Longitudinal Modeling Workshop

University of Cambridge – March 2023

Assistant Professor, Methodology and Statistics
Leiden University

@E_M_McCormick
e.m.mccormick@fsw.leidenuniv.nl
https://e-m-mccormick.github.io/

Goals and Organization of this Workshop

Goals for Our Time Together

- Theoretical Knowledge
  - Longitudinal Methods
    - Model choices, handling time, covariates, and nesting
  - Evaluating which methods are best suited to a hypothesis

- Practical
  - Translate theoretical understanding into fitting these models in R on real data (ideally your own!)
  - Interpreting the output from the models
  - Reporting these findings in publications

Organization of Our Time Together

- Monday Morning
  - Organization and Goals (You are here)
  - Introduction to Longitudinal Modeling Frameworks
  - Approaches to Incorporating Time
    - Sampling designs, time-coding effects, time constructs
- Monday Afternoon
  - Optimal Shape of Change
  - Hack-a-thon Part 1
  - Q&A

Organization of Our Time Together

- Tuesday Morning
  - Covariates and Distal Outcomes
    - Time-invariant and time-varying predictors, multivariate models
  - Nested Data

- Tuesday Afternoon
  - Q&A & Wrap-up
  - Hack-a-thon Part 2

Disclaimers

- For many of you, this will be the first time you see these models
  - You <u>are not</u> meant to be an expert by tomorrow afternoon
  - Anytime you feel "stupid", realize that these models are – quite literally – my day job, and I've been doing it a while
    - Interrupt me if things are unclear!

- I have a habit of saying "just…"
  - This is not to undersell the complexity
  - More that I think there is often a mystique to these (and other) advanced methods
    - Try to use clearer language and link to prior knowledge

Data Sources

1. Introduction to Longitudinal Modeling

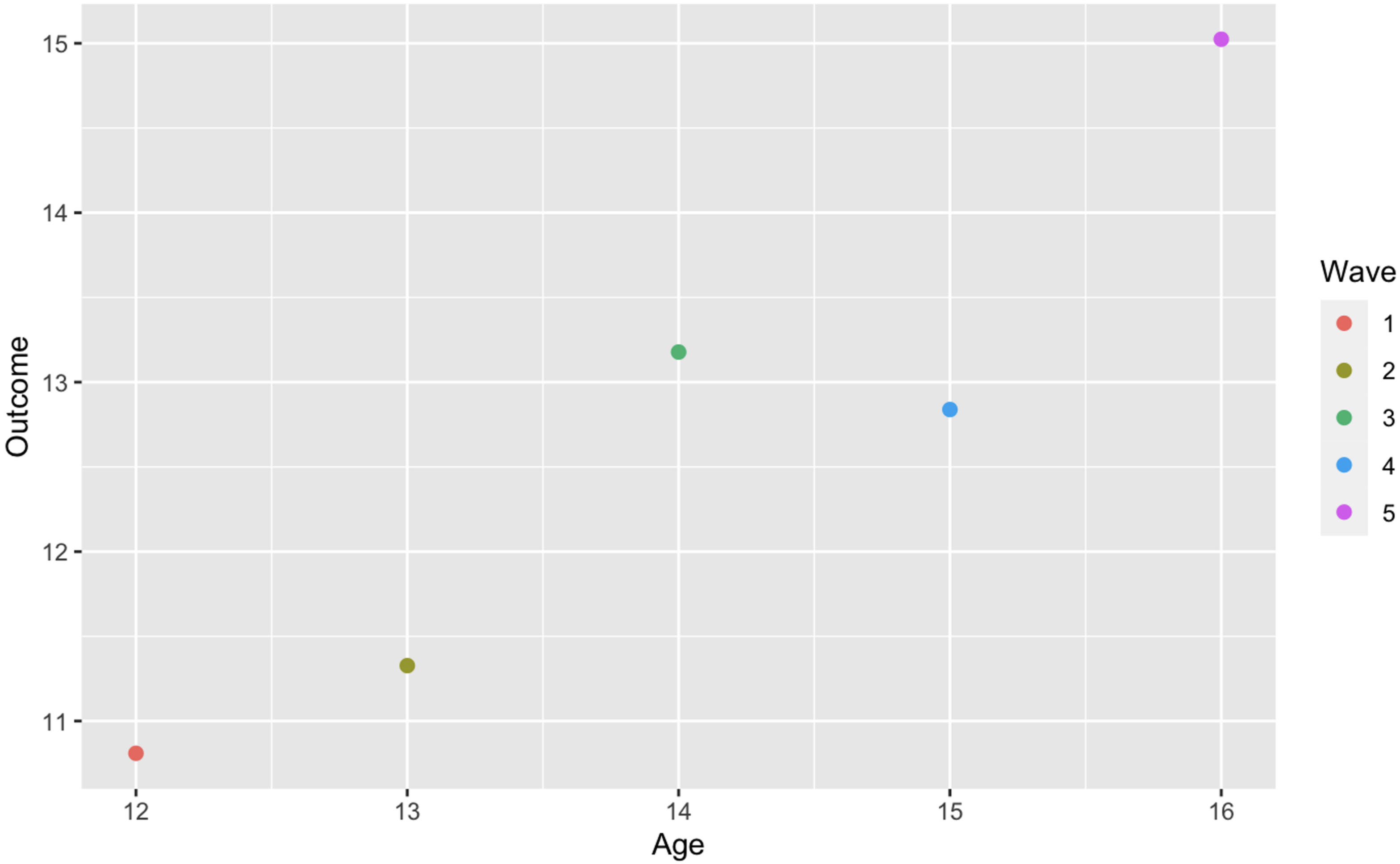
Goals of Longitudinal Modeling

A foundational goal underlying the developmental sciences is the systematic construction of a reliable and valid understanding of the course, causes, and consequences of human behavior.

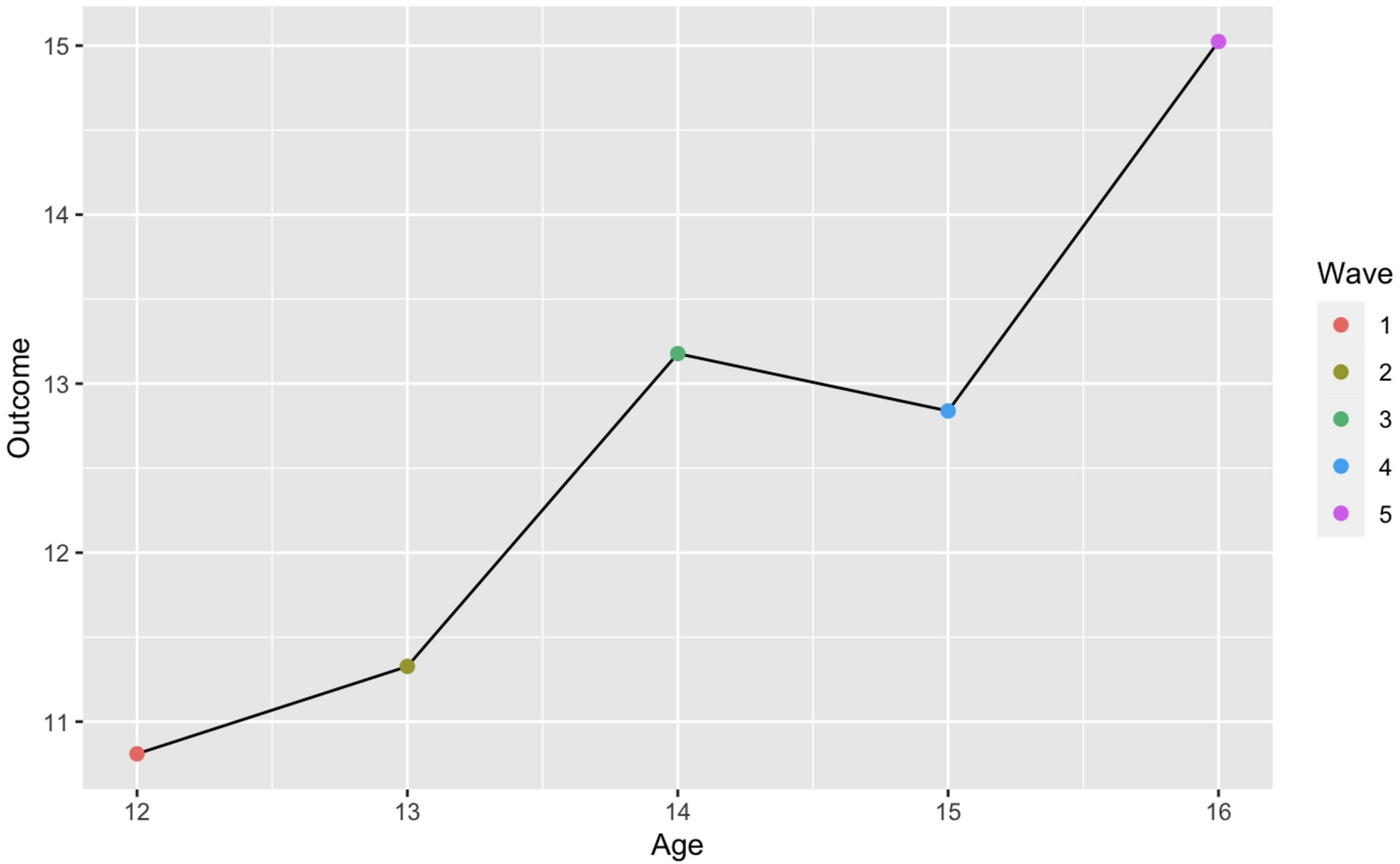
Curran, Obeidat, & Losardo, 2010

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121-136.

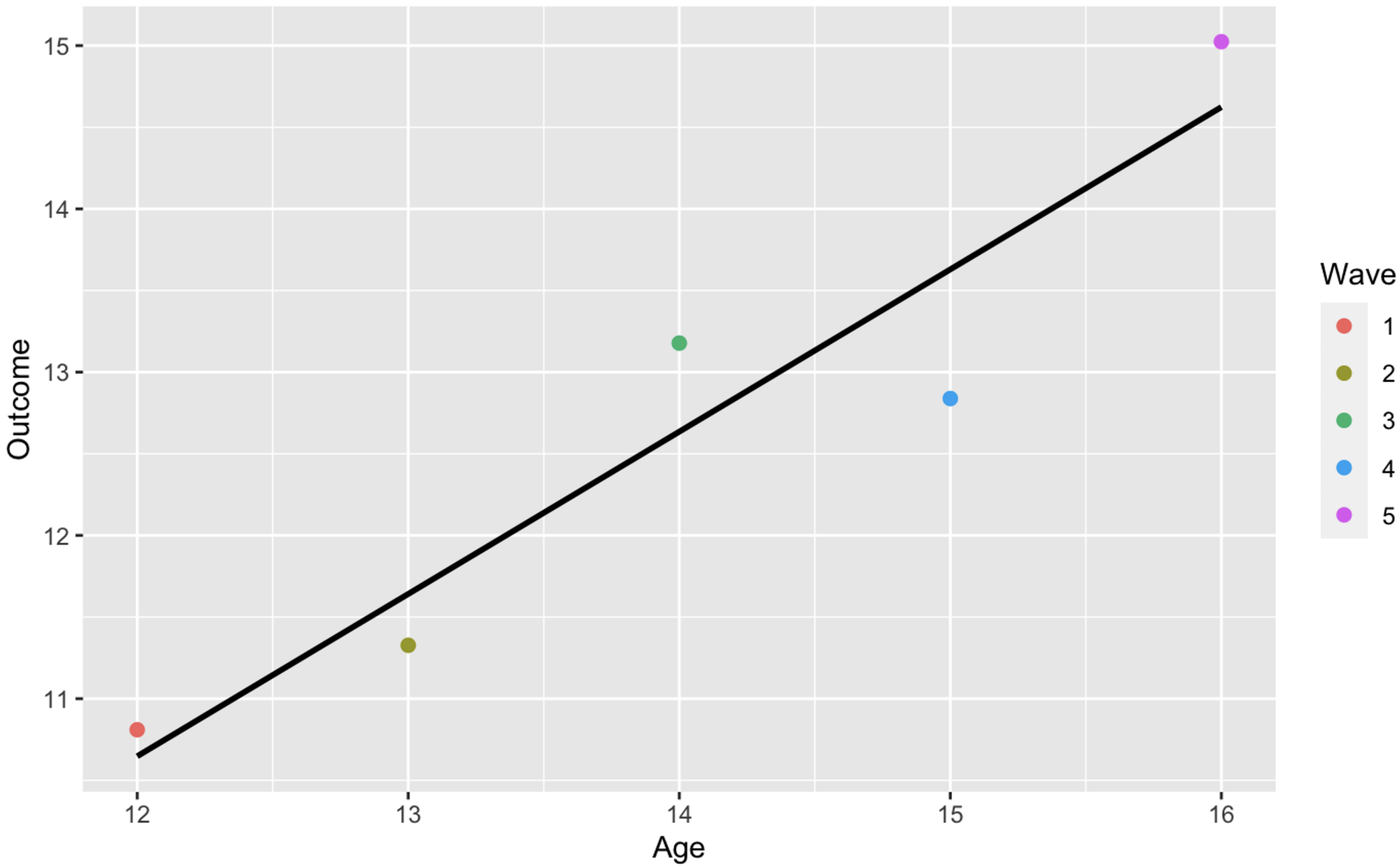
Goals of Longitudinal Modelling



Goals of Longitudinal Modelling



Goals of Longitudinal Modelling



Selecting a Longitudinal Modeling Approach

- Overwhelming number of ways to model repeated measures data
  - Often told “use theory”, but not *how* to do that

- Different frameworks*
  - Mixed-effects & structural equation models
  - A lot of historical baggage

- Need a conceptual decision tree to adjudicate between model choices

King, K. M., Littlefield, A. K., McCabe, C. J., Mills, K. L., Flournoy, J., & Chassin, L. (2018). Longitudinal modeling in developmental neuroimaging research: Common challenges, and solutions from developmental psychology. Developmental Cognitive Neuroscience, 33, 54-72.

The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods

Ethan M. McCormick*[1,2,3], Michelle L. Byrne[4,5], John C. Flournoy[6], Kathryn L. Mills[5], and Jennifer H. Pfeifer[5]

[1]Methodology & Statistics Department, Institute of Psychology, Leiden University, Leiden, Netherlands
[2]Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill, United States
[3]Cognitive Neuroscience Department, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, Netherlands
[4]Turner Institute for Brain and Mental Health, School of Psychological Sciences, Monash University, Clayton, Australia
[5]Department of Psychology, University of Oregon, Eugene, United States
[6]Department of Psychology, Harvard University, Cambridge, United States

January 25, 2023

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Longitudinal Primer + Codebook

The Hitchhiker's Guide to Longitudinal Models

Code Companion

Ethan M. McCormick

Published: 13 February, 2023

About

The following document is a code companion to The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods, https://osf.io/bn6yu/.

Some general notes about this code companion:

- We believe in the importance of using real data in our examples of longitudinal models. However, some of the models we discuss can not yet be fit using publicly-available neuroimaging data (most often due to a limited number of observations). To bridge this gap, we have synthesized data from a number of sources, detailed in Datasets. Variable names and identification codes have been changed to protect the innocent and to provide examples that will be familiar to developmental cognitive neuroscience researchers. However, one limitation of synthesized data is that model fits are often significantly worsened compared to the real data. As such, for pedagogical purposes, we will fit (and sometimes interpret) results from models that we would usually reject in practice based

Longitudinal Modeling in Pre-collected Data

- Matching models to theory is often complicated by the need to use pre-existing data sources
  - Other people have made decisions for you

- But with more datasets becoming publicly available, there may be options in the future

Using large, publicly available data sets to study adolescent development: opportunities and challenges
Rogier A. Kievit[2,3], Ethan M. McCormick[2,3,a], Delia Fuhrmann[1,4,a], Marie K. Deserno[5,6,a] and Amy Orben[1,a]

Kievit, R. A., McCormick, E. M., Fuhrmann, D., Deserno, M. K., & Orben, A. (2022). Using large, publicly available data sets to study adolescent development: opportunities and challenges. *Current Opinion in Psychology, 44*, 303-308.

Model Notation (a.k.a. the language of longitudinal modeling)

Terms and Notation Cheat Sheet

1 Matrix Conventions

Matrix expressions are a powerful and compact way to arrange information with great computational advantages. However, many researchers are familiar with simple equations for only a single outcome that can be written on a single line – we will call these scalar expressions – and matrix expressions are often foreign. Here we attempt to clarify the notational conventions for matrix expressions. We will highlight these conventions using a set of simple ordinary least squares (OLS) regression equations.

The most familiar form of an equation is in *scalar* form, where each term either refers to a single estimated parameter, or to a single observed variable in our data frame. Scalar here just refers to individual numerical values. For instance, in a single-predictor regression, we have the following:

estimated parameters residual

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

observed variables

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Mixed Effects Models

Modeling Frameworks for Longitudinal Data

- Mixed-effects models (MEMs)
  - Multilevel and generalized additive models
  - Originated in hierarchically-structured data
    - E.g., kids in classrooms
  - Classically univariate models

- Structural Equation models (SEMs)
  - Latent Curve and Latent Change Score models
  - Originated from psychometric models
    - E.g., scale scores, latent variables
  - Classically multi-variate models

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Multilevel Models (MLMs)

- Originated in the field of education
  - Children within a given classroom are more similar to one another than we would expect by chance
  - This logic extends naturally into the longitudinal context
    - Repeated measures from a single individual are not independent over time

- Multi-levels
  - Level 1: describes the within-unit relationships
  - Level 2: describes the between-unit relationships
    - “average” differences of within-unit observations between units

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101*(1), 147-158.

Multilevel Models

- In longitudinal models:
  - Level 1: describes the within-person relationships
    - Predictors explain occasion-to-occasion variation for individuals with respect to their own level
  - Level 2: describes the between-person relationships
    - Average levels of the predictors explain average levels of the outcome for each individual

- Terminology: “repeated measures nested within individuals”
  - Level 1 nested within Level 2

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker’s Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Multilevel Models

- Notation

Level 1:

$$\overbrace{y_{ti}}^{\text{repeated-measures outcome}} = \underbrace{\beta_{0i}}_{\text{intercept (where all predictors are 0)}} + \overbrace{\beta_{1i}}^{\text{effect of x}} \underbrace{x_{ti}}_{\text{observed measure of time or growth}} + \overbrace{r_{ti}}^{\text{time-specific residual}}$$

Level 2:

$$\overbrace{\beta_{0i}}^{\text{from Equation 1}} = \overbrace{\gamma_{00}}^{\text{fixed effects}} + u_{0i}$$

$$\beta_{1i} = \gamma_{10} + \underbrace{u_{1i}}_{\text{random effects}}$$

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Multilevel Models

- Notation

Level 1: Within-person equation

repeated-measures outcome　　effect of x　　time-specific residual

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

intercept (where all predictors are 0)　　observed measure of time or growth

Level 2: Between-person equation

from Equation 1　　fixed effects

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

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

random effects

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

- Notation

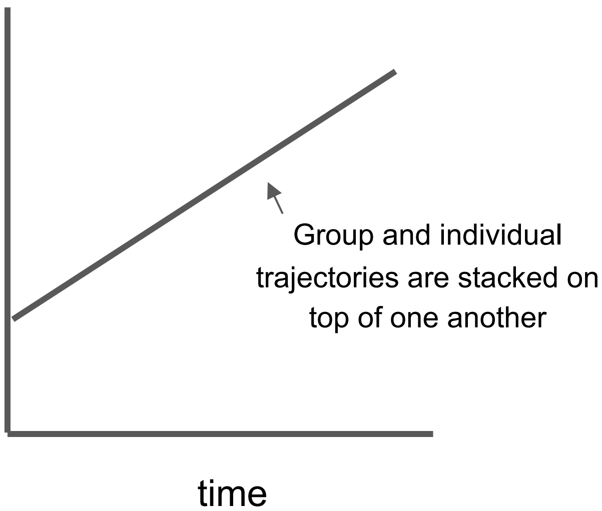
Reduced Form Equation

$$y_{ti} = \underbrace{\gamma_{0i} + \gamma_{1i}x_{ti}}_{\text{fixed effects}} + \underbrace{u_{0i} + u_{1i}x_{ti}}_{\text{random effects}} + r_{ti}$$

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

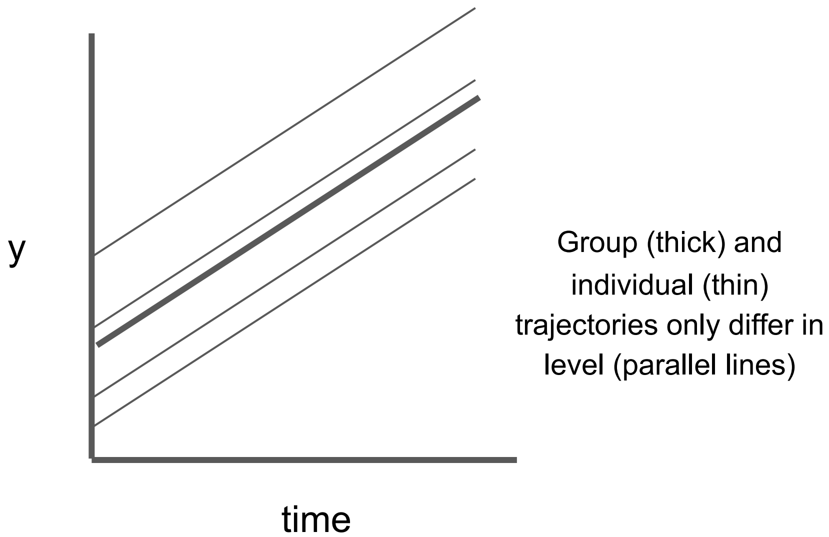
Fixed Versus Random Effects

Fixed Intercept, Fixed Slope



You would never do this, but it's basically OLS regression

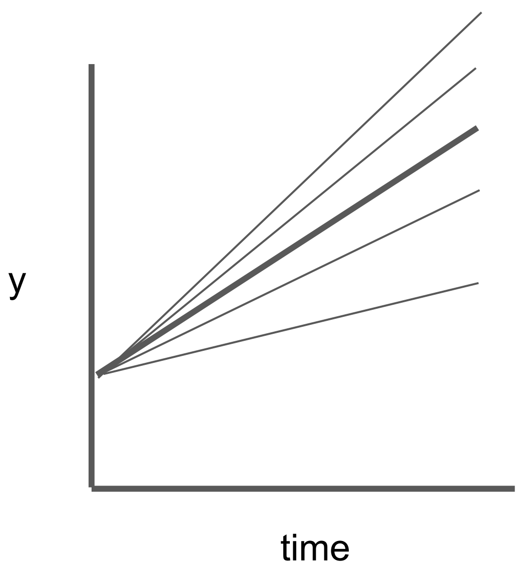
Random Intercept, Fixed Slope



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Fixed Versus Random Effects

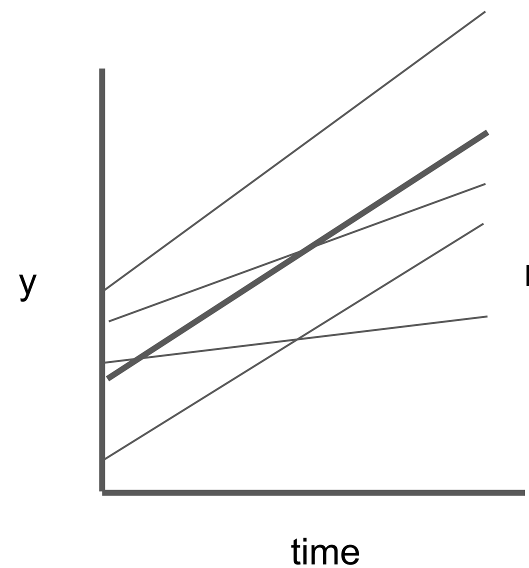
Fixed Intercept, Random Slope



Non-parallel lines radiate from a common intercept

Relatively uncommon but sometimes makes sense

Random Intercept, Random Slope



Individuals differ both in intercept and rate of change over time

Notice that rank order is NOT preserved across time

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Multilevel Models

- Covariance Matrix

Normal (Gaussian) distribution

variances of the random effects

$$\begin{bmatrix} u_0 \\ u_1 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

the means are assumed to be zero

covariance between the random effects

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Generalized Additive Mixed Models (GAMMs)

- All the effects in the MLM are linear
  - Effects enter the equation additively

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

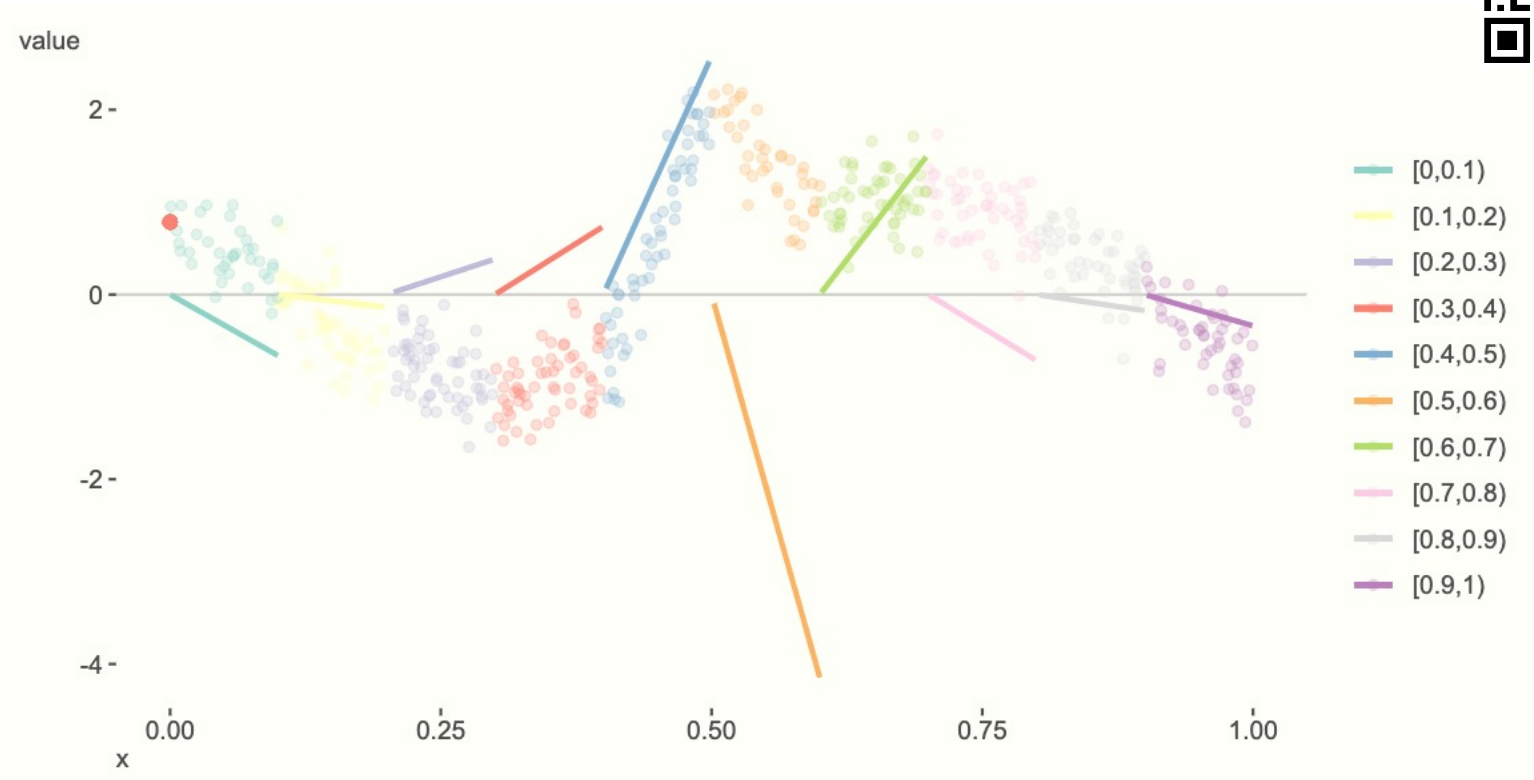
- If we want to include non-linear terms, we can wrap the predictor in a general function

$$y_{ti} = \gamma_{00} + f(x_{ti}) + u_{0i} + r_{ti}$$

  - Many approaches for estimating this general function (e.g., splines, basis functions)

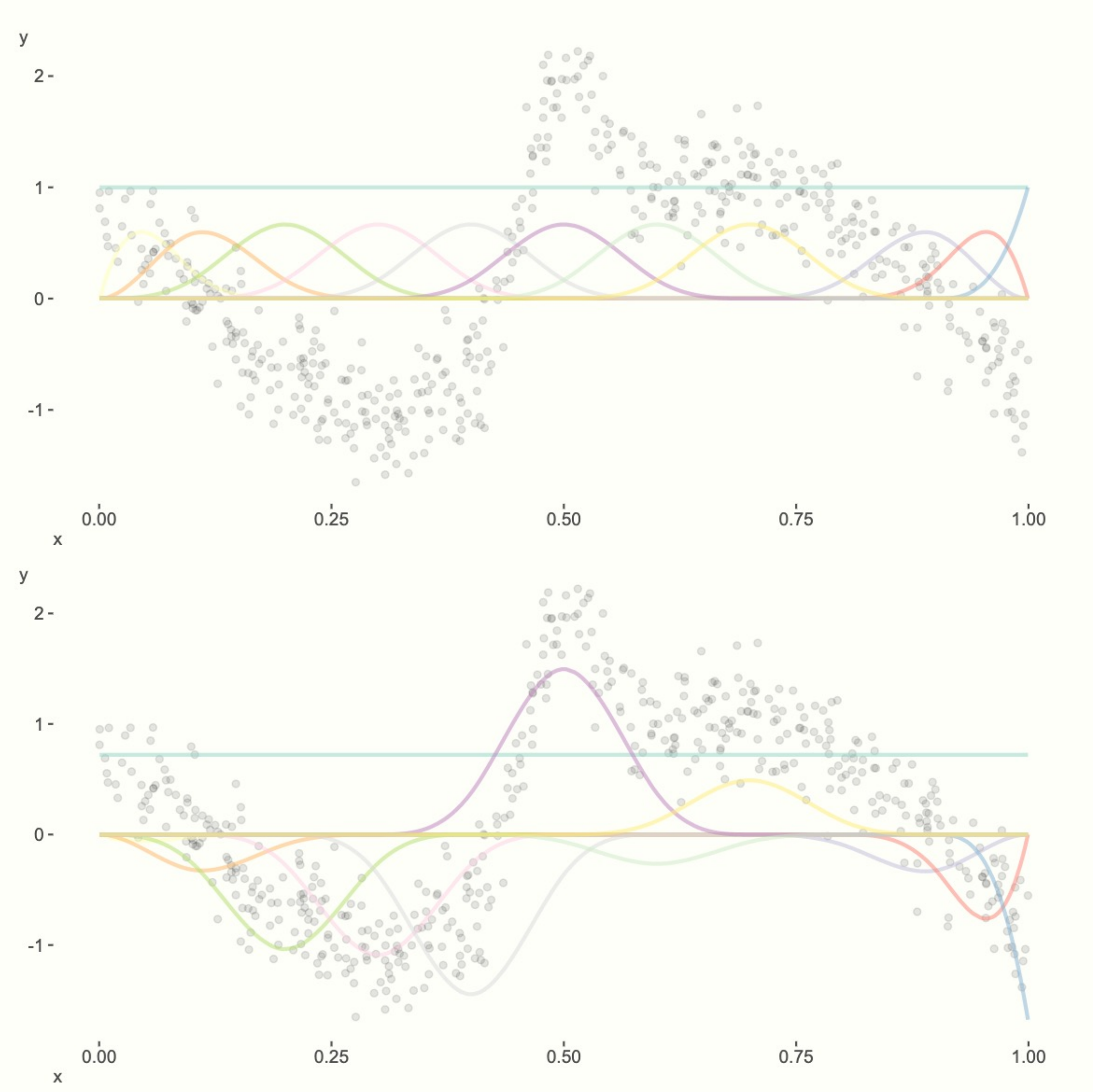
Lin, X., & Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61*(2), 381-400.

Generalized Additive Mixed Models



https://m-clark.github.io/generalized-additive-models/technical.html#a-detailed-example

Generalized Additive Mixed Models



https://m-clark.github.io/generalized-additive-models/technical.html#a-detailed-example

Generalized Additive Mixed Models

- ”Wiggliness” of the function characterizes how flexibly it fits the data

- Preventing overfitting
  - The more basis functions we use, the better we can approximate the curves
  - *but* can start of to overfit noise

- Deal with this typically through penalization

Lin, X., & Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61*(2), 381-400.

