Review of Multilevel Models for Longitudinal Data

- Topics:
  - **Concepts in longitudinal multilevel modeling**
  - Describing within-person fluctuation using ACS models
  - Describing within-person change using random effects
  - Likelihood estimation in random effects models
  - Describing nonlinear patterns of change
  - Time-invariant predictors

What is a Multilevel Model (MLM)?

- Same as other terms you have heard of:
  - **General Linear Mixed Model** (if you are from statistics)
    - *Mixed* = Fixed and Random effects
  - **Random Coefficients Model** (also if you are from statistics)
    - Random coefficients = Random effects
  - **Hierarchical Linear Model** (if you are from education)
    - Not the same as hierarchical regression

- Special cases of MLM:
  - Random Effects ANOVA or Repeated Measures ANOVA
  - (Latent) Growth Curve Model (where "Latent" → SEM)
  - Within-Person Fluctuation Model (e.g., for daily diary data)
  - Clustered/Nested Observations Model (e.g., for kids in schools)
  - Cross-Classified Models (e.g., "value-added" models)

The Two Sides of Any Model

- **Model for the Means:**
  - *Aka* **Fixed Effects**, Structural Part of Model
  - What you are used to **caring about for testing hypotheses**
  - How the expected outcome for a given observation varies as a function of values on predictor variables

- **Model for the Variance:**
  - *Aka* **Random Effects and Residuals**, Stochastic Part of Model
  - What you are used to **making assumptions about** instead
  - How residuals are distributed and related across observations (persons, groups, time, etc.) → these relationships are called "dependency" and ***this is the primary way that multilevel models differ from general linear models*** *(e.g., regression)*

Review: Variances and Covariances

Variance:
Dispersion of y

$$\text{Variance}(y_t) = \frac{\sum_{i=1}^{N}(y_{ti} - \hat{y}_{ti})^2}{N - k}$$

Covariance:
How y's go together, unstandardized

$$\text{Covariance}(y_1, y_2) = \frac{\sum_{i=1}^{N}(y_{1i} - \hat{y}_{1i})(y_{2i} - \hat{y}_{2i})}{N - k}$$

Correlation:
How y's go together, standardized (−1 to 1)

$$\text{Correlation}(y_1, y_2) = \frac{\text{Covariance}(y_1, y_2)}{\sqrt{\text{Variance}(y_1)} * \sqrt{\text{Variance}(y_2)}}$$

N = # people, t = time, i = person
k = # fixed effects, $\hat{y}_{ti}$ = y predicted from fixed effects

Dimensions for Organizing Models

- Outcome type: General (normal) vs. General*ized* (not normal)
- Dimensions of sampling: One (so one variance term per outcome) vs. **Multiple** (so multiple variance terms per outcome) → **OUR WORLD**

- **General Linear Models:** conditionally normal outcome distribution, **fixed effects** (identity link; only one dimension of sampling)
- **General*ized* Linear Models:** **any conditional outcome distribution**, **fixed** effects through **link functions**, no random effects (one dimension)

Note: Least Squares is only for GLM

- **General Linear Mixed Models:** conditionally normal outcome distribution, **fixed and random effects** (identity link, but multiple sampling dimensions)
- **General*ized* Linear Mixed Models:** **any conditional outcome distribution**, **fixed and random effects** through **link functions** (multiple dimensions)

- "Linear" means the fixed effects predict the *link-transformed* DV in a linear combination of (effect*predictor) + (effect*predictor)…

What can MLM do for you?

1. **Model dependency across observations**
   - Longitudinal, clustered, and/or cross-classified data? No problem!
   - Tailor your model of sources of correlation to your data

2. **Include categorical or continuous predictors at any level**
   - Time-varying, person-level, group-level predictors for each variance
   - Explore reasons for dependency, don't just control for dependency

3. **Does not require same data structure for each person**
   - Unbalanced or missing data? No problem!

4. **You already know how (or you will soon)!**
   - Use SPSS Mixed, **SAS Mixed**, Stata, Mplus, R, HLM, MlwiN…
   - What's an intercept? What's a slope? What's a pile of variance?

Levels of Analysis in Longitudinal Data

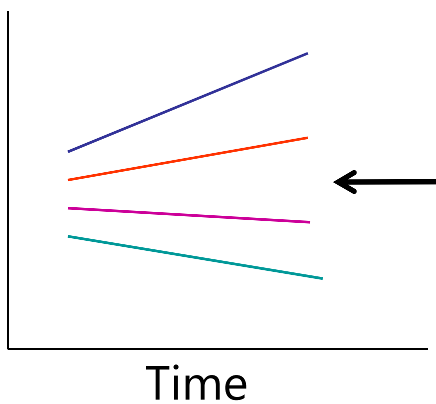
- Between-Person (BP) Variation:
  - **Level-2** – "**INTER**-individual Differences" – Time-Invariant
  - All longitudinal studies begin as cross-sectional studies
- Within-Person (WP) Variation:
  - **Level-1** – "**INTRA**-individual Differences" – Time-Varying
  - Only longitudinal studies can provide this extra information
- Longitudinal studies allow examination of both types of relationships simultaneously (and their interactions)
  - Any variable measured over time usually has both BP and WP variation
  - BP = more/less than other people; WP = more/less than one's average
- I use "person" here, but level-2 can be anything that is measured repeatedly (like animals, schools, countries…)

A Longitudinal Data Continuum

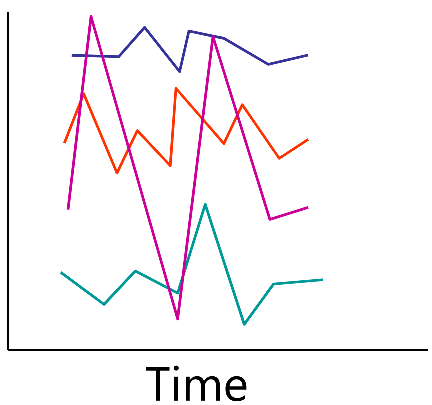
- **Within-Person Change**: Systematic change
  - Magnitude or direction of change can be different across individuals
  - "Growth curve models" → Time is meaningfully sampled

- **Within-Person Fluctuation**: No systematic change
  - Outcome just varies/fluctuates over time (e.g., emotion, stress)
  - Time is just a way to get lots of data per individual

Pure WP Change



Pure WP Fluctuation



The Two Sides of a (BP) Model

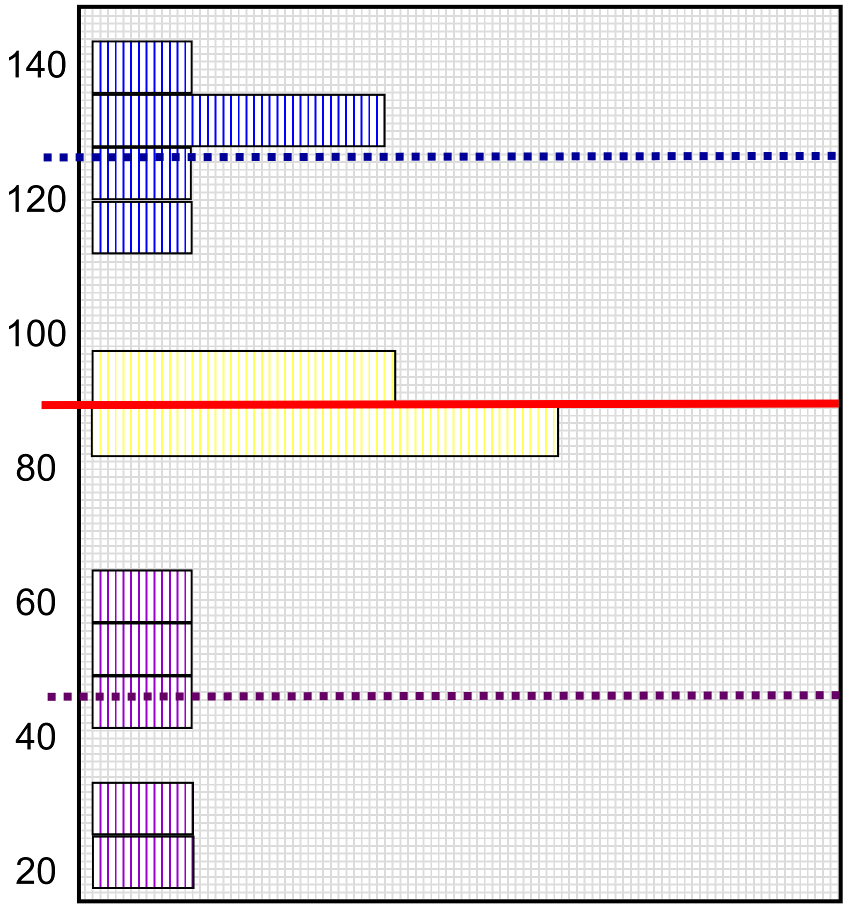
$$y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + e_i$$

Our focus today

- **Model for the Means (Predicted Values):**
  - Each person's expected (predicted) outcome is a weighted linear function of his/her values on X and Z (and here, their interaction), each measured once per person (i.e., this is a between-person model)
  - Estimated parameters are called fixed effects (here, β_0, β_1, β_2, and β_3)

- **Model for the Variance ("Piles" of Variance):**
  - $e_i \sim N(0, \sigma_e^2)$ → ONE residual (unexplained) deviation
  - e_i has a mean of 0 with some estimated constant variance σ_e^2, is normally distributed, is unrelated to X and Z, and is unrelated across people (across all observations, just people here)
  - **Estimated parameter is residual variance only in above BP model**

Empty +Within-Person Model



Start off with Mean of Y as "best guess" for any value:

= Grand Mean

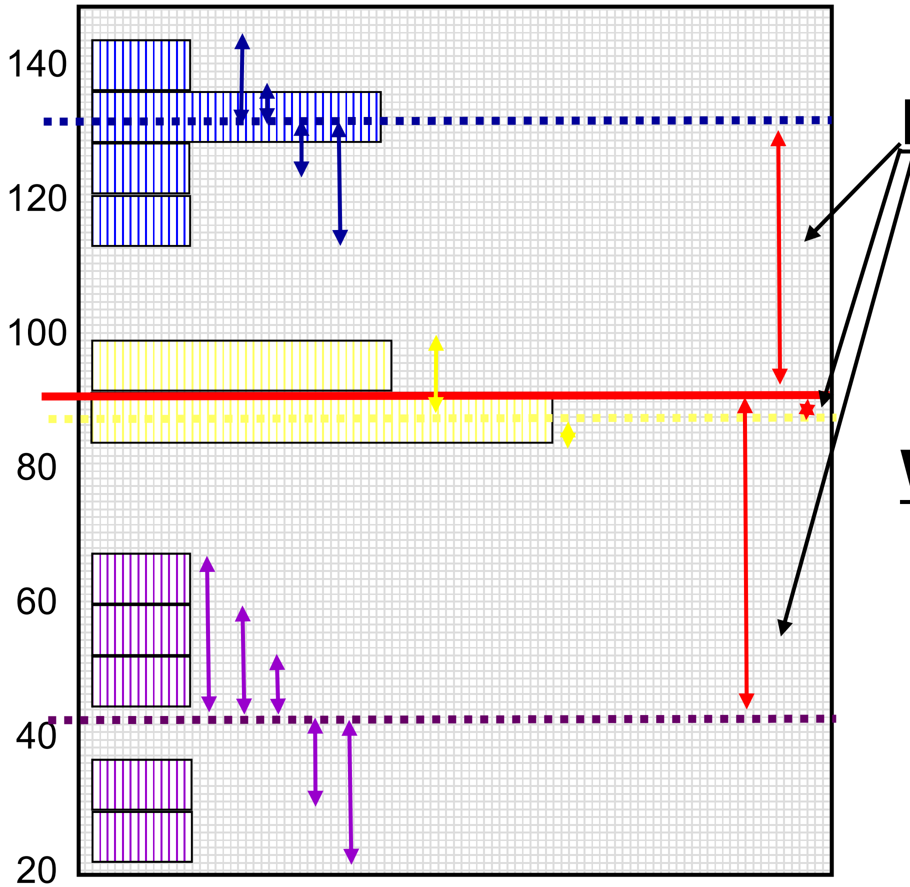
= Fixed Intercept

Can make better guess by taking advantage of repeated observations:

= Person Mean

→ Random Intercept

Empty +Within-Person Model



Variance of Y → 2 sources:

Between-Person (BP) Variance:

