The American Statistical Association

BOSTON CHAPTER
Invited Short Course, 20 May 2005

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

Boston Chapter of the American Statistical Association Short Course

www.amstat.org/chapters/boston

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.

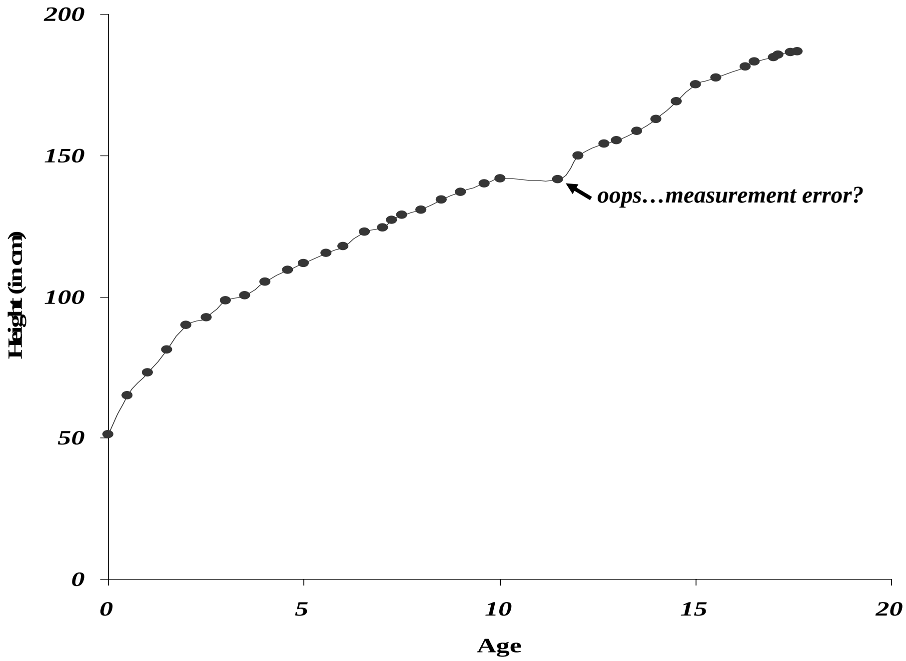
Abstract Based on their book, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* (Oxford, 2003), Singer and Willett will give an accessible yet in-depth presentation of multilevel models for individual change. Using real data sets from published studies, the instructors will take 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)

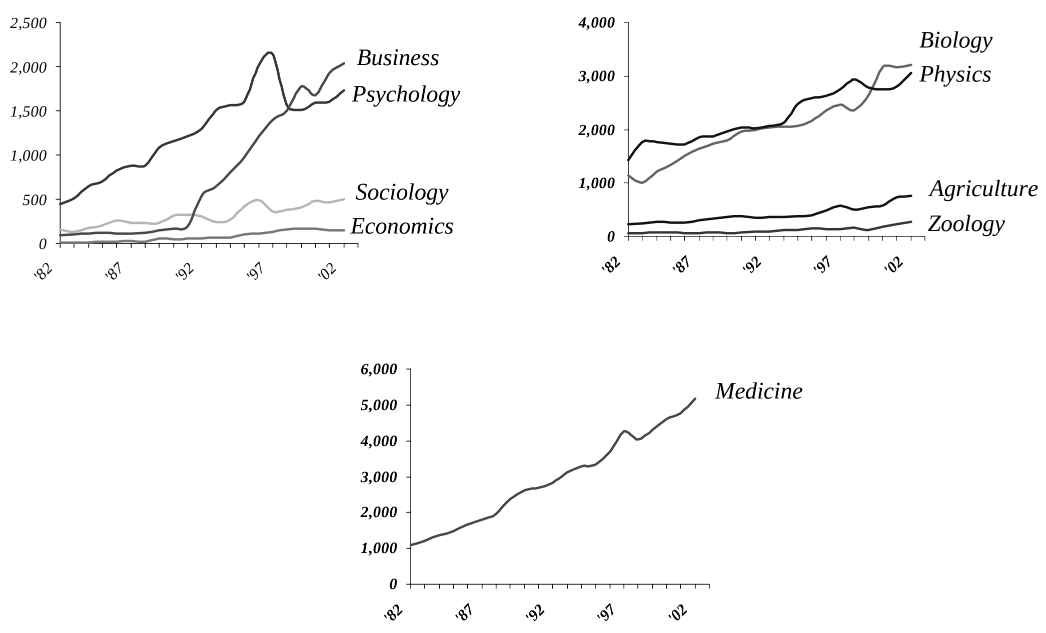
Recorded his son's height every six months from his birth in 1759 until his 18th birthday



Scammon, RE (1927) The first seriation study of human growth, Am J of Physical Anthropology, 10, 329-336.

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

Read 150 articles published in 10 APA journals in 1999 and 2003

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

	'99	'03
% longitudinal	*33%*	*47%*
2 waves	*36%*	*26%*
3 waves	*26%*	*29%*
4 or more waves	*38%*	*45%*

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

	'99	'03
Growth modeling	*7%*	*15%*
Survival analysis	*2%*	*5%*
Repeated measures ANOVA	*40%*	*29%*
Wave-to-wave regression	*38%*	*32%*
Separate but parallel analyses	*8%*	*17%*
"Simplifying" analyses by…		
Setting aside waves	*8%*	*7%*
Combining waves	*6%*	*8%*
Ignoring age-heterogeneity	*6%*	*9%*

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.

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

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.

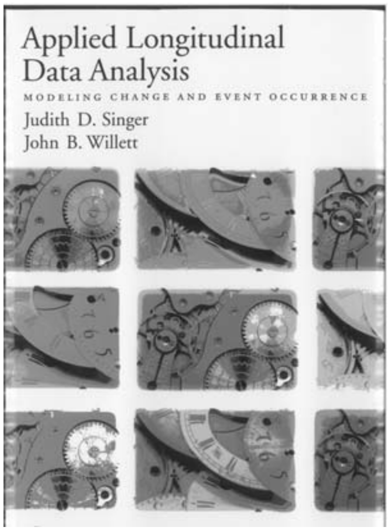
1. ***Within-person summary:*** *When are boys most at risk of having sex for the 1st time?*
2. ***Between-person comparison:*** *How does this risk vary by teen characteristics?*

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 Submodel for Individual Change* 49
3.3 *The Level-2 Submodel 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, Revisited* 85
4.4 *First Steps: Fitting Two Unconditional Multilevel Models for Change* 92
4.5 *Practical Data Analytic Strategies for Model Building* 104
4.6 *Comparing Models Using Deviance Statistics* 116

5 Treating *TIME* More Flexibly 138
5.1 *Variably Spaced Measurement Occasions* 139
5.2 *Varying Numbers of Measurement Occasions* 146
5.3 *Time-Varying Predictors* 159
5.4 *Recentering the Effect of* TIME 181

6 Modeling Discontinuous and Nonlinear Change 189
6.1 *Discontinuous Individual Change* 190
6.2 *Using Transformations to Model Nonlinear Individual Change* 208
6.3 *Representing Individual Change Using a Polynomial Function of* TIME 213
6.4 *Truly Nonlinear Trajectories* 223

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

A word about programming and software

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

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

Introducing the Multilevel Model for Change:
ALDA, Chapter Three

"When you're finished changing, you're finished"
Benjamin Disraeli

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

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*

(ALDA, Chapter 3 intro, p. 45)

Illustrative example: The effects of early intervention on children's IQ

Data source: Peg Burchinal and colleagues (2000) Child Development.

- *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?*

(ALDA, Section 3.1, pp. 46-49)

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

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")

ID	*AGE*	*COG*	*PROGRAM*
68	1.0	103	1
68	1.5	119	1
68	2.0	96	1
70	1.0	106	1
70	1.5	107	1
70	2.0	96	1
71	1.0	112	1
71	1.5	86	1
71	2.0	73	1
72	1.0	100	1
72	1.5	93	1
72	2.0	87	1
...	...	...	...
902	1.0	119	0
902	1.5	93	0
902	2.0	99	0
904	1.0	112	0
904	1.5	98	0
904	2.0	79	0
906	1.0	89	0
906	1.5	66	0
906	2.0	81	0
908	1.0	117	0
908	1.5	90	0
908	2.0	76	0
...	...	...	...

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*

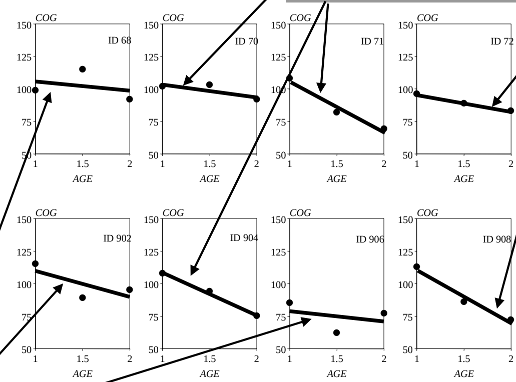
(ALDA, Section 3.1, pp. 46-49)

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*

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

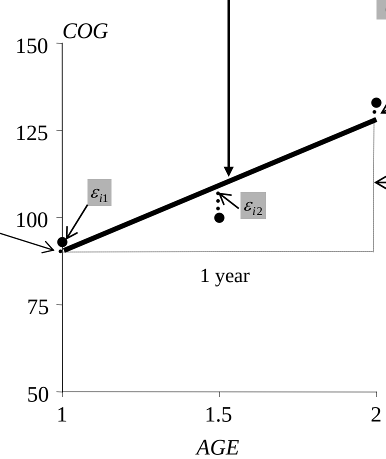
Key assumption: In the population, COG_{ij} *is a linear function of child i's AGE on occasion j*

$$COG_{ij} = \left[\boxed{\pi_{0i}} + \boxed{\pi_{1i}}(AGE_{ij} - 1)\right] + \left[\varepsilon_{ij}\right]$$

- ***i indexes persons (i=1 to 103)***
- ***j indexes occasions (j=1 to 3)***

Individual i's hypothesized true change trajectory

ε_{i1}, ε_{i2}, *and* ε_{i3} *are deviations of i's true change trajectory from linearity on each occasion (measurement error)*

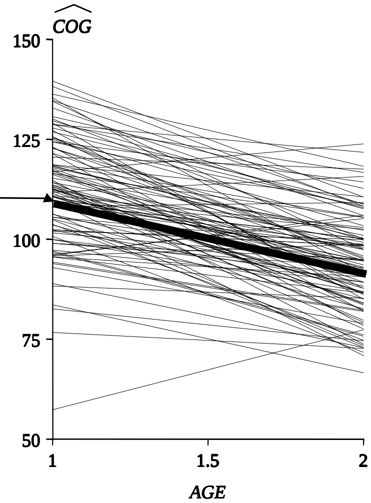


π_{0i} *is the* ***intercept*** *of i's true change trajectory, his true value of COG at AGE=1, his "true initial status"*

π_{1i} *is the* ***slope*** *of i's true change trajectory, his yearly rate of change in true COG, his true "annual rate of change"*

Examining 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 ≅ 110-10(AGE-1)

But there's also great variation in these OLS estimates

Fitted initial status

```
14   0
13*  5568
13.  00134
12*  5556778999
12.  02233344
11*  55667777888889
11.  0001111122222233334444
10*  55666688999
10.  0012222244
 9*  6666677799
 9.  344
 8*  89
 8.  34
 7*  7
 7.
 6*
 6.
 5*  7
```

Fitted rate of change

```
 2.  0
 1*
 1.  0
 0*  79
 0.  134
-0*  4444332
-0.  99998888777765
-1*  4333322211000
-1.  99888877666655
-2*  44322211110000
-2.  9999877776655
-3*  443322100000
-3.  987
-4*  443111
```

Residual variance

```
46  8
44
42
40  00
38
36  8
34
32  3
30
28  4
26  7
24  1444
22  8
20
18  3
16  00011
14
12  21
10  44433
 8  1118886666
 6  77744
 4  333844
 2  044448888333388888888
 0  000011112223333444444446666811111114447
```

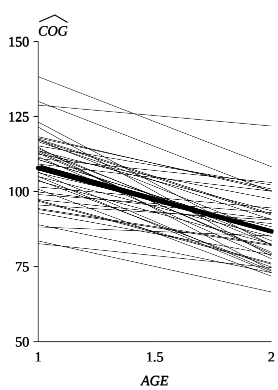
What does this behavior mean for a level-2 model?

- *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)*

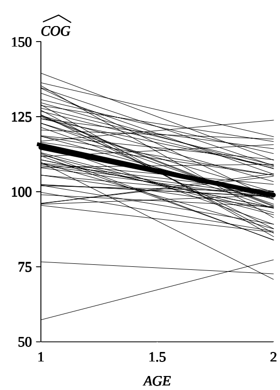
(ALDA, Section 3.2.3, pp. 55-56)

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

PROGRAM=0



PROGRAM=1



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*

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

1. *Outcomes are the level-1 individual growth parameters π_{0i} and π_{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*

(ALDA, Section 3.3, pp. 57-60)

Level-2 submodels for systematic interindividual differences in change

For the level-1 intercept *(initial status)*

$$\pi_{0i} = \boxed{\gamma_{00}} + \boxed{\gamma_{01}} PROGRAM + \zeta_{0i}$$

For the level-1 slope *(rate of change)*

$$\pi_{1i} = \boxed{\gamma_{10}} + \boxed{\gamma_{11}} PROGRAM + \zeta_{1i}$$

Symbol	Definition	Illustrative interpretation
γ_{00}	Population average of the level-1 intercepts, π_{0i}, for individuals with a level-2 predictor value of 0.	Population average true initial status for nonparticipants.
γ_{01}	Population average difference in level-1 intercept, π_{0i}, for a 1-unit difference in the level-2 predictor.	Difference in population average true initial status between participants and nonparticipants.
γ_{10}	Population average of the level-1 slopes, π_{1i}, for individuals with a level-2 predictor value of 0.	Population average annual rate of true change for nonparticipants.
γ_{11}	Population average difference in level-1 slope, π_{1i}, for a 1-unit difference in the level-2 predictor.	Difference in population average annual rate of true change between participants and non-participants.

