Individual Growth Modeling:
Modern Methods for Studying Change

Judith D. Singer & John B. Willett
Harvard Graduate School of Education

“Time is the one immaterial object which we cannot influence—neither speed up nor slow down, add to nor diminish.”
Maya Angelou

You may download these slides and supporting materials at:
http://gseacademic.harvard.edu/alda/
http://gseacademic.harvard.edu/~willetjo/
http://gseweb.harvard.edu/~faculty/singer/
http://www.ats.ucla.edu/stat/examples/alda/
http://www.oup.com/us/singerwillettbook

Individual Growth Modeling: Modern Methods for Studying Change
Judith D. Singer and John B. Willett
Harvard Graduate School of Education

Date & Time
Friday, May 20, 2005
8:30 AM – 9:00 AM Check-in
9:00 AM – 5:00 PM Course

Location
Larsen Hall Room G-08
Harvard Graduate School of Education
Appian Way, Harvard Square Cambridge, MA

Cost
$100 for chapter members, $130 for non-members, and $70 for students (ID must be presented at check-in, or send a copy with your advance registration). This will cover the cost of the course, morning coffee, lunch, and course materials.

Registration
Limited to 90 participants. Mail a check (along with your name and e-mail address) for the course fee, payable to BCASA, addressed to BCASA, c/o Tom Lane, 128 Bingham Rd., Carlisle, MA 01741. Registrations will be accepted until the course fills, but should arrive no later than May 13. If space remains, on-site registration will be allowed. No refunds after May 13 unless you have someone else to fill the space. Receipts will be available at the event. Inquiries can be sent to tlane@mathworks.com.

Directions
See www.gse.harvard.edu/~admit/directions.html for directions to the Ed School campus. This website includes campus maps, subway information, and a list of local parking garages.

Abstract
Based on their book, Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence (Oxford, 2003), Singer and Willett will give an accessible yet in-depth presentation of multilevel models for individual change. Using real data sets from published studies, the instructors will take participants step-by-step through complete analyses, from simple exploratory displays that reveal underlying patterns through sophisticated specifications of complex statistical models. All concepts will be illustrated using real data sets from recent studies. Implementation using a variety of software packages will also be discussed (including SAS, Stata, SPSS, Splus, MLwiN and HLM). The course’s emphasis is data analytic, focusing on five linked phases of work: articulating research questions; postulating an appropriate model and understanding its assumptions; choosing a sound method of estimation; interpreting analytic results; and presenting findings—in words, tables, and graphs—to both technical and non-technical audiences. Thoughtful analysis can be difficult and messy, raising delicate problems of model specification and parameter interpretation. The default options in most computer packages do not fit the statistical models people generally want. The course’s goal is to provide you with the short-term guidance needed to start using the methods quickly, as well as with long-term advice to support your work wisely once begun. The morning session will begin with descriptive and exploratory methods, followed by a detailed discussion of basic model specification, model fitting, and parameter interpretation. The afternoon session will extend these principles to the messy arena of real world applications, delving into topics such as centering predictors, handling variably spaced measurement occasions and varying numbers of waves, including time-varying predictors, and fitting discontinuous and non-linear change trajectories. The target audience is professionals who have yet to fully exploit these longitudinal approaches. Some participants may be comfortable with multilevel modeling, although we assume no familiarity with the topic. Although methodological colleagues are not the prime audience, they, too, should find much of interest.

Book
The course is based on the first half of the instructors’ recent book, known by the acronym ALDA. You can learn more about ALDA at gseacademic.harvard.edu/~alda/. Participants are strongly encouraged to obtain copies of ALDA in advance of the workshop from either www.amazon.com or Oxford University Press www.oup.com. We are also investigating the possibility of having copies for sale at the event. Check with Tom Lane, tlane@mathworks.com, to determine if the book will be available at the course. ALDA is supported by a companion website at the UCLA Academic Technology Services, www.ats.ucla.edu/stat/examples/alda/. There you can download the many data sets used throughout the book and code for reproducing all the book’s analyses, using your preferred major software package.

Instructors
Judith D. Singer is the James Bryant Conant Professor of Education and John B. Willett is the Charles William Eliot Professor of Education, both at the Harvard Graduate School of Education. Singer holds a PhD in Statistics from Harvard University; Willett holds a PhD in Quantitative Methods from Stanford University. Collaborators for 20 years, their professional lives focus on improving the quantitative methods used in social, educational and behavioral research. Singer and Willett are best known for their contributions to the practice of individual growth modeling, survival analysis, and multilevel modeling, and to making these and other statistical methods accessible to empirical researchers. You can learn more about the instructors on their home pages: gseweb.harvard.edu/~faculty/singer/ and gseacademic.harvard.edu/~willetjo
The first longitudinal study of growth: Filibert Guéneau de Montbeillard (1720-1785)

Recorded his son’s height every six months from his birth in 1759 until his 18th birthday.


In most fields, the quantity of longitudinal research is exploding

Annual searches for keyword 'longitudinal' in 9 OVID databases, between 1982 and 2002.
Quality, however, can be another matter

First, the good news:
More longitudinal studies are being published, and an increasing percentage of these are "truly" longitudinal

<table>
<thead>
<tr>
<th>% longitudinal</th>
<th>'99</th>
<th>'03</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 waves</td>
<td>36%</td>
<td>26%</td>
</tr>
<tr>
<td>3 waves</td>
<td>26%</td>
<td>29%</td>
</tr>
<tr>
<td>4 or more waves</td>
<td>38%</td>
<td>45%</td>
</tr>
</tbody>
</table>

Now, the bad news:
Very few of these longitudinal studies use "modern" analytic methods

<table>
<thead>
<tr>
<th>Growth modeling</th>
<th>'99</th>
<th>'03</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survival analysis</td>
<td>2%</td>
<td>5%</td>
</tr>
<tr>
<td>Repeated measures ANOVA</td>
<td>40%</td>
<td>29%</td>
</tr>
<tr>
<td>Wave-to-wave regression</td>
<td>38%</td>
<td>32%</td>
</tr>
<tr>
<td>Separate but parallel analyses</td>
<td>8%</td>
<td>17%</td>
</tr>
<tr>
<td>&quot;Simplifying&quot; analyses by...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Setting aside waves</td>
<td>8%</td>
<td>7%</td>
</tr>
<tr>
<td>Combining waves</td>
<td>6%</td>
<td>8%</td>
</tr>
<tr>
<td>Ignoring age-heterogeneity</td>
<td>6%</td>
<td>9%</td>
</tr>
</tbody>
</table>

Part of the problem may well be reviewers’ ignorance

Comments received last year from two reviewers for Developmental Psychology of a paper that fit individual growth models to 3 waves of data on vocabulary size among young children:

Reviewer A:
“I do not understand the statistics used in this study deeply enough to evaluate their appropriateness. I imagine this is also true of 99% of the readers of Developmental Psychology. ... Previous studies in this area have used simple correlation or regression which provide easily interpretable values for the relationships among variables. ... In all, while the authors are to be applauded for a detailed longitudinal study, ... the statistics are difficult. ... I thus think Developmental Psychology is not really the place for this paper.”

Reviewer B:
“The analyses fail to live up to the promise...of the clear and cogent introduction. I will note as a caveat that I entered the field before the advent of sophisticated growth-modeling techniques, and they have always aroused my suspicion to some extent. I have tried to keep up and to maintain an open mind, but parts of my review may be naïve, if not inaccurate.”
**What kinds of research questions require longitudinal methods?**

- **Questions about systematic change over time**
  - Curran et al (1997) studied alcohol use:
    - 82 teens interviewed at ages 14, 15, & 16—alcohol use tended to increase over time.
    - Children of Alcoholics (COAs) drank more but had no steeper rates of increase over time.

- **Questions about whether and when events occur**
  - Capaldi et al (1996) studied age of 1st sex:
    - 180 boys interviewed annually from 7th to 12th grade (30% remained virgins at end of study).
    - Boys who experienced early parental transitions were more likely to have had sex.

1. **Within-person summary:** How does a teen's alcohol consumption change over time?
2. **Between-person comparison:** How do these trajectories vary by teen characteristics?

**Individual Growth Model/ Multilevel Model for Change**

**Discrete- and Continuous-Time Survival Analysis**

---

**Four important advantages of modern longitudinal methods**

1. **You have great flexibility in research design**
   - Not everyone needs the same rigid data collection schedule—cadence can be person specific.
   - Not everyone needs the same number of waves—can use all cases, even those with just one wave.
   - Design can be experimental or observational.
   - Designs can be single level (individuals only) or multilevel (e.g. students within classes/schools; physicians within hospitals).

2. **You can identify temporal patterns in the data**
   - Does the outcome increase, decrease, or remain stable over time?
   - Is the general pattern linear or non-linear?
   - Are there abrupt shifts at substantively interesting moments?

3. **You can include time varying predictors** (those whose values vary over time)
   - Participation in an intervention.
   - Family circumstances (income, parental status, etc).

4. **You can include interactions with time** (to test whether a predictor’s effect varies over time)
   - Some effects dissipate—they wear off.
   - Some effects increase—they become more important.
   - Some effects are especially pronounced at particular times.
What we’re going to cover today

3 Introducing the Multilevel Model for Change 45
  3.1 What is the Purpose of the Multilevel Model for Change? 46
  3.2 The Level 1 Model for Individual Change 49
  3.3 The Level 2 Model for Systematic Interindividual Differences in Change 57
  3.4 Fitting the Multilevel Model for Change to Data 63
  3.5 Examining Estimated Fixed Effects 68
  3.6 Examining Estimated Variance Components 72

4 Doing Data Analysis with the Multilevel Model for Change 75
  4.1 Example: Changes in Adolescent Alcohol Use 76
  4.2 The Composite Specification of the Multilevel Model for Change 80
  4.3 Methods of Estimation, Recapitulated 85
  4.4 First Steps: Placing Two Unconditional Multilevel Models for Change 92
  4.5 Practical Data Analysis: Strategies for Model Building 104
  4.6 Comparing Models Using Distance Statistics 116

5 Treating TIME More Flexibly 128
  5.1 Variable Spaced Measurement Occasions 139
  5.2 Varying Numbers of Measurement Occasions 146
  5.3 Time-Varying Parameters 159
  5.4 Revisiting the Effect of Taste 181

6 Modeling Discontinuous and Nonlinear Change 189
  6.1 Discontinuous Individual Change 196
  6.2 Using Transformations to Model Nonlinear Individual Change 198
  6.3 Representing Discontinuous Change Using a Polynomial Function of Time 213
  6.4 Truly Nonlinear Trajectories 232

Go to www.oup.com/us/singerwillettbook for a 20% discount coupon

© Singer & Willett, page 7

A word about programming and software

UCLA Academic Technology Services
www.ats.ucla.edu/stat/examples/alda

<table>
<thead>
<tr>
<th>Chapter</th>
<th>SPSS</th>
<th>SAS</th>
<th>S-Plus</th>
<th>Stata</th>
<th>HLM</th>
<th>MLwiN</th>
<th>Mplus</th>
<th>Table of contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ch 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A framework for investigating change over time</td>
</tr>
<tr>
<td>Ch 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exploring longitudinal data on change</td>
</tr>
<tr>
<td>Ch 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Introducing the multilevel model for change</td>
</tr>
<tr>
<td>Ch 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Doing data analysis with the multilevel model for change</td>
</tr>
<tr>
<td>Ch 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Treating time more flexibly</td>
</tr>
<tr>
<td>Ch 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Modeling discontinuous and nonlinear change</td>
</tr>
<tr>
<td>Ch 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Examining the multilevel model's error covariance structure</td>
</tr>
<tr>
<td>Ch 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Modeling change using covariance structure analysis</td>
</tr>
<tr>
<td>Ch 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A framework for investigating event occurrence</td>
</tr>
<tr>
<td>Ch 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Describing discrete-time event occurrence data</td>
</tr>
<tr>
<td>Ch 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fitting basic discrete-time hazard models</td>
</tr>
<tr>
<td>Ch 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Extending the discrete-time hazard model</td>
</tr>
<tr>
<td>Ch 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Describing continuous-time event occurrence data</td>
</tr>
<tr>
<td>Ch 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fitting the Cox regression model</td>
</tr>
<tr>
<td>Ch 15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Extending the Cox regression model</td>
</tr>
</tbody>
</table>

© Singer & Willett, page 8
Chapter 3: Introducing the multilevel model for change

General Approach: We’ll go through a worked example from start to finish saving practical data analytic advice for the next session

- The level-1 submodel for individual change (§3.2)—examining empirical growth trajectories and asking what population model might have given rise these observations?
- The level-2 submodels for systematic interindividual differences in change (§3.3)—what kind of population model should we hypothesize to represent the behavior of the parameters from the level-1 model?
- Fitting the multilevel model for change to data (§3.4)—there are now many options for model fitting, and more practically, many software options.
- Interpreting the results of model fitting (§3.5 and §3.6) Having fit the model, how do we sensibly interpret and display empirical results?
  - Interpreting fixed effects
  - Interpreting variance components
  - Plotting prototypical trajectories
Illustrative example: The effects of early intervention on children’s IQ


Sample: 103 African American children born to low income families
- 58 randomly assigned to an early intervention program
- 45 randomly assigned to a control group

Research design
- Each child was assessed 12 times between ages 6 and 96 months
- Here, we analyze only 3 waves of data, collected at ages 12, 18, and 24 months

Research question
- What is the effect of the early intervention program on children’s cognitive performance?

What does the person-period data set look like?

