Advances in Longitudinal Data Analysis: Longitudinal Mixture Modeling

Katie Witkiewitz, PhD
Department of Psychology
Washington State University

April 7th, 2009

Workshop at the 31st Annual Meeting of the Society of Behavioral Medicine
Goal of seminar is to introduce basic concepts, terminology and potential applications of longitudinal mixture models.

What will not be covered?

Level 1: $y_{it} = \eta_{0i} + \eta_{1i} x_t + \varepsilon_{it}$

Level 2: $\eta_{0i} = \alpha_0 + \zeta_{0i}$

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

Or...

$$P(W = w) = \sum_{c_1=1}^{C_1} \sum_{d_1=1}^{D_1} \cdots \sum_{c_T=1}^{C_T} \sum_{d_T=1}^{D_T} \alpha_{c_1} \beta_{d_1|c_1} \left( \prod_{j=1}^{q} \prod_{k=1}^{r_j} I(w_{1j} = k) \right) \times$$

$$\prod_{t=2}^{T} \left( \epsilon_{c_t|c_{t-1}d_{t-1}}^{(t-1)} \eta_{d_t|c_{t-1}d_{t-1}}^{(t-1)} \prod_{j=1}^{q} \prod_{k=1}^{r_j} I(w_{tj} = k) \right),$$
I. Motivating Examples
Why learn about longitudinal mixture modeling?

Depends on your research question...
Question 1: How do individuals change their drinking behavior following treatment?
Individual drinking patterns \((n = 395)\) following an initial post-treatment lapse.
Question 2: Does the number of cigarettes smoked per day change during telephone tobacco counseling?
Cigarettes per day during telephone smoking cessation calls (n = 11,927)
Question 3: Are there changes in quit status after telephone tobacco counseling?
Quit status 6- and 12-months following telephone tobacco counseling (n = 7,624)
Question 4: How do we examine the effects of weight loss intervention when some individuals in the treatment group are non-responders?
Response to weight-loss intervention

The non-responders in the intervention group lessen the effect size of the intervention in comparison to the control group.
Examples of Longitudinal Mixture Modeling


II. The Basics
Latent Variable

= unobserved; unmeasured; based on relationships between observed variables

observed variable (x1)

E1

observed variable (x2)

E2

observed variable (x3)

E3

= observed; measured variables

= error or “residual,” not explained by shared variance
Latent variables can be continuous or categorical; two representations of the same reality

Continuous latent variable – correlation explained by underlying factor
Ex. structural equation models, factor models, growth curve models, multilevel models

Categorical latent variable – correlation reflects difference between discrete groups on mean levels of observed variables
Ex. latent class analysis, mixture analyses, latent transition analysis, latent profile analysis
Types of models

<table>
<thead>
<tr>
<th>Categorical latent variable</th>
<th>Categorical &amp; continuous latent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-sectional</td>
<td>Latent class analysis</td>
</tr>
<tr>
<td>Latent profile analysis</td>
<td></td>
</tr>
<tr>
<td>Longitudinal</td>
<td>Latent transition analysis</td>
</tr>
<tr>
<td>Latent Markov model</td>
<td></td>
</tr>
<tr>
<td>Latent class growth analysis</td>
<td></td>
</tr>
</tbody>
</table>
Finite Mixture Modeling
Direct and Indirect Applications

1. Direct: identify distinct subpopulations of individuals
   - Obtain estimates of the conditional distributions
   - Calculate posterior probabilities that individual $n$ came from subpopulation $k$

2. Indirect: approximate intractable distribution with a small number of simpler component distributions
Cross-Sectional Latent Mixture Models

latent profile analysis
latent class analysis
Latent profile analysis = categorical latent variable observed continuous variables

“Profiles”

% heavy drinking days

E

Latent profile analysis = categorical latent variable observed continuous variables

Class 1 = infrequent drinking (78%)
Class 2 = occasional drinking (12%)
Class 3 = frequent drinking (10%)

Profiles

% heavy drinking days

% drinking days post-treatment

Most likely profile

Class 1
Class 2
Class 3

0
20
40
60
80
100
Latent class analysis = categorical latent variable observed categorical variables

<table>
<thead>
<tr>
<th>Responses</th>
<th>Joe</th>
<th>Jane</th>
<th>Bob</th>
<th>Betty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any smoking</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cigarettes per day = 0</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cigarettes per day &lt;20</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cigarettes per day  20+</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Contemplation stage</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Preparation stage</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Action stage</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Latent class analysis = categorical latent variable observed categorical variables

Class 1 = non smoking, action stage (10%)
Class 2 = smoking, 12-20 CPD, preparation stage (13%)
Class 3 = smoking, 20+ CPD, preparation stage (77%)

<table>
<thead>
<tr>
<th>Probability of endorsing</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any smoking</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0 CPD</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>12-20 CPD</td>
<td>0.00</td>
<td>1.00</td>
<td>0.34</td>
</tr>
<tr>
<td>20+ CPD</td>
<td>0.00</td>
<td>0.00</td>
<td>0.38</td>
</tr>
<tr>
<td>Contemplation</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Preparation</td>
<td>0.11</td>
<td>0.99</td>
<td>0.93</td>
</tr>
<tr>
<td>Action</td>
<td>0.89</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Longitudinal Latent Mixture Models

Latent Growth Mixture Model
Latent Markov Model
Latent Transition Analysis
Repeated Measures Latent Class Analysis
Latent Growth Curve = longitudinal continuous latent variable

\[ y_{it} = \eta_{oi} + \eta_{it}x_t + \epsilon_{it} \]
Latent growth model
Latent Growth Mixture Model: Combined continuous and categorical latent variable model

