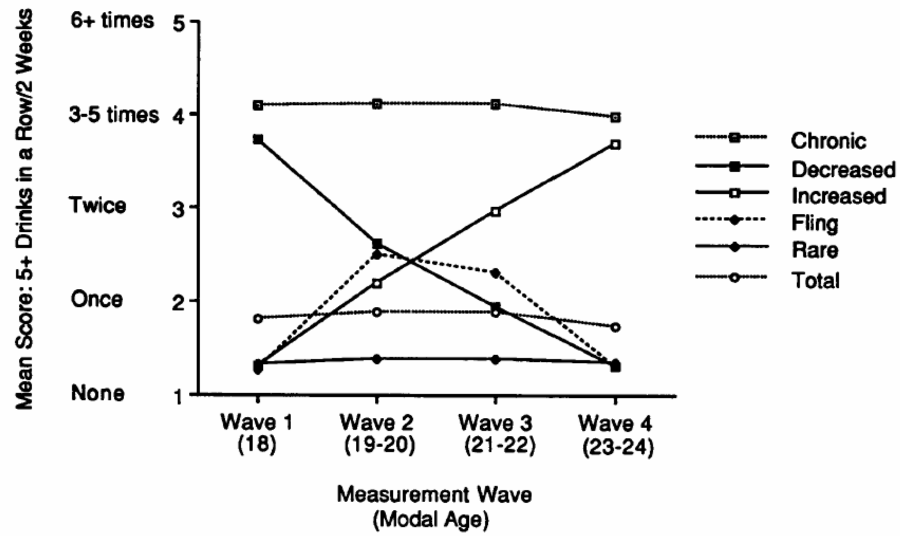
(Meehl, 1994)

Empirical Examples

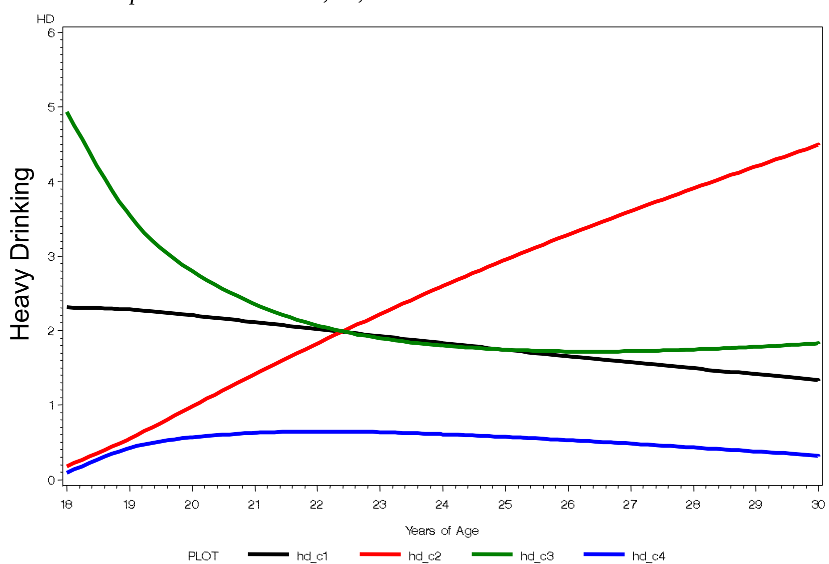
Colder, C.R., Mehta, P., Balanda, K., Campbell, R.T., Mayhew, K.P., Stanton, W.R., Pentz, M.A. & Flay, B.R. (2001). Identifying trajectories of adolescent smoking: An application of latent growth mixture modeling. *Health Psychology, 20*, 127-135.



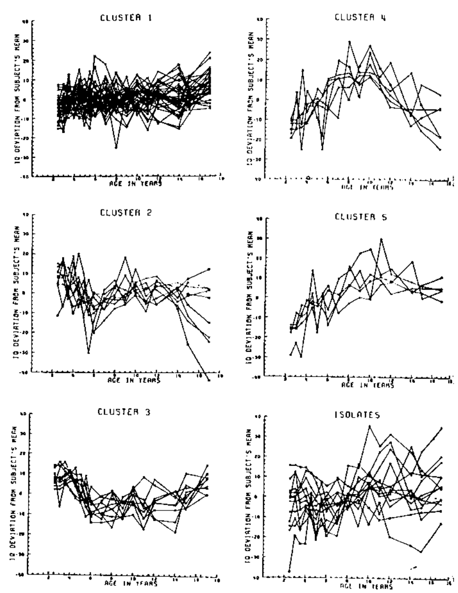
Schulenberg, J., O'Malley, P. M., Bachman, J. G., Wadsworth, K. N. & Johnston, L. D. (1996). Getting drunk and growing up: Trajectories of frequent binge drinking during the transition to young adulthood. *Journal of Studies on Alcohol, 57*, 289-304.



Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882-891.



McCall, R.B., Appelbaum, M.I., & Hogarty, P.S. (1973). *Developmental Changes in Mental Performance*. Monographs for the Society for Research in Child Development.



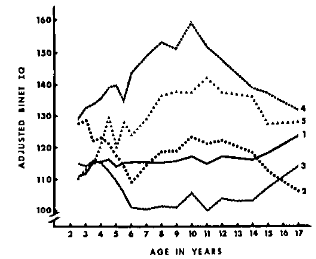


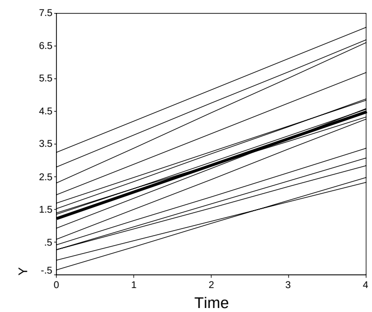
FIGURE 5.—Mean IQ (adjusted for differences between Binet revision) over age for the five IQ clusters.

The Growth Mixture Model

The Random Coefficient Model

- A standard random coefficient growth model can be written:

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$$



Where

$\mathbf{y}_i$ is the vector of repeated measures

$\boldsymbol{\Lambda}$ is the design matrix (here assumed balanced)

$\boldsymbol{\eta}_i$ is the vector of random coefficients for the individual trajectories (i.e., intercept, slope)

$\boldsymbol{\varepsilon}_i$ is the vector of time-specific residuals from the trajectories

The Random Coefficient Model

- On the assumption that

$$\begin{bmatrix} \boldsymbol{\eta}_i \\ \boldsymbol{\varepsilon}_i \end{bmatrix} \sim MVN\left(\begin{bmatrix} \boldsymbol{\alpha} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta} \end{bmatrix}\right)$$

the pdf (or marginal model) for $\mathbf{y}_i$ is:

$$f(\mathbf{y}_i) = \phi\left(\mathbf{y}_i; \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta})\right)$$

where
$$\boldsymbol{\mu}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\alpha}$$
$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}$$

The Growth Mixture Model

- Now assume that there are K groups, each with their own random coefficient growth model.

- Retain the assumption of normality of $(\boldsymbol{\eta}_i \quad \boldsymbol{\varepsilon}_i)'$ *within groups*.

- The marginal model for $\mathbf{y}_i$ is now a mixture of normals with structured means and covariances, or

$$f(\mathbf{y}_i) = \sum_{k=1}^{K} \pi_k \phi_k\left(\mathbf{y}_i; \boldsymbol{\mu}_k(\boldsymbol{\theta}_k), \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_k)\right)$$

where
$$\boldsymbol{\mu}_k(\boldsymbol{\theta}_k) = \boldsymbol{\Lambda}_k\boldsymbol{\alpha}_k$$
$$\boldsymbol{\Sigma}_k(\boldsymbol{\theta}_k) = \boldsymbol{\Lambda}_k\boldsymbol{\Psi}_k\boldsymbol{\Lambda}_k' + \boldsymbol{\Theta}_k$$

Applications of Mixtures

Titterington, Smith & Makov (1985, pp. 2-3):

> "By a **direct application**, we have in mind a situation where we believe, more or less, in the existence of *k* underlying categories or sources, such that the experimental unit on which the observation *X* is made belongs to one of these categories..."
>
> "By an **indirect application**, we have in mind a situation where the finite mixture form is simply being used as a mathematical device in order to provide an indirect means of obtaining a flexible, tractable form of analysis."

We will now consider the assumptions of the growth mixture model, and how they reflect on both possible types of applications.

Assumptions of the GMM

- The Component Distributions are Normal
- The Growth Model is Correctly Specified
- The Relationships Among $\mathbf{y}$ and η are Linear

What problems will violating these assumptions create for direct applications?

Are these 'problems' in fact opportunities for indirect applications?