General structure:
A person-period data set has one row of data for each period when that particular person was observed

<table>
<thead>
<tr>
<th>ID</th>
<th>AGE</th>
<th>COG</th>
<th>PROGRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>68</td>
<td>1.0</td>
<td>105</td>
<td>1</td>
</tr>
<tr>
<td>68</td>
<td>1.5</td>
<td>119</td>
<td>1</td>
</tr>
<tr>
<td>68</td>
<td>2.0</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>1.0</td>
<td>106</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>1.5</td>
<td>107</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>2.0</td>
<td>96</td>
<td>1</td>
</tr>
<tr>
<td>71</td>
<td>1.0</td>
<td>112</td>
<td>1</td>
</tr>
<tr>
<td>71</td>
<td>1.5</td>
<td>86</td>
<td>1</td>
</tr>
<tr>
<td>71</td>
<td>2.0</td>
<td>78</td>
<td>1</td>
</tr>
<tr>
<td>72</td>
<td>1.0</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>72</td>
<td>1.5</td>
<td>93</td>
<td>1</td>
</tr>
<tr>
<td>72</td>
<td>2.0</td>
<td>87</td>
<td>1</td>
</tr>
<tr>
<td>902</td>
<td>1.0</td>
<td>119</td>
<td>0</td>
</tr>
<tr>
<td>902</td>
<td>1.5</td>
<td>95</td>
<td>0</td>
</tr>
<tr>
<td>905</td>
<td>2.0</td>
<td>99</td>
<td>0</td>
</tr>
<tr>
<td>905</td>
<td>1.0</td>
<td>112</td>
<td>0</td>
</tr>
<tr>
<td>904</td>
<td>1.5</td>
<td>98</td>
<td>0</td>
</tr>
<tr>
<td>904</td>
<td>2.0</td>
<td>79</td>
<td>0</td>
</tr>
<tr>
<td>906</td>
<td>1.0</td>
<td>89</td>
<td>0</td>
</tr>
<tr>
<td>906</td>
<td>1.5</td>
<td>66</td>
<td>0</td>
</tr>
<tr>
<td>906</td>
<td>2.0</td>
<td>51</td>
<td>0</td>
</tr>
<tr>
<td>909</td>
<td>1.0</td>
<td>117</td>
<td>0</td>
</tr>
<tr>
<td>909</td>
<td>1.5</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>909</td>
<td>2.0</td>
<td>76</td>
<td>0</td>
</tr>
</tbody>
</table>

PROGRAM is a dummy variable indicating whether the child was randomly assigned to the special early childhood program (1) or not (0)

COG is a nationally normed scale
- Declines within empirical growth records
- Instead of asking whether the growth rate is higher among program participants, we’ll ask whether the rate of decline is lower

Fully balanced, 3 waves per child
AGE=1.0, 1.5, and 2.0 (clocked in years—instead of months—so that we assess “annual rate of change”)
Examining empirical growth plots to help suggest a suitable individual growth model
(by superimposing fitted OLS trajectories)

Overall impression:
COG declines over time, but there’s some variation in the fit

Many trajectories are smooth and systematic
(70, 71, 72, 904, 908)

Other trajectories are scattered, irregular and perhaps curvilinear?
(68, 902, 906)

Q: What type of population individual growth model might have generated these sample data?
• Linear or curvilinear?
• Smooth or jagged?
• Continuous or disjoint?

Postulating a simple linear level-1 submodel for individual change:
Examining its structural and stochastic portions

Structural portion, which embodies our hypothesis about the shape of each person’s true trajectory of change over time

\[ COG_{ij} = [\pi_0] + [\pi_1] (AG_i - 1) + [E_{ij}] \]

Key assumption: In the population, COG, is a linear function of child i’s AGE on occasion j

Stochastic portion, which allows for the effects of random error from the measurement of person i on occasion j. Usually assume \( E_{ij} \sim N(0, \sigma^2_{ij}) \)

\( \pi_0 \) is the intercept of i’s true change trajectory, its true value of COG at AGE=1, its “true initial status”

\( \pi_1 \) is the slope of i’s true change trajectory, his yearly rate of change in true COG, his true “annual rate of change”
Experiencing fitted OLS trajectories to help suggest a suitable level-2 model

Most children decline over time (although there are a few exceptions)

Average OLS trajectory across the full sample
\[ \cong 110 - 10(\text{AGE}-1) \]

But there’s also great variation in these OLS estimates

<table>
<thead>
<tr>
<th>Fitted initial status</th>
<th>Fitted rate of change</th>
<th>Residual variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

What does this behavior mean for a level-2 model?
- The level-2 model must capture both the averages and the variation about these averages
- And...it must allow for systematic interindividual differences in change according to variation in predictor(s) (here, PROGRAM participation)

Further developing the level-2 submodel for interindividual differences in change

Four desired features of the level-2 submodel(s)

1. Outcomes are the level-1 individual growth parameters \( \pi_{0i} \) and \( \pi_{1i} \)
2. Need two level-2 submodels, one per growth parameter (one for initial status, one for change)
3. Each level-2 submodel must specify the relationship between a level-1 growth parameter and predictor(s), here PROGRAM
   - We need to specify a functional form for these relationships at level-2 (beginning with linear but ultimately becoming more flexible)
4. Each level-2 submodel should allow individuals with common predictor values to nevertheless have different individual change trajectories
   - We need stochastic variation at level-2, too
   - Each level-2 model will need its own error term, and we will need to allow for covariance across level-2 errors

Program participants tend to have:
- Higher scores at age 1 (higher initial status)
- Less steep rates of decline (shallower slopes)
- But these are only overall trends—there’s great interindividual heterogeneity
Level-2 submodels for systematic interindividual differences in change

For the level-1 intercept (initial status)
\[ \pi_{0i} = \gamma_{00} + \gamma_{01} \text{PROGRAM} + \zeta_{0i} \]

For the level-1 slope (rate of change)
\[ \pi_{1i} = \gamma_{10} + \gamma_{11} \text{PROGRAM} + \zeta_{1i} \]

Key to remembering subscripts on the gammas (the \( \gamma \)'s)
- First subscript indicates role in level-1 model (0 for intercept; 1 for slope)
- Second subscript indicates role in level-2 model (0 for intercept; 1 for slope)

What about the zetas (the \( \zeta \)'s)?
- They're level-2 residuals that permit the level-1 individual growth parameters to vary stochastically across people
- As with most residuals, we're less interested in their values than their population variances and covariances

Understanding the stochastic components of the level-2 submodels

General idea:
- Model posits the existence of an average population trajectory for each program group
- Because of the level-2 residuals, each child \( i \) has his own true change trajectory (defined by \( \pi_{0i} \) and \( \pi_{1i} \))
- Shading suggests the existence of many true population trajectories, one per child

Assumptions about the level-2 residuals:

\[
\begin{bmatrix}
\zeta_{0i} \\
\zeta_{1i}
\end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{10} & \sigma_{11} \end{bmatrix}\right)
\]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Illustrative interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \pi_{0i} )</td>
<td>Population average of the level-1 intercepts, ( \pi_{0i} ) for individuals with a level-2 predictor value of 0.</td>
<td>Population average true initial status for nonparticipants.</td>
</tr>
<tr>
<td>( \gamma_{00} )</td>
<td>Population average difference in level-1 intercept, ( \pi_{0i} ) for a 1-unit difference in the level-2 predictor.</td>
<td>Difference in population average true initial status between participants and nonparticipants.</td>
</tr>
<tr>
<td>( \gamma_{10} )</td>
<td>Population average of the level-1 slopes, ( \pi_{1i} ) for individuals with a level-2 predictor value of 0.</td>
<td>Population average annual rate of true change for nonparticipants.</td>
</tr>
<tr>
<td>( \gamma_{11} )</td>
<td>Population average difference in level-1 slope, ( \pi_{1i} ) for a 1-unit difference in the level-2 predictor.</td>
<td>Difference in population average annual rate of true change between participants and non-participants.</td>
</tr>
</tbody>
</table>

Assumptions about the level-2 residuals:

\[
\begin{bmatrix}
\zeta_{0i} \\
\zeta_{1i}
\end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{00} & \sigma_{01} \\ \sigma_{10} & \sigma_{11} \end{bmatrix}\right)
\]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
<th>Illustrative interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sigma_{0i} )</td>
<td>Level-2 residual variance in true intercept, ( \pi_{0i} ) across all individuals in the population.</td>
<td>Population residual variance of true initial status, controlling for program participation.</td>
</tr>
<tr>
<td>( \sigma_{1i} )</td>
<td>Level-2 residual variance in true slope, ( \pi_{1i} ) across all individuals in the population.</td>
<td>Population residual variance of true rate of change, controlling for program participation.</td>
</tr>
<tr>
<td>( \alpha_{01} )</td>
<td>Level-2 residual covariance between true intercept, ( \pi_{0i} ) and true slope, ( \pi_{1i} ) across all individuals in the population.</td>
<td>Population residual covariance between true initial status and true annual rate of change, controlling for program participation.</td>
</tr>
</tbody>
</table>
Fitting the multilevel model for change to data
Three general types of software options (whose numbers are increasing over time)

- Programs expressly designed for multilevel modeling
  - MLwiN
- Multipurpose packages with multilevel modeling modules
  - SAS, SPSS, STATA
- Specialty packages originally designed for another purpose that can also fit some multilevel models
  - Mplus, LIMDEP, LISREL, Egret, aML

Two sets of issues to consider when comparing (and selecting) packages

8 practical considerations
(that affect ease of use/pedagogic value)
- Data input options—level-1/level-2 vs. person-period; raw data or xyz.dataset
- Programming options—graphical interfaces and/or scripts
- Availability of other statistical procedures
- Model specification options—level-1/level-2 vs. composite; random effects
- Automatic centering options
- Wisdom of program’s defaults
- Documentation & user support
- Quality of output—text & graphics

8 technical considerations
(that affect research value)
- # of levels that can be handled
- Range of assumptions supported (for the outcomes & effects)
- Types of designs supported (e.g., cross-nested designs; latent variables)
- Estimation routines—full vs. restricted; ML vs. GLS—more on this later…
- Ability to handle design weights
- Quality and range of diagnostics
- Speed
- Strategies for handling estimation problems (e.g., boundary constraints)

Advice: Use whatever package you’d like but be sure to invest the time and energy to learn to use it well.

Visit http://www.ats.ucla.edu/stat/examples/alda for data, code in the major packages, and more
Examining estimated fixed effects

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Parameter</th>
<th>Estimate</th>
<th>( \text{a.e.} )</th>
<th>( z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial status, ( \pi_0 )</td>
<td>Intercept</td>
<td>( \gamma_0 )</td>
<td>107.84***</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td>PROGRAM</td>
<td>( \gamma_i )</td>
<td>6.85*</td>
<td>2.71</td>
</tr>
<tr>
<td>Rate of change, ( \pi_1 )</td>
<td>Intercept</td>
<td>( \gamma_0 )</td>
<td>-21.13***</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>PROGRAM</td>
<td>( \gamma_i )</td>
<td>5.27*</td>
<td>2.52</td>
</tr>
</tbody>
</table>

~\( p < .10; * p < .05; ** p < .01; *** p < .00 \)

In the population from which this sample was drawn we estimate that...

True initial status (COG at age 1) for the average non-participant is 107.84
For the average participant, it is 6.85 higher

Fitted model for initial status
\[ \hat{\pi}_{0i} = 107.84 + 6.85\text{PROGRAM}_i \]

Fitted model for rate of change
\[ \hat{\pi}_{1i} = -21.13 + 5.27\text{PROGRAM}_i \]

True annual rate of change for the average non-participant is -21.13
For the average participant, it is 5.27 higher

**Advice:** As you’re learning these methods, take the time to actually write out the fitted level-1/level-2 models before interpreting computer output—It’s the best way to learn what you’re doing!

Plotting prototypical change trajectories

**General idea:** Substitute prototypical values for the level-2 predictors (here, just \text{PROGRAM}=0 or 1) into the fitted models

\[ \hat{\pi}_{00} = 107.84 + 6.85(0) = 107.84 \]
\[ \hat{\pi}_{10} = -21.13 + 5.27(0) = -21.13 \]
so: \( \text{COG} = 107.84 - 21.13 \text{AGE} \)

\[ \hat{\pi}_{01} = 107.84 + 6.85(1) = 114.69 \]
\[ \hat{\pi}_{11} = -21.13 + 5.27(1) = -15.86 \]
so: \( \text{COG} = 114.69 - 15.86 \text{AGE} \)

**Tentative conclusion:** Program participants appear to have higher initial status and slower rates of decline.

**Question:** Might these differences be due to nothing more than sampling variation?
Testing hypotheses about fixed effects using single parameter tests

For initial status:
- Average non-participant had a non-zero level of COG at age 1 (surprise!)
- Program participants had higher initial status, on average, than non-participants (probably because the intervention had already started)

General formulation:
\[ z = \frac{\hat{\gamma}}{\text{ase}(\hat{\gamma})} \]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>se</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial status, ( \pi_0 ) Intercept</td>
<td>( \gamma_0 )</td>
<td>10.78***</td>
<td>2.04</td>
</tr>
<tr>
<td>PROGRAM</td>
<td>( \gamma_1 )</td>
<td>6.85*</td>
<td>2.71</td>
</tr>
<tr>
<td>Rate of change, ( \pi_{1t} ) Intercept</td>
<td>( \gamma_{0t} )</td>
<td>-21.13***</td>
<td>1.89</td>
</tr>
<tr>
<td>PROGRAM</td>
<td>( \gamma_{1t} )</td>
<td>5.23*</td>
<td>2.52</td>
</tr>
</tbody>
</table>

\( * p < .10; ** p < .05; *** p < .01; **** p < .00 \)

For rate of change:
- Average non-participant had a non-zero rate of decline (depressing)
- Program participants had slower rates of decline, on average, than non-participants (the "program effect").

Examining estimated variance components

General idea:
- Variance components quantify the amount of residual variation left—at either level-1 or level-2—that is potentially explainable by other predictors not yet in the model.
- Interpretation is easiest when comparing different models that each have different predictors (which we will soon do...).