- Growth factors (i.e., random effects) are indicators of latent trajectory “classes”
Latent Growth Mixture Model
Latent Class Growth Analysis (i.e., semi-parametric group-based modeling): Combined continuous and categorical latent variable model

CLASS ("mixture")

Intercept

Slope

Post-treatment
Wee k 10
Week 26
Week 52

ε₁
ε₂
ε₃
ε₄

All variances in trajectories explained by classification! With no variation around trajectories…
Latent Class Growth Analysis

The graph shows the percentage of heavy drinking days over weeks following treatment. There are four classes:

- **Class 1, 68.1%**: The percentage remains relatively stable over time.
- **Class 2, 15.1%**: The percentage increases over time, peaking around 26 weeks.
- **Class 3, 7.5%**: The percentage decreases sharply over time, starting high and ending lower.
- **Class 4, 9.3%**: The percentage decreases over time, starting low and ending lower than Class 3.

The graph uses different symbols for each class, making it easy to distinguish between them.
Latent Markov model = estimate probability of transitioning over time
ex. Percent heavy drinking days (%HD)
Class 1 = infrequent drinking (78%)
Class 2 = occasional drinking (12%)
Class 3 = frequent drinking (10%)
Latent Markov model = estimate probability of transitioning over time
ex. Percent heavy drinking days (%HD)
### Latent Markov Model

<table>
<thead>
<tr>
<th>Week 0</th>
<th>Week 10</th>
<th>Week 26</th>
<th>Week 52</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infreq. (n\approx1006)</td>
<td>Occas. (n\approx162)</td>
<td>Freq. (n\approx128)</td>
<td>Infreq. (n\approx831)</td>
</tr>
<tr>
<td>0.84</td>
<td>0.13</td>
<td>0.03</td>
<td>0.90</td>
</tr>
<tr>
<td>0.18</td>
<td>0.58</td>
<td>0.24</td>
<td>0.16</td>
</tr>
<tr>
<td>0.08</td>
<td>0.13</td>
<td>0.79</td>
<td>0.08</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 10</th>
<th>Week 26</th>
<th>Week 52</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infreq. (n\approx887)</td>
<td>Occas. (n\approx238)</td>
<td>Freq. (n\approx171)</td>
</tr>
<tr>
<td>0.85</td>
<td>0.11</td>
<td>0.04</td>
</tr>
<tr>
<td>0.17</td>
<td>0.59</td>
<td>0.24</td>
</tr>
<tr>
<td>0.11</td>
<td>0.13</td>
<td>0.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 26</th>
<th>Week 52</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infreq. (n\approx256)</td>
<td>Freq. (n\approx218)</td>
</tr>
<tr>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>0.16</td>
<td>0.54</td>
</tr>
<tr>
<td>0.08</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Latent transition analysis = estimate probability of transitioning from one class to another over time
Class 1 = non smoking, action stage (10%)
Class 2 = smoking, 12-20 CPD, preparation stage (13%)
Class 3 = smoking, 20+ CPD, preparation stage (77%)

<table>
<thead>
<tr>
<th>Probability of endorsing</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any smoking</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0 CPD</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>12-20 CPD</td>
<td>0.00</td>
<td>1.00</td>
<td>0.34</td>
</tr>
<tr>
<td>20+ CPD</td>
<td>0.00</td>
<td>0.00</td>
<td>0.38</td>
</tr>
<tr>
<td>Contemplation</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Preparation</td>
<td>0.11</td>
<td>0.99</td>
<td>0.93</td>
</tr>
<tr>
<td>Action</td>
<td>0.89</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>
## Latent transition analysis

<table>
<thead>
<tr>
<th>Probability of endorsing</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any smoking</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0 CPD</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>12-20 CPD</td>
<td>0.00</td>
<td>1.00</td>
<td>0.34</td>
</tr>
<tr>
<td>20+ CPD</td>
<td>0.00</td>
<td>0.00</td>
<td>0.38</td>
</tr>
<tr>
<td>Contemplation</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Preparation</td>
<td>0.11</td>
<td>0.99</td>
<td>0.93</td>
</tr>
<tr>
<td>Action</td>
<td>0.89</td>
<td>0.01</td>
<td>0.00</td>
</tr>
</tbody>
</table>

### Call 1

<table>
<thead>
<tr>
<th>C1 (33%)</th>
<th>Call 2</th>
<th>C3 (57%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 (10%)</td>
<td>0.43</td>
<td>0.00</td>
</tr>
<tr>
<td>C2 (13%)</td>
<td>0.39</td>
<td>0.51</td>
</tr>
<tr>
<td>C3 (77%)</td>
<td>0.00</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Repeated measures latent class analysis = estimate % of individuals who follow similar pattern over time
Repeated measures latent class analysis
III. Data and Model Considerations
Prior to running any analyses...

Is your data appropriate?
• Check your distributions
• Sample size and power
• Time points and time scale
Check your distributions

Sample size and power

- Most LMMs are estimated with large samples
- No set guidelines for parameter-to-n ratio
- Monte Carlo study to determine sample size for given model (Muthen & Muthen, (2002) *Structural Equation Modeling*)
Time, it’s on your side...

- # of time points?
  - Growth mixture models - 3 is enough, but not ideal
  - Latent Markov/transition analysis - 2 is enough

- Time scale?
  - Age
  - Calendar time
  - Clinically meaningful scaling
Model selection & estimation

- Class enumeration and model fit evaluation
- Type of mixture model?
- Parameter restrictions
- Starting values
Class Enumeration???
Fit Indices

- Information Criteria – based on loglikelihood and # of parameters
  - Akaike’s Information Criteria (AIC)
  - Bayes Information Criteria (BIC)
  - Adjusted BIC

- Lo-Mendell-Rubin likelihood ratio test (LMR) and bootstrapped likelihood ratio test (BLRT)
  - k-1 vs. k class model
Classification precision

- Model entropy criteria – range from 0 to 1
- Average Latent Class Probabilities for Most Likely Latent Class Membership by Latent Class

Example.