Key to remembering subscripts on the gammas (the γ'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 ζ'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* ➔

(ALDA, Section 3.3.1, pp. 60-61)

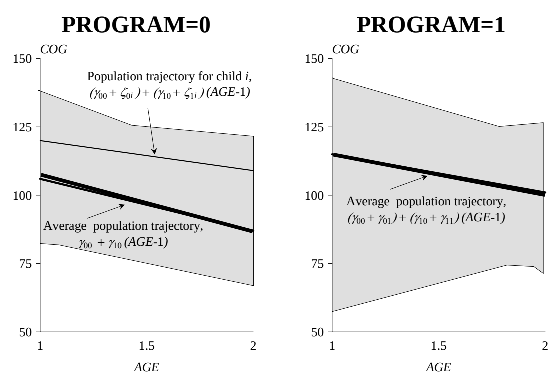
Understanding the stochastic components of the level-2 submodels

$$\pi_{0i} = \gamma_{00} + \gamma_{01} PROGRAM + \boxed{\zeta_{0i}}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11} PROGRAM + \boxed{\zeta_{1i}}$$

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 π_{0i} and π_{1i})***
- ***Shading suggests the existence of many true population trajectories, one per child***



Assumptions about the level-2 residuals:

initial status / rate of change

$$\begin{bmatrix}\zeta_{0i}\\ \zeta_{1i}\end{bmatrix} \sim N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\boxed{\sigma_0^2} & \sigma_{01}\\ \boxed{\sigma_{10}} & \boxed{\sigma_1^2}\end{bmatrix}\right)$$

Symbol	Definition	Illustrative interpretation
σ_0^2	Level-2 residual variance in true intercept, π_{0i}, across all individuals in the population.	Population residual variance of true initial status, controlling for program participation.
σ_1^2	*Level-2 residual variance in true slope*, π_{1i}, across all individuals in the population.	Population residual variance of true rate of change, controlling for program participation.
σ_{01}	Level-2 residual covariance between true intercept, π_{0i}, and true slope, π_{1i}, across all individuals in the population.	Population residual covariance between true initial status and true annual rate of change, controlling for program participation.

(ALDA, Section 3.3.2, pp. 61-63)

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

Specialty packages originally designed for another purpose that can also fit some multilevel models

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

		Parameter	Estimate	*ase*	*z*
Fixed Effects					
Initial status, π_{0i}	Intercept	γ_{00}	107.84***	2.04	52.97
	PROGRAM	γ_{01}	6.85*	2.71	2.53
Rate of change, π_{1i}	Intercept	γ_{10}	−21.13***	1.89	−11.18
	PROGRAM	γ_{11}	5.27*	2.52	2.09

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .00$

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 $\hat{\pi}_{0i} = 107.84 + 6.85 PROGRAM_i$

Fitted model for rate of change $\hat{\pi}_{1i} = -21.13 + 5.27 PROGRAM_i$

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!*

(ALDA, Section 3.5, pp. 68-71)

Plotting prototypical change trajectories

*<u>General idea</u>: **Substitute prototypical values for the level-2 predictors** (here, just PROGRAM=0 or 1) **into the fitted models***

$$\hat{\pi}_{0i} = 107.84 + 6.85 PROGRAM_i$$

$$\hat{\pi}_{1i} = -21.13 + 5.27 PROGRAM_i$$

PROGRAM = 0

$\hat{\pi}_{0i} = 107.84 + 6.85(\mathbf{0}) = 107.84$

$\hat{\pi}_{1i} = -21.13 + 5.27(\mathbf{0}) = -21.13$

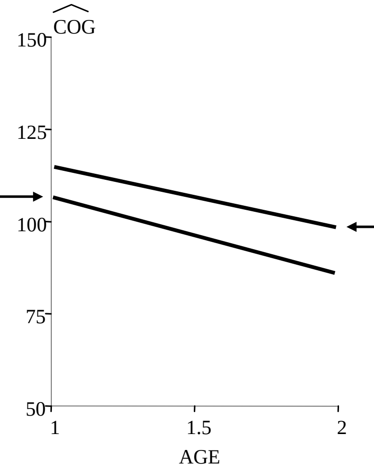
$so: \quad \hat{COG} = 107.84 - 21.13 AGE$

PROGRAM = 1

$\hat{\pi}_{0i} = 107.84 + 6.85(1) = 114.69$

$\hat{\pi}_{1i} = -21.13 + 5.27(1) = -15.86$

$so: \quad \hat{COG} = 114.69 - 15.86 AGE$



***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?*

(ALDA, Section 3.5.1, pp. 69-71)

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:

$$z = \frac{\hat{\gamma}}{ase(\hat{\gamma})}$$

		Parameter	Estimate	*ase*	*z*
Fixed Effects					
Initial status, π_{0i}	Intercept	γ_{00}	107.84***	2.04	52.97
	PROGRAM	γ_{01}	6.85*	2.71	2.53
Rate of change, π_{1i}	Intercept	γ_{10}	−21.13***	1.89	−11.18
	PROGRAM	γ_{11}	5.27*	2.52	2.09

~$p < .10$; *$p < .05$; **$p < .01$; ***$p < .00$

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").*

Careful:

Most programs provide appropriate tests but… ***different programs use different terminology***

Terms like z-statistic, t-statistic, t-ratio, quasi-t-statistic—which are not the same—are used interchangeably

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

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*

		Parameter	Estimate	*ase*	*z*
Variance Components					
Level 1:	Within-person, ε_{ij}	σ_ε^2	74.24***	10.34	7.17
Level 2:	In initial status, ζ_{0i}	σ_0^2	124.64***	27.38	4.55
	In rate of change, ζ_{1i}	σ_1^2	12.29	30.50	0.40
	Covariance between ζ_{0i} and ζ_{1i}	σ_{01}	−36.41	22.74	−1.60

~$p < .10$; *$p < .05$; **$p < .01$; ***$p < .001$.

Level-2 residual variance:

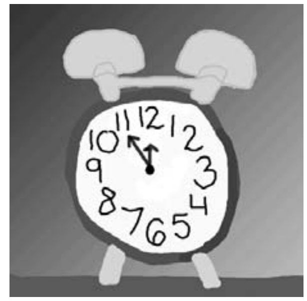
$$\begin{bmatrix} 124.64^{***} & -36.41 \\ -36.41 & 12.29 \end{bmatrix}$$

- *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)*

(ALDA, Section 3.6, pp. 72-74)

Doing data analysis with the multilevel model for change

ALDA, Chapter Four



"We are restless because of incessant change, but we would be frightened if change were stopped"
Lyman Bryson

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

Chapter 4: Doing data analysis with the multilevel model for change

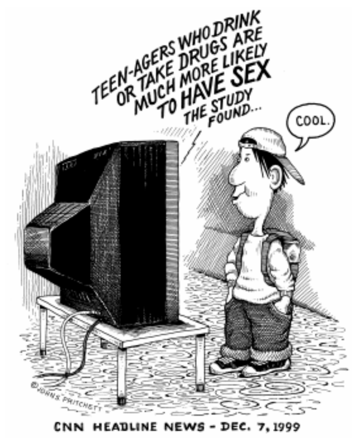
General Approach: ***Once again, we'll go through a worked example, but now we'll delve into the practical data analytic details***

- ***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

Data source: *Pat Curran and colleagues (1997)*
Journal of Consulting and Clinical Psychology.

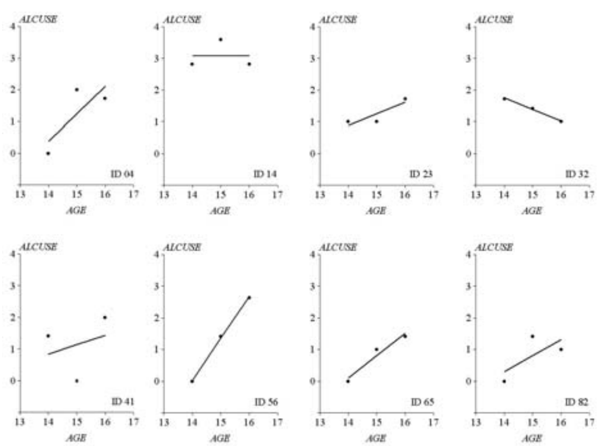
- *Sample: 82 adolescents*
  - *37 are children of an alcoholic parent (COAs)*
  - *45 are non-COAs*
- *Research design*
  - *Each was assessed 3 times—at ages 14, 15, and 16*
  - *The outcome, ALCUSE, was computed as follows:*
    - *4 items: (1) drank beer/wine; (2) hard liquor; (3) 5 or more drinks 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…

$$ALCUSE_{ij} = \pi_{0i} + \pi_{1i}(AGE_{ij} - 14) + \varepsilon_{ij} \quad \text{where } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

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

i's true initial status (ie, when TIME=0)

i's true rate of change per unit of TIME

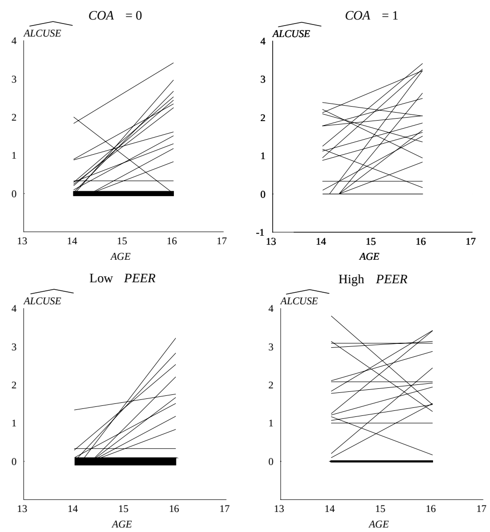
portion of i's outcome unexplained on occasion j

(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

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

$$\pi_{0i} = \gamma_{00} + \gamma_{01} COA_i + \zeta_{0i}$$ *(for initial status)*

$$\pi_{1i} = \gamma_{10} + \gamma_{11} COA_i + \zeta_{1i}$$ *(for rate of change)*

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

$$\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$$

(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

$$\pi_{0i} = \gamma_{00} + \gamma_{01} COA_i + \zeta_{0i} \qquad \pi_{1i} = \gamma_{10} + \gamma_{11} COA_i + \zeta_{1i}$$

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

$$Y_{ij} = \left(\gamma_{00} + \gamma_{01} COA_i + \zeta_{0i}\right) + \left(\gamma_{10} + \gamma_{11} COA_i + \zeta_{1i}\right) \mathrm{TIME}_{ij} + \varepsilon_{ij}$$

$$Y_{ij} = [\gamma_{00} + \gamma_{10} TIME_{ij} + \gamma_{01} COA_i + \gamma_{11} (COA_i \times TIME_{ij})] + [\zeta_{0i} + \zeta_{1i} TIME_{ij} + \varepsilon_{ij}]$$

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)*

The composite specification also:

- *Demonstrates the complexity of the composite residual—this is not regular OLS regression*
- *Is the specification used by most software packages for multilevel modeling*
- *Is the specification that maps most easily onto the person-period data set…* ⇨⇨⇨

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

The person-period data set and its relationship to the composite specification

ID	ALCUSE	AGE-14	COA	COA*(AGE-14)
3	1.00	0	1	0
3	2.00	1	1	1
3	3.32	2	1	2
4	0.00	0	1	0
4	2.00	1	1	1
4	1.73	2	1	2
44	0.00	0	0	0
44	1.41	1	0	0
44	3.00	2	0	0
66	1.41	0	0	0
66	3.46	1	0	0
66	3.00	2	0	0

$$\boxed{ALCUSE_{ij}} = [\gamma_{00} + \gamma_{10}\boxed{(AGE-14)_{ij}} + \gamma_{01}\boxed{COA_i} + \gamma_{11}\boxed{(COA_i \times (AGE-14)_{ij})}]$$
$$+[\zeta_{0i} + \zeta_{1i}(AGE-14)_{ij} + \varepsilon_{ij}]$$

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: **<u>Two unconditional models</u>***

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.*

<u>What these unconditional models tell us</u>:

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: $Y_{ij} = \pi_{0i} + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

Level-2 Model: $\pi_{0i} = \gamma_{00} + \zeta_{0i}$, where $\zeta_{0i} \sim N(0, \sigma_0^2)$

Composite Model: $Y_{ij} = \boxed{\gamma_{00}} + \boxed{\zeta_{0i}} + \boxed{\varepsilon_{ij}}$

Grand mean across individuals and occasions

Person-specific means

Within-person deviations

Within-person variance

Between-person variance

		Parameter	Model A
Fixed Effects			
Initial status, π_{0i}	Intercept	γ_{00}	0.922*** (0.096)
Variance Components			
Level 1	Within-person	σ_ε^2	0.562*** (0.062)
Level 2	In initial status	σ_0^2	0.564*** (0.119)

Let's look more closely at these variances….