Level-1 residual variance (74.24***):
- Summarizes within-person variability in outcomes around individuals’ own trajectories (usually non-zero)
- Here, we conclude there is some within-person residual variability

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>se</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1: Within-person, ( \varepsilon_t )</td>
<td>( \sigma^2 )</td>
<td>74.24***</td>
<td>10.34</td>
</tr>
<tr>
<td>Level 2: In initial status, ( \zeta_{0t} )</td>
<td>( \sigma^2 )</td>
<td>27.38</td>
<td>4.55</td>
</tr>
<tr>
<td>In rate of change, ( \zeta_{1t} )</td>
<td>( \sigma^2 )</td>
<td>12.29</td>
<td>4.00</td>
</tr>
<tr>
<td>Covariance between ( \zeta_{0t} ) and ( \zeta_{1t} )</td>
<td>( \alpha_{01} )</td>
<td>-36.41</td>
<td>22.74</td>
</tr>
</tbody>
</table>

\( * p < .10; ** p < .05; *** p < .01; **** p < .001 \)

Level-2 residual variance:
- Summarizes between-person variability in change trajectories (here, initial status and growth rates) after controlling for predictor(s) (here, PROGRAM)
- No residual variance in rates of change to be explained (nor is there a residual covariance)

© Singer & Willett, page 15

(ALDA, Section 3.5.2, pp. 71-72)

© Singer & Willett, page 16

(ALDA, Section 3.6, pp. 72-74)
Chapter 4: Doing data analysis with the multilevel model for change

General Approach: Once again, we'll go through a worked example, but now we'll delve into the practical data analytic details

- Composite specification of the multilevel model for change (§4.2) and how it relates to the level-1/level-2 specification just introduced
- First steps: unconditional means model and unconditional growth model (§4.4)
  - Intraclass correlation
  - Quantifying proportion of outcome variation “explained”
- Practical model building strategies (§4.5)
  - Developing and fitting a taxonomy of models
  - Displaying prototypical change trajectories
  - Recentering to improve interpretation
- Comparing models (§4.6)
  - Using deviance statistics
  - Using information criteria (AIC and BIC)
Illustrative example: The effects of parental alcoholism on adolescent alcohol use


Sample: 82 adolescents
- 37 are children of an alcoholic parent (COAs)
- 45 are non-COAs

Research design
- Each was assessed 3 times—at ages 14, 15, and 16
- The outcome, ALCUSE, was computed as follows:
  - 4 items: (1) drank beer/wine; (2) hard liquor; (3) 5 or more drinks in a row; and (4) got drunk
  - Each item was scored on an 8 point scale (0=“not at all” to 7=“every day”)
  - ALCUSE is the square root of the sum of these 4 items
- At age 14, PEER, a measure of peer alcohol use was also gathered

Research question
- Do trajectories of adolescent alcohol use differ by:
  - (1) parental alcoholism; and (2) peer alcohol use?

What’s an appropriate functional form for the level-1 submodel?
(Examining empirical growth plots with superimposed OLS trajectories)

3 features of these plots:
1. Often approximately linear (but not always increasing over time)
2. Some OLS trajectories fit well (23, 32, 56, 65)
3. Other OLS trajectories show more scatter (04, 14, 41, 82)

A linear model makes sense...

$ALCUSE_{ij} = \pi_{0i} + \pi_{1i}(AGE_{ij} - 14) + \epsilon_{ij}$ where $\epsilon_{ij} \sim N(0, \sigma^2)$

$Y_{ij} = \pi_{0i} + \pi_{1i}TIME_{ij} + \epsilon_{ij}$

- $i$’s true initial status (ie, when $TIME=0$)
- $i$’s true rate of change per unit of $TIME$
- Portion of $i$’s outcome unexplained on occasion $j$

(ALDA, Section 4.1, pp.76-80)
Specifying the level-2 submodels for individual differences in change

Examining variation in OLS-fitted level-1 trajectories by:
- **COA**: COAs have higher intercepts but no steeper slopes
- **PEER (split at mean)**: Teens whose friends at age 14 drink more have higher intercepts but shallower slopes

**Level-2 intercepts**
Population average initial status and rate of change for a non-COA

\[
\pi_{0i} = \gamma_{00} + \gamma_{01} \text{COA}_i + \xi_{0i}
\]
(for initial status)

**Level-2 slopes**
Effect of COA on initial status and rate of change

\[
\pi_{1i} = \gamma_{10} + \gamma_{11} \text{COA}_i + \xi_{1i}
\]
(for rate of change)

**Level-2 residuals**
Deviations of individual change trajectories around predicted averages

\[
[\begin{bmatrix} \xi_{0i} \\ \xi_{1i} \end{bmatrix}] \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} , \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix} \right)
\]

(AlDA, Section 4.1, pp. 76-80)

Developing the composite specification of the multilevel model for change by substituting the level-2 submodels into the level-1 individual growth model

\[
\pi_{0i} = \gamma_{00} + \gamma_{01} \text{COA}_i + \xi_{0i}
\]
\[
\pi_{1i} = \gamma_{10} + \gamma_{11} \text{COA}_i + \xi_{1i}
\]
\[
Y_{ij} = \pi_{0i} + \pi_{1i} \text{TIME}_{ij} + \epsilon_{ij}
\]
\[
Y_{ij} = (\gamma_{00} + \gamma_{01} \text{COA}_i + \xi_{0i}) + (\gamma_{10} + \gamma_{11} \text{COA}_i + \xi_{1i}) \text{TIME}_{ij} + \epsilon_{ij}
\]
\[
Y_{ij} = \gamma_{00} + \gamma_{01} \text{COA}_i + \gamma_{10} \text{TIME}_{ij} + \gamma_{11} \text{COA}_i \text{TIME}_{ij}
\]
\[
+ \xi_{0i} + \xi_{1i} \text{TIME}_{ij} + \epsilon_{ij}
\]

**The composite specification shows how the outcome depends simultaneously on:**
- the level-1 predictor TIME and the level-2 predictor COA as well as
- the cross-level interaction, COA*TIME

This tells us that the effect of one predictor (TIME) differs by the levels of another predictor (COA)

(AlDA, Section 4.2, pp. 80-83)

© Singer & Willett, page 5

© Singer & Willett, page 6
The person-period data set and its relationship to the composite specification

\[ ALCUSE_{ij} = \gamma_0 + \gamma_1 (AGE - 14)_{ij} + \gamma_0 COA + \gamma_1 (COA \times (AGE - 14))_{ij} + [\xi_0 + \xi_1 (AGE - 14)_{ij} + \epsilon_{ij}] \]

<table>
<thead>
<tr>
<th>ID</th>
<th>ALCUSE</th>
<th>AGE-14</th>
<th>COA</th>
<th>COA*(AGE-14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1.00</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>2.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3.32</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2.00</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1.73</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>44</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>1.41</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>44</td>
<td>3.00</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>66</td>
<td>1.41</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>66</td>
<td>3.46</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>66</td>
<td>3.00</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Words of advice before beginning data analysis

Be sure you’ve examined empirical growth plots and fitted OLS trajectories. You don’t want to begin data analysis without being reasonably confident that you have a sound level-1 model.

Be sure your person-period data set is correct.

- Run simple diagnostics in whatever general purpose program you’re comfortable with.
- Once again, you don’t want to invest too much data analytic effort in a mis-formed data set.

Don’t jump in by fitting a range of models with substantive predictors. Yes, you want to know “the answer,” but first you need to understand how the data behave, so instead you should...

First steps: Two unconditional models

1. Unconditional means model—a model with no predictors at either level, which will help partition the total outcome variation.
2. Unconditional growth model—a model with TIME as the only level-1 predictor and no substantive predictors at level 2, which will help evaluate the baseline amount of change.

What these unconditional models tell us:

1. Whether there is systematic variation in the outcome worth exploring and, if so, where that variation lies (within or between people).
2. How much total variation there is both within- and between-persons, which provides a baseline for evaluating the success of subsequent model building (that includes substantive predictors).

(ALDA, Section 4.4, p. 92+)
The Unconditional Means Model (Model A)  
Partitioning total outcome variation between and within persons

Level-1 Model: \( Y_{ij} = \pi_0 + \varepsilon_{ij} \), where \( \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \)

Level-2 Model: \( \pi_{0i} = \gamma_{00} + \zeta_0 + \zeta_{0i} \), where \( \zeta_{0i} \sim N(0, \sigma_\zeta^2) \)

Composite Model: \( Y_{ij} = \gamma_{00} + \zeta_0 + \zeta_{0i} + \varepsilon_{ij} \)

Let's look more closely at these variances....

Intraclass correlation compares the relative magnitude of these VCs by estimating the proportion of total variation in Y that lies “between” people

\[ \rho = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_{\varepsilon}^2} \]

An estimated 50% of the total variation in alcohol use is attributable to differences between adolescents

But what role does TIME play?

(ALDA, Section 4.4.1, p. 92-97)
The Unconditional Growth Model (Model B)
A baseline model for change over time

Level-1 Model: \[ Y_{ij} = \beta_{0i} + \beta_{1i} \text{TIME}_{ij} + \epsilon_{ij} \], where \( \epsilon_{ij} \sim N(0, \sigma^2_{\epsilon}) \)

Level-2 Model:
\[ \pi_{0i} = \gamma_{00} + \zeta_{0i} \]
\[ \pi_{1i} = \gamma_{10} + \zeta_{1i} \]
where \( \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\zeta 0} & \sigma_{\zeta 10} \\ \sigma_{10} & \sigma^2_{\zeta 1} \end{bmatrix} \right) \)

Composite Model:
\[ Y_{ij} = \gamma_{00} + \gamma_{10} \text{TIME}_{ij} + \left( \zeta_{0i} + \zeta_{1i} \text{TIME}_{ij} + \epsilon_{ij} \right) \]

Average true initial status at AGE 14
Average true rate of change

Parameter Model B

| Fixed Effects |  
| Initial status: \( \pi_{0i} \) | Intercept \( \hat{\gamma}_{00} \), \( 0.651^{* * * } \), \( 0.105 \)  
| Rate of change, \( \pi_{1i} \) | Intercept \( \hat{\gamma}_{10} \), \( 0.271^{* * * } \), \( 0.069 \) |

What about the variance components?

The unconditional growth model: Interpreting the variance components

Level-1 (within person):
There is still unexplained within-person residual variance

Level-2 (between-persons):
- There is between-person residual variance in initial status (but careful, because initial status has changed)
- There is between-person residual variance in rate of change (should consider adding a level-2 predictor)
- Estimated res. covariance between initial status and change is n.s.

So...what has been the effect of moving from an unconditional means model to an unconditional growth model?

(ALDA, Section 4.4.2, pp 97-102)
Quantifying the proportion of outcome variation explained

For later: Extending the idea of proportional reduction in variance components to Level-2 (to estimate the percentage of between-person variation in ALCUSE associated with predictors)

\[
R^2 = \left( \frac{\sigma^2_\text{Total} - \sigma^2_\text{Model}}{\sigma^2_\text{Total}} \right) = 0.40
\]

40% of the within-person variation in ALCUSE is associated with linear time

\[
R^2_{Y,\hat{Y}} = \left( \frac{\sigma^2_\text{Uncond Growth Model} - \sigma^2_{\hat{Y}} \text{ (Later Growth Model)}}{\sigma^2_\text{Uncond Growth Model}} \right) = 0.043
\]

4.3% of the total variation in ALCUSE is associated with linear time

Careful: Don’t do this comparison with the unconditional means model.

Where we’ve been and where we’re going...

What these unconditional models tell us:
1. About half the total variation in ALCUSE is attributable to differences among teens
2. About 40% of the within-teen variation in ALCUSE is explained by linear TIME
3. There is significant variation in both initial status and rate of change—so it pays to explore substantive predictors (COA & PEER)

How do we build statistical models?
- Use all your intuition and skill you bring from the cross sectional world
  - Examine the effect of each predictor separately
  - Prioritize the predictors,
    - Focus on your “question” predictors
    - Include interesting and important control predictors
  - Progress towards a “final model” whose interpretation addresses your research questions

But because the data are longitudinal, we have some other options...
- Multiple level-2 outcomes (the individual growth parameters)—each can be related separately to predictors
- Two kinds of effects being modeled:
  - Fixed effects
  - Variance components
  - Not all effects are required in every model
What will our analytic strategy be?

Because our research interest focuses on the effect of COA, essentially treating PEER as a control, we’re going to proceed as follows...

Model C: COA predicts both initial status and rate of change.

Model D: Adds PEER to both Level-2 sub-models in Model C.

Model E: Simplifies Model D by removing the non-significant effect of COA on change.

Model C: Assessing the uncontrolled effects of COA (the question predictor)

Fixed effects

- Est. initial value of ALCUSE for non-COA is 0.316 (p<.001)
- Est. differential in initial ALCUSE between COAs and non-COAs is 0.743 (p<.001)
- Est. annual rate of change in ALCUSE for non-COAs is 0.293 (p<.001)
- Estimated differential in annual rate of change between COAs and non-COAs is −0.049 (ns)

Variance components

- Within person VC is identical to B’s because no predictors were added
- Initial status VC declines from B: COA “explains” 22% of variation in initial status (but still stat sig, suggesting need for level-2 pred’s)
- Rate of change VC unchanged from B: COA “explains” no variation in change (but also still sig suggesting need for level-2 pred’s)

Next step?