<table>
<thead>
<tr>
<th>Latent class assignment</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most likely latent class</td>
<td>1</td>
<td>0.875</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.060</td>
<td>0.913</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.021</td>
<td>0.051</td>
</tr>
</tbody>
</table>
Assessing Model Fit in Practice

Example: Drinking trajectories following an initial lapse
- Type: Growth mixture model
- Class enumeration: 1-5 classes estimated
  - Sample-size adjusted BIC (aBIC) prefer lower values
  - Bootstrapped likelihood ratio test (k vs. k-1)
Example 1: Observed data…

(n = 395; % days abstinent per month, during the months following an initial post-treatment lapse)
Two Class Growth Mixture Model % Days Abstinent

Entropy = .95

aBIC = 30510.36

How does it compare to a one class Model?

aBIC = 30707.75
BLRT = 200.24 (p < .0005)

frequent drinkers, 24.3%
light drinkers, 75.7%
Three Class Growth Mixture Model: % Days Abstinent

Entropy = .91
aBIC = 30458.31
2-Class Model
aBIC = 30510.36
BLRT = 70.23 (p = .20)
Four Class Growth Mixture Model: % Days Abstinent

Entropy = .92
aBIC = 30316.96

3-Class Model
aBIC = 30458.31
BLRT = 147.64 (p < .0005)

BLRT of 5 class model = 25.76 (p = .85)
Type of mixture model???
Example 2: Observed data…

(n = 1,383; % heavy drinking days post-treatment)
Comparison of latent growth mixture and Markov models

- **Growth mixture:** 3-class model based on LMR
  - Entropy = 0.98
  - BIC (3-class GMM) = 43886.02

- **Markov model:** 3-class model based on LMR
  - Entropy = 0.95
  - BIC (3-class LMM) = 41045.22

Witkiewitz, Maisto, & Donovan (under review)
Latent Growth Mixture Model

![Graph showing percentage of heavy drinking days over weeks following treatment for three classes: Class 1 (12.2%), Class 2 (77.9%), and Class 3 (9.9%). Each class has a distinct line on the graph, with different symbols and colors to represent data points at various weeks (0, 10, 26, 52).]
Growth mixture model: “Infrequent heavy drinking class”

Weeks following treatment
Growth mixture model: “Occasional heavy drinking class”
Growth mixture model: “Frequent heavy drinking class”
Latent Markov model: “Infrequent heavy drinking class pattern”
Latent Markov model: “Infrequent to occasional heavy drinking class pattern”
Latent Markov model: “Frequent heavy drinking class pattern”
Cross-classification of models

Most likely class membership in latent growth mixture model:
- Occasional heavy class
- Infrequent heavy class
- Frequent heavy class

Infrequent drinking pattern based on Markov model

Percent of individuals in growth mixture classes

- No
- Yes
A few more sticky issues

- Parameter restrictions
  - Equality constraints
  - Setting parameters to specific values (0, 1, known value)

- Starting values
  - Local solutions - erratic likelihood functions
  - Substantively different results
  - Multiple random starting values can help examine likelihood space (Hipp & Bauer, 2006)
IV. Incorporating Covariates
Common research questions

Do different levels of x correspond to a greater likelihood of being in class y?

Motivation

What is the relationship between x and within class change over time?

Motivation

Does class membership predict outcome y?

3 year drinking

CLASS

Intercept

Slope

Time 1

ε₁

Time 2

ε₂

Time 3

ε₃

Time 4

ε₄

69
Growth Mixture Model with Covariates

- Coping
- Cognition
- Affect
- Distal Risk

"Class"

Intercept

Slope

Simplified Growth Mixture Mode
PDA = % days abstinent
Mo = follow-up month
Growth Mixture Model with Covariates predicting class membership

<table>
<thead>
<tr>
<th>Drinking Class</th>
<th>Coping</th>
<th>Affect</th>
<th>Distal risk</th>
<th>Cognitive risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infrequent</td>
<td>1.84*</td>
<td>-.03</td>
<td>-.05*</td>
<td>.06*</td>
</tr>
<tr>
<td>Occasional</td>
<td>1.45*</td>
<td>.005</td>
<td>-.01</td>
<td>.02</td>
</tr>
<tr>
<td>Prolapsers</td>
<td>1.24*</td>
<td>.06</td>
<td>-.08*</td>
<td>.05*</td>
</tr>
<tr>
<td>Frequent</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>**</td>
</tr>
</tbody>
</table>

Frequent drinkers vs. other classes

- ✓ significantly lower coping scores vs. all classes
- ✓ significantly higher distal risk vs. light drinkers & prolapsers
- ✓ significantly lower cognitive risk vs. light drinkers & prolapsers

* = (p<.05)
Growth Mixture Model with covariates predicting within class growth
Covariates predicting within class growth