Cheat Sheet for MEMs

3 Common Symbols/Notations and their Definitions

3.1 Mixed-Effects Models (MEM) Terminology & Notation

Many things are common across both multilevel and generalized additive mixed models. Hence, we will cover most notations under the banner of the MLM (because that is how it is encountered in the manuscript) and then note the unique additional notations for the GAMM.

3.1.1 Commonly-Encountered Terms

Level 1 Model equations that structure the within-unit data and effects. In longitudinal models, this level typically models the repeated measures within individuals (i.e., time-varying).

Level 2 Model equations that structure the between-unit data and effects by writing additional equations for Level 1 parameters. In longitudinal models, this level typically models the average differences between individuals (i.e., time-invariant).

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Structural Equation Models

Latent Curve Models (LCMs)

- Originated in the field of psychometrics
  - A very specialized form of confirmatory factor analysis (CFA)
  - By fixing the factor loadings to pre-determined values, we can represent the intercept and slope (or other features) as latent variables

- Single-level, multivariate model
  - Formally equivalent to the MLM in many cases
  - Factor variance-covariance matrix (Psi) is analogous to the Tau matrix in MLMs

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107-122.

Latent Curve Models

- Notation and Path Model

$$\overbrace{\mathbf{y}_{ti}}^{\text{observed repeated measures}} = \underbrace{\mathbf{\Lambda}}_{\text{factor loading matrix (contains values of time)}} \overbrace{\boldsymbol{\eta}_i}^{\text{latent growth factor(s)}} + \underbrace{\boldsymbol{\varepsilon}_{ti}}_{\text{time-specific residuals}}$$

$$\overbrace{\boldsymbol{\eta}_i}^{\text{from Equation 8}} = \overbrace{\boldsymbol{\alpha}}^{\text{factor means (fixed effects)}} + \underbrace{\boldsymbol{\zeta}_i}_{\text{factor variances (random effects)}}$$

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Cheat Sheet for Matrix Expressions

$$
\begin{aligned}
\mathbf{y}_i &= \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \\
\boldsymbol{\eta}_i &= \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i
\end{aligned} \tag{5}
$$

We can go through and recognize which terms are **vectors** and **MATRICES** based on the formatting conventions we have outlined previously.

$$
\overset{\textbf{vector}\text{ of observed outcomes}}{\mathbf{y}_i} = \underset{\textbf{MATRIX}\text{ of factor loadings}}{\boldsymbol{\Lambda}}\ \overset{\textbf{vector}\text{ of latent factors}}{\boldsymbol{\eta}_i} + \underset{\textbf{vector}\text{ of residuals}}{\boldsymbol{\varepsilon}_i}
$$

$$
\boldsymbol{\eta}_i = \overset{\textbf{vector}\text{ of factor means}}{\boldsymbol{\alpha}} + \underset{\textbf{MATRIX}\text{ of regression coefficients of } \boldsymbol{\eta}_i \text{ on } \mathbf{x}_i}{\boldsymbol{\Gamma}}\ \overset{\textbf{vector}\text{ of observed covariates}}{\mathbf{x}_i} + \underset{\textbf{vector}\text{ of disturbances}}{\boldsymbol{\zeta}_i} \tag{6}
$$

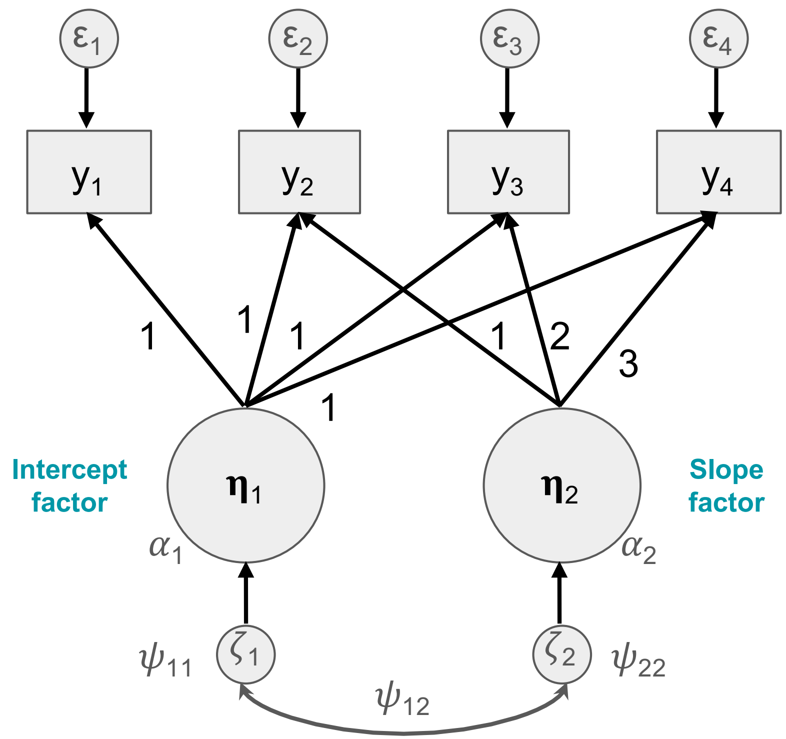
We can finally expand each of these vectors and matrices to show how each *scalar* value would be arranged, resulting in the following:

$$
\begin{bmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ y_{4i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \end{bmatrix} \tag{7}
$$

$$
\begin{bmatrix} \eta_{1i} \\ \eta_{2i} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} \begin{bmatrix} x_{1i} \end{bmatrix} + \begin{bmatrix} \zeta_{1i} \\ \zeta_{2i} \end{bmatrix}
$$

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

- Notation and Path Model



Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55*(1), 107-122.

- Covariance Matrix

$$\underbrace{\boldsymbol{\Psi}}_{\text{variance-covariance matrix}} = \begin{bmatrix} \psi_{11} & \\ \psi_{21} & \psi_{22} \end{bmatrix}$$

variances of the latent factors: ψ_{11}, ψ_{22}

factor covariance: ψ_{21}

MLM Tau Matrix:

$$\begin{bmatrix} u_0 \\ u_1 \end{bmatrix} \sim \underbrace{N}_{\text{Normal (Gaussian) distribution}} \left(\underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\text{the means are assumed to be zero}}, \begin{bmatrix} \tau_{00} & \\ \tau_{10} & \tau_{11} \end{bmatrix} \right)$$

variances of the random effects: τ_{00}, τ_{11}

covariance between the random effects: τ_{10}

Curran, P. J. (2003). Have multilevel models been structural equation models all along?. *Multivariate Behavioral Research, 38*(4), 529-569.

Latent Change Score Models (LCSMs)

- We can think of a trajectory described by the LCM as a continuous, smooth path
  - But we might also want to model it as a set of discrete steps that sum into that trajectory
  - This is how the LCS models change over time

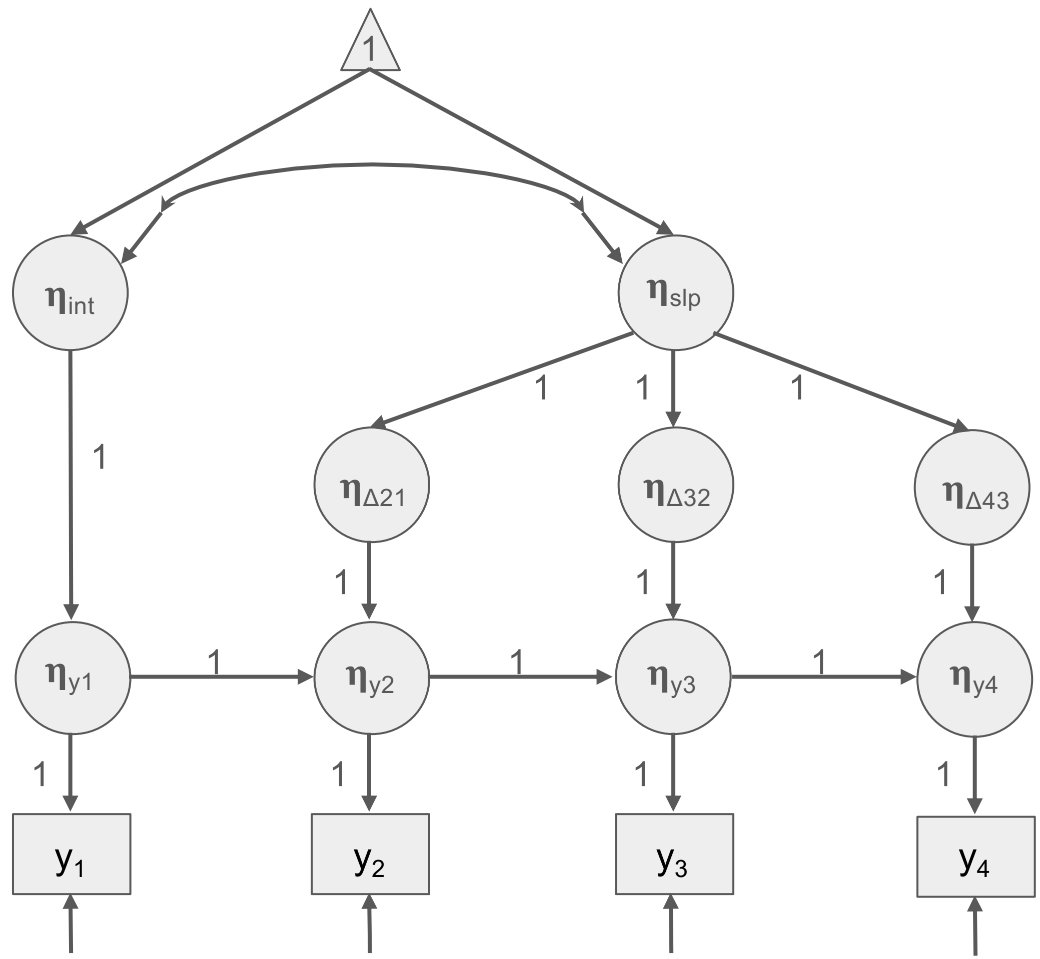
- Formally equivalent to the LCM in many applications, but with some additional options

Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B., & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural Equation Modeling: A Multidisciplinary Journal, 19*, 268-292.

Latent Change Score Models

- Notation and Path Model

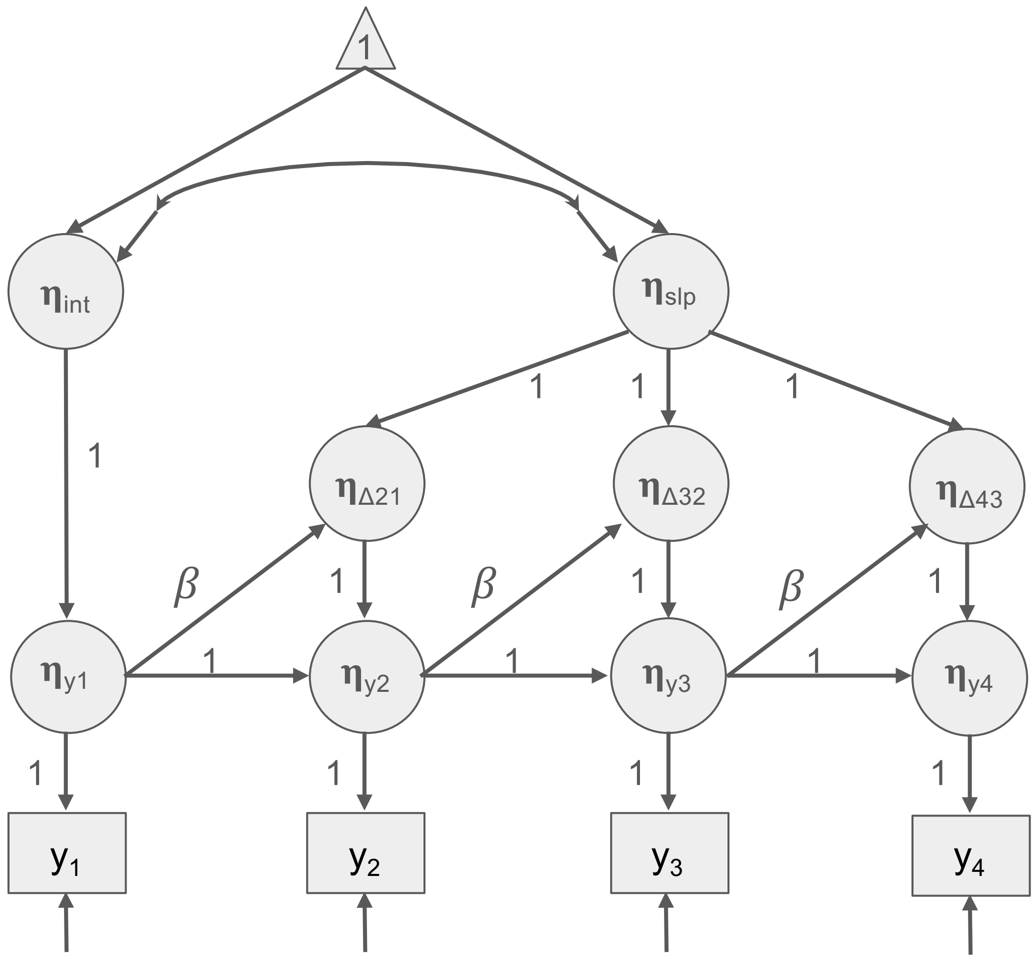
$$\Delta y_{t,t-1} = y_t - y_{t-1}$$



Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B., & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural Equation Modeling: A Multidisciplinary Journal, 19*, 268-292.

- Notation and Path Model

$$\Delta y_{t,t-1} = y_t - y_{t-1}$$



Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B., & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural Equation Modeling: A Multidisciplinary Journal, 19*, 268-292.

Cheat Sheet for SEMs

3.2 Structrual Equation Models (SEM) Terminology & Notation

Like with the MEM notation, many things are common across both latent curve and latent change score models. Hence, we will cover most notations under the banner of the LCM (because that is how it is encountered in the manuscript) and then note the unquie additional notations for the LCSM.

T	Test statistic comparing a specified model to a baseline comparison model.
ν	An intercept (nu) of an observed outcome. These are analogous to the β_0 intercept mixed-effect growth model (or any standard regression equation).
$\boldsymbol{\nu}$	A vector of item intercepts, where the length of the vector is equal to the number of observed items.
λ	A factor loading (lambda). These typically* represent the regression of an observed outcome on a latent factor. They are interpreted just like regression coefficients: "A 1-unit change in the latent factor is associated with a λ-unit change in the outcome."
$\boldsymbol{\Lambda}$	A matrix of factor loadings, where the number of rows is equal to the number of observed items, and the number of columns is equal to the number of latent factors.
η	A latent factor/variable (eta). These are variables for which we do not have observed values in our data, but rather we infer their existence from the pattern of covariance between observed items.

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Comparing Modeling Frameworks

- MEM vs. SEM
  - MEM:
    - Strengths: Continuous & individually-varying time, multiple covariates, REML for small samples
    - Weaknesses: Just-identified model, really oriented towards univariate models, no missing data
  - SEM:
    - Strengths: Metrics of absolute model fit, ideally suited for multivariate models, missing data allowed
    - Weaknesses: Needs discrete forms time on some level, larger numbers of covariates challenging

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Comparing Modeling Frameworks

- MLM vs. GAMM
  - MLM is probably the default unless you really need that non-linear basis function
  - MLM more suited to random effects – splines tend to be fixed effects only (in classic longitudinal data)

- LCM vs. LCSM:
  - Because the LCSM is often equivalent to the LCM, the LCM is usually the recommended default
  - However, proportional change is only available in the LCSM
    - Really useful for certain forms of exponential change

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Canonical Models

What follows are canonical versions of growth models in each of the four different frameworks. These models represent basic implementations of a linear growth trajectory with random effects for both the intercept and slope, with the exception of the GAMM, where a non-linear spline model is implemented (otherwise it would just re-capitulate the MLM results). This will be the longest chapter of the codebook since we will cover syntax and model output more in-depth than in later chapters. Remain calm and clutch your towel as necessary.

First, we need to read in the datasets we will use in this chapter.

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Questions?

@E_M_McCormick
e.m.mccormick@fsw.leidenuniv.nl
https://e-m-mccormick.github.io/

2. Approaches to Incorporating Time

Thinking about Time

- Time* is the foundation of all longitudinal models

- Time is a special kind of predictor
  - Monotonic and changes consistently across individuals
  - Terminology can be confusing
    - We call a model "unconditional" when it only includes time

- How we include time in the model is both dependent on the modeling framework of choice and our goals

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

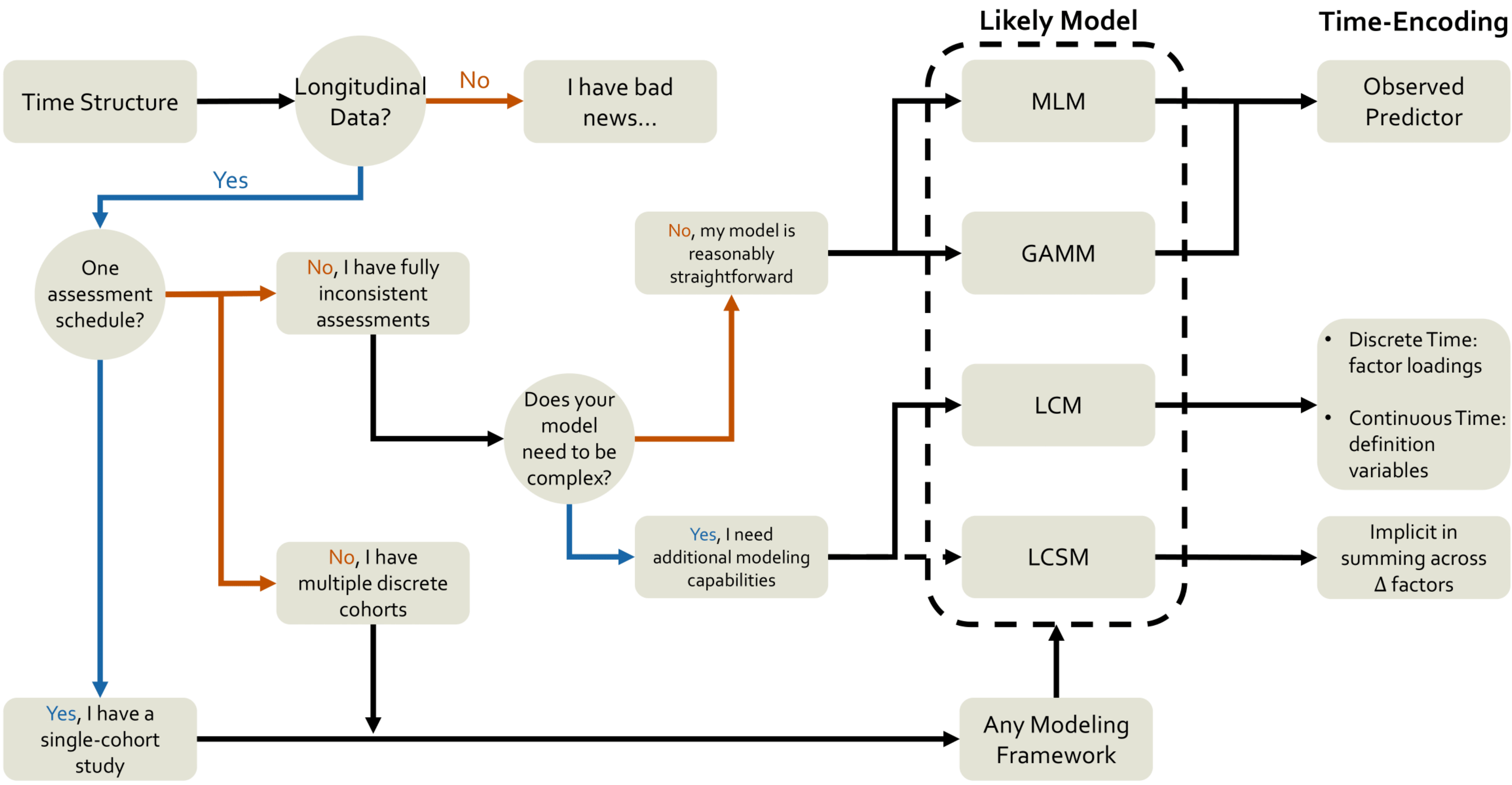
Thinking about Time

- We will talk about time in several ways:

1. Time in study design
   a. Assessment schedules
   b. Missing data
2. Time in the model
   a. Observed predictor versus model parameter
   b. Time coding
3. Alternative time predictors
   a) Multiple forms of time
   b) Time-dependent residuals

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Thinking about Time



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Time in Study Design

Time in Study Design

- There are several possibilities for structuring time within our study
  - Selecting developmental window to study
  - Spacing and duration of assessments
  - Structure of assessment across individuals

- **How you set up your study determines all subsequent modeling decisions**

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Time in Study Design

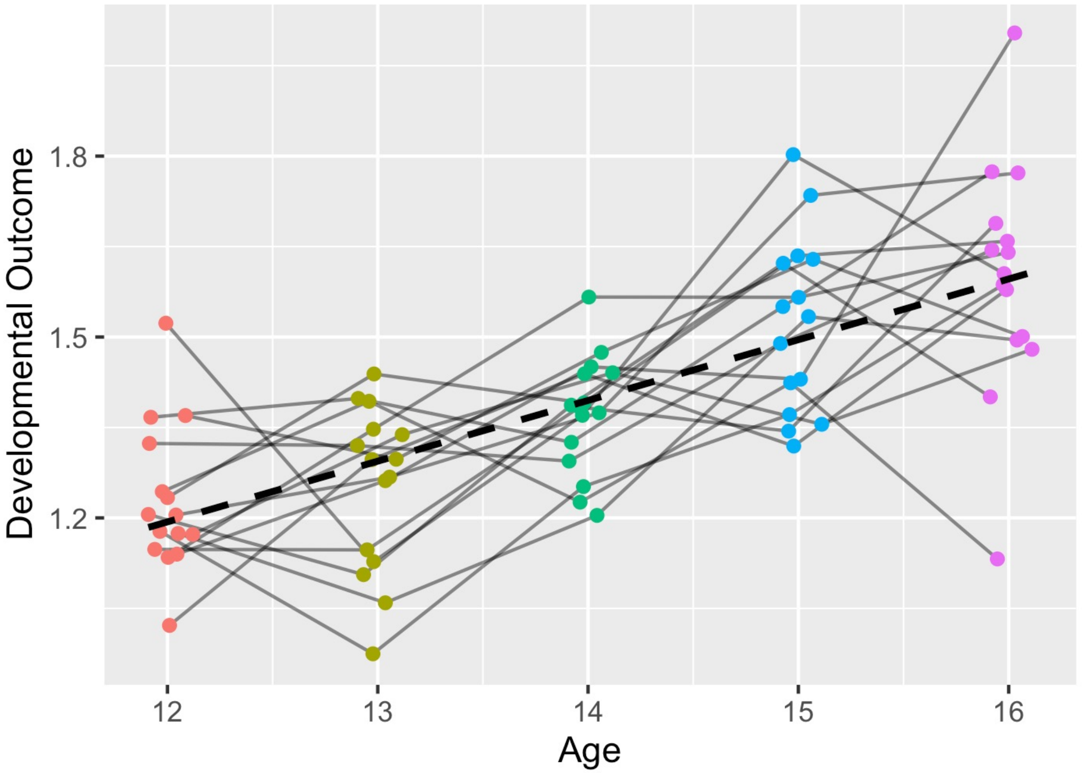
- Assessment Schedules
  - A combination of how often and how consistently individuals are assessed within a longitudinal study

- 3 broad cases (with some fuzziness)
  - Single-Cohort
  - Multiple-Cohort (or cohort-sequential)
  - Accelerated

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Single-Cohort Design

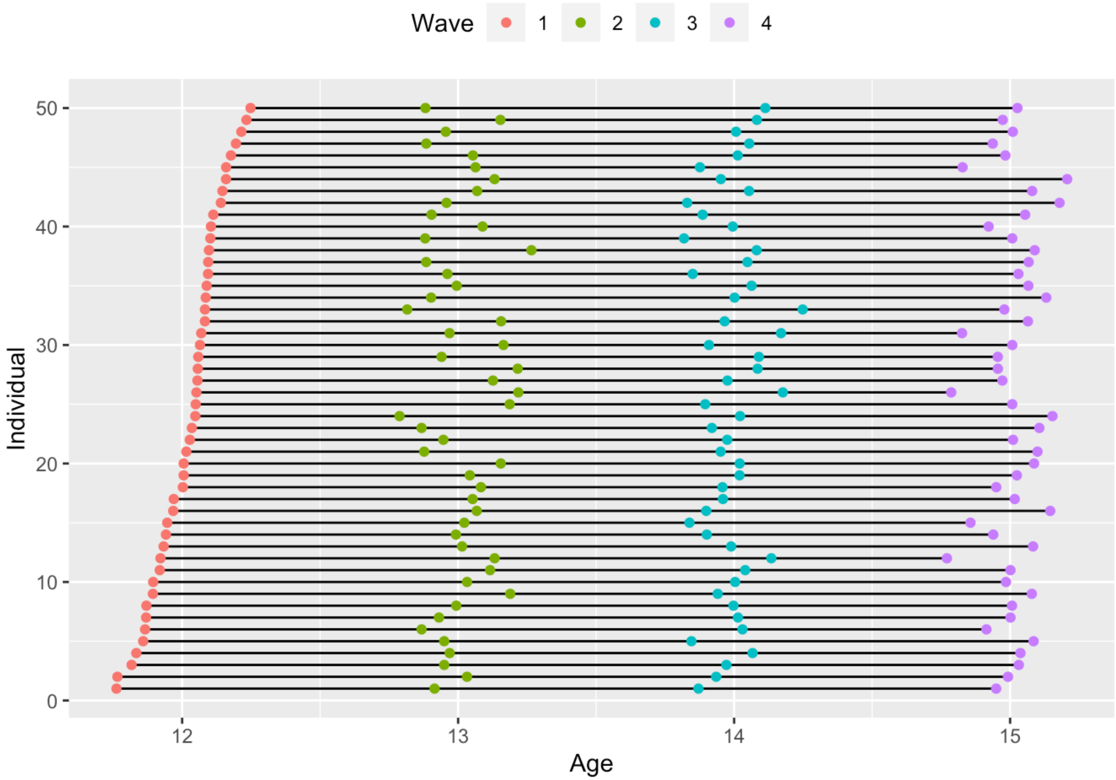
- This is the classic longitudinal study
  - All individuals are observed at the same defined ages



McCormick, E. M. (2021). Multi-Level Multi-Growth Models: New opportunities for addressing developmental theory using advanced longitudinal designs with planned missingness. *Developmental Cognitive Neuroscience*, *51*, 101001.

Single-Cohort Design

- This is the classic longitudinal study
  - All individuals are observed at the same defined ages



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

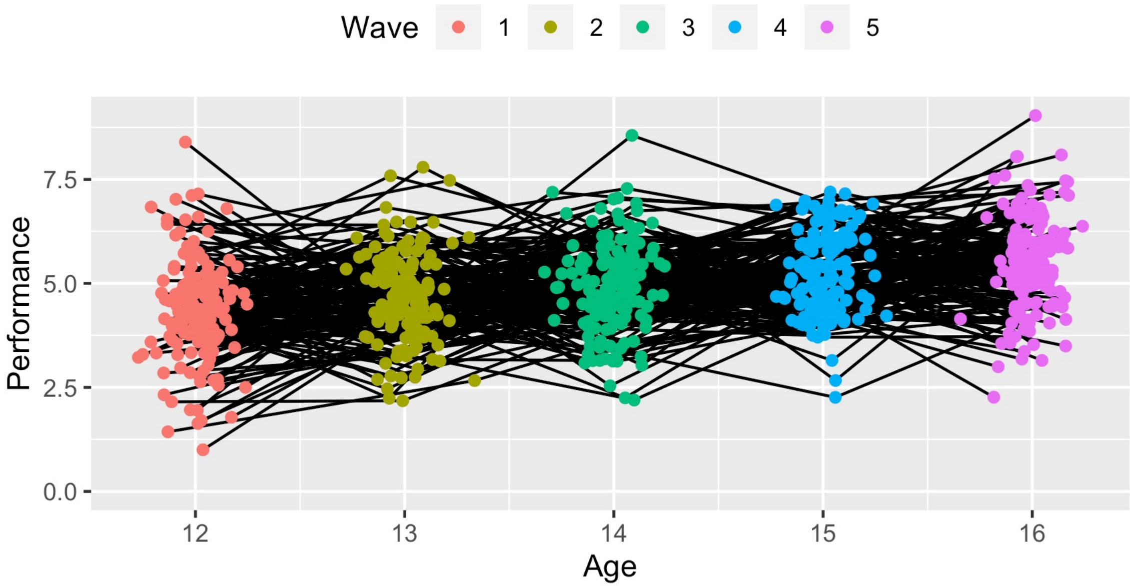
Single-Cohort Design

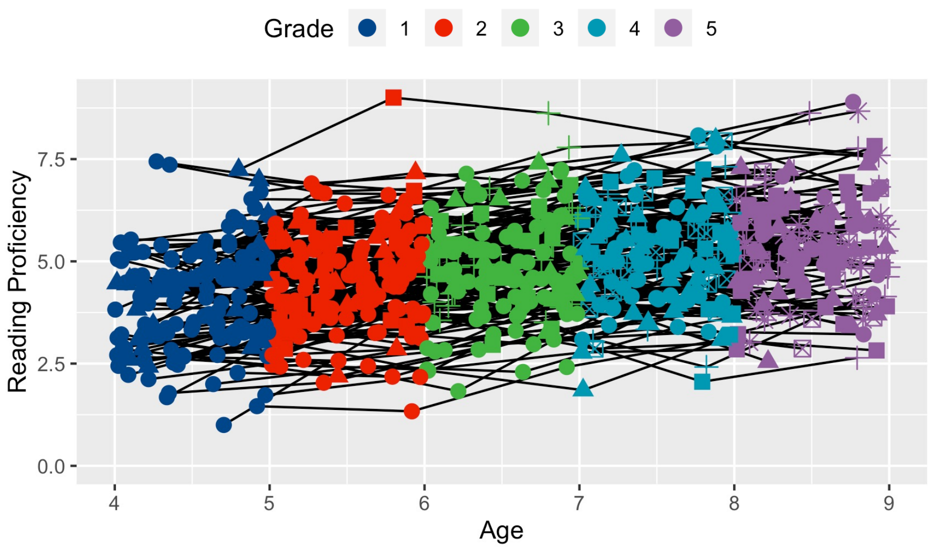
- Advantages
  - Relatively simple to implement study design
  - High degree of power at each observed age
  - Straightforward ways to estimate individual differences

- Disadvantages
  - Relatively short developmental windows
  - How much within-observation jitter is acceptable?
    - Compression of heterogeneity

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Single-Cohort Design

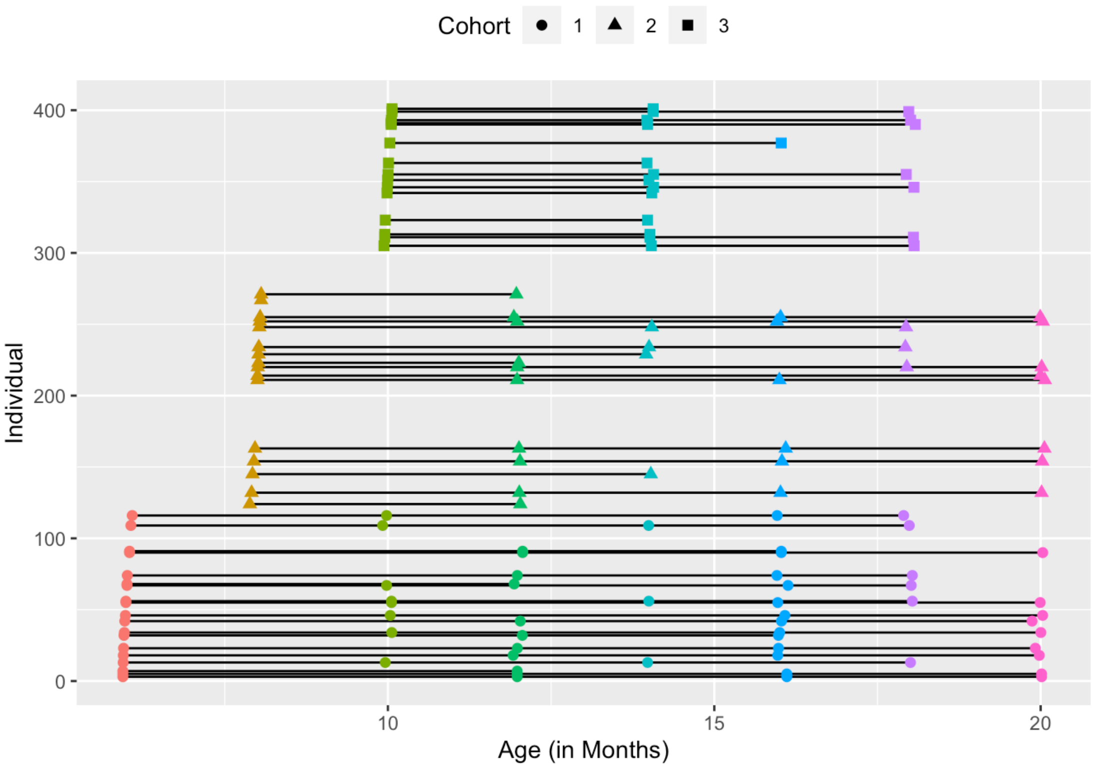




McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Multiple-Cohort Design

- Multiple-cohort designs essentially duplicate a single-cohort design, but separate groups of individuals start at different ages