- Remove COA? Not yet—question predictor
- Add PEER—Yes, to examine controlled effects of COA

(Model C, Section 4.5.2, pp 107-108)
Model D: Assessing the controlled effects of COA (the question predictor)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model C</th>
<th>Model D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial status, ( \mu_0 )</td>
<td>( \gamma_0 )</td>
<td>( \gamma_0 )</td>
</tr>
<tr>
<td></td>
<td>( 0.316^{***} )</td>
<td>( -0.317^{***} )</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.148)</td>
</tr>
<tr>
<td>COA</td>
<td>( \gamma_1 )</td>
<td>( \gamma_1 )</td>
</tr>
<tr>
<td></td>
<td>0.745***</td>
<td>0.570***</td>
</tr>
<tr>
<td></td>
<td>(0.195)</td>
<td>(0.160)</td>
</tr>
<tr>
<td>PEER</td>
<td>( \gamma_2 )</td>
<td>( \gamma_2 )</td>
</tr>
<tr>
<td></td>
<td>0.694***</td>
<td>0.694***</td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.112)</td>
</tr>
<tr>
<td>Rate of change, ( \mu_v )</td>
<td>( \gamma_0 )</td>
<td>( \gamma_0 )</td>
</tr>
<tr>
<td></td>
<td>0.296***</td>
<td>0.425***</td>
</tr>
<tr>
<td></td>
<td>(0.084)</td>
<td>(0.144)</td>
</tr>
<tr>
<td>COA</td>
<td>( \gamma_1 )</td>
<td>( \gamma_1 )</td>
</tr>
<tr>
<td></td>
<td>-0.049</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
<td>(0.125)</td>
</tr>
<tr>
<td>PEER</td>
<td>( \gamma_2 )</td>
<td>( \gamma_2 )</td>
</tr>
<tr>
<td></td>
<td>-0.150</td>
<td>-0.150</td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(0.086)</td>
</tr>
<tr>
<td>Variance Components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>( \sigma_i^2 )</td>
<td>( \sigma_i^2 )</td>
</tr>
<tr>
<td>Within person</td>
<td>0.337***</td>
<td>0.337***</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>In initial status</td>
<td>( \sigma_i^2 )</td>
<td>( \sigma_i^2 )</td>
</tr>
<tr>
<td></td>
<td>0.488**</td>
<td>0.241**</td>
</tr>
<tr>
<td></td>
<td>(0.128)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>In rate of change</td>
<td>( \sigma_i^2 )</td>
<td>( \sigma_i^2 )</td>
</tr>
<tr>
<td></td>
<td>0.151*</td>
<td>0.139*</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Covariance</td>
<td>( \sigma_{i,j} )</td>
<td>( \sigma_{i,j} )</td>
</tr>
<tr>
<td></td>
<td>0.056</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Pseudo R² Statistics and Goodness-of-fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>.150</td>
<td>.291</td>
</tr>
<tr>
<td>AIC</td>
<td>621.2</td>
<td>588.7</td>
</tr>
<tr>
<td>BIC</td>
<td>657.5</td>
<td>628.7</td>
</tr>
<tr>
<td>Deviance</td>
<td>614.0</td>
<td>614.0</td>
</tr>
<tr>
<td>AIC</td>
<td>608.7</td>
<td>608.7</td>
</tr>
<tr>
<td>BIC</td>
<td>632.8</td>
<td>628.4</td>
</tr>
</tbody>
</table>

\(-p < .10; * p < .05; ** p < .01; ***p < .001\)

Next step?

- If we had other predictors, we’d add them because the VCs are still significant
- Simplify the model? Since COA is not associated with rate of change, why not remove this term from the model?

Model E: Removing the non-significant effect of COA on rate of change

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model D</th>
<th>Model E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial status, ( \mu_0 )</td>
<td>( \gamma_0 )</td>
<td>( \gamma_0 )</td>
</tr>
<tr>
<td></td>
<td>(-0.317^{***} )</td>
<td>(-0.314^{***} )</td>
</tr>
<tr>
<td></td>
<td>(0.148)</td>
<td>(0.148)</td>
</tr>
<tr>
<td>COA</td>
<td>( \gamma_1 )</td>
<td>( \gamma_1 )</td>
</tr>
<tr>
<td></td>
<td>0.579***</td>
<td>0.571***</td>
</tr>
<tr>
<td></td>
<td>(0.162)</td>
<td>(0.146)</td>
</tr>
<tr>
<td>PEER</td>
<td>( \gamma_2 )</td>
<td>( \gamma_2 )</td>
</tr>
<tr>
<td></td>
<td>0.694***</td>
<td>0.695***</td>
</tr>
<tr>
<td></td>
<td>(0.112)</td>
<td>(0.111)</td>
</tr>
<tr>
<td>Rate of change, ( \mu_v )</td>
<td>( \gamma_0 )</td>
<td>( \gamma_0 )</td>
</tr>
<tr>
<td></td>
<td>0.429***</td>
<td>0.425***</td>
</tr>
<tr>
<td></td>
<td>(0.114)</td>
<td>(0.106)</td>
</tr>
<tr>
<td>COA</td>
<td>( \gamma_1 )</td>
<td>( \gamma_1 )</td>
</tr>
<tr>
<td></td>
<td>-0.014</td>
<td>-0.014</td>
</tr>
<tr>
<td></td>
<td>(0.125)</td>
<td>(0.111)</td>
</tr>
<tr>
<td>PEER</td>
<td>( \gamma_2 )</td>
<td>( \gamma_2 )</td>
</tr>
<tr>
<td></td>
<td>-0.151</td>
<td>-0.151</td>
</tr>
<tr>
<td></td>
<td>(0.086)</td>
<td>(0.085)</td>
</tr>
<tr>
<td>Variance Components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1</td>
<td>( \sigma_i^2 )</td>
<td>( \sigma_i^2 )</td>
</tr>
<tr>
<td>Within person</td>
<td>0.337***</td>
<td>0.347***</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>In initial status</td>
<td>( \sigma_i^2 )</td>
<td>( \sigma_i^2 )</td>
</tr>
<tr>
<td></td>
<td>0.241**</td>
<td>0.241**</td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>In rate of change</td>
<td>( \sigma_i^2 )</td>
<td>( \sigma_i^2 )</td>
</tr>
<tr>
<td></td>
<td>0.159*</td>
<td>0.139*</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Covariance</td>
<td>( \sigma_{i,j} )</td>
<td>( \sigma_{i,j} )</td>
</tr>
<tr>
<td></td>
<td>-0.006</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>Pseudo R² Statistics and Goodness-of-fit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NS</td>
<td>.291</td>
<td>.291</td>
</tr>
<tr>
<td>AIC</td>
<td>621.2</td>
<td>688.7</td>
</tr>
<tr>
<td>BIC</td>
<td>657.5</td>
<td>606.7</td>
</tr>
<tr>
<td>Deviance</td>
<td>614.0</td>
<td>614.0</td>
</tr>
<tr>
<td>AIC</td>
<td>608.7</td>
<td>608.7</td>
</tr>
<tr>
<td>BIC</td>
<td>632.8</td>
<td>628.4</td>
</tr>
</tbody>
</table>

\(-p < .10; * p < .05; ** p < .01; ***p < .001\)
Where we’ve been and where we’re going…

- Let’s call Model E our tentative “final model” (based on not just these results but many other analyses not shown here)
- Controlling for the effects of PEER, the estimated differential in ALCUSE between COAs and nonCOAs is 0.571 (p<.001)
- Controlling for the effects of COA, for each 1-pt difference in PEER: the average initial ALCUSE is 0.695 higher (p<.001) and average rate of change is 0.151 lower (p<.10)

Displaying prototypical trajectories
Recentering predictors to improve interpretation
Alternative strategies for hypothesis testing: Comparing models using Deviance statistics and information criteria
Additional comments about estimation

Displaying analytic results: Constructing prototypical fitted plots

Key idea: Substitute prototypical values for the predictors into the fitted models to yield prototypical fitted growth trajectories

Review of the basic approach (with one dichotomous predictor)

Model C:
\[
\hat{\pi}_0 = 0.316 + 0.743 \text{COA} \\
\hat{\pi}_1 = 0.293 - 0.049 \text{COA}
\]

1. Substitute observed values for COA (0 and 1)
   - When COA = 0: \[
   \hat{\pi}_0 = 0.316 + 0.743(0) = 0.316 \\
   \hat{\pi}_1 = 0.293 - 0.049(0) = 0.293
   \]
   - When COA = 1: \[
   \hat{\pi}_0 = 0.316 + 0.743(1) = 1.059 \\
   \hat{\pi}_1 = 0.293 - 0.049(1) = 0.244
   \]

2. Substitute the estimated growth parameters into the level-1 growth model
   - when COA = 0: \[
   \hat{Y}_0 = 0.316 + 0.293 \text{TIME}
   \]
   - when COA = 1: \[
   \hat{Y}_0 = 1.059 + 0.244 \text{TIME}
   \]

What happens when the predictors aren’t all dichotomous?
Constructing prototypical fitted plots when some predictors are continuous

Key idea: Substitute “interesting” values of the continuous predictors into the fitted model and plot prototypical trajectories, by choosing
- Substantively interesting values (e.g., 12 and 16 yrs of education)
- A sensible range of percentiles (e.g., 10th, 50th, and 90th)
- The sample mean ± .5 (or 1) standard deviation
- The sample mean itself if you want to simply control for a predictor’s effect (instead of displaying its effect)

PEER: mean = 1.018, sd = 0.726

Low PEER: 1.018 - 0.695(0.726) = 0.655
Low PEER: 1.018 + 0.695(0.726) = 1.381

Model E

\[ \hat{y}_i = -0.314 + 0.695 \text{PEER} + 0.571 \text{COA} \]

\[ \hat{y}_i = 0.425 - 0.151 \text{PEER} \]

Intercepts for plotting
Slopes for plotting

What’s the effect of re-centering predictors?

At level-1, re-centering TIME is usually beneficial
- Ensures that the individual intercepts are easily interpretable, corresponding to status at a specific age
- Often use “initial status,” but as we’ll see, we can center TIME on any sensible value

Model F centers only PEER
Model G centers PEER and COA

As expected, centering the level-2 predictors changes the level-2 intercepts

Our preference: Here we prefer model F because it leaves the dichotomous question predictor COA uncentered

© Singer & Willett, page 22
Hypothesis testing: What we’ve been doing and an alternative approach

Single parameter hypothesis tests
- Simple to conduct and easy to interpret—making them very useful in hands-on data analysis (as we’ve been doing)
- However, statisticians disagree about their nature, form, and effectiveness
- Disagreement is so strong that some software packages (e.g., MLwiN) won’t output them
- Their behavior is poorest for tests on variance components

Deviance based hypothesis tests
- Based on the log likelihood (LL) statistic that is maximized under Maximum Likelihood estimation
- Have superior statistical properties (compared to the single parameter tests)
- Special advantage: permit joint tests on several parameters simultaneously
- You need to do the tests “manually” because automatic tests are rarely what you want

Deviance = -2[LL_{current model} – LL_{saturated model}]

Quantifies how much worse the current model is in comparison to a saturated model
A model with a small deviance statistic is nearly as good; a model with large deviance statistic is much worse (we obviously prefer models with smaller deviance)

Simplification: Because a saturated model fits perfectly, its LL = 0 and the second term drops out, making Deviance = -2LL_{current}

Hypothesis testing using Deviance statistics

You can use deviance statistics to compare two models if two criteria are satisfied:
1. Both models are fit to the same exact data—beware missing data
2. One model is nested within the other—we can specify the less complex model (e.g., A) by imposing constraints on one or more parameters in the more complex model (e.g., B), usually, but not always, setting them to 0)

If these conditions hold, then:
- Difference in the two deviance statistics is asymptotically distributed as \( \chi^2 \)
- \( df = \) # of independent constraints

1. We can obtain Model A from Model B by invoking 3 constraints:
   \[ H_0: \sigma_{11} = 0, \sigma_{12} = 0, \sigma_{01} = 0 \]

2. Compute difference in Deviance statistics and compare to appropriate \( \chi^2 \) distribution
   \[ \Delta \text{Deviance} = 33.55 \text{ (3 df, } p < .001 \) \]
   \( \Rightarrow \) reject \( H_0 \)
Using deviance statistics to test more complex hypotheses

Key idea: Deviance statistics are great for simultaneously evaluating the effects of adding predictors to both level-2 models.

We can obtain Model B from Model C by invoking 2 constraints:

\[ H_0 : \gamma_{01} = 0, \gamma_{11} = 0 \]

2: Compute difference in Deviance statistics and compare to appropriate \( \chi^2 \) distribution

\[ \Delta \text{Deviance} = 15.41 (2 \text{ df}, p < .001) \]

\[ \Rightarrow \text{reject } H_0 \]

The pooled test does not imply that each level-2 slope is on its own statistically significant.

Comparing non-nested multilevel models using AIC and BIC

Information Criteria: AIC and BIC

You can (supposedly) compare non-nested multilevel models using information criteria:

- The AIC penalty accounts for the number of parameters in the model.
- The BIC penalty goes further and also accounts for sample size.

Smaller values of AIC & BIC indicate better fit.

Here’s the taxonomy of multilevel models that we ended up fitting, in the ALCUSE example.....

Model E has the lowest AIC and BIC statistics.

Interpreting differences in BIC across models (Raftery, 1995):

- 0-2: Weak evidence
- 2-6: Positive evidence
- 6-10: Strong evidence
- >10: Very strong

Models need not be nested, but datasets must be the same.

© Singer & Willett, page 25

© Singer & Willett, page 26
A final comment about estimation and hypothesis testing

Two most common methods of estimation

Maximum likelihood (ML):
Seeks those parameter estimates that maximize the likelihood function, which assess the joint probability of simultaneously observing all the sample data actually obtained (implemented, e.g., in HLM and SAS Proc Mixed).

Generalized Least Squares (GLS): (Iterative GLS)
Iteratively seeks those parameter estimates that minimize the sum of squared residuals (allowing them to be autocorrelated and heteroscedastic) (implemented, e.g., in MLwiN and stata xreg).

A more important distinction: Full vs. Restricted (ML or GLS)

Full: Simultaneously estimate the fixed effects and the variance components.
- Default in MLwiN & HLM

Restricted: Sequentially estimate the fixed effects and then the variance components
- Default in SAS Proc Mixed & stata xtmixed

Goodness of fit statistics apply to the entire model (both fixed and random effects)
This is the method we’ve used in both the examples shown so far

Goodness of fit statistics apply to only the random effects
So we can only test hypotheses about VCs (and the models being compared must have identical fixed effects)

Other topics covered in Chapter Four

- Using Wald statistics to test composite hypotheses about fixed effects (§4.7)—generalization of the “parameter estimate divided by its standard error” approach that allows you to test composite hypotheses about fixed effects, even if you’ve used restricted estimation methods

- Evaluating the tenability of the model’s assumptions (§4.8)
  - Checking functional form
  - Checking normality
  - Checking homoscedasticity

- Model-Based (empirical Bayes) estimates of the individual growth parameters (§4.9) Superior estimates that combine OLS estimates with population average estimates that are usually your best bet if you would like to display individual growth trajectories for particular sample members
Chapter 5: Treating TIME more flexibly

General idea: Although all our examples have been equally spaced, time-structured, and fully balanced, the multilevel model for change is actually far more flexible

- Variously spaced measurement occasions (§5.1)—each individual can have his or her own customized data collection schedule
- Varying numbers of waves of data (§5.2)—not everyone need have the same number of waves of data
  - Allows us to handle missing data
  - Can even include individuals with just one or two waves
- Including time-varying predictors (§5.3)
  - The values of some predictors vary over time
  - They’re easy to include and can have powerful interpretations
- Re-centering the effect of TIME (§5.4)
  - Initial status is not the only centering constant for TIME
  - Recentering TIME in the level-1 model improves interpretation in the level-2 model

“Change is a measure of time”
Edwin Way Teale
### Example for handling variably spaced waves: Reading achievement over time

**Data source:** Children of the National Longitudinal Survey of Youth (CNLSY)

- **Sample:** 89 children
  - Each approximately 6 years old at study start

- **Research design**
  - 3 waves of data collected in 1986, 1988, and 1990, when the children were to be “in their 6th yr,” “in their 8th yr,” and “in their 10th yr”
  - Of course, not each child was tested on his/her birthday or half-birthday, which creates the variably spaced waves
  - The outcome, PIAT, is the child’s unstandardized score on the reading portion of the Peabody Individual Achievement Test
    - Not standardized for age so we can see growth over time
    - No substantive predictors to keep the example simple

- **Research question**
  - How do PIAT scores change over time?