<table>
<thead>
<tr>
<th>Covariate</th>
<th>PDD intercept $B$ (SE)</th>
<th>PDD slope $B$ (SE)</th>
<th>DDD intercept $B$ (SE)</th>
<th>DDD slope $B$ (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>-0.04 (0.04)</td>
<td>-0.01 (0.01)</td>
<td>-2.94 (1.13)*</td>
<td>-0.08 (0.26)</td>
</tr>
<tr>
<td>Family history</td>
<td>0.002 (0.003)</td>
<td>-0.001 (0.001)</td>
<td>0.004 (0.05)</td>
<td>-0.02 (0.01)</td>
</tr>
<tr>
<td>Years of drinking</td>
<td>0.00 (0.002)</td>
<td>0.00 (0.001)</td>
<td>-0.002 (0.07)</td>
<td>0.00 (0.02)</td>
</tr>
<tr>
<td>ADS total scores</td>
<td>0.003 (0.003)</td>
<td>0.001 (0.001)</td>
<td>0.22 (0.08)*</td>
<td>0.03 (0.02)*</td>
</tr>
<tr>
<td>Month of lapse</td>
<td>0.02 (0.01)*</td>
<td>-0.002 (0.003)</td>
<td>-0.01 (0.22)</td>
<td>0.07 (0.08)</td>
</tr>
<tr>
<td>Coping</td>
<td>-0.09 (0.02)*</td>
<td>-0.004 (0.01)</td>
<td>0.41 (0.58)</td>
<td>-0.28 (0.16)*</td>
</tr>
</tbody>
</table>

*Note. PDD = percentage of drinking days; DDD = drinks per drinking day.  
*p < .05.

Witkiewitz & Masyn (2008). *Psych Addictive Behaviors*
“Predicting” a distal outcome

Does class membership predict outcome $y$?

Latent class

Any smoking call 1

Any smoking call 2

Any smoking call 3

Any smoking call 4

Any smoking call 5

12-month point prevalence
Predict 12-month point prevalence

<table>
<thead>
<tr>
<th>Equality tests of mean 12-month point prevalence across classes using posterior probability-based multiple imputations</th>
<th>$\chi^2$, p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1 (mean = .23) vs. Class 2 (mean = .46)</td>
<td>72.54, p&lt;0.001</td>
</tr>
<tr>
<td>Class 1 vs. Class 3 (mean = .52)</td>
<td>83.73, p&lt;0.001</td>
</tr>
<tr>
<td>Class 2 vs. Class 3</td>
<td>2.24, p = 0.34</td>
</tr>
</tbody>
</table>
Latent transition analysis with time-invariant covariate predicting classes and transitions between classes
Latent transition analysis with time-invariant covariate predicting classes and transitions between classes

<table>
<thead>
<tr>
<th>Class membership</th>
<th>Posttreatment</th>
<th>Month 6</th>
<th>Month 12</th>
<th>Posttreatment</th>
<th>Month 6</th>
<th>Month 12</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>Non. vs heavy&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.87 (0.78-0.96)*</td>
<td>1.01 (0.90-1.13)</td>
<td>0.87 (0.93-1.02)</td>
<td>0.94 (0.79-1.11)</td>
<td>0.81 (0.70-0.94)*</td>
<td>0.97 (0.86-1.10)</td>
</tr>
<tr>
<td>Mod. vs heavy&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.64 (0.54-0.77)*</td>
<td>0.78 (0.65-0.92)*</td>
<td>0.70 (0.54-0.91)*</td>
<td>0.82 (0.60-1.13)</td>
<td>0.71 (0.59-0.85)*</td>
<td>0.85 (0.69-1.05)</td>
</tr>
<tr>
<td>Mod. vs non.&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.74 (0.63-0.87)*</td>
<td>0.77 (0.66-0.88)*</td>
<td>0.81 (0.68-0.96)*</td>
<td>0.88 (0.72-1.08)</td>
<td>0.87 (0.75-1.01)</td>
<td>0.88 (0.73-1.06)</td>
</tr>
</tbody>
</table>

Notes: Heavy = heavy frequent drinking class; mod. = moderate infrequent drinking class; non. = nondrinking class. <sup>a</sup>Reference class in multinomial logistic regression, p < .05. *p < .05.
V. Limitations and Criticisms
Limitations: Longitudinal Mixture Models

• User specified number of latent classes
  • Classes are not real and true # of classes never known
• All sorts of modeling issues
  • See Bauer (2007). Multivariate Behav Research
  Sampson & Laub (2005). Criminology
• Adding covariates
  • Non-normality and nonlinearity can result in spurious class-by-predictor relationships
  • Can change underlying latent classes
Summary of modeling issues

- Over-extraction of latent classes (Bauer & Curran, 2003)
- Reification of groups that do not exist (Raudenbush, 2005)
- Loss of power - artificially dividing growth trajectories into classes
- Spurious covariate by class interactions that do not exist (Bauer & Curran, 2003)
Bauer and Curran (2003):

“All of our models are wrong, and it is quite possible that there is no “right” model to discern whatsoever. The real task at hand is to decide which model is most useful” (p. 388).
VI. Model Estimation
Software

- Mplus – [www.statmodel.com](http://www.statmodel.com)
- Mx - [http://www.vcu.edu/mx/](http://www.vcu.edu/mx/)
- SAS ProcTraj - [http://www.andrew.cmu.edu/user/bjones/](http://www.andrew.cmu.edu/user/bjones/)
- Open Mx - [http://openmx.psyc.virginia.edu/](http://openmx.psyc.virginia.edu/)
- Others?
Mplus Version 6
Coming Soon

Latest News

Mplus Version 6 will be available the week of April 26. Preorders will be taken starting Monday, April 12.

The major new feature in Mplus Version 6 is Bayesian analysis using MCMC. This includes multiple imputation for missing data as well as plausible values for latent variables. Other additions include replicate weights for complex survey data, convenience features for modeling with missing data, and survival analysis models and plots.