(ALDA, Section 4.4.1, p. 92-97)

Using the unconditional means model to estimate the Intraclass Correlation Coefficient (ICC or ρ)

Major purpose of the ***unconditional means model:***
To partition the variation in Y into two components

Quantifies the amount of <u>*variation*</u> <u>***between*** *individuals regardless of time*</u>

$\hat{\sigma}_0^2 = 0.564^{***}$
Estimated between-person variance

$\hat{\sigma}_\varepsilon^2 = 0.562^{***}$
Estimated within-person variance

Quantifies the amount of <u>*variation*</u> <u>***within*** *individuals over time*</u>

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

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

$$\hat{\rho} = \frac{0.564}{0.564 + 0.562} = 0.50$$

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: $Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

Level-2 Model: $\pi_{0i} = \gamma_{00} + \zeta_{0i}$, $\pi_{1i} = \gamma_{10} + \zeta_{1i}$ *where* $\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$

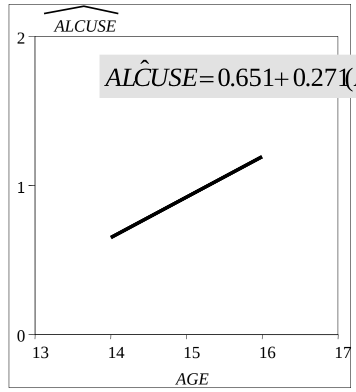
Composite Model: $Y_{ij} = \boxed{\gamma_{00}} + \boxed{\gamma_{10}} TIME_{ij} + \boxed{[\zeta_{0i} + \zeta_{1i} TIME_{ij} + \varepsilon_{ij}]}$

Composite residual

Average true initial status at AGE 14

Average true rate of change

		Parameter	Model B
Fixed Effects			
Initial status, π_{0i}	Intercept	γ_{00}	0.651*** (0.105)
Rate of change, π_{1i}	Intercept	γ_{10}	0.271*** (0.062)



What about the variance components?

(ALDA, Section 4.4.2, pp 97-102)

The unconditional growth model: Interpreting the variance components

		Parameter	Model B
Variance Components			
Level 1	Within-person	σ_ε^2	0.337*** (0.053)
Level 2	In initial status	σ_0^2	0.624*** (0.148)
	In rate of change	σ_1^2	0.151** (0.056)
	Covariance	σ_{01}	–0.068 (0.070)

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

		Parameter	Model A	Model B
Variance Components				
Level 1	Within-person	σ_ε^2	0.562*** (0.062)	0.337*** (0.053)
Level 2	In initial status	σ_0^2		
	In rate of change	σ_1^2		0.151** (0.056)
	Covariance	σ_{01}		−0.068 (0.070)
Pseudo R^2 Statistics and Goodness-of-fit				
		$R^2_{y\hat{y}}$		.043
		R^2_ε		.40

$$R_\varepsilon^2 = \left(\begin{array}{c} \text{Proportional reduction in the} \\ \text{Level-1 variance component} \end{array} \right)$$

$$= \left(\frac{0.562 - 0.337}{.562} \right) = 0.40$$

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

$$R^2_{Y,\hat{Y}} = \left(\hat{r}_{Y,\hat{Y}}\right)^2 = (0.21)^2 = 0.043$$

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

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)*

$$Pseudo R_\zeta^2 = \frac{\hat{\sigma}_\zeta^2(\mathbf{Uncond\,Growth\,Model}) - \hat{\sigma}_\zeta^2(\mathbf{Later\,Growth\,Model})}{\hat{\sigma}_\zeta^2(\mathbf{Uncond\,Growth\,Model})}$$

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

(ALDA, Section 4.4.3, pp 102-104)

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*

(ALDA, Section 4.5.1, pp 105-106)

What will our analytic strategy be?

Because our research interest ***focuses on the effect of COA, essentially treating PEER is a control****, we're going to proceed as follows…*

CHILDREN OF ALCOHOLICS FOUNDATION

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.

	Level-1/level-2 specification	
Model	level-1 model	level-2 model
C	$Y_{ij} = \pi_{0i} + \pi_{1i}TIME_{ij} + \varepsilon_{ij}$	$\pi_{0i} = \gamma_{00} + \gamma_{01}COA_i + \zeta_{0i}$ $\pi_{1i} = \gamma_{10} + \gamma_{11}COA_i + \zeta_{1i}$
D	$Y_{ij} = \pi_{0i} + \pi_{1i}TIME_{ij} + \varepsilon_{ij}$	$\pi_{0i} = \gamma_{00} + \gamma_{01}COA_i + \gamma_{02}PEER_i + \zeta_{0i}$ $\pi_{1i} = \gamma_{10} + \gamma_{11}COA_i + \gamma_{12}PEER_i + \zeta_{1i}$
E	$Y_{ij} = \pi_{0i} + \pi_{1i}TIME_{ij} + \varepsilon_{ij}$	$\pi_{0i} = \gamma_{00} + \gamma_{01}COA_i + \gamma_{02}PEER_i + \zeta_{0i}$ $\pi_{1i} = \gamma_{10} + \gamma_{12}PEER_i + \zeta_{1i}$

(ALDA, Section 4.5.1, pp 105-106)

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

		Parameter	Model B	Model C
Fixed Effects				
Initial status, π_{0i}	Intercept	γ_{00}	0.651*** (0.105)	0.316*** (0.131)
	COA	γ_{01}		0.743*** (0.195)
	PEER	γ_{02}		
Rate of change, π_{2i}	Intercept	γ_{10}	0.271*** (0.062)	0.293*** (0.084)
	COA	γ_{11}		−0.049 (0.125)
	PEER	γ_{12}		
Variance Components				
Level 1	Within-person	σ_ε^2	0.337*** (0.053)	0.337*** (0.053)
Level 2	In initial status	σ_0^2	0.624*** (0.148)	0.488** (0.128)
	In rate of change	σ_1^2	0.151** (0.056)	0.151* (0.056)
	Covariance	σ_{01}	−0.068 (0.070)	−0.059 (0.066)
Pseudo R^2 Statistics and Goodness-of-fit				
	$R^2_{y\hat{y}}$		.043	.150
	R^2_ε		.40	.40
	R^2_0			.218
	R^2_1			.000
	Deviance		636.6	621.2
	AIC		648.6	637.2
	BIC		663.0	656.5

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Fixed effects

- *Est. initial value of ALCUSE for non-COAs 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*

(ALDA, Section 4.5.2, pp 107-108)

Model D: Assessing the controlled effects of COA (the question predictor)

		Parameter	Model C	Model D
Fixed Effects				
Initial status, π_{0i}	Intercept	γ_{00}	0.316*** (0.131)	−0.317*** (0.148)
	COA	γ_{01}	0.743*** (0.195)	0.579*** (0.162)
	PEER	γ_{02}		0.694*** (0.112)
Rate of change, π_{2i}	Intercept	γ_{10}	0.293*** (0.084)	0.429*** (0.114)
	COA	γ_{11}	−0.049 (0.125)	−0.014 (0.125)
	PEER	γ_{12}		−0.150~ (0.086)
Variance Components				
Level 1	Within-person	σ^2_ε	0.337*** (0.053)	0.337*** (0.053)
Level 2	In initial status	σ^2_0	0.488** (0.128)	0.241** (0.093)
	In rate of change	σ^2_1	0.151* (0.056)	0.139* (0.055)
	Covariance	σ_{01}	−0.059 (0.066)	−0.006 (0.055)
Pseudo R^2 Statistics and Goodness-of-fit				
	$R^2_{y\hat{y}}$		.150	.291
	R^2_ε		.40	.40
	R^2_0		.218	.614
	R^2_1		.000	.079
	Deviance		621.2	588.7
	AIC		637.2	608.7
	BIC		656.5	632.8

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Fixed effects of COA

- *Est. diff in ALCUSE between COAs and non-COAs, controlling for PEER, is 0.579 (p<.001)*
- *No sig. Difference in rate of change*

Fixed effects of PEER

- *Teens whose peers drink more at 14 also drink more at 14 (initial status)*
- *Modest neg effect on rate of change (p<.10)*

Variance components

- *Within person VC unchanged (as expected)*
- *Still sig. variation in both initial status and change—need other level-2 predictors*
- *Taken together, PEER and COA explain*
  - *61.4% of the variation in initial status*
  - *7.9% of the variation in rates of change*

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?*

(ALDA, Section 4.5.2, pp 108-109)

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

		Parameter	Model D	Model E
Fixed Effects				
Initial status, π_{0i}	Intercept	γ_{00}	−0.317*** (0.148)	−0.314*** (0.146)
	COA	γ_{01}	0.579*** (0.162)	0.571*** (0.146)
	PEER	γ_{02}	0.694*** (0.112)	0.695*** (0.111)
Rate of change, π_{2i}	Intercept	γ_{10}	0.429*** (0.114)	0.425*** (0.106)
	COA	γ_{11}	−0.014 (0.125)	
	PEER	γ_{12}	−0.150~ (0.086)	−0.151~ (0.085)
Variance Components				
Level 1	Within-person	σ^2_ε	0.337*** (0.053)	0.337*** (0.053)
Level 2	In initial status	σ^2_0	0.241** (0.093)	0.241** (0.093)
	In rate of change	σ^2_1	0.139* (0.055)	0.139* (0.055)
	Covariance	σ_{01}	−0.006 (0.055)	−0.006 (0.055)
Pseudo R^2 Statistics and Goodness-of-fit				
	$R^2_{y\hat{y}}$		.291	.291
	R^2_ε		.40	.40
	R^2_0		.614	.614
	R^2_1		.079	.079
	Deviance		588.7	588.7
	AIC		608.7	606.7
	BIC		632.8	628.4

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Fixed effects of COA

Controlling for PEER, the estimated diff in ALCUSE between COAs and non-COAs is 0.571 (p<.001)

Fixed effects of PEER

Controlling for COA, for each 1 pt difference in PEER, initial ALCUSE is 0.695 higher (p<.001) but rate of change in ALCUSE is 0.151 lower (p<.10)

***Variance components are unchanged** suggesting little is lost by eliminating the main effect of COA on rate of change (although there is still level-2 variance left to be predicted by other variables)*

***Partial covariance is indistinguishable from 0**. After controlling for PEER and COA, initial status and rate of change are unrelated*

(ALDA, Section 4.5.2, pp 109-110)

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

(ALDA, Section 4.5.1, pp 105-106)

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)

$$Model\ \ C: \begin{array}{l} \hat{\pi}_{0i} = 0.316 + 0.743COA \\ \hat{\pi}_{1i} = 0.293 - 0.049COA \end{array}$$

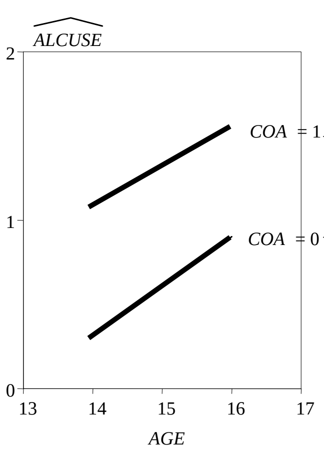
1. Substitute observed values for COA (0 and 1)

$$When\ COA_i = 0: \begin{cases} \hat{\pi}_{0i} = 0.316 + 0.743\boxed{(0)} = \boxed{0.316} \\ \hat{\pi}_{1i} = 0.293 - 0.049\boxed{(0)} = \boxed{0.293} \end{cases}$$

$$When\ COA_i = 1 \begin{cases} \hat{\pi}_{0i} = 0.316 + 0.743\boxed{(1)} = \boxed{1.059} \\ \hat{\pi}_{1i} = 0.293 - 0.049\boxed{(1)} = \boxed{0.244} \end{cases}$$

2. Substitute the estimated growth parameters into the level-1 growth model

$$when\ COA_i = 0: \quad \hat{Y}_{ij} = 0.316 + 0.293TIME$$
$$when\ COA_i = 1: \quad \hat{Y}_{ij} = 1.059 + 0.244TIME$$



What happens when the predictors aren't all dichotomous?

(ALDA, Section 4.5.3, pp 110-113)

Constructing prototypical fitted plots when some predictors are continuous

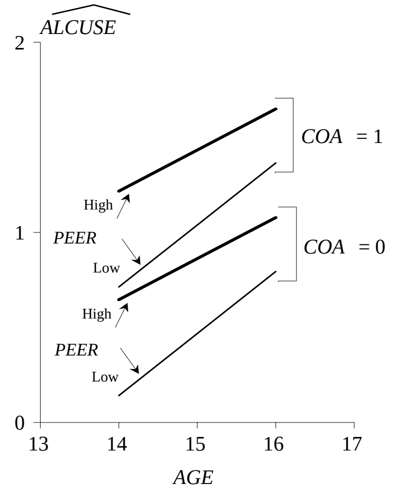
Key idea: ***Substitute "interesting" values of the continuous predictors into the fitted model and plot prototypical trajectories, by choosing***

- *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-.5(0.726) = ***0.655***

High PEER: 1.018+.5(0.726) = ***1.381***



Model E

$\hat{\pi}_{0i} = -0.314 + 0.695 PEER + 0.571 COA \qquad \hat{\pi}_{1i} = 0.425 - 0.151 PEER$

PEER	*COA*	Initial status ($\hat{\pi}_{0i}$)	Rate of change ($\hat{\pi}_{1i}$)
Low	No	−0.314 + 0.695(0.655) + 0.571(0) = 0.142	0.425 − 0.151(0.655) = 0.326
Low	Yes	−0.314 + 0.695(0.655) + 0.571(1) = 0.713	0.425 − 0.151(0.655) = 0.326
High	No	−0.314 + 0.695(1.381) + 0.571(0) = 0.646	0.425 − 0.151(1.381) = 0.216
High	Yes	−0.314 + 0.695(1.381) + 0.571(1) = 1.217	0.425 − 0.151(1.381) = 0.216

Intercepts for plotting

Slopes for plotting

(ALDA, Section 4.5.3, pp 110-113)

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*

		Parameter	Model E	Model F (*CPEER*)	Model G (*CCOA* & *CPEER*)
Fixed Effects Initial status, π_{0i}	Intercept	γ_{00}	−0.314*** (0.146)	0.394*** (0.104)	0.651*** (0.080)
	COA	γ_{01}	0.571*** (0.146)	0.571*** (0.146)	0.571*** (0.146)
	PEER	γ_{02}	0.695*** (0.111)	0.695*** (0.111)	0.695*** (0.111)
Rate of change, π_{1i}	Intercept	γ_{10}	0.425*** (0.106)	0.271*** (0.061)	0.271*** (0.061)
	COA	γ_{11}			
	PEER	γ_{12}	−0.151~ (0.085)	−0.151~ (0.085)	−0.151~ (0.085)
Variance Components Level 1	Within-person	σ_ε^2	0.337*** (0.053)	0.337*** (0.053)	0.337*** (0.053)
Level 2	In initial status	σ_0^2	0.241** (0.093)	0.241** (0.093)	0.241** (0.093)
	In rate of change	σ_1^2	0.139* (0.055)	0.139* (0.055)	0.139* (0.055)
	Covariance	σ_{01}	−0.006 (0.055)	−0.006 (0.055)	−0.006 (0.055)