### What does the person-period data set look like when waves are variably spaced?

<table>
<thead>
<tr>
<th>ID</th>
<th>WAVE</th>
<th>AGEGRP</th>
<th>AGE</th>
<th>PIAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>04</td>
<td>1</td>
<td>6.5</td>
<td>6.00</td>
<td>18</td>
</tr>
<tr>
<td>04</td>
<td>2</td>
<td>8.5</td>
<td>8.60</td>
<td>31</td>
</tr>
<tr>
<td>04</td>
<td>3</td>
<td>10.5</td>
<td>10.67</td>
<td>50</td>
</tr>
<tr>
<td>27</td>
<td>1</td>
<td>8.5</td>
<td>8.20</td>
<td>19</td>
</tr>
<tr>
<td>27</td>
<td>2</td>
<td>10.5</td>
<td>10.92</td>
<td>57</td>
</tr>
<tr>
<td>31</td>
<td>1</td>
<td>8.5</td>
<td>8.83</td>
<td>34</td>
</tr>
<tr>
<td>31</td>
<td>2</td>
<td>10.5</td>
<td>10.92</td>
<td>57</td>
</tr>
<tr>
<td>31</td>
<td>3</td>
<td>10.5</td>
<td>10.92</td>
<td>57</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>6.5</td>
<td>6.55</td>
<td>19</td>
</tr>
<tr>
<td>35</td>
<td>2</td>
<td>8.5</td>
<td>8.92</td>
<td>34</td>
</tr>
<tr>
<td>35</td>
<td>3</td>
<td>10.5</td>
<td>10.75</td>
<td>29</td>
</tr>
<tr>
<td>41</td>
<td>1</td>
<td>8.5</td>
<td>8.35</td>
<td>28</td>
</tr>
<tr>
<td>41</td>
<td>2</td>
<td>10.5</td>
<td>10.83</td>
<td>36</td>
</tr>
<tr>
<td>49</td>
<td>1</td>
<td>6.5</td>
<td>6.59</td>
<td>19</td>
</tr>
<tr>
<td>49</td>
<td>2</td>
<td>8.5</td>
<td>8.75</td>
<td>32</td>
</tr>
<tr>
<td>49</td>
<td>3</td>
<td>10.5</td>
<td>10.67</td>
<td>48</td>
</tr>
<tr>
<td>69</td>
<td>1</td>
<td>6.5</td>
<td>6.67</td>
<td>25</td>
</tr>
<tr>
<td>69</td>
<td>2</td>
<td>8.5</td>
<td>8.17</td>
<td>47</td>
</tr>
<tr>
<td>69</td>
<td>3</td>
<td>10.5</td>
<td>11.35</td>
<td>45</td>
</tr>
<tr>
<td>77</td>
<td>1</td>
<td>8.5</td>
<td>8.83</td>
<td>17</td>
</tr>
<tr>
<td>77</td>
<td>2</td>
<td>10.5</td>
<td>10.60</td>
<td>28</td>
</tr>
<tr>
<td>87</td>
<td>1</td>
<td>6.5</td>
<td>6.92</td>
<td>28</td>
</tr>
<tr>
<td>87</td>
<td>2</td>
<td>8.5</td>
<td>9.42</td>
<td>49</td>
</tr>
<tr>
<td>87</td>
<td>3</td>
<td>10.5</td>
<td>11.50</td>
<td>64</td>
</tr>
</tbody>
</table>

Person-period data sets are easy to construct even with variably spaced waves.

Three different ways of coding TIME:

- **WAVE**—reflects design but has no substantive meaning
- **AGEGRP**—child’s “expected” age on each occasion
- **AGE**—child’s actual age (to the day) on each occasion—notice “occasion creep”—later waves are more likely to be even later in a child’s life

© Singer & Willett, page 4
Comparing OLS trajectories fit using AGEGRP and AGE

For many children—especially those assessed near the half-years—it makes little difference. For some children though—there’s a big difference in slope, which is our conceptual outcome (rate of change).

Why ever use rounded AGE?
Note that this what we did in the past two examples, and so do lots of researchers!!!

Verifying models fit with AGEGRP and AGE

Level-1 Model: \[ Y_{ij} = \pi_0 + \pi_1 \text{TIME}_{ij} + \epsilon_{ij}, \] where \( \epsilon_{ij} \sim N(0, \sigma^2) \)

Level-2 Model: \[
\begin{align*}
\pi_0 &= \gamma_00 + \zeta_00 \\
\pi_1 &= \gamma_{10} + \zeta_{10} \\
\end{align*}
\]

where \[
\begin{bmatrix}
\zeta_{00} \\
\zeta_{10}
\end{bmatrix}
\sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)
\]

Composite Model: \[ Y_{ij} = \gamma_{00} + \gamma_{10} \text{TIME}_{ij} + [\zeta_{0i} + \zeta_{1i} \text{TIME}_{ij} + \epsilon_{ij}] \]

AIC and BIC better with AGE
Example for handling varying numbers of waves: Wages of HS dropouts

**Data source:** Murnane, Boudett and Willett (1999), *Evaluation Review*

- **Sample:** 888 male high school dropouts
  - Based on the National Longitudinal Survey of Youth (NLSY)
  - Tracked from first job since HS dropout, when the men varied in age from 14 to 17

- **Research design**
  - Each interviewed between 1 and 13 times
    - Interviews were approximately annual, but some were every 2 years
    - Each wave’s interview conducted at different times during the year
  - Both variable number and spacing of waves
  - Outcome is log(WAGES), inflation adjusted natural logarithm of hourly wage

- **Research question**
  - How do log(WAGES) change over time?
  - Do the wage trajectories differ by ethnicity and highest grade completed?

---

**Examining a person-period data set with varying numbers of waves of data per person**

<table>
<thead>
<tr>
<th>ID</th>
<th>EXPER</th>
<th>LNW</th>
<th>BLACK</th>
<th>HGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>206</td>
<td>1.874</td>
<td>2.082</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>206</td>
<td>2.814</td>
<td>2.397</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>206</td>
<td>4.314</td>
<td>2.482</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>332</td>
<td>0.125</td>
<td>1.630</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>332</td>
<td>1.025</td>
<td>1.476</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>332</td>
<td>2.413</td>
<td>1.804</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>332</td>
<td>3.903</td>
<td>1.439</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>332</td>
<td>4.470</td>
<td>1.748</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>332</td>
<td>5.178</td>
<td>1.526</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>332</td>
<td>6.082</td>
<td>2.044</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>332</td>
<td>7.048</td>
<td>2.179</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>332</td>
<td>8.197</td>
<td>2.186</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>332</td>
<td>9.092</td>
<td>4.055</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

**ID 206 has 3 waves**

<table>
<thead>
<tr>
<th>ID</th>
<th>EXPER</th>
<th>LNW</th>
<th>BLACK</th>
<th>HGC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1028</td>
<td>0.004</td>
<td>0.872</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1028</td>
<td>0.035</td>
<td>0.903</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1028</td>
<td>0.515</td>
<td>1.389</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1028</td>
<td>1.483</td>
<td>2.324</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1028</td>
<td>2.141</td>
<td>1.494</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1028</td>
<td>3.161</td>
<td>1.705</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>1028</td>
<td>4.105</td>
<td>2.343</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

**ID 332 has 10 waves**

**ID 1028 has 7 waves**

**Covariates:**
- Race and Highest Grade Completed

**LNW in constant dollars seems to rise over time**

(ALDA, Section 5.2.1, pp 146-148)
Fitting multilevel models for change when data sets have varying numbers of waves

Everything remains the same—there’s really no difference!

Unconditional growth model: On average, a dropout’s hourly wage increases with work experience.

\[100(e^{0.0457}-1) = 4.7\] is the percentage change in Y per annum.

Fully specified growth model (both HGC & BLACK)

- HGC is associated with initial status (but not change)
- BLACK is associated with change (but not initial status)

⇒ Fit Model C, which removes non-significant parameters

Prototypical wage trajectories of HS dropouts

Race
- At dropout, no racial differences in wages
- Racial disparities increase over time because wages for Blacks increase at a slower rate

Highest grade completed
- Those who stay in school longer have higher initial wages
- This differential remains constant over time (lines remain parallel)
Practical advice: Problems can arise when analyzing unbalanced data sets

The multilevel model for change is designed to handle unbalanced data sets, and in most circumstances, it does its job well, however...

- When imbalance is severe, or lots of people have just 1 or 2 waves of data, problems can occur:
  - You may not estimate some parameters (well)
  - Iterative fitting algorithms may not converge
  - Some estimates may hit boundary constraints
  - Problem is usually manifested via VCs not fixed effects (because the fixed portion of the model is like a 'regular regression model').

- Software packages may not issue clear warning signs:
  - If you’re lucky, you’ll get negative variance components
  - Another sign is too much time to convergence (or no convergence)
  - Most common problem: your model is overspecified
  - Most common solution: simplify the model

Many practical strategies discussed in ALDA, Section 5.2.2

Another major advantage of the multilevel model for change: How easy it is to include time-varying predictors

Example for illustrating time-varying predictors: Unemployment & depression

Source: Liz Ginexi and colleagues (2000), J of Occupational Health Psychology

- Sample: 254 people identified at unemployment offices.
- Research design: Goal was to collect 3 waves of data per person at 1, 3 and 11 months of job loss. In reality, however, data set is not time-structured:
  - Interview 1 was within 1 day and 2 months of job loss
  - Interview 2 was between 3 and 8 months of job loss
  - Interview 3 was between 10 and 16 months of job loss
  - In addition, not everyone completed the 2nd and 3rd interview.
- Time-varying predictor: Unemployment status (UNEMP)
  - 132 remained unemployed at every interview
  - 61 were always working after the 1st interview
  - 41 were still unemployed at the 2nd interview, but working by the 3rd
  - 19 were working at the 2nd interview, but were unemployed again by the 3rd
- Outcome: CES-D scale—20 4-pt items (score of 0 to 80)
- Research question
  - How does unemployment affect depression symptomatology?

(ALDA, Section 5.2.2, pp151-156)
A person-period data set with a time-varying predictor

<table>
<thead>
<tr>
<th>ID</th>
<th>MONTHS</th>
<th>CES-D</th>
<th>UNEMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>7589</td>
<td>1.5142</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>7589</td>
<td>5.0924</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>7589</td>
<td>11.7047</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>55697</td>
<td>1.3471</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>55697</td>
<td>5.7823</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>65641</td>
<td>0.5585</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>65641</td>
<td>4.1008</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>65641</td>
<td>10.9405</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>65441</td>
<td>1.0842</td>
<td>27</td>
<td>1</td>
</tr>
<tr>
<td>65441</td>
<td>4.6982</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>53782</td>
<td>11.2690</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>53782</td>
<td>0.4271</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>53782</td>
<td>4.2382</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>53782</td>
<td>11.0719</td>
<td>21</td>
<td>1</td>
</tr>
</tbody>
</table>

**ID 7589** has 3 waves, all unemployed

**ID 65641** has 3 waves, re-employed after 1st wave

**ID 53782** has 3 waves, re-employed at 2nd, unemployed again at 3rd

**TIME=MONTHS since job loss**

**UNEMP (by design, must be 1 at wave 1)**

---

**Analytic approach: We’re going to sequentially fit 4 increasingly complex models**

Goal is to both explain the use of TV predictors and illustrate how you do practical data analysis

**Model A: An individual growth model with no substantive predictors**

\[ Y_{ij} = \pi_{0i} + \pi_{1i} \text{TIME}_{ij} + \epsilon_{ij} \text{, where } \epsilon_{ij} \sim N(0, \sigma^2_\epsilon) \]

**Model B: Adding the main effect of UNEMP**

\[ Y_{ij} = \gamma_0 + \gamma_{10} \text{TIME}_{ij} + \gamma_{20} \text{UNEMP}_{ij} + [\zeta_{0i} + \zeta_{1i} \text{TIME}_{ij} + \epsilon_{ij}] \]

**Model C: Allowing the effect of UNEMP to vary over TIME**

\[ Y_{ij} = \gamma_0 + \gamma_{10} \text{TIME}_{ij} + \gamma_{20} \text{UNEMP}_{ij} + \gamma_{30} \text{UNEMP}_{ij} \times \text{TIME}_{ij} + [\zeta_{0i} + \zeta_{1i} \text{TIME}_{ij} + \epsilon_{ij}] \]

**Model D: Also allows the effect of UNEMP to vary over TIME, but does so in a very particular way**

\[ Y_{ij} = \gamma_0 + \gamma_{20} \text{UNEMP}_{ij} + \gamma_{30} \text{UNEMP}_{ij} \times \text{TIME}_{ij} + \gamma_{30} \text{UNEMP}_{ij} \times \text{TIME}_{ij} + [\zeta_{0i} + \zeta_{2i} \text{UNEMP}_{ij} + \zeta_{3i} \text{UNEMP}_{ij} \times \text{TIME}_{ij} + \epsilon_{ij}] \]

(ALDA, Section 5.3.1, pp 159-164)
First step: Model A: The unconditional growth model

Let’s get a sense of the data by ignoring UNEMP and fitting the usual unconditional growth model

Level-1 Model: \[ Y_{ij} = \pi_{0i} + \pi_{1i} \text{TIME}_{ij} + \varepsilon_{ij} \] where \( \varepsilon_{ij} \sim N(0, \sigma^{2}_{\varepsilon}) \)

Level-2 Model: \[ \pi_{0i} = \gamma_{00} + \zeta_{0i} \] where \( \zeta_{0i} \sim N \left( 0, \begin{bmatrix} \sigma^{2}_{\zeta_{0i}} & \sigma_{0i} \\ \sigma_{0i} & \sigma^{2}_{\zeta_{i}} \end{bmatrix} \right) \)

Composite Model: \[ Y_{ij} = \gamma_{00} + \gamma_{10} \text{TIME}_{ij} + [\zeta_{0i} + \zeta_{1i} \text{TIME}_{ij} + \varepsilon_{ij}] \]

Fixed Effects
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_{00} )</td>
<td>17.6094*** (0.7756)</td>
</tr>
<tr>
<td>( \gamma_{10} )</td>
<td>-0.4220*** (0.0630)</td>
</tr>
</tbody>
</table>

Variance Components
| Level-1: Within-person | \( \sigma^{2}_{\varepsilon} \) |
| Level-2: In intercept | 68.85*** |
| In rate of change | 96.86*** |

~ \( p < .10; * p < .05; ** p < .01; *** p < .001 \).