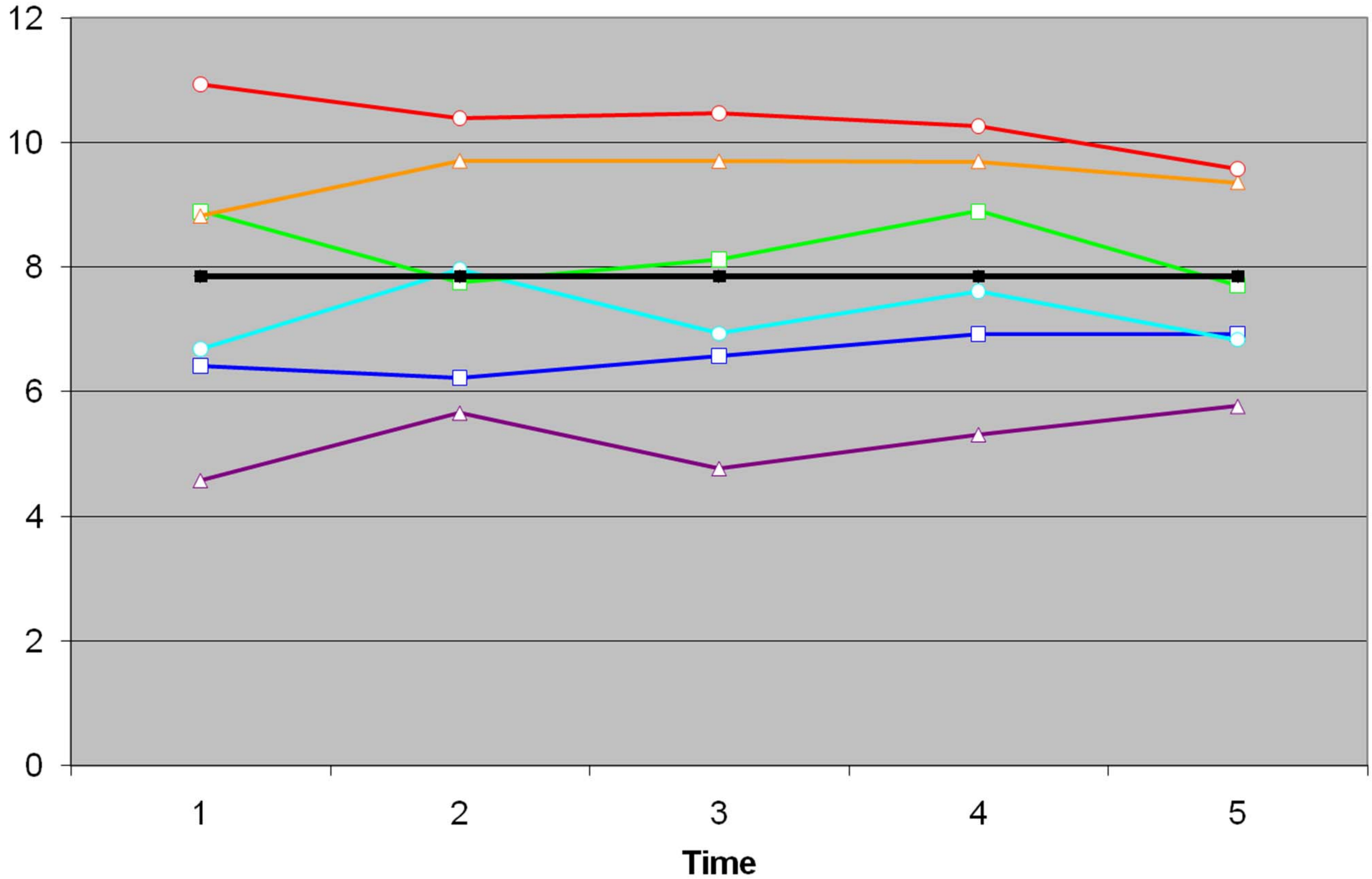
- → Differences from **GRAND** mean
- → **INTER**-Individual Differences

Within-Person (WP) Variance:

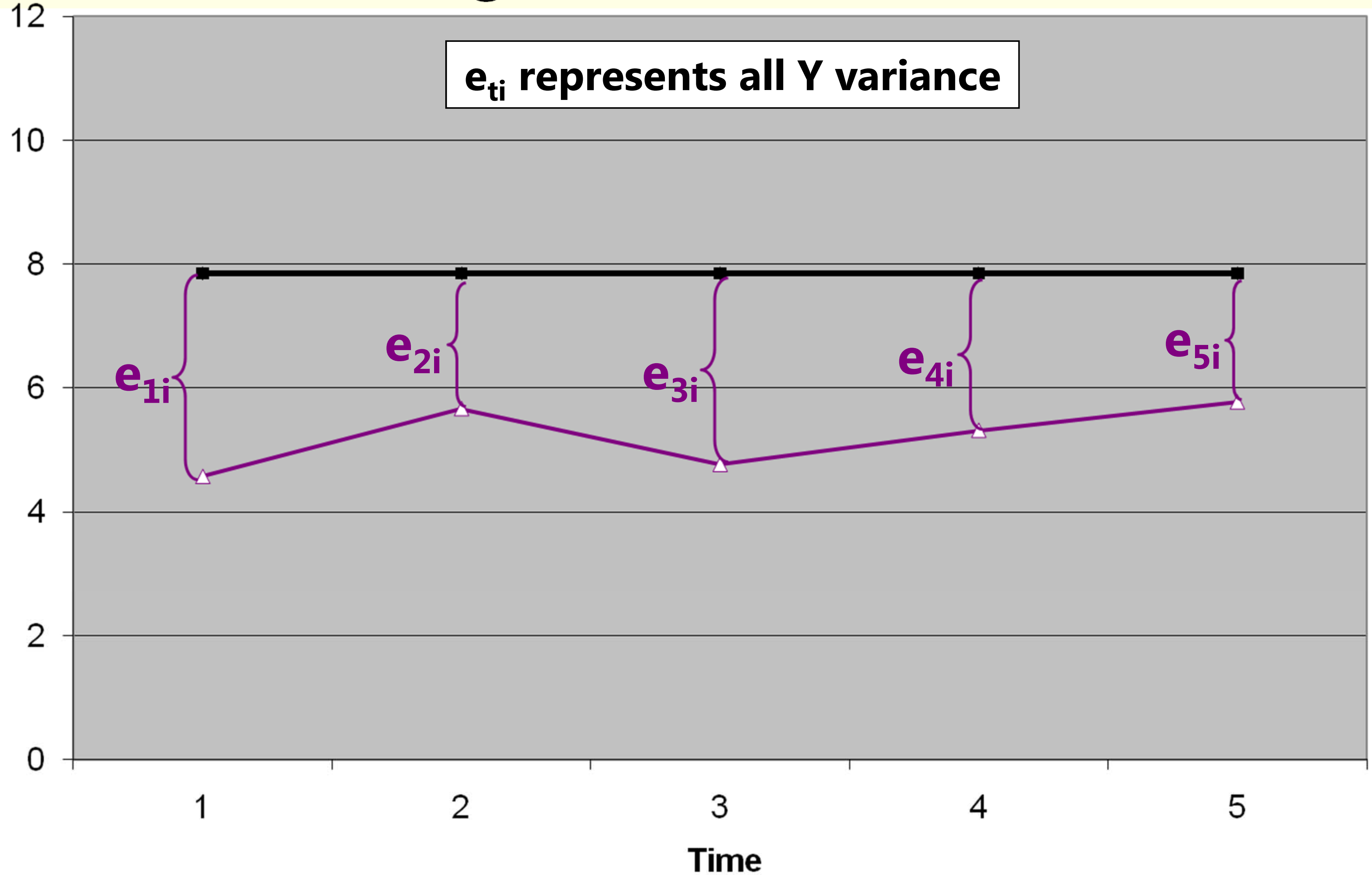
- → Differences from **OWN** mean
- → **INTRA**-Individual Differences
- → This part is only observable through longitudinal data.

Now we have 2 piles of variance in Y to predict.

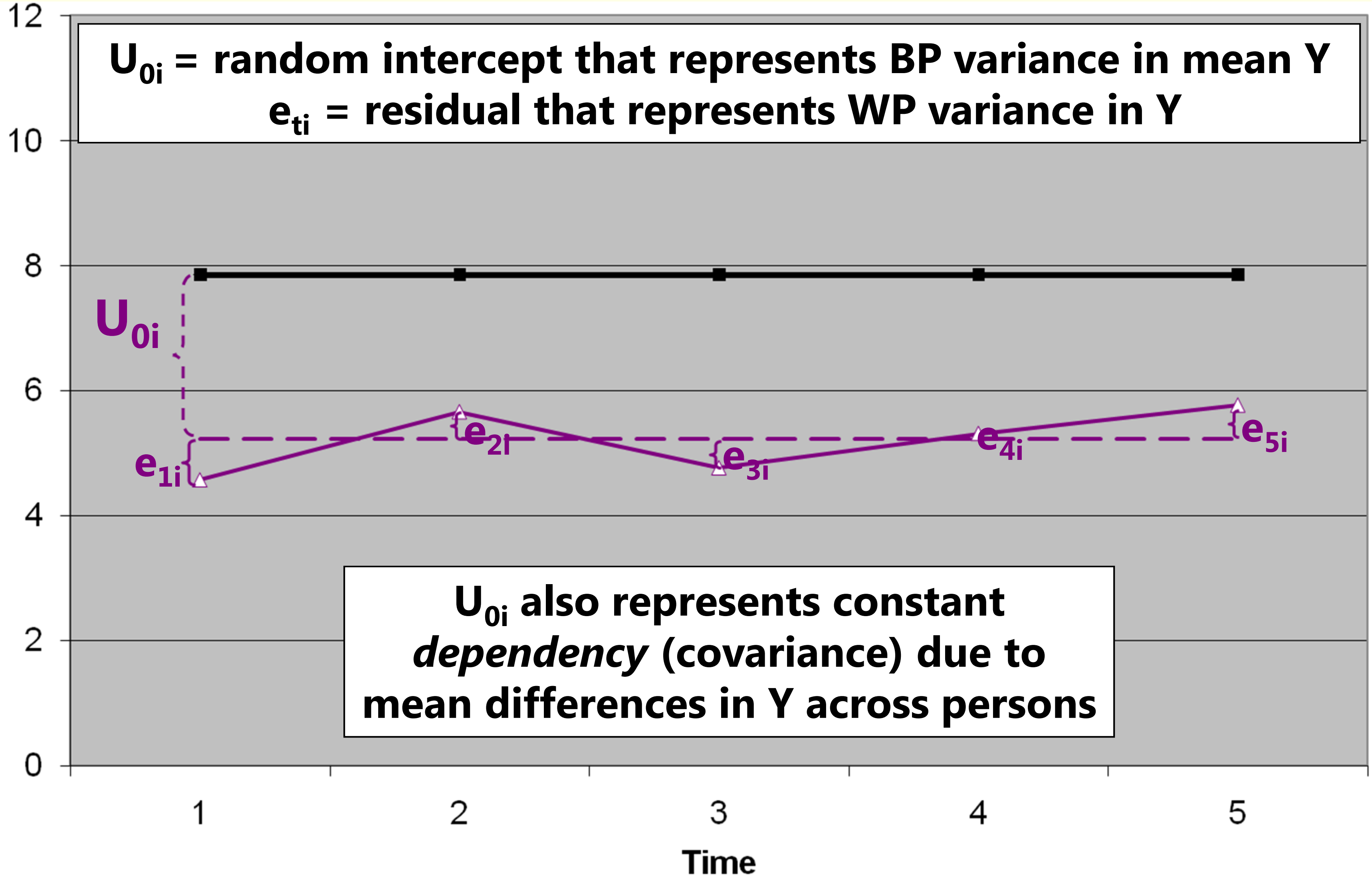
Hypothetical Longitudinal Data



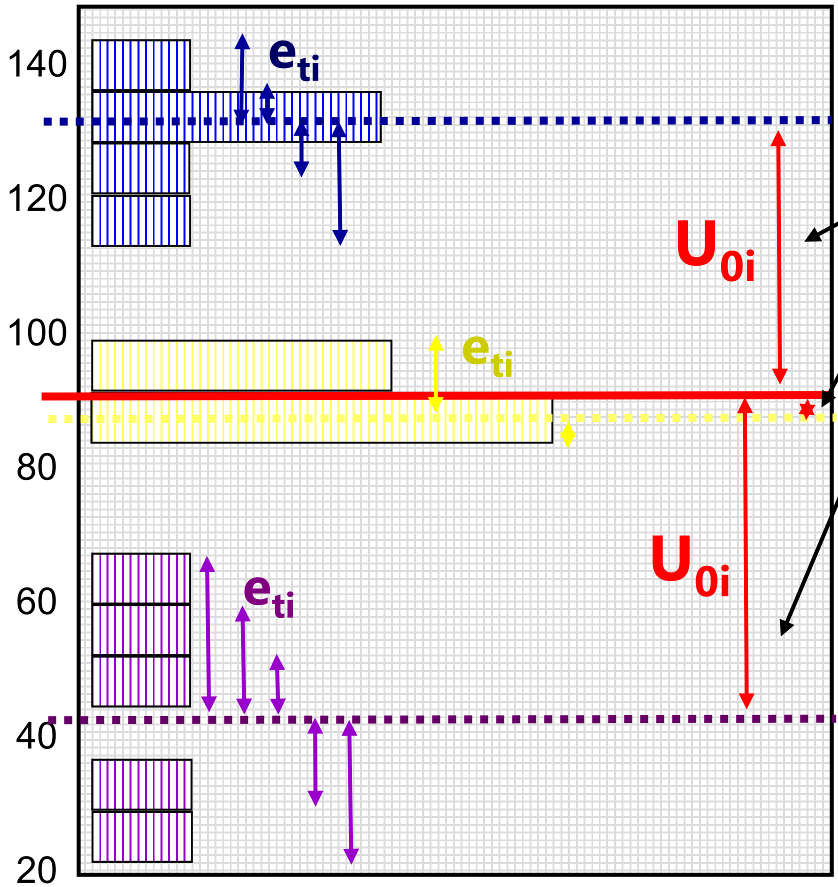
"Error" in a BP Model for the Variance: Single-Level Model



"Error" in a +WP Model for the Variance: Multilevel Model



Empty +Within-Person Model



Variance of Y → 2 sources:

