Introduction to Growth Curve Modeling: An Overview and Recommendations for Practice

Patrick J. Curran & Daniel J. Bauer
University of North Carolina at Chapel Hill

Goals for the Morning

- Brief review of traditional methods for analyzing change
- Conceptualizing a growth curve
- Estimating individual trajectories using multilevel modeling
  - Including time-invariant covariates to predict trajectories
- Estimating individual trajectories using structural equation model
  - Including time-invariant covariates to predict
- Examples incorporated throughout based on repeated measures of antisocial behavior in children
- Conclude with model extensions recommendations for practice

- Overall goal is a general overview of many options for analyzing repeated measures data -- by necessity, many details are omitted

Section 1

Defining a Latent Growth Curve

Objectives

- Briefly review traditional models for panel data
- Discuss pros and cons of traditional methods
- Introduce general concept of a latent growth curve
- Describe a growth curve for a single individual
- Describe a set of growth curves for multiple individuals

Repeated Measures ANOVA/MANOVA

- Repeated measures ANOVA and MANOVA were developed to address the fully balanced longitudinal design.
  - N participants are recruited.
  - All participants are assessed on T occasions.
  - The assessment schedule is identical for all participants.
  - There is no missing data.
- Fully balanced designs more common for true experiments, and less common for observational-type studies
- Time is a within-subjects factor (nominal predictor, classification variable) with T levels.
- Compares time-specific mean values, but imposes highly restrictive structure on residual variances and covariances

Assumed Residual Covariance Matrix

ANOVA:

$$\begin{pmatrix} r_{1i} \\ r_{2i} \\ r_{3i} \\ r_{4i} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 + \sigma_1 & & & \\ \sigma_1 & \sigma^2 + \sigma_1 & & \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 & \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix} \right)$$

MANOVA:

$$\begin{pmatrix} r_{1i} \\ r_{2i} \\ r_{3i} \\ r_{4i} \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & & & \\ \sigma_{21} & \sigma_2^2 & & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix} \right)$$

Statistical Limitations of ANOVA/MANOVA

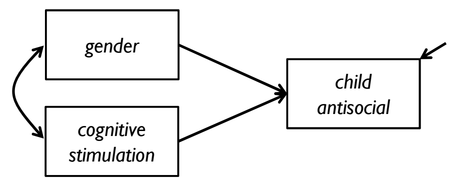
- Require that the assessment schedule be identical for all participants.
- Require that there be no missing data.
  - Listwise deletion of cases with missing data results in potentially low power and requires MCAR assumption.
- RM-ANOVA makes unrealistic assumptions about the residual covariance matrix (compound symmetry).
  - However, can use Greenhouse-Geisser or Huynh-Feldt corrections
- RM-MANOVA makes no assumptions on residual covariance matrix, but this reduces power.

Conceptual Limitations of ANOVA/MANOVA

- Focus on changes in means of repeated measures over time
- Many substantive theories are concerned with individual change, not simply mean change.
  - particularly inter-individual differences in intra-individual change
- Individual change often thought to follow continuous time trajectory that is not captured by treating *time* as nominal predictor
- ANOVA / MANOVA don't correspond well with many contemporary theories of individual stability and change
- In part because of these limitations, regression-based models were developed

Cross-sectional Regression Model

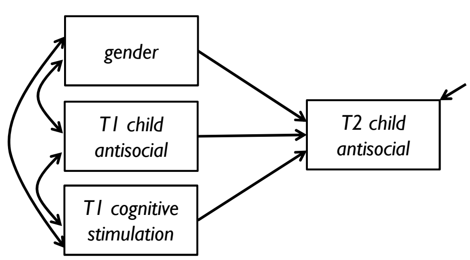
- Consider 2-predictor cross-sectional regression where antisocial behavior is DV & child gender and cognitive stimulation are IVs



- Limited in that cannot establish temporal precedence

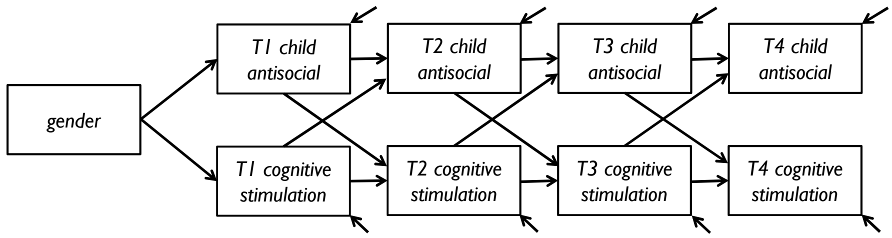
Two Time Point Regression Model

- Regression model can be extended to include two time points
- The T2 assessment is the dependent variable, and the T1 assessment of the same variable is an additional predictor
  - sometimes called *residualized change* model



Autoregressive Path Analytic Model

- Can expand to more than two assessments



- Primarily captures time-adjacent relations among set of RMs
- Does not allow for estimation of continuous trajectory of change
- Often disjoint between theoretical model and statistical model

Advantages and Disadvantages

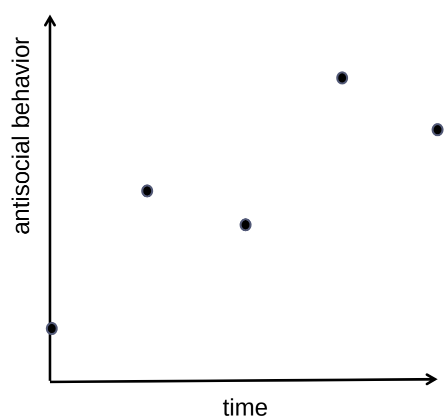
- Many advantages of traditional approaches
  - straightforward to estimate as standard path models
  - obtain overall tests of fit and tests direct and indirect effects
  - corresponds to some models of change posited by some theories
- But also many disadvantages
  - only considers change in a series of two time-point comparisons
  - assumes all measures have perfect reliability
  - can be biased if systematic growth exists in repeated measures
  - poorly corresponds to most theoretical models of change
- We must consider a very different approach to these data

Conceptualizing a Growth Curve

- To capture *continuous trajectory of change*, will approach precisely same data structure from different perspective
- Will build model for data that estimates change over time within each individual and then compare change across individuals
  - estimate *inter-individual* variability in *intra-individual* change
- This is core concept behind a growth curve
  - also sometimes called latent trajectories, latent curves, growth trajectories, or time paths

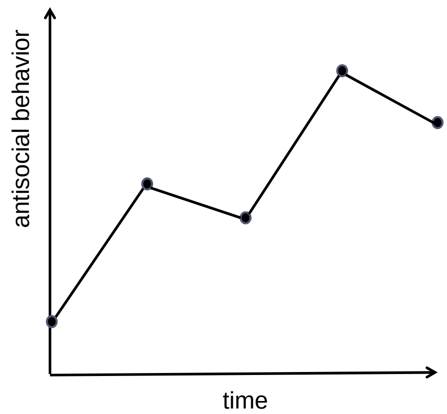
Repeated Measures for One Person

- Consider hypothetical case where we have five repeated measures assessing antisocial behavior for a single child



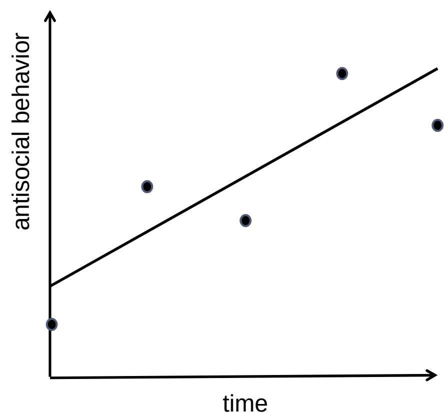
Repeated Measures for One Person

- Could connect observations to see time-adjacent changes



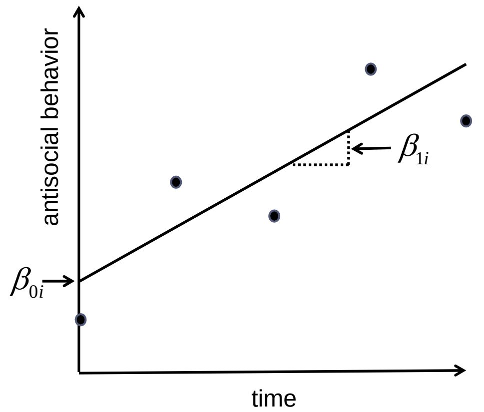
A Growth Curve for One Person

- Could instead "smooth over" repeated measures and estimate a line of best fit for this individual



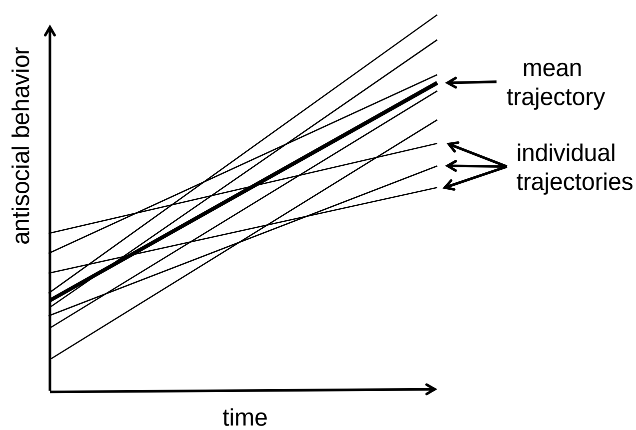
A Growth Curve for One Person

- Can summarize line by two pieces of information
  - the intercept (β_{0i}) and the slope (β_{1i}) unique to individual i



Growth Curves for Multiple Persons

- Rarely interested in one individual, but in a sample of individuals
  - can extend to 8 trajectories (but would use 100 or more in practice)



The Latent Growth Curve

- Characteristics of the latent trajectories captured in two ways
- Trajectory means
  - the average value of the parameters that define the growth trajectory pooling over all individuals in the sample
    - e.g., mean starting point and mean rate of change for the entire sample
- Trajectory variances
  - the variability of individual cases around the mean trajectory parameters
    - e.g., individual variability in starting point and rate of change over time
  - larger variances reflect greater variability in growth

Exemplar Motivating Questions

- What is the mean course of change over time?
- Are there individual differences in the course of change?
- Are there time-invariant predictors of change?
- Are there time-varying predictors of change?
- Do two constructs travel together through time?