Many estimates are unaffected by centering

At level-2, you can re-center by subtracting out:

- *The sample mean, which causes the level-2 intercepts to represent* ***average*** *fitted values (mean PEER=1.018; mean COA=0.451)*
- *Another meaningful value, e.g., 12 yrs of ed, IQ of 100*

*As expected, c****entering the level-2 predictors changes the level-2 intercepts***

F's intercepts describe an "average" non-COA

G's intercepts describe an "average" teen

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

(ALDA, Section 4.5.4, pp 113-116)

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 do 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$_{current}$*

(ALDA, Section 4.6, p 116)

Hypothesis testing using Deviance statistics

		Parameter	Model A	Model B
Fixed Effects				
Initial status, π_{0i}	Intercept	γ_{00}	0.922*** (0.096)	0.651*** (0.105)
	COA	γ_{01}		
	PEER	γ_{02}		
Rate of change, π_{1i}	Intercept	γ_{10}	☑	0.271*** (0.062)
	COA	γ_{11}		
	PEER	γ_{12}		
Variance Components				
Level 1	Within-person	σ_ε^2	0.562*** (0.062)	0.337*** (0.053)
Level 2	In initial status	σ_0^2	0.564*** (0.119)	0.624*** (0.148)
	In rate of change	σ_1^2	☑	0.151** (0.056)
	Covariance	σ_{01}	☑	−0.068 (0.070)
Pseudo R^2 Statistics and Goodness-of-fit				
	$R^2_{y\hat{y}}$			.043
	R^2_ε			.40
	R^2_0			
	R^2_1			
	Deviance		670.2	636.6
	AIC		676.2	648.6
	BIC		683.4	663.0

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

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 χ^2*
- *df = # of independent constraints*

1. We can obtain Model A from Model B by invoking 3 constraints:

$$H_0: \gamma_{10} = 0, \sigma_1^2 = 0, \sigma_{01} = 0$$

2: Compute difference in Deviance statistics and compare to appropriate χ^2 distribution

Δ Deviance = 33.55 (3 df, p<.001)

⇨ *reject H_0*

(ALDA, Section 4.6.1, pp 116-119)

Using deviance statistics to test more complex hypotheses

		Parameter	Model B	Model C
Fixed Effects				
Initial status, π_{0i}	Intercept	γ_{00}	0.651*** (0.105)	0.316*** (0.131)
	COA	γ_{01}	☑	0.743*** (0.195)
	PEER	γ_{02}		
Rate of change, π_{2i}	Intercept	γ_{10}	0.271*** (0.062)	0.293*** (0.084)
	COA	γ_{11}	☑	–0.049 (0.125)
	PEER	γ_{12}		
Variance Components				
Level 1	Within-person	σ_ε^2	0.337*** (0.053)	0.337*** (0.053)
Level 2	In initial status	σ_0^2	0.624*** (0.148)	0.488** (0.128)
	In rate of change	σ_1^2	0.151** (0.056)	0.151* (0.056)
	Covariance	σ_{01}	–0.068 (0.070)	–0.059 (0.066)
Pseudo R^2 Statistics and Goodness-of-fit				
	$R^2_{y\hat{y}}$		.043	.150
	R^2_ε		.40	.40
	R^2_0			.218
	R^2_1			.000
	Deviance		636.6	621.2
	AIC		648.6	637.2
	BIC		663.0	656.5

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

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 χ^2 distribution

Δ Deviance = 15.41 (2 df, p<.001)

⇨ reject H_0

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

(ALDA, Section 4.6.1, pp 116-119)

Comparing non-nested multilevel models using AIC and BIC

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

Information Criteria: AIC and BIC

Each information criterion "penalizes" the log-likelihood statistic for "excesses" in the structure of the current model

- *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

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

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

	Model A	Model B	Model C	Model D	Model E	Model F (*CPEER*)	Model G (*CCOA* & *CPEER*)
Deviance	670.2	636.6	621.2	588.7	588.7	588.7	588.7
AIC	676.2	648.6	637.2	608.7	606.7	606.7	606.7
BIC	683.4	663.0	656.5	632.8	628.4	628.4	628.4

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*

Careful: Gelman & Rubin (1995) declare these statistics and criteria to be "off-target and only by serendipity manage to hit the target"

(ALDA, Section 4.6.4, pp 120-122)

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 xtreg).*

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)

(ALDA, Section, 3.4, pp 63-68; Section 4.3, pp 85-92)

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*

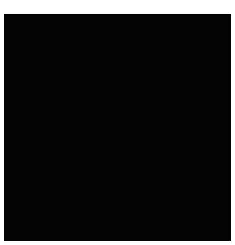
Extending the multilevel model for change
ALDA, Chapter Five

"Change is a measure of time"
Edwin Way Teale

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

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



- *Variably 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*

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?

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

ID	WAVE	AGEGRP	AGE	PIAT
04	1	6.5	6.00	18
04	2	8.5	8.50	31
04	3	10.5	10.67	50
27	1	6.5	6.25	19
27	2	8.5	9.17	36
27	3	10.5	10.92	57
31	1	6.5	6.33	18
31	2	8.5	8.83	31
31	3	10.5	10.92	51
33	1	6.5	6.33	18
33	2	8.5	8.92	34
33	3	10.5	10.75	29
41	1	6.5	6.33	18
41	2	8.5	8.75	28
41	3	10.5	10.83	36
49	1	6.5	6.50	19
49	2	8.5	8.75	32
49	3	10.5	10.67	48
69	1	6.5	6.67	26
69	2	8.5	9.17	47
69	3	10.5	11.33	45
77	1	6.5	6.83	17
77	2	8.5	8.08	19
77	3	10.5	10.00	28
87	1	6.5	6.92	22
87	2	8.5	9.42	49
87	3	10.5	11.50	64
...	...	...	...	...

We could build models of PIAT scores over time using ANY of these 3 measures for TIME—so ***which should we use****?*

Three different ways of coding TIME

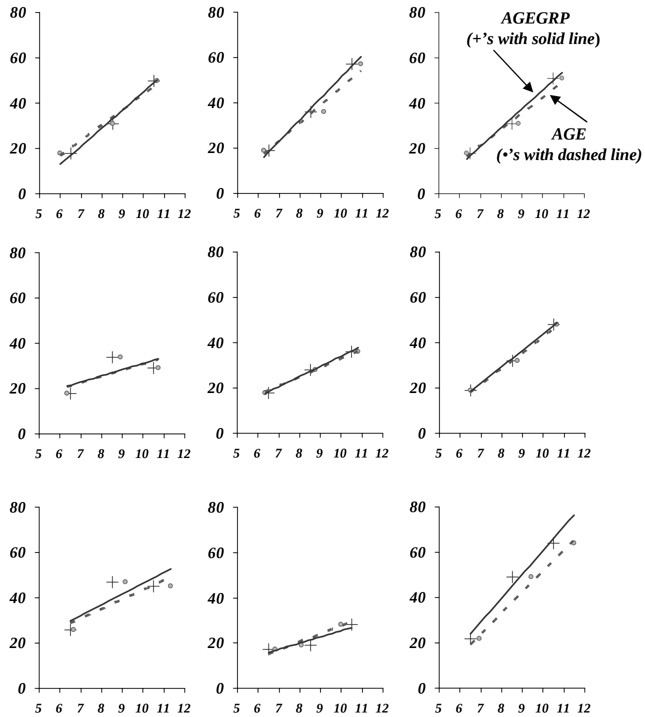
WAVE*—reflects design but has* **no substantive meaning**

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*

(ALDA, Section 5.1.1, pp 139-144)

Comparing OLS trajectories fit using AGEGRP and AGE



For many children—especially those assessed near the half-years—it makes little difference

?

Why ever use rounded AGE?

Note that this what we did in the past two examples, and so do lots of researchers!!!

For some children though—there's a big difference in slope, which is our conceptual outcome (rate of change)

(ALDA, Figure 5.1 p. 143)

Comparing models fit with AGEGRP and AGE

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

Level-2 Model: $\begin{matrix} \pi_{0i} = \gamma_{00} + \zeta_{0i} \\ \pi_{1i} = \gamma_{10} + \zeta_{1i} \end{matrix}$ where $\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$

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

By writing the level-1 model using the generic predictor TIME, the specification is identical

		Parameter	Predictor representing *TIME*: *AGEGRP* – 6.5	*AGE* – 6.5
Fixed Effects				
Initial status, π_{0i}	Intercept	γ_{00}	21.1629*** (0.6143)	21.0608*** (0.5593)
Rate of Change, π_{1i}	Intercept	γ_{10}	5.0309*** (0.2956)	4.5400*** (0.2606)
Variance Components				
Level-1:	within-person	σ_ε^2	27.04***	27.45***
Level-2:	In initial status	σ_0^2	11.05*	5.11
	In rate of change	σ_1^2	4.40***	3.30***
Goodness-of-fit				
	Deviance		1819.8	1803.9
	AIC		1831.9	1815.9
	BIC		1846.9	1830.8

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

AIC and BIC better with AGE

Some parameter estimates are virtually identical

Other est's larger with AGEGRP

- $\hat{\gamma}_{10}$ *, the slope, is ½ pt larger*
- *cumulates to a 2 pt diff over 4 yrs*
- *Level-2 VCs are also larger*
- *AGEGRP associates the data from later waves with earlier ages than observed, making the slope steeper*
- *Unexplained variation for initial status is associated with real AGE*

Treating an unstructured data set as structured introduces error into the analysis

(ALDA, Section 5.1.2, pp 144-146)

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

ID 206 *has 3 waves*

ID 332 *has 10 waves*

ID 1028 *has 7 waves*

ID	EXPER	LNW	BLACK	HGC
206	1.874	2.028	0	10
206	2.814	2.297	0	10
206	4.314	2.482	0	10
332	0.125	1.630	0	8
332	1.625	1.476	0	8
332	2.413	1.804	0	8
332	3.393	1.439	0	8
332	4.470	1.748	0	8
332	5.178	1.526	0	8
332	6.082	2.044	0	8
332	7.043	2.179	0	8
332	8.197	2.186	0	8
332	9.092	4.035	0	8
1028	0.004	0.872	1	8
1028	0.035	0.903	1	8
1028	0.515	1.389	1	8
1028	1.483	2.324	1	8
1028	2.141	1.484	1	8
1028	3.161	1.705	1	8
1028	4.103	2.343	1	8

# waves	*N men*
1	*38*
2	*39*
3-4	*82*
5-6	*166*
7-8	*226*
9-10	*240*
>10	*97*

EXPER = specific moment (to the nearest day) in each man's labor force history
- *Varying # of waves*
- *Varying spacing*

LNW in constant dollars seems to rise over time

Covariates: Race and Highest Grade Completed

(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 %age change in Y per annum

		Parameter	Model A	Model B	Model C
Fixed Effects					
Initial status, π_{0i}	Intercept	γ_{00}	1.7156*** (0.0108)	1.7171*** (0.0125)	1.7215*** (0.0107)
	$(HGC-9)$	γ_{01}		0.0349*** (0.0079)	0.0384*** (0.0064)
	BLACK	γ_{02}		0.0154 (0.0239)	
Rate of change, π_{1i}	Intercept	γ_{10}	0.0457*** (0.0023)	0.0493*** (0.0026)	0.0489*** (0.0025)
	$(HGC-9)$	γ_{11}		0.0013 (0.0017)	
	BLACK	γ_{12}		−0.0182** (0.0055)	−0.0161*** (0.0045)
Variance Components					
Level-1:	within-person	σ_ε^2	0.0951***	0.0952***	0.0952***
Level-2:	In initial status	σ_0^2	0.0543***	0.0518***	0.0518***
	In rate of change	σ_1^2	0.0017***	0.0016***	0.0016***
Goodness-of-fit					
	Deviance		4921.4	4873.8	4874.7
	AIC		4933.4	4893.8	4890.7
	BIC		4962.1	4941.7	4929.0

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Model C: an intermediate "final" model
- *Almost identical Deviance as Model B*
- *Effect of HGC—dropouts who stay in school longer earn higher wages on labor force entry (~4% higher per yr of school)*
- *Effect of BLACK—in contrast to Whites and Latinos, the wage of Black males increase less rapidly with labor force experience*
  - *Rate of change for Whites and Latinos is $100(e^{0.489}-1)=5.0\%$*
  - *Rate of change for Blacks is $100(e^{0.489-0.0161}-1)=3.3\%$*
- *Significant level-2 VCs indicate that there's still unexplained variation—this is hardly a 'final' model*

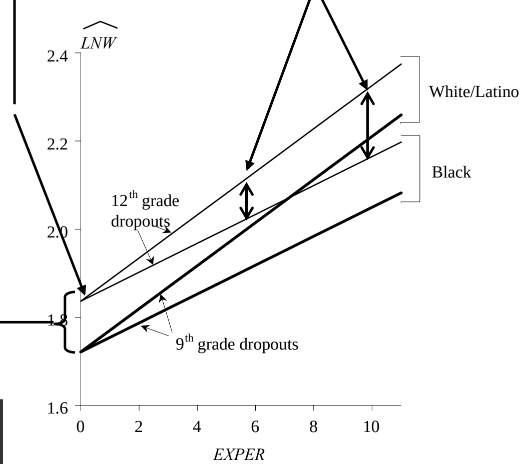
Fully specified growth model (both HGC & BLACK)
- *HGC is associated with initial status (but not change)*
- *BLACK is associated with change (but not initial status)*