Level 2 Random Intercept Variance (of U_{0i}, as $\tau^2_{U_0}$):

- → **Between**-Person Variance
- → Differences from **GRAND** mean
- → **INTER**-Individual Differences

Level 1 Residual Variance (of e_{ti}, as σ^2_e):

- → **Within**-Person Variance
- → Differences from **OWN** mean
- → **INTRA**-Individual Differences

BP vs. +WP Empty Models

- Empty **Between-Person** Model (used for 1 occasion):

$$y_i = \beta_0 + e_i$$

  - β_0 = fixed intercept = grand mean
  - e_i = residual deviation from GRAND mean

- Empty **+Within-Person** Model (>1 occasions):

$$y_{ti} = \beta_0 + U_{0i} + e_{ti}$$

  - β_0 = fixed intercept = grand mean
  - U_{0i} = random intercept = individual deviation from GRAND mean
  - e_{ti} = time-specific residual deviation from OWN mean

Intraclass Correlation (ICC)

Intraclass Correlation (ICC):

$$\text{ICC} = \frac{\text{BP}}{\text{BP} + \text{WP}} = \frac{\text{Intercept Var.}}{\text{Intercept Var.} + \text{Residual Var.}} = \frac{\boldsymbol{\tau}^2_{\mathbf{U_0}}}{\boldsymbol{\tau}^2_{\mathbf{U_0}} + \boldsymbol{\sigma}^2_{\mathbf{e}}}$$

$$\text{Corr}(y_1, y_2) = \frac{\text{Cov}(y_1, y_2)}{\sqrt{\text{Var}(y_1)} * \sqrt{\text{Var}(y_2)}}$$

R matrix

$$\begin{bmatrix} \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} \end{bmatrix}$$

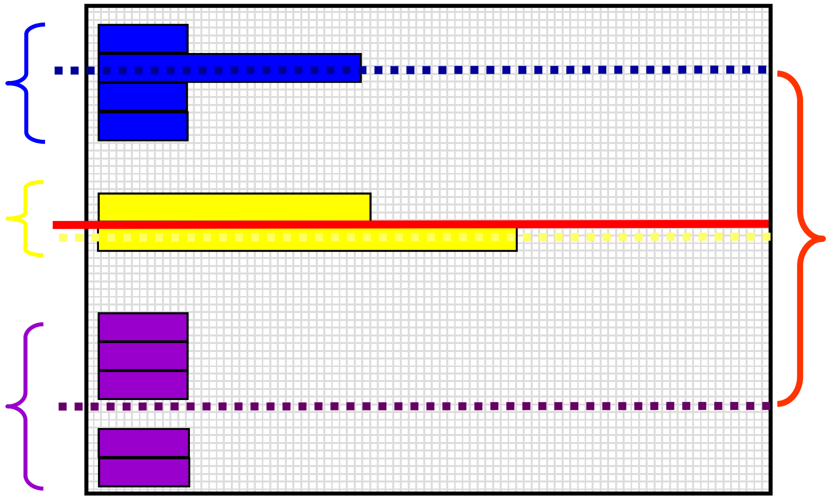
R CORR Matrix

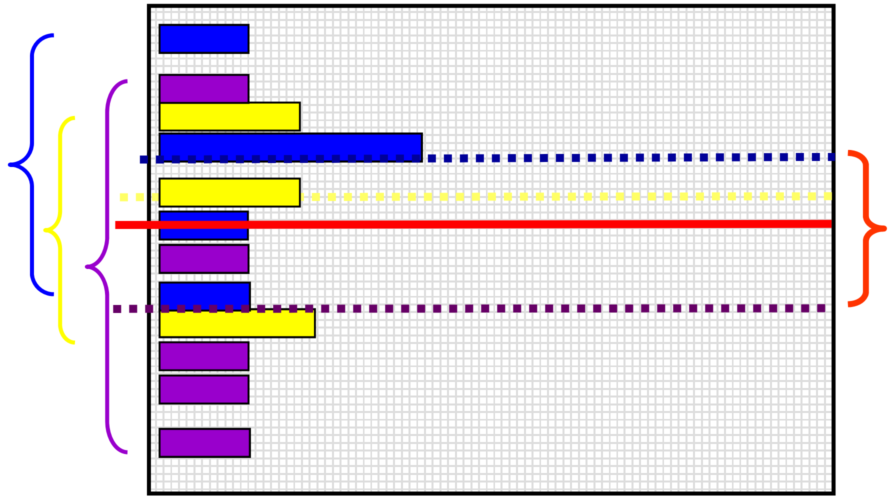
$$\begin{bmatrix} 1 & \text{ICC} & \text{ICC} \\ \text{ICC} & 1 & \text{ICC} \\ \text{ICC} & \text{ICC} & 1 \end{bmatrix}$$

- ICC = Proportion of total variance that is between persons
- ICC = Average correlation among occasions (in RCORR)
- ICC is a standardized way of expressing how much we need to worry about *dependency due to person mean differences* **(i.e., ICC is an effect size for constant person dependency)**

$$ICC = \frac{Between-Person}{Between-Person + Within-Person}$$

<u>Counter-Intuitive:</u> Between-Person Variance is in the numerator, but the ICC is the correlation over time!





<u>ICC</u> = BTW / BTW + within

→ Large ICC

→ Large correlation over time

<u>ICC</u> = btw / btw + WITHIN

→ Small ICC

→ Small correlation over time

BP and +WP Conditional Models

- Multiple Regression, **Between-Person** ANOVA: **1 PILE**
  - $y_i = (\beta_0 + \beta_1 X_i + \beta_2 Z_i ...) + e_i$
  - e_i → ONE residual, assumed uncorrelated with equal variance across observations (here, just persons) → "**BP (all) variation**"

- Repeated Measures, **Within-Person** ANOVA: **2 PILES**
  - $y_{ti} = (\beta_0 + \beta_1 X_i + \beta_2 Z_i ...) + U_{0i} + e_{ti}$
  - U_{0i} → A random intercept for differences in person means, assumed uncorrelated with equal variance across persons → "**BP (mean) variation**" = $\tau^2_{U_0}$ is now "leftover" after predictors
  - e_{ti} → A residual that represents remaining time-to-time variation, usually assumed uncorrelated with equal variance across observations (now, persons and time) → "**WP variation**" = σ^2_e is also now "leftover" after predictors

ANOVA for longitudinal data?

- There are 3 possible "kinds" of ANOVAs we could use:
  - Between-Persons/Groups, Univariate RM, and Multivariate RM
- **NONE OF THEM ALLOW:**
  - **Missing occasions** (do listwise deletion due to least squares)
  - **Time-varying predictors** (covariates are BP predictors only)
- Each includes the same model for the means for time: all possible mean differences (so 4 parameters to get to 4 means)
  - **"Saturated means model":** $\boldsymbol{\beta_0 + \beta_1(T_1) + \beta_2(T_2) + \beta_3(T_3)}$
  - **The *Time* variable must be balanced and discrete in ANOVA!**
- These ANOVAs differ by what they predict for the correlation across outcomes from the same person in the model for the variance…
  - i.e., **how they "handle dependency"** due to persons, or what they says the variance and covariance of the y_{ti} residuals should look like…

1. Between-Groups ANOVA

- **Uses e_{ti} only** (total variance = a single variance term of σ_e^2)
- **Assumes no covariance** at all among observations from the same person: *Dependency? What dependency?*
- Will usually be **very, very wrong** for longitudinal data
  - WP effects tested against wrong residual variance (significance tests will often be way too conservative)
  - Will also tend to be wrong for clustered data, but less so *(because the correlation among persons from the same group is not as strong as the correlation among occasions from the same person)*
- Predicts a variance-covariance matrix over time (here, 4 occasions) like this, called "**Variance Components**" (**R** matrix is TYPE=VC on REPEATED):

R matrix

$$\begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

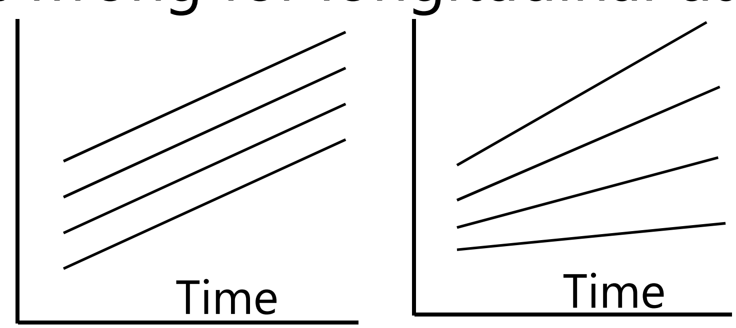
2a. Univariate Repeated Measures

- Separates total variance into **two** sources:
  - **Between-Person** (mean differences due to U_{0i}, or $\boldsymbol{\tau^2_{U_0}}$)
  - **Within-Person** (remaining variance due to e_{ti}, or $\boldsymbol{\sigma^2_e}$)
- Predicts a variance-covariance matrix over time (here, 4 occasions) like this, called "**Compound Symmetry**" (**R** matrix is TYPE=**CS** on REPEATED):
  - **Mean differences from U_{0i} are the only reason why occasions are correlated**

R matrix

$$\begin{bmatrix} \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} & \tau^2_{u_0} \\ \tau^2_{u_0} & \tau^2_{u_0} & \tau^2_{u_0} & \sigma^2_e + \tau^2_{u_0} \end{bmatrix}$$

- Will usually be at least somewhat wrong for longitudinal data
  - If people change at different rates, the variances and covariances over time have to change, too



The Problem with Univariate RM ANOVA

- Univ. RM ANOVA $\left(\tau^2_{U_0} + \sigma^2_e\right)$ predicts **compound symmetry:**
  - All variances and all covariances are equal across occasions
  - In other words, the amount of error observed should be the same at any occasion, so a single, pooled error variance term makes sense
  - If not, tests of fixed effects may be biased (i.e., sometimes tested against too much or too little error, if error is not really constant over time)
  - **COMPOUND SYMMETRY RARELY FITS FOR LONGITUDINAL DATA**
- But to get the correct tests of the fixed effects, the data must only meet a less restrictive assumption of **sphericity**:
  - In English → **pairwise differences** between adjacent occasions have equal variance and covariance (satisfied by default with only 2 occasions)
  - If compound symmetry is satisfied, so is sphericity (but see above)
  - Significance test provided in ANOVA for where data meet sphericity assumption
  - **Other RM ANOVA approaches are used when sphericity fails…**

The Other Repeated Measures ANOVAs…

- 2b. **Univariate RM ANOVA with sphericity corrections**
  - Based on ε → how far off sphericity (from 0-1, 1=spherical)
  - Applies an overall correction for model df based on estimated ε, but it doesn't really address the problem that data ≠ model

- 3. **Multivariate Repeated Measures ANOVA**
  - All variances and covariances are estimated separately over time (here, 4 occasions), called "**Unstructured**" (**R** matrix is TYPE=UN on REPEATED)—it's not a model, it IS the data:

R matrix

$$\begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44}^2 \end{bmatrix}$$

  - Because it can never be wrong, UN can be useful for **complete and balanced longitudinal data** with few occasions (e.g., 2-4)
  - Parameters = $\frac{\#\text{occasions} * (\#\text{occasions} + 1)}{2}$ so can be hard to estimate
  - Unstructured can also be specified to include random intercept variance $\tau_{U_0}^2$
  - Every other model for the variances is nested within Unstructured (we can do model comparisons to see if all other models are NOT WORSE)

Summary: ANOVA approaches for longitudinal data are "one size fits most"

- **Saturated Model for the Means** (balanced time required)
  - All possible mean differences
  - Unparsimonious, but best-fitting (is a description, not a model)

- **3 kinds of Models for the Variances** (complete data required)
  - BP ANOVA (σ_e^2 only) → assumes independence and constant variance over time
  - Univ. RM ANOVA $\left(\tau_{U_0}^2 + \sigma_e^2\right)$ → assumes constant variance and covariance
  - Multiv. RM ANOVA (whatever) → no assumptions; is a description, not a model

there is no structure that shows up in a scalar equation (i.e., the way U_{0i} + e_{ti} does)

- **MLM will give us more flexibility in both parts of the model:**
  - Fixed effects that *predict* the pattern of means (polynomials, pieces)
  - Random intercepts and slopes and/or alternative covariance structures that *predict* intermediate patterns of variance and covariance over time

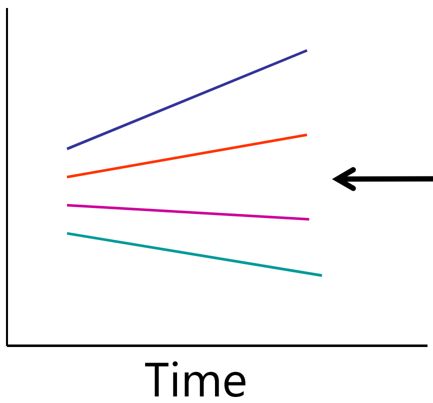
Review of Multilevel Models for Longitudinal Data

- Topics:
  - Concepts in longitudinal multilevel modeling
  - **Describing within-person fluctuation using ACS models**
  - Describing within-person change using random effects
  - Likelihood estimation in random effects models
  - Describing nonlinear patterns of change
  - Time-invariant predictors

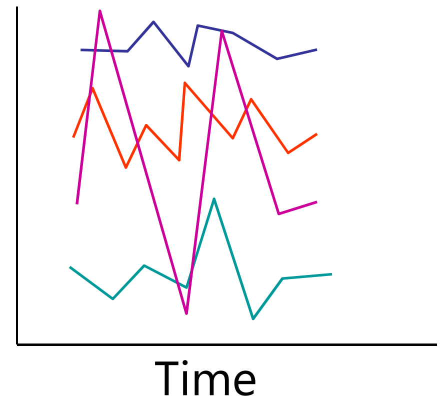
Modeling Change vs. Fluctuation

Our focus right now → Pure WP Fluctuation

Pure WP Change



Pure WP Fluctuation



Model for the Means:

- WP Change → describe pattern of *average* change (over "time")
- **WP Fluctuation** → *may* not need anything (if no systematic change)

Model for the Variances:

- WP Change → describe *individual differences* in change (random effects)
  → this allows variances and covariances to differ over time
- **WP Fluctuation** → describe pattern of variances and covariances over time

Big Picture Framework: Models for the Variance in Longitudinal Data

Compound Symmetry (CS)

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Most useful model: likely somewhere in between!