Distributional Assumptions

A mixture of normals is necessarily nonnormal

(except in degenerate cases)

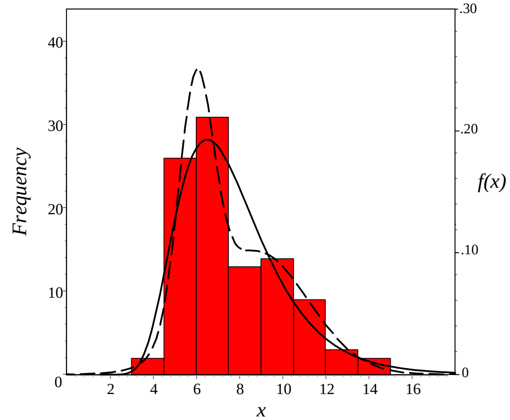
Nonnormality does not necessarily reflect a normal mixture

The Conundrum

Pearson (1895, p. 394):

"The question may be raised, how are we to discriminate between a true curve of skew type and a compound curve [or mixture]."

2 Groups or Just an Approximation?



Testing The Distributional Assumptions of Growth Mixture Models

Bauer, D.J. & Curran, P.J. (2003). Distributional assumptions of growth mixture models: Implications for over-extraction of latent trajectory classes. Forthcoming in *Psychological Methods*.

Hypothesis 1

With multivariate normal data, it should be *difficult* to estimate two trajectory classes (with random effects);

With multivariate *nonnormal* data, it should be *easy* to estimate at least two trajectory classes.

Hypothesis 2

A two-class model should generally only produce a significant increase in model fit for *nonnormal* data.

Testing The Distributional Assumptions of Structured Normal Mixture Models

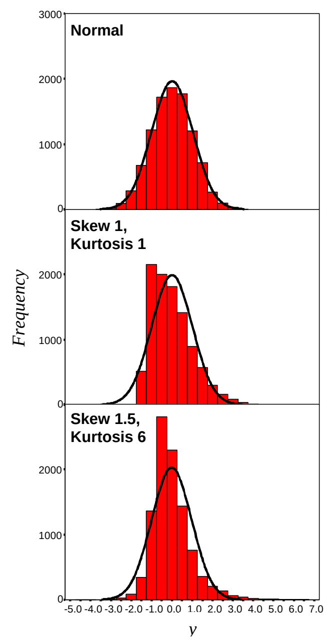
- 500 Datafiles Generated From a K=1 Random Coefficient Growth Model (N=200 or N=600)

$$\begin{pmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \eta_{0i} \\ \eta_{1i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{0i} \\ \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \end{pmatrix}$$

$$\boldsymbol{\alpha} = (1.0 \quad .80)' \qquad \boldsymbol{\Psi} = \begin{pmatrix} 1.0 & .11 \\ .11 & .20 \end{pmatrix}$$

$$\boldsymbol{\Theta} = DIAG(1.00, 1.42, 2.25, 3.47, 5.09)$$

Marginal Distributions



Hypotheses

- With normal data, it should be *difficult* to extract two classes, With nonnormal data, it should be *easy* to extract two classes.
- A two-class model should generally only produce a significant increase in model fit for *nonnormal data.*

Results (N=600)

Normal Data
18% Converged on Proper Solutions (89 Samples)
- BIC favored 2 Classes 0% of the time.
- CLC favored 2 Classes 2% of the time.

Skew 1, Kurtosis 1 Data
94% Converged on Proper Solutions (471 Samples)
- BIC favored 2 Classes 100% of the time.
- CLC favored 2 Classes 99% of the time.

Skew 1.5, Kurtosis 6 Data
97% Converged on Proper Solutions (485 Samples)
- BIC favored 2 Classes 100% of the time.
- CLC favored 2 Classes 100% of the time.

Problem for Direct Applications

- In practice, it will be difficult to know whether estimated classes reflect a true mixture or simply serve to accommodate violation of distributional assumptions.

Opportunity for Indirect Applications

- GMMs can be used to provide a semiparametric approximation to nonnormal distributions of repeated measures (random effects, residuals).
- GMMs avoid the traditional arbitrary assumption of normality.