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Multiple-Cohort Design

- Multiple-cohort designs essentially duplicate a single-cohort design, but separate groups of individuals start at different ages

- Two common approaches
  - Same starting age, but in different years
    - E.g., follow ages 12-15, starting in 1990, 1992, & 1993
    - Common in educational settings
  - Starting age for one cohort is the ending age for another
    - E.g., one cohort is followed 10 – 15, second cohort is followed 15 - 20
    - Maximizes developmental coverage

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Multiple-Cohort Design

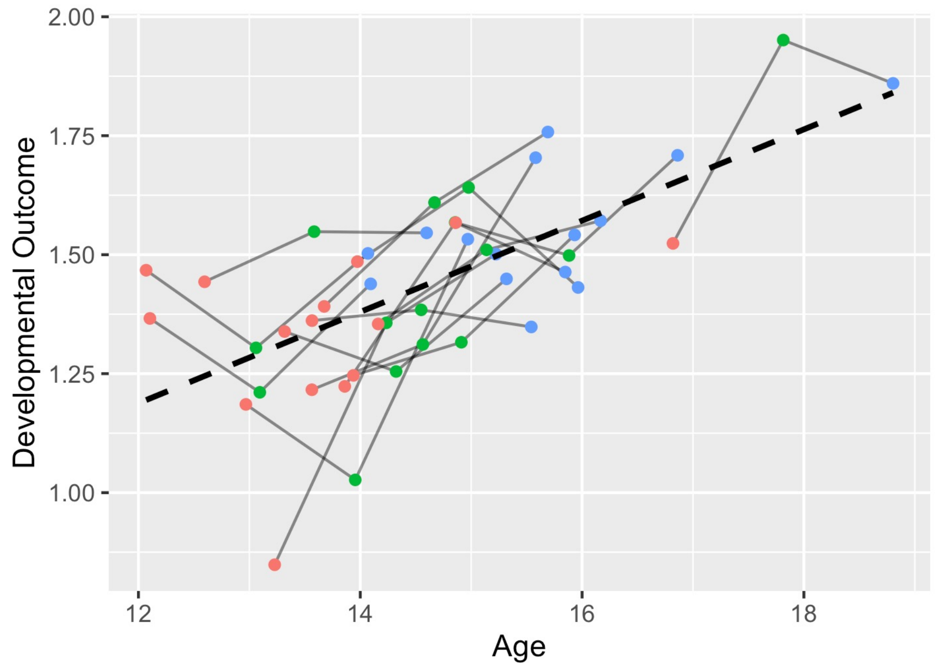
- Advantages
  - No single individual needs to be assessed across the entire developmental window
  - (still) Relatively high power at each observed age
  - Missing data (at least the planned kind) does not bias model parameters – it is Missing Completely At Random

- Disadvantages
  - We need to assume or establish cohort exchangeability
  - Less power than a single-cohort design
  - Might not be able to capture slow individual differences

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Accelerated Design

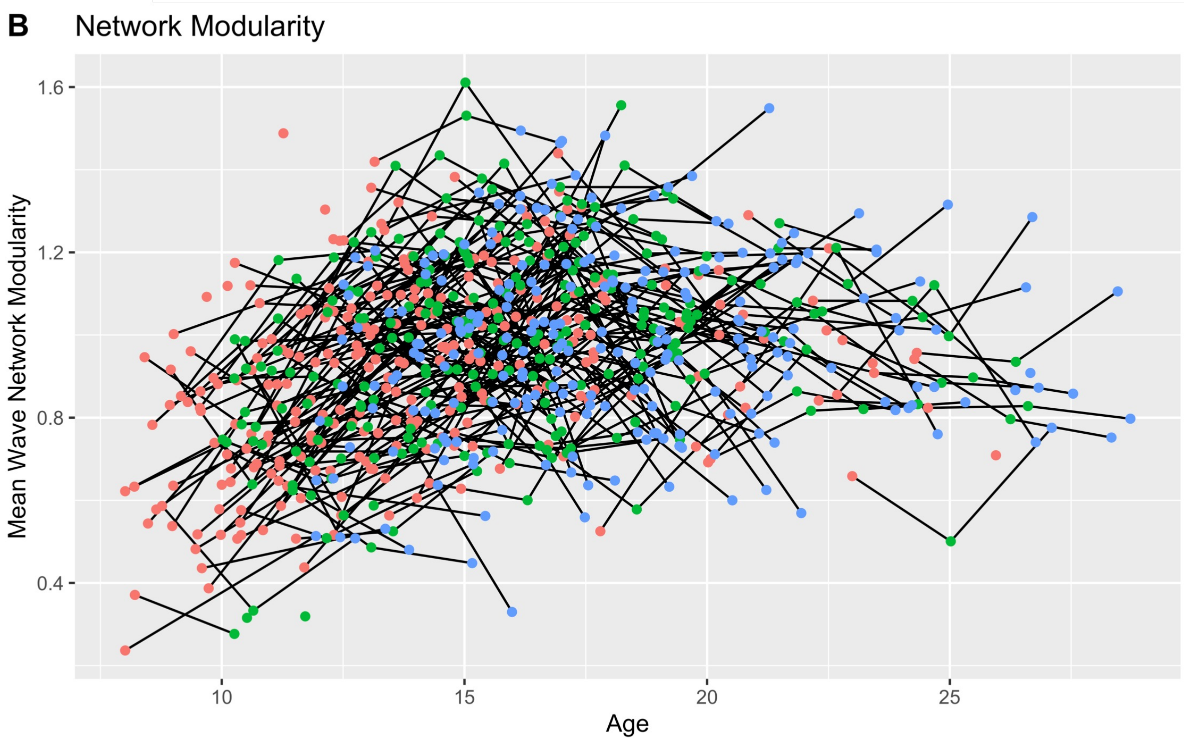
- Individualized assessment schedules
  - No two individuals need be assessed at the same ages, lag, or for the same number of observations



McCormick, E. M. (2021). Multi-Level Multi-Growth Models: New opportunities for addressing developmental theory using advanced longitudinal designs with planned missingness. *Developmental Cognitive Neuroscience*, *51*, 101001.

Accelerated Design

- Advantages
  - Much greater developmental coverage possible



McCormick, E. M., Peters, S., Crone, E. A., & Telzer, E. H. (2021). Longitudinal network re-organization across learning and development. *NeuroImage*, *229*, 117784.

Accelerated Design

- Advantages
  - Much greater developmental coverage possible
  - To some degree all longitudinal studies are accelerated
    - We never get observations at *exactly* 12, 13, etc.
    - No compression of age heterogeneity

- Disadvantages
  - Strong exchangeability assumptions
  - All individuals have missing data (and lots of it)
  - Restricts the kinds of models available
  - Ability to detect individual differences is weaker

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

The Problem and Advantage of Missing Data

- When we plan missing data (e.g., multiple-cohort or accelerated study), the missingness does not bias parameter estimates (i.e., MCAR)
  - Allows us to do many things because we unhook maturation or age from observation number

- BUT missing data that we do **not** plan for biases parameter estimates
  - Need MAR, but mostly have MNAR (worst acronyms ever)

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Time in the Model

- To fit a longitudinal model, we need to translate the timing of observations into a predictor of the repeated measures

Mixed-Effect Model

repeated-measures outcome

effect of x

time-specific residual

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

intercept (where all predictors are 0)

observed measure of time or growth

Structural Equation Model

observed repeated measures

latent growth factor(s)

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

factor loading matrix (contains values of time)

time-specific residuals

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Model Encoding of Time

- Mixed-effects models include time like any other continuous predictor (only you know it is "special")
  - We could substitute time for something like depression or cognitive ability
  - While the model is somewhat more complex, this is essentially "just" a regression equation

repeated-measures outcome — effect of x — time-specific residual

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

intercept (where all predictors are 0) — observed measure of time or growth

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Model Encoding of Time

- Mixed-effects models include time like any other continuous predictor (only you know it is “special”)
  - We could substitute time for something like depression or cognitive ability
  - While the model is somewhat more complex, this is essentially “just” a regression equation

```r
mlm.lme4 <- lme4::lmer(dlpfc ~ 1 + wave + (1 + wave | id),
                       na.action = na.omit,
                       REML = TRUE,
                       data = executive.function.long)
```


McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker’s Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Model Encoding of Time

- Structural equation models, but contrast, are less intuitive in how they encode time
  - Can be encoded as a parameter of the model (LCM), or implicit in how the model is constructed (LCSM)

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Model Encoding of Time

- In the LCM, values of time are fixed in the factor loading matrix

observed repeated measures

latent growth factor(s)

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

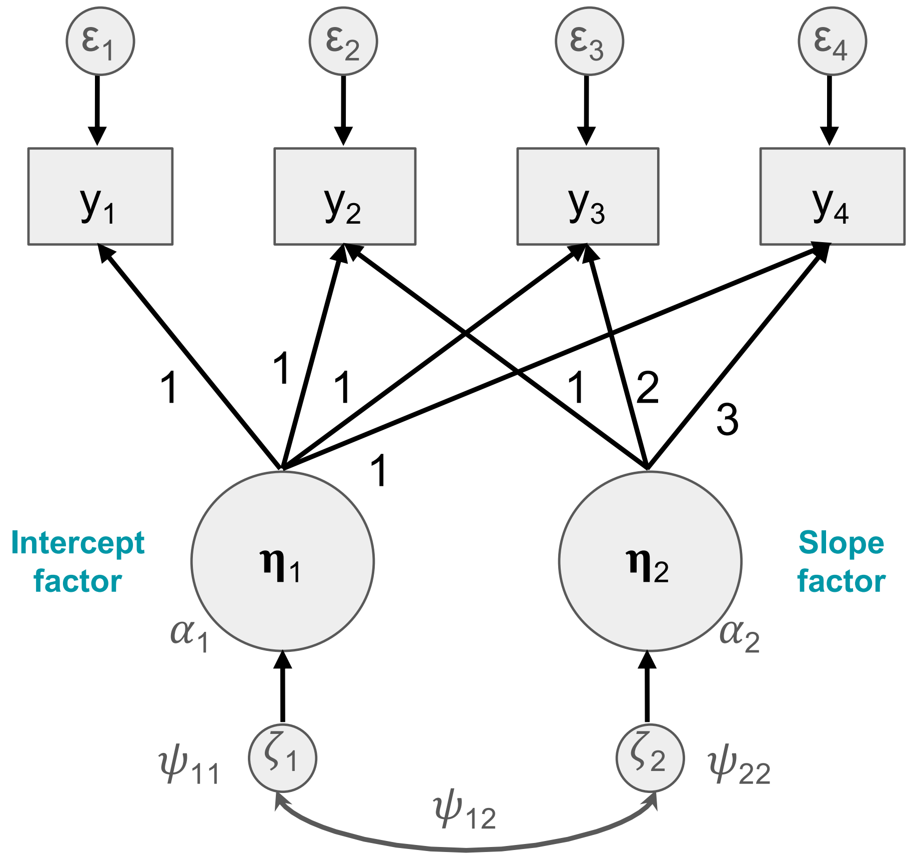
factor loading matrix (contains values of time)

time-specific residuals

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Model Encoding of Time

- In the LCM, values of time are fixed in the factor loading matrix



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Model Encoding of Time

- In the LCM, values of time are fixed in the factor loading matrix

```
linear.lcm <- "
               # Define the Latent Variables
               int =~ 1*dlpfc1 + 1*dlpfc2 + 1*dlpfc3 + 1*dlpfc4
               slp =~ 0*dlpfc1 + 1*dlpfc2 + 2*dlpfc3 + 3*dlpfc4
```

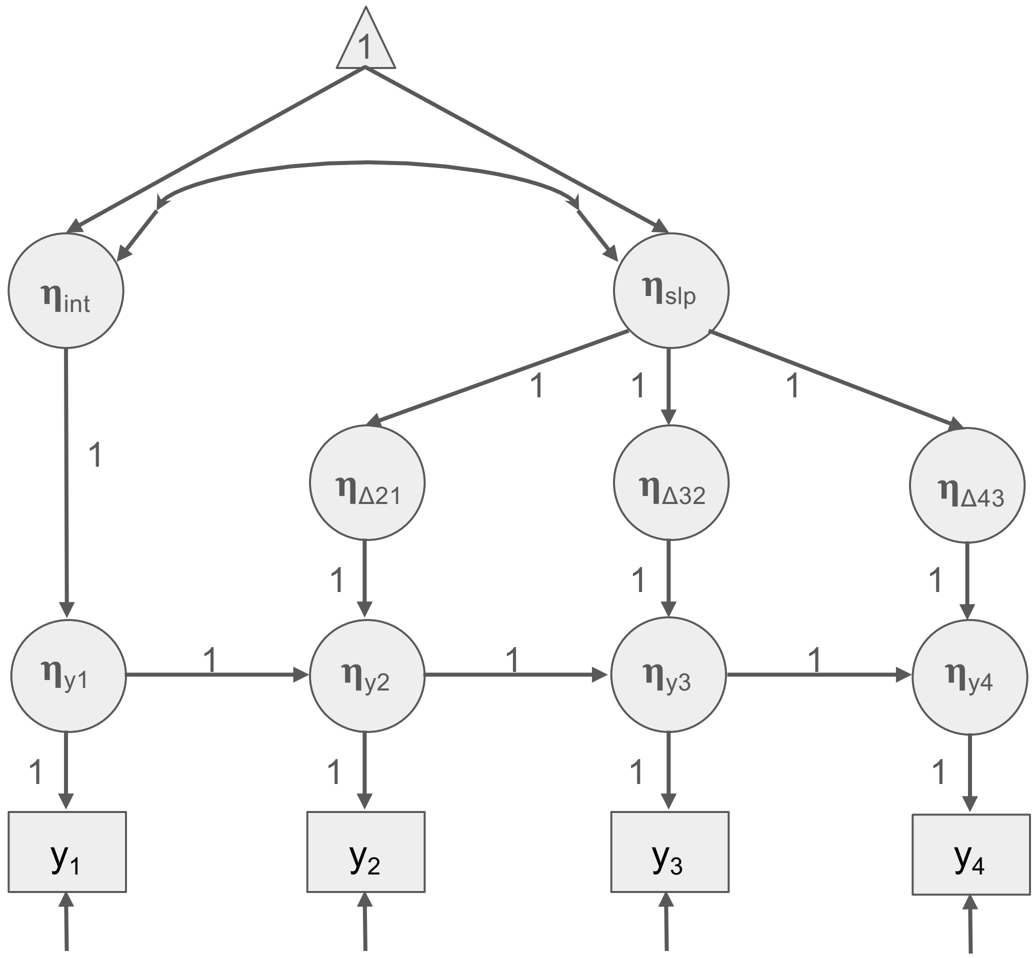

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Model Encoding of Time

- Even more oddly, the LCSM does not contain values of time anywhere in the model
  - Rather time is built in by summing up the difference scores

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Model Encoding of Time



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Implications of the Relative Approaches

- For MEMs, we have maximum flexibility
  - No two individuals need to share the same values of time
  - Allows for modeling any type of assessment schedule by default
- For SEMs, we are more often restricted to cohort designs (single or multiple)
  - Each repeated measure is treated as its own outcome variable
    - Need to compute a unique residual
  - Model needs special tools (e.g., definition variables) to work for accelerated designs

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Time Coding

- For all models, where we center time will influence the parameter estimates we recover
  – Very common to have 0 be at the first time point, but reasonable alternatives exist
  – **All time coding models fit the data identically**
    - But your exact parameter estimates will change to reflect the specification

- While of little concern for unconditional models, this becomes important for including covariates and distal outcomes

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Time Coding

One point of frequent confusion when modeling data (nevermind longitudinal models) is the role of centering the predictors for model results/fit/etc. In general, centering predictors does not change the fundamental information contained within the model, although sometimes it is necessary for practical reason (e.g., reducing collinearity between main effects and product terms). In longitudinal models, the main centering concern is where to place the intercept (i.e., where time is coded 0). While many of our parameter estimates will indeed change based on where we choose to estimate the intercept (most notably the…wait for it…intercept, as well as covariances with the intercept). Here we will demonstrate with the LCM framework since the factor-loading matrix makes what is happening very explicit, but you could replicate these results with any of the other approaches.

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Alternative Forms of Time

- Age is **by far** the most common form of time in longitudinal modeling
  - Easy to measure precisely, intuitive

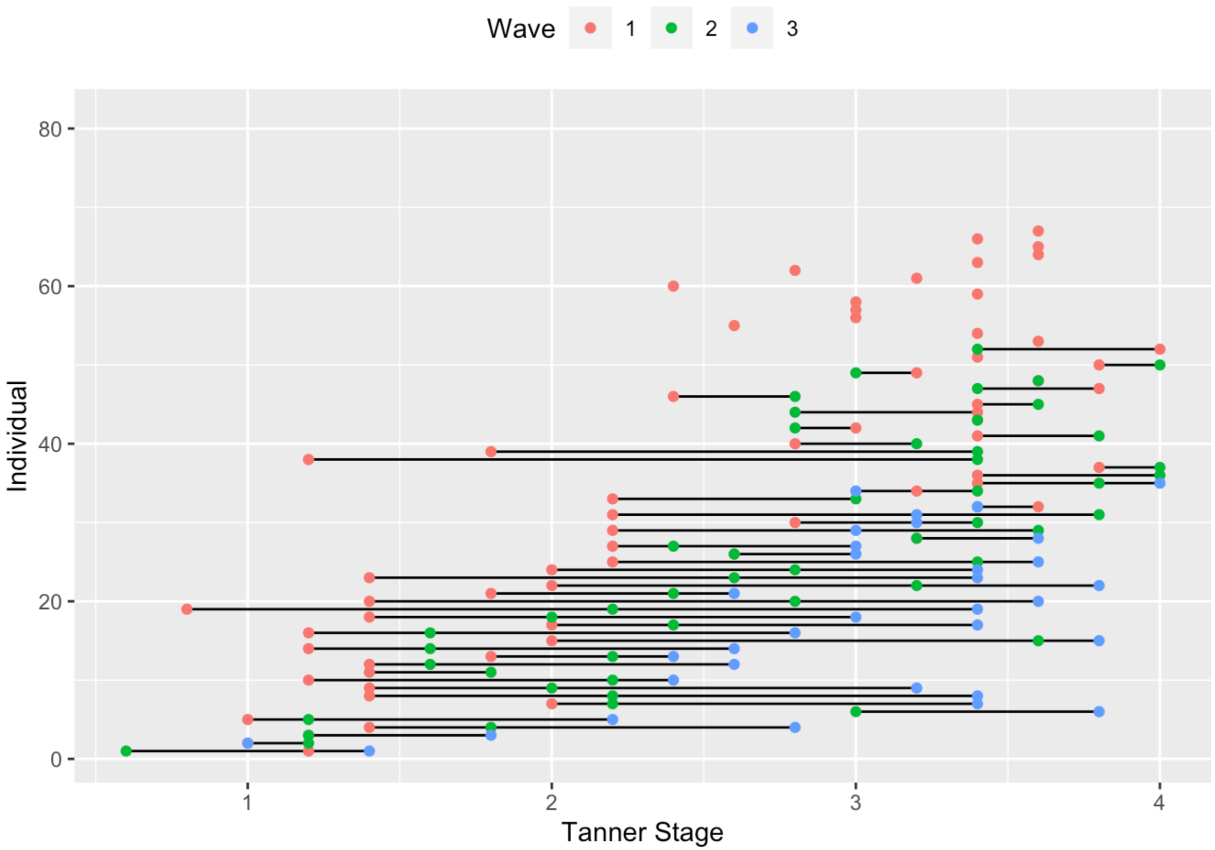
- But in reality, many developmental processes influence behaviors/cognitions of interest
  - Age: opportunity differences, peer effects, (often a catch-all)
  - Puberty: tuning of physiology
  - Experience: practice, consolidation

- In longitudinal studies, we also worry about retest effects (e.g., practice, scanner anxiety)

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Alternative Forms of Time

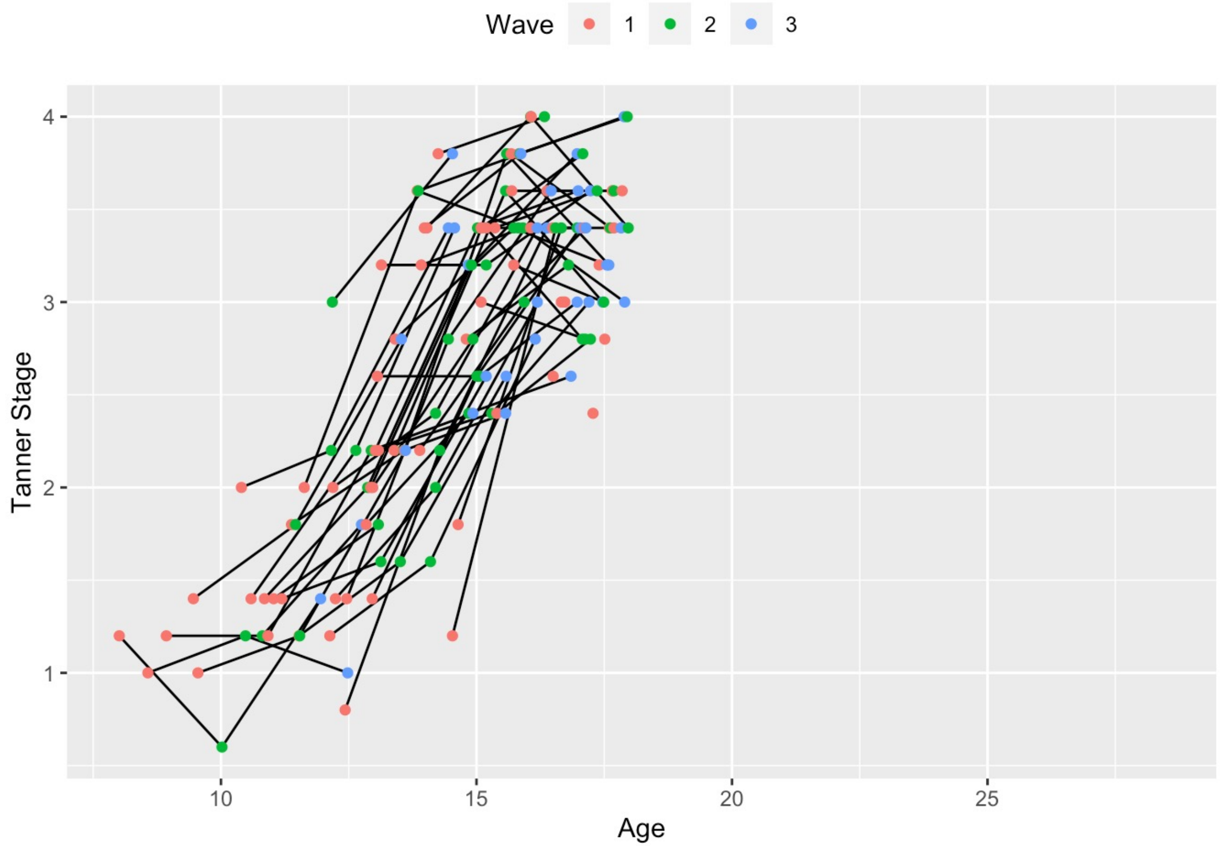
- For instance, we could structure our longitudinal model based on pubertal status



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Alternative Forms of Time

- Can take advantage of the within-age heterogeneity in pubertal status (dense vs. long observations)



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Residual Estimates

- As we touched on in the model considerations, SEMs tend to estimate unique residuals at each time point
  – MLMs by contrast tend to estimate a single residual for the outcome over time

- These are simply conventions, and we can do either approach with each model type
  – Constrain unique residuals equal in the SEMs
  – Create unique estimates per time point in MEM

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Questions?

@E_M_McCormick
e.m.mccormick@fsw.leidenuniv.nl
https://e-m-mccormick.github.io/

3. Determining the Optimal Shape of Change Over Time

The Shape of Change

- A foundational challenge in longitudinal modeling is determining the functional form of change over time
  – We cannot have valid predictions of or from change if we do not know how the change occurs

- Two primary approaches
  – Known functional forms
    - Polynomials, piecewise, nonlinear functions
  – Data-driven methods
    - Splines, latent basis, proportional change

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

The Shape of Change

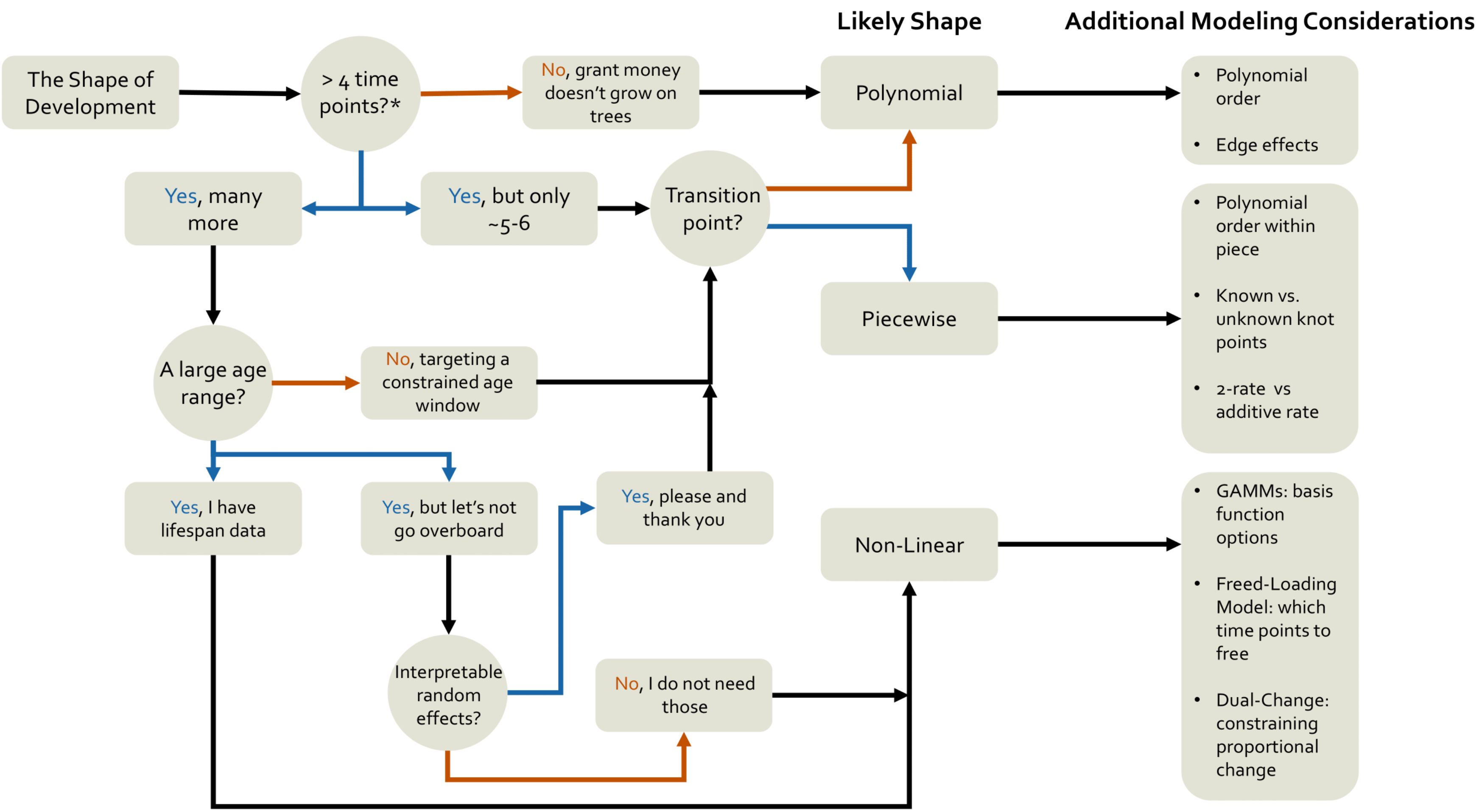
- When selecting a model, we need to keep several factors in mind:

1. Number of observations available
   - Constraining factor for more complex models
   - Is change consistent over time?
   - Fixed versus random effects

2. Interpretability versus flexibility
   - Tension between explanation & prediction
   - Generalizability

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

The Shape of Change



* Do not have to be exclusively within person

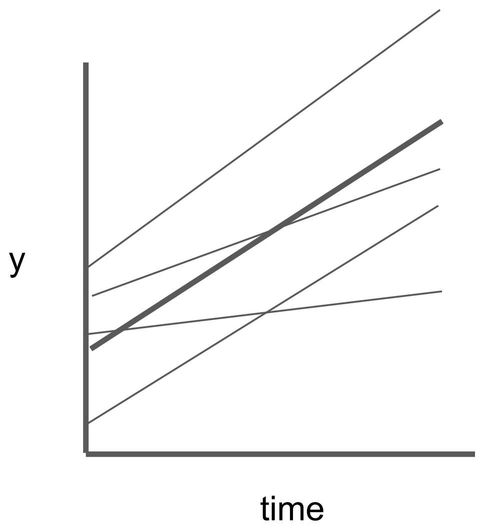
McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Fixed Versus Random Effects

- Fixed Effect:
  - Average or population effect
  - Comparable with standard regression slope

- Random Effect:
  - Individual deflection from the fixed effect
  - We don't estimate these directly
    - Rather the variance of the individual effects

Random Intercept, Random Slope



Notice that rank order is NOT preserved across time

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Polynomial Trajectories

Polynomial Models

- Polynomials are models using powered (i.e., exponents) versions of time/age
  - Similarly fit by both MLMs & SEMs

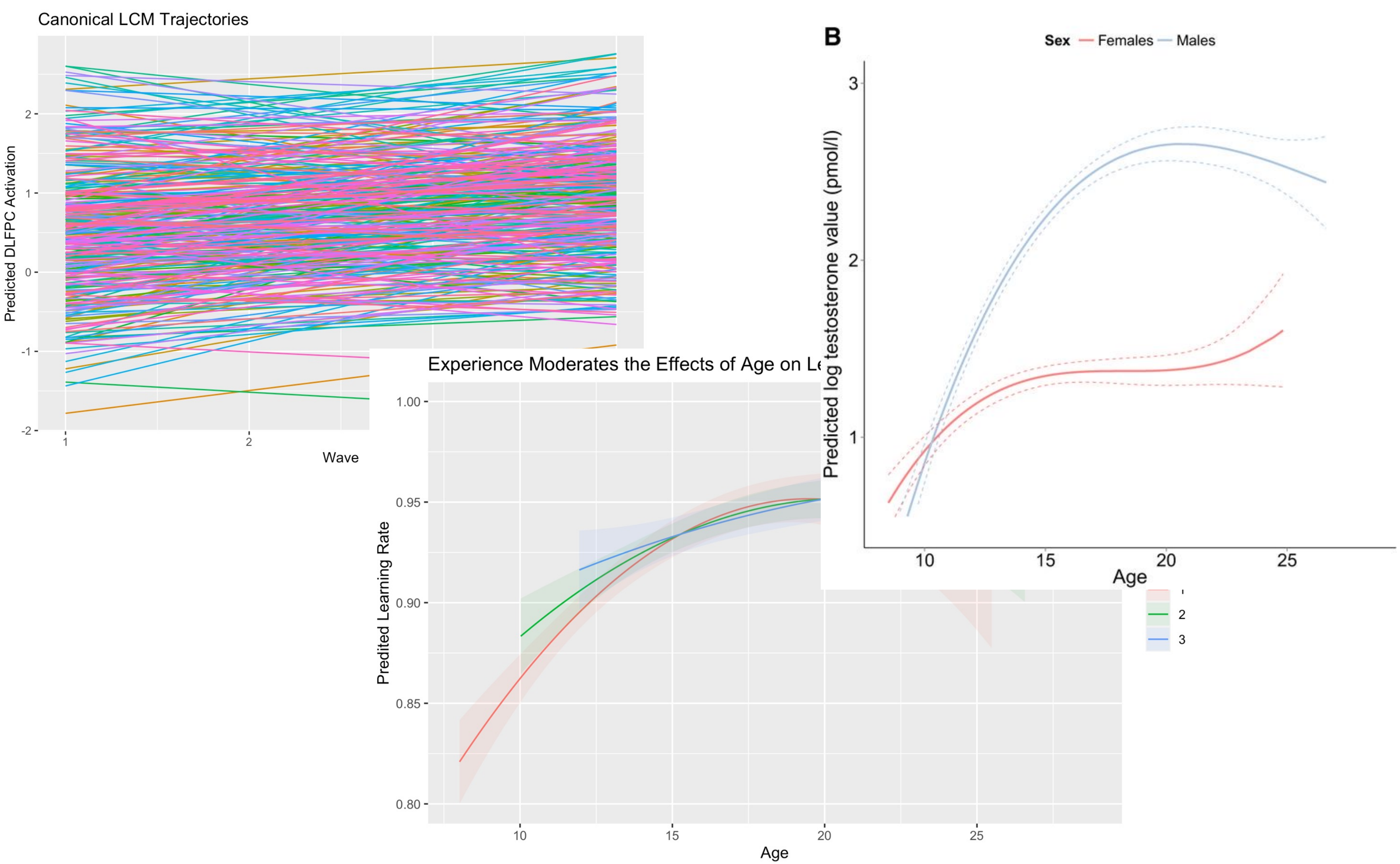
- The most common polynomial is the linear model

$$y_{ti} = \underbrace{\gamma_{0i} + \gamma_{1i} x_{ti}}_{\text{fixed effects}} + \underbrace{u_{0i} + u_{1i} x_{ti}}_{\text{random effects}} + r_{ti}$$