Summary

- Motivation is hypothesis that there exists underlying continuous trajectory of change that was not directly observed
- Use observed repeated measures to infer underlying trajectory
- Means capture overall parameters that define growth trajectory
- Variances capture individual variability in growth trajectory
- Goal is to then build a model to incorporate predictors of individual variability in growth
- Will do this using both the MLM and SEM analytical frameworks

Section 2

Trajectory Estimation: Multilevel Model

Objectives

- Define the equation for a growth curve for a single individual.
- Define the equation for a growth curve for multiple individuals.
- Describe the explication of a growth curve within the general multilevel model
- Fit growth model to developmental trajectories of antisocial behavior in children

Observations Nested Within Groups

- Multilevel modeling designed to allow for nested data structures
- Classic example is children nested within classrooms
- Can define nested structure as Level 1 and 2 equations
  - Children (Level 1) are nested within classrooms (Level 2)
- Level 1 (Individual nested within-classroom):

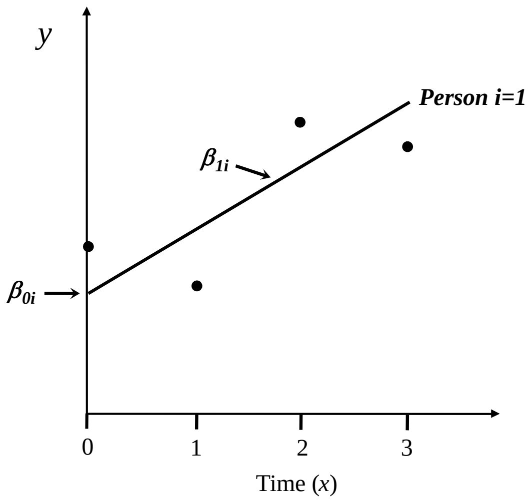
$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij}$$

- Level 2 (Between-person):

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$

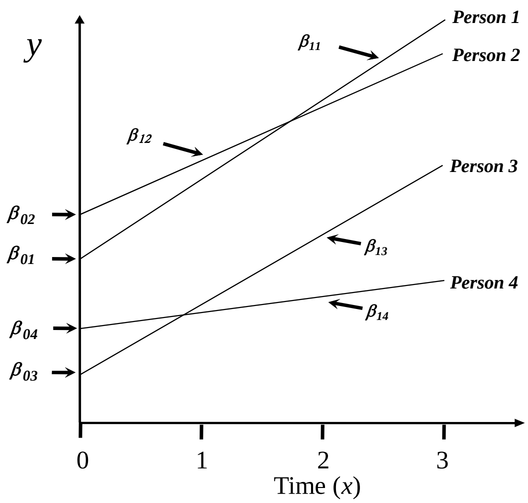
- Framework extends naturally to time nested within individual

Fitting a Single Linear Trajectory



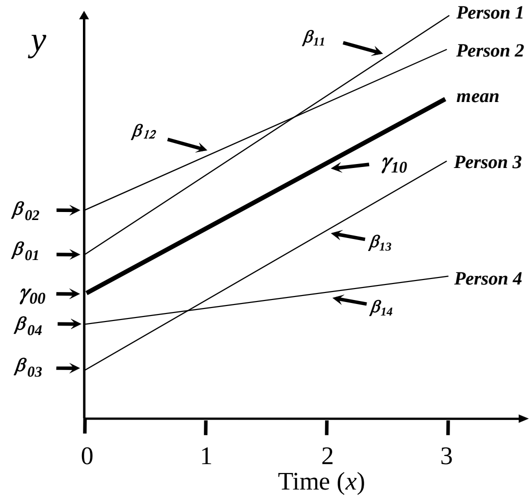
$$y_{ti} = \beta_{0i} + \beta_{1i}\, x_{ti} + r_{ti}$$
$$x_{ti} = 0,\ 1,\ 2,\ 3$$

Fitting Multiple Linear Trajectories



$$y_{ti} = \beta_{0i} + \beta_{1i}\, x_{ti} + r_{ti}$$
$$x_{ti} = 0,\ 1,\ 2,\ 3$$

The Mean Trajectory



$$y_{ti} = \beta_{0i} + \beta_{1i} x_{ti} + r_{ti}$$

$$\beta_{0i} = \gamma_{00} + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + u_{1i}$$

$$x_{ti} = 0,\ 1,\ 2,\ 3$$

Level 1 and 2 Equations

- Growth model naturally fits within multilevel modeling framework, articulated as Level 1 and 2 equations
  - Repeated measures (Level 1) nested within person (Level 2)
- Level 1 (Within-person):

$$y_{ti} = \beta_{0i} + \beta_{1i} x_{ti} + r_{ti}$$

- Level 2 (Between-person):

$$\beta_{0i} = \gamma_{00} + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + u_{1i}$$

The Reduced-Form Equation

- The Level 1 and 2 equations define a single model
- Substituting the Level 2 equations into the Level 1 equation, we obtain the reduced-form or mixed-model equation:

$$y_{ti} = \underbrace{(\gamma_{00} + u_{0i})}_{\beta_{0i}} + \underbrace{(\gamma_{10} + u_{1i})}_{\beta_{1i}} x_{ti} + r_{ti}$$

- Rearranging, we have

$$y_{ti} = \underbrace{(\gamma_{00} + \gamma_{10} x_{ti})}_{\text{Fixed Effects}} + \underbrace{(u_{0i} + u_{1i} x_{ti})}_{\text{Random Effects}} + \underbrace{r_{ti}}_{\text{Residual}}$$

Fixed and Random Effects

$$y_{ti} = \underbrace{(\gamma_{00} + \gamma_{10} x_{ti})}_{\text{Fixed Effects}} + \underbrace{(u_{0i} + u_{1i} x_{ti})}_{\text{Random Effects}} + \underbrace{r_{ti}}_{\text{Residual}}$$

- *Fixed effects* are constants in the population
  - Represented by greek "gamma"
  - Capture the mean structure in the data (i.e., $\gamma_{00} + \gamma_{10} x_{ti}$ is the mean trajectory)
- *Random effects* vary across units of the population
  - Level-2 random effects represented by u, vary over individuals i.
  - Level-1 random effect (residual) represented by r, varies over individuals i and within individuals over time t.
  - Capture between- and within-person variability, respectively

Variance Components

- We must assume a distribution for the random effects.
  - We're used to making such an assumption in regular regression, i.e., assuming the residuals to be *iid* normal.
  - Now we make similar assumptions for the two levels of our data
- Typical to assume the random effects are normally distributed, or

$$r_{ti} \sim N\left(0, \sigma^2\right) \qquad \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$$

- We estimate the variance-covariance parameters of these distributions, sometimes referred to as the variance components of the model.