Upcoming Mplus Short Courses

Mplus Short Courses, IDEC: Universitat Pompeu Fabra, Barcelona, Spain, May 31–June 1, 2010

Two short courses that cover growth modeling, survival analysis, and missing data analysis will be given at IDEC-Universitat Pompeu Fabra, Barcelona, Spain, May 31 and June 1, 2010. Those are Topics 3 and 4 of the 8-topic Mplus course sequence. Click here for more information and to sign up for the full two-day course. Click here to sign up for the first day only. Click here to sign up for the
TITLE:
Latent growth mixture model
DATA:
FILE IS lgm.dat;

VARIABLE:
  NAMES ARE t16_phd t26_phd t52_phd t69_phd;
  USEVAR ARE t16_phd t26_phd t52_phd t69_phd;
  CLASSES ARE c(3);
  MISSING ARE ALL (999);

ANALYSIS:
  TYPE = MIXTURE;
  STARTS = 400 100; !setting random starting values
  PROCESSORS = 4; !this command divides estimation across processors

MODEL:
  %OVERALL%
i q | T16_phd@-1 T26_phd@0 T52_phd@2.6 T68_phd@4.2;
i s*; !variances of intercept and slope set to be class invariant
i with s*; !covariance between intercept and slope set to be class invariant
q@0; !setting variance of q at zero because EM results in -variance estimates

  %C#1%
  [i s* q*];
  %C#2%
  [i s* q*];
  %C#3%
  [i s* q*];

OUTPUT: TECH1 TECH7 TECH10 TECH11 STANDARDIZED;
  !TECH1=parameter specifications and starting values for free parameters
  !TECH7=sample statistics for each class using raw data weighted by
  !the estimated posterior probabilities for each class
  !TECH10=univariate, bivariate, and response patter for categorical DVs
  !TECH11=Lo-Mendel-Rubin likelihood ratio test
  !STANDARDIZED = standardized values

PLOT:
  TYPE IS plot3;
  SERIES IS t4_phd-t68_phd(s);
WARNING: WHEN ESTIMATING A MODEL WITH MORE THAN TWO CLASSES, IT MAY BE NECESSARY TO INCREASE THE NUMBER OF RANDOM STARTS USING THE STARTS OPTION TO AVOID LOCAL MAXIMA.

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT
Loglikelihood
   H0 Value         -21907.093
   H0 Scaling Correction Factor  1.949
   for MLR
Information Criteria
   Number of Free Parameters  18
   Akaike (AIC)              43850.186
   Bayesian (BIC)            43943.193
   Sample-Size Adjusted BIC  43886.015
   \( n^* = (n + 2) / 24 \) 
### FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES BASED ON THE ESTIMATED MODEL

<table>
<thead>
<tr>
<th>Latent Classes</th>
<th>Count</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>158.22360</td>
<td>0.12209</td>
</tr>
<tr>
<td>2</td>
<td>1009.38814</td>
<td>0.77885</td>
</tr>
<tr>
<td>3</td>
<td>128.38827</td>
<td>0.09907</td>
</tr>
</tbody>
</table>

### FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASS PATTERNS BASED ON ESTIMATED POSTERIOR PROBABILITIES

<table>
<thead>
<tr>
<th>Latent Classes</th>
<th>Count</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>158.22360</td>
<td>0.12209</td>
</tr>
<tr>
<td>2</td>
<td>1009.38813</td>
<td>0.77885</td>
</tr>
<tr>
<td>3</td>
<td>128.38827</td>
<td>0.09907</td>
</tr>
</tbody>
</table>

### CLASSIFICATION QUALITY

<table>
<thead>
<tr>
<th>Entropy</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>0.979</td>
</tr>
</tbody>
</table>
CLASSIFICATION OF INDIVIDUALS BASED ON THEIR MOST LIKELY LATENT CLASS MEMBERSHIP

Class Counts and Proportions

Latent Classes

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>156</td>
<td>1011</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>0.12037</td>
<td>0.78009</td>
<td>0.09954</td>
</tr>
</tbody>
</table>

Average Latent Class Probabilities for Most Likely Latent Class Membership (Row) by Latent Class (Column)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.970</td>
<td>0.027</td>
<td>0.002</td>
</tr>
<tr>
<td>2</td>
<td>0.006</td>
<td>0.994</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>0.008</td>
<td>0.000</td>
<td>0.992</td>
</tr>
</tbody>
</table>
**MODEL RESULTS**