- Can range from very simple (e.g., intercept-only) to increasingly complex (e.g., cubic, inverse)

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Polynomial Models



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Polynomial Models

- Polynomials are stable known-function models
  - We know the exact value to specify for each unit of time in order to give a known trajectory shape
  - Highly interpretable

$$\mathbf{\Lambda}_{lin} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \quad \mathbf{\Lambda}_{quad} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \quad \mathbf{\Lambda}_{cub} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{bmatrix} \quad \mathbf{\Lambda}_{inv} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1/2 \\ 1 & 2 & 2/3 \\ 1 & 3 & 3/4 \\ 1 & 4 & 4/5 \end{bmatrix}$$

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Polynomial Models

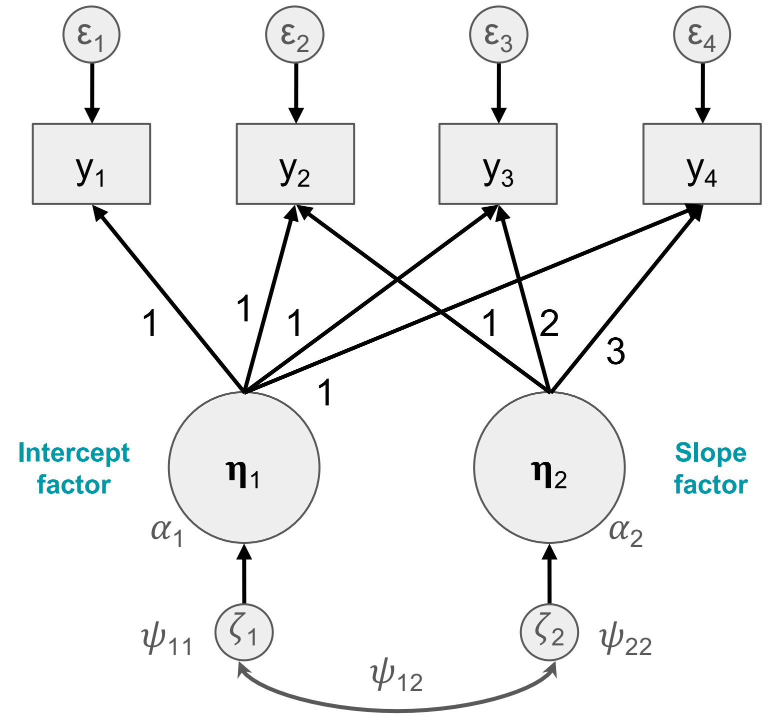
MLM:

$$y_{ti} = \underbrace{\gamma_{0i} + \gamma_{1i} x_{ti}}_{\text{fixed effects}} + \underbrace{u_{0i} + u_{1i} x_{ti}}_{\text{random effects}} + r_{ti}$$

LCM:

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

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



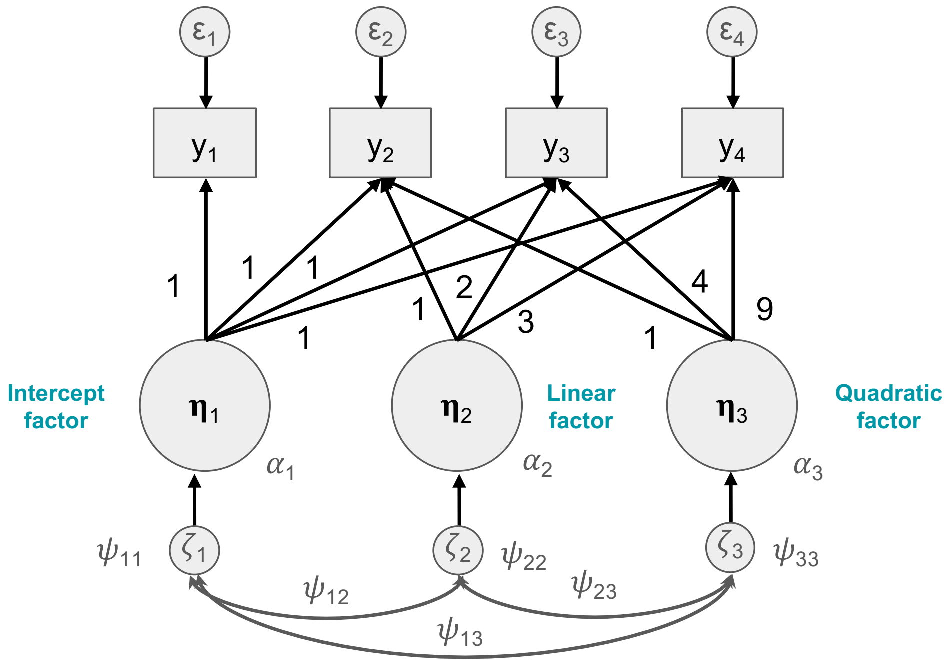
McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Fitting an Quadratic Model

MLM:

$$y_{ti} = \gamma_{00} + \gamma_{10}age_{ti} + \gamma_{20}age_{ti}^2 + u_{0i} + u_{1i}age_{ti} + u_{2i}age_{ti}^2 + r_{ti}$$

LCM:



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Polynomial Models

- Polynomials are incredibly stable known-function models
  - We know the exact value to specify for each unit of time in order to give a known trajectory shape

- Disadvantages
  - Function unbounded and extends outside the data range
    - Unrealistic over longer periods of time
    - Data points at the edges of the age-range tend to have high leverage
  - Higher-order polynomials have much lower replicability/generalizability

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Polynomial Models

- A major decision point is determining the complexity of the polynomial that is sufficient to explain the data without overfitting
  - This can involve both formal model comparisons and informal checks

- Somewhat limited by the requirements of higher-order polynomials
  - Cohort studies often do not have sufficient time points for higher order models
  - In accelerated studies, we can fit fixed-effects for higher-order effects

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Polynomial Models

- A quick side-note on time-point requirements
  - Often said that 3 repeated measures is needed for linear models, 4 for quadratic, etc.
  - **These are necessary but unlikely to be sufficient** conditions for random effects models

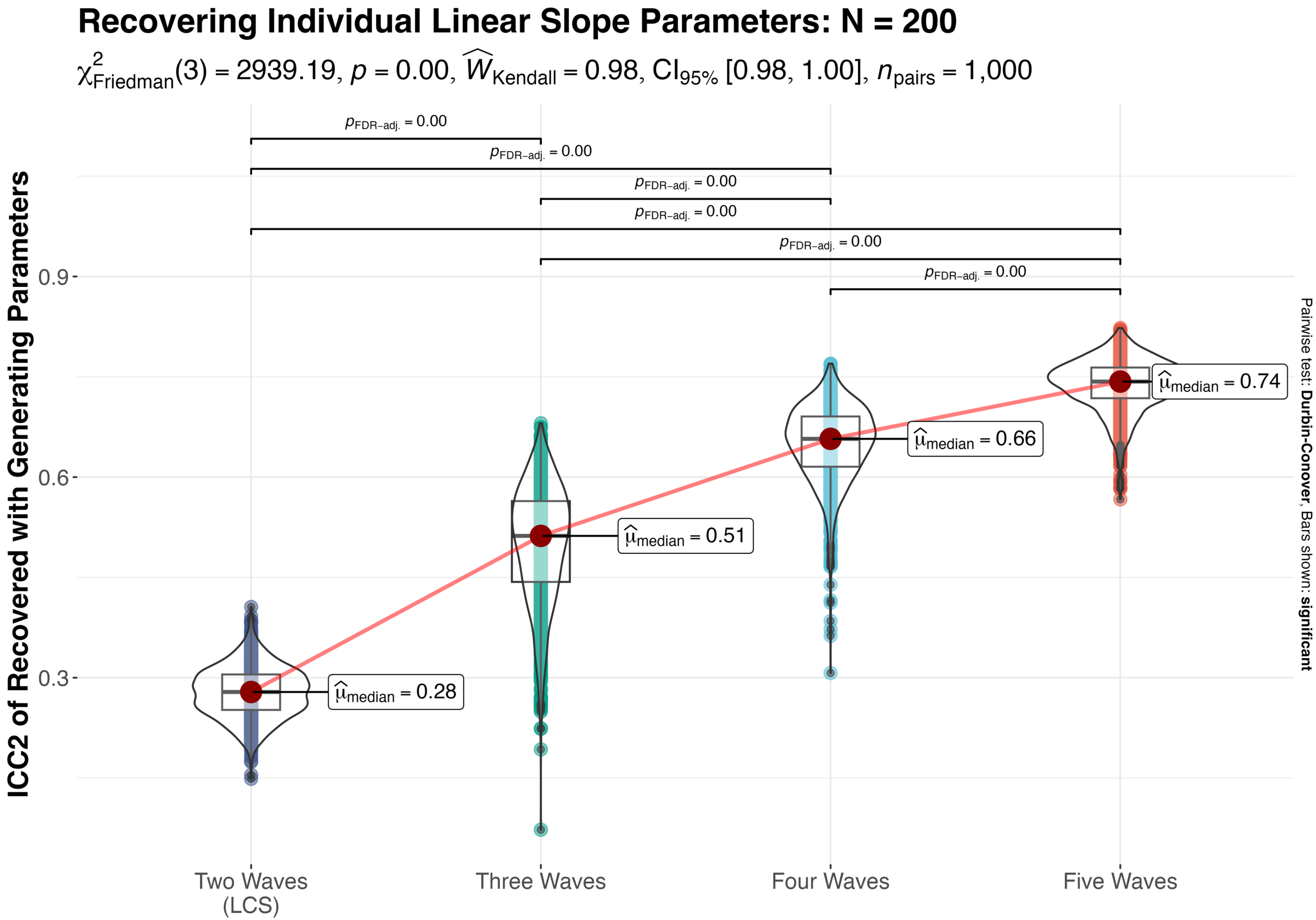
Two timepoints poorly capture trajectories of change: A warning for longitudinal neuroscience

Sam Parsons[1] and Ethan M. McCormick*[1,2,3]

[1]Cognitive Neuroscience Department, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, Netherlands
[2]Methodology & Statistics Department, Institute of Psychology, Leiden University, Leiden, Netherlands
[3]Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill, United States

Parsons, S. & McCormick, E. M. (2023, March 23). Two timepoints poorly capture trajectories of change: A warning for longitudinal neuroscience. https://doi.org/10.31234/osf.io/96ph3

Polynomial Models



Parsons, S. & McCormick, E. M. (2023, March 23). Two timepoints poorly capture trajectories of change: A warning for longitudinal neuroscience. https://doi.org/10.31234/osf.io/96ph3

Polynomial Models

Polynomial Trajectories

Like we discussed in the main text, polynomial trajectories are far and away the most common trajectories modeled with longitudinal data. They require relatively few unique timepoints, are straightforward to model, and offer easily-interpretable parameter estimates.

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Piecewise Trajectories

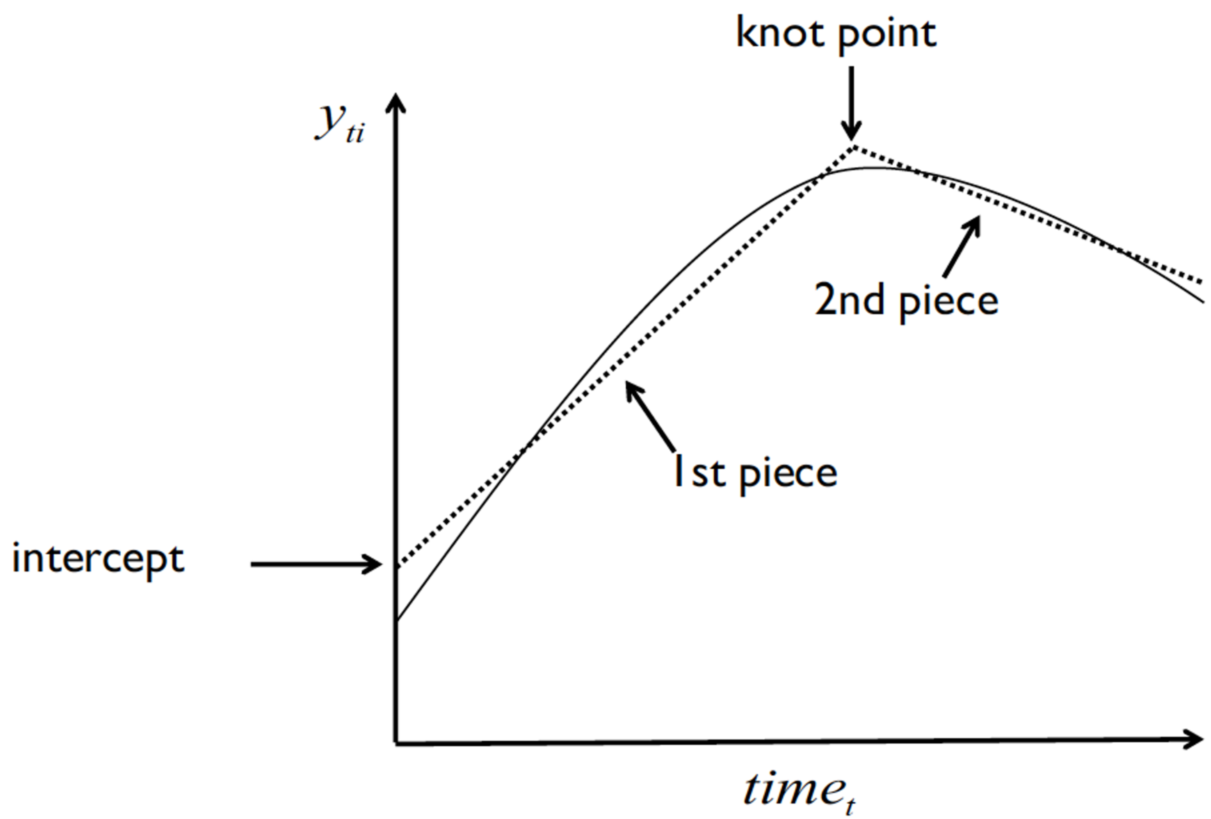
Piecewise Models

- One of the major downsides to classic polynomial models is that they are unbounded an extend infinitely with the same shape
  - Might not be able to approximate complex trends
  - But they have nice interpretation features we like

- One alternative is to stitch together two or more polynomial models
  - Can maintain local interpretability but also allow for complex transitions in the trajectory

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Piecewise Models



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Piecewise Models

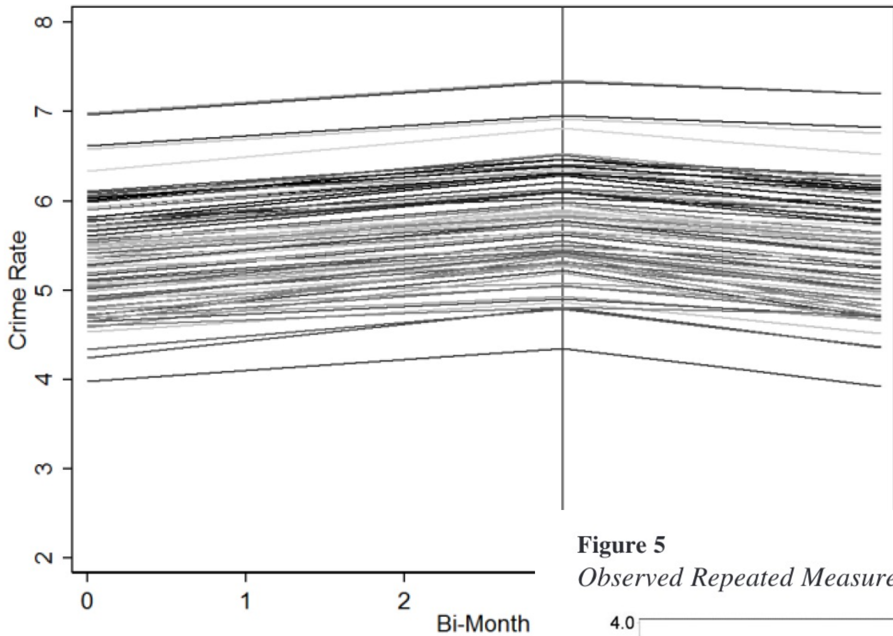
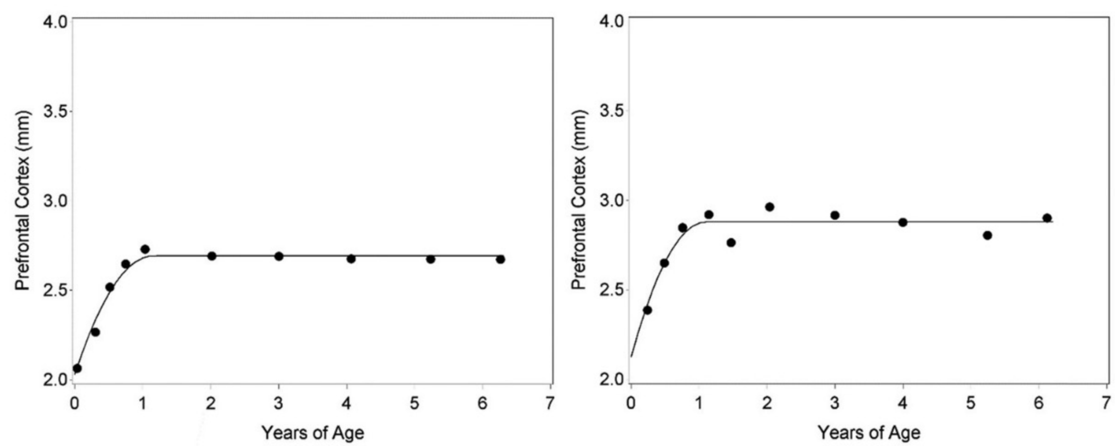


Figure 5

Observed Repeated Measures (Circles) and Implied Trajectories for Six Representative Children



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Piecewise Models

- Where we join two function is known as the **knot point**
  - Placement of this point is a key consideration in these models

- Constraints
  - We must have sufficient time points to specify the desired functions on either side of the knot point
  - Cannot place the knot points at the edges of the trajectory
  - Example: linear-linear models require 5 time points
    - 3 for each piece with one shared time point in the center
    - Only one place for the knot point

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

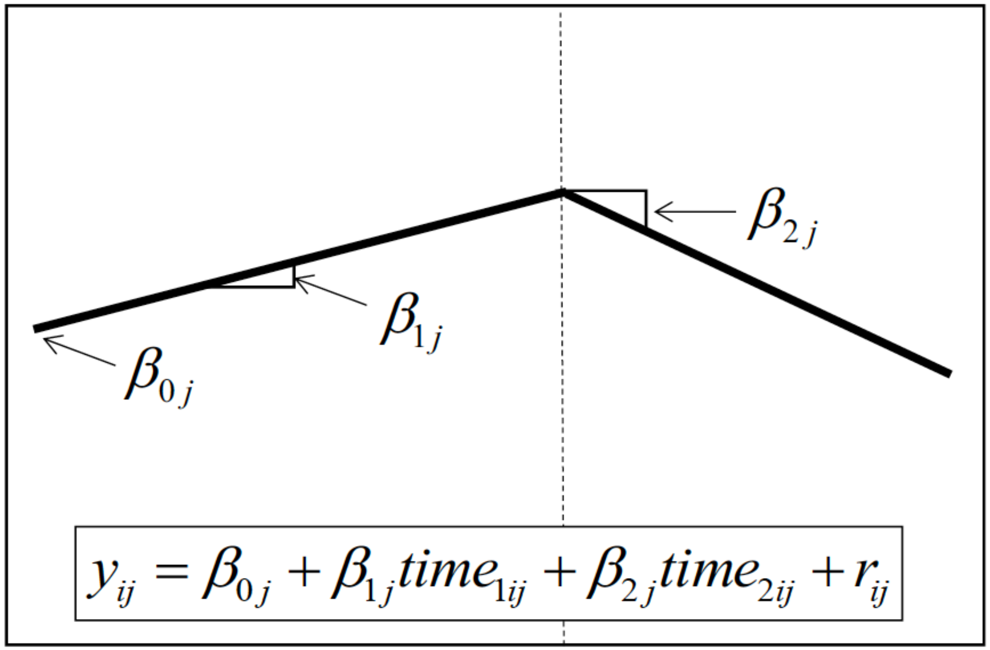
Linear-Linear Piecewise Model

- For linear-linear models, there are two kinds of specifications we might adopt

- 2-rate parameterization
  - Most intuitive (and most common)
  - Two linear slopes that can be interpreted independently

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Linear-Linear Piecewise Model

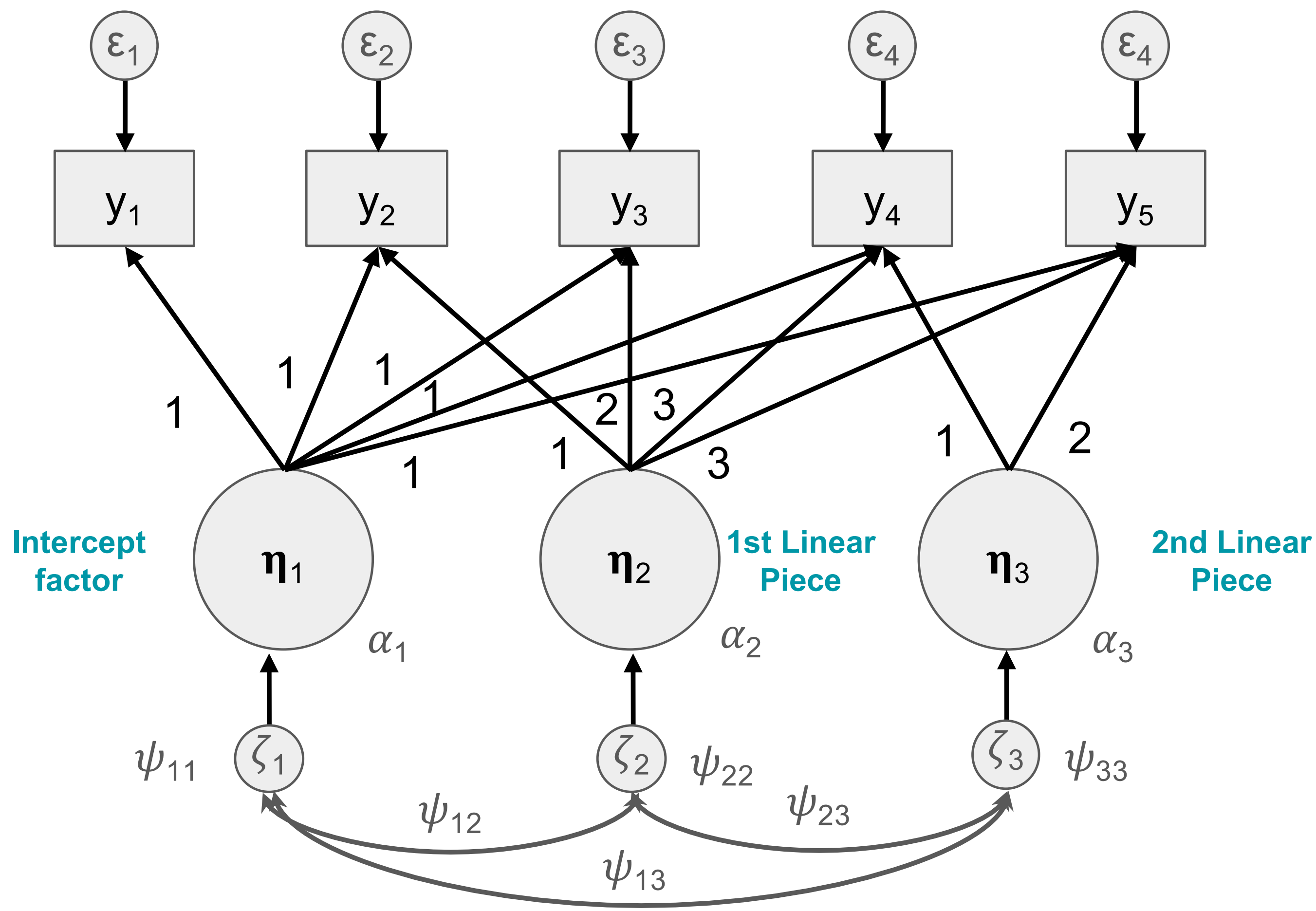


time	0	1	2	3	4	5	6	7	8	9	10	11	12
$time_1$	0	1	2	3	4	5	6	7	7	7	7	7	7
$time_2$	0	0	0	0	0	0	0	0	1	2	3	4	5

β_{1j} and β_{2j} are the slopes in the two periods, respectively

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Linear-Linear Piecewise Model



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Linear-Linear Piecewise Model

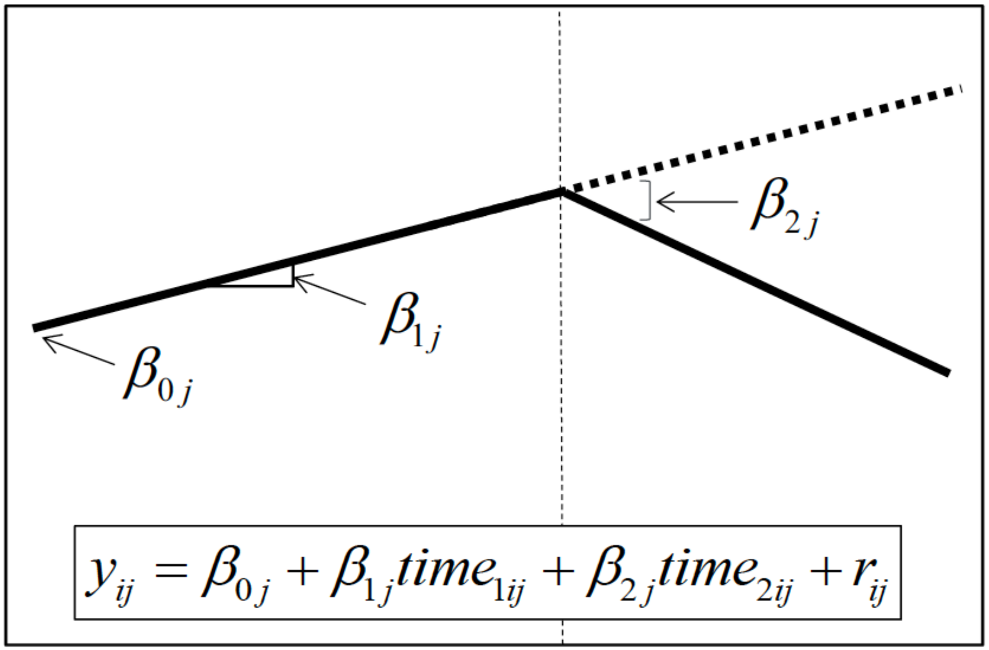
- For linear-linear models, there are two kinds of specifications we might adopt

- 2-rate parameterization
  - Most intuitive (and most common)
  - Two linear slopes that can be interpreted independently

- Added-rate parameterization
  - Second slope is the difference (or deflection) from the first slope

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Linear-Linear Piecewise Model

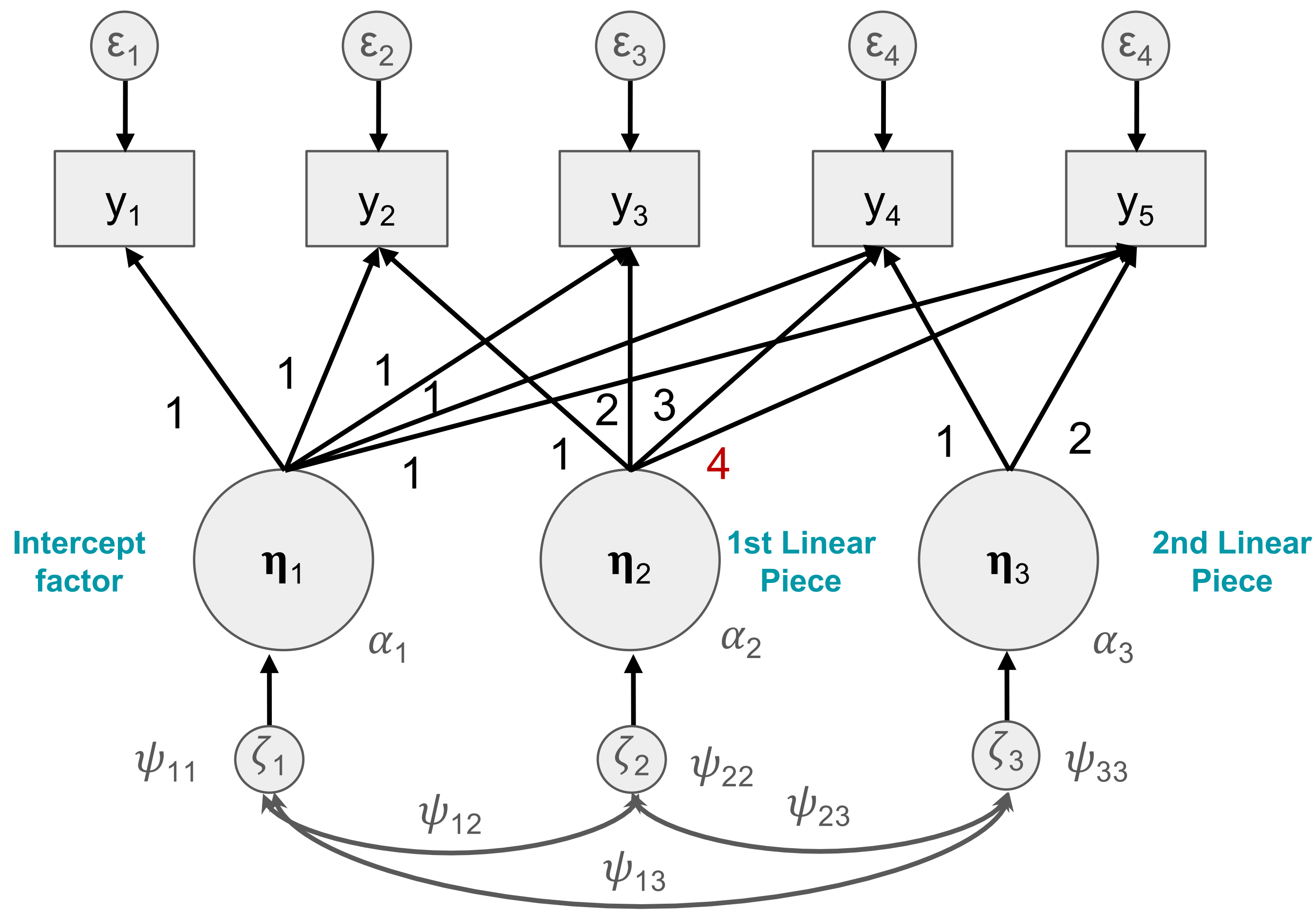


β_{2j} is the change in slope from β_{1j}

time	0	1	2	3	4	5	6	7	8	9	10	11	12
$time_1$	0	1	2	3	4	5	6	7	8	9	10	11	12
$time_2$	0	0	0	0	0	0	0	0	1	2	3	4	5