Summary of Multilevel Growth Model

- Model: $y_{ti} = \underbrace{\left(\gamma_{00} + \gamma_{10} x_{ti}\right)}_{\text{Fixed Effects}} + \underbrace{\left(u_{0i} + u_{1i} x_{ti}\right)}_{\text{Random Effects}} + \underbrace{r_{ti}}_{\text{Residual}}$

- Assumptions:

$$r_{ti} \sim N\left(0, \sigma^2\right) \qquad \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{pmatrix}\right]$$

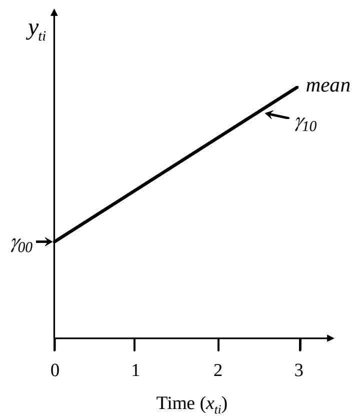
- Parameters we estimate:

γ_{00}, γ_{10} — Fixed Effects

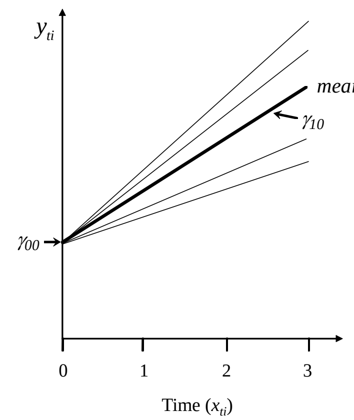
σ^2; $\tau_{00}, \; \tau_{10}, \; \tau_{11}$ — (Co)Variance Parameters

Combinations of Fixed and Random Effects

– fixed intercept
– fixed slope

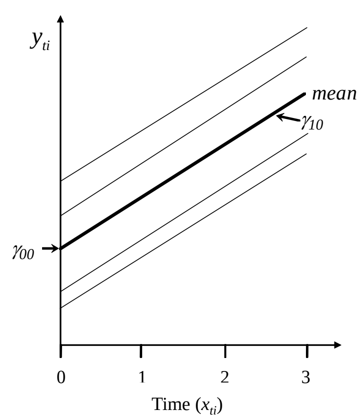


– fixed intercept
– random slope

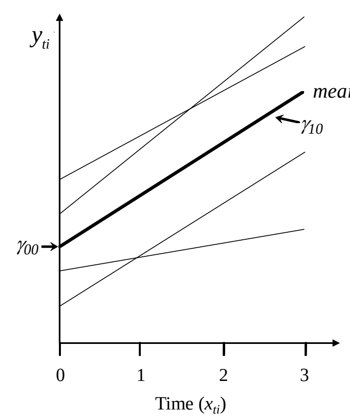


Combinations of Fixed and Random Effects

– random intercept
– fixed slope



– random intercept
– random slope



Adding Predictors to the Model

- There are two types of predictors in growth models
- *Time-invariant covariates* (TICs) are assumed to be constant over time
  - e.g., biological sex, country of origin, DNA characteristics
  - sometimes might vary over time but only baseline measures are assessed e.g., baseline anxiety or substance use
- *Time-varying covariates* (TVCs) change as a function of time
  - e.g., days of work missed per month, hours of sleep per night, marital status over time
- TVCs are introduced in the Level 1 equation, TICs in the Level 2 equations
- We'll focus on TICs here

The Inclusion of TICs in Level-2

$$y_{ti} = \beta_{0i} + \beta_{1i} x_{ti} + r_{ti}$$

The Level 1 (within person) equation is unchanged.

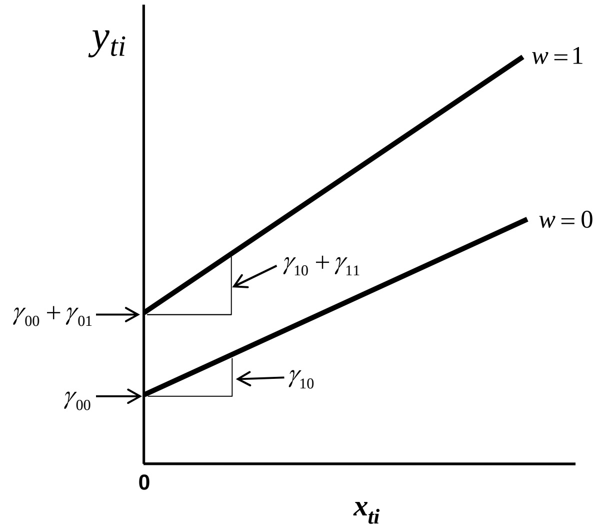
$$\beta_{0i} = \gamma_{00} + \gamma_{01} w_i + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11} w_i + u_{1i}$$

Here we expand the Level 2 (between persons) equations to include a single person-level predictor denoted *w*.

$$y_{ti} = \left(\gamma_{00} + \gamma_{10} x_{ti} + \gamma_{01} w_i + \gamma_{11} x_{ti} w_i\right) + \left(u_{0i} + u_{1i} x_{ti} + r_{ti}\right)$$

- Notice additional fixed effects and the *cross-level interaction* between *x* and *w*.

Binary Person-Level Predictor



Example: Antisocial Behavior

- N=405 cases drawn from the National Longitudinal Survey of Youth (NLSY); N=221 complete data.
- Age 6-8 at first assessment; re-assessed up to three more times every other year.
- Mother report of child antisocial behavior on six items, each with a 0,1,2 response scale. Sum score ranges from 0 to 12.
- Predictor: Child Gender
- Research question: What are the characteristics of trajectories of antisocial behavior, and can these trajectories be predicted by child-level measures?

Data Structure for Multilevel Analysis

id	male	anti	age
1	1	1	6
1	1	0	9
1	1	1	11
1	1	0	13
2	1	1	7
2	1	1	10
2	1	0	12
2	1	1	14
3	0	5	8
3	0	0	11
3	0	5	13
3	0	3	14
4	1	1	7
4	1	1	10
5	1	2	6
5	1	3	9
5	1	3	11
5	1	1	13

note this case has only two observations → (id 4)

This type of data structure is sometimes called "long" because there are multiple records for each case, one for each assessment

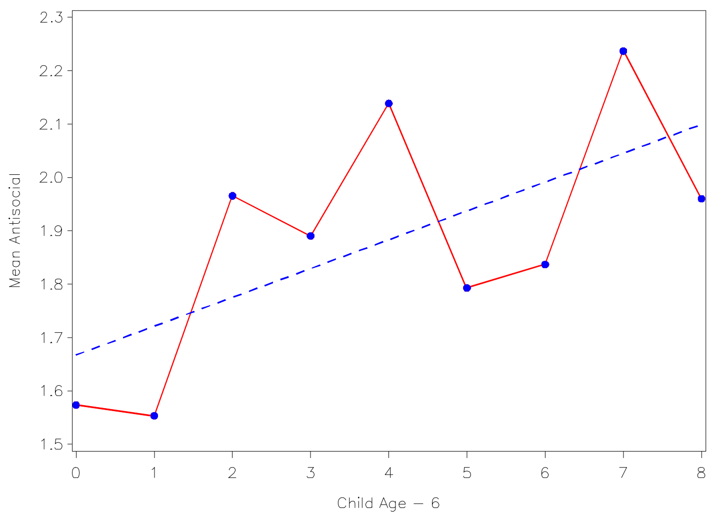
Age-Specific Means

- Note that although any individual case has between 1 and 4 repeated assessments, there are nine unique ages

Antisocial Behavior

age	*N*	*Mean*	*Std Dev*	*Variance*
6	122	1.57	1.66	2.77
7	168	1.55	1.55	2.41
8	146	1.97	1.79	3.23
9	192	1.89	1.95	3.81
10	151	2.14	2.13	4.56
11	174	1.79	1.94	3.79
12	135	1.84	1.77	3.16
13	173	2.24	2.27	5.15
14	101	1.96	2.08	4.35

Plotting Age-Specific Means



Flat Trajectory Model

- Level-1 model

$$anti_{ti} = \beta_{0i} + r_{ti} \qquad r_{ti} \sim N(0, \sigma^2)$$

- Level-2 model

$$\beta_{0i} = \gamma_{00} + u_{0i} \qquad u_{0i} \sim N(0, \tau_{00})$$

- Reduced-Form model

$$anti_{ti} = \gamma_{00} + u_{0i} + r_{ti}$$

- This model implies no-change over time (although from the sample statistics we do not believe this to be true)

Flat Trajectory Model

$$anti_{ti} = \gamma_{00} + u_{0i} + r_{ti}$$

Covariance Parameter Estimates

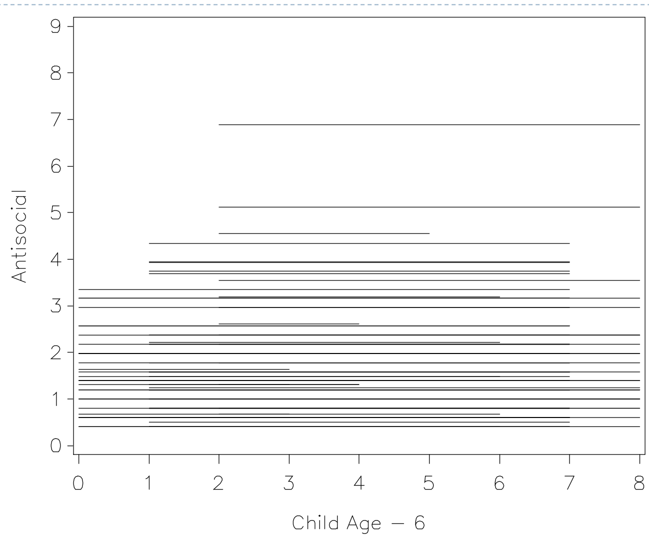
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	id	1.7974	0.1723	10.43	<.0001
Residual		1.9572	0.08936	21.90	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	1.8975	0.07736	404	24.53	<.0001	0.05	1.7454	2.0496

http://curranbauer.org/SRA2016/

Model-Implied Trajectories



Random Intercept Linear Growth Model

- Level-1 model

$$anti_{ti} = \beta_{0i} + \beta_{1i} age_{ti} + r_{ti} \qquad r_{ti} \sim N(0, \sigma^2)$$

- Level-2 model

$$\beta_{0i} = \gamma_{00} + u_{0i} \qquad u_{0i} \sim N(0, \tau_{00})$$
$$\beta_{1i} = \gamma_{10}$$