<table>
<thead>
<tr>
<th>Latent Class 1</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>Two-Tailed P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T26_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T52_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T68_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4_PHD</td>
<td>-1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T26_PHD</td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T52_PHD</td>
<td>2.600</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T68_PHD</td>
<td>4.200</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T26_PHD</td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T52_PHD</td>
<td>6.760</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T68_PHD</td>
<td>17.640</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>Latent Class 1</td>
<td>Estimate</td>
<td>S.E.</td>
<td>Est./S.E.</td>
<td>P-Value</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>S WITH</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>32.714</td>
<td>2.262</td>
<td>14.460</td>
<td>0.000</td>
</tr>
<tr>
<td>Means</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>46.136</td>
<td>2.268</td>
<td>20.339</td>
<td>0.000</td>
</tr>
<tr>
<td>S</td>
<td>3.534</td>
<td>1.426</td>
<td>2.479</td>
<td>0.013</td>
</tr>
<tr>
<td>Q</td>
<td>-0.718</td>
<td>0.310</td>
<td>-2.320</td>
<td>0.020</td>
</tr>
<tr>
<td>Intercepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4_PHD</td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T26_PHD</td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T52_PHD</td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T68_PHD</td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>86.389</td>
<td>6.205</td>
<td>13.923</td>
<td>0.000</td>
</tr>
<tr>
<td>S</td>
<td>26.446</td>
<td>1.933</td>
<td>13.680</td>
<td>0.000</td>
</tr>
<tr>
<td>Q</td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>Residual Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4_PHD</td>
<td>10.086</td>
<td>4.484</td>
<td>2.249</td>
<td>0.024</td>
</tr>
<tr>
<td>T26_PHD</td>
<td>342.876</td>
<td>25.345</td>
<td>13.528</td>
<td>0.000</td>
</tr>
<tr>
<td>T52_PHD</td>
<td>310.972</td>
<td>21.594</td>
<td>14.401</td>
<td>0.000</td>
</tr>
<tr>
<td>T68_PHD</td>
<td>123.691</td>
<td>27.015</td>
<td>4.579</td>
<td>0.000</td>
</tr>
<tr>
<td>Latent Class 2</td>
<td>Estimate</td>
<td>S.E.</td>
<td>Est./S.E.</td>
<td>P-Value</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T26_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T52_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T68_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4_PHD</td>
<td>-1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T26_PHD</td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T52_PHD</td>
<td>2.600</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T68_PHD</td>
<td>4.200</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4_PHD</td>
<td>1.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T26_PHD</td>
<td>0.000</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T52_PHD</td>
<td>6.760</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>T68_PHD</td>
<td>17.640</td>
<td>0.000</td>
<td>999.000</td>
<td>999.000</td>
</tr>
<tr>
<td>S WITH I</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>32.714</td>
<td>2.262</td>
<td>14.460</td>
<td>0.000</td>
</tr>
<tr>
<td>Means</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>9.608</td>
<td>0.554</td>
<td>17.328</td>
<td>0.000</td>
</tr>
<tr>
<td>S</td>
<td>5.368</td>
<td>0.310</td>
<td>17.313</td>
<td>0.000</td>
</tr>
<tr>
<td>Q</td>
<td>-0.825</td>
<td>0.078</td>
<td>-10.584</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Random Starts Specifications for the k-1 Class Analysis Model

Number of initial stage random starts 1000
Number of final stage optimizations 200

Vuong-Lo-Mendell-Rubin Likelihood Ratio Test for 2 (H0) Versus 3 Classes

H0 Loglikelihood Value -22251.816
2 Times the Loglikelihood Difference 689.446
Difference in the Number of Parameters 4
Mean 3.678
Standard Deviation 162.623
P-Value 0.0005

Lo-Mendell-Rubin Adjusted LRT Test
Value 666.207
P-Value 0.0006
TITLE:
Latent Markov Model 3-classes

DATA:
FILE IS LMM.dat;

VARIABLE:
NAMES ARE t16_phd t26_phd t52_phd t69_phd;
USEVAR ARE t16_phd t26_phd t52_phd t69_phd;
CLASSES ARE c16(3) c26(3) c52(3) c68(3);
MISSING ARE ALL (999);

ANALYSIS:
TYPE = MIXTURE;
STARTS = 400 100; !setting random starting values
LRTSTARTS = 5 5 50 20; !more random starts for LRT estimation
ALGORITHM = INTEGRATION; !numerical integration is necessary for many models
PROCESSORS = 4; !this command divides estimation across processors

MODEL:
%OVERALL%
c68#1 c68#2 on c52#1 c52#2; !latent transition probabilities
c52#1 c52#2 on c26#1 c26#2;
c26#1 c26#2 on c16#1 c16#2;

MODEL c16:
%C16#1% !classes defined by mean values of phd
[t16_PHD*](1); !(#) places equality constraints so classes are invariant
%C16#2%
[t16_PHD*](2);
%C16#3%
[t16_PHD*](3);

MODEL c26:
%C26#1%
[T26_PHD*](1);
%C26#2%
[T26_PHD*](2);
%C26#3%
[T26_PHD*](3);
Latent Markov model continued

%LET PHI = (1);
%LET PHI = (2);
%LET PHI = (3);

MODEL c52:
%LET PHI = (1);
%LET PHI = (2);
%LET PHI = (3);

MODEL c68:
%LET PHI = (1);
%LET PHI = (2);
%LET PHI = (3);

OUTPUT: TECH1 TECH7 TECH10 sampstat standardized;

PLOT:
TYPE IS plot3;
SERIES IS t16_phd(0) t26_phd(1) t52_phd(2) t68_phd(3) ;

SAVEDATA:
FILE is LMHC3.sav;
FORMAT is FREE;
SAVE = cprobabilities;
LATENT TRANSITION PROBABILITIES BASED ON THE ESTIMATED MODEL

<table>
<thead>
<tr>
<th></th>
<th>C16 Classes (Rows) by C26 Classes (Columns)</th>
<th></th>
<th>C26 Classes (Rows) by C52 Classes (Columns)</th>
<th></th>
<th>C52 Classes (Rows) by C68 Classes (Columns)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1               2          3</td>
<td>1               2          3</td>
<td>1               2          3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1               0.578    0.242    0.180</td>
<td>1               0.587    0.238    0.174</td>
<td>1               0.537    0.162    0.301</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2               0.135    0.788    0.077</td>
<td>2               0.128    0.764    0.108</td>
<td>2               0.154    0.765    0.081</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3               0.131    0.029    0.840</td>
<td>3               0.113    0.036    0.851</td>
<td>3               0.086    0.017    0.897</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TITLE:
Latent Transition Analysis 3-class model

DATA:
File is lta.dat;

VARIABLE:
NAMES ARE  t16_fdro t26_fdro t52_fdro t68_fdro t16_dro t26_dro t52_dro t68_dro;
USEVAR ARE t16_fdro t26_fdro t52_fdro t68_fdro t16_dro t26_dro t52_dro t68_dro;
CLASSES ARE c16(3) c26(3) c52(3) c68(3);
MISSING ARE ALL (999);

CATEGORICAL ARE
t16_fdro t26_fdro t52_fdro t68_fdro t16_dro t26_dro t52_dro t68_dro;
!fdro and dro are categorical variables with 3 categories each