Unstructured (UN)

$$\begin{bmatrix} \sigma^2_{T1} & \sigma_{T12} & \sigma_{T13} & \sigma_{T14} \\ \sigma_{T21} & \sigma^2_{T2} & \sigma_{T23} & \sigma_{T24} \\ \sigma_{T31} & \sigma_{T32} & \sigma^2_{T3} & \sigma_{T43} \\ \sigma_{T41} & \sigma_{T42} & \sigma_{T43} & \sigma^2_{T4} \end{bmatrix}$$

Parsimony ⟵⟶ **Good fit**

Univariate RM ANOVA

NAME ...THAT ... STRUCTURE!

Multivariate RM ANOVA

What is the pattern of variance and covariance over time?

CS and UN are just two of the many, many options available within MLM, including ***random effects models*** (for change) and ***alternative covariance structure models*** (for fluctuation).

Alternative Covariance Structure Models

- Useful in predicting patterns of variance and covariance that arise from <u>fluctuation</u> in the outcome over time:
  - **Variances**: Same (homogeneous) or different (heterogeneous)?
  - **Covariances**: Same or different? If different, what is the pattern?
    - Models with heterogeneous variances predict correlation instead of covariance
  - Often don't need any fixed effects for systematic effects of time in the model for the means (although this is always an empirical question)

- Limitations for most of the ACS models:
  - Require **equal-interval** occasions (they are based on idea of "time lag")
  - Require **balanced** time across persons (no intermediate time values)
  - But **do not require complete data** (unlike when CS and UN are estimated via least squares in ANOVA instead of ML/REML in MLM)

- ACS models do require some new terminology to introduce…

Two Families of ACS Models

- So far, we've referred to the variance and covariance matrix of the longitudinal outcomes as the **R** matrix
  - We now refer to these as **"R-only models"** (use **REPEATED** statement only)
  - Although the **R** matrix is actually specified per individual, ACS models usually assume the same **R** matrix for everyone
  - **R** matrix is symmetric with dimensions *n* x *n*, in which *n* = # occasions per person (although people can have missing data, the same set of *possible* occasions is required across people to use most **R**-only models)

- **3 other matrices we'll see in "G and R combined" ACS models:**
  - **G** = matrix of random effects variances and covariances (stay tuned)
  - **Z** = matrix of values for predictors that have random effects (stay tuned)
  - **V** = symmetric *n* x *n* matrix of **total** variance and covariance over time
    - If the model includes random effects, then **G** and **Z** get combined with **R** to make **V** as $\mathbf{V} = \mathbf{ZGZ}^{\mathrm{T}} + \mathbf{R}$ (accomplished by adding the **RANDOM** statement)
    - If the model does NOT include random effects in **G**, then $\mathbf{V} = \mathbf{R}$… *so, **R**-only*

R-Only ACS Models

- The **R-only** models to be presented next are all specified using the **REPEATED** statement only (no RANDOM statement)

- They are explained by showing their predicted **R** matrix, which provides the **total** variances and covariances across occasions
  - Total variance per occasion on diagonal
  - Total covariances across occasions on off-diagonals
  - I've included in " " the labels SAS uses for each parameter

- Correlations across occasions can be calculated given variances and covariances, which would be shown in the **RCORR** matrix (available in SAS PROC MIXED)
  - 1's on diagonal (standardized variables), correlations on off-diagonal

- **Unstructured (TYPE=UN) will always fit best by −2LL**
  - All ACS models are nested within Unstructured (UN = the data)
  - Goal: find an ACS model that is **simpler** but **not worse fitting** than UN

R-Only ACS Models: CS/CSH

- **Compound Symmetry: TYPE=CS**
  - 2 parameters:
    - **1 "residual" variance σ_e^2**
    - **1 "CS" covariance** across occasions
  - Constant total variance: $CS + \sigma_e^2$
  - Constant total covariance: CS

$$\begin{bmatrix} CS+\sigma_e^2 & CS & CS & CS \\ CS & CS+\sigma_e^2 & CS & CS \\ CS & CS & CS+\sigma_e^2 & CS \\ CS & CS & CS & CS+\sigma_e^2 \end{bmatrix}$$

- **Compound Symmetry Heterogeneous: TYPE=CSH**
  - *n*+1 parameters:
    - ***n* separate "Var(*n*)" total variances σ_{Tn}^2**
    - **1 "CSH" total correlation across occasions**
  - Separate total variances are estimated directly
  - Still constant total correlation: CSH (but has non-constant covariances)

$$\begin{bmatrix} \sigma_{T1}^2 & CSH\sigma_{T1}\sigma_{T2} & CSH\sigma_{T1}\sigma_{T3} & CSH\sigma_{T1}\sigma_{T4} \\ CSH\sigma_{T2}\sigma_{T1} & \sigma_{T2}^2 & CSH\sigma_{T2}\sigma_{T3} & CSH\sigma_{T2}\sigma_{T4} \\ CSH\sigma_{T3}\sigma_{T1} & CSH\sigma_{T3}\sigma_{T2} & \sigma_{T3}^2 & CSH\sigma_{T3}\sigma_{T4} \\ CSH\sigma_{T4}\sigma_{T1} & CSH\sigma_{T4}\sigma_{T2} & CSH\sigma_{T4}\sigma_{T3} & \sigma_{T4}^2 \end{bmatrix}$$

R-Only ACS Models: AR1/ARH1

- **1st Order Auto-Regressive: TYPE=AR(1)**
  - 2 parameters:
    - **1 constant total variance σ_T^2 (mislabeled "residual")**
    - **1 "AR1" total auto-correlation r_T** across occasions
    - r_T^1 is lag-1 correlation, r_T^2 is lag-2 correlation, r_T^3 is lag-3 correlation….

$$\begin{bmatrix} \sigma_T^2 & r_T^1\sigma_T^2 & r_T^2\sigma_T^2 & r_T^3\sigma_T^2 \\ r_T^1\sigma_T^2 & \sigma_T^2 & r_T^1\sigma_T^2 & r_T^2\sigma_T^2 \\ r_T^2\sigma_T^2 & r_T^1\sigma_T^2 & \sigma_T^2 & r_T^1\sigma_T^2 \\ r_T^3\sigma_T^2 & r_T^2\sigma_T^2 & r_T^1\sigma_T^2 & \sigma_T^2 \end{bmatrix}$$

- **1st Order Auto-Regressive Heterogeneous: TYPE=ARH(1)**
  - *n*+1 parameters:
    - ***n* separate "Var(*n*)" total variances σ_{Tn}^2**
    - **1 "ARH1" total auto-correlation r_T** across occasions
    - r_T^1 is lag-1 correlation, r_T^2 is lag-2 correlation, r_T^3 is lag-3 correlation….

$$\begin{bmatrix} \sigma_{T1}^2 & r_T^1\sigma_{T1}\sigma_{T2} & r_T^2\sigma_{T1}\sigma_{T3} & r_T^3\sigma_{T1}\sigma_{T4} \\ r_T^1\sigma_{T2}\sigma_{T1} & \sigma_{T2}^2 & r_T^1\sigma_{T2}\sigma_{T3} & r_T^2\sigma_{T2}\sigma_{T4} \\ r_T^2\sigma_{T3}\sigma_{T1} & r_T^1\sigma_{T3}\sigma_{T2} & \sigma_{T3}^2 & r_T^1\sigma_{T3}\sigma_{T4} \\ r_T^3\sigma_{T4}\sigma_{T1} & r_T^2\sigma_{T4}\sigma_{T2} & r_T^1\sigma_{T4}\sigma_{T3} & \sigma_{T4}^2 \end{bmatrix}$$

R-Only ACS Models: TOEP*n*/TOEPH*n*

- **Toeplitz(*n*): TYPE=TOEP(*n*)**
  - *n* parameters:
    - **1 constant total variance σ^2_T (mislabeled "residual")**
    - ***n*−1 "TOEP(*lag*)" c_{Tn} banded total covariances** across occasions
    - c_{T1} is lag-1 covariance, c_{T2} is lag-2 covariance, c_{T3} is lag-3 covariance….

$$\begin{bmatrix} \sigma^2_T & & & \\ c_{T1} & \sigma^2_T & & \\ c_{T2} & c_{T1} & \sigma^2_T & \\ c_{T3} & c_{T2} & c_{T1} & \sigma^2_T \end{bmatrix}$$

- **Toeplitz Heterogeneous(*n*): TYPE=TOEPH(*n*)**
  - *n* + (*n*−1) parameters:
    - ***n* separate "Var(*n*)" total variances σ^2_{Tn}**
    - ***n*−1 "TOEPH(*lag*)" r_{Tn} banded total correlations** across occasions
    - r_{T1} is lag-1 correlation, r_{T2} is lag-2 correlation, r_{T3} is lag-3 correlation….

$$\begin{bmatrix} \sigma^2_{T1} & r_{T1}\sigma_{T1}\sigma_{T2} & r_{T2}\sigma_{T1}\sigma_{T3} & r_{T3}\sigma_{T1}\sigma_{T4} \\ r_{T1}\sigma_{T2}\sigma_{T1} & \sigma^2_{T2} & r_{T1}\sigma_{T2}\sigma_{T3} & r_{T2}\sigma_{T2}\sigma_{T4} \\ r_{T2}\sigma_{T3}\sigma_{T1} & r_{T1}\sigma_{T3}\sigma_{T2} & \sigma^2_{T3} & r_{T1}\sigma_{T3}\sigma_{T4} \\ r_{T3}\sigma_{T4}\sigma_{T1} & r_{T2}\sigma_{T4}\sigma_{T2} & r_{T1}\sigma_{T4}\sigma_{T3} & \sigma^2_{T4} \end{bmatrix}$$

Comparing **R**-only ACS Models

- Baseline models: **CS =simplest**, **UN = most complex**
  - Relative to CS, more complex models fit "better" or "not better"
  - Relative to UN, less complex models fit "worse" or "not worse"

- Other rules of nesting and model comparisons:
  - Homogeneous variance models are nested within heterogeneous variance models (e.g., CS in CSH, AR1 in ARH1, TOEP in TOEPH)
  - CS and AR1 are each nested within TOEP (i.e., TOEP can become CS or AR1 through restrictions of its covariance patterns)
  - CS and AR1 are not nested (because both have 2 parameters)
  - **R**-only models differ in unbounded parameters, so can be compared using regular $-2\Delta LL$ tests (instead of mixture $-2\Delta LL$ tests)
  - Good idea to start by assuming heterogeneous variances until you settle on the covariance pattern, then test if het. var. are still necessary
  - When in doubt, just compare AIC and BIC (useful even with $-2\Delta LL$ tests)

The Other Family of ACS Models

- **R**-only models *directly* predict the **total** variance and covariance
- **G** and **R** models *indirectly* predict the total variance and covariance through **between-person (BP)** and **within-person (WP)** sources of variance and covariance → So, for this model: $\mathbf{y_{ti}} = \boldsymbol{\beta_0} + \mathbf{U_{0i}} + \mathbf{e_{ti}}$
  - **BP** = **G** matrix of **level-2 random effect ($\mathbf{U_{0i}}$)** variances and covariances
    - Which effects get to be random (whose variance and covariances are then included in **G**) is specified using the **RANDOM** statement (always TYPE=UN)
    - Our ACS models have a random intercept only, so **G** is 1x1 scalar of $[\boldsymbol{\tau}^2_{\mathbf{U_0}}]$
  - **WP** = **R** matrix of **level-1 ($\mathbf{e_{ti}}$) residual** variances and covariances
    - The *n* x *n* **R** matrix of **residual** variances and covariances **that remain** after controlling for random intercept variance is then modeled with **REPEATED**
  - **Total** = **V** = *n* x *n* matrix of **total** variance and covariance over time that results from putting **G** and **R** together: $\mathbf{V} = \mathbf{ZGZ}^{\mathrm{T}} + \mathbf{R}$
    - **Z** is a matrix that holds the values of predictors with random effects, but **Z** will be an *n* x 1 column of 1's for now (random intercept only)

A "Random Intercept" (**G** and **R**) Model

Total Predicted Data Matrix is called V Matrix

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Level 2, BP Variance

Unstructured **G Matrix**
(RANDOM statement)
Each person has same **1 x 1 G** matrix (no covariance across persons in two-level model)

Random Intercept Variance only

$$\begin{bmatrix} \tau^2_{U_0} \end{bmatrix}$$

Level 1, WP Variance

Diagonal (VC) **R Matrix**
(REPEATED statement)
Each person has same ***n* x *n* R** matrix → **equal variances and 0 covariances** across time (no covariance across persons)

Residual Variance only

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

CS as a "Random Intercept" Model

<u>RI and DIAG</u>: Total predicted data matrix is called V matrix, created from the G [TYPE=UN] and R [TYPE=VC] matrices as follows:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix} = \begin{bmatrix} \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e \end{bmatrix}$$

Z represents *n* per person

Does the end result V look familiar? It should: CS = τ^2_{U0}

$$\begin{bmatrix} CS + \sigma^2_e & CS & CS & CS \\ CS & CS + \sigma^2_e & CS & CS \\ CS & CS & CS + \sigma^2_e & CS \\ CS & CS & CS & CS + \sigma^2_e \end{bmatrix}$$

So if the **R-only CS model** (the simplest baseline) can be specified equivalently using **G and R**, can we do the same for the **R-only UN model** (the most complex baseline)?