⇒ Fit Model C, which removes non-significant parameters

(ALDA, Table 5.4 p. 149)

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)*

(ALDA, Section 5.2.1 and 5.2.2, pp150-156)

Practical advice: Problems can arise when analyzing unbalanced data sets

The multilevel model for change is <u>designed</u> to handle unbalanced data sets, and in most circumstances, it does its job well, however…

- *<u>When imbalance is severe, or lots of people have just 1 or 2 waves of data, problems can occur</u>*
  - *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").*
- *<u>Software packages may not issue clear warning signs</u>*
  - *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**

(ALDA, Section 5.2.2, pp151-156)

Example for illustrating time-varying predictors: Unemployment & depression

<u>Source</u>: *Liz Ginexi and colleagues (2000), <u>J of Occupational Health Psychology</u>*

- *<u>Sample: 254 people identified at unemployment offices.</u>*
- *<u>Research design:</u> Goal was to collect 3 waves of data per person at 1, 5 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)*
- *<u>Research question</u>*
  - *How does unemployment affect depression symptomatology?*

(ALDA, Section 5.3..1, pp160-161)

A person-period data set with a time-varying predictor

TIME=MONTHS since job loss

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

***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*

ID	*MONTHS*	*CES-D*	*UNEMP*
7589	1.3142	36	1
7589	5.0924	40	1
7589	11.7947	39	1
55697	1.3471	7	1
55697	5.7823	4	1
65641	0.3285	32	1
65641	4.1068	9	0
65641	10.9405	10	0
65441	1.0842	27	1
65441	4.6982	15	1
65441	11.2690	7	0
53782	0.4271	22	1
53782	4.2382	15	0
53782	11.0719	21	1

(ALDA, Table 5.6, p161)

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

$$Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \varepsilon_{ij} \text{, where } \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

Model B: Adding the main effect of UNEMP

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

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

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

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

$$Y_{ij} = \gamma_{00} + \gamma_{20} UNEMP_{ij} + \gamma_{30} UNEMP_{ij} \times TIME_{ij} + [\zeta_{0i} + \zeta_{2i} UNEMP_{ij} + \zeta_{3i} UNEMP_{ij} \times TIME_{ij} + \varepsilon_{ij}]$$

(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: $Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

Level-2 Model: $\begin{matrix} \pi_{0i} = \gamma_{00} + \zeta_{0i} \\ \pi_{1i} = \gamma_{10} + \zeta_{1i} \end{matrix}$ $\quad where \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$

How can it go at level-2???

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

It seems like it can go here

		Parameter	Model A
Fixed Effects			
Composite model	Intercept (initial status)	γ_{00}	17.6694** (0.7756)
	TIME (rate of change)	γ_{10}	−0.4220*** (0.0830)
Variance Components			
Level-1:	Within-person	σ_ε^2	68.85***
Level-2:	In intercept	σ_0^2	86.85***
	In rate of change	σ_1^2	0.36*

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

On the first day of job loss, the average person has an estimated CES-D of 17.7

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 do we add the time-varying predictor UNEMP?

(ALDA, Section 5.3.1, pp 159-164)

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

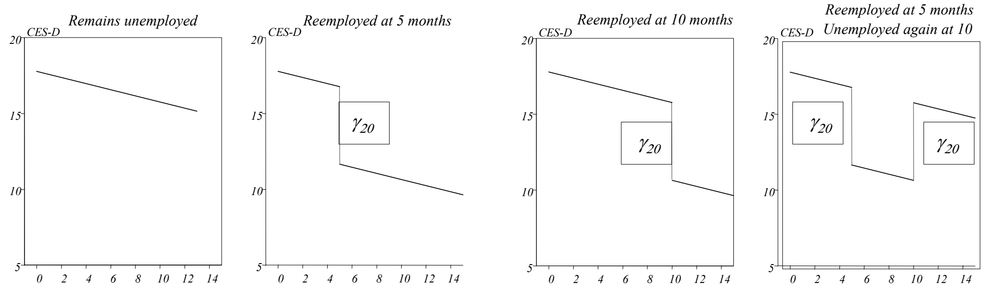
$$Y_{ij} = \boxed{\gamma_{00}} + \boxed{\gamma_{10}} TIME_{ij} + \boxed{\gamma_{20}} UNEMP_{ij} + [\zeta_{0i} + \zeta_{1i} TIME_{ij} + \varepsilon_{ij}]$$

Logical impossibility

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

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

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:*



What happens when we fit Model B to data?

(ALDA, Section 5.3.1, pp 159-164)

Fitting and interpreting Model B, which includes the TV predictor UNEMP

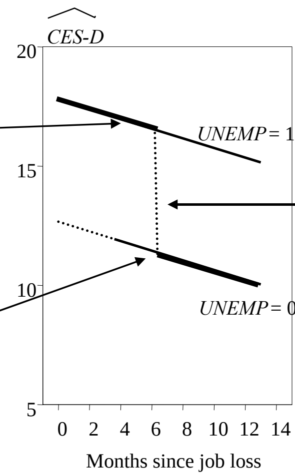
		Parameter	Model A	Model B
Fixed Effects				
Composite model	Intercept (initial status)	γ_{00}	17.6694** (0.7756)	12.6656*** (1.2421)
	TIME (rate of change)	γ_{10}	−0.4220*** (0.0830)	−0.2020* (0.0933)
	UNEMP	γ_{20}		5.1113*** (0.9888)
	UNEMP by *TIME*	γ_{30}		
Goodness-of-fit				
	Deviance		5133.1	5107.6
	AIC		5145.1	5121.6
	BIC		5166.3	5146.4

~$p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$.

Monthly rate of decline is cut in half by controlling for UNEMP (still sig.)

UNEMP has a large and stat sig effect

Model A is a much poorer fit (Δ Deviance = 25.5, 1 df, p<.001)



Consistently unemployed (UNEMP=1):

$$\hat{Y}_j = (12.6656 + 5.1113) - 0.2020 MONTHS_j$$

$$\hat{Y}_j = 17.7769 - 0.2020 MONTHS_j$$

Consistently employed (UNEMP=0):

$$\hat{Y}_j = 12.6656 - 0.2020 MONTHS_j$$

What about people who get a job?

What about the variance components?

(ALDA, Section 5.3.1, pp. 162-167)

Variance components behave differently when you're working with TV predictors

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!*

		Parameter	Model A	Model B
Fixed Effects				
Composite model	Intercept (initial status)	γ_{00}	17.6694** (0.7756)	12.6656*** (1.2421)
	TIME (rate of change)	γ_{10}	−0.4220*** (0.0830)	−0.2020* (0.0933)
	UNEMP	γ_{20}		5.1113*** (0.9888)
Variance Components				
Level-1:	Within-person	σ_ε^2	68.85***	62.39***
Level-2:	In intercept	σ_0^2	86.85***	93.52***
	In rate of change	σ_1^2	0.36*	0.46**

~$p<.10$; * $p<.05$; ** $p<.01$; *** $p<.001$.

Look what happened to the Level-2 VC's

- *In this example, they've **increased!***
- *Why?: Because including a TV predictor changes the meaning of the individual growth parameters (e.g., the intercept now refers to the value of the outcome when **all** level-1 predictors, including UNEMP are 0).*

Level-1 VC, σ_ε^2

- *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*

We can clarify what's happened by decomposing the composite specification back into a Level 1/Level-2 representation

(ALDA, Section 5.3.1, pp. 162-167)

Decomposing the composite specification of Model B into a L1/L2 specification

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

Level-1 Model: $Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \boxed{\pi_{2i} UNEMP_{ij}} + \varepsilon_{ij}$

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

Level-2 Models:

$$\pi_{0i} = \gamma_{00} + \zeta_{0i}$$
$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$
$$\pi_{2i} = \gamma_{20} \boxed{}$$

- *Model B's level-2 model for π_{2i} has no residual!*
- *Model B automatically assumes that π_{2i} 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 ➢➢*

(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: $Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \pi_{2i} UNEMP_{ij} + \varepsilon_{ij}$

Level-2 Models:

$$\pi_{0i} = \gamma_{00} + \zeta_{0i}$$
$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$
$$\pi_{2i} = \gamma_{20} + \boxed{\zeta_{2i}}$$

- *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….*

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!!*

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \quad and \quad \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \\ \zeta_{2i} \end{bmatrix} \sim N\left(\begin{matrix} 0 \\ 0, \\ 0 \end{matrix} \begin{bmatrix} \sigma_0^2 & \sigma_{01} & \sigma_{02} \\ \sigma_{10} & \sigma_1^2 & \sigma_{12} \\ \sigma_{20} & \sigma_{21} & \sigma_2^2 \end{bmatrix} \right)$$

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?***

(ALDA, Section 5.3.1, pp. 169-171)

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

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

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)

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

		Parameter	Model B	Model C
Fixed Effects				
Composite model	Intercept (initial status)	γ_{00}	12.6656*** (1.2421)	9.6167*** (1.8893)
	TIME (rate of change)	γ_{10}	−0.2020* (0.0933)	0.1620 (0.1937)
	UNEMP	γ_{20}	5.1113*** (0.9888)	8.5291*** (1.8779)
	UNEMP by *TIME*	γ_{30}		−0.4652* (0.2172)
Variance Components				
Level-1:	Within-person	σ_ε^2	62.39***	62.03***
Level-2:	In intercept	σ_0^2	93.52***	93.71***
	In rate of change	σ_1^2	0.46**	0.45**
Goodness-of-fit				
	Deviance		5107.6	5103.0
	AIC		5121.6	5119.7
	BIC		5146.4	5147.3

~p < .10; *p < .05; **p < .01; ***p < .001.

Main effect of TIME is now positive (!) & not stat sig ?!?!?!?!?!?!?!?!

*UNEMP*TIME interaction is stat sig (p<.05)*

Model B is a much poorer fit than C (Δ Deviance = 4.6, 1 df, p<.05)

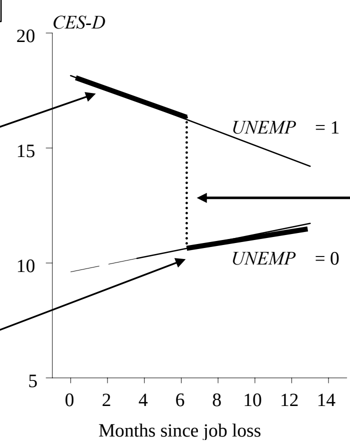
Consistently unemployed (UNEMP=1)

$$\hat{Y}_j = (9.6167 + 8.5291) + 0.(0.1620 - 0.4652) MONTHS_j$$

$$\hat{Y}_j = 18.1458 - 0.3032 MONTHS_j$$

Consistently employed (UNEMP=0)

$$\hat{Y}_j = 9.6167 + 0.1620 MONTHS_j$$



What about people who get a job?

Should the trajectory for the reemployed be constrained to 0?

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

How should we constrain the individual growth trajectory for the re-employed?

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

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

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

$$Y_{ij} = \gamma_{00} + \boxed{\gamma_{20} UNEMP_{ij}} + \boxed{\gamma_{30} UNEMP_{ij} \times TIME_{ij}} + [\zeta_{0i} + \boxed{\zeta_{3i} UNEMP_{ij} \times TIME_{ij}} + \varepsilon_{ij}]$$

*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)!

Model D:

$$Y_{ij} = \gamma_{00} + \boxed{\gamma_{20} UNEMP_{ij}} + \boxed{\gamma_{30} UNEMP_{ij} \times TIME_{ij}}$$
$$+ [\zeta_{0i} + \boxed{\zeta_{2i} UNEMP_{ij}} + \boxed{\zeta_{3i} UNEMP_{ij} \times TIME_{ij}} + \varepsilon_{ij}]$$

UNEMP has both a fixed & random effect

*UNEMP*TIME has both a fixed & random effect*

What happens when we fit Model D to data?

(ALDA, Section 5.3.2, pp. 172-173)

Model D: Constraining the individual growth trajectory among the reemployed

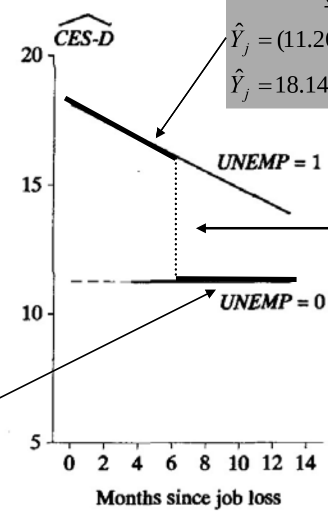
		Parameter	Model C	Model D
Fixed Effects				
Composite model	Intercept (initial status)	γ_{00}	9.6167***	11.2666***
			(1.8893)	(0.7690)
	TIME (rate of change)	γ_{10}	0.1620	
			(0.1937)	
	UNEMP	γ_{20}	8.5291***	6.8795***
			(1.8779)	(0.9133)
	UNEMP by *TIME*	γ_{30}	−0.4652*	−0.3254**
			(0.2172)	(0.1105)
Variance Components				
Level-1:	Within-person	σ_ε^2	62.03***	62.43***
Level-2:	In intercept	σ_0^2	93.71***	41.52***
	In rate of change	σ_1^2	0.45**	—
	In *UNEMP*	σ_2^2	—	40.45*
	In *UNEMP* by *TIME*	σ_3^2	—	0.71**
Goodness-of-fit				
	Deviance		5103.0	5093.6
	AIC		5119.7	5113.6
	BIC		5147.3	5148.9

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

Best fitting model (lowest AIC and BIC)

Consistently unemployed

$$\hat{Y}_j = (11.2666 + 6.8795) - 0.3254 MONTHS_j$$
$$\hat{Y}_j = 18.1461 - 0.3254 MONTHS_j$$



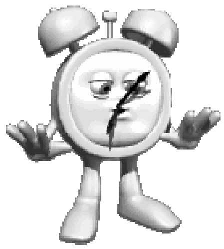
What about people who get a job?