ANALYSIS:
TYPE = MIXTURE;
STARTS = 400 100; !setting random starting values
PROCESSORS = 4; !this command divides estimation across processors

MODEL:
%OVERALL%
c68#1 c68#2 on c52#1 c52#2; !latent transition probabilities
c52#1 c52#2 on c26#1 c26#2;
c26#1 c26#2 on c16#1 c16#2;
c16#1;
%c16#1%
[t16_fdro$1](1); !thresholds of fdro and dro are class-varying
[t16_fdro$2](2); !(#) places equality constraints so classes are invariant
[t16_dro$1](3);
[t16_dro$2](4);
c16#2;
%c16#2%
[t16_fdro$1](10); !thresholds of fdro and dro are class-varying
[t16_fdro$2](20); !(#) places equality constraints so classes are invariant
[t16_dro$1](30);
[t16_dro$2](40);
c16#3;
%c16#3%
[t16_fdro$1](100); !thresholds of fdro and dro are class-varying
[t16_fdro$2](200); !(#) places equality constraints so classes are invariant
[t16_dro$1](300);
Latent transition analysis continued
Latent transition analysis continued

OUTPUT:  TECH1 TECH7 TECH10 TECH11 SAMPSTAT STANDARDIZED;
!TECH1=parameter specifications and starting values for free parameters
!TECH7=sample statistics for each class using raw data weighted by
!the estimated posterior probabilities for each class
!TECH10=univariate, bivariate, and response patter for categorical DVs
!TECH11=Lo-Mendel-Rubin likelihood ratio test
!SAMPSTAT = sample statistics
!STANDARDIZED = standardized values

PLOT:
TYPE IS plot3;
SERIES IS t16_fdro(0) t26_fdro(1) t52_fdro(2) t68_fdro(3) |
t16_dro(0) t26_dro(1) t52_dro(2) t68_dro(3) ;
TITLE:
Repeated Measures Latent Class Analysis 3-class model

DATA:
File is rmlca.dat;

VARIABLE:
NAMES ARE  t4_fdro t26_fdro t52_fdro t68_fdro t4_dro t26_dro t52_dro t68_dro;
USEVAR ARE t4_fdro t26_fdro t52_fdro t68_fdro t4_dro t26_dro t52_dro t68_dro;
CLASSES ARE  C(3) ;
MISSING ARE ALL (999);

CATEGORICAL ARE
t4_fdro t26_fdro t52_fdro t68_fdro t4_dro t26_dro t52_dro t68_dro;
!fdro and dro are categorical variables with 3 categories each

ANALYSIS:
TYPE = MIXTURE;
STARTS = 400 100;  !setting random starting values
LRTSTARTS = 5 5 50 20;  !more random starts for LRT estimation
ALGORITHM = INTEGRATION;  !numerical integration is necessary for many models
PROCESSORS = 4;  !this command divides estimation across processors

MODEL:
%OVERALL%
%C#1%
[t4_fdro$1];  !thresholds of fdro and dro are class-varying
[t4_fdro$2];
[t4_dro$1];
[t4_dro$2];
[t26_fdro$1];
[t26_fdro$2];
[t26_dro$1];
[t26_dro$2];
[t52_fdro$1];
[t52_fdro$2];
[t52_dro$1];
[t52_dro$2];
[t68_fdro$1];
[t68_fdro$2];
Repeated measures latent class analysis continued
## Covariates in Mplus

**Do different levels of x correspond to a greater likelihood of being in class y?**

**What is the relationship between x and change over time across classes?**

**What is the relationship between x and change over time within classes?**

**Does class membership predict outcome?**

<table>
<thead>
<tr>
<th>C#1 C#2 on motiv;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ictpt slope on tx;</td>
</tr>
<tr>
<td>%C#1% ictpt slope on tx;</td>
</tr>
<tr>
<td>%C#2% ictpt slope on tx;</td>
</tr>
<tr>
<td>%C#1% [drink_3];</td>
</tr>
<tr>
<td>%C#2% [drink_3];</td>
</tr>
<tr>
<td>OR</td>
</tr>
<tr>
<td>AUXILIARY = drink_3(e);</td>
</tr>
</tbody>
</table>
Common Mplus error messages

**Error**
WARNING: WHEN ESTIMATING A MODEL WITH MORE THAN TWO CLASSES, IT MAY BE NECESSARY TO INCREASE THE NUMBER OF RANDOM STARTS USING THE STARTS OPTION TO AVOID LOCAL MAXIMA.

**Solution**
Make sure you are using STARTS =
*Note, you will still get this warning with STARTS command

**Error**
WARNING: THE BEST LOGLIKELIHOOD VALUE WAS NOT REPLICATED. THE SOLUTION MAY NOT BE TRUSTWORTHY DUE TO LOCAL MAXIMA. INCREASE THE NUMBER OF RANDOM STARTS.

**Solution**
Increase the number of random starts.
Common Mplus error messages

Error
WARNING: THE LATENT VARIABLE COVARIANCE MATRIX (PSI) IN CLASS 1 IS NOT POSITIVE DEFINITE. THIS COULD INDICATE A NEGATIVE VARIANCE/RESIDUAL VARIANCE FOR A LATENT VARIABLE, A CORRELATION GREATER OR EQUAL TO ONE BETWEEN TWO LATENT VARIABLES, OR A LINEAR DEPENDENCY AMONG MORE THAN TWO LATENT VARIABLES. CHECK THE TECH4 OUTPUT FOR MORE INFORMATION. PROBLEM INVOLVING VARIABLE S.