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2021, December 7). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Linear-Linear Piecewise Model



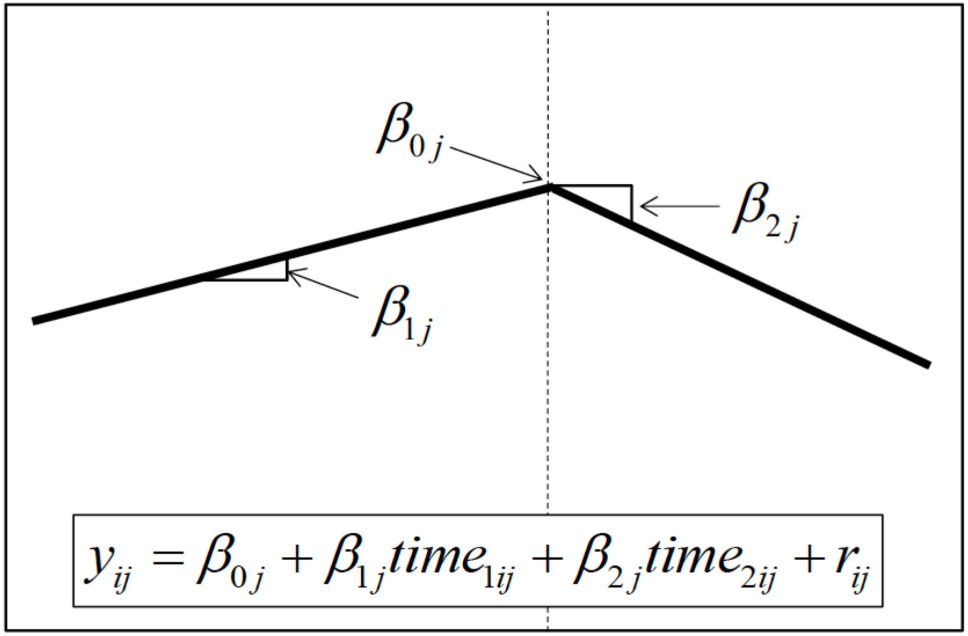
McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Linear-Linear Piecewise Model

- Time centering
  - While we can put the intercept wherever we wish - similar to other growth models – it is often of most theoretical interest to model individual variability at the transition (knot) point

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Linear-Linear Piecewise Model



time	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5
$time_1$	-7	-6	-5	-4	-3	-2	-1	0	0	0	0	0	0
$time_2$	0	0	0	0	0	0	0	0	1	2	3	4	5

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Linear-Linear Piecewise Model

- Time centering
  - While we can put the intercept wherever we wish - similar to other growth models – it is often of most theoretical interest to model individual variability at the transition (knot) point

- Depending on the number of time-points we have, we could estimate and additional intercept to indicate some sort of discontinuity (i.e., due to treatment)

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Comparing 2- and Added-Rate Models

- Fit of the two model is **identical**
  - Re-packaging the same information depending on what we want to interpret
  - Trajectories implied will also be identical

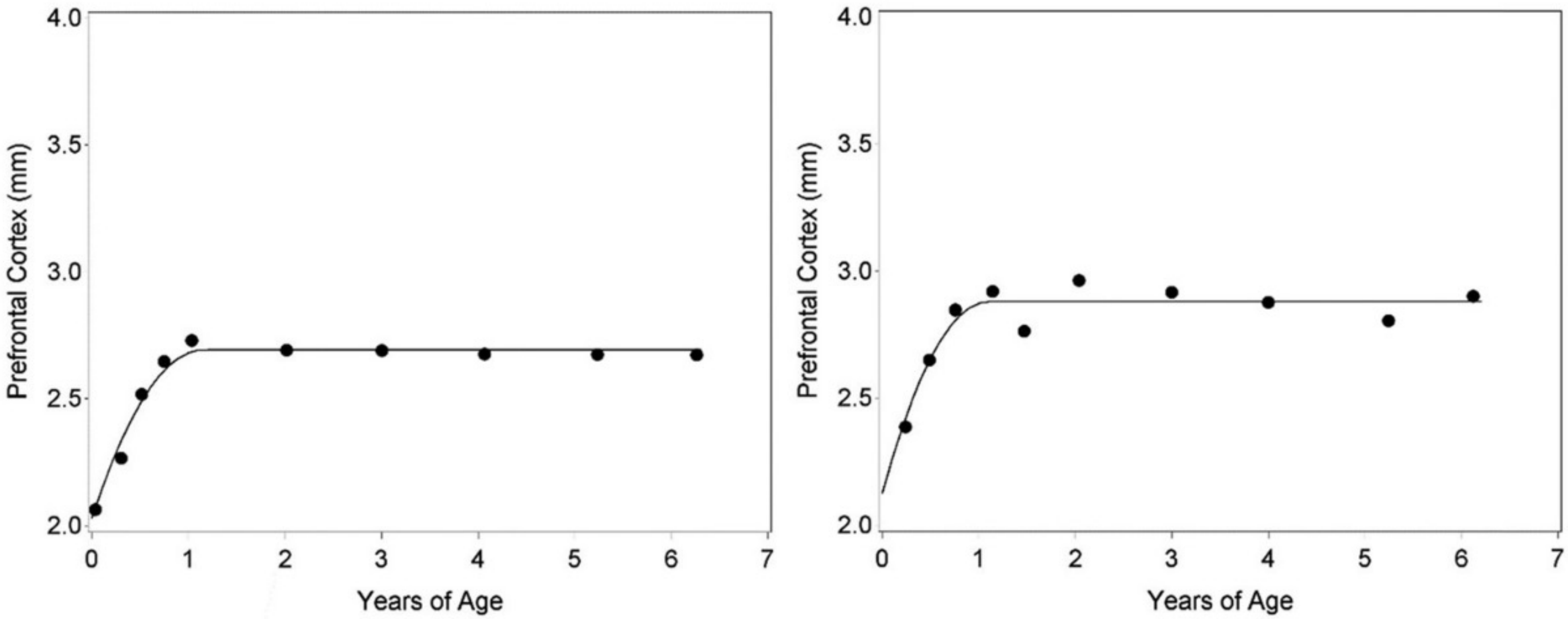
- Negative slp2 in added-rate
  - Indicates that slp1 is more steeply positive, so it negatively deflects from the initial trajectory
  - Often need to plot the data to make sense of the added rate parameter

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Quadratic-Linear Model

Figure 5

Observed Repeated Measures (Circles) and Implied Trajectories for Six Representative Children



McNeish, D., Bauer, D. J., Dumas, D., Clements, D. H., Cohen, J. R., Lin, W., Sarama, J., & Sheridan, M. A. (2021). Modeling individual differences in the timing of change onset and offset. *Psychological Methods.* Advance online publication.

Extensions of Piecewise Models

- Additional knot points
  - With more observations, we can potentially have multiple knot points
  - Less common with cohort studies, but possible as fixed effects with accelerated studies (or with more intensive data)

- Unknown knot points (see work by Bob Cudeck)
  - Fixed effects: estimate where the knot point should go
  - Random effects: allow for individual transition points

Kohli, N., Harring, J. R., & Hancock, G. R. (2013). Piecewise linear–linear latent growth mixture models with unknown knots. *Educational and Psychological Measurement*, *73*(6), 935-955.

Fitting a Piecewise Model

Piecewise Trajectories

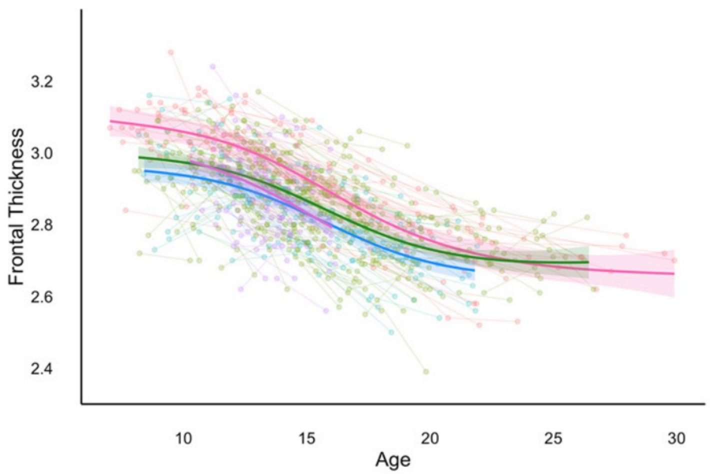
If we do not think that a single polynomial function can sufficiently capture a complex trajectory, we might consider bolting two (or more) polynomial functions together using a piecewise approach. Here we will use the `adversity` data which covers 8 years of childhood (ages $4 - 11$). The simplest piecewise trajectory can be constructed two distinct linear pieces joined at a knot point. We need at least 3 time points to specify a line but the pieces can share a time point at the knot point. This means we need a minimum of 5 time points in order to fit even the simplest piecewise model. Note that with this minimum, the knot point is constrained to be at the middle time point, and the knot can never be placed at the first or last two time points because of the 3 time point requirement to estimate the linear slope. Note that as we discussed before, these time point requirements can be accomodated at the group level, and no one individual need be observed 5 or more times. Indeed this is the case here, where no individual is measured more than 4 times.

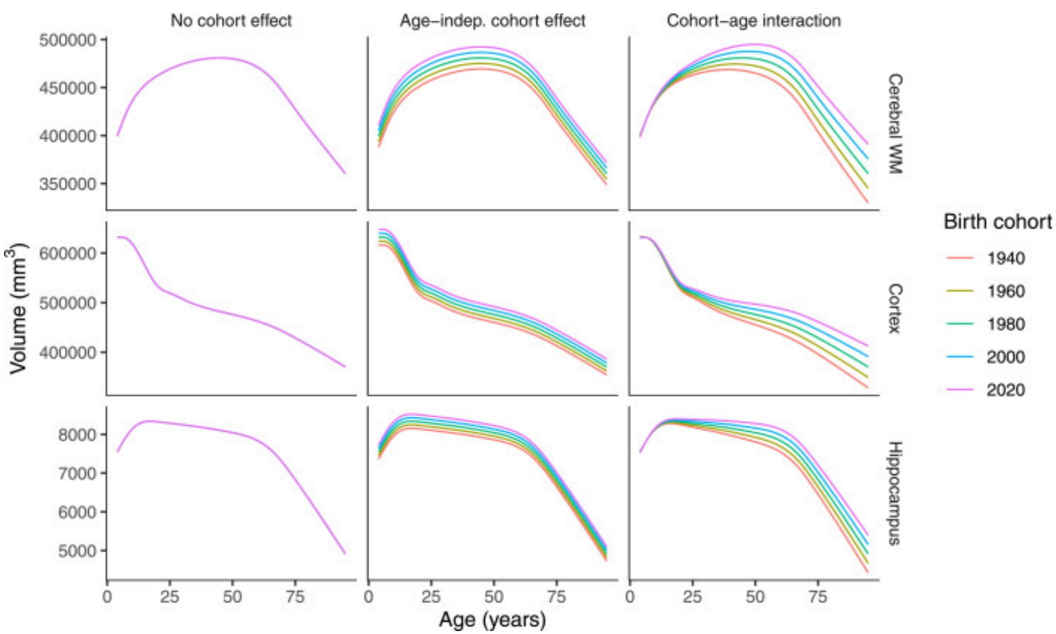
McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nonlinear Trajectories

Why Nonlinear Trajectory Models?

- Linear models are by far the most common in substantive applications
  - Pro: simple, stable, and easy to interpret
  - Con: unlikely to be true over longer time windows*



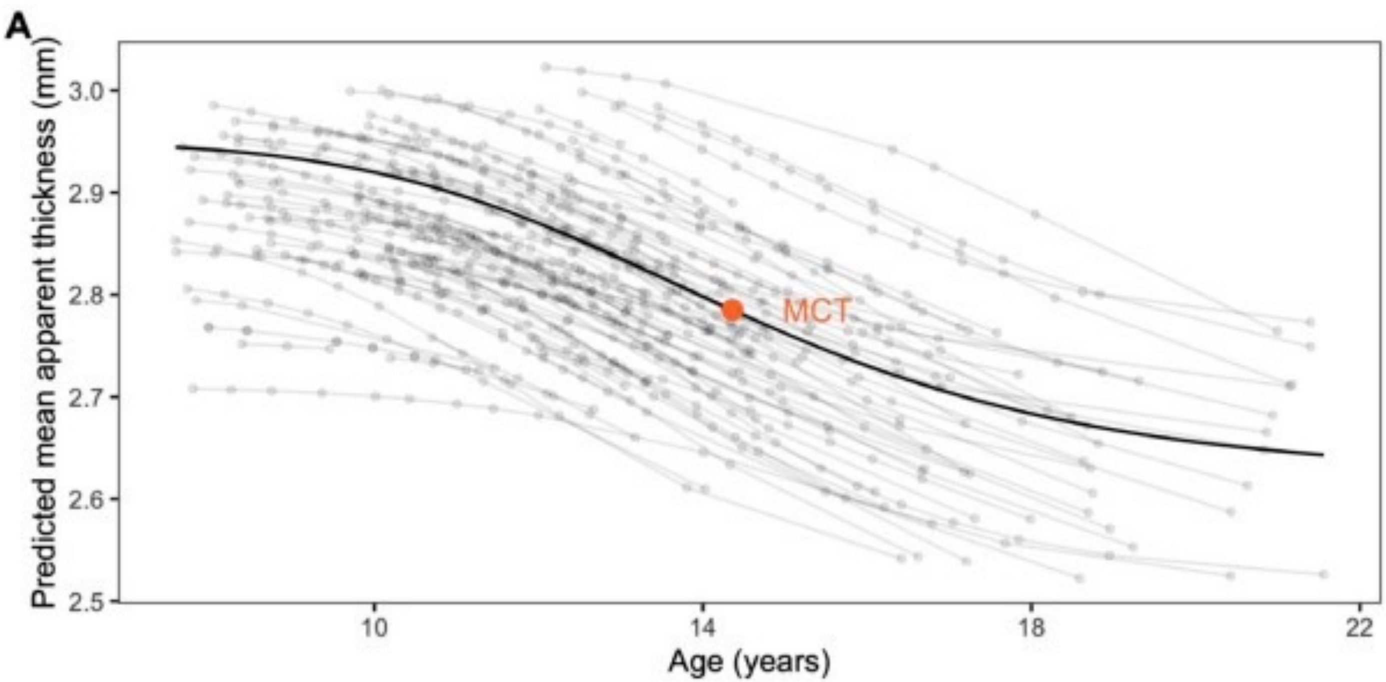


Tamnes, C. K., et. al., (2017). Development of the cerebral cortex across adolescence: a multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. *Journal of Neuroscience*, 37(12), 3402-3412.

Sørensen, Ø., Walhovd, K. B., & Fjell, A. M. (2021). A recipe for accurate estimation of lifespan brain trajectories, distinguishing longitudinal and cohort effects. *NeuroImage*, 226, 117596.

Nonlinear Trajectory Models

- Non-linear models are an attractive alternative
  - Pro: reality is messy, why should our models not be?
  - Con: data-demands, overfitting, interpretability



4PL

$$y = d + \frac{(a-d)}{\left[1+\left(x/c\right)^{b}\right]}$$

Fuhrmann, D., Madsen, K. S., Johansen, L. B., Baaré, W. F., & Kievit, R. A. (2022). The midpoint of cortical thinning between late childhood and early adulthood differs between individuals and brain regions: Evidence from longitudinal modelling in a 12-wave neuroimaging sample. *NeuroImage*, 119507.

What Does Nonlinear Mean?

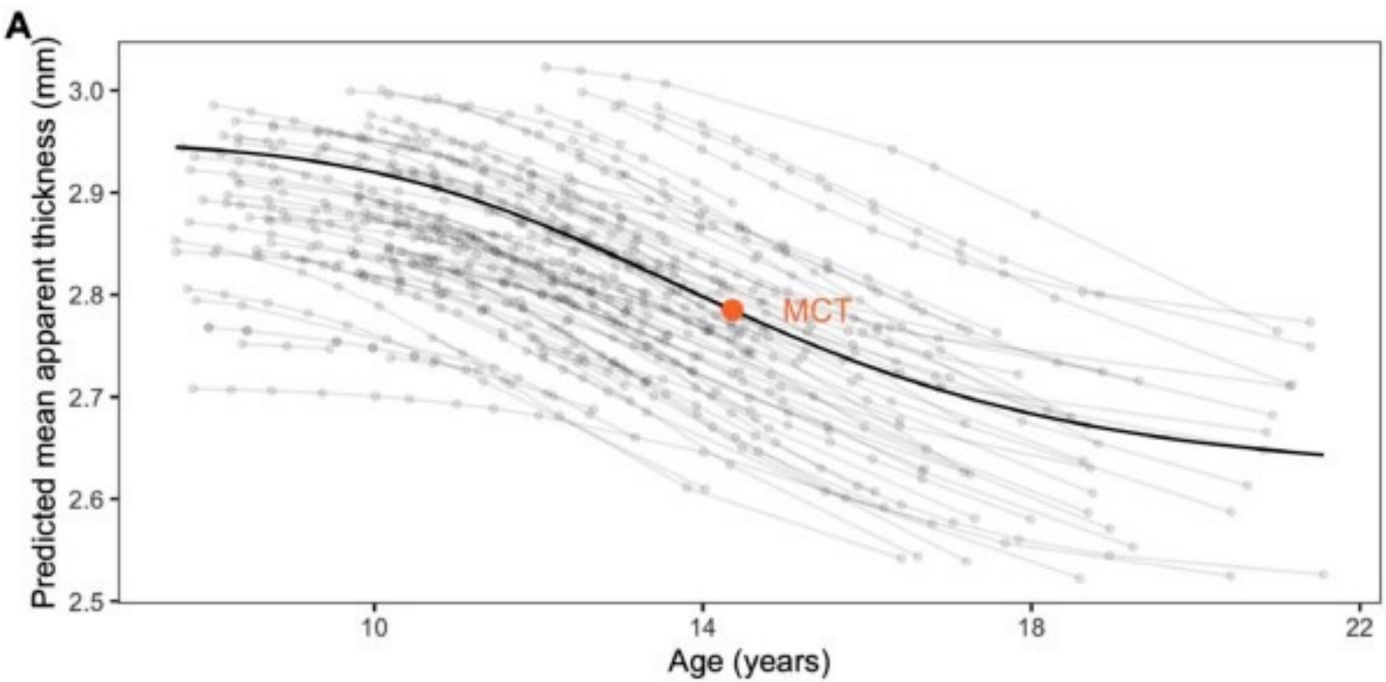
- Multiple meanings of nonlinear models
  - Non-linear in the shape of the trajectory
    - Would include higher-order polynomials
    - Mostly what we will talk about (conveniently ignoring polynomials)
  - Non-linear with respect to the parameters
    - Estimated parameters are not additive
    - Includes the more obvious things like logistics
    - Also categorical outcome models

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nonlinear Trajectories with Known Functional Forms

Nonlinear Models – Known Functional Form

- Known functional form models
  - These are models which have non-linear shapes over time, but we know the equation that describes them



4PL

$$y = d + \frac{(a-d)}{\left[1+\left(x/c\right)^{b}\right]}$$

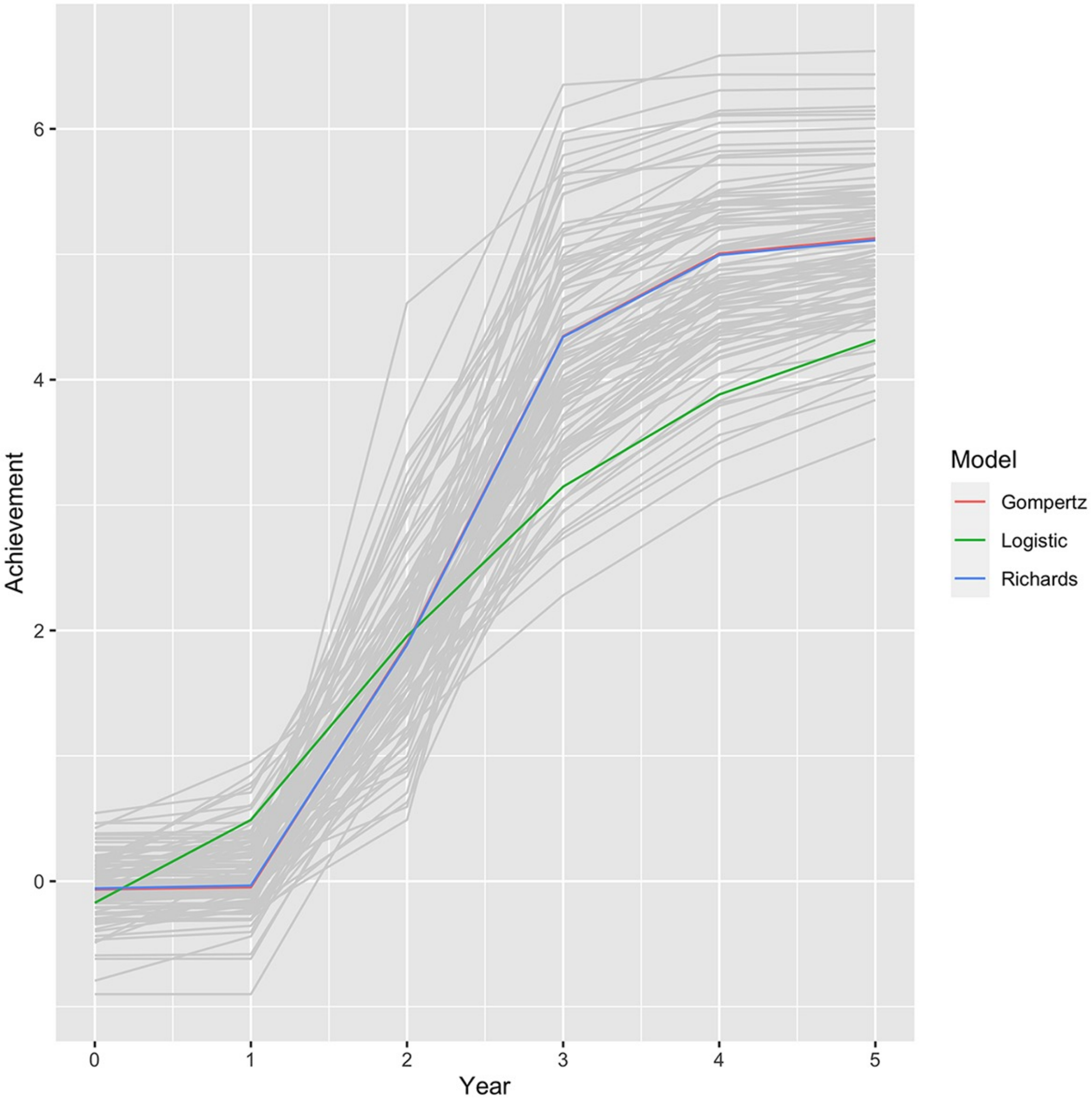
McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nonlinear Models – Known Functional Form

- Known functional form models
  - These are models which have nonlinear shapes over time, but we know the equation that describes them
  - Can be challenging to estimate but the results are often reasonably interpretable
    - Parameters often have known theoretical meaning
    - Consistent trade off between fitting an optimal curve and ease of interpretation
  - Primary way nonlinearities are modeled in MLMs

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nonlinear Models – Known Functional Form



Comets, E., Lavenu, A., & Lavielle, M. (2017). Parameter estimation in nonlinear mixed effects models using saemix, an R implementation of the SAEM algorithm. *Journal of Statistical Software, 80*(3), 1-41.

Boedeker, P. (2021). Nonlinear Mixed-Effects Growth Models: A Tutorial Using'saemix'in R. *Methodology, 17*(4), 250-270.

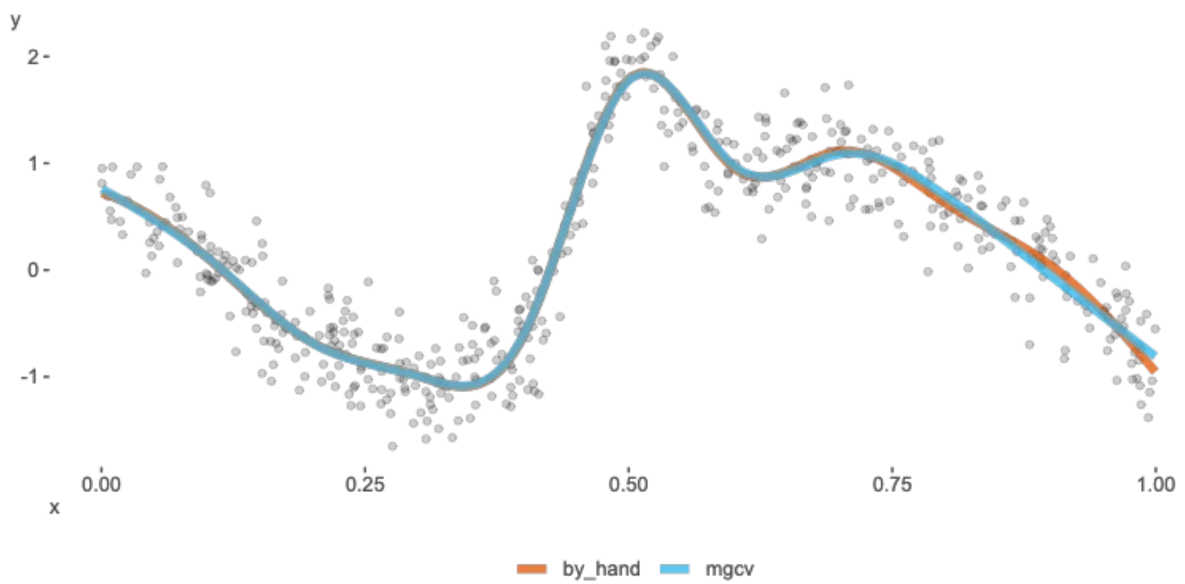
Nonlinear Trajectories with Unknown Functional Forms

Nonlinear Models – Unknown Functional Form

- Unknown functional form models
  - These are models which have nonlinear shapes over time, and we **don't** know the equation that describes them
  - Need to estimate this shape from the data

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

- Unknown functional form models
  - These are models which have nonlinear shapes over time, and we **don't** know the equation that describes them
  - Need to estimate this shape from the data



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nonlinear Models – Unknown Functional Form

- Unknown functional form models
  - These are models which have nonlinear shapes over time, and we **don't** know the equation that describes them
  - Need to estimate this shape from the data

- While estimation can be a challenge, more often the issue is avoiding overfitting
  - Data-driven methods are "greedy"
  - Want to use as few components as possible to adequately fit the curve
    - model complexity, generalizability, and interpretability

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nonlinear Models – Unknown Functional Form

- There are degrees of data-driven, and we can discuss models which increase in their flexibility
  - And therefore, tendency to overfit

1. Dual Change Latent Change Score Models

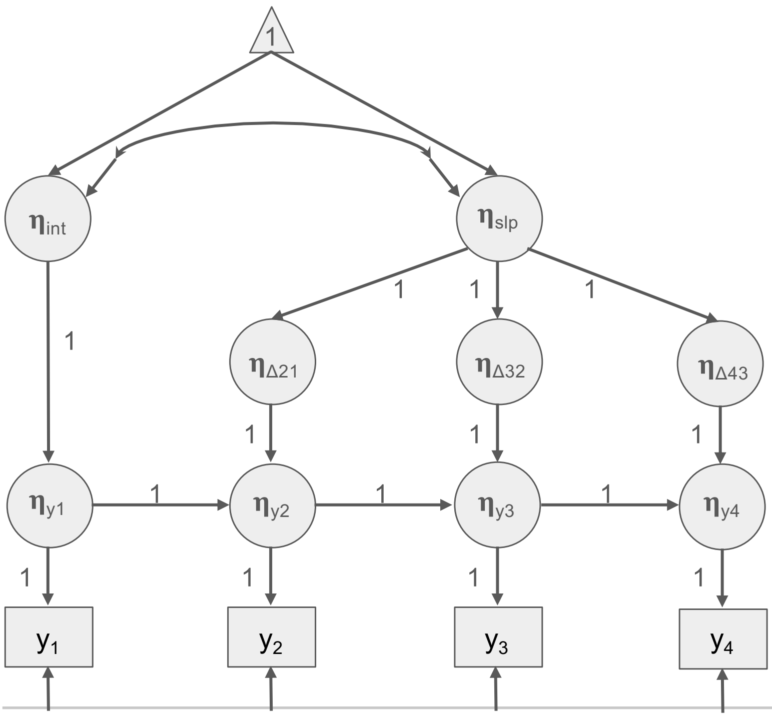
2. Free-Loading Latent Curve Models

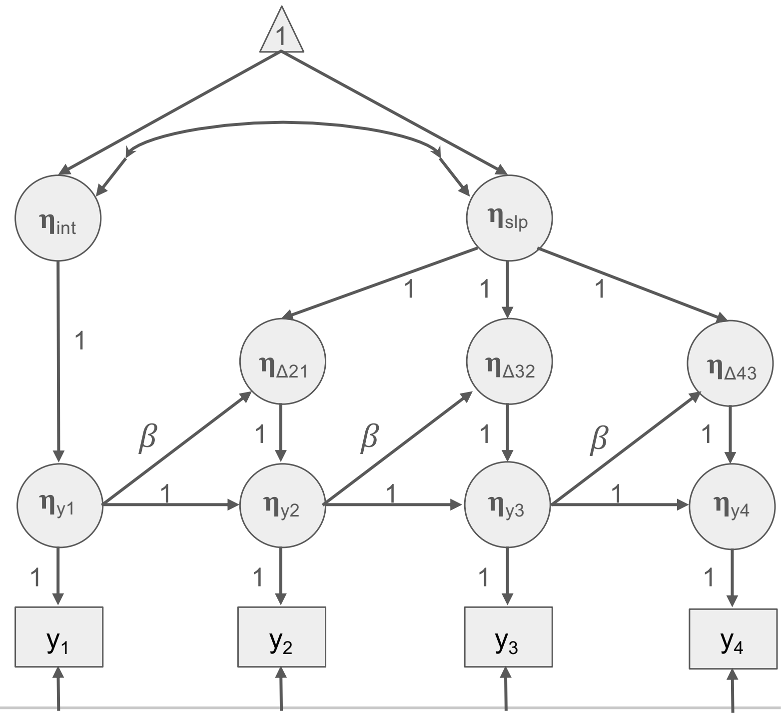
3. Generalized Additive Mixed Models

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Latent Change Score Model: Dual Change

- Until now, we have mostly ignored the latent change score model
  - Mostly because without the feature that allows for non-linearities, it's largely equivalent to the LCM