On the first day of job loss, the average person has an estimated CES-D of 17.6.

On average, CES-D declines by 0.42/mo.

There’s significant residual within-person variation.

There’s significant variation in initial status and rates of change.

How can we understand this graphically? Although the magnitude of the TV predictor’s effect remains constant, the TV nature of UNEMP implies the existence of many possible population average trajectories, such as:

How do we add the time-varying predictor UNEMP?

Model B: Adding time-varying UNEMP to the composite specification

\[ Y_{ij} = \gamma_{00} + \gamma_{10} \text{TIME}_{ij} + \gamma_{20} \text{UNEMP}_{ij} + [\zeta_{0i} + \zeta_{1i} \text{TIME}_{ij} + \varepsilon_{ij}] \]

Logical impossibility

Population average rate of change in CES-D, controlling for UNEMP

Population average difference, over time, in CES-D by UNEMP status

What happens when we fit Model B to data?
Fitting and interpreting Model B, which includes the TV predictor UNEMP

When analyzing time-invariant predictors, we know which VCs will change and how:
- Level-1 VCs will remain relatively stable because time-invariant predictors cannot explain much within-person variation
- Level-2 VCs will decline if the time-invariant predictors explain some of the between-person variation

When analyzing time-varying predictors, all VCs can change, but
- Although you can interpret a decrease in the magnitude of the Level-1 VCs
- Changes in Level-2 VCs may not be meaningful!

Level-1 VC, $\sigma^2_1$
- Adding UNEMP to the unconditional growth model (A) reduces its magnitude 68.85 to 62.39
- UNEMP “explains” 9.4% of the variation in CES-D scores

Level-2 VC, $\sigma^2_2$
- $\sigma^2_2$ changes

What about the variance components?

What about people who get a job?

© Singer & Willett, page 17
Decomposing the composite specification of Model B into a L1/L2 specification

\[ Y_{ij} = \gamma_{00} + \gamma_{10} TIME_{ij} + \gamma_{20} UNEMP_{ij} + [\xi_{0i} + \xi_{1i} TIME_{ij} + \epsilon_{ij}] \]

**Level-1 Model:**
\[ Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \pi_{2i} UNEMP_{ij} + \epsilon_{ij} \]

**Level-2 Models:**
\[
\begin{align*}
\pi_{0i} &= \gamma_{00} + \xi_{0i} \\
\pi_{1i} &= \gamma_{10} + \xi_{1i} \\
\pi_{2i} &= \gamma_{20} + \xi_{2i}
\end{align*}
\]

Unlike time-invariant predictors, TV predictors go into the level-1 model.

*Model B’s level-2 model for \( \pi_2 \) has no residual!*

*Model B automatically assumes that \( \pi_2 \) is “fixed” (that it has the same value for everyone).*

**Should we accept this constraint?**
- Should we assume that the effect of the person-specific predictor is constant across people?
- When predictors are time-invariant, we have no choice
- When predictors are time-varying, we can try to relax this assumption

(\textit{ALDA, Section 5.3.1, pp. 168-169})

---

Trying to add back the “missing” level-2 stochastic variation in the effect of UNEMP

**Level-1 Model:**
\[ Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \pi_{2i} UNEMP_{ij} + \epsilon_{ij} \]

**Level-2 Models:**
\[
\begin{align*}
\pi_{0i} &= \gamma_{00} + \xi_{0i} \\
\pi_{1i} &= \gamma_{10} + \xi_{1i} \\
\pi_{2i} &= \gamma_{20} + \xi_{2i}
\end{align*}
\]

- It’s easy to allow the effect of UNEMP to vary randomly across people by adding in a level-2 residual
- Check your software to be sure you know what you’re doing....

\[ \epsilon_{ij} \sim N(0, \sigma_{\epsilon}^2) \] and
\[ \begin{bmatrix} \xi_{0i} \\ \xi_{1i} \\ \xi_{2i} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{0}^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_{1}^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_{2}^2 \end{bmatrix} \right) \]

*But, you pay a price you may not be able to afford*
- Adding this one term adds 3 new VCs
- If you have only a few waves, you may not have enough data
- Here, we can’t actually fit this model!!

\[ \sigma_{0}^2, \ \sigma_{1}^2, \ \sigma_{2}^2 \]

Moral: The multilevel model for change can easily handle TV predictors, but...
- Think carefully about the consequences for both the structural and stochastic parts of the model.
- Don’t just “buy” the default specification in your software.
- *‘‘Until you’re sure you know what you’re doing, always write out your model before specifying code to a computer package*

So...
Are we happy with Model B as the final model??
Is there any other way to allow the effect of UNEMP to vary – if not across people, across TIME?

(\textit{ALDA, Section 5.3.1, pp. 169-171})

© Singer & Willett, page 20
Model C: Might the effect of a TV predictor vary over time?

When analyzing the effects of time-invariant predictors, we automatically allowed predictors to affect the trajectory’s slope.

Because of the way in which we’ve constructed the models with TV predictors, we’ve automatically constrained UNEMP to have only a “main effect” influencing just the trajectory’s level.

To allow the effect of the TV predictor to vary over time, just add its interaction with TIME:

\[ Y_{ij} = \gamma_{00} + \gamma_{10} TIME_{ij} + \gamma_{20} UNEMP_{ij} + \gamma_{30} UNEMP_{ij} \times TIME_{ij} + [\zeta_{0i} + \zeta_{1i} TIME_{ij} + \varepsilon_{ij}] \]

Two possible (equivalent) interpretations:
- The effect of UNEMP differs across occasions.
- The rate of change in depression differs by unemployment status.

But you need to think very carefully about the hypothesized error structure:
- We’ve basically added another level-1 parameter to capture the interaction.
- Just like we asked for the main effect of the TV predictor UNEMP, should we allow the interaction effect to vary across people?
- We won’t right now, but we will in a minute.

What happens when we fit Model C to data?

(ALDA, Section 5.3.2, pp. 171-172)

© Singer & Willett, page 21

Model C: Allowing the effect of a TV predictor to vary over time

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Model B</th>
<th>Model C</th>
<th>UNEMP TIME interaction is stat sig. (p&lt;.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td>Intercept (initial status)</td>
<td>( \gamma_0 = 12.6656^{**} )</td>
<td>( \gamma_0 = 9.6167 )</td>
</tr>
<tr>
<td></td>
<td>TIME (rate of change)</td>
<td>( \gamma_1 = -0.0209^{*} )</td>
<td>( \gamma_1 = 0.1620 )</td>
</tr>
<tr>
<td></td>
<td>UNEMP</td>
<td>( \gamma_2 = 5.1113^{***} )</td>
<td>( \gamma_2 = 8.3291^{***} )</td>
</tr>
<tr>
<td></td>
<td>UNEMP by TIME</td>
<td>( \gamma_3 = 0.4052^{*} )</td>
<td>( \gamma_3 = -0.4652^{*} )</td>
</tr>
<tr>
<td>Variance Components</td>
<td>Level1: Within-person ( \sigma^2 )</td>
<td>62.29^{***}</td>
<td>62.09^{***}</td>
</tr>
<tr>
<td></td>
<td>Level1: In intercept ( \sigma^2 )</td>
<td>95.52^{***}</td>
<td>93.71^{***}</td>
</tr>
<tr>
<td></td>
<td>Level1: In rate of change ( \sigma^2 )</td>
<td>0.46^{**}</td>
<td>0.45^{**}</td>
</tr>
<tr>
<td>Goodness-of-fit</td>
<td>Deviance</td>
<td>5107.6</td>
<td>5145.0</td>
</tr>
<tr>
<td></td>
<td>AIC</td>
<td>5121.6</td>
<td>5159.7</td>
</tr>
<tr>
<td></td>
<td>BIC</td>
<td>5146.4</td>
<td>5147.3</td>
</tr>
</tbody>
</table>

Main effect of TIME is now positive (\( \gamma_1 \)) & not stat sig. (\( p>.05 \))

Model C is a much poorer fit than B (\( \Delta \)Deviance \( = 4.6, 1 \text{ df} \), p<.05)

\( \hat{Y}_{ij} \) for the reemployed constrained to 0?

What about people who get a job?

© Singer & Willett, page 22
How should we constrain the individual growth trajectory for the re-employed?

Model D:

\[ Y_{ij} = \gamma_0 + \gamma_{20} \text{UNEMP}_{ij} + \gamma_{30} \text{UNEMP}_{ij} \times \text{TIME}_{ij} + [\zeta_{0i} + \zeta_{1i} \text{TIME}_{ij} + \epsilon_{ij}] \]

Should we remove the main effect of TIME? (which is the slope when UNEMP=0)

Yes, but this creates a lack of congruence between the model’s fixed and stochastic parts

So, let’s better align the parts by having UNEMP TIME be both fixed and random

If we’re allowing the UNEMP TIME slope to vary randomly, might we also need to allow the effect of UNEMP itself to vary randomly?

But, this actually fits worse (larger AIC & BIC)!

UNEMP has both a fixed & random effect

UNEMP TIME has both a fixed & random effect

What happens when we fit Model D to data?

Model D: Constraining the individual growth trajectory among the reemployed

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model C</th>
<th>Model D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\gamma_0$</td>
<td>$\gamma_0$</td>
</tr>
<tr>
<td>TIME</td>
<td>$\gamma_0$</td>
<td>$\gamma_0$</td>
</tr>
<tr>
<td>UNEMP</td>
<td>$\gamma_0$</td>
<td>$\gamma_0$</td>
</tr>
<tr>
<td>UNEMP by TIME</td>
<td>$\gamma_0$</td>
<td>$\gamma_0$</td>
</tr>
<tr>
<td>Variance Components</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 1: Within-person</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>Level 2: In intercept</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>In rate of change</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>In UNEMP</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
<tr>
<td>In UNEMP by TIME</td>
<td>$\sigma^2$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

Best fitting model (lowest AIC and BIC)

Consistently unemployed

\[ \hat{Y}_j = (11.2666, 6.8795) - 0.3254 \text{MONTHS}_j \]

Consistently employed

\[ \hat{Y}_j = 11.2666 \]

What about people who get a job?
Recentering the effects of TIME

All our examples so far have centered TIME on the first wave of data collection

- Allows us to interpret the level-1 intercept as individual i’s true initial status
- While commonplace and usually meaningful, this approach is not sacrosanct.

We always want to center TIME on a value that ensures that the level-1 growth parameters are meaningful, but there are other options

- Middle TIME point—focus on the “average” value of the outcome during the study
- Endpoint—focus on “final status”
- Any inherently meaningful constant can be used

Example for recentering the effects of TIME

Data source: Tomarken & colleagues (1997) American Psychological Society Meetings

Sample: 73 men and women with major depression who were already being treated with non-pharmacological therapy

- Randomized trial to evaluate the efficacy of supplemental antidepressants (vs. placebo)

Research design

- Pre-intervention night, the researchers prevented all participants from sleeping
- Each person was electronically paged 3 times a day (at 8 am, 3 pm, and 10 pm) to remind them to fill out a mood diary
- With full compliance—which didn’t happen, of course—each person would have 21 mood assessments (most had at least 16 assessments, although 1 person had only 2 and 1 only 12)
- The outcome, POS is the number of positive moods

Research question:

- How does POS change over time?
- What is the effect of medication on the trajectories of change?
How might we clock and code TIME?