- Reduced-Form model

$$anti_{ti} = (\gamma_{00} + \gamma_{10} age_{ti}) + (u_{0i} + r_{ti})$$

- This model implies that antisocial behavior is changing over time, but that all children are changing at exactly the same rate

Random Intercept Linear Growth Model

$$anti_{ti} = (\gamma_{00} + \gamma_{10} age_{ti}) + (u_{0i} + r_{ti})$$

Covariance Parameter Estimates

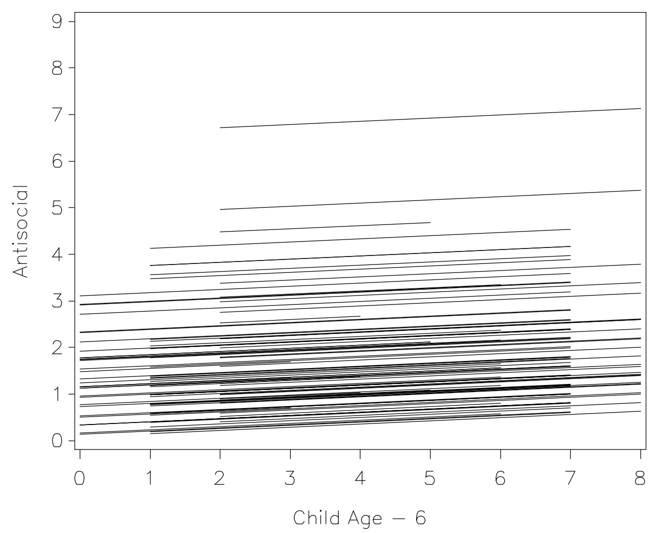
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	id	1.8197	0.1732	10.51	<.0001
Residual		1.9199	0.08774	21.88	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	1.6384	0.09888	404	16.57	<.0001	0.05	1.4440	1.8328
age	0.06853	0.01623	956	4.22	<.0001	0.05	0.03669	0.1004

http://curranbauer.org/SRA2016/

Model-Implied Trajectories



Fully Random Linear Growth Model

- Level-1 model

$$anti_{ti} = \beta_{0i} + \beta_{1i} age_{ti} + r_{ti} \qquad r_{ti} \sim N(0, \sigma^2)$$

- Level-2 model

$$\beta_{0i} = \gamma_{00} + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + u_{1i}$$

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

- Reduced-Form model

$$anti_{ti} = (\gamma_{00} + \gamma_{10} age_{ti}) + (u_{0i} + u_{1i} age_{ti} + r_{ti})$$

- This model implies that antisocial behavior is changing over time, and the amount of change varies randomly over individual

Fully Random Linear Growth Model

$$anti_{ti} = (\gamma_{00} + \gamma_{10} age_{ti}) + (u_{0i} + u_{1i} age_{ti} + r_{ti})$$

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
Intercept	id	1.0134	0.2312	4.38	<.0001
covariance	id	0.07303	0.03831	1.91	0.0566
Slope	id	0.02295	0.01009	2.27	0.0115
Residual		1.7518	0.1017	17.23	<.0001

Estimated Correlation Matrix

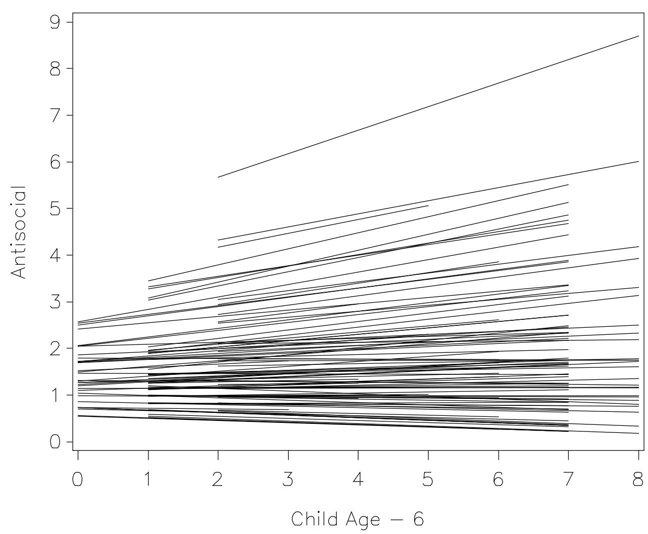
Effect		
Intercept	1.0000	0.4788
Slope	0.4788	1.0000

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	1.6289	0.08573	404	19.00	<.0001	0.05	1.4604	1.7975
age	0.07425	0.01774	956	4.18	<.0001	0.05	0.03943	0.1091

http://curranbauer.org/SRA2016/

Model-Implied Growth Trajectories



Predicting Individual Trajectories

- Level 1 model

$$anti_{ti} = \beta_{0i} + \beta_{1i} age_{ti} + r_{ti} \qquad r_{ti} \sim N(0, \sigma^2)$$

- Level 2 model

$$\beta_{0i} = \gamma_{00} + \gamma_{01} male_i + u_{0i}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11} male_i + u_{1i}$$
$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

- Reduced-Form model

$$anti_{ij} = \left(\gamma_{00} + \gamma_{10} age_{ij} + \gamma_{01} male_i + \gamma_{11} male_i age_{ij} \right) + \left(u_{0j} + u_{1j} age_{ij} \right) + r_{ij}$$

Dichotomous Predictor of Antisocial

Solution for Fixed Effects

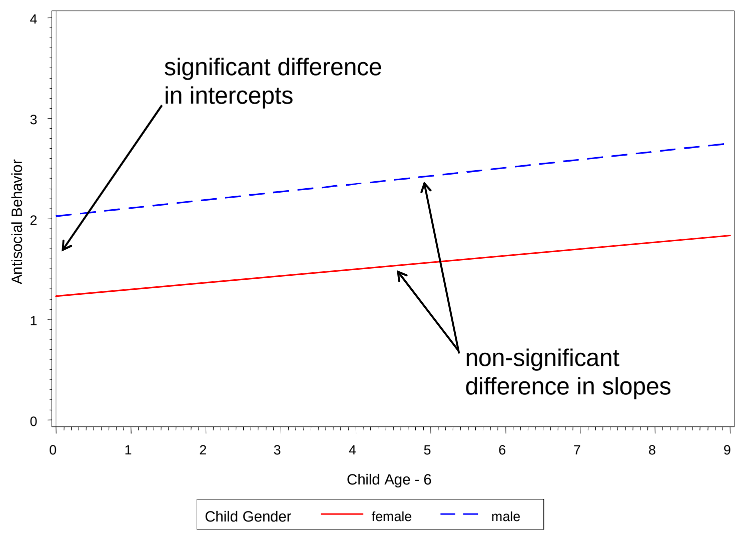
Effect	*Estimate*	*Standard Error*	*DF*	*t Value*	*Pr > |t|*	*Alpha*	*Lower*	*Upper*
Intercept	1.2269	0.1189	403	10.32	<.0001	0.05	0.9931	1.4607
age	0.06736	0.02559	955	2.63	0.0086	0.05	0.01714	0.1176
male	0.7987	0.1668	403	4.79	<.0001	0.05	0.4708	1.1266
*age*male*	0.01291	0.03550	955	0.36	0.7161	0.05	-0.05674	0.08257

- Model-implied mean trajectories for males and females:

$$E\left(antisocial \mid male \right) = \left(1.227 + .799 \times male \right) + \left(.067 + .013 \times male \right) age$$
$$E\left(antisocial \mid male = 0 \right) = \left(1.227 \right) + \left(.067 \right) age$$
$$E\left(antisocial \mid male = 1 \right) = \left(1.227 + .799 \right) + \left(.067 + .013 \right) age$$

http://curranbauer.org/SRA2016/

Model-Implied Mean Trajectories



Summary

- Instead of comparing time-specific means (as in ANOVA/MANOVA) or time-adjacent prediction (as in ARCL) growth models estimate smoothed time trajectories within-individual
- Because growth trajectory can be conceptualized as repeated measures nested within individual, we can bring the multilevel model to bear on the analysis of repeated measures data
  - Fixed effects capture mean-level change, random effects capture individual-level change
- Example showed
  - Antisocial behavior increases from age 6 to 14
  - Large individual differences in initial levels and rates of change over time
  - On average, boys have higher levels of antisocial behavior than girls, but both groups show comparable increases over time.