Solution
Check the residual variances under Model Results

<table>
<thead>
<tr>
<th>Variances</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Est./S.E.</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>7.326</td>
<td>3.033</td>
<td>2.416</td>
<td>0.016</td>
</tr>
<tr>
<td>S</td>
<td>-0.709</td>
<td>0.510</td>
<td>-1.392</td>
<td>0.164</td>
</tr>
</tbody>
</table>
Common Mplus error messages

Error(s)

THE STANDARD ERRORS OF THE MODEL PARAMETER ESTIMATES MAY NOT BE TRUSTWORTHY FOR SOME PARAMETERS DUE TO A NON-POSITIVE DEFINITE FIRST-ORDER DERIVATIVE PRODUCT MATRIX. THIS MAY BE DUE TO THE STARTING VALUES BUT MAY ALSO BE AN INDICATION OF MODEL NONIDENTIFICATION. THE CONDITION NUMBER IS 0.186D-16. PROBLEM INVOLVING PARAMETER 18.

ONE OR MORE MULTINOMIAL LOGIT PARAMETERS WERE FIXED TO AVOID SINGULARITY OF THE INFORMATION MATRIX. THE SINGULARITY IS MOST LIKELY BECAUSE THE MODEL IS NOT IDENTIFIED, OR BECAUSE OF EMPTY CELLS IN THE JOINT DISTRIBUTION OF THE CATEGORICAL LATENT VARIABLES AND ANY INDEPENDENT VARIABLES. THE FOLLOWING PARAMETERS WERE FIXED: 23

Potential Solutions

- Look at TECH1 output to determine what parameter is causing the problem
- Increase number of random starts
- Reduce complexity of model via parameter constraints
VII. Advanced Topics
Advanced longitudinal mixture models

- Discrete time survival mixture analysis
- Multilevel mixture modeling
- Associative latent transition analysis
- Markov switching models
- Alternative model specifications
  - Two-part
  - Piecewise
  - Zero-inflated Poisson
BASIC SETUP

- Programs can be simple (taking advantage of default settings) or complex (by making specific model statements)
- .dat data files
- .inp input files
- .out output files
- .gph graphic files
COMMAND LINES

TITLE:
DATA:
VARIABLE:
DEFINE:
ANALYSIS:
MODEL:
PLOT:
OUTPUT:
SAVEDATA:

- All lines end with ;
- Only 80 characters per line
- "-" can shorten variable lists (y1-y6)

!comments can be used anywhere;
EXAMPLE PROGRAM

**TITLE:** Drinking example 1: 2 classes - only linear slope - No covariates
**DATA:** FILE IS lgmm.example.dat;
**VARIABLE:**
NAMES ARE pda_gm1 pda_gm2 pda_gm3 pda_gm4 pda_gm5;
USEVAR ARE pda_gm1-pda_gm5;
CLASSES = c(2);
**ANALYSIS:** TYPE = MIXTURE;
STARTS = 500 250;
**MODEL:**
%OVERALL%
i s | pda_gm1@0 pda_gm2@1 pda_gm3@2 pda_gm4@3
   pda_gm5@4;
i* s*;
i with s*;
%C#1%
[i* s*];

%C#2%
[i* s*];

**PLOT:** TYPE IS plot3;
SERIES IS pda_gm1(0) pda_gm2(1) pda_gm3(2) pda_gm4(3)
   pda_gm5(4);
**OUTPUT:** standardized sampstat tech11 tech14;
**VARIABLE COMMAND**

<table>
<thead>
<tr>
<th>VARIABLE COMMAND</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAMES ARE</td>
<td>names of variables in the data set;</td>
</tr>
<tr>
<td>USEOBSERVATIONS ARE</td>
<td>conditional statement to select observations;</td>
</tr>
<tr>
<td>USEVARIABLES ARE</td>
<td>names of analysis variables;</td>
</tr>
<tr>
<td>MISSING ARE</td>
<td>variable (#);</td>
</tr>
<tr>
<td>GROUPING IS</td>
<td>name of grouping variable (labels);</td>
</tr>
<tr>
<td>IDVARIABLE IS</td>
<td>name of ID variable;</td>
</tr>
<tr>
<td>KNOWNCLASS =</td>
<td>name of categorical latent variable with known class membership (labels);</td>
</tr>
<tr>
<td>CLASSES =</td>
<td>names of categorical latent variables (number of latent classes);</td>
</tr>
</tbody>
</table>
ANALYSIS COMMAND

- Many options

**TYPE IS** mixture, meanstructure, complex, twolevel

**ESTIMATOR** = ML, GLS, WLS, MLR, WLSMV

- Each question may require a different analysis type and estimator
MODEL COMMAND

PART OF THE MODEL
%OVERALL% - indicates the overall model
%C#1% - indicates class #1 model
%C#2% - indicates class #2 model

DEFINING GROWTH FACTORS
i s | y1@0 y2@1 y3@2 y4@3
i = intercept
s = slope
| = names and defines growth factors

GROWTH FACTOR MEANS
[i* s*] [means] i* = intercept mean is estimated
[i@0 s@1] i@0 = intercept mean is fixed at 0

GROWTH FACTOR VARIANCES
i* s* no brackets=variances, i* = intercept variance is estimated
i@0 s@1 no brackets=variances, i@0 = intercept variance is fixed at 0

COVARIANCES
i WITH s WITH=covariance of i and s

REGRESSION
i ON s ON=regression of i on s
OUTPUT COMMAND

- Many options for output, including: confidence intervals, standard errors, modification indices, standardized coefficients, residuals, “tech1-tech15”
- Tech11 = Lo-Mendell-Rubin Likelihood ratio test
- Tech13 = tests for assessing multivariate skew and kurtosis
- Tech14 = bootstrapped likelihood ratio test