Absolutely! ...*with one small catch*

UN via a "Random Intercept" Model

RI and UN$n-1$: Total predicted data matrix is called V matrix, created from the G [TYPE=UN] and R [TYPE=UN($n-1$)] matrices as follows:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma^2_{e1} & \sigma_{e12} & \sigma_{e13} & 0 \\ \sigma_{e21} & \sigma^2_{e2} & \sigma_{e23} & \sigma_{e24} \\ \sigma_{e31} & \sigma_{e32} & \sigma^2_{e3} & \sigma_{e34} \\ 0 & \sigma_{e42} & \sigma_{e43} & \sigma^2_{e4} \end{bmatrix} = \begin{bmatrix} \tau^2_{U_0} + \sigma^2_{e1} & \tau^2_{U_0} + \sigma_{e12} & \tau^2_{U_0} + \sigma_{e13} & \tau^2_{U_0} \\ \tau^2_{U_0} + \sigma_{e21} & \tau^2_{U_0} + \sigma^2_{e2} & \tau^2_{U_0} + \sigma_{e23} & \tau^2_{U_0} + \sigma_{e24} \\ \tau^2_{U_0} + \sigma_{e31} & \tau^2_{U_0} + \sigma_{e32} & \tau^2_{U_0} + \sigma^2_{e3} & \tau^2_{U_0} + \sigma_{e34} \\ \tau^2_{U_0} & \tau^2_{U_0} + \sigma_{e42} & \tau^2_{U_0} + \sigma_{e43} & \tau^2_{U_0} + \sigma^2_{e4} \end{bmatrix}$$

This **RI and UN$n-1$ model** is equivalent to (makes same predictions as) the **R-only UN model**. But it shows the *residual* (not total) covariances.

Because we can't estimate all possible variances and covariances in the **R** matrix and also estimate the random intercept variance $\tau^2_{U_0}$ in the **G** matrix, we have to eliminate the last **R** matrix covariance by setting it to 0.

Accordingly, in the **RI and UN$n-1$** model, the random intercept variance $\tau^2_{U_0}$ takes on the value of the covariance for the first and last occasions.

Rationale for **G** and **R** ACS models

- Modeling WP fluctuation traditionally involves using **R** only (no **G**) → **Total** BP + WP variance described by just **R** matrix (so **R**=**V**)
  - Correlations would still be expected even at distant time lags because of constant individual differences (i.e., the BP random intercept)
  - Resulting **R**-only model may require lots of estimated parameters as a result e.g., 8 time points? Pry need a 7-lag Toeplitz(8) model

- **Why not take out the primary reason for the covariance across occasions (the random intercept variance) and see what's left?**
  - Random intercept variance $\boldsymbol{\tau}^2_{\mathbf{U_0}}$ in **G** → control for person mean differences
  - THEN predict just the **residual** variance/covariance in **R**, not the **total**
  - Resulting model may be more parsimonious (e.g., maybe only lag1 or lag2 occasions are still related after removing $\boldsymbol{\tau}^2_{\mathbf{U_0}}$ as a source of covariance)
  - Has the advantage of still distinguishing BP from WP variance (useful for descriptive purposes and for calculating effect sizes later)

Random Intercept + Diagonal **R** Models

RI and DIAG: V is created from **G [TYPE=UN] and R [TYPE=VC]:**
homogeneous *residual variances;* ***no*** *residual covariances*

Same fit as R-only CS

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix} = \begin{bmatrix} \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_e \end{bmatrix}$$

RI and DIAGH: V is created from **G [TYPE=UN] and R [TYPE=UN(1)]:**
heterogeneous *residual variances;* ***no*** *residual covariances*

NOT same fit as R-only CSH

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma^2_{e1} & 0 & 0 & 0 \\ 0 & \sigma^2_{e2} & 0 & 0 \\ 0 & 0 & \sigma^2_{e3} & 0 \\ 0 & 0 & 0 & \sigma^2_{e4} \end{bmatrix} = \begin{bmatrix} \tau^2_{U_0} + \sigma^2_{e1} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_{e2} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_{e3} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} + \sigma^2_{e4} \end{bmatrix}$$

Random Intercept + AR1 **R** Models

RI and AR1: V is created from **G [TYPE=UN] and R [TYPE=AR(1)]:**
homogeneous *residual variances;* ***auto-regressive lagged*** *residual covariances*

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_e & r^1_e\sigma^2_e & r^2_e\sigma^2_e & r^3_e\sigma^2_e \\ r^1_e\sigma^2_e & \sigma^2_e & r^1_e\sigma^2_e & r^2_e\sigma^2_e \\ r^2_e\sigma^2_e & r^1_e\sigma^2_e & \sigma^2_e & r^1_e\sigma^2_e \\ r^3_e\sigma^2_e & r^2_e\sigma^2_e & r^1_e\sigma^2_e & \sigma^2_e\end{bmatrix} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}+r^1_e\sigma^2_e & \tau^2_{U_0}+r^2_e\sigma^2_e & \tau^2_{U_0}+r^3_e\sigma^2_e \\ \tau^2_{U_0}+r^1_e\sigma^2_e & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}+r^1_e\sigma^2_e & \tau^2_{U_0}+r^2_e\sigma^2_e \\ \tau^2_{U_0}+r^2_e\sigma^2_e & \tau^2_{U_0}+r^1_e\sigma^2_e & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}+r^1_e\sigma^2_e \\ \tau^2_{U_0}+r^3_e\sigma^2_e & \tau^2_{U_0}+r^2_e\sigma^2_e & \tau^2_{U_0}+r^1_e\sigma^2_e & \tau^2_{U_0}+\sigma^2_e\end{bmatrix}$$

RI and ARH1: V is created from **G [TYPE=UN] and R [TYPE=ARH(1)]:**
heterogeneous *residual variances;* ***auto-regressive lagged*** *residual covariances*

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_{e1} & r^1_e\sigma_{e1}\sigma_{e2} & r^2_e\sigma_{e1}\sigma_{e3} & r^3_e\sigma_{e1}\sigma_{e4} \\ r^1_e\sigma_{e2}\sigma_{e1} & \sigma^2_{e2} & r^1_e\sigma_{e2}\sigma_{e3} & r^2_e\sigma_{e2}\sigma_{e4} \\ r^2_e\sigma_{e3}\sigma_{e1} & r^1_e\sigma_{e3}\sigma_{e2} & \sigma^2_{e3} & r^1_e\sigma_{e3}\sigma_{e4} \\ r^3_e\sigma_{e4}\sigma_{e1} & r^2_e\sigma_{e4}\sigma_{e2} & r^1_e\sigma_{e4}\sigma_{e3} & \sigma^2_{e4}\end{bmatrix} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_{e1} & \tau^2_{U_0}+r^1_e\sigma_{e1}\sigma_{e2} & \tau^2_{U_0}+r^2_e\sigma_{e1}\sigma_{e3} & \tau^2_{U_0}+r^3_e\sigma_{e1}\sigma_{e4} \\ \tau^2_{U_0}+r^1_e\sigma_{e2}\sigma_{e1} & \tau^2_{U_0}+\sigma^2_{e2} & \tau^2_{U_0}+r^1_e\sigma_{e2}\sigma_{e3} & \tau^2_{U_0}+r^2_e\sigma_{e2}\sigma_{e4} \\ \tau^2_{U_0}+r^2_e\sigma_{e3}\sigma_{e1} & \tau^2_{U_0}+r^1_e\sigma_{e3}\sigma_{e2} & \tau^2_{U_0}+\sigma^2_{e3} & \tau^2_{U_0}+r^1_e\sigma_{e3}\sigma_{e4} \\ \tau^2_{U_0}+r^3_e\sigma_{e4}\sigma_{e1} & \tau^2_{U_0}+r^2_e\sigma_{e4}\sigma_{e2} & \tau^2_{U_0}+r^1_e\sigma_{e4}\sigma_{e3} & \tau^2_{U_0}+\sigma^2_{e4}\end{bmatrix}$$

Random Intercept + TOEP*n*−*1* R Models

RI and TOEP*n*−*1*: V is created from **G [TYPE=UN] and R [TYPE=TOEP(*n*−*1*)]:**
homogeneous *residual variances;* ***banded*** *residual covariances*

Same fit as R-only TOEP(*n*)

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_e & c_{e1} & c_{e2} & 0\\ c_{e1} & \sigma^2_e & c_{e1} & c_{e2}\\ c_{e2} & c_{e1} & \sigma^2_e & c_{e1}\\ 0 & c_{e2} & c_{e1} & \sigma^2_e\end{bmatrix} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}+c_{e1} & \tau^2_{U_0}+c_{e2} & \tau^2_{U_0}\\ \tau^2_{U_0}+c_{e1} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}+c_{e1} & \tau^2_{U_0}+c_{e2}\\ \tau^2_{U_0}+c_{e2} & \tau^2_{U_0}+c_{e1} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}+c_{e1}\\ \tau^2_{U_0} & \tau^2_{U_0}+c_{e2} & \tau^2_{U_0}+c_{e1} & \tau^2_{U_0}+\sigma^2_e\end{bmatrix}$$

Because of τ^2_{U0}, highest lag covariance in **R** must be set to 0 for model to be identified

RI and TOEPH*n*−*1*: V is created from **G [TYPE=UN] and R [TYPE=TOEPH(*n*−*1*)]:**
homogeneous *residual variances;* ***banded*** *residual covariances*

NOT same fit as R-only TOEPH(*n*)

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_{e1} & r_{e1}\sigma_{e1}\sigma_{e2} & r_{e2}\sigma_{e1}\sigma_{e3} & 0\\ r_{e1}\sigma_{e2}\sigma_{e1} & \sigma^2_{e2} & r_{e1}\sigma_{e2}\sigma_{e3} & r_{e2}\sigma_{e2}\sigma_{e4}\\ r_{e2}\sigma_{e3}\sigma_{e1} & r_{e1}\sigma_{e3}\sigma_{e2} & \sigma^2_{e3} & r_{e1}\sigma_{e3}\sigma_{e4}\\ 0 & r_{e2}\sigma_{e4}\sigma_{e2} & r_{e1}\sigma_{e4}\sigma_{e3} & \sigma^2_{e4}\end{bmatrix} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_{e1} & \tau^2_{U_0}+r_{e1}\sigma_{e1}\sigma_{e2} & \tau^2_{U_0}+r_{e2}\sigma_{e1}\sigma_{e3} & \tau^2_{U_0}\\ \tau^2_{U_0}+r_{e1}\sigma_{e2}\sigma_{e1} & \tau^2_{U_0}+\sigma^2_{e2} & \tau^2_{U_0}+r_{e1}\sigma_{e2}\sigma_{e3} & \tau^2_{U_0}+r_{e2}\sigma_{e2}\sigma_{e4}\\ \tau^2_{U_0}+r_{e2}\sigma_{e3}\sigma_{e1} & \tau^2_{U_0}+r_{e1}\sigma_{e3}\sigma_{e2} & \tau^2_{U_0}+\sigma^2_{e3} & \tau^2_{U_0}+r_{e1}\sigma_{e3}\sigma_{e4}\\ \tau^2_{U_0} & \tau^2_{U_0}+r_{e2}\sigma_{e4}\sigma_{e2} & \tau^2_{U_0}+r_{e1}\sigma_{e4}\sigma_{e3} & \tau^2_{U_0}+\sigma^2_{e4}\end{bmatrix}$$

Random Intercept + TOEP*2* R Models

RI and TOEP*2*: **V** is created from **G [TYPE=UN] and R [TYPE=TOEP(*2*)]:**
homogeneous *residual variances;* ***banded*** *residual covariance at* ***lag1*** *only*

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_e & c_{e1} & 0 & 0\\ c_{e1} & \sigma^2_e & c_{e1} & 0\\ 0 & c_{e1} & \sigma^2_e & c_{e1}\\ 0 & 0 & c_{e1} & \sigma^2_e\end{bmatrix} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}+c_{e1} & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0}+c_{e1} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}+c_{e1} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0}+c_{e1} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}+c_{e1}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+c_{e1} & \tau^2_{U_0}+\sigma^2_e\end{bmatrix}$$