Extensions of the MLM

- Can estimate nonlinear trajectories over time
- Can include time-varying covariates (TVC)
  - TVC is not person-specific (like biological sex) but can vary with time
  - e.g., symptom count, diagnosis, hours slept, alcohol consumed
- Multivariate MLM can include growth in two outcomes
  - TVC model assumes covariate not changing as function of time
  - multivariate MLM allows growth in two or more processes
- MLM naturally extends to three or more levels of nesting
  - e.g., time nested within child nested within classroom
  - can examine within- and across-level predictors for all three levels
- Can estimate with discrete outcomes (e.g., binary, ordinal, count)
- MLM a powerful and flexible method for longitudinal data, but SEM approach offers several advantages as well

Section 3

Trajectory Estimation: Latent Curve Model

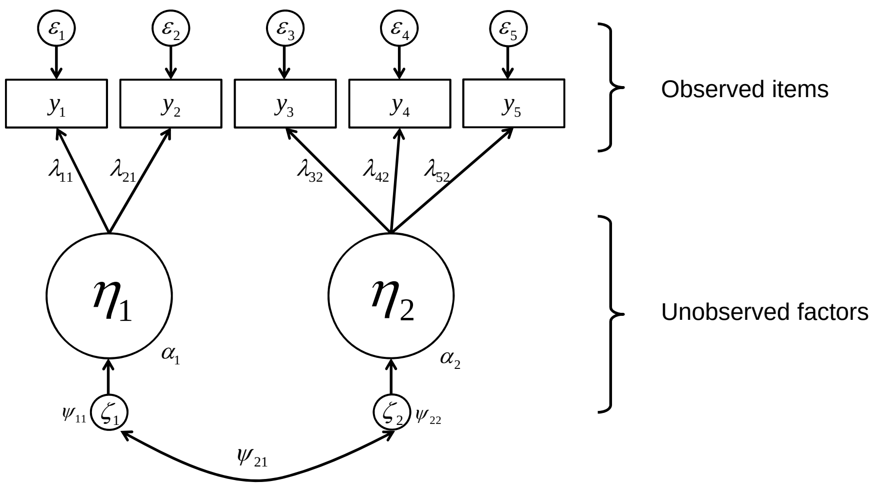
Objectives

- Show how a growth model can be cast as a special type of structural equation model, the latent curve model
- Discuss the correspondence between multilevel growth models and latent curve models
- Provide an example

Confirmatory Factor Analysis

- LCM is a special case of confirmatory factor analysis (CFA)
- CFA is primarily theory-driven: test model that specifies number and nature of the latent factors behind set of observed measures
  - e.g., latent depression and anxiety underlie set of 20 symptom items
- Model identified through restrictions on parameters
- Number of latent factors determined by theory
- Factor pattern matrix is restricted by analyst to reflect theory
  - e.g., some loadings freely estimated, others fixed to zero
- Attention paid to global and local fit of model to data

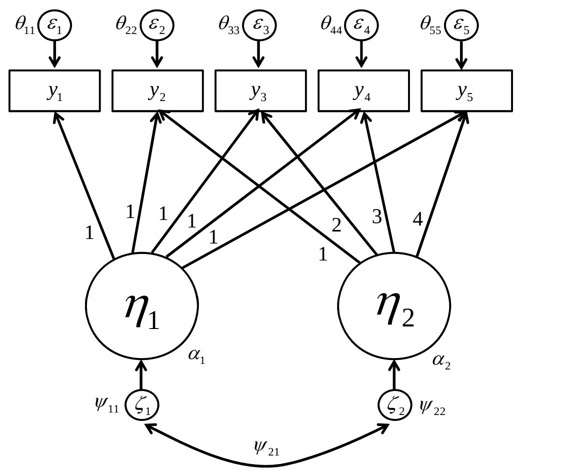
Blueprint of the General CFA



Capturing Growth as a Latent Factor

- Theory posits existence of unobserved continuous trajectory
- Cannot directly observe, but can infer existence based on set of repeated measures
- Growth curve thus fits naturally into latent variable model
  - e.g., depression, self esteem, worker productivity, etc.
- Can draw on strengths of confirmatory factor analysis (CFA) to define latent curve model
- The LCM is fundamentally a highly restricted CFA model
- In MLM, *time* enters the model as a numerical predictor variable
- in SEM, *time* enters the model as fixed or estimated values in the factor loading matrix

A Linear LCM as a Path Diagram



Repeated Measures

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

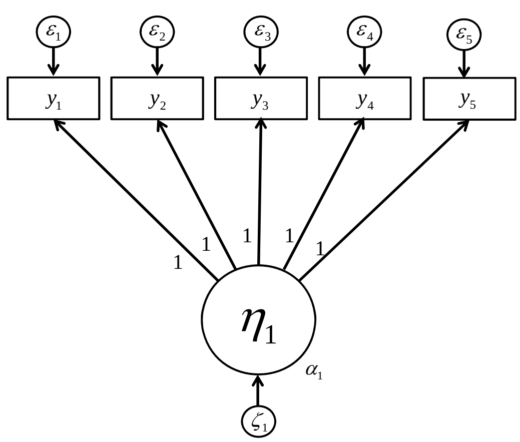
Latent Growth Factors

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i$$

- We can examine different expressions of this model…

Intercept-only LCM

- Implies outcome does not change as function of time
  - all individual trajectories horizontal but at different levels



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

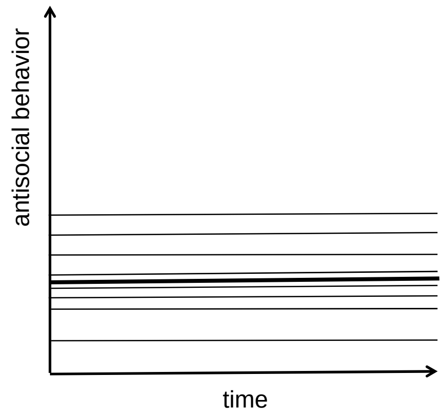
$$\begin{bmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \\ y_{5i} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \left[\eta_{1i}\right] + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \\ \varepsilon_{5i} \end{bmatrix}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i$$

$$\left[\eta_{1i}\right] = \left[\alpha_1 + \zeta_{1i}\right]$$

Intercept-only LCM: Trajectories

- Intercept-only LCM implies between-person variability in overall level of outcome, but outcome does not change with time



Intercept-only LCM: Factor Mean

- Just one factor mean because only one factor is defined to represent latent intercept

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i$$

$$E(\boldsymbol{\eta}_i) = \boldsymbol{\alpha}$$

$$\boldsymbol{\alpha} = [\alpha_1]$$

- This simply reflects the mean level of all repeated measures pooled over all individuals

Intercept-only LCM: Factor Variance

- Just one factor variance again because only one factor is defined to represent latent intercept

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i$$

$$\boldsymbol{\zeta}_i \sim N(\mathbf{0}, \boldsymbol{\Psi})$$

$$\boldsymbol{\Psi} = [\psi_{11}]$$

- This represents the individual variability around the overall mean

Intercept-only LCM: Residual Variance

- Finally, time-specific residuals for RMs allowed to obtain a unique value at each time point and are typically uncorrelated over time

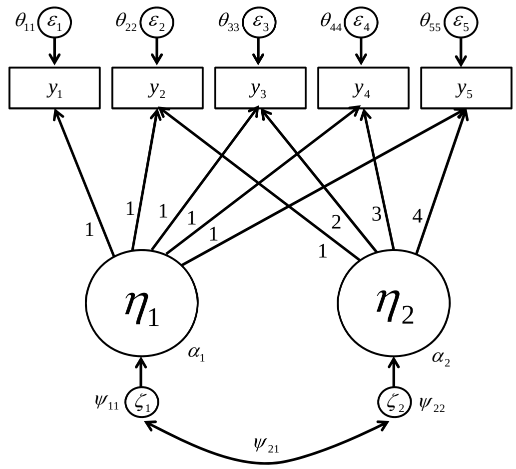
$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$$

$$\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Theta})$$

$$\boldsymbol{\Theta} = \begin{bmatrix} \theta_{11} & & & & \\ 0 & \theta_{22} & & & \\ 0 & 0 & \theta_{33} & & \\ 0 & 0 & 0 & \theta_{44} & \\ 0 & 0 & 0 & 0 & \theta_{55} \end{bmatrix}$$

Linear LCM

- Can add a second correlated factor to capture linear change:



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

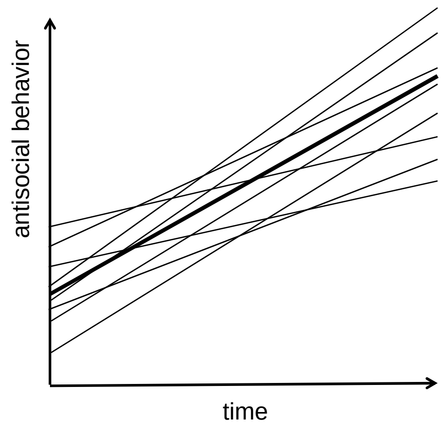
$$\begin{bmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \\ y_{5i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \\ \varepsilon_{5i} \end{bmatrix}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i$$

$$\begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} = \begin{bmatrix} \alpha_1 + \zeta_{1i} \\ \alpha_2 + \zeta_{2i} \end{bmatrix}$$

Linear LCM: Trajectories

- Intercept and linear slope LCM implies individual differences in both level and rate of change



Linear LCM: Mean Structure

- Two factor means, one for intercept and one for linear slope

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i$$

$$E\left(\boldsymbol{\eta}_i\right) = \boldsymbol{\alpha}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}$$

- Reflects mean starting point and mean linear rate of change
  - defines starting point because we set first value of time to be zero
  - can also have zero-point at middle or end of trajectory

Linear LCM: Variance Components

- Factor variance now expressed as a covariance matrix with variance of intercept and slope and covariance between the two