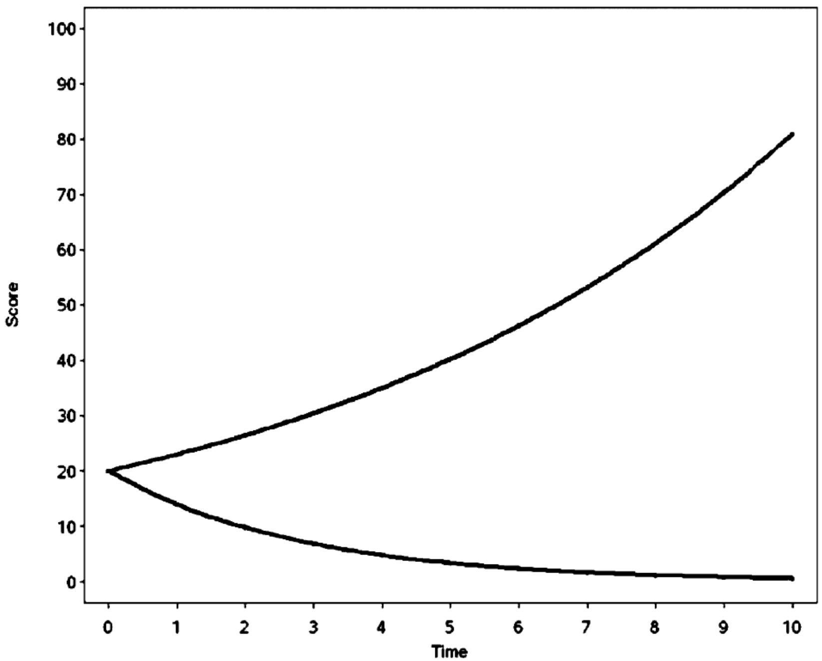


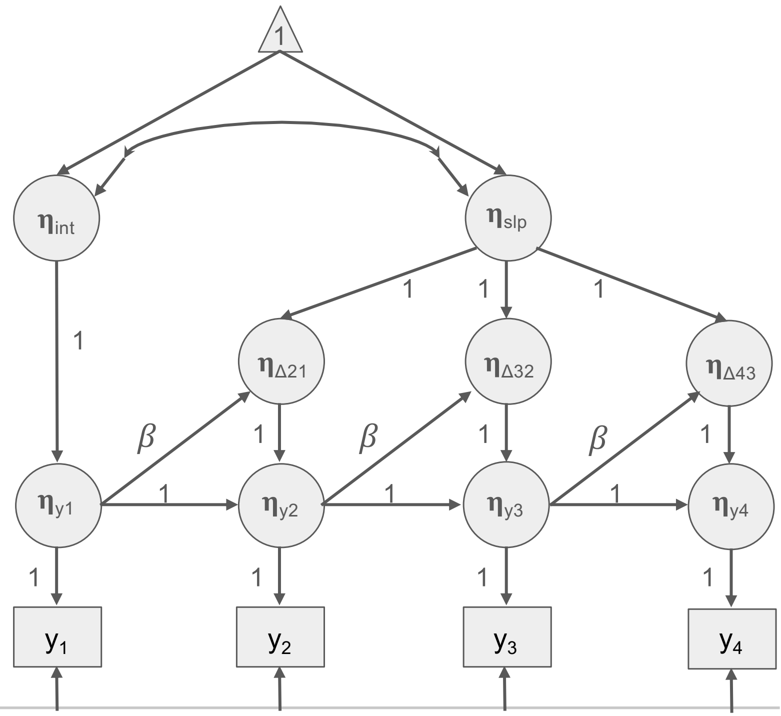


McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Latent Change Score Model: Dual Change

- The dual-change model allows us to model exponential trends in the data because prior status predicts amount of change

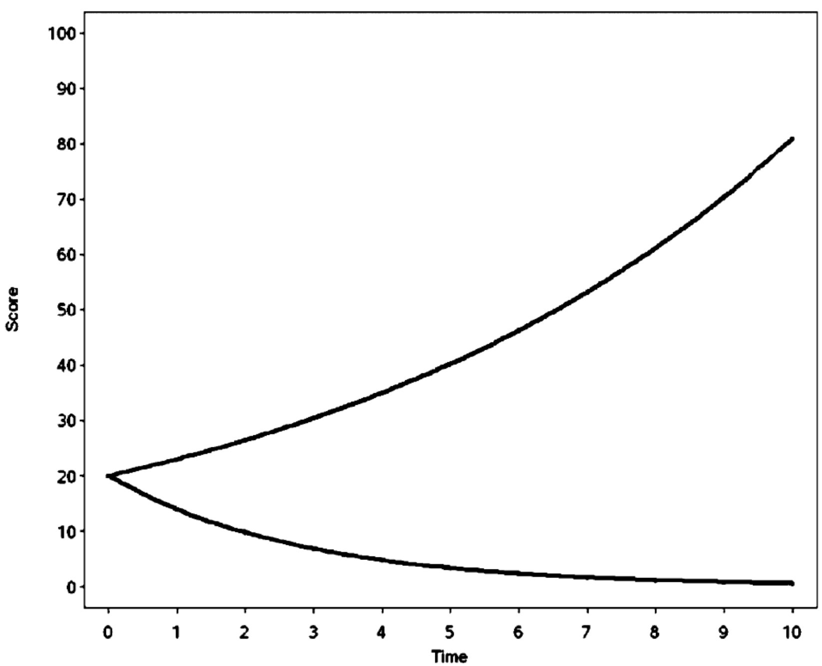


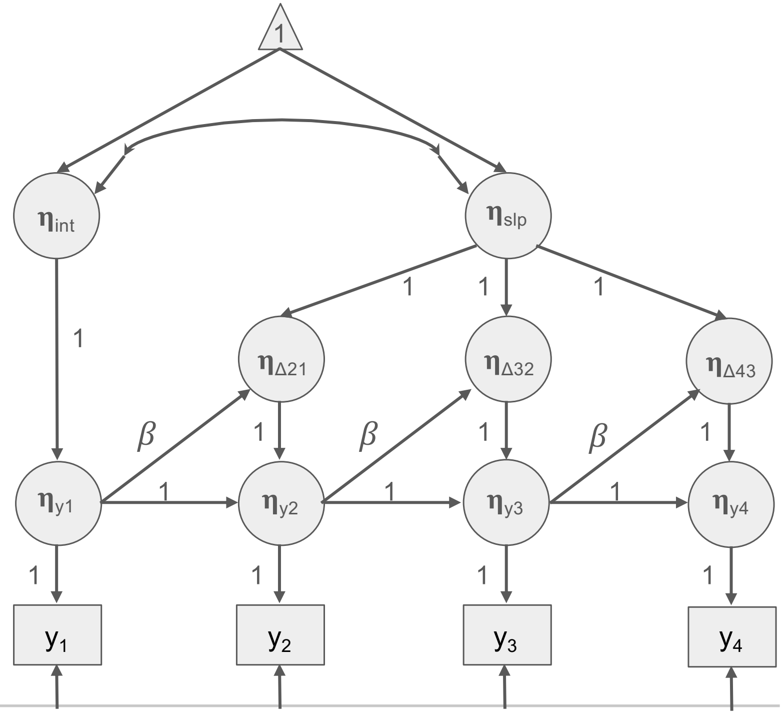


McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Latent Change Score Model: Dual Change

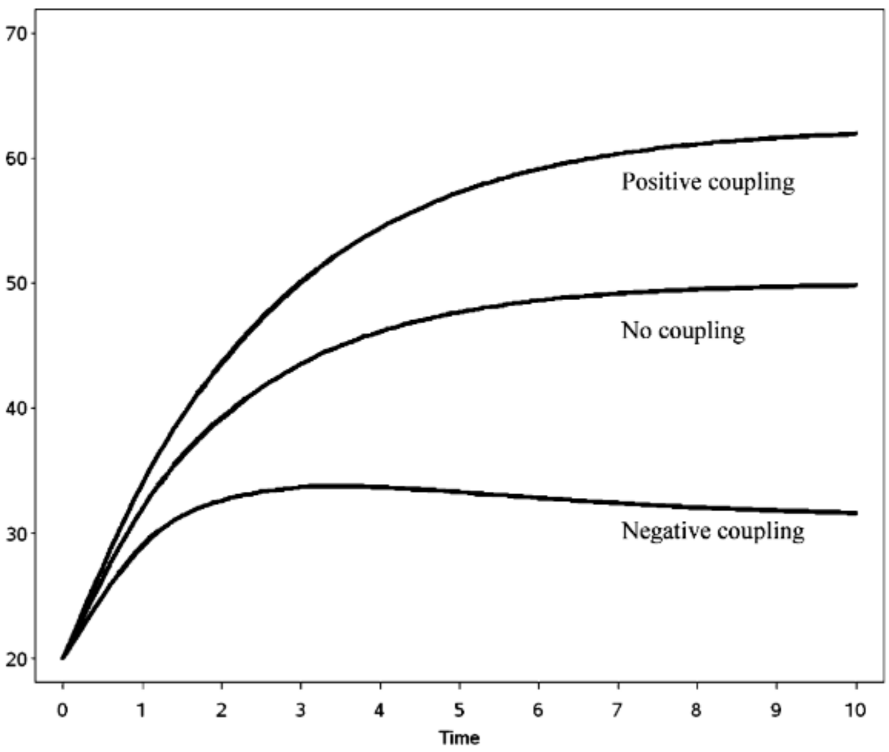
- The dual-change model allows us to model exponential trends in the data because prior status predicts amount of change
  - If $\beta = 0$, then the trajectory is linear

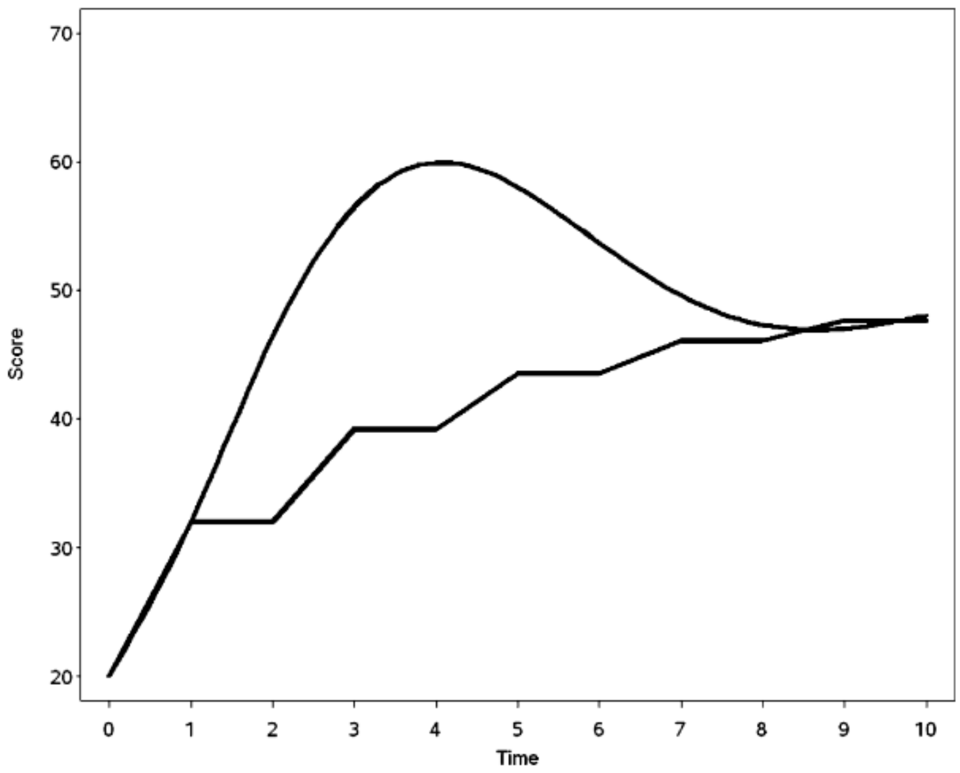




McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Fitting a Dual Change LCSM

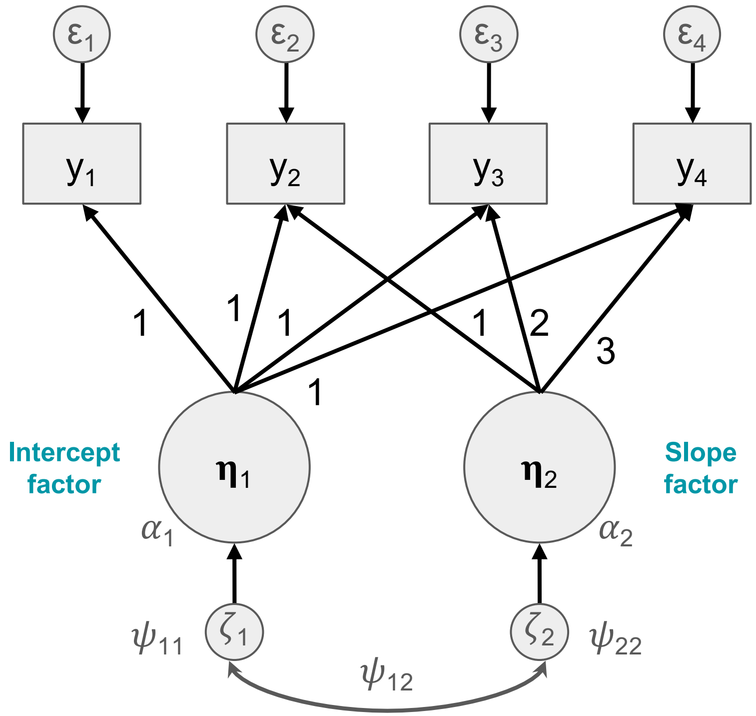




Grimm, K. J., An, Y., McArdle, J. J., Zonderman, A. B., & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(2), 268-292.

Free-Loading LCM

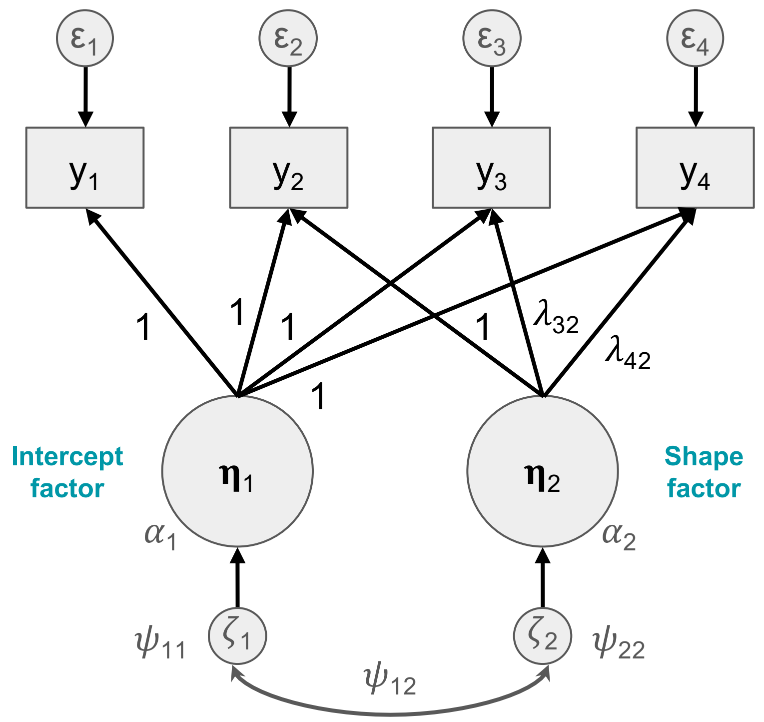
- Recall in the linear LCM we fix all the factor loadings to specified values, unlike in a CFA where factor loadings are estimated



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Free-Loading LCM

- But we can combine the strengths of the two approaches and estimate certain subsets of the factor loadings



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Free-Loading LCM

- But we can combine the strengths of the two approaches and estimate certain subsets of the factor loadings

fixed factor loading

$$\boldsymbol{\Lambda}_{free} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & \lambda_{32} \\ 1 & \lambda_{42} \\ 1 & \lambda_{52} \end{bmatrix} \qquad \boldsymbol{\Lambda}_{free'} = \begin{bmatrix} 1 & 0 \\ 1 & \lambda_{22} \\ 1 & \lambda_{32} \\ 1 & \lambda_{42} \\ 1 & 1 \end{bmatrix}$$

freely-estimated parameter

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Free-Loading LCM

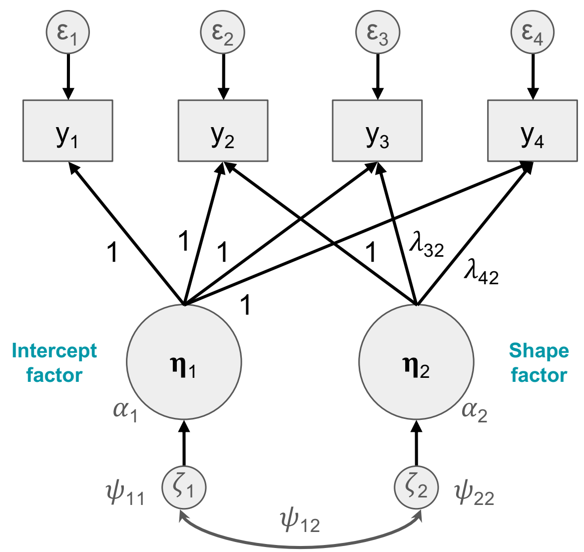
- But we can combine the strengths of the two approaches and estimate certain subsets of the factor loadings

- Allows us to have factor loadings which deviate from the linear 1-2-3-4 and
  - We can capture complex non-linear change
  - Lose the interpretability of a slope factor
    - Beta is no longer a change per unit of time
    - Talk about "degree to which individuals change like the fixed effect over time"

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Free-Loading LCM

- Two potential specifications:
  1. Set the first two factor loadings to 0 & 1, then estimate the rest
     - Change is then scaled to relative to the change that occurs between the first two time points



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

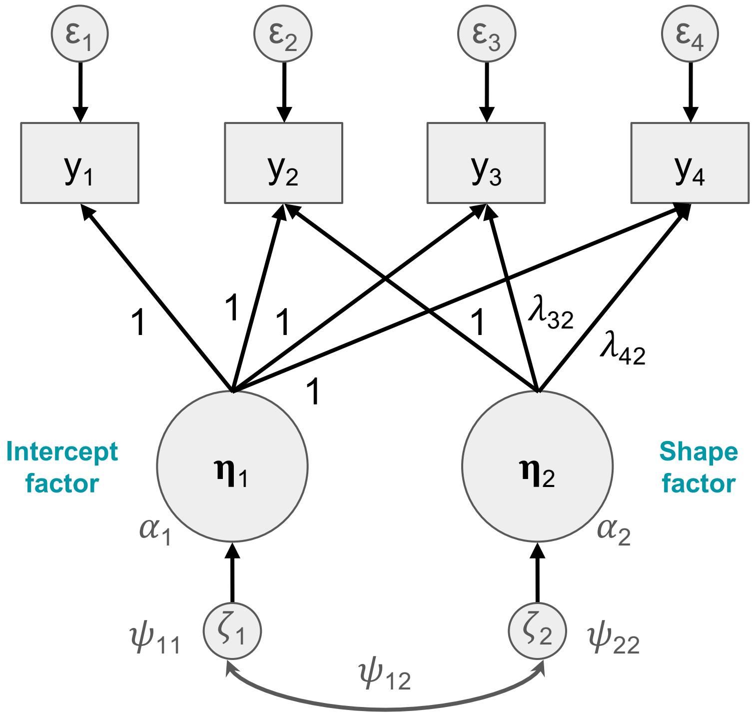
Free-Loading LCM

- Two potential specifications:
  1. Set the first two factor loadings to 0 & 1, then estimate the rest
     - Change is then scaled to relative to the change that occurs between the first two time points

  2. Set the first and **last** factor loadings to 0 & 1, then estimate the rest
     - Change is scaled to be relative to the total change that occurs across the trajectory
     - Generally less interpretable
       - Can get negative values or values > 1

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Fitting a Free-Loading LCM




```
free.load1 <- "int   =~ 1*fmin4 + 1*fmin5 + 1*fmin6 + 1*fmin7 +
                        1*fmin8 + 1*fmin9 + 1*fmin10 + 1*fmin11
               basis =~ 0*fmin4 + 1*fmin5 + l3*fmin6 + l4*fmin7 +
                        l5*fmin8 + l6*fmin9 + l7*fmin10 + l8*fmin11
"

free.load1.fit <- growth(free.load1,
                         data = adversity,
                         estimator = "ML",
                         missing = "FIML")
```

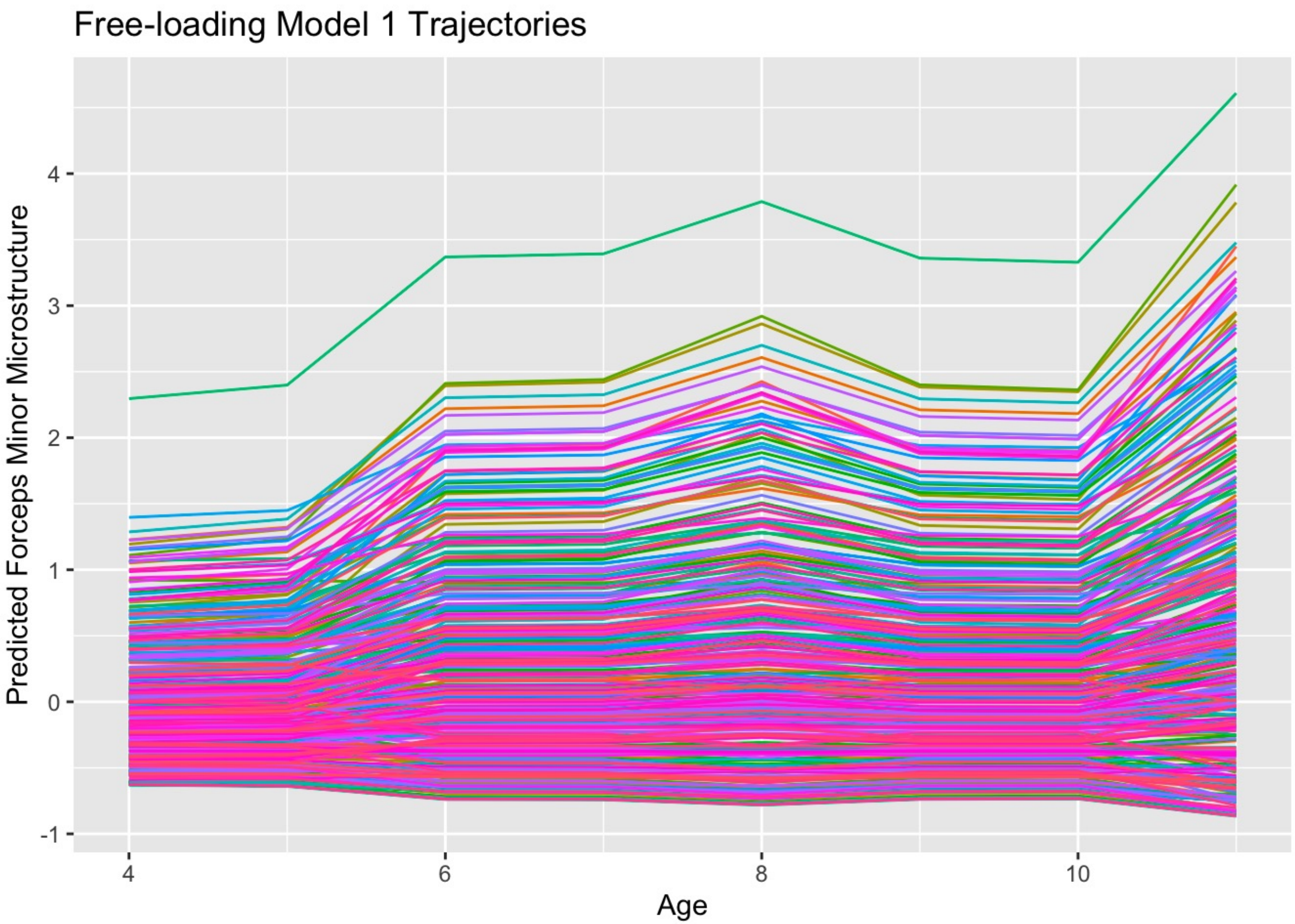

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Free-Loading LCM

```
## Latent Variables:
##                    Estimate  Std.Err  z-value  P(>|z|)   Std.lv  Std.all
##   int =~
##     fmin4             1.000                                0.600    0.591
##     fmin5             1.000                                0.600    0.649
##     fmin6             1.000                                0.600    0.549
##     fmin7             1.000                                0.600    0.517
##     fmin8             1.000                                0.600    0.468
##     fmin9             1.000                                0.600    0.511
##     fmin10            1.000                                0.600    0.531
##     fmin11            1.000                                0.600    0.431
##   basis =~
##     fmin4             0.000                                0.000    0.000
##     fmin5             1.000                                0.043    0.046
##     fmin6     (l3)   10.433   29.658    0.352    0.725    0.446    0.408
##     fmin7     (l4)   10.668   29.930    0.356    0.722    0.456    0.393
##     fmin8     (l5)   14.511   41.748    0.348    0.728    0.621    0.484
##     fmin9     (l6)   10.347   28.954    0.357    0.721    0.443    0.377
##     fmin10    (l7)   10.041   28.272    0.355    0.722    0.429    0.380
##     fmin11    (l8)   22.493   65.844    0.342    0.733    0.962    0.691
```


McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Free-Loading LCM



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Free-Loading LCM

- Challenge of overfitting
  - Clearly, we can fit a very complex trend
  - But how much noise are we picking up in our trajectory vs signal?
    - Would a line be sufficient?

- Free-loading LCMs are often good-fitting models
  - Will out-compete a linear-only model (they are greedy)
  - But (probably) won't replicate/generalize as well as a linear model across samples

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

GAMMs

- If we are looking for a method to describe a complex nonlinear trajectory, GAMMs are probably the best game in town
  - Note the use of “describe”
  - Quasi machine learning approach

- One advantage the GAMM has compared to something like the free-loading LCM is the use of regularization
  - Automated approach to reducing model complexity (# of splines, smoothness), without sacrificing fit
  - Penalization occurs during optimization

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker’s Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

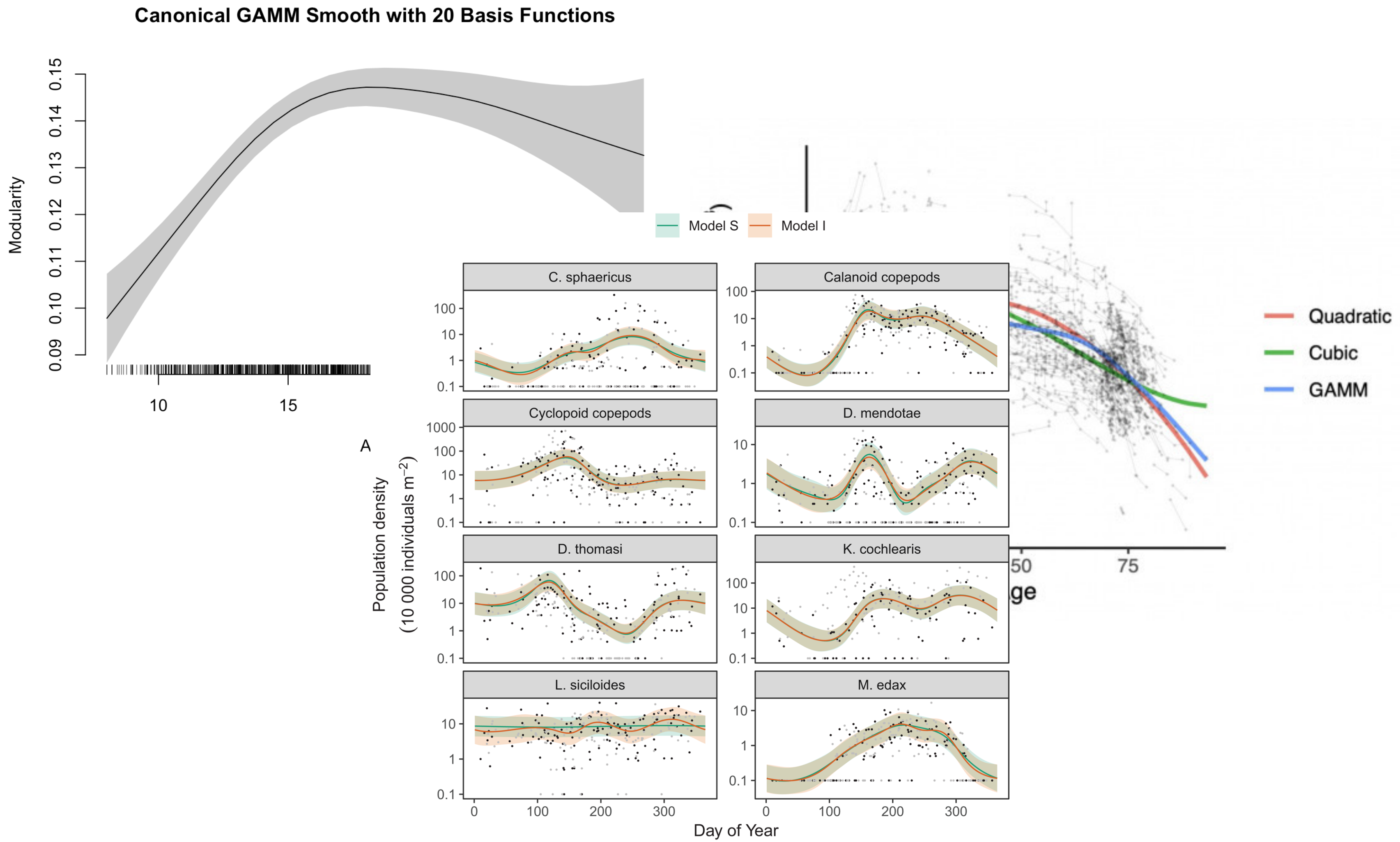
GAMMs

- Most useful with either dense observations (within-person effects) or across large age ranges (mixed effects)
  - One of the only feasible options in lifespan work
  - In longitudinal models, the overwhelming use is with mixed effects rather than purely within-person effects

- Can combine nonlinear functions with linear predictors
  - Need a wide range of predictor values to apply the function options

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

GAMMs



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Fitting a GAMM

- Amazingly simple on the user end

```r
gamm <- gamm4(scale(modularity) ~ 1 + s(age),
              random = ~ (1 | id),
              data = feedback.learning)

gamtabs(gamm$gam, type = "html",
        pnames = c("Intercept"), snames = c("s(Age)"),
        caption = "Modularity as a Function of Age")
```


McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Fitting a GAMM

- Can report model output, but not **terribly** informative in this model

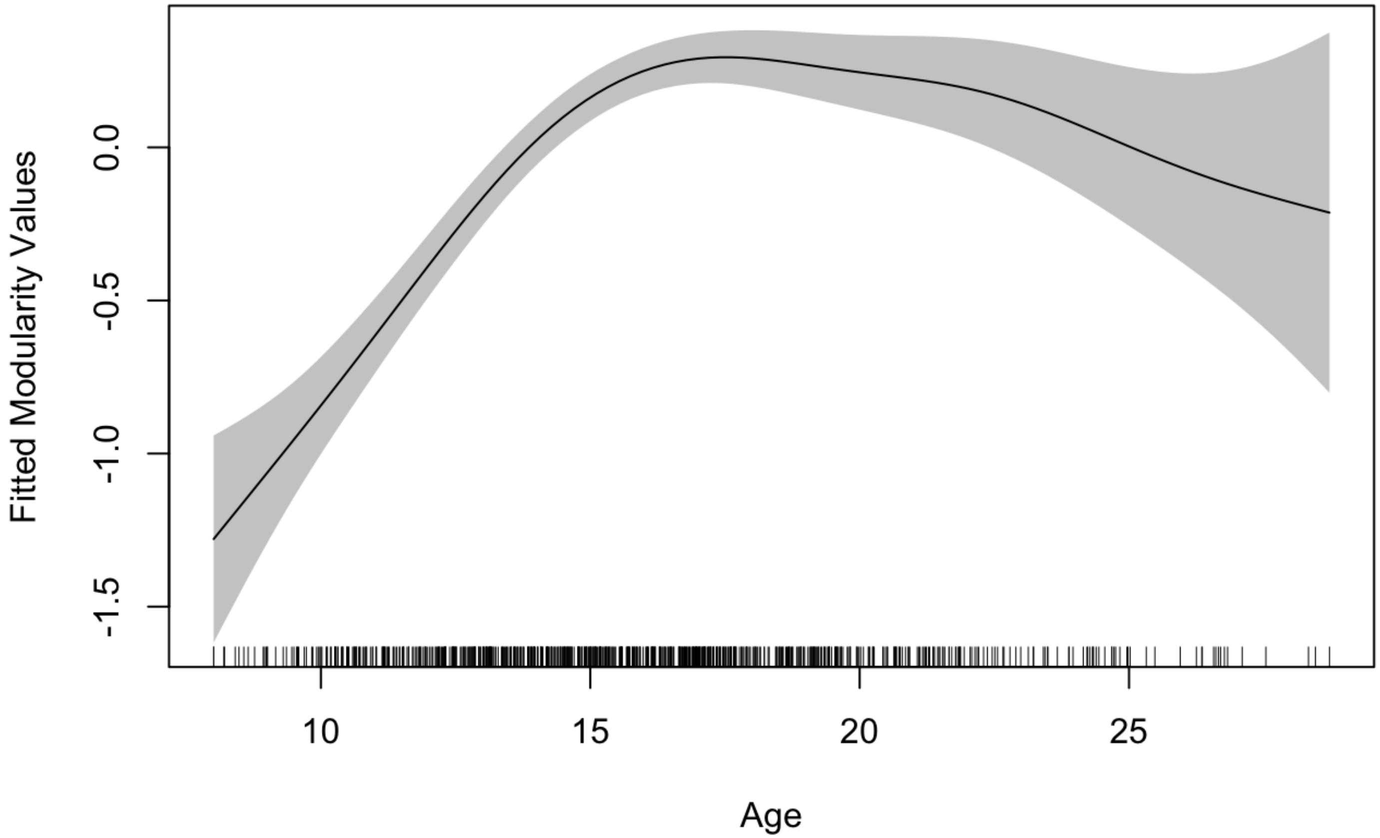
A. parametric coefficients	Estimate	Std. Error	t-value	p-value
Intercept	-0.0095	0.0464	-0.2052	0.8375
B. smooth terms	edf	Ref.df	F-value	p-value
s(Age)	4.6471	4.6471	28.1838	< 0.0001

Modularity as a Function of Age

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Fitting a GAMM

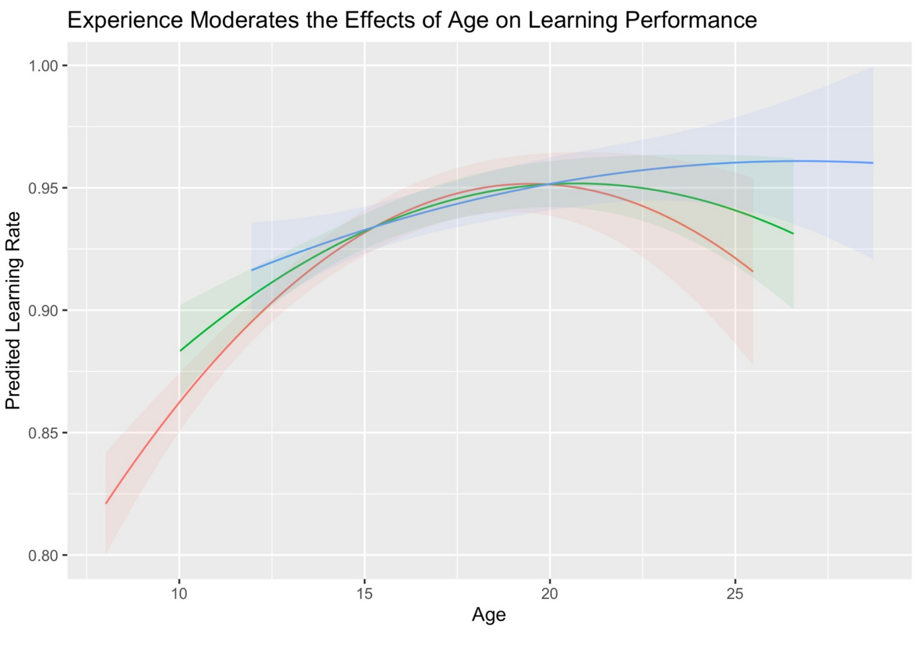
- Visualization is key with these models

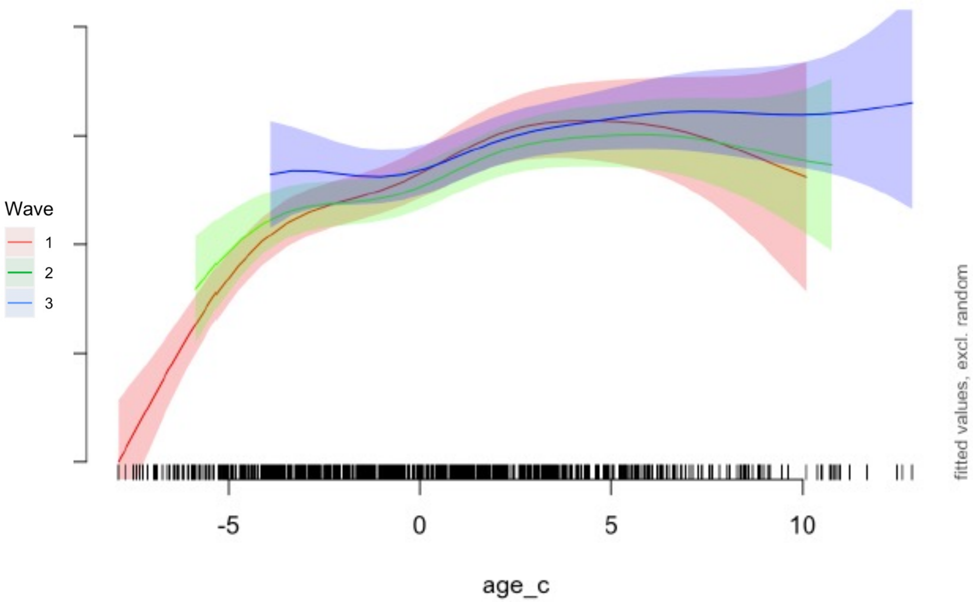


McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Fitting a GAMM

- One question might be how necessary the GAMM is versus some more constrained/interpretable model
  - Can do an informal type of sensitivity analysis
  - Would you change your discussion section?