Now we can test the need for residual covariances at higher lags

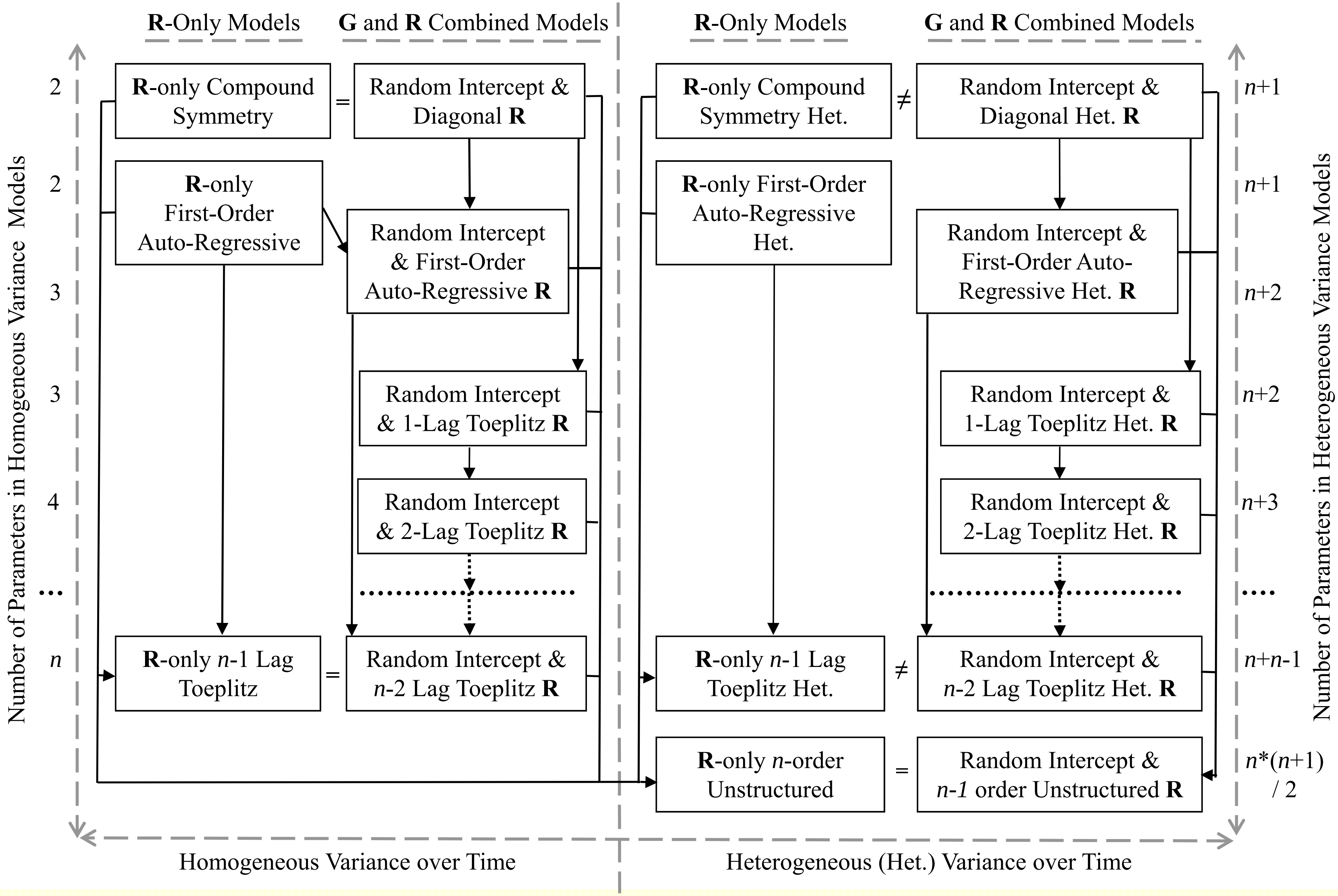
RI and TOEPH*1*: **V** is created from **G [TYPE=UN] and R [TYPE=TOEPH(*2*)]:**
homogeneous *residual variances;* ***banded*** *residual covariance at* ***lag1*** *only*

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_{e1} & r_{e1}\sigma_{e1}\sigma_{e2} & 0 & 0\\ r_{e1}\sigma_{e2}\sigma_{e1} & \sigma^2_{e2} & r_{e1}\sigma_{e2}\sigma_{e3} & 0\\ 0 & r_{e1}\sigma_{e3}\sigma_{e2} & \sigma^2_{e3} & r_{e1}\sigma_{e3}\sigma_{e4}\\ 0 & 0 & r_{e1}\sigma_{e4}\sigma_{e3} & \sigma^2_{e4}\end{bmatrix} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_{e1} & \tau^2_{U_0}+r_{e1}\sigma_{e1}\sigma_{e2} & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0}+r_{e1}\sigma_{e2}\sigma_{e1} & \tau^2_{U_0}+\sigma^2_{e2} & \tau^2_{U_0}+r_{e1}\sigma_{e2}\sigma_{e3} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0}+r_{e1}\sigma_{e3}\sigma_{e2} & \tau^2_{U_0}+\sigma^2_{e3} & \tau^2_{U_0}+r_{e1}\sigma_{e3}\sigma_{e4}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+r_{e1}\sigma_{e4}\sigma_{e3} & \tau^2_{U_0}+\sigma^2_{e4}\end{bmatrix}$$

Map of **R**-only and **G** and **R** ACS Models

Arrows indicate nesting (end is more complex model)



Stuff to Watch Out For…

- **If using a random intercept, don't forget to drop 1 parameter in:**
  - ***n*-1 order UN R**: Can't get all possible elements in **R**, plus $\tau^2_{U_0}$ in **G**
  - **TOEP*n*−1**: Have to eliminate last lag covariance

- If using a random intercept…
  - Can't do RI + CS **R**: Can't get a constant in **R**, and then another constant in **G**
  - Can often test if random intercept helps (e.g., AR1 is nested within RI + AR1)

- If "**time**" is treated as **continuous** in the fixed effects, you will need another variable for **time** that is **categorical** to use in the syntax:
  - "Continuous Time" → on MODEL statement
  - "Categorical Time" → on CLASS and REPEATED statements

- Most alternative covariance structure models assume **time is balanced across persons with equal intervals across occasions**
  - If not, holding correlations of same lag equal doesn't make sense
  - Other structures can be used for unbalanced time
    - SP(POW)(time) = AR1 for unbalanced time (see SAS REPEATED statement for others)

Summary: Two Families of ACS Models

- **R**-only models:
  - Specify **R** model on REPEATED statement without any random effects variances in **G** (so no RANDOM statement is used)
  - Include UN, CS, CSH, AR1, AR1H, TOEP*n*, TOEPH*n* (among others)
  - *Total* variance and *total* covariance kept in **R**, so **R** = **V**
  - Other than CS, does not partition total variance into BP vs. WP

- **G** and **R** combined models (so **G** and **R** → **V**):
  - Specify random intercept variance $\tau^2_{U_0}$ in **G** using RANDOM statement, then specify **R** model using REPEATED statement
  - **G** matrix = Level-2 BP variance and covariance due to U_{0i}, so **R** = Level-1 WP variance and covariance of the e_{ti} residuals
  - **R** models what's left after accounting for mean differences between persons (via the random intercept variance $\tau^2_{U_0}$ in **G**)

Syntax for Models for the Variance

- Does your model include r**andom intercept variance $\tau^2_{U_0}$ (for U_{0i})** ?
  - Use the **RANDOM** statement → **G matrix**
  - Random intercept models BP interindividual differences in mean Y
- What about **residual variance σ^2_e (for e_{ti})** ?
  - Use the **REPEATED** statement → **R matrix**
    - **WITHOUT a RANDOM statement: R is BP and WP variance together** = σ^2_T
      → Total variances and covariances (to model all variation, so **R** = **V**)
    - **WITH a RANDOM statement: R is WP variance only = σ^2_e**
      → Residual variances and covariances to model WP intraindividual variation
      → **G** and **R** put back together = **V matrix** of total variances and covariances
- The **REPEATED** statement is always there implicitly…
  - Any model **always** has at least one residual variance in **R** matrix
- But the **RANDOM** statement is only there if you write it
  - **G** matrix isn't always necessary (don't always need random intercept)

Wrapping Up: ACS Models

- Even if you just expect fluctuation over time rather than change, you still should be concerned about accurately predicting the variances and covariances across occasions

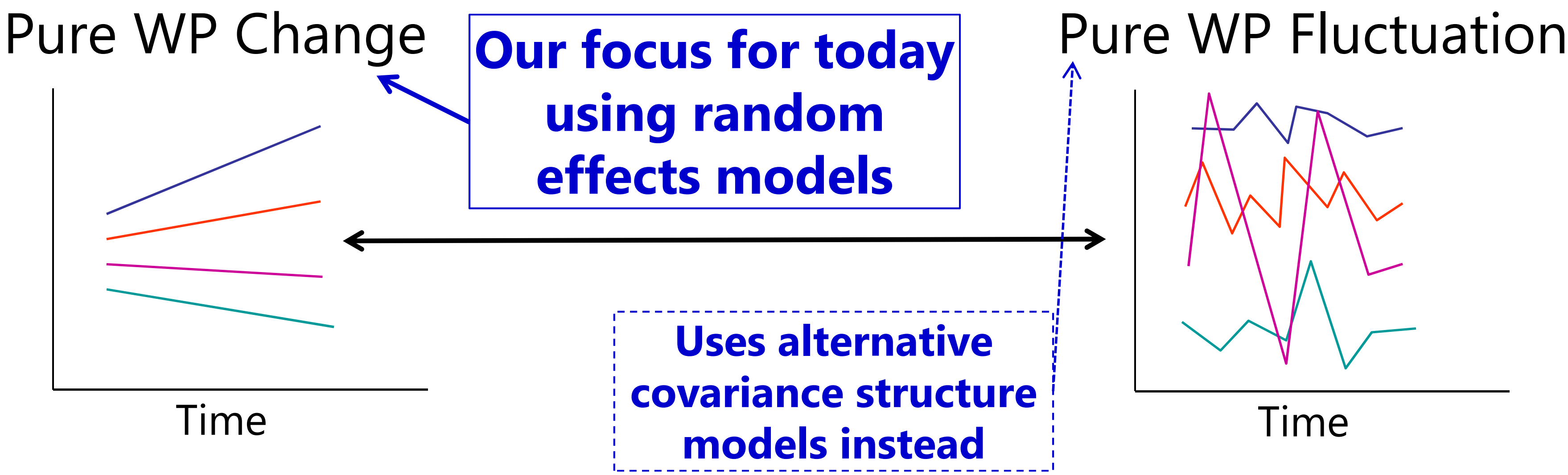
- Baseline models (from ANOVA least squares) are CS & UN:
  - Compound Symmetry: Equal variance and covariance over time
  - Unstructured: All variances & covariances estimated separately
  - CS and UN via ML or REML estimation allows missing data

- MLM gives us choices in the middle
  - Goal: Get as close to UN as parsimoniously as possible
  - **R**-only: Structure TOTAL variation in one matrix (**R** only)
  - **G**+**R**: Put constant covariance due to random intercept in **G**, then structural RESIDUAL covariance in **R** (so that **G** and **R** → **V** TOTAL)

Review of Multilevel Models for Longitudinal Data

- Topics:
  - Concepts in longitudinal multilevel modeling
  - Describing within-person fluctuation using ACS models
  - **Describing within-person change using random effects**
  - Likelihood estimation in random effects models
  - Describing nonlinear patterns of change
  - Time-invariant predictors

Modeling Change vs. Fluctuation



Model for the Means:

- **WP Change** → describe pattern of *average* change (over "time")
- WP Fluctuation → *may* not need anything (if no systematic change)

Model for the Variance:

- **WP Change** → describe *individual differences* in change (random effects)
  → this allows variances and covariances to differ over time
- WP Fluctuation → describe pattern of variances and covariances over time

The Big Picture of Longitudinal Data: **Models for the Means**

- What kind of change occurs on average over "time"? There are two baseline models to consider:
  - "**Empty**" → only a fixed intercept (predicts no change)
  - "**Saturated**" → all occasion mean differences from time 0 (ANOVA model that uses # fixed effects= *n*) *** *may not be possible in unbalanced data*

Parsimony ← **Empty Model:** Predicts NO change over time — **1 Fixed Effect**

In-between options: polynomial slopes, piecewise slopes, nonlinear models…

Saturated Means: Reproduces mean at each occasion — **# Fixed Effects = # Occasions** → **Good fit**

Name… that… Trajectory!

The Big Picture of Longitudinal Data: **Models for the Variance**

Compound Symmetry (CS)

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Most useful model: likely somewhere in between!

Unstructured (UN)

$$\begin{bmatrix} \sigma^2_{T1} & \sigma_{T12} & \sigma_{T13} & \sigma_{T14} \\ \sigma_{T21} & \sigma^2_{T2} & \sigma_{T23} & \sigma_{T24} \\ \sigma_{T31} & \sigma_{T32} & \sigma^2_{T3} & \sigma_{T43} \\ \sigma_{T41} & \sigma_{T42} & \sigma_{T43} & \sigma^2_{T4} \end{bmatrix}$$

Parsimony ⟵⟶ **Good fit**

Univariate RM ANOVA — ***Name …that … Structure!*** — Multivariate RM ANOVA

What is the pattern of variance and covariance over time?

CS and UN are just two of the many, many options available within MLM, including ***random effects models*** (for change) and ***alternative covariance structure models*** (for fluctuation).