<table>
<thead>
<tr>
<th>Wave</th>
<th>Day</th>
<th>Reading</th>
<th>Time of Day</th>
<th>Time</th>
<th>(Time - 3.33)</th>
<th>(Time - 6.67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>08 A.M.</td>
<td>0.00</td>
<td>0.00</td>
<td>-3.33</td>
<td>-6.67</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>05 P.M.</td>
<td>0.33</td>
<td>0.33</td>
<td>-3.00</td>
<td>-6.33</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>10 A.M.</td>
<td>0.67</td>
<td>0.67</td>
<td>-2.67</td>
<td>-6.00</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>08 A.M.</td>
<td>0.00</td>
<td>1.00</td>
<td>-2.33</td>
<td>-5.67</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>10 A.M.</td>
<td>0.33</td>
<td>1.33</td>
<td>-2.00</td>
<td>-5.33</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>10 P.M.</td>
<td>0.67</td>
<td>1.67</td>
<td>-1.67</td>
<td>-5.00</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>10 P.M.</td>
<td>3.33</td>
<td>0.00</td>
<td></td>
<td>-3.33</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>08 A.M.</td>
<td>0.00</td>
<td>5.00</td>
<td>1.67</td>
<td>-1.67</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>05 P.M.</td>
<td>0.33</td>
<td>5.33</td>
<td>2.00</td>
<td>-1.33</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>10 P.M.</td>
<td>0.67</td>
<td>5.67</td>
<td>2.33</td>
<td>-1.00</td>
</tr>
<tr>
<td>19</td>
<td>6</td>
<td>08 A.M.</td>
<td>0.00</td>
<td>6.00</td>
<td>2.67</td>
<td>-0.67</td>
</tr>
<tr>
<td>20</td>
<td>6</td>
<td>10 A.M.</td>
<td>0.33</td>
<td>6.33</td>
<td>3.00</td>
<td>-0.33</td>
</tr>
<tr>
<td>21</td>
<td>6</td>
<td>10 P.M.</td>
<td>0.67</td>
<td>6.67</td>
<td>3.33</td>
<td>0.00</td>
</tr>
</tbody>
</table>

WAVE—Great for data processing—no intuitive meaning

DAY—Intuitively appealing, but doesn’t distinguish readings each day

TIME OF DAY—quantifies 3 distance between readings (could also make unequal)

(TIME - 3.33)

Same as TIME but now centered on the study’s midpoint

(TIME - 6.67)

Same as TIME but now centered on the study’s endpoint

READING—right idea, but how to quantify?

TIME—days since study began (centered on first wave of data collection)

TIME - 3.33

Same as TIME but now centered on the study’s midpoint

TIME - 6.67

Same as TIME but now centered on the study’s endpoint

Understanding what happens when we recenter TIME

Instead of writing separate models depending upon the representation for TIME, let use a generic form:

**Level-1 Model:**

\[ Y_{ij} = \pi_{0i} + \pi_{1i}(\text{TIME}_{ij} - c) + \varepsilon_{ij} \]

where \( \varepsilon_{ij} \sim N(0, \sigma^2_{\varepsilon}) \)

**Level-2 Model:**

\[
\begin{align*}
\pi_{0i} &= \gamma_{00} + \gamma_{01}\text{TREAT}_{i} + \zeta_{0i} \\
\pi_{1i} &= \gamma_{10} + \gamma_{11}\text{TREAT}_{i} + \zeta_{1i}
\end{align*}
\]

\[
\begin{bmatrix}
\zeta_{0i} \\
\zeta_{1i}
\end{bmatrix} \sim N\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\zeta} & \sigma_{\zeta\gamma} \\ \sigma_{\gamma\zeta} & \sigma^2_{\gamma} \end{bmatrix} \right)
\]

**Notice how changing the value of the centering constant, c, changes the definition of the intercept in the level-1 model:**

- \( Y_{ij} = \pi_{0i} + \pi_{1i}\text{TIME}_{ij} + \varepsilon_{ij} \)
  - When \( c = 0 \):
    - \( \pi_{0i} \) is the individual mood at \( \text{TIME}=0 \)
    - Usually called “initial status”
  - When \( c = 3.33 \):
    - \( \pi_{0i} \) is the individual mood at \( \text{TIME}=3.33 \)
    - Useful to think of as “mid-experiment status”
  - When \( c = 6.67 \):
    - \( \pi_{0i} \) is the individual mood at \( \text{TIME}=6.67 \)
    - Useful to think about as “final status”
Comparing the results of using different centering constants for TIME

What are affected are the level-1 intercepts

\( \gamma_0 \) assesses level of POS at time \( c \) for the control group (TREAT = 0)

\( \gamma_1 \) assesses the difference in POS between the groups (TREATment effect)

-3.11 (ns) at study beginning
15.35 (ns) at study midpoint
33.80 * at study conclusion

The choice of centering constant has no effect on:

- Goodness of fit indices
- Estimates for rates of change
- Within person residual variance
- Between person residual variance in rate of change

You can extend the idea of recentering TIME in lots of interesting ways

Example: Instead of focusing on rate of change, parameterize the level-1 model so it produces one parameter for initial status and one parameter for final status…

\[ Y_{ij} = \pi_0 \left( \frac{6.67 - \text{TIME}_{ij}}{6.67} \right) + \pi_1 \left( \frac{\text{TIME}_{ij}}{6.67} \right) + \epsilon_{ij} \]

Individual Initial Status Parameter

Individual Final Status Parameter

Advantage: You can use all your longitudinal data to analyze initial and final status simultaneously.
Chapter 6: Modeling discontinuous and nonlinear change

**General idea:** All our examples so far have assumed that individual growth is smooth and linear. But the multilevel model for change is much more flexible:

- **Discontinuous individual change** (§6.1)—especially useful when discrete shocks or time-limited treatments affect the life course
- **Using transformations to model non-linear change** (§6.2)—perhaps the easiest way of fitting non-linear change models
  - Can transform either the outcome or TIME
  - We already did this with ALCUSE (which was a square root of a sum of 4 items)
- **Using polynomials of TIME to represent non-linear change** (§6.3)
  - While admittedly atheoretical, it’s very easy to do
  - Probably the most popular approach in practice
- **Truly non-linear trajectories** (§6.4)
  - Logistic, exponential, and negative exponential models, for example
  - A world of possibilities limited only by your theory (and the quality and amount of data)
Example for discontinuous individual change: Wage trajectories & the GED

**Data source:** Murnane, Boudett and Willett (1999), Evaluation Review

1. **Sample:** the same 888 male high school dropouts (from before)
2. **Research design**
   - Each was interviewed between 1 and 13 times after dropping out
   - 34.6% (n=307) earned a GED at some point during data collection
3. **OLD research questions**
   - How do log(WAGES) change over time?
   - Do the wage trajectories differ by ethnicity and highest grade completed?
4. **Additional NEW research questions:** What is the effect of GED attainment? Does earning a GED:
   - affect the wage trajectory’s elevation?
   - affect the wage trajectory’s slope?
   - create a discontinuity in the wage trajectory?

(ALDA, Section 6.1.1, pp 190-193)

First steps: Think about how GED receipt might affect an individual’s wage trajectory

Let’s start by considering four plausible effects of GED receipt by imagining what the wage trajectory might look like for someone who got a GED 3 years after labor force entry (post dropout)

![Graph showing wage trajectories with labels for A: No effect of GED whatsoever, B: An immediate shift in elevation; no difference in rate of change, C: An immediate shift in rate of change; no difference in elevation, D: Immediate shifts in both elevation & rate of change, F: Immediate shifts in both elevation & rate of change)](image)

(ALDA, Figure 6.1, p 193)
Including a discontinuity in elevation, not slope (Trajectory B)

Key idea: It's easy; simply include GED as a time-varying effect at level-1

\[ Y_{ij} = \pi_{0i} + \pi_{1i} \text{EXPER}_{ij} + \pi_{2i} \text{GED}_{ij} + \varepsilon_{ij} \]

Post-GED (GED=1):
\[ Y_{ij} = (\pi_{0i} + \pi_{2i}) + \pi_{1i} \text{EXPER}_{ij} + \varepsilon_{ij} \]

Pre-GED (GED=0):
\[ Y_{ij} = \pi_{0i} + \pi_{1i} \text{EXPER}_{ij} + \varepsilon_{ij} \]

Including a discontinuity in slope, not elevation (Trajectory D)

Using an additional temporal predictor to capture the “extra slope” post-GED receipt

\[ Y_{ij} = \pi_{0i} + \pi_{1i} \text{EXPER}_{ij} + \pi_{3i} \text{POSTEXP}_{ij} + \varepsilon_{ij} \]

Post-GED (POSTEXP clocked in same cadence as EXPER):
\[ Y_{ij} = \pi_{0i} + \pi_{1i} \text{EXPER}_{ij} + \pi_{3i} \text{POSTEXP} + \varepsilon_{ij} \]

\[ Y_{ij} = \pi_{0i} + (\pi_{1i} + \pi_{3i}) \text{EXPER} + \varepsilon_{ij} \]

POSTEXP$_{ij}$ = 0 prior to GED
POSTEXP$_{ij}$ = “Post GED experience,” a new TV predictor that clocks “TIME since GED receipt” (in the same cadence as EXPER)
Including a discontinuities in both elevation and slope (Trajectory F)

Simple idea: Combine the two previous approaches

\[ Y_{ij} = \pi_{0i} + \pi_{1i}EXPER_{ij} + \pi_{2i}GED + \pi_{3i}POSTEXP_{ij} + \varepsilon_{ij} \]

**Post-GED**

\[ Y_{ij} = (\pi_{0i} + \pi_{2i}) + (\pi_{1i} + \pi_{3i})EXPER + \varepsilon_{ij} \]

**Pre-GED**

\[ Y_{ij} = \pi_{0i} + \pi_{1i}EXPER_{ij} + \varepsilon_{ij} \]

Many other types of discontinuous individual change trajectories are possible

Just like a regular regression model, the multilevel model for change can include discontinuities, non-linearities and other ‘non-standard’ terms

- Generally more limited by data, theory, or both, than by the ability to specify the model
- Extra terms in the level-1 model translate into extra parameters to estimate

What kinds of other complex trajectories could be used?

- Effects on elevation and slope can depend upon timing of GED receipt (ALDA pp. 199-201)
- You might have non-linear changes before or after the transition point
- The effect of GED receipt might be instantaneous but not endure
- The effect of GED receipt might be delayed
- Might there be multiple transition points (e.g., on entry in college for GED recipients)

Think carefully about what kinds of discontinuities might arise in your substantive context

(© Singer & Willett, page 7)

(© Singer & Willett, page 8)
Let's start with a “baseline model” (Model A) against which we'll compare alternative discontinuous trajectories.

(UERATE-7) is the local area unemployment rate (added in previous chapter as an example of a TV predictor), centered around 7% for interpretability.

\[ Y_{ij} = \pi_{0i} + \pi_{1i} \text{EXPER}_{ij} + \pi_{2i} (\text{UERATE}_{ij} - 7) + \epsilon_{ij} \]

\[ \pi_{0i} = \gamma_{00} + \gamma_{01} (\text{HGC}_{i} - 9) + \xi_{0i} \]

\[ \pi_{1i} = \gamma_{10} + \gamma_{11} \text{BLACK}_{i} + \xi_{1i} \]

\[ \pi_{2i} = \gamma_{20} \]

\[ \epsilon_{ij} \sim N(0, \sigma_{\epsilon}^2) \] and \[ \begin{bmatrix} \xi_{0i} \\ \xi_{1i} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\xi0}^2 & \sigma_{\xi0\xi1} \\ \sigma_{\xi0\xi1} & \sigma_{\xi1}^2 \end{bmatrix} \right) \]

To appropriately compare this deviance statistic to more complex models, we need to know how many parameters have been estimated to achieve this value of deviance.

Benchmark against which we’ll evaluate discontinuous models:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td>Intercept</td>
</tr>
<tr>
<td>Initial status, ( \pi_{0i} )</td>
<td>( \gamma_{00} )</td>
</tr>
<tr>
<td>( \text{HGC}_{i} - 9 )</td>
<td>( \gamma_{00} )</td>
</tr>
<tr>
<td>( \text{UERATE}_{ij} - 7 )</td>
<td>( \pi_{0i} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of change, ( \pi_{i} )</td>
<td>Intercept</td>
</tr>
<tr>
<td>( \text{BLACK}_{i} )</td>
<td>( \pi_{0i} )</td>
</tr>
<tr>
<td>( \text{UERATE}_{ij} - 7 )</td>
<td>( \pi_{0i} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance Components</td>
<td>Level 1: within-person</td>
</tr>
<tr>
<td>Level 2: in rate of change</td>
<td>( \sigma_{\xi0}^2 )</td>
</tr>
</tbody>
</table>

Goodness-of-fit:

- Deviance: 4830.5
- AIC: 4848.5
- BIC: 4893.6

How we’re going to proceed...

Instead of constructing tables of (seemingly endless) parameter estimates, we’re going to construct a summary table that presents the...

- specific terms in the model
- \( n \) parameters (for d.f.)
- deviance statistic (for model comparison)

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed effects parameters</th>
<th>Variance components parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Intercept, EXPER, HGC</td>
<td>Intercept, EXPER, HGC</td>
</tr>
<tr>
<td>B</td>
<td>Model A + GED</td>
<td>Intercept, EXPER, GED</td>
</tr>
<tr>
<td>C</td>
<td>Model B w/o GED</td>
<td>Model B w/o GED</td>
</tr>
<tr>
<td>D</td>
<td>Model A + POSTEXP</td>
<td>Intercept, EXPER, POSTEXP</td>
</tr>
<tr>
<td>E</td>
<td>Model A + POSTEXP</td>
<td>Model D w/o POSTEXP</td>
</tr>
<tr>
<td>F</td>
<td>Model A + POSTEXP</td>
<td>Intercept, EXPER, POSTEXP</td>
</tr>
<tr>
<td>G</td>
<td>Model F</td>
<td>Model F w/o POSTEXP</td>
</tr>
<tr>
<td>H</td>
<td>Model F</td>
<td>Model F w/o GED</td>
</tr>
<tr>
<td>I</td>
<td>Model A + GED x POSTEXP</td>
<td>Intercept, GED x POSTEXP</td>
</tr>
<tr>
<td>J</td>
<td>Model 1</td>
<td>Model 1 w/o GED x POSTEXP</td>
</tr>
</tbody>
</table>

- \( p < .10 \), * \( p < .05 \), ** \( p < .01 \), *** \( p < .001 \)
First steps: Investigating the discontinuity in elevation by adding the effect of GED

B: Add GED as both a fixed and random effect
(1 extra fixed parameter; 3 extra random)
\( \Delta \text{Deviance}=25.0, 4 \text{ df}, p<.001 \) — keep GED effect

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Model} & \text{Fixed effects} & \text{Variance components (in addition to ef)} & \text{Fixed effects} & \text{Variance components} & \text{Deviance} & \text{Comparison model:} \\
\hline
A & \text{Intercept, EXPERT, HGC - 9, BLACK* EXPERT, URBANIZE - 7} & \text{Intercept, EXPERT} & 5 & 4 & 4890.5 & \text{—} \\
B & \text{Model A + GED} & \text{Intercept, EXPERT, GED} & 6 & 7 & 4865.5 & A: 25.0** (4) \\
C & \text{Model B} & \text{Model B w/o GED} & 6 & 4 & 4818.5 & B: 12.5** (3) \\
\end{array}
\]