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \boldsymbol{\zeta}_i$$

$$\boldsymbol{\zeta}_i \sim N\left(\mathbf{0}, \boldsymbol{\Psi}\right)$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix}$$

- Represents individual variability around starting point and rate of change, and covariance between starting point and rate of change
  - covariance can be standardized to a correlation coefficient

Linear LCM: Variance Components

- Interestingly, the covariance structure among the time-specific residuals is precisely the same as with the intercept-only LCM

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

$$\boldsymbol{\varepsilon}_i \sim N\left(\mathbf{0}, \mathbf{\Theta}\right)$$

$$\mathbf{\Theta} = \begin{bmatrix} \theta_{11} & & & & \\ 0 & \theta_{22} & & & \\ 0 & 0 & \theta_{33} & & \\ 0 & 0 & 0 & \theta_{44} & \\ 0 & 0 & 0 & 0 & \theta_{55} \end{bmatrix}$$

Heteroscedastic Residuals

- Standard LCM assumes each time-specific repeated measure is defined by a unique residual variance
  - called heteroscedasticity

$$\mathbf{\Theta} = \begin{bmatrix} \theta_{11} & & & & \\ 0 & \theta_{22} & & & \\ 0 & 0 & \theta_{33} & & \\ 0 & 0 & 0 & \theta_{44} & \\ 0 & 0 & 0 & 0 & \theta_{55} \end{bmatrix}$$

- A simplifying condition is to assume error variances are equal
  - called homoscedasticity

Homoscedastic Residuals

- Can impose equality constraint on residuals over time

$$\mathbf{\Theta} = \begin{bmatrix} \theta & & & & \\ 0 & \theta & & & \\ 0 & 0 & \theta & & \\ 0 & 0 & 0 & \theta & \\ 0 & 0 & 0 & 0 & \theta \end{bmatrix}$$

- This is a testable hypothesis
  - homoscedasticity more parsimonious, but may not correspond to characteristics of the observed data
- Typically want to identify most parsimonious structure that does not significantly contribute to model misfit

Example: Antisocial Behavior

- We use precisely the same data as before
- N=405 cases drawn from National Longitudinal Survey of Youth
- Age 6-8 at first assessment; re-assessed up to 3 more times every other year
- Mother report of child antisocial behavior on six items, each with a 0,1,2 response scale. Sum score ranges from 0 to 12
- Initial research questions:
  - what is the optimal mean functional form of the trajectory over time?
  - is there individual child-to-child variability around these mean values?
- Later we will consider child-specific predictors of the trajectories, but here we focus on establishing the optimal functional form

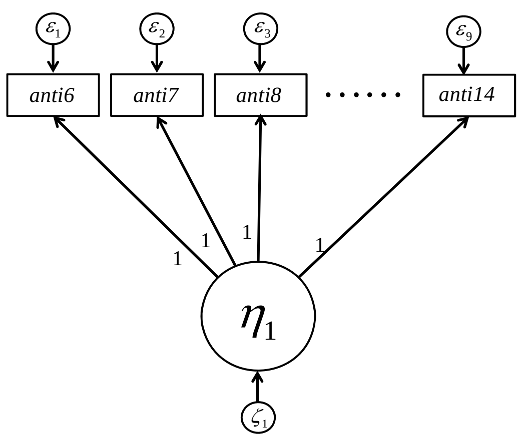
The Typical Data Structure

- Recall MLM data was in "long" format
  - one line of data per-individual, per-assessment
- For or SEM, we will use "wide" format
  - one line of data per-individual
- All data for each case is on a single line, and variable names demarcate time at which measure was assessed

Obs	*id*	*male*	*age86*	*age88*	*age90*	*age92*	*anti86*	*anti88*	*anti90*	*anti92*
1	1	1	6	9	11	13	1	0	1	0
2	2	1	7	10	12	14	1	1	0	1
3	3	0	8	11	13	14	5	0	5	3
4	4	1	7	10	.	.	1	1	.	.
5	5	1	6	9	11	13	2	3	3	1

Intercept-Only Model

- We will begin by estimating an intercept-only model
  - although we expect this to fit poorly given the sample means



$$\Lambda = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

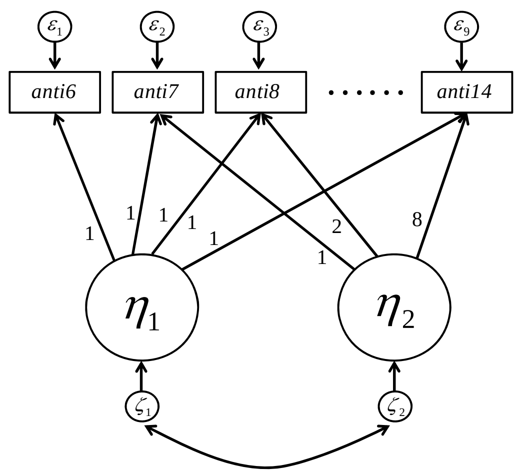
Intercept-Only Model

- As expected, this model fit the data quite poorly
  - $\chi^2(39) = 101.7$, $p < .0001$
  - CFI=.82
  - TLI=.85 } typically want values > .90
  - RMSEA=.06 ← typically want values < .05
  - SRMR=.16 ← typically want values < .08
- No index supports adequate fit of the model to the sample data
- Next expand model to include a linear slope component
  - because intercept-only model is nested within the linear model, can conduct a likelihood ratio test (LRT) to assess improvement in model fit

http://curranbauer.org/SRA2016/

Intercept and Linear Slope Model

- We added a slope factor to the intercept-only model



$$\Lambda = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}$$

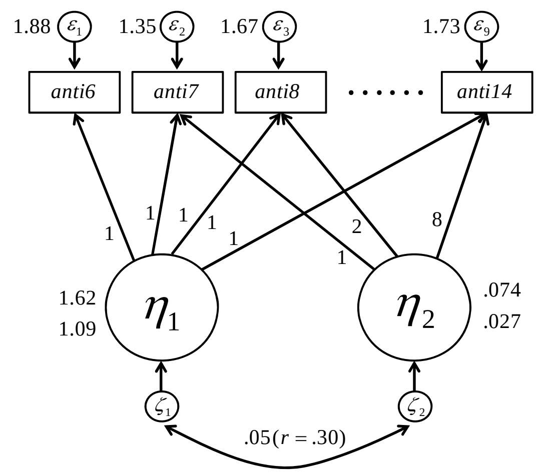
Intercept and Slope Model

- This model fit the data significantly better compared to the intercept-only model
  - intercept model: $\chi^2(39) = 101.7$
  - intercept + slope model: $\chi^2(36) = 52.6$
  - LRT difference: $\chi^2(3) = 49.1$ ($p<.0001$)
- Fit indices reflects that linear model fits data rather well
  - $\chi^2(36)=52.6$, $p=.04$;
  - CFI=.95; TLI=.96; RMSEA=.03;
  - SRMR=.11
- Could also test nonlinear function but don't show here
  - nonlinear does not significantly improve fit of model

http://curranbauer.org/SRA2016/

Linear LCM: Parameter Estimates

- The latent factor means are listed first followed by the variances



Summary

- Linear trajectory deemed optimal functional form of growth
  - multiple ways to model nonlinearity, but do not show these here
- Significant latent factor means for intercept and slope
  - on average, children are reporting antisocial behaviors of 1.62 at age 6 and are increasing at .07 units per year
- Significant latent factor variances for intercept and slope
  - there is significant individual variability in the initial levels of antisocial behavior and in the rates of increase in antisocial behavior over time
- Non-significant covariance between intercept and slope factors
  - on average, individual variability in starting point is not related to individual variability in rate of change over time
- Can next consider predictors of growth over time

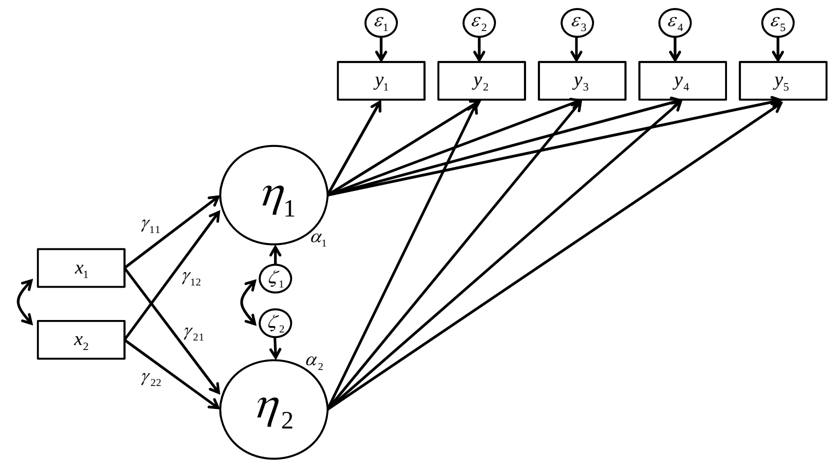
Time-Invariant Predictors

- TICs do not influence the measurement equation, but instead enter via the structural equation
  - e.g., for two TICs the structural equation is