Empty Means, Random Intercept Model

GLM Empty Model:

- $y_i = \beta_0 + e_i$

MLM Empty Model:

- Level 1:

  $$y_{ti} = \beta_{0i} + e_{ti}$$

- Level 2:

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

Fixed Intercept =grand mean (because no predictors yet)

Random Intercept = individual-specific deviation from predicted intercept

Residual = time-specific deviation from individual's predicted outcome

3 Total Parameters:

Model for the Means (1):

- Fixed Intercept γ_{00}

Model for the Variance (2):

- Level-1 Variance of e_{ti} → σ_e^2
- Level-2 Variance of U_{0i} → $\tau_{U_0}^2$

Composite equation:

$$y_{ti} = (\gamma_{00} + U_{0i}) + e_{ti}$$

Augmenting the empty means, random intercept model with *time*

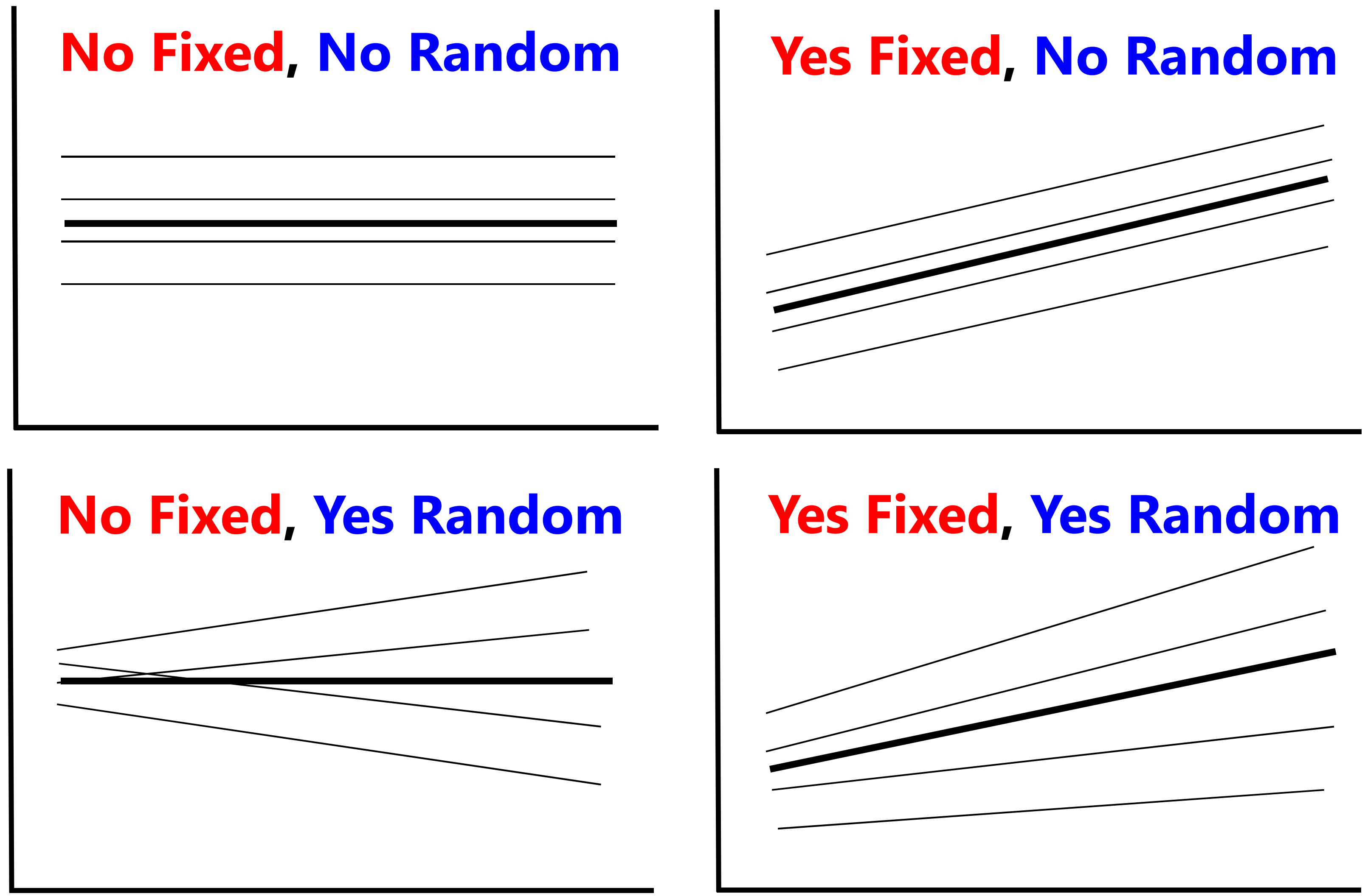
- 2 questions about the possible effects of *time*:

1. **Is there an effect of time on average?**
   - If the line describing the sample means not flat?
   - Significant **FIXED** effect of time

2. **Does the average effect of time vary across individuals?**
   - Does each individual need his or her own line?
   - Significant **RANDOM** effect of time

Fixed and Random Effects of Time

(Note: The intercept is random in every figure)



Fixed Linear Time, Random Intercept Model
(4 total parameters: effect of time is FIXED only)

Multilevel Model

Residual = time-specific deviation from individual's predicted outcome → estimated variance of σ_e^2

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

Fixed Intercept = predicted mean outcome at time 0

Fixed Linear Time Slope = predicted mean rate of change per unit time

Level 2: $\beta_{0i} = \gamma_{00} + U_{0i} \qquad \beta_{1i} = \gamma_{10}$

Random Intercept = individual-specific deviation from fixed intercept → estimated variance of $\tau_{U_0}^2$

Composite Model

$$y_{ti} = \underbrace{(\gamma_{00} + U_{0i})}_{\beta_{0i}} + \underbrace{(\gamma_{10})}_{\beta_{1i}}(Time_{ti}) + e_{ti}$$

Because the effect of time is **fixed**, everyone is predicted to change at exactly the same rate.

Explained Variance from Fixed Linear Time

- Most common measure of effect size in MLM is Pseudo-R^2
  - Is supposed to be variance accounted for by predictors
  - Multiple piles of variance mean multiple possible values of pseudo R^2 (can be calculated per variance component or per model level)
  - A fixed linear effect of time will reduce level-1 residual variance σ_e^2 in **R**
  - By how much is the residual variance σ_e^2 reduced?

$$\text{Pseudo } R_e^2 = \frac{\text{residual variance}_{\text{fewer}} - \text{residual variance}_{\text{more}}}{\text{residual variance}_{\text{fewer}}}$$

  - If time varies between persons, then level-2 random intercept variance $\tau_{U_0}^2$ in **G** may also be reduced:

$$\text{Pseudo } R_{U0}^2 = \frac{\text{random intercept variance}_{\text{fewer}} - \text{random intercept variance}_{\text{more}}}{\text{random intercept variance}_{\text{fewer}}}$$

  - But you are likely to see a (net) INCREASE in $\tau_{U_0}^2$ instead…. Here's why:

Increases in Random Intercept Variance

- Level-2 random intercept variance $\tau^2_{U_0}$ will often increase as a consequence of reducing level-1 residual variance σ^2_e
- Observed level-2 $\tau^2_{U_0}$ is NOT just between-person variance
  - Also has a small part of within-person variance (level-1 σ^2_e), or: **Observed $\tau^2_{U_0}$ = True $\tau^2_{U_0}$ + (σ^2_e/*n*)**
    - As *n* occasions increases, bias of level-1 σ^2_e is minimized
  - Likelihood-based estimates of "true" $\tau^2_{U_0}$ use (σ^2_e/*n*) as correction factor: **True $\tau^2_{U_0}$ = Observed $\tau^2_{U_0}$ – (σ^2_e/*n*)**
- For example: observed level-2 $\tau^2_{U_0}$=4.65, level-1 σ^2_e=7.06, *n*=4
  - True $\tau^2_{U_0}$= 4.65 −(**7.60**/4) = **2.88** in empty means model
  - Add fixed linear time slope → reduce σ^2_e from 7.06 to 2.17 (R^2 = .69)
  - But now True $\tau^2_{U_0}$= 4.65 −(**2.17**/4) = **4.10** in fixed linear time model

Random Intercept Models Imply…

- **People differ from each other systematically in only ONE way**— in intercept (U_{0i}), which implies **ONE kind of BP variance**, which translates to **ONE source of person dependency** (covariance or correlation in the outcomes from the same person)
- If so, after controlling for BP intercept differences (by estimating the variance of U_{0i} as $\tau^2_{U_0}$ in the **G** matrix), the **e_{ti} residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with homogeneous variance across time**, as shown:

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau^2_{U_0} \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

G and **R** matrices combine to create a total **V** matrix with CS pattern

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Matrices in a Random Intercept Model

Total predicted data matrix is called V matrix, created from the G [TYPE=UN] and R [TYPE=VC] matrices as follows:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R} = \mathbf{V}$$

$$\mathbf{V} = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\begin{bmatrix}\tau^2_{U_0}\end{bmatrix}\begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix} + \begin{bmatrix}\sigma^2_e & 0 & 0 & 0\\ 0 & \sigma^2_e & 0 & 0\\ 0 & 0 & \sigma^2_e & 0\\ 0 & 0 & 0 & \sigma^2_e\end{bmatrix} = \begin{bmatrix}\tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0}\\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e\end{bmatrix}$$

VCORR then provides the intraclass correlation, calculated as:
ICC = $\tau^2_{U_0} / (\tau^2_{U_0} + \sigma^2_e)$

$$\begin{bmatrix}1 & ICC & ICC & ICC\\ ICC & 1 & ICC & ICC\\ ICC & ICC & 1 & ICC\\ ICC & ICC & ICC & 1\end{bmatrix}$$

assumes a constant correlation over time

For any random effects model:

G matrix = BP variances/covariances

R matrix = WP variances/covariances

Z matrix = values of predictors with random effects (just intercept here), which can vary per person

V matrix = Total variance/covariance

Random Linear Time Model (6 total parameters)

Multilevel Model

Residual = time-specific deviation from individual's predicted outcome → estimated variance of σ_e^2

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}$

Fixed Intercept = predicted mean outcome at time 0

Fixed Linear Time Slope = predicted mean rate of change per unit time

Level 2: $\beta_{0i} = \gamma_{00} + U_{0i}$ $\quad \beta_{1i} = \gamma_{10} + U_{1i}$

Random Intercept = individual-specific deviation from fixed intercept at time 0 → estimated variance of $\tau_{U_0}^2$

Random Linear Time Slope = individual-specific deviation from fixed linear time slope → estimated variance of $\tau_{U_1}^2$

Also has an estimated **covariance** of random intercepts and slopes of $\tau_{U_{01}}$

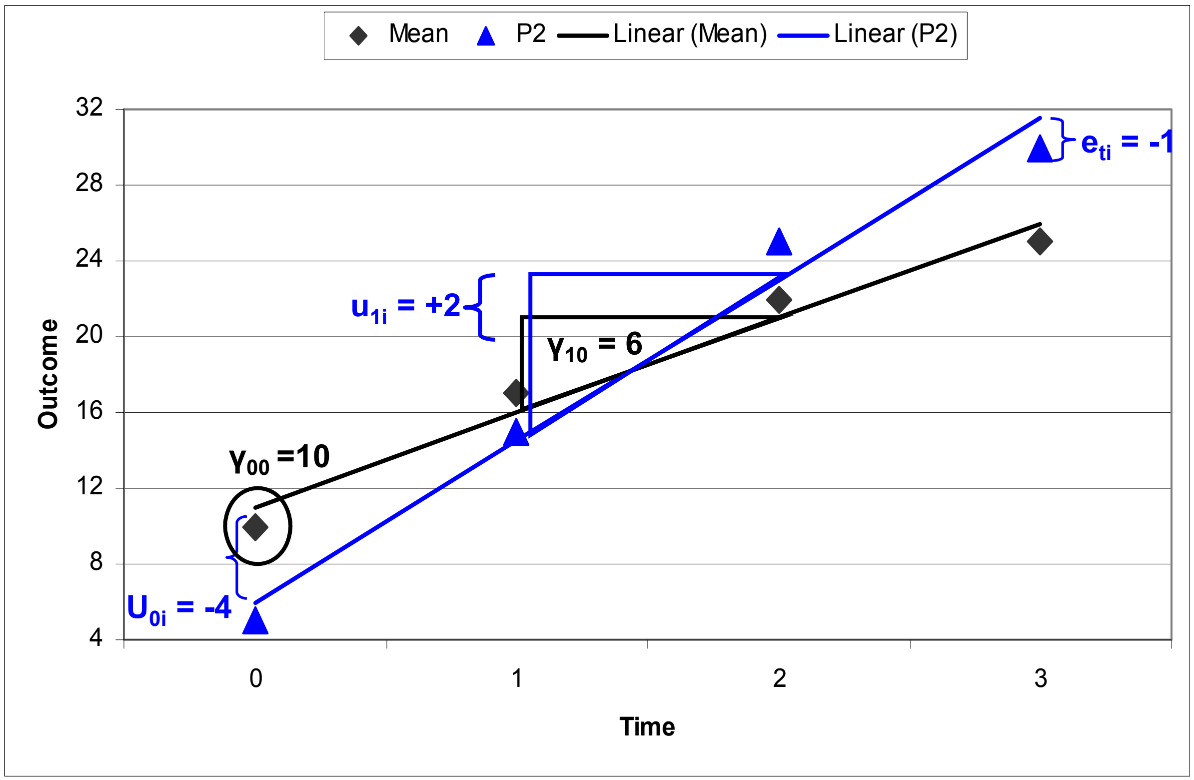
Composite Model

$$y_{ti} = \underbrace{(\gamma_{00} + U_{0i})}_{\beta_{0i}} + \underbrace{(\gamma_{10} + U_{1i})}_{\beta_{1i}}(Time_{ti}) + e_{ti}$$