C: But does the GED discontinuity vary across people?
(Do we need to keep the extra VCs for the effect of GED?)
\( \Delta \text{Deviance}=12.8, 3 \text{ df}, p<.01 \) — keep VCs

Next steps: Investigating the discontinuity in slope by adding the effect of POSTEXP
(without the GED effect producing a discontinuity in elevation)

D: Adding POSTEXP as both a fixed and random effect
(1 extra fixed parameter; 3 extra random)
\( \Delta \text{Deviance}=13.1, 4 \text{ df}, p<.05 \) — keep POSTEXP effect

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\text{Model} & \text{Fixed effects} & \text{Variance components (in addition to ef)} & \text{Fixed effects} & \text{Variance components} & \text{Deviance} & \text{Comparison model:} \\
\hline
A & \text{Intercept, EXPERT, HGC - 9, BLACK* EXPERT, URBANIZE - 7} & \text{Intercept, EXPERT} & 5 & 4 & 4890.5 & \text{—} \\
B & \text{Model A + POSTEXP} & \text{Intercept, EXPERT, POSTEXP} & 6 & 7 & 4861.7 & A: 13.1** (4) \\
C & \text{Model B} & \text{Model B w/o POSTEXP} & 6 & 4 & 4818.5 & B: 3.5 (ns) (3) \\
\end{array}
\]

E: But does the POSTEXP slope vary across people?
(Do we need to keep the extra VCs for the effect of POSTEXP?)
\( \Delta \text{Deviance}=3.3, 3 \text{ df}, \text{ns} \) — don’t need the POSTEXP random effects
(but in comparison with A still need POSTEXP fixed effect)

What if we include both types of discontinuity?
Examining both discontinuities simultaneously

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed effects</th>
<th>Variance components (in addition to $\sigma^2$)</th>
<th>Fixed effects</th>
<th>Variance components</th>
<th>Deviance</th>
<th>Comparison model: $\Delta$Deviance (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Intercept, EXP, GED, $-9$, BLACK$\times$EXP, URATE$-7$</td>
<td>Intercept, EXP, GED</td>
<td>5</td>
<td>4</td>
<td>4930.5</td>
<td>—</td>
</tr>
<tr>
<td>B</td>
<td>Model A + GED</td>
<td>Intercept, EXP, GED</td>
<td>6</td>
<td>4</td>
<td>4865.3</td>
<td>A: 25.4*** (4)</td>
</tr>
<tr>
<td>C</td>
<td>Model B</td>
<td>Model B w/o GED</td>
<td>6</td>
<td>4</td>
<td>4818.5</td>
<td>B: 12.8** (3)</td>
</tr>
<tr>
<td>D</td>
<td>Model A + POSTEXP</td>
<td>Intercept, EXP, POSTEXP</td>
<td>6</td>
<td>4</td>
<td>4817.4</td>
<td>A: 15.1** (4)</td>
</tr>
<tr>
<td>E</td>
<td>Model D</td>
<td>Model D w/o POSTEXP</td>
<td>6</td>
<td>4</td>
<td>4820.7</td>
<td>D: 3.5 (ns) (5)</td>
</tr>
<tr>
<td>F</td>
<td>Model A + GED and POSTEXP</td>
<td>Intercept, EXP, GED, POSTEXP</td>
<td>6</td>
<td>4</td>
<td>4799.4</td>
<td>B: 16.2** (4)</td>
</tr>
</tbody>
</table>

- F: Add GED and POSTEXP simultaneously (each as both fixed and random effects)
- comp. with B shows significance of POSTEXP
- comp. with D shows significance of GED

(ALDA, Section 6.1.2, pp 204-205)

Can we simplify this model by eliminating the VCs for POSTEXP (G) or GED (H)?

<table>
<thead>
<tr>
<th>Model</th>
<th>Fixed effects</th>
<th>Variance components (in addition to $\sigma^2$)</th>
<th>Fixed effects</th>
<th>Variance components</th>
<th>Deviance</th>
<th>Comparison model: $\Delta$Deviance (df)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Model A + GED and POSTEXP</td>
<td>Intercept, EXP, GED, POSTEXP</td>
<td>7</td>
<td>11</td>
<td>4799.4</td>
<td>B: 16.2** (5)</td>
</tr>
<tr>
<td>G</td>
<td>Model F</td>
<td>Model F w/o POSTEXP</td>
<td>7</td>
<td>7</td>
<td>4802.7</td>
<td>F: 15.5*** (4)</td>
</tr>
<tr>
<td>H</td>
<td>Model F</td>
<td>Model F w/o GED</td>
<td>7</td>
<td>7</td>
<td>4812.6</td>
<td>F: 25.3*** (4)</td>
</tr>
</tbody>
</table>

- Each results in a worse fit, suggesting that Model F (which includes both random effects) is better (even though Model E suggested we might be able to eliminate the VC for POSTEXP)

(ALDA, Section 6.1.2, pp 204-205)
Displaying prototypical discontinuous trajectories
(Log Wages for HS dropouts pre- and post-GED attainment)

1.6
1.8
2
2.2
2.4
0 2 4 6 8 10
EXPERIENCE

GED receipt has two effects
• Upon GED receipt, wages rise immediately by 4.2%
• Post-GED receipt, wages rise annually by 5.2% (vs. 4.2% pre-receipt)

Highest grade completed
• Those who stay longer have higher initial wages
• This differential remains constant over time

Race
• At dropout, no racial differences in wages
• Racial disparities increase over time because wages for Blacks increase at a slower rate

© Singer & Willett, page 15

Modeling non-linear change using transformations

When facing obviously non-linear trajectories, we usually begin by trying transformation:

A straight line—even on a transformed scale—is a simple form with easily interpretable parameters
Since many outcome metrics are ad hoc, transformation to another ad hoc scale may sacrifice little

The prototypical individual growth trajectories are now non-linear:
By transforming the outcome before analysis, we have effectively modeled non-linear change over time

So...how do we know what variable to transform using what transformation?

Earlier, we modeled ALCUSE, an outcome that we formed by taking the square root of the researchers’ original alcohol use measurement
We can ‘detransform’ the findings and return to the original scale, by squaring the predicted values of ALCUSE and re-plotting

© Singer & Willett, page 16
Step 1: What kinds of transformations do we consider?

- Ladder of Powers
  - Expand scale: \( V, V^2, V^3, \ldots \)
  - Compress scale: \( \log V, V^{-1}, V^{-2}, \ldots \)

Step 2: How do we know when to use which transformation?
1. Plot many empirical growth trajectories
2. Find linearizing transformations by moving “up” or “down” in the direction of the “bulge”

The effects of transformation for a single child in the Berkeley Growth Study

- IQ vs. TIME
- IQ^{(2.3)} vs. TIME
- IQ vs. TIME^{(1/2.3)}

How else might we model non-linear change?
### Representing individual change using a polynomial function of TIME

<table>
<thead>
<tr>
<th>Shape</th>
<th>Level-1 model</th>
<th>Parameter values</th>
<th>Plot of the true change trajectory</th>
</tr>
</thead>
<tbody>
<tr>
<td>No change</td>
<td>$Y_i = \beta_0 + \epsilon_i$</td>
<td>$\beta_0 = 71$</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>Linear change</td>
<td>$Y_i = \beta_0 + \beta_1 TIME_i + \epsilon_i$</td>
<td>$\beta_0 = 71, \beta_1 = 1.2$</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>Quadratic change</td>
<td>$Y_i = \beta_0 + \beta_1 TIME_i + \beta_2 TIME_i^2 + \epsilon_i$</td>
<td>$\beta_0 = 71, \beta_1 = 3.8, \beta_2 = -0.65$</td>
<td><img src="image" alt="Graph" /></td>
</tr>
<tr>
<td>Cubic change</td>
<td>$Y_i = \beta_0 + \beta_1 TIME_i + \beta_2 TIME_i^2 + \beta_3 TIME_i^3 + \epsilon_i$</td>
<td>$\beta_0 = 71, \beta_1 = 10, \beta_2 = -2, \beta_3 = 0.0042$</td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>

#### Polynomial of the “zero order” (because TIME$^0=1$)
- Like including a constant predictor $\epsilon$ in the level-1 model
- Intercept represents vertical elevation
- Different people can have different elevations

#### Polynomial of the “first order” (because TIME$^1=TIME$)
- Familiar individual growth model
- Varying intercepts and slopes yield criss-crossing lines

#### Second order polynomial for quadratic change
- Includes both TIME and TIME$^2$
- $\beta_0 =$ intercept, but now both TIME and TIME$^2$ must be 0
- $\beta_1 =$ instantaneous rate of change when TIME = 0 (there is no longer a constant slope)
- $\beta_2 =$ curvature parameter; larger its value, more dramatic its effect
- Peak is called a “stationary point”—a quadratic has 1.

#### Third order polynomial for cubic change
- Includes TIME, TIME$^2$ and TIME$^3$
- Can keep on adding powers of TIME
- Each extra polynomial adds another stationary point—a cubic has 2

---

### Example for illustrating use of polynomials in TIME to represent change

**Source:** Margaret Keiley & colleagues (2000). *J of Abnormal Child Psychology*

- **Sample:** 45 boys and girls identified in 1st grade
- **Goal:** was to study behavior changes over time (until 6th grade)

#### Research design
- At the end of every school year, teachers rated each child’s level of externalizing behavior using Achenbach’s Child Behavior Checklist:
  - 3 point scale (0 = rarely/never; 1 = sometimes; 2 = often)
  - 24 aggressive, disruptive, or delinquent behaviors
- **Outcome:** EXTERNAL—ranges from 0 to 68 (simple sum of these scores)
- **Predictor:** FEMALE—are there gender differences?

#### Research question
- How does children’s level of externalizing behavior change over time?
- Do the trajectories of change differ for boys and girls?
Selecting a suitable level-1 polynomial trajectory for change

Examining empirical growth plots (which invariably display great variability in temporal complexity)

- Quadratic change (but with varying curvatures)
- Linear decline (at least until 4th grade)
- Little change over time (flat line?)
- Two stationary points? (suggests a cubic)
- Three stationary points? (suggests a quartic!!!)

When faced with so many different patterns, how do you select a common polynomial for analysis?

Examining alternative fitted OLS polynomial trajectories

Order optimized for each child (solid curves) and a common quartic across children (dashed line)

- First impression: Most fitted trajectories provide a reasonable summary for each child’s data
- Second impression: Maybe these ad hoc decisions aren’t the best?
- Third realization: We need a common polynomial across all cases (and might the quartic be just too complex)?
- Using sample data to draw conclusions about the shape of the underlying true trajectories is tricky—let’s compare alternative models →
Using model comparisons to test higher order terms in a polynomial level-1 model

Add polynomial functions of TIME to person period data set

Compare goodness of fit (accounting for all the extra parameters that get estimated)

A: significant between- and within-child variation

B: no fixed effect of TIME but significant var comps
\[ \Delta \text{Deviance} = 18.5, 3\text{df}, p < .01 \]

C: no fixed effects of TIME & TIME^2 but significant var comps
\[ \Delta \text{Deviance} = 16.0, 4\text{df}, p < .01 \]

D: still no fixed effects for TIME terms, but now VCs are ns also
\[ \Delta \text{Deviance} = 11.1, 5\text{df}, \text{ns} \]
Quadatic (C) is best choice—and it turns out there are no gender differentials at all.

Example for truly non-linear change

Data source: Terry Tivnan (1980) Dissertation at Harvard Graduate School of Education

Sample: 17 1st and 2nd graders

During a 3 week period, Terry repeatedly played a two-person checkerboard game called Fox 'n Geese, (hopefully) learning from experience
- Fox is controlled by the experimenter, at one end of the board
- Children have four geese, that they use to try to trap the fox

Great for studying cognitive development because:
- There exists a strategy that children can learn that will guarantee victory
- This strategy is not immediately obvious to children
- Many children can deduce the strategy over time

Research design

Each child played up to 27 games (each game is a "wave")
- The outcome, NMOVES is the number of moves made by the child before making a catastrophic error (guaranteeing defeat)—ranges from 1 to 20

Research question:

How does NMOVES change over time?
What is the effect of a child’s reading (or cognitive) ability?—READ (score on a standardized reading test)
Selecting a suitable level-1 nonlinear trajectory for change
Examining empirical growth plots (and asking what features should the hypothesized model display?)

A lower asymptote, because everyone makes at least 1 move and it takes a while to figure out what’s going on

An upper asymptote, because a child can make only a finite # moves each game

A smooth curve joining the asymptotes, that initially accelerates and then decelerates

These three features suggest a level-1 logistic change trajectory, which unlike our previous growth models will be non-linear in the individual growth parameters

Understanding the logistic individual growth trajectory
(which is anything but linear in the individual growth parameters)

Upper asymptote in this particular model is constrained to be 20 (1+19)

$\pi_0$ is related to, and determines, the intercept

$Y_{ij} = \frac{19}{1 + e^{\pi_0 \pi_1 \text{TIME}_{ij}}} + \epsilon_{ij}$

$\pi_1$ determines the rapidity with which the trajectory approaches the upper asymptote

When $\pi_1$ is large, the trajectory rises more rapidly

Higher the value of $\pi_0$, the lower the intercept

When $\pi_1$ is small, the trajectory rises slowly (often not reaching an asymptote)

Models can be fit in usual way using provided your software can do it ⇒
Results of fitting logistic change trajectories to the Fox 'n Geese data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model A</th>
<th>Model B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>$\beta_0$</td>
<td>$\beta_0$</td>
</tr>
<tr>
<td>Slope</td>
<td>$\beta_1$</td>
<td>$\beta_1$</td>
</tr>
</tbody>
</table>

Not statistically significant (note small n's), but better READers approach asymptote more rapidly.

(Alda, Section 6.4.2, pp 229-232)

A limitless array of non-linear trajectories awaits…
(each is illustrated in detail in Alda, Section 6.4.3)