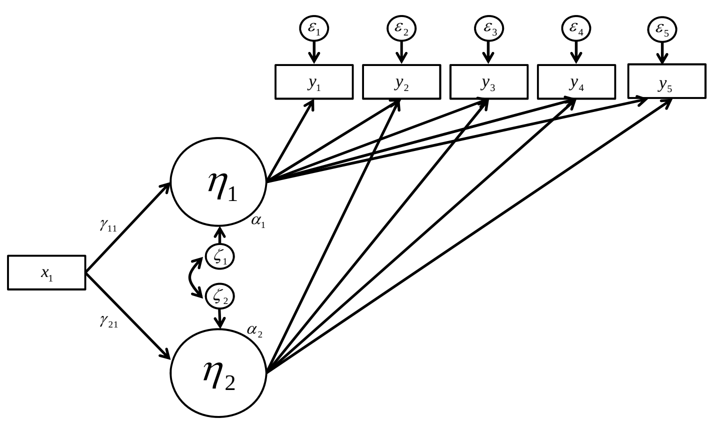
$$\begin{aligned}\boldsymbol{\eta}_i &= \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i \\ &= \begin{bmatrix}\alpha_1 \\ \alpha_2\end{bmatrix} + \begin{bmatrix}\gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22}\end{bmatrix}\begin{bmatrix}x_{1i} \\ x_{2i}\end{bmatrix} + \begin{bmatrix}\zeta_{1i} \\ \zeta_{2i}\end{bmatrix} \\ &= \begin{bmatrix}\alpha_1 + \gamma_{11}x_{1i} + \gamma_{12}x_{2i} + \zeta_{1i} \\ \alpha_2 + \gamma_{21}x_{1i} + \gamma_{22}x_{2i} + \zeta_{2i}\end{bmatrix}\end{aligned}$$

- Really no different than a two-predictor regression equation, but latent intercept and slope are dependent variables

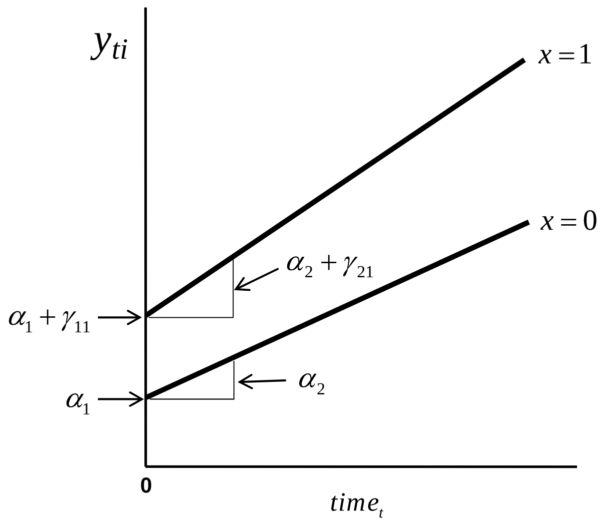
Time-Invariant Predictors



Probing Conditional Effects: 1 Predictor

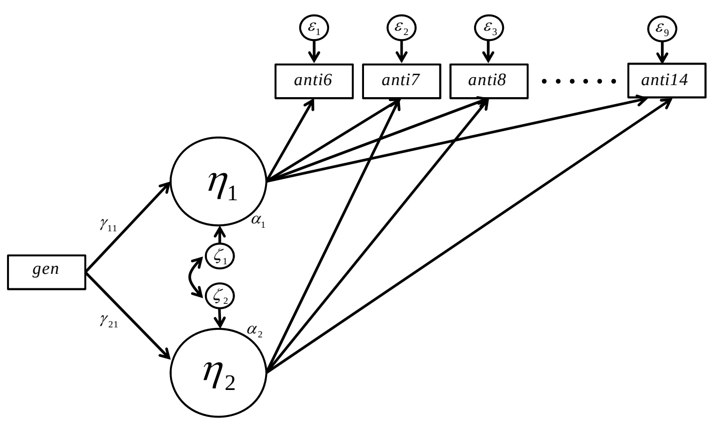


Binary Person-Level Predictor



Example: Antisocial Behavior

- Begin by regressing linear LCM on binary measure of gender
  - zero represents girls and one represents boys

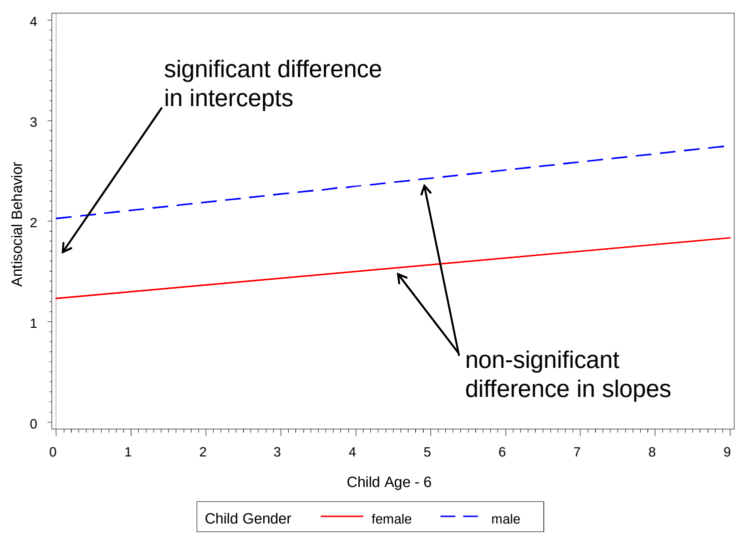


Example: Antisocial Behavior

- The model fit the data well
  - $\chi^2(43)$ =59.9, p=.045; CFI=.96; TLI=.96; RMSEA=.03; SRMR=.11
- Was a significant prediction of intercept ($\hat{\gamma}_{01} = .79,\ p < .001$) but not of slope ($\hat{\gamma}_{11} = .01,\ p = .70$)
  - boys reported initial levels of antisocial behavior .79-units higher than did girls
  - girls and boys did not differ in the rates of change in antisocial behavior over time
- We can compute the simple intercepts and simple slopes for boys and girls and examine this effect graphically

http://curranbauer.org/SRA2016/

Model-Implied Simple Trajectories



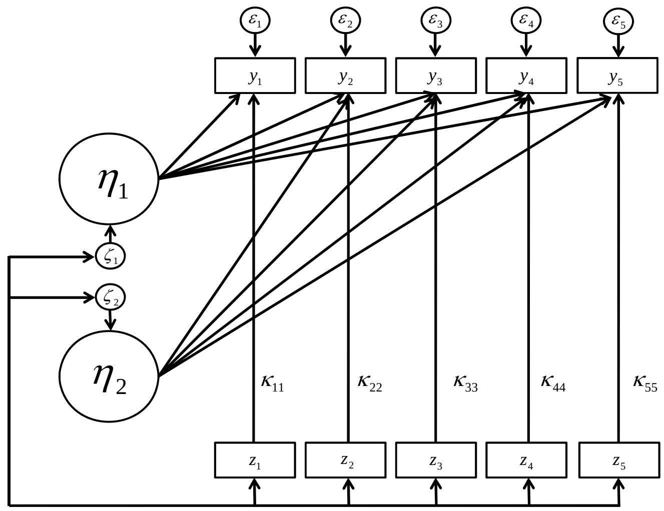
Summary

- Unconditional LCM shows there is significant variability in both starting point and positive rate of change over time
- Conditional LCM shows that males start significantly higher but do not increase significantly steeper over time
- Only focused on binary predictor, but could easily expand to include multiple predictors that are binary or continuous
- Can also estimate interactions among TICs themselves
  - e.g., examine interaction between gender and cognitive stimulation in the home in the prediction of trajectories of antisocial behavior
- Many extensions possible the LCM

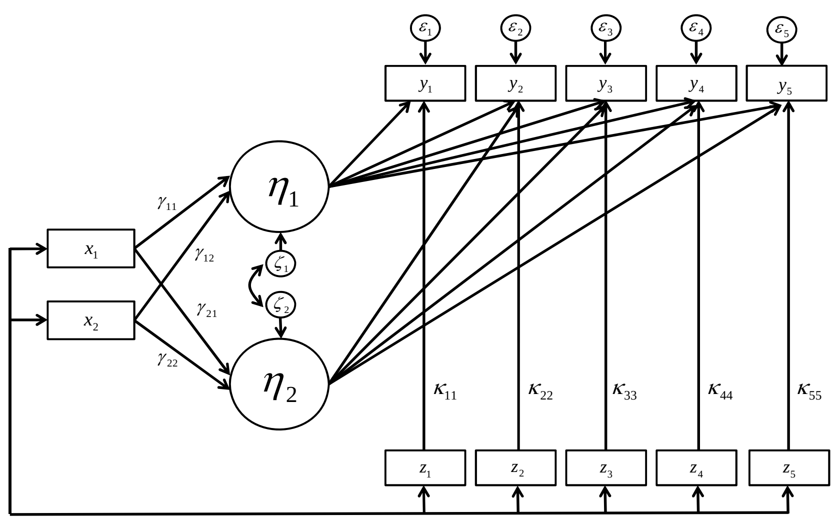
Extensions of the LCM

- Many options for modeling nonlinearity
  - polynomials, freed loadings, piecewise linear, exponential, etc.
- Can estimate complex forms of mediation
- Can use latent factors to model measurement error
- Can estimate models as function of observed or unobserved discrete groups (i.e., "latent classes")
- There are a variety of methods for modeling the unfolding of two or more constructs over time
  - these are most easily seen in path diagram form...