Consistently employed

$$\hat{Y}_j = 11.2666$$

(ALDA, Section 5.3.2, pp. 172-173)

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*

(ALDA, Section 5.4, pp. 181-182)

Example for recentering the effects of TIME

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

- *<u>Sample: 73 men and women with major depression who were already being treated with non-pharmacological therapy</u>*
  - *Randomized trial to evaluate the efficacy of supplemental antidepressants (vs. placebo)*
- *<u>Research design</u>*
  - *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*
- *<u>Research question:</u>*
  - *How does POS change over time?*
  - *What is the effect of medication on the trajectories of change?*

(ALDA, Section 5.4, pp. 181-183)

How might we clock and code TIME?

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***

WAVE— Great for data processing—no intuitive meaning

WAVE	*DAY*	*READING*	*TIME OF DAY*	*TIME*	(*TIME* − 3.33)	(*TIME* − 6.67)
1	0	8 A.M.	0.00	$\boxed{0.00}$	−3.33	−6.67
2	0	3 P.M.	0.33	0.33	−3.00	−6.33
3	0	10 P.M.	0.67	0.67	−2.67	−6.00
4	1	8 A.M.	0.00	1.00	−2.33	−5.67
5	1	3 P.M.	0.33	1.33	−2.00	−5.33
6	1	10 P.M.	0.67	1.67	−1.67	−5.00
. . .						
11	3	3 P.M.	0.33	3.33	$\boxed{0.00}$	−3.33
. . .						
16	5	8 A.M.	0.00	5.00	1.67	−1.67
17	5	3 P.M.	0.33	5.33	2.00	−1.33
18	5	10 P.M.	0.67	5.67	2.33	−1.00
19	6	8 A.M.	0.00	6.00	2.67	−0.67
20	6	3 P.M.	0.33	6.33	3.00	−0.33
21	6	10 P.M.	0.67	6.67	3.33	$\boxed{0.00}$

READING—right idea, but how to quantify?

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

(TIME-6.67) *Same as TIME but now* ***centered on the study's endpoint***

(ALDA, Section 5.4, pp 181-183)

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: $Y_{ij} = \pi_{0i} + \pi_{1i}(TIME_{ij} - c) + \varepsilon_{ij}$, where $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

Level-2 Model:
$$\pi_{0i} = \gamma_{00} + \gamma_{01} TREAT_i + \zeta_{0i}$$
$$\pi_{1i} = \gamma_{10} + \gamma_{11} TREAT_i + \zeta_{1i}$$
$$\text{where} \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix} \right)$$

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

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

When c = 0:

- π_{0i} *is the individual mood at TIME=0*
- *Usually called "initial status"*

$Y_{ij} = \pi_{0i} + \pi_{1i}(TIME_{ij} - 3.33) + \varepsilon_{ij}$

When c = 3.33:

- π_{0i} *is the individual mood at TIME=3.33*
- *Useful to think of as"mid-experiment status"*

$Y_{ij} = \pi_{0i} + \pi_{1i}(TIME_{ij} - 6.67) + \varepsilon_{ij}$

When c = 6.67:

- π_{0i} *is the individual mood at TIME=6.67*
- *Useful to think about as "final status"*

(ALDA, Section 5.4, pp 182-183)

Comparing the results of using different centering constants for TIME

		Parameter	Temporal predictor in level-1 model: *TIME*	(*TIME* – 3.33)	(*TIME* – 6.67)
Fixed Effects					
Level-1 intercept, π_{0i}	Intercept	γ_{00}	167.46*** (9.33)	159.40*** (8.76)	151.34*** (11.54)
	TREAT	γ_{01}	−3.11 (12.33)	15.35 (11.54)	33.80* (15.16)
Rate of change, π_{1i}	Intercept	γ_{10}	−2.42 (1.73)	−2.42 (1.73)	−2.42 (1.73)
	TREAT	γ_{11}	5.54* (2.28)	5.54* (2.28)	5.54* (2.28)
Variance Components					
Level-1:	within-person	σ_ε^2	1229.93***	1229.93***	1229.93***
Level-2:	In level-1 intercept	σ_0^2	2111.33***	2008.72***	3322.45***
	In rate of change	σ_1^2	63.74***	63.74***	63.74***
	Covariance	σ_{01}	−121.62*	90.83	303.28***
Goodness-of-fit					
	Deviance		12680.5	12680.5	12680.5
	AIC		12696.5	12696.5	12696.5
	BIC		12714.8	12714.8	12714.8

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

The choice of centering constant has no effect on:

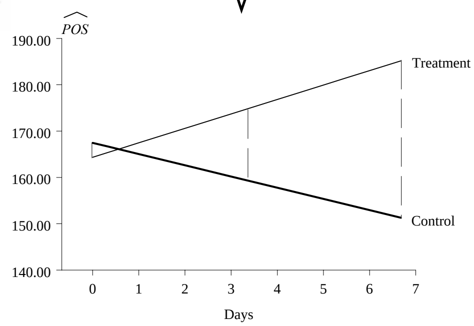
- *Goodness of fit indices*
- *Estimates for rates of change*
- *Within person residual variance*
- *Betw person res variance in rate of change*

<u>What are affected are the level-1 intercepts</u>

γ_{00} *assesses level of POS at time c for the control group (TREAT=0)*

γ_{01} *assesses the diff. in POS between the groups (TREATment effect)*

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



(ALDA, Section 5.4, pp 183-186)

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…

$$Y_{ij} = \pi_{0i}\left(\frac{6.67 - TIME_{ij}}{6.67}\right) + \pi_{1i}\left(\frac{TIME_{ij}}{6.67}\right) + \varepsilon_{ij}$$

Individual Initial Status Parameter

Individual Final Status Parameter

<u>Advantage</u>: You can use all your longitudinal data to analyze initial and final status simultaneously.

(ALDA, Section 5.4, pp 186-188)

Modeling discontinuous and nonlinear change
ALDA, Chapter Six

“Things have changed”
Bob Dylan

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

Chapter 6: Modeling discontinuous and nonlinear change

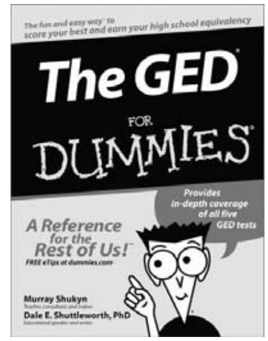
General idea: All our examples so far have assumed that individual growth is smooth and linear. But the multilevel model for change is much more flexible:

- ***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*

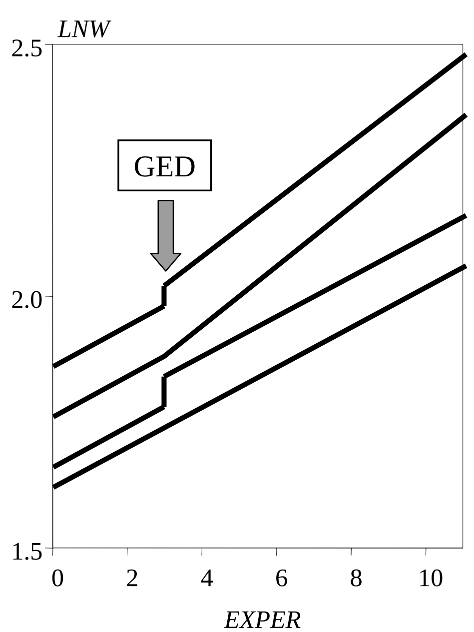
- *Sample: the same 888 male high school dropouts (from before)*
- *Research design*
  - *Each was interviewed between 1 and 13 times after dropping out*
  - *34.6% (n=307) earned a GED at some point during data collection*
- *OLD research questions*
  - *How do log(WAGES) change over time?*
  - *Do the wage trajectories differ by ethnicity and highest grade completed?*
- *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)*



F: Immediate shifts in both elevation & rate of change

D: An immediate shift in rate of change; no difference in elevation

B: An immediate shift in elevation; no difference in rate of change

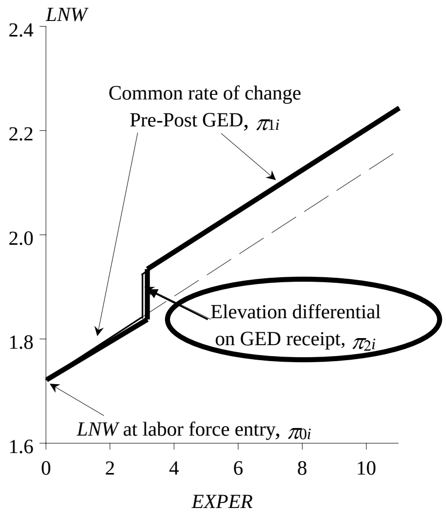
A: No effect of GED whatsoever

*How do we model trajectories like these within the context of a **linear** growth model??? →*

(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



$$Y_{ij} = \pi_{0i} + \pi_{1i} EXPER_{ij} + \pi_{2i} GED_{ij} + \varepsilon_{ij}$$

Post-GED (GED=1):

$$Y_{ij} = (\pi_{0i} + \pi_{2i}) + \pi_{1i} EXPER_{ij} + \varepsilon_{ij}$$

Pre-GED (GED=0):

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

(ALDA, Section 6.1.1, pp 194-195)

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

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

ID	*LNW*	*EXPER*	*GED*	*POSTEXP*
206	2.028	1.874	0	0
206	2.297	2.814	0	0
206	2.482	4.314	0	0
2365	1.782	0.660	0	0
2365	1.763	1.679	0	0
2365	1.710	2.737	0	0
2365	1.736	3.679	0	0
2365	2.192	4.679	1	0
2365	2.042	5.718	1	1.038
2365	2.320	6.718	1	2.038
2365	2.665	7.872	1	3.192
2365	2.418	9.083	1	4.404
2365	2.389	10.045	1	5.365
2365	2.485	11.122	1	6.442
2365	2.445	12.045	1	7.365
4384	2.859	0.096	0	0
4384	1.532	1.039	0	0
4384	1.590	1.726	1	0
4384	1.969	3.128	1	1.402
4384	1.684	4.282	1	2.556
4384	2.625	5.724	1	3.998
4384	2.583	6.024	1	4.298

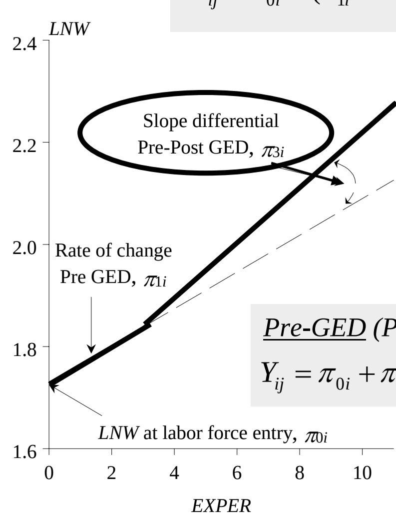
POSTEXP$_{ij}$ *= 0 prior to GED*
POSTEXP$_{ij}$ *= "Post GED experience," a new TV predictor that clocks "TIME since GED receipt" (in the same cadence as EXPER)*

$$Y_{ij} = \pi_{0i} + \pi_{1i} EXPER_{ij} + \pi_{3i} POSTEXP_{ij} + \varepsilon_{ij}$$

Post-GED (POSTEXP clocked in same cadence as EXPER):

$$Y_{ij} = \pi_{0i} + \pi_{1i} EXPER_{ij} + \pi_{3i} POSTEXP + \varepsilon_{ij}$$

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



Pre-GED (POSTEXP=0):

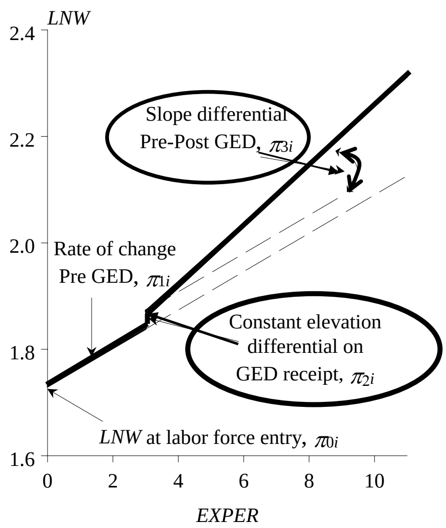
$$Y_{ij} = \pi_{0i} + \pi_{1i} EXPER_{ij} + \varepsilon_{ij}$$

(ALDA, Section 6.1.1, pp 195-198)

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

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

(ALDA, Section 6.1.1, pp 195-198)

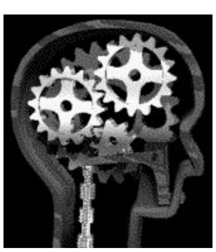
Many other types of discontinuous individual change trajectories are possible

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

How do we select among the alternative discontinuous models?