Model Specification

Effect of Misspecification: Bivariate Case

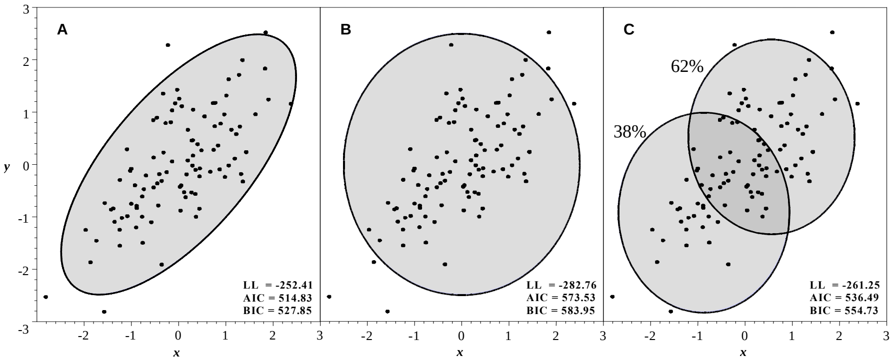
Population Model: Bivariate Standard Normal Distribution, $(\rho^2 = .50)$

Fitted Models:

A. Unrestricted One Class Model

B. Restricted One Class Model $(\hat{r}^2 = 0)$

C. Restricted Two Class Model $(\hat{r}_1^2 = 0;\ \hat{r}_2^{\,2} = 0;\ \hat{\sigma}_{x1}^2 = \hat{\sigma}_{x2}^2;\ \hat{\sigma}_{y1}^2 = \hat{\sigma}_{y2}^2)$



Misspecification of the Structural Model and Latent Class Extraction

Model for Means and Covariances of Aggregate Population:

$$\boldsymbol{\mu}(\boldsymbol{\pi}, \boldsymbol{\theta}_1 \cdots \boldsymbol{\theta}_K) = \sum_{k=1}^{K} \pi_k \boldsymbol{\mu}_k(\boldsymbol{\theta}_k)$$

$$\boldsymbol{\Sigma}(\boldsymbol{\pi}, \boldsymbol{\theta}_1 \cdots \boldsymbol{\theta}_K) = \sum_{k=1}^{K} \sum_{l=k+1}^{K} \pi_k \pi_l \left[\boldsymbol{\mu}_k(\boldsymbol{\theta}_k) - \boldsymbol{\mu}_l(\boldsymbol{\theta}_l)\right]\left[\boldsymbol{\mu}_k(\boldsymbol{\theta}_k) - \boldsymbol{\mu}_l(\boldsymbol{\theta}_l)\right]' + \sum_{k=1}^{K} \pi_k \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_k)$$

- The aggregate covariance matrix is partitioned into components reflecting *class differences in mean growth* and *within class covariance due to the random effects*.
- If the within-class growth model is misspecified, then the estimation of spurious latent classes with different mean growth may improve recovery of Σ.

The Impact of Misspecification of the Growth Model in Growth Mixture Models

- A popular variant of the GMM assumes variances and covariances of growth parameters are zero within classes (Nagin, 1999).

- If there is individual variability around class mean trajectories, this is a misspecified model.

- **Two wrongs make a right:**
  Misspecification of both the within-class growth model and the number of classes can lead to overall good fit.

Hypothesis

- Even with multivariate normally distributed data, spurious latent trajectory classes may be estimated and appear optimal if the growth model is misspecified.

Results (from exemplar replication)

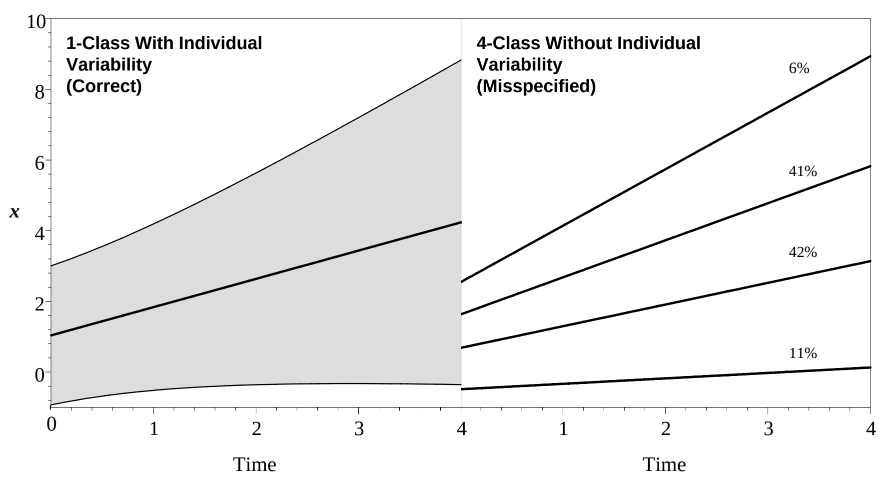
Individual Trajectories Distributed Around Mean Trajectory