McCormick, E. M., Peters, S., Crone, E. A., & Telzer, E. H. (2021). Longitudinal network re-organization across learning and development. *NeuroImage*, *229*, 117784.

Nonlinear Trajectory Models: Summary

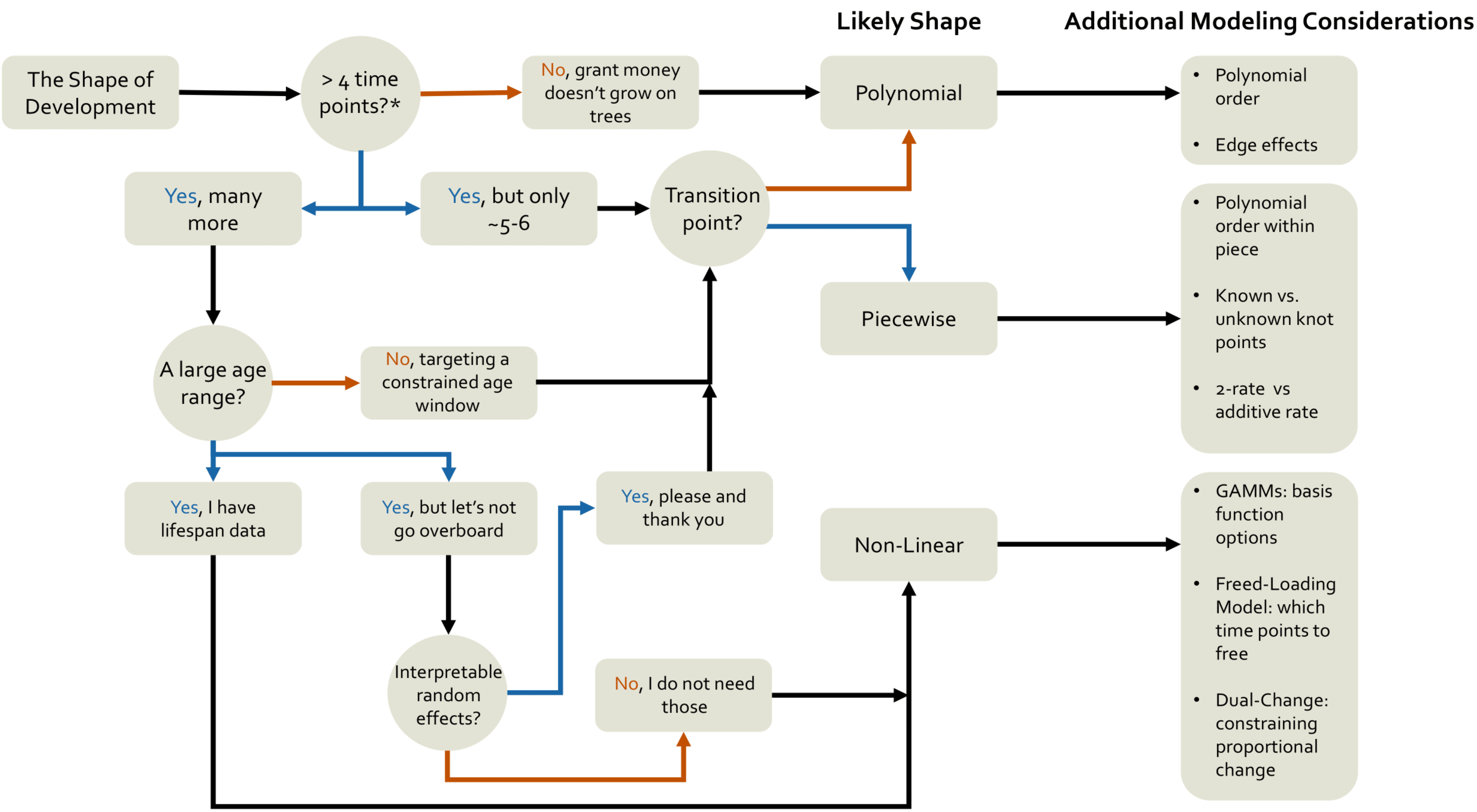
- Development is complex, and nonlinear models can help capture that complexity

- BUT: the more flexibility we give our model, the less like it it is to replicate/generalize
  - Always a balance between parsimony and explanatory power

- Enormous amount of work remains to improve the use of these models
  - Sometimes we have to make compromises, be transparent about these

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

The Shape of Change



* Do not have to be exclusively within person

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Questions?

@E_M_McCormick
e.m.mccormick@fsw.leidenuniv.nl
https://e-m-mccormick.github.io/

4. Hack-a-thon #1

Hack-a-thon #1

- Let's dig into your data and your questions
  - Briefly talk about what we are working on and the kinds of model we are trying to fit
  - Dig into the code and data

- I'll be available for questions, troubleshooting, and moral/emotional support

Questions?

@E_M_McCormick
e.m.mccormick@fsw.leidenuniv.nl
https://e-m-mccormick.github.io/

5. Covariates, Multivariate Models, and Distal Outcomes

Covariates in Longitudinal Models

- Our previous topic concerned charting the shape of developmental
  - While critically important, this tends to be more descriptive in nature
- Covariates establish (hopefully) casual explanations of <u>why</u> individuals differ over time
  - E.g.,: What is the effect of early maltreatment on emotional development?
  - Covariates can effect the whole trajectory or individual observations

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Quick Side Note on Causal Inference

- Like many causal questions, selecting the correct model is necessary, but never sufficient
  - E.g., If we mess up the shape of change, we are predicting the wrong features
  - If we include the wrong predictors, we are missing the relevant causal factors

- Far far more important is the need to bring design information to bear
  - temporal ordering, randomization, natural experiments

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

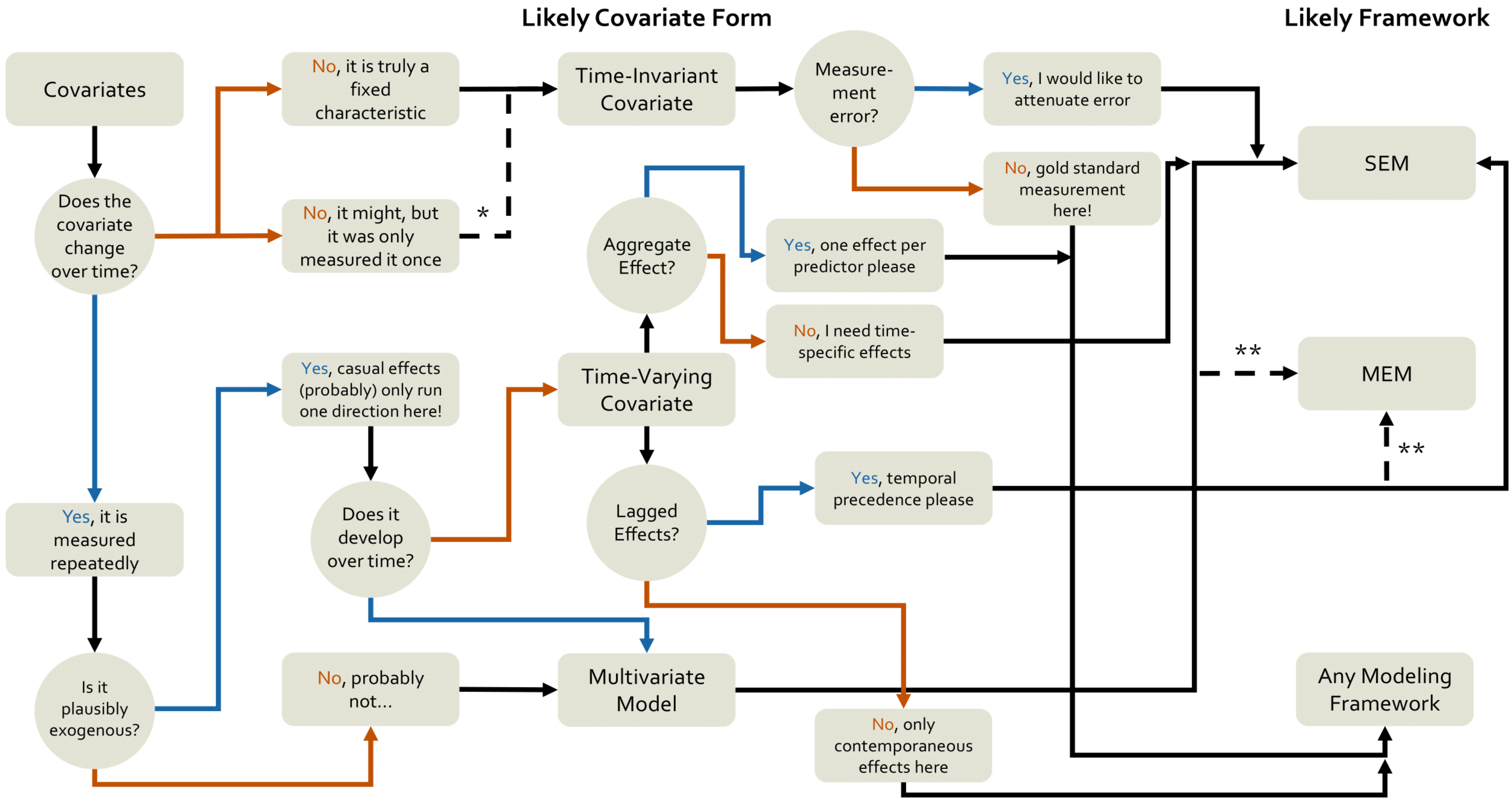
Covariates in Longitudinal Models

- Like many causal questions, selecting the correct model is necessary, but never sufficient
  - E.g., If we mess up the shape of change, we are predicting the wrong features
  - If we include the wrong predictors, we are missing the relevant causal factors

- Far far more important is the need to bring design information to bear
  - temporal ordering, randomization, natural experiments

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

The Shape of Change



* this should affect your interpretations

** a more complicated option

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Covariates in Longitudinal Models

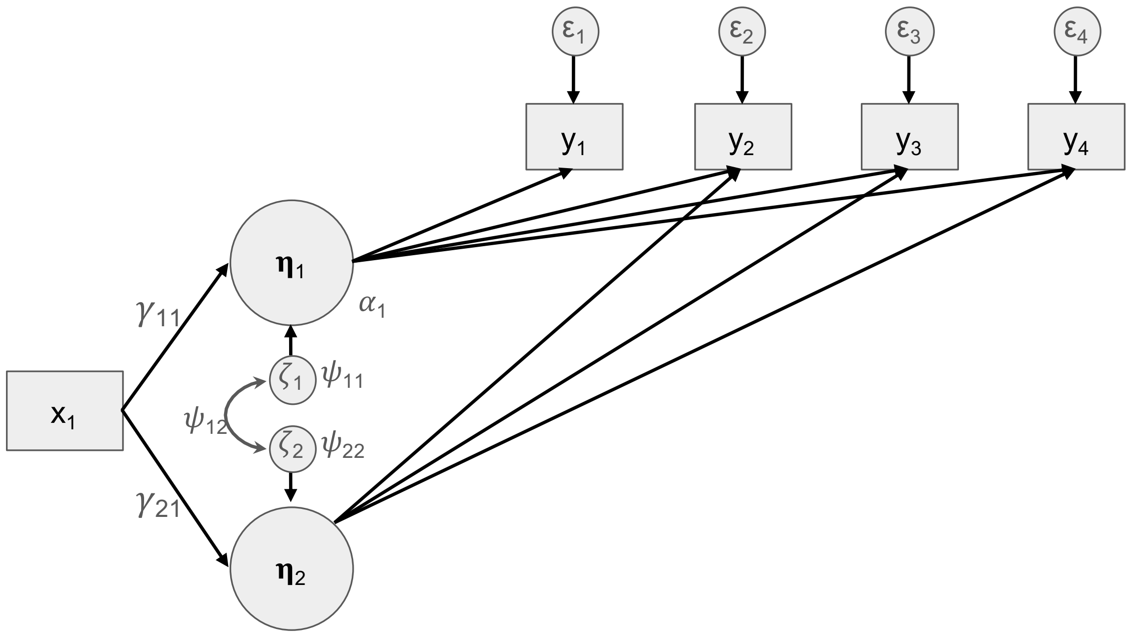
- Covariates in longitudinal models can be anything we would use in a regression context
  - No distributional assumptions
  - Can include interactions or other moderators
  - In SEMs, we can use latent factors as covariates

- Typically thought of in two categories based on where they influence the model
  - Time-invariant covariates (TICs)
  - Time-varying covariates (TVCs)

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

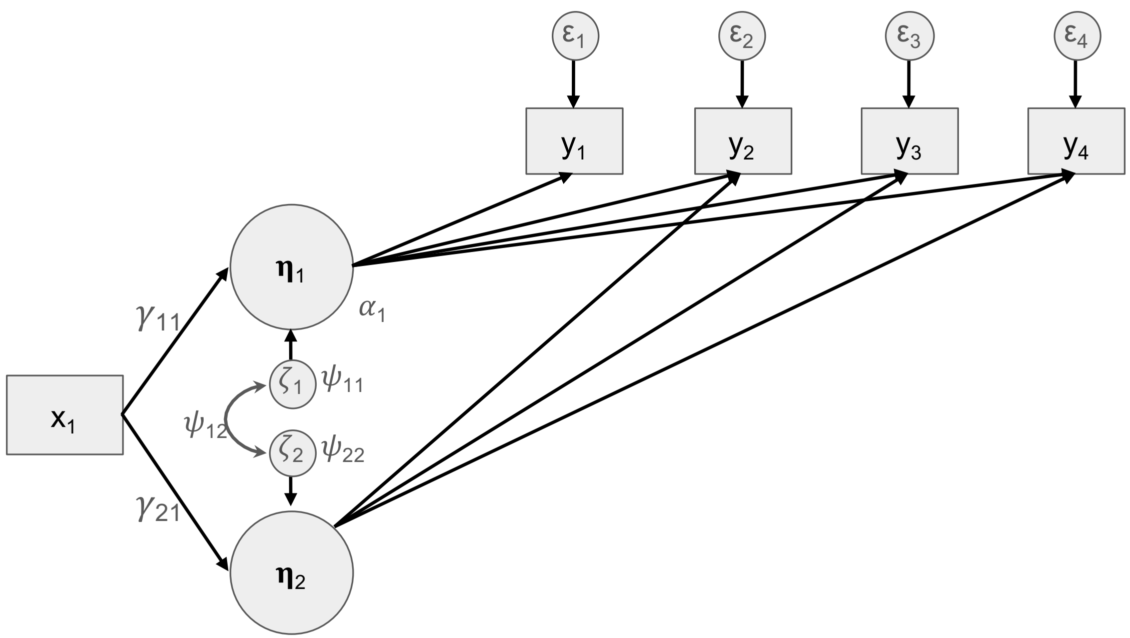
Time-Invariant Covariates

- Time-invariant covariate (TIC) models predict between-person differences in growth trajectory parameters
  - E.g., higher/lower intercepts, faster/slower growth
- <u>NO</u> temporal precedence unless imposed by design



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

- TICs only influence the repeated measures *indirectly*
  - Equivalent to an interaction if x_1 predicts the slope (a.k.a., cross-level interaction)



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Time-Invariant Covariates

- Some TICs are truly time-invariant
  - Although that number shrinks the more you think about things
  - E.g., country of origin, maternal age at first birth, early-life experiences

- Some TICs really do change over time, but we either don't measure them repeatedly, or they don't change within the developmental window under consideration
  - Need to be much more cautious about interpreting these covariates

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Time-Invariant Covariates

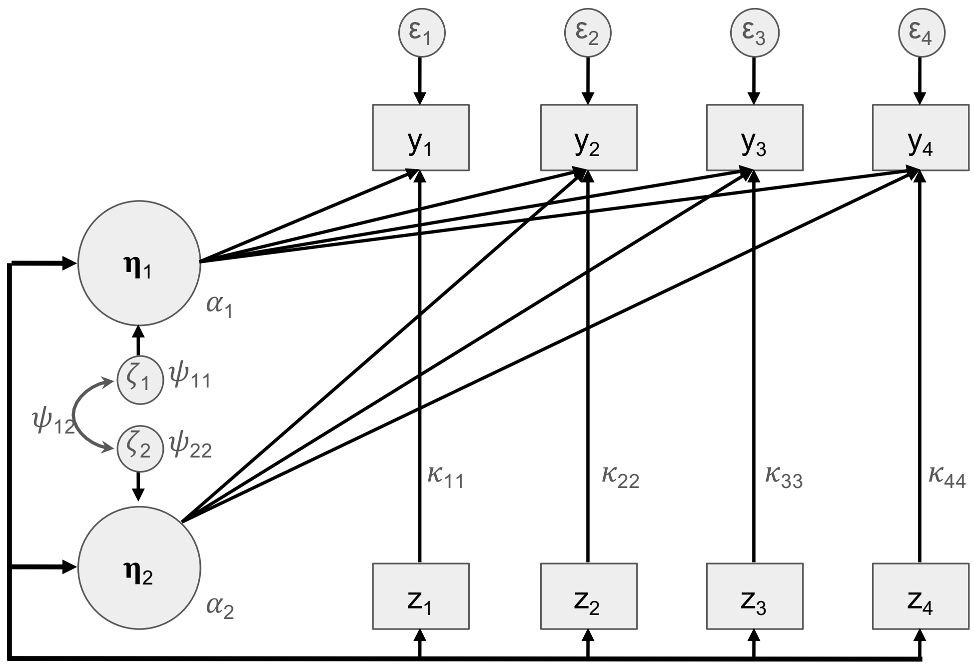
Time Invariant Covariates

We will begin with the relatively straightforward time-invariant covariate (TIC). These covariates operate at the level of the growth process: meaning that they predict the random effects (MEMs) or latent factors (SEMs). As the name implies, we will have a single value of each TIC per individual. While we often obtain that value at the first observation, in theory we could have measured that variable at any time and we should get the same value (otherwise it isn't time-invariant, is it?). One common misconception is that there is any temporal ordering inherent at this level of the model. The TIC and the intercept/slope parameters are time-invariant and so we cannot establish temporal precedence unless we bring other knowledge of the data to bear (e.g., the TIC measures something early in life and the growth parameters are fit to adolescent data). This is the same reason why regressing the slope of one outcome on the intercept of the other is very theoretically dubious unless the growth processes are truly separated in time. At this level, we essentially have cross-sectional regressions (again unless we know something extra about our variables temporally). As such, inclusion of TICs is relatively simple compared with what we will consider for the remainder of this section. We will demonstrate our examples in the MLM and LCM because of the relatively simple syntax, but the principles extend naturally to the GAMM and LCSM.

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Time-Varying Covariates

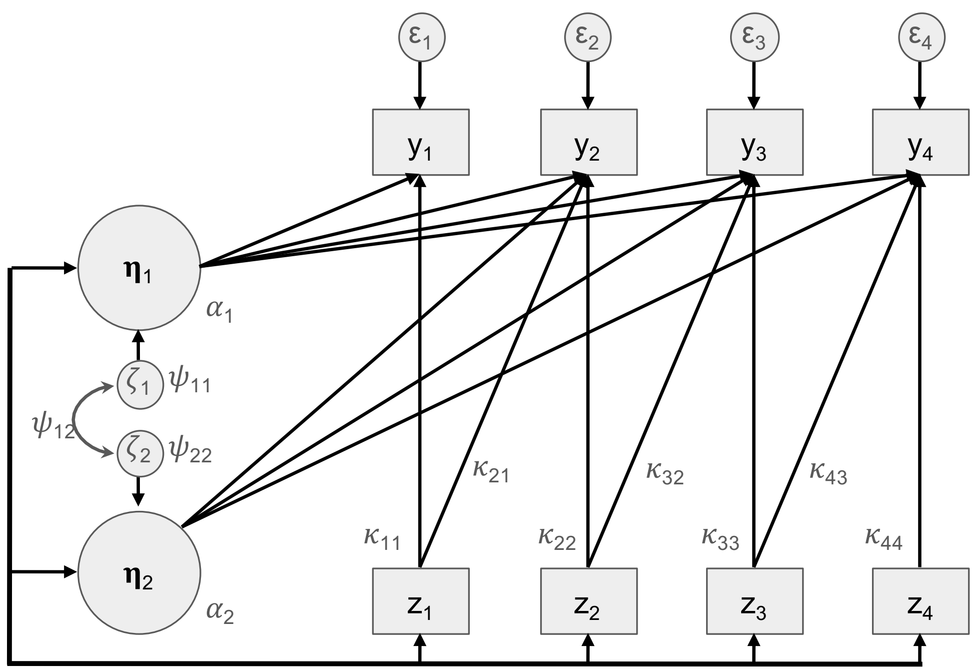
- Time-varying covariate (TVC) capture dynamic predictions on the individual repeated measures
  - E.g., time-specific deviations from overall trend
  - Temporal precedence imposed through specification



Curran, P. J., & Hancock, G. R. (2021). The challenge of modeling co-developmental processes over time. *Child Development Perspectives*, 15(2), 67-75.

Time-Varying Covariates

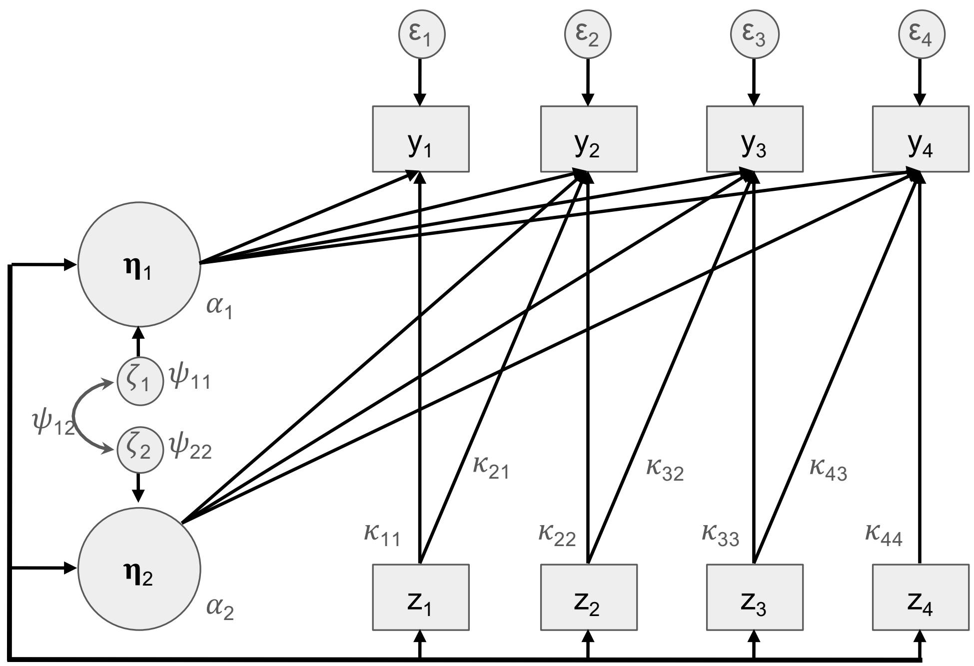
- Time-varying covariate (TVC) capture dynamic predictions on the individual repeated measures
  - E.g., time-specific deviations from overall trend
  - Temporal precedence imposed through specification



Curran, P. J., & Hancock, G. R. (2021). The challenge of modeling co-developmental processes over time. *Child Development Perspectives*, 15(2), 67-75.

Time-Varying Covariates

- Time-varying covariate (TVC) capture dynamic predictions on the individual repeated measures
  - Relationships can be random effects (typically MLMs) or time-specific (typically SEMs)



Curran, P. J., & Hancock, G. R. (2021). The challenge of modeling co-developmental processes over time. *Child Development Perspectives*, 15(2), 67-75.

Time Varying Covariates

While TICs can provide important insights into what factors might contribute to individual differences in the growth trajectories, they fundamentally represent cross-sectional regressions of the latent factors on a set of exogenous variables. However, if we want to test the effects of variables that themselves change over time, we can instead move the TVC model. While TICs predict the growth components (and therefore the repeated measures *indirectly*; hence cross-level interactions), a TVC directly predicts the individual repeated observations. In this section, we will return to our `executive.function` dataset. Here we can model the impact of DLPFC activation on executive function scores. As with the TIC model, we will first consider the MLM specification of this approach.

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

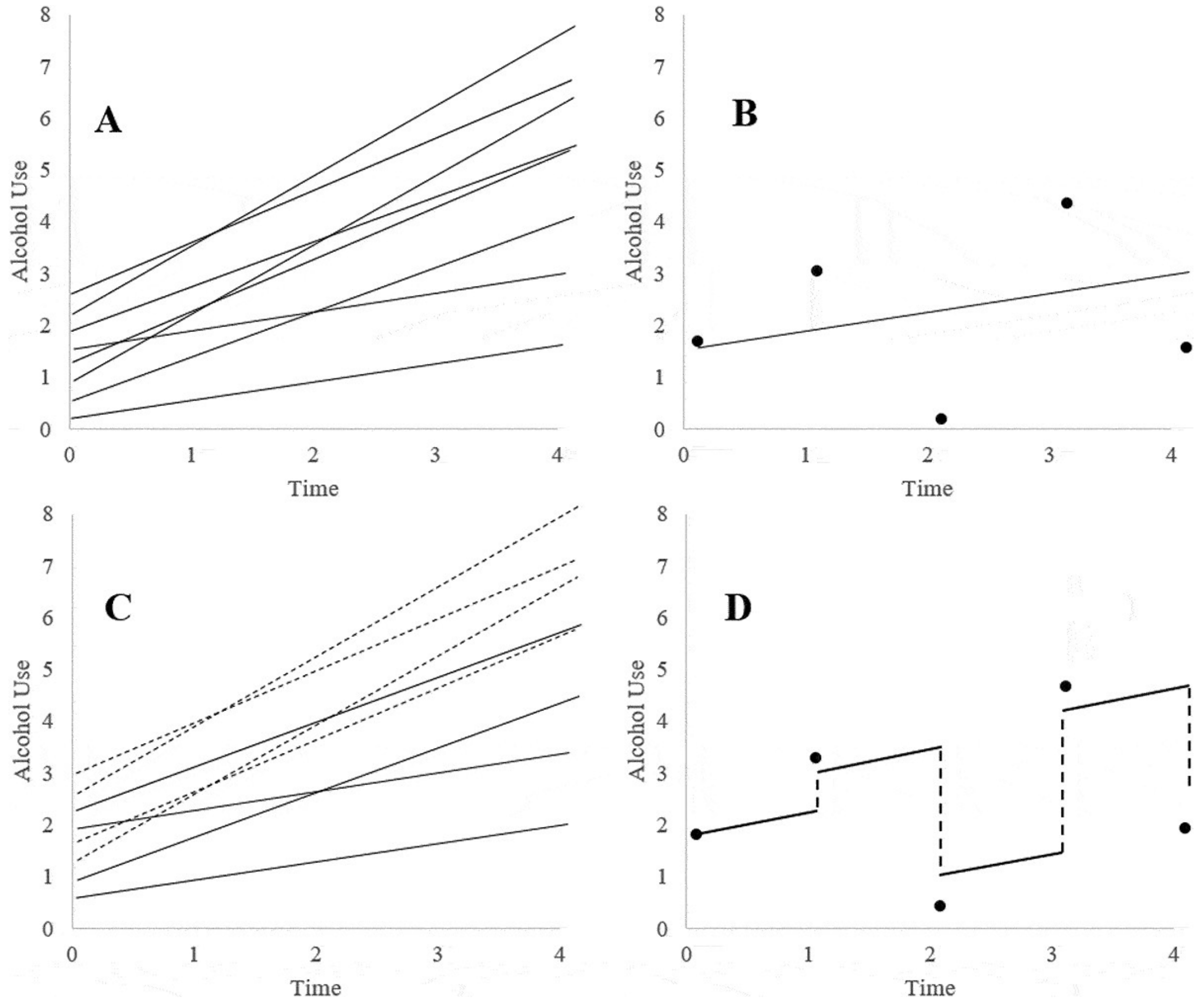
Within- and Between-Person Variance

- TICs explain only between-person differences
  - Covariates and growth parameters are characteristics of the person and are therefore invariant across time

- With TVCs, we want to only explain within-person differences

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Within- and Between-Person Variance



McNeish, D., & Matta, T. H. (2020). Flexible treatment of time-varying covariates with time unstructured data. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(2), 298-317.