(ALDA, Section 6.1.1, pp199-201)

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*

Benchmark against which we'll evaluate discontinuous models

$$Y_{ij} = \pi_{0i} + \pi_{1i} EXPER_{ij} + \boxed{\pi_{2i}(UERATE_{ij} - 7)} + \varepsilon_{ij}$$

$$\pi_{0i} = \gamma_{00} + \gamma_{01}(HGC_i - 9) + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \gamma_{11} BLACK_i + \zeta_{1i}$$

$$\pi_{2i} = \gamma_{20}$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2) \quad \text{and} \quad \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \boxed{\sigma_{01}} \\ \sigma_{10} & \sigma_1^2 \end{bmatrix}\right)$$

		Parameter	Model A
Fixed Effects			
Initial status, π_{0i}	Intercept	γ_{00}	1.7490*** (0.0114)
	$(HGC - 9)$	γ_{01}	0.0400*** (0.0064)
	UERATE -7	γ_{20}	−0.0120*** (0.0018)
Rate of change, π_{1i}	Intercept	γ_{10}	0.0441*** (0.0026)
	BLACK	γ_{12}	−0.0182*** (0.0045)
Variance Components			
Level-1:	within-person	σ_ε^2	0.0948***
Level-2:	In initial status	σ_0^2	0.0506***
	In rate of change	σ_1^2	0.0016***
Goodness-of-fit			
	Deviance		4830.5
	AIC		4848.5
	BIC		4891.6

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

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

4 random effects

5 fixed effects

(ALDA, Section 6.1.2, pp 201-202)

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)

Baseline just shown

Model	Fixed effects	Variance components (in addition to σ_ε^2)	*n* parameters for . . . Fixed effects	Variance components	Deviance	Comparison model: ΔDeviance (*df*)
A	Intercept, *EXPER*, *HGC* – 9, *BLACK* × *EXPER*, *UERATE* – 7	Intercept, *EXPER*	5	4	4830.5	—
B	Model A + *GED*	Intercept, *EXPER*, *GED*	6	7	4805.5	A: 25.0*** (4)
C	Model B	Model B w/o *GED*	6	4	4818.3	B: 12.8** (3)
D	Model A + *POSTEXP*	Intercept, *EXPER*, *POSTEXP*	6	7	4817.4	A: 13.1** (4)
E	Model D	Model D w/o *POSTEXP*	6	4	4820.7	D: 3.3 (ns) (3)
F	Model A + *GED* and *POSTEXP*	Intercept, *EXPER*, *GED*, *POSTEXP*	7	11	4789.4	B: 16.2** (5) D: 28.1*** (5)
G	Model F	Model F w/o *POSTEXP*	7	7	4802.7	F: 13.3** (4)
H	Model F	Model F w/o *GED*	7	7	4812.6	F: 23.3*** (4)
I	Model A + *GED* and *GED* × *EXPER*	Intercept, *EXPER*, *GED*, *GED* × *EXPER*	7	11	4787.0	B: 18.5*** (5)
J	Model I	Model I w/o *GED* × *EXPER*	7	7	4804.6	I: 17.6** (4)

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

(ALDA, Section 6.1.2, pp 202-203)

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)
*ΔDeviance=25.0, 4 df, p<.001—**keep GED effect***

Model	Fixed effects	Variance components (in addition to σ_ε^2)	*n* parameters for . . . Fixed effects	Variance components	Deviance	Comparison model: ΔDeviance (*df*)
A	Intercept, *EXPER*, *HGC* – 9, *BLACK* × *EXPER*, *UERATE* – 7	Intercept, *EXPER*	5	4	4830.5	—
B	Model A + *GED*	Intercept, *EXPER*, *GED*	6	7	4805.5	A: 25.0*** (4)
C	Model B	Model B w/o *GED*	6	4	4818.3	B: 12.8** (3)

C: But does the GED discontinuity vary across people?
(do we need to keep the extra VCs for the effect of GED?)
*ΔDeviance=12.8, 3 df, p<.01— **keep VCs***

What about the discontinuity in slope?

(ALDA, Section 6.1.2, pp 202-203)

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)
*ΔDeviance=13.1, 4 df, p<.05— **keep POSTEXP effect***

Model	Fixed effects	Variance components (in addition to σ_ε^2)	*n* parameters for . . . Fixed effects	Variance components	Deviance	Comparison model: ΔDeviance (*df*)
A	Intercept, *EXPER*, *HGC* – 9, *BLACK* × *EXPER*, *UERATE* – 7	Intercept, *EXPER*	5	4	4830.5	—
B	Model A + *GED*	Intercept, *EXPER*, *GED*	6	7	4805.5	A: 25.0*** (4)
C	Model B	Model B w/o *GED*	6	4	4818.3	B: 12.8** (3)
D	Model A + *POSTEXP*	Intercept, *EXPER*, *POSTEXP*	6	7	4817.4	A: 13.1** (4)
E	Model D	Model D w/o *POSTEXP*	6	4	4820.7	D: 3.3 (ns) (3)

E: But does the POSTEXP slope vary across people?
(do we need to keep the extra VCs for the effect of POSTEXP?)
ΔDeviance=3.3, 3 df, 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?

(ALDA, Section 6.1.2, pp 203-204)

Examining both discontinuities simultaneously

Model	Fixed effects	Variance components (in addition to σ_ε^2)	*n* parameters for . . . Fixed effects	Variance components	Deviance	Comparison model: ΔDeviance (*df*)
A	Intercept, *EXPER*, *HGC* – 9, *BLACK* × *EXPER*, *UERATE* – 7	Intercept, *EXPER*	5	4	4830.5	—
B	Model A + *GED*	Intercept, *EXPER*, *GED*	6	7	4805.5	A: 25.0*** (4)
C	Model B	Model B w/o *GED*	6	4	4818.3	B: 12.8** (3)
D	Model A + *POSTEXP*	Intercept, *EXPER*, *POSTEXP*	6	7	4817.4	A: 13.1** (4)
E	Model D	Model D w/o *POSTEXP*	6	4	4820.7	D: 3.3 (ns) (3)
F	Model A + *GED* and *POSTEXP*	Intercept, *EXPER*, *GED*, *POSTEXP*	7	11	4789.4	B: 16.2** (5) D: 28.1*** (5)

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)?

Model	Fixed effects	Variance components (in addition to σ_ε^2)	*n* parameters for . . . Fixed effects	Variance components	Deviance	Comparison model: ΔDeviance (*df*)
F	Model A + *GED* and *POSTEXP*	Intercept, *EXPER*, *GED*, *POSTEXP*	7	11	4789.4	B: 16.2** (5) D: 28.1*** (5)
G	Model F	Model F w/o *POSTEXP*	7	7	4802.7	F: 13.3** (4)
H	Model F	Model F w/o *GED*	7	7	4812.6	F: 23.3*** (4)

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)

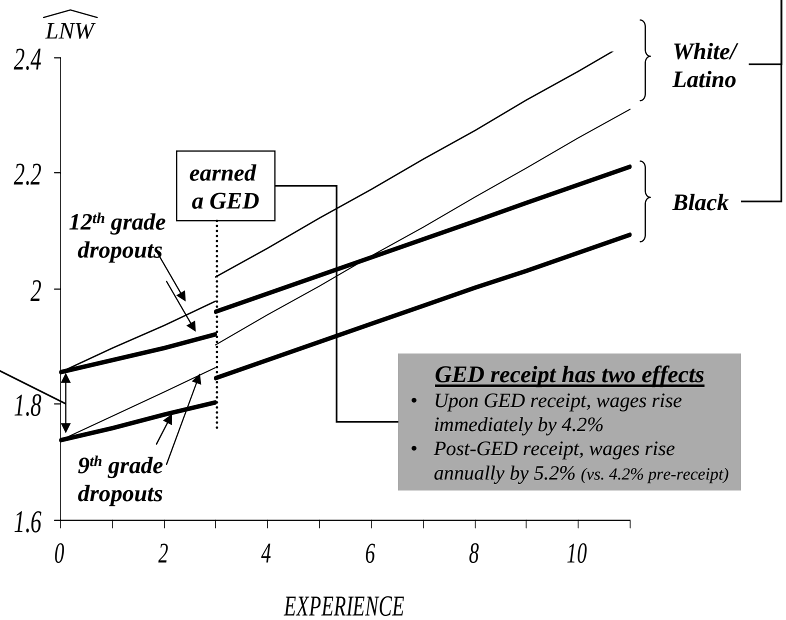
We actually fit several other possible models (see ALDA) but F was the best alternative—so…how do we display its results?

(ALDA, Section 6.1.2, pp 204-205)

Displaying prototypical discontinuous trajectories

(Log Wages for HS dropouts pre- and post-GED attainment)

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



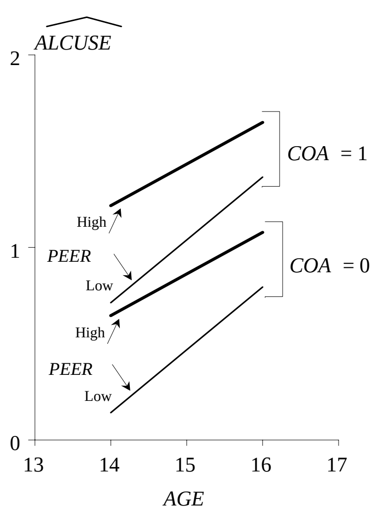
Highest grade completed
- *Those who stay longer have higher initial wages*
- *This differential remains constant over time*

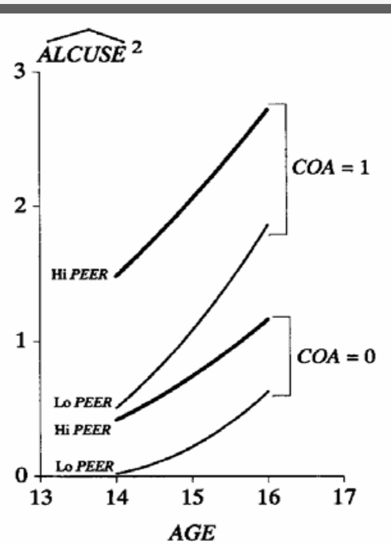
GED receipt has two effects
- *Upon GED receipt, wages rise immediately by 4.2%*
- *Post-GED receipt, wages rise annually by 5.2% (vs. 4.2% pre-receipt)*

(ALDA, Section 6.1.2, pp 204-206)

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*

(ALDA, Section 6.2, pp 208-210)

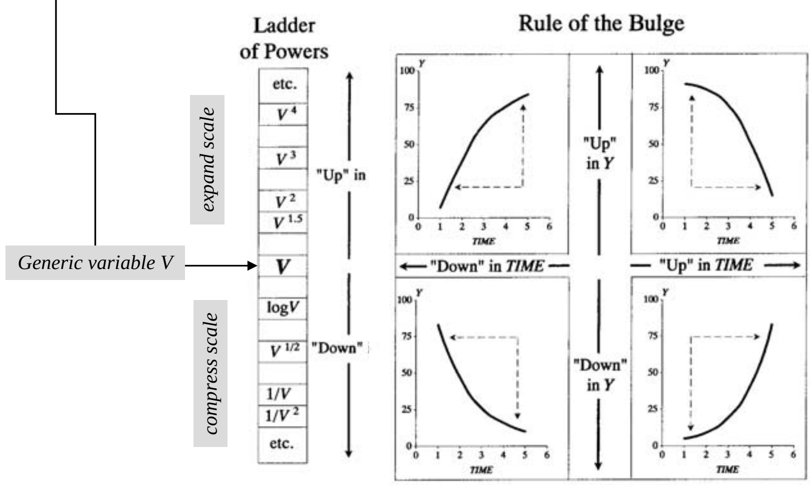
The “Rule of the Bulge” and the “Ladder of Transformations”

Mosteller & Tukey (1977): EDA techniques for straightening lines

Step 1: ***What kinds of transformations do we consider?***

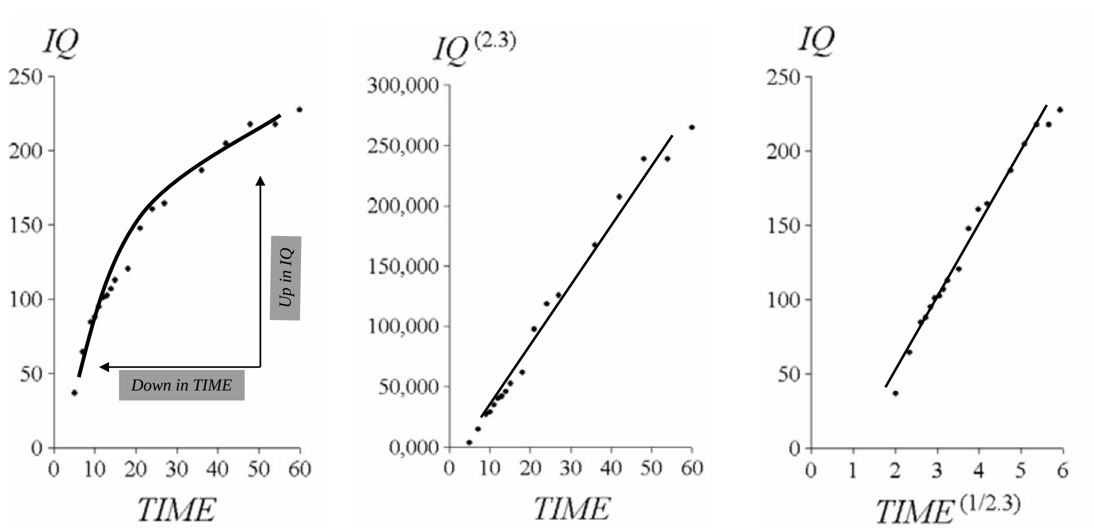
Step 2: ***How do we know when to use which transformation?***

1. *Plot many empirical growth trajectories*
2. *You find linearizing transformations by moving “up” or “down” in the direction of the “bulge”*



(ALDA, Section 6.2.1, pp. 210-212)

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



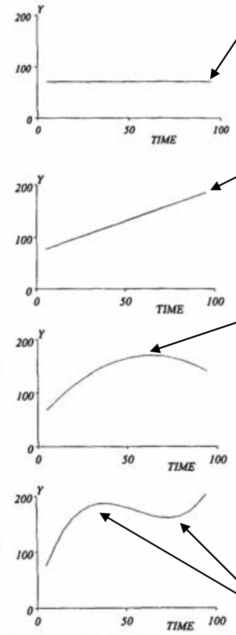
How else might we model non-linear change?