Model	LL	AIC	BIC	SRMR
1-Class	6050.73	12121.46	12165.43	.027

No Individual Trajectories – Class Mean Trajectories Only

Model	LL	AIC	BIC	SRMR
1-Class	6492.43	12998.85	13029.63	.375
2-Class	6175.36	12380.72	12446.67	.111
3-Class	6089.23	12224.47	12325.6	.059
4-Class	6059.81	12181.62	12317.93	.049
5-Class	6043.50	12165.00	12336.48	.034
6-Class	6030.25	12154.50	12361.15	.032
7-Class	6014.40	12138.79	12380.63	.029
8-Class	6009.77	12145.54	12422.54	.028

Misspecification Leading to Spurious Latent Classes



Problem for Direct Applications

- In practice, one must guard against the estimation of spurious classes due to misspecification of the within-class growth model.

Opportunity for Indirect Applications

- Nagin's (1999) model can be used to identify modal patterns in growth and to examine local conditions relating to those patterns.

Nonlinear Relationships

Nonlinear Relationships

Nonlinear Relationships may be Viewed in Two Ways

(1) Distributional Assumptions are Violated

- Multivariate Normality → Linearity
- NonLinearity → Multivariate Nonnormality

(2) The Model is Misspecified

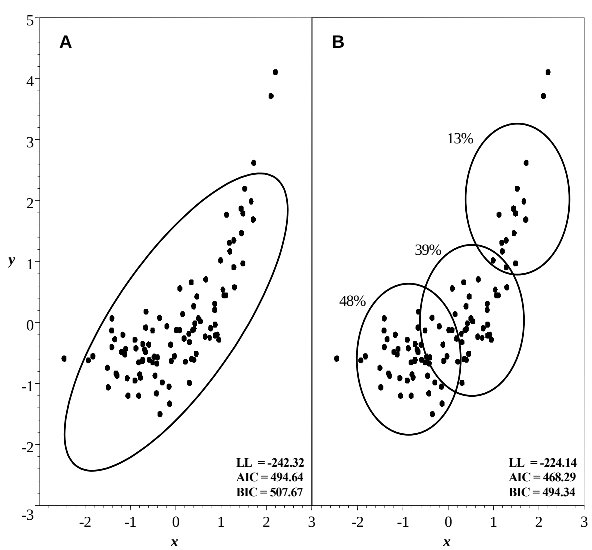
- Relationships modeled as linear when actually nonlinear
- Nonlinear function could be adequately approximated by inclusion of polynomial terms?

From either perspective, latent classes will be needed to accommodate nonlinearity, even if only one group exists in the population.

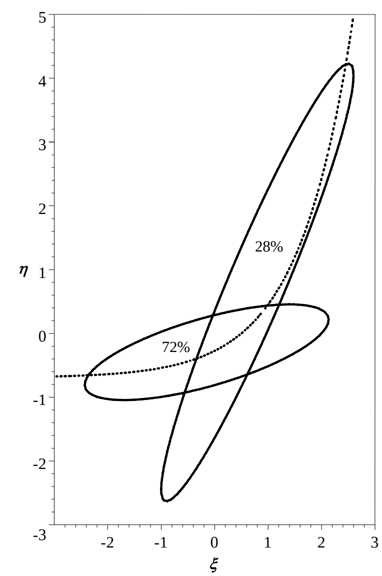
Effect of Nonlinearity: Bivariate Case

Latent Classes accommodate nonlinearity in the same way as nonnormality and model misfit

Here, differences in location accommodate nonlinearity



Effect of Nonlinearity: Bivariate Case



Here, differences in location *and* orientation of class distributions accomodate the nonlinear relationship.

Bivariate distribution could be the distribution of individual intercepts and slopes in a linear growth model.

Problem for Direct Applications

- In practice, latent classes may reflect nonlinear relationships among random effects rather than a true mixture.

Opportunity for Indirect Applications

- GMMs may be used to provide a semiparametric appoximation to possibly nonlinear relationships among random effects.
- Avoids the arbitrary assumption of linearity.

Conclusions

- Growth Mixture Models offer a number of new modeling possibilities, bringing both new problems and new opportunities for analysis.
  - The violation of several key assumptions of the traditional random coefficient growth model can induce the estimation of spurious classes.
    - The problem this poses for direct applications is that it may be difficult to discern the true function that the latent classes are serving (true clusters or not?).
      - The opportunity is that indirect applications of the GMM can recover features of the growth process that might otherwise go unmodeled (and typically constitute assumption violations).