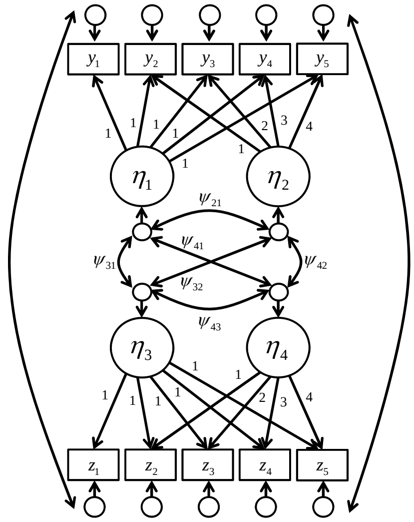
Time-Varying Covariates: Unconditional



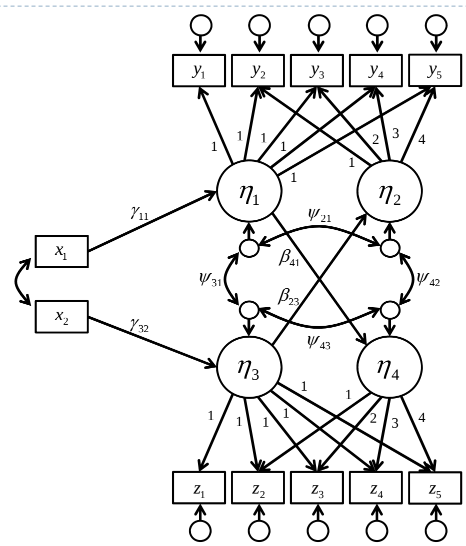
Conditional TVC Model



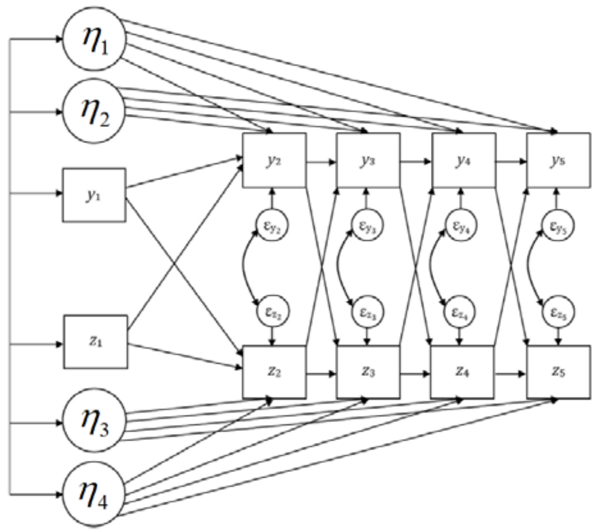
Multivariate LCM: Correlated Factors



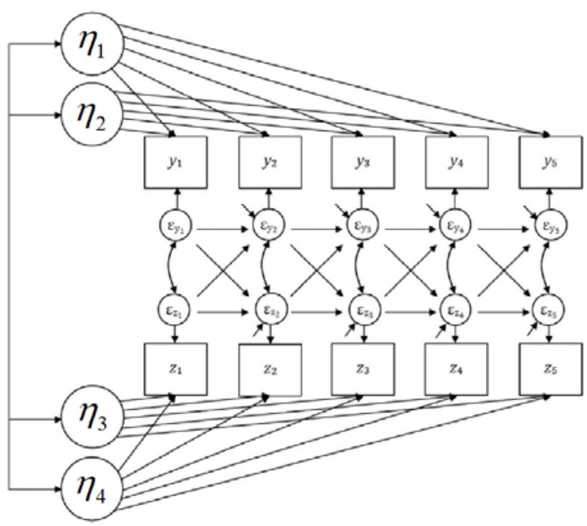
Multivariate LCM: Conditioned on TICs



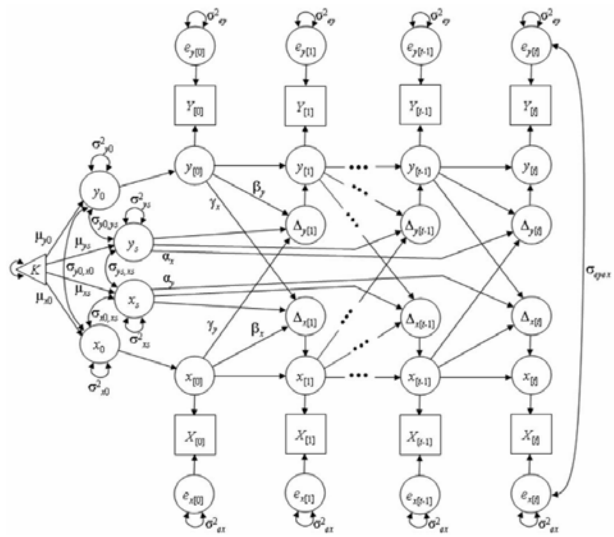
Bivariate Unconditional ALT Model



LCM with Structured Residuals



Latent Change Score Model



Selecting Between MLM and LCM

- For many basic growth models, the LCM and MLM are *precisely the same model* fit within different modeling architectures
  - they are not just similar but are *mathematically equivalent*
- In general, the MLM is better suited for higher-order nesting
  - 3-level models of time nested within person nested within group
  - cross-classified models where can be nested in one of two groups at same level (e.g., neighborhood or school or both)
  - also handles time-intensive longitudinal data very well
- In general, the SEM is better suited for multivariate models
  - tests of multi-chain mediation
  - multiple indicator latent factors defining a measurement model
  - many different types of multivariate growth models
- MLM or SEM is not better or worse -- they are simply different

What to do Next

- Many excellent articles, books, chapters, and online resources
  - see attached list for just a few suggestions
- Try to seek out more traditional classroom-based classes
  - regression, multivariate, MLM, SEM, factor analysis, etc.
- There are an increasing number of summer workshops available
- We offer workshops in MLM, SEM, longitudinal SEM, & mixtures
  - *curranbauer.org*
- But there are many other options available around the country
- Excellent strategy is to get data and experiment with models
  - both fitting various models & thinking about results relative to theory
- Key thing to realize is all of us can fit these models -- they are intuitive & straightforward and can become standard practice

A non-Exhaustive Sampling of Further Reading

Bauer, D.J. (2007). Observations on the use of growth mixture models in psychological research. Multivariate Behavioral Research, 42, 757-786.

Bauer, D.J. & Curran, P.J. (2004). The integration of continuous and discrete latent variable models: potential problems and promising opportunities. Psychological Methods, 9, 3-29.

Bauer, D. J. & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. Multivariate Behavioral Research, 40, 373-400.

Biesanz, J.C., Deeb-Sossa, N., Papadakis, A.A., Bollen, K.A., & Curran, P.J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9,* 30-52.

Bolger, N. Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579-616.

Bollen, K.A., & Curran, P.J. (2006). *Latent Curve Models: A Structural Equation Approach*. Wiley Series on Probability and Mathematical Statistics. John Wiley & Sons: New Jersey.

Cole, D.A., Martin, N.C., & Steiger, J.H. (2005). Empirical and conceptual problems with longitudinal trait-state models: Support for a Trait-State-Occasion model. *Psychological Methods, 10,* 3-20.

Curran, P.J., & Bauer, D.J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. Annual Review of Psychology, 62, 583-619.

Curran, P.J., Howard, A.L., Bainter, S.A., Lane, S.T., & McGinley J.S. (2013). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. Journal of Consulting and Clinical Psychology, 82, 879-894.

Curran, P.J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. Journal of Cognition and Development, 11, 121-136.

Ferrer, E., & McArdle, J.J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science, 19*, 149-154.

Hedeker, D., & Gibbons, R. D. (2006). Longitudinal Data Analysis. New York: Wiley.

Hoffman, L. (2014). Longitudinal Analysis: Modeling Within-Person Fluctuation and Change. Routledge, NY, NY.

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), Handbook of multivariate experimental psychology (pp. 561-614). New York: Plenum.

McArdle, J.J. (1989). Structural modeling experiments using multiple growth functions. In P. Ackerman, R. Kanfer & R. Cudeck (Eds.), *Learning and individual differences: abilities, motivation, and methodology* (pp. 71-117). Hillsdale, NJ: Erlbaum.

McArdle, J.J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60,* 577-605. DOI: 10.1146/annurev.psych.60.110707.163612

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110–133.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55,* 107-122.

Muthén, B.O., & Curran, P.J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2,* 371-402.

Peugh, J. L., & Enders, C. K. (2005). Using the SPSS mixed procedure to fit cross-sectional and longitudinal multilevel models. Educational and Psychological Measurement, 65, 717-741

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interaction effects in multiple linear regression, multilevel modeling, and latent curve analysis. Journal of Educational and Behavioral Statistics, 31, 437-448.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. Journal of Educational and Behavioral Statistics, 23, 323-355

Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. Oxford Univ. Press

Willett, J.B., & Sayer, A.G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. Psychological Bulletin, 116, 363-381.