Within- and Between-Person Variance

- TICs explain only between-person differences
  - Covariates and growth parameters are characteristics of the person and are therefore invariant across time

- With TVCs, we want to only explain within-person differences
  - **But** TVCs also contain between-person information (i.e., the means of the TVCs over time)
  - Need to use centered versions of the TVC to achieve pure within-person estimates

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

Within- and Between-Person Variance

Within- and Between-Person Variance

In longitudinal models, we are deeply interested in understanding within-person processes (i.e., how individuals change over time). However, we need to be on guard against *between*-person differences masquerading as these within-person effects. For instance, there is a substantively different understanding of adolescent substance use if it is guided by within-person effects (e.g., adolescent are more likely to use illicit drugs when they are with their peers) versus between-person effects (e.g., adolescents with more friends are more likely to use illicit drugs). Unfortunately, we can often unintentionally conflate these two types of effects in our models. This happens because variables at level 1 (e.g., TVCs) contain information about both within- (level 1) and between- (level 2) person differences. To take our example thus far, individuals might show higher or lower DLPFC activation relative to the last time they played the task (i.e., within-person differences) but also some individuals might consistently show higher (or lower) than average DLPFC activation relative to other individuals in the sample (i.e., between-person differences). We can return to our contemporaneous TVC model to demonstrate this principle. Below is the model we fit before.

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

TVC versus Multivariate Models

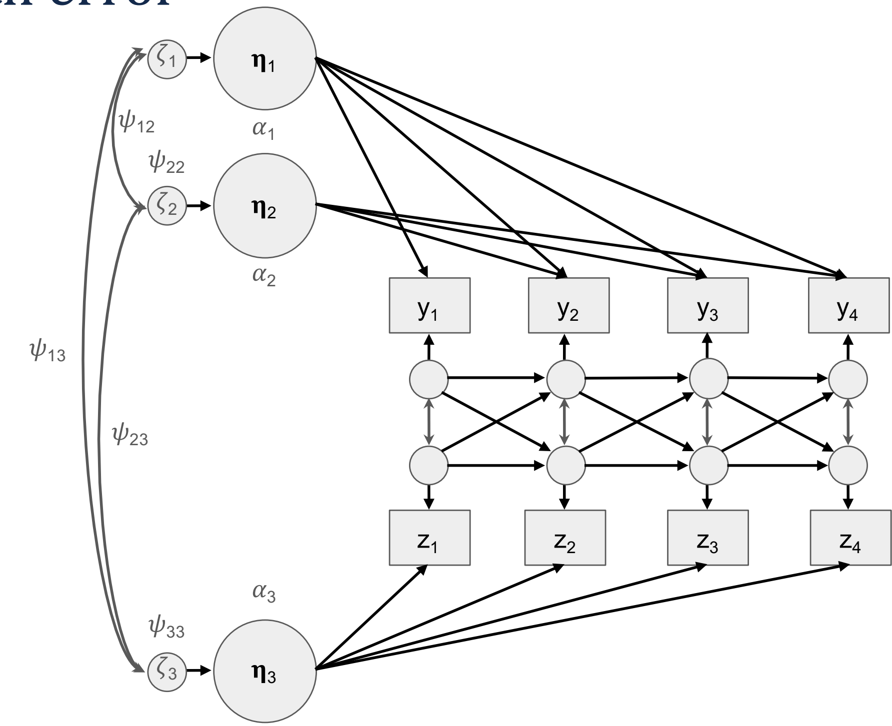
- Time-varying Covariates
  - TVCs are exogenous to the model
    - Only one causal direction
    - TVCs are treated as “fixed and known”
  - TVC does not systematically change

- Does this truly describe many variables we tend to encounter?

McNeish, D., & Matta, T. H. (2020). Flexible treatment of time-varying covariates with time unstructured data. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(2), 298-317.

Multivariate Models

- Multivariate Models
  - Both variables are endogenous to the model
    - Reciprocal relationships over time
    - TVC assumed measured with error
  - Both variables changing*
    - Different functional forms
  - Real limitations for MEMs
    - SEMs predominate

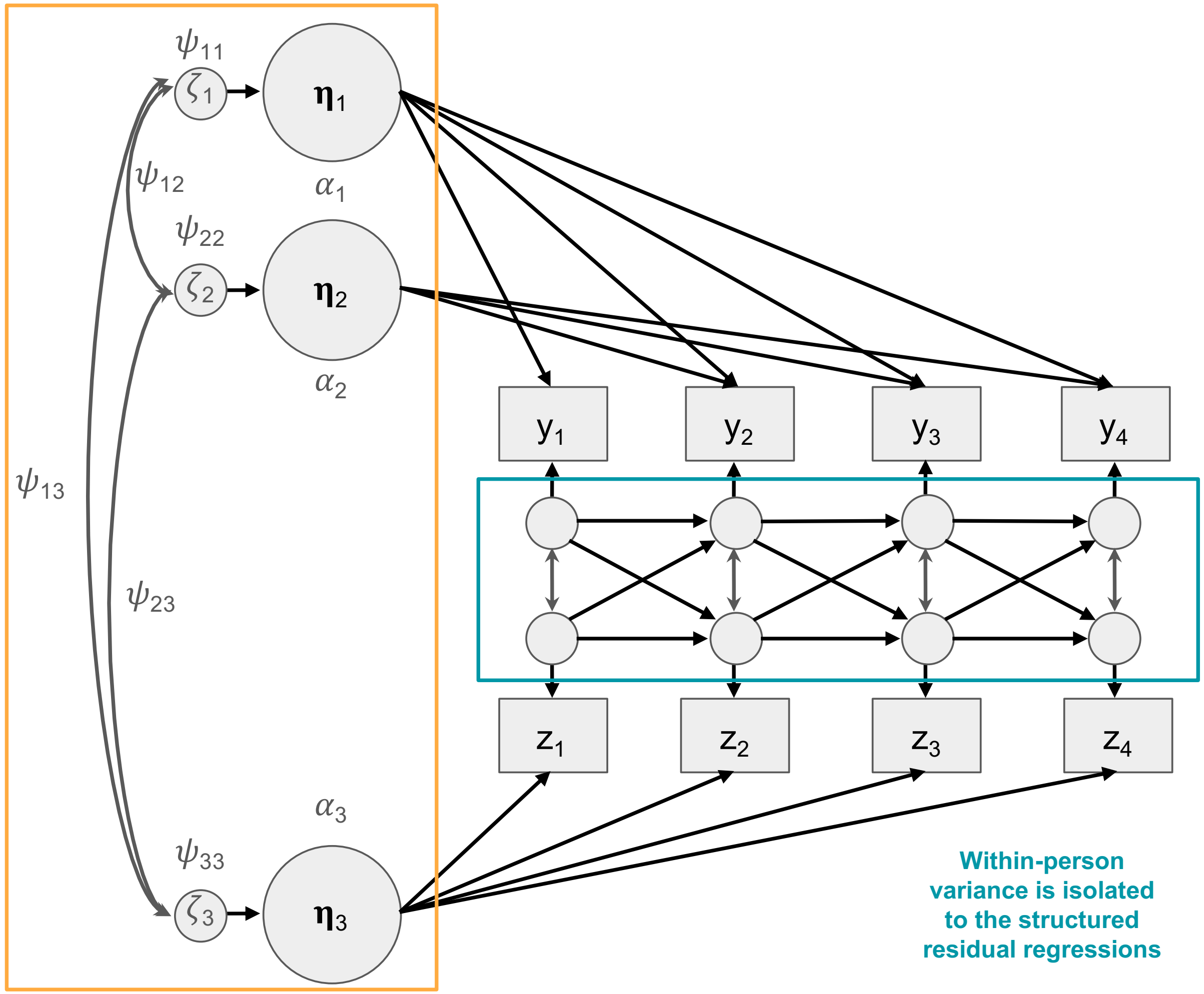


Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: a latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology, 82*(5), 879.

- Multivariate Models
  - Allow us to look at reciprocal coupling
  - Can impose constraints to see if that coupling increase/decreases over time
  - Within- and between-person variance separation

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: a latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology, 82*(5), 879.

Multivariate Models



Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: a latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology, 82*(5), 879.

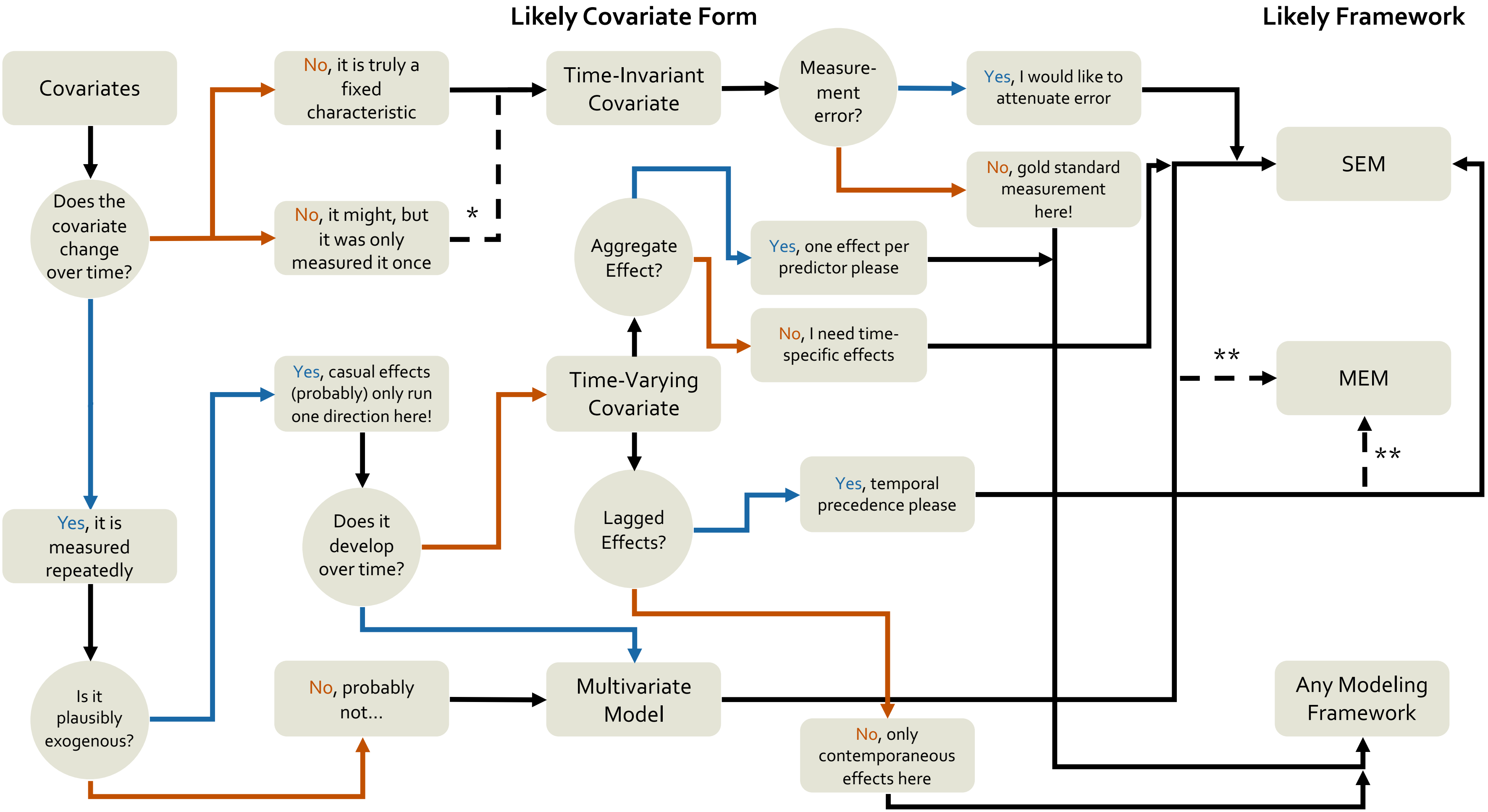
Covariates and Multivariate Models

- Very little in the models themselves allows us to establish that our assumptions are met
  - TICs are just cross-sectional regressions*
  - TVC model makes sense if temperature is our covariate, but what about depression?

- Dynamic relationships are challenging, and we might need to combine long-term longitudinal models with short-term intensive longitudinal models to better understand them

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Covariates and Multivariate Models



* this should affect your interpretations ** a more complicated option

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Distal Outcomes

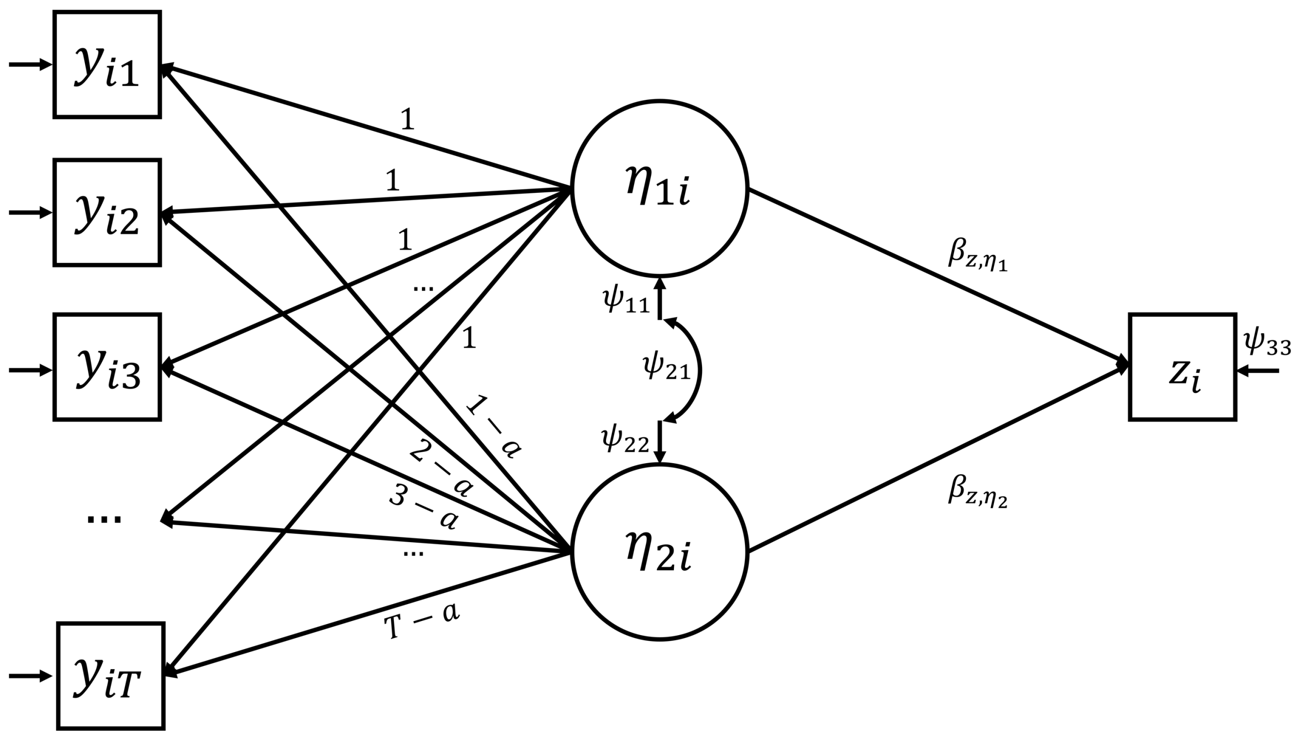
- Covariates and multivariate models are concerned with predicting individual differences, but we can also use growth models to predict future variables of interest
  - E.g., chronic health problems, survival rates, incarceration

- This is what I think of as the "*so what?*" of longitudinal models

- Purely predictive and explanatory approaches possible

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

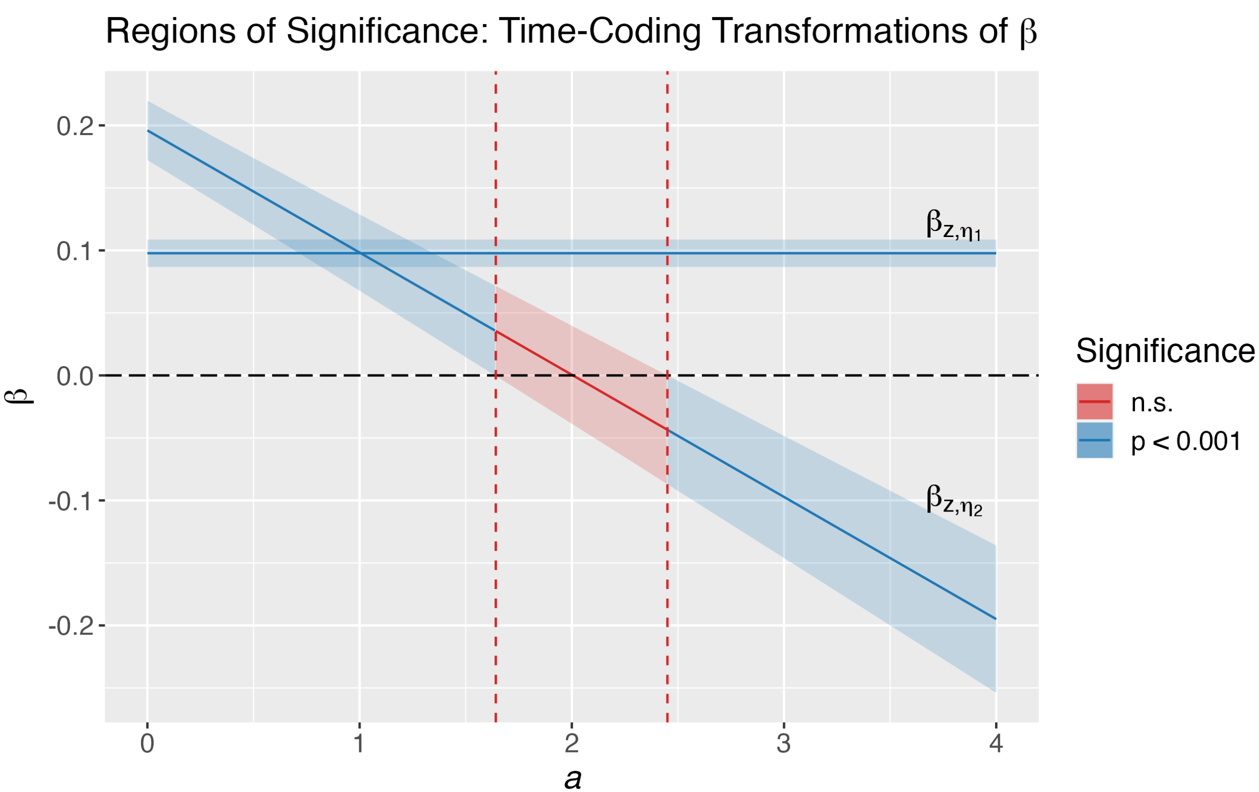
Distal Outcomes



McCormick, E.M., Curran, P.J., & Hancock, G.R. (*in prep*). Latent Growth Factors as Predictors of Distal Outcomes: Completing the Triad.

Distal Outcomes

- Area of active research
  - Some additional complications for time-coding



McCormick, E.M., Curran, P.J., & Hancock, G.R. (*in prep*). Latent Growth Factors as Predictors of Distal Outcomes: Completing the Triad.

Distal Outcomes

- Area of active research
  - Some additional complications for time-coding
  - Predicting from the intercept is generally of less interest, but is the invariant parameter
- Mixed Effects models generally take a 2-step approach
  - 1) estimate growth model, 2) extract parameters and use them in a subsequent analysis
- SEMs can do a 1- or 2-step approach
  - 1-step is generally better

McCormick, E.M., Curran, P.J., & Hancock, G.R. (*in prep*). Latent Growth Factors as Predictors of Distal Outcomes: Completing the Triad.

Distal Outcomes

- Despite the incredible importance of long-term outcomes for validating theories and contextualizing development, few studies include truly distal outcomes
  - Cost and effort, lack of theoretical clarity
- Balancing interpretability and predictive validity is key
  - Growth features provide explanation for predictive relationships
  - Might want more meaningful model features
- More soon!

McNeish, D., & Matta, T. H. (2020). Flexible treatment of time-varying covariates with time unstructured data. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(2), 298-317.

Questions?

@E_M_McCormick
e.m.mccormick@fsw.leidenuniv.nl
https://e-m-mccormick.github.io/

6. Nested Data

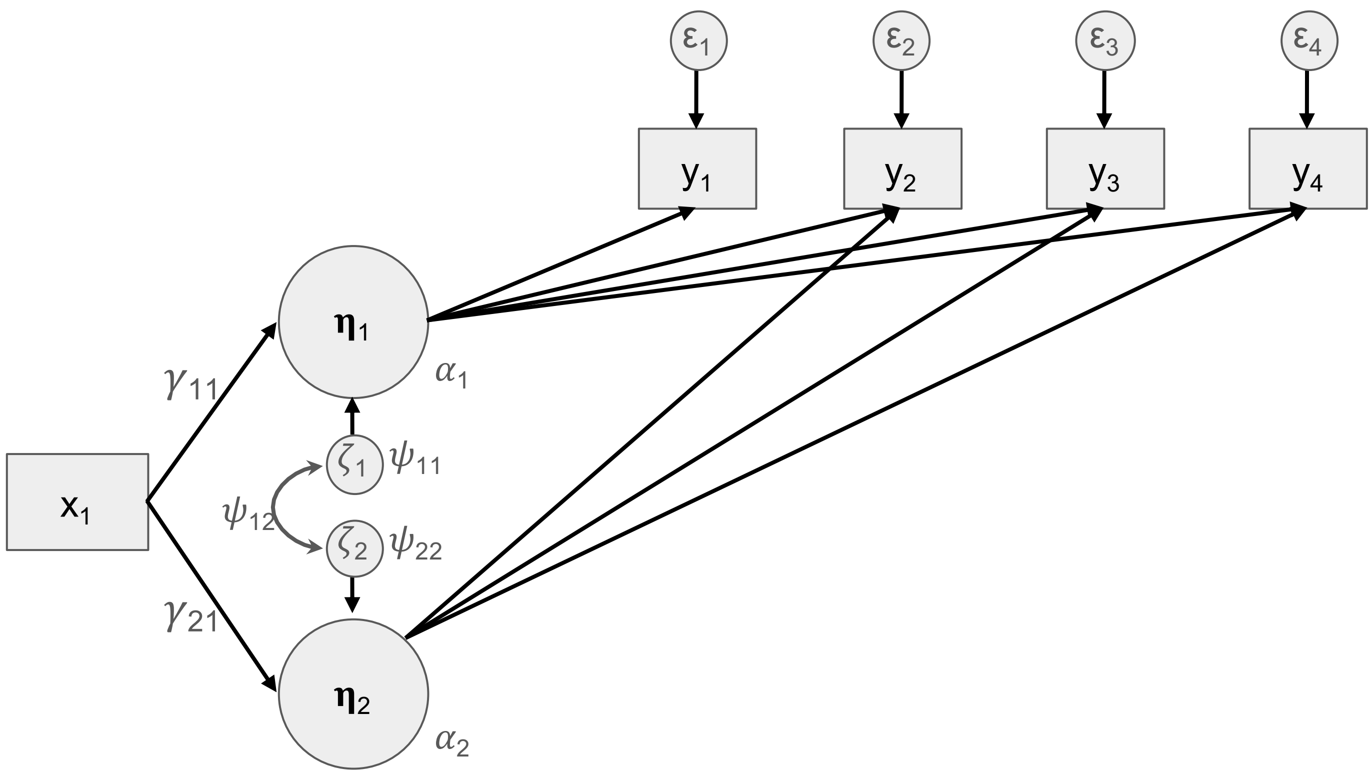
Nested Data

- All longitudinal data represents nesting of repeated measures nested within person
  - Easier to see in the mixed-effect models, but the SEMs are equivalent ways to deal with this nesting
- Here we will talk about models which account for additional forms of nesting
  - E.g., individuals within classrooms or families
  - Essentially accounting for additional dependencies in the data

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data

- Nominal covariates



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data

- Nominal covariates
  - Members of the same group have more similar model parameters
  - Allows us to account for additional heterogeneity in the model

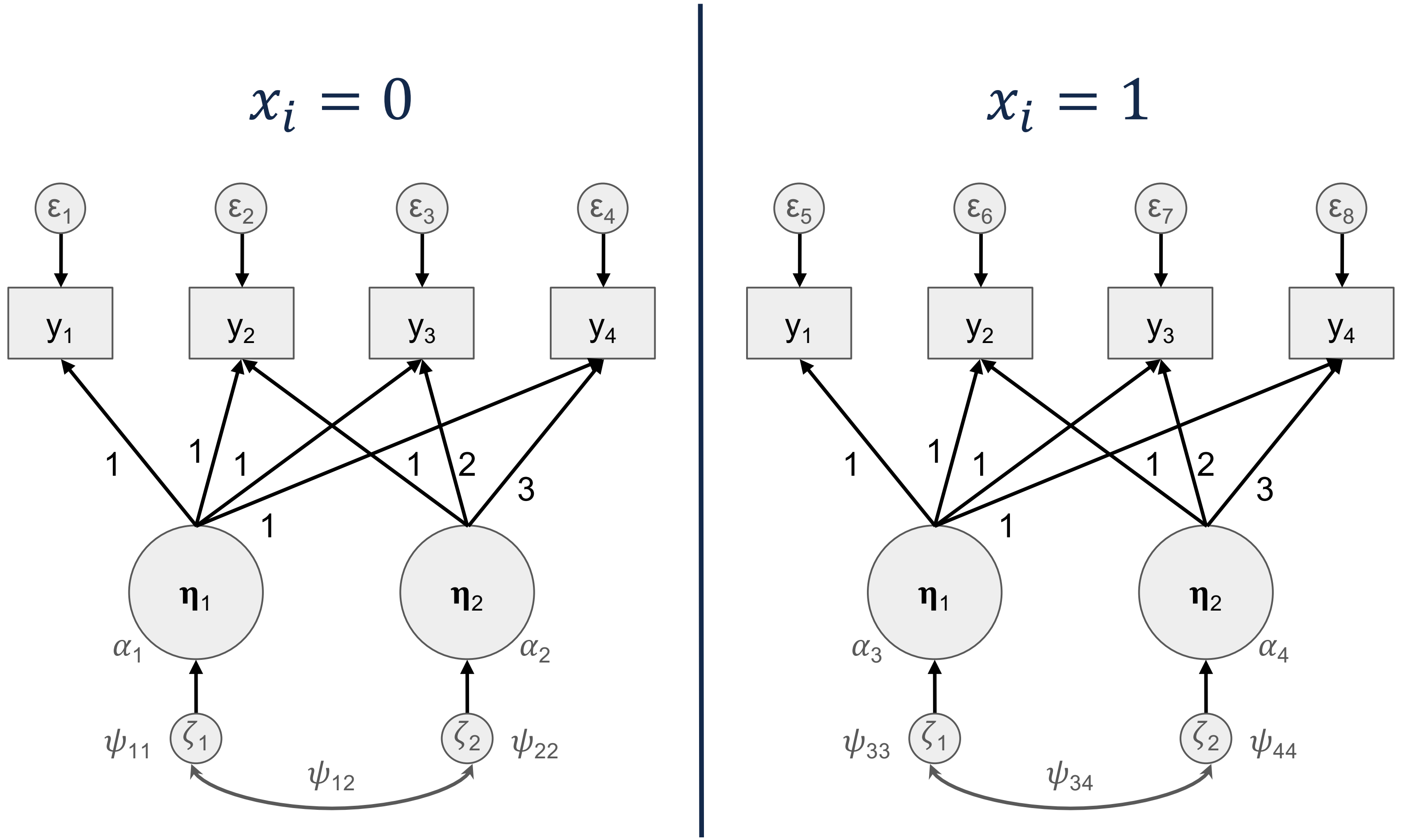
- Primarily used with binary groups

- Can be extended to multiple groups or more-general parameter moderation models

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data

- Multiple Groups Models



McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data: Fixed and Random Effect Models

- More traditional forms of nesting
  - Often not a predictor of interest per se, but rather a consequence of sampling design

- Two ways to conceptualize these nesting units
  - A set of discrete groupings
    - Fixed effect approach
    - E.g., country, religious group, data collection site
  - A random sample from a population of possible units
    - Random effects approach
    - E.g., families, classrooms, individuals
    - Requires distributional assumptions

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data: Fixed and Random Effect Models

- Fixed-Effects Models
  - When we have a set of discrete units that are not a random sample from a possible population, we can instead treat groups as a sequence of fixed effects
  - Removes *all* variance associated with differences in the mean of the outcome related to group
    - No matter what causes them to differ

$$y_{ti} = \beta_1 Cntrl + \beta_2 Treat_1 + \beta_3 Treat_2 + r_{ti}$$

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data: Fixed and Random Effect Models

- Fixed-Effects Models
  - If we think the effect of time (or other covariate) differs across groups, we can also remove these effects with interactions
    - More often only remove intercept effects
    - Models expand rapidly, but does cleanly isolate the group-specific effects

$$y_{ti} = \beta_1 Cntrl + \beta_2 (Cntrl \times Time) + \beta_3 Treat_1 + \beta_4 (Treat_1 \times Time) +$$
$$\beta_5 Treat_2 + \beta_6 (Treat_2 \times Time) + r_{ti}$$

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data: Fixed and Random Effect Models

- Fixed-Effects Models
  - Advantages
    - Does not require distributional assumptions
    - Can index the entire population of nesting units
    - Differences across groups are controlled for
  - Disadvantages
    - Models get messy fast
    - Cannot predict group heterogeneity

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data: Fixed and Random Effect Models

- Random-Effects Models
  - When we want to characterize continuous differences from a population of groups, we can specify group as a higher-order random effect
    - Groups we observe are assumed to be a random sample of possible groups
    - We are already doing this at the individual level when we use a mixed-effect or structural equation longitudinal model
  - Instead of estimating unique group effects, we estimate the variance of effects

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data: Fixed and Random Effect Models

- Random-Effects Models
  - When we want to characterize continuous differences from a population of groups, we can specify group as a higher-order random effect
    - Groups we observe are assumed to be a random sample of possible groups
    - We are already doing this at the individual level when we use a mixed-effect or structural equation longitudinal model
  - Instead of estimating unique group effects, we estimate the variance of effects

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data: Fixed and Random Effect Models

- Random-Effects Models
  - Advantages
    - Can accommodate *many* more groups
    - Relatively parsimonious by estimating variances
    - It matches the longitudinal model approach
  - Disadvantages
    - How often to do we have *many* higher-level groups?
    - Higher-order variances are very hard to estimate and likely pretty unreliable
    - Do we have a hypothesis at the higher level?
    - Normality assumption

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data: Modeling versus Correction

- Un-modeled clustering does not tend to bias point estimates
  - It is the standard errors (i.e., inferences) that are inappropriately small

- As such, if we do not want to test a higher-order cluster-level hypothesis, we could just fit the un-clustered model and correct for cluster membership on the standard errors
  - A relatively elegant solution
  - Avoids fitting an unstable or possibly mis-specified random effect

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

Nested Data

The final topic we will consider is the way we account for nesting within longitudinal data. While technically all of longitudinal modeling involves nested data (i.e., multiple observations nested within the same individual), here we focus more on between-person nesting groups (e.g., classroom, family, data collection sites). As we will see, there are several forms of the models that we have already used that deal with nesting. The `achieve` dataset we will work with here are from a 4-wave school-based assessment of math and science achievement across ages 12 - 17. Schools were drawn from 5 different metropolitan areas and assessments were conducted by separate research teams in each city. Here we will mostly focus on the science achievement data, but may draw examples from math achievement if they differ in interesting ways.

McCormick, E. M., Byrne, M. L., Flournoy, J. C., Mills, K. L., & Pfeifer, J. H. (2023, January 25). The Hitchhiker's Guide to Longitudinal Models: A Primer on Model Selection for Repeated-Measures Methods. https://doi.org/10.31234/osf.io/ga4qz

7. Workshop Review

Workshop Review

- We have covered an amount of of material in 2 days
  - If you don't feel like an expert yet, don't be discouraged!
  - Hopefully your brain is full and ready to digest

- We are capable of running an analysis before we fully understand it (everyone is like this)
  - Iterating between theory and practice will help solidify

- Keep me in the loop
  - Always enjoy hearing how you will apply these models
  - Tell your friends!

Thank you for your time and attention!

Questions?

@ E_M_McCormick

e.m.mccormick@fsw.leidenuniv.nl

https://e-m-mccormick.github.io/

8. Hack-a-thon #2

Hack-a-thon #2

- Let's dig into your data and your questions
  - Dig into the code and data

- I'll be available for questions, troubleshooting, and moral/emotional support