(ALDA, Section 6.2.1, pp. 211-213)

Representing individual change using a polynomial function of TIME

Shape	Level-1 model	Parameter values
No change	$Y_{ij} = \pi_{0i} + \varepsilon_{ij}$	$\pi_{0i} = 71$
Linear change	$Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \varepsilon_{ij}$	$\pi_{0i} = 71$, $\pi_{1i} = 1.2$
Quadratic change	$Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \pi_{2i} TIME_{ij}^2 + \varepsilon_{ij}$	$\pi_{0i} = 50$, $\pi_{1i} = 3.8$, $\pi_{2i} = -0.03$
Cubic change	$Y_{ij} = \pi_{0i} + \pi_{1i} TIME_{ij} + \pi_{2i} TIME_{ij}^2 + \pi_{3i} TIME_{ij}^3 + \varepsilon_{ij}$	$\pi_{0i} = 30$, $\pi_{1i} = 10$, $\pi_{2i} = -.2$, $\pi_{3i} = .0012$

Illustrative example — Plot of the true change trajectory



Polynomial of the "zero order" (because $TIME^0=1$)
- *Like including a constant predictor **1** 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 $TIME^2$*
- *π_{0i}=intercept, but now both TIME and $TIME^2$ must be 0*
- *π_{1i}=instantaneous rate of change when TIME=0 (there is no longer a constant slope)*
- *π_{2i}=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*

(ALDA, Section 6.3.1, pp. 213-217)

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?*

(ALDA, Section 6.3.2, p. 217)

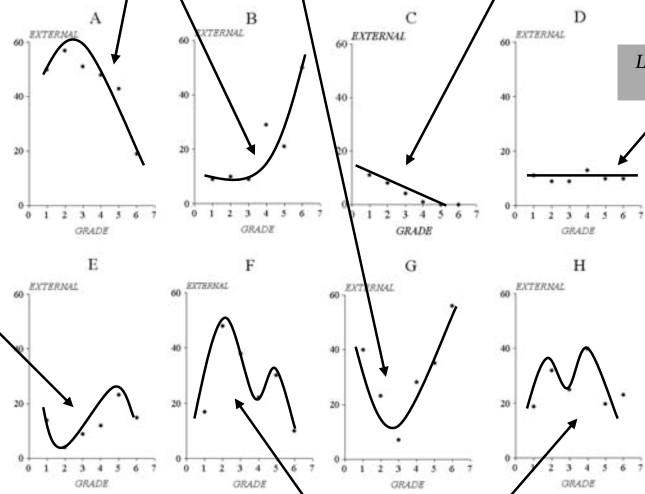
Selecting a suitable level-1 polynomial trajectory for change
Examining empirical growth plots (which invariably display great variability in temporal complexity)

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?

(ALDA, Section 6.3.2, pp 217-220)

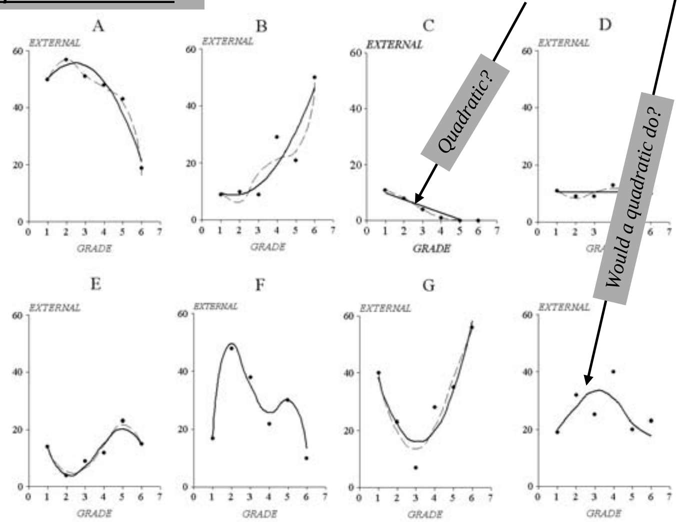
Examining alternative fitted OLS polynomial trajectories
Order optimized for each child (solid curves) and a common quartic across children (dashed line)

*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?***

Quadratic?

Would a quadratic do?



*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 ➔

(ALDA, Section 6.3.2, pp 217-220)

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)

		Parameter	Model A No change	Model B Linear change	Model C Quadratic change	Model D Cubic change
Fixed Effects						
Composite model	Intercept (1st grade status)	γ_{00}	12.96***	13.29***	13.97***	13.79***
	TIME (linear term)	γ_{10}		−0.13	−1.15	−0.35
	TIME^2 (quadratic term)	γ_{20}			0.20	−0.23
	TIME^3 (cubic term)	γ_{30}				0.06
Variance Components						
Level-1:	Within-person	σ_ε^2	70.20***	53.72***	41.98***	40.10***
Level-2:	In 1st grade status	σ_0^2	87.42***	123.52***	107.08***	126.09***
	Linear term					
	variance	σ_1^2		4.69**	24.60*	88.71
	covar with 1st grade status	σ_{01}		−12.54*	−3.69	−51.73
	Quadratic term					
	variance	σ_2^2			1.22*	11.35
	covar with 1st grade status	σ_{02}			−1.36	22.83~
	covar with linear term	σ_{12}			−4.96*	−31.62
	Cubic term					
	variance	σ_3^2				0.08
	covar with 1st grade status	σ_{03}				−3.06~
	covar with linear term	σ_{13}				2.85
	covar with quadratic term	σ_{23}				−0.97
Goodness-of-fit						
	Deviance statistic		2010.3	1991.8	1975.8	1967.0
	AIC		2016.3	2003.8	1995.8	1997.0
	BIC		2021.9	2015.0	2014.5	2025.1

~ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.

A: significant between- and within-child variation

B: no fixed effect of TIME but significant var comps
ΔDeviance=18.5, 3df, p<.01

C: no fixed effects of TIME & TIME^2 but significant var comps
ΔDeviance=16.0, 4df, p<.01

D: still no fixed effects for TIME terms, but now VCs are ns also
ΔDeviance=11.1, 5df, ns

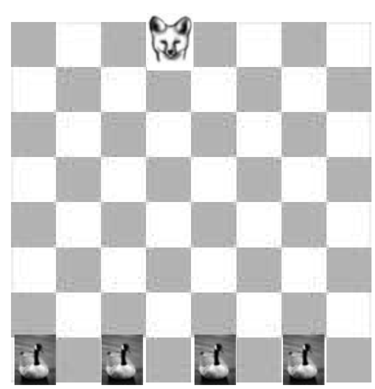
Quadratic (C) is best choice—and it turns out there are no gender differentials at all.

(ALDA, Section 6.3.3, pp 220-223)

Example for truly non-linear change

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

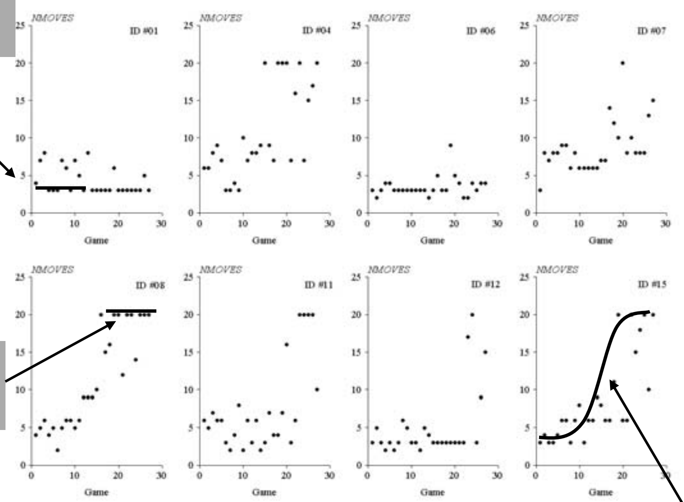
- *Sample: 17 1st and 2nd graders*
  - *During a 3 week period, Terry repeatedly played a two-person checkerboard game called Fox 'n Geese, (hopefully) learning from experience*
    - *Fox is controlled by the experimenter, at one end of the board*
    - *Children have four geese, that they use to try 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)*



(ALDA, Section 6.4.1, pp. 224-225)

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 *<u>level-1 logistic change trajectory</u>*, *which unlike our previous growth models will be* ***non-linear in the individual growth parameters***

(ALDA, Section 6.4.2, pp. 225-228)

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

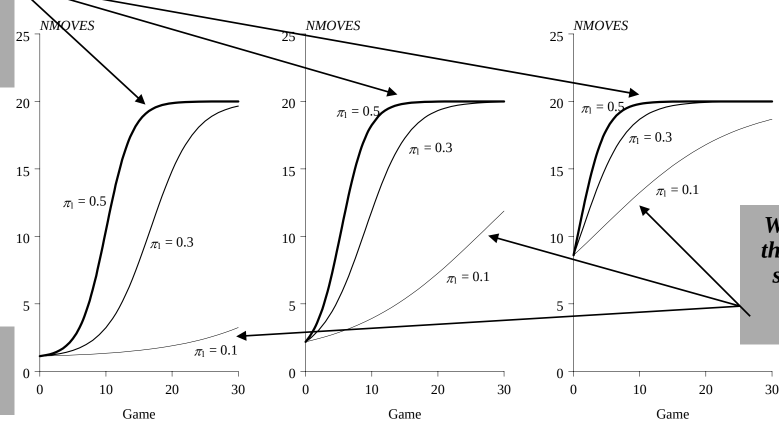
Upper asymptote in this particular model is constrained to be 20 (1+19)

π_{0i} is related to, and determines, the intercept

$$Y_{ij} = 1 + \frac{\boxed{19}}{1 + \boxed{\pi_{0i}} e^{\boxed{-\pi_{1i}} TIME_{ij}}} + \varepsilon_{ij}$$

π_{1i} determines the rapidity with which the trajectory approaches the upper asymptote

When π_{1i} is large, the trajectory rises more rapidly



When π_{1i} is small, the trajectory rises slowly *(often not reaching an asymptote)*

Higher the value of π_{0i}, the lower the intercept

π_0 = 150 π_0 = 15 π_0 = 1.5

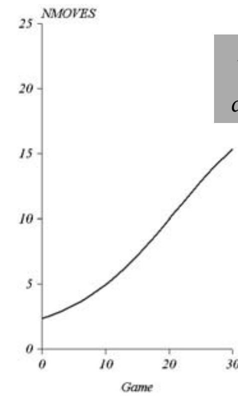
Models can be fit in usual way using provided your software can do it ⇒

(ALDA, Section 6.4.2, pp 226-230)

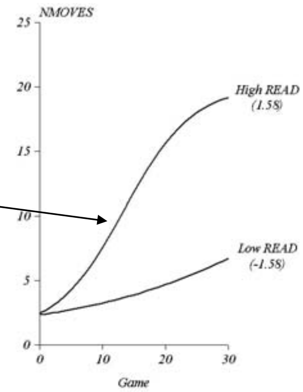
Results of fitting logistic change trajectories to the Fox 'n Geese data

		Parameter	Model A	Model B
Fixed Effects				
"Intercept," π_{0i}	Intercept	γ_{00}	12.9551***	12.8840***
	$(READ - \overline{READ})$	γ_{01}		$\boxed{-0.3745}$
"Slope," π_{0i}	Intercept	γ_{10}	0.1227***	0.1223***
	$(READ - \overline{READ})$	γ_{11}		$\boxed{0.0405}$
Variance Components				
Level-1:	Within-person	σ_ε^2	13.4005***	13.4165***
Level-2:	In "intercept"	σ_0^2	0.6761	0.5610
	In "slope"	σ_1^2	0.0072*	0.0060~
	Covariance between "intercept" and "slope"	σ_{01}	−0.0586	−0.0469
Goodness-of-fit				
	Deviance		2479.7	2477.8
	AIC		2491.7	2493.8
	BIC		2496.7	2500.5

~$p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$.



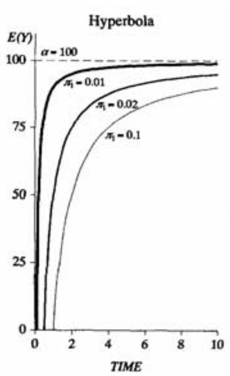
Begins low and rises smoothly and non-linearly



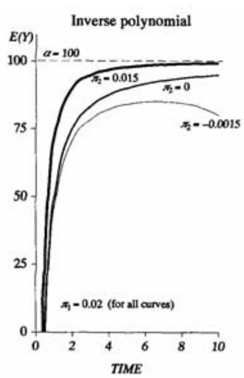
Not statistically significant (note small n's), but better READers approach asymptote more rapidly

(ALDA, Section 6.4.2, pp 229-232)

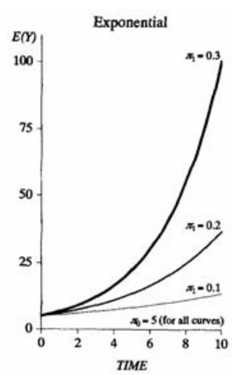
A limitless array of non-linear trajectories awaits… (each is illustrated in detail in ALDA, Section 6.4.3)



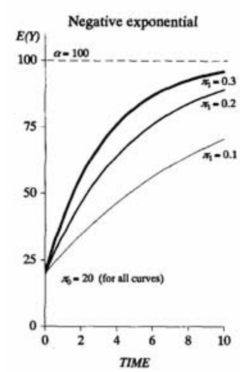
$$Y_{ij} = \alpha_i - \frac{1}{\pi_{1i} TIME_{ij}} + \varepsilon_{ij}$$



$$Y_{ij} = \alpha_i - \frac{1}{(\pi_{1i} TIME_{ij} + \pi_{2i} TIME_{ij}^2)} + \varepsilon_{ij}$$



$$Y_{ij} = \pi_{0i} e^{\pi_{1i} TIME_{ij}} + \varepsilon_{ij}$$



$$Y_{ij} = \alpha_i - (\alpha_i - \pi_{0i}) e^{-\pi_{1i} TIME_{ij}} + \varepsilon_{ij}$$

(ALDA, Section 6.4.3, pp 232-242)