Random Linear Time Model

$$y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})(Time_{ti}) + e_{ti}$$

- γ_{00}: Fixed Intercept
- U_{0i}: Random Intercept Deviation
- γ_{10}: Fixed Slope
- U_{1i}: Random Slope Deviation
- e_{ti}: error for person i at time t



6 Parameters:

2 Fixed Effects:

γ_{00} Intercept, γ_{10} Slope

2 Random Effects Variances:

U_{0i} Intercept Variance $= \tau^2_{U_0}$

U_{1i} Slope Variance $= \tau^2_{U_1}$

Int-Slope Covariance $= \tau_{U_{01}}$

1 e_{ti} Residual Variance $= \sigma^2_e$

Quantification of Random Effects Variances

- We can test if a random effect variance is significant, but the variance estimates are not likely to have inherent meaning
  - e.g., "I have a significant fixed linear time effect of $\gamma_{10} = 1.72$, so people increase by 1.72/time on average. I also have a significant random linear time slope variance of $\tau^2_{U_1} = 0.91$, so people need their own slopes (people change differently). But how much is a variance of **0.91**, really?"
- **95% Random Effects Confidence Intervals** can tell you
  - Can be calculated for each effect that is random in your model
  - Provide range around the fixed effect within which 95% of your sample is predicted to fall, based on your random effect variance:

$$\text{Random Effect 95\% CI} = \text{fixed effect} \pm \left(1.96 * \sqrt{\text{Random Variance}}\right)$$

$$\text{Linear Time Slope 95\% CI} = \gamma_{10} \pm \left(1.96 * \sqrt{\tau^2_{U_1}}\right) \rightarrow 1.72 \pm \left(1.96 * \sqrt{0.91}\right) = -0.15 \text{ to } 3.59$$

  - So although people improve on average, individual slopes are predicted to range from −0.15 to 3.59 (so some people may actually decline)

Random Linear Time Models Imply:

- **People differ from each other systematically in TWO ways**—in intercept (U_{0i}) and slope (U_{1i}), which implies **TWO kinds of BP variance**, which translates to **TWO sources of person dependency** (covariance or correlation in the outcomes from the same person)
- If so, after controlling for both BP intercept and slope differences (by estimating the $\tau^2_{U_0}$ and $\tau^2_{U_1}$ variances in the **G** matrix), the **e_{ti} residuals** (whose variance and covariance are estimated in the **R** matrix) should be **uncorrelated with homogeneous variance across time**, as shown:

Level-2 **G** matrix: RANDOM TYPE=UN

$$\begin{bmatrix} \tau^2_{U_0} & \tau_{U_{10}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix}$$

Level-1 **R** matrix: REPEATED TYPE=VC

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

G and **R** combine to create a total **V** matrix whose per-person structure depends on the specific time occasions each person has (very flexible for unbalanced time)

Random Linear Time Model

(6 total parameters: effect of time is now RANDOM)

- How the model predicts each element of the **V** matrix:

Level 1: $\mathbf{y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}}$

Level 2: $\mathbf{\beta_{0i} = \gamma_{00} + U_{0i}}$

$\mathbf{\beta_{1i} = \gamma_{10} + U_{0i}}$

Composite Model: $\mathbf{y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{0i})(Time_{ti}) + e_{ti}}$

Predicted *Time-Specific* Variance:

$$\mathrm{Var}\left[y_{ti}\right] = \mathrm{Var}\left[\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10} + U_{1i}\right)\left(\mathrm{Time}_i\right) + e_{ti}\right]$$

$$= \mathrm{Var}\left[\left(U_{0i}\right) + \left(U_{1i} * \mathrm{Time}_i\right) + e_{ti}\right]$$

$$= \left\{\mathrm{Var}\left(U_{0i}\right)\right\} + \left\{\mathrm{Var}\left(U_{1i} * \mathrm{Time}_i\right)\right\} + \left\{2 * \mathrm{Cov}\left(U_{0i}, U_{1i} * \mathrm{Time}_i\right)\right\} + \left\{\mathrm{Var}\left(e_{ti}\right)\right\}$$

$$= \left\{\mathrm{Var}\left(U_{0i}\right)\right\} + \left\{\mathrm{Time}_i^2 * \mathrm{Var}\left(U_{1i}\right)\right\} + \left\{2 * \mathrm{Time}_i * \mathrm{Cov}\left(U_{0i}, U_{1i}\right)\right\} + \left\{\mathrm{Var}\left(e_{ti}\right)\right\}$$

$$= \left\{\tau_{U_0}^2\right\} + \boxed{\left\{\mathrm{Time}_i^2 * \tau_{U_1}^2\right\} + \left\{2 * \mathrm{Time}_i * \tau_{U_{01}}\right\}} + \left\{\sigma_e^2\right\}$$

Random Linear Time Model
(6 total parameters: effect of time is now RANDOM)

- How the model predicts each element of the **V** matrix:

Level 1: $\mathbf{y_{ti} = \beta_{0i} + \beta_{1i}(Time_{ti}) + e_{ti}}$

Level 2: $\mathbf{\beta_{0i} = \gamma_{00} + U_{0i}}$

$\mathbf{\beta_{1i} = \gamma_{10} + U_{1i}}$

Composite Model: $\mathbf{y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})(Time_{ti}) + e_{ti}}$

Predicted *Time-Specific* Covariances (Time A with Time B):

$$\mathrm{Cov}\left[y_{Ai}, y_{Bi}\right] = \mathrm{Cov}\left[\left\{\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10} + U_{1i}\right)\left(A_i\right) + e_{Ai}\right\}, \left\{\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10} + U_{1i}\right)\left(B_i\right) + e_{Bi}\right\}\right]$$

$$= \mathrm{Cov}\left[\left\{U_{0i} + \left(U_{1i}A_i\right)\right\}, \left\{U_{0i} + \left(U_{1i}B_i\right)\right\}\right]$$

$$= \mathrm{Cov}\left[U_{0i}, U_{0i}\right] + \mathrm{Cov}\left[U_{0i}, U_{1i}B_i\right] + \mathrm{Cov}\left[U_{0i}, U_{1i}A_i\right] + \mathrm{Cov}\left[U_{1i}A_i, U_{1i}B_i\right]$$

$$= \left\{\mathrm{Var}\left(U_{0i}\right)\right\} + \left\{\left(A_i + B_i\right) * \mathrm{Cov}\left(U_{0i}, U_{1i}\right)\right\} + \left\{\left(A_i B_i\right)\mathrm{Var}\left(U_{1i}\right)\right\}$$

$$= \left\{\tau^2_{U_0}\right\} + \boxed{\left\{\left(A_i + B_i\right)\tau_{U_{01}}\right\} + \left\{\left(A_i B_i\right)\tau^2_{U_1}\right\}}$$

Random Linear Time Model

(6 total parameters: effect of time is now RANDOM)

- Scalar "mixed" model equation per person:

$$\mathbf{Y}_i = \mathbf{X}_i * \boldsymbol{\gamma} + \mathbf{Z}_i * \mathbf{U}_i + \mathbf{E}_i$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} U_{0i} \\ U_{1i} \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) \\ \gamma_{00} + \gamma_{10}(1) \\ \gamma_{00} + \gamma_{10}(2) \\ \gamma_{00} + \gamma_{10}(3) \end{bmatrix} + \begin{bmatrix} U_{0i} + U_{1i}(0) \\ U_{0i} + U_{1i}(1) \\ U_{0i} + U_{1i}(2) \\ U_{0i} + U_{1i}(3) \end{bmatrix} + \begin{bmatrix} e_{0i} \\ e_{1i} \\ e_{2i} \\ e_{3i} \end{bmatrix}$$

$$\begin{bmatrix} y_{0i} \\ y_{1i} \\ y_{2i} \\ y_{3i} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{10}(0) + U_{0i} + U_{1i}(0) + e_{0i} \\ \gamma_{00} + \gamma_{10}(1) + U_{0i} + U_{1i}(1) + e_{1i} \\ \gamma_{00} + \gamma_{10}(2) + U_{0i} + U_{1i}(2) + e_{2i} \\ \gamma_{00} + \gamma_{10}(3) + U_{0i} + U_{1i}(3) + e_{3i} \end{bmatrix}$$

$\mathbf{X}_i$ = *n* x *k* values of **predictors with fixed effects**, so can differ per person (*k* = 2: intercept, linear time)

$\boldsymbol{\gamma}$ = *k* x 1 estimated **fixed effects**, so will be the same for all persons (γ_{00} = intercept, γ_{10} = linear time)

$\mathbf{Z}_i$ = *n* x *u* values of **predictors with random effects**, so can differ per person (*u* = 2: intercept, linear time)

$\mathbf{U}_i$ = *u* x 2 estimated individual **random effects**, so can differ per person

$\mathbf{E}_i$ = *n* x *n* time-specific residuals, so can differ per person

Random Linear Time Model
(6 total parameters: effect of time is now RANDOM)

- Predicted total variances and covariances per person:

$$\mathbf{V}_i = \mathbf{Z}_i * \mathbf{G}_i * \mathbf{Z}_i^T + \mathbf{R}_i$$

$$\mathbf{V}_i = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

$\mathbf{V}_i$ matrix: Variance$[y_{time}]$

$$= \tau^2_{U_0} + \left[(time)^2 \tau^2_{U_1}\right] + \left[2(time)\tau_{U_{01}}\right] + \sigma^2_e$$

$\mathbf{V}_i$ matrix = complicated ☺

$\mathbf{V}_i$ matrix: Covariance$[y_A, y_B]$

$$= \tau^2_{U_0} + \left[(A+B)\tau_{U_{01}}\right] + \left[(AB)\tau^2_{U_1}\right]$$

$\mathbf{Z}_i$ = *n* x *u* values of **predictors with random effects**, so can differ per person (*u* = 2: int., time slope)

$\mathbf{Z}_i^T$ = *u* x *n* values of predictors with random effects (just $\mathbf{Z}_i$ transposed)

$\mathbf{G}_i$ = *u* x *u* estimated **random effects variances and covariances**, so will be the same for all persons ($\tau^2_{U_0}$ = int. var., $\tau^2_{U_1}$ = slope var.)

$\mathbf{R}_i$ = *n* x *n* **time-specific residual variances and covariances**, so will be same for all persons (here, just diagonal σ^2_e)

Building **V** across persons: Random Linear Time Model

- **V** for two persons with **unbalanced time** observations:

$$\mathbf{V} = \mathbf{Z} \quad * \quad \mathbf{G} \quad * \quad \mathbf{Z}^{T} \quad + \quad \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 2.3 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 2.3 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- The giant combined **V** matrix across persons is how the multilevel or mixed model is actually estimated
- Known as "**block diagonal**" structure → predictions are given for each person, but 0's are given for the elements that describe relationships between persons (because persons are supposed to be independent here!)

Building **V** across persons: Random Linear Time Model

- **V** for two persons also with ***different n*** per person:

$$\mathbf{V} = \mathbf{Z} * \mathbf{G} * \mathbf{Z}^T + \mathbf{R}$$

$$\mathbf{V} = \begin{bmatrix} 1 & 0.0 & 0 & 0 \\ 1 & 1.0 & 0 & 0 \\ 1 & 2.0 & 0 & 0 \\ 1 & 3.0 & 0 & 0 \\ 0 & 0 & 1 & 0.2 \\ 0 & 0 & 1 & 1.4 \\ 0 & 0 & 1 & 3.5 \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} & 0 & 0 \\ \tau_{U_{01}} & \tau^2_{U_1} & 0 & 0 \\ 0 & 0 & \tau^2_{U_0} & \tau_{U_{01}} \\ 0 & 0 & \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0.0 & 1.0 & 2.0 & 3.0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0.2 & 1.4 & 3.5 \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2_e & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

- The “block diagonal” does not need to be the same size or contain the same time observations per person…
- **R** matrix can also include non-0 covariance or differential residual variance across time (as in ACS models), although the models based on the idea of a “lag” won’t work for unbalanced or unequal-interval time

G, **R**, and **V**: The Take-Home Point

- The partitioning of variance into piles…
  - **Level 2 = BP → G** matrix of random effects variances/covariances
  - **Level 1 = WP → R** matrix of residual variances/covariances
  - **G** and **R** combine via **Z** to create **V** matrix of total variances/covariances
  - Many flexible options that allows the variances and covariances to vary in a time-dependent way that better matches the actual data
    - Can allow differing variance and covariance due to other predictors, too

Compound Symmetry (CS)

$$\begin{bmatrix} \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e & \tau^2_{U_0} \\ \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0} & \tau^2_{U_0}+\sigma^2_e \end{bmatrix}$$

Random effects models use **G** and **R** to predict something in-between!

Unstructured (UN)

$$\begin{bmatrix} \sigma^2_{T1} & \sigma_{T12} & \sigma_{T13} & \sigma_{T14} \\ \sigma_{T21} & \sigma^2_{T2} & \sigma_{T23} & \sigma_{T24} \\ \sigma_{T31} & \sigma_{T32} & \sigma^2_{T3} & \sigma_{T43} \\ \sigma_{T41} & \sigma_{T42} & \sigma_{T43} & \sigma^2_{T4} \end{bmatrix}$$

Parsimony ← Univariate RM ANOVA ——— Multivariate RM ANOVA → **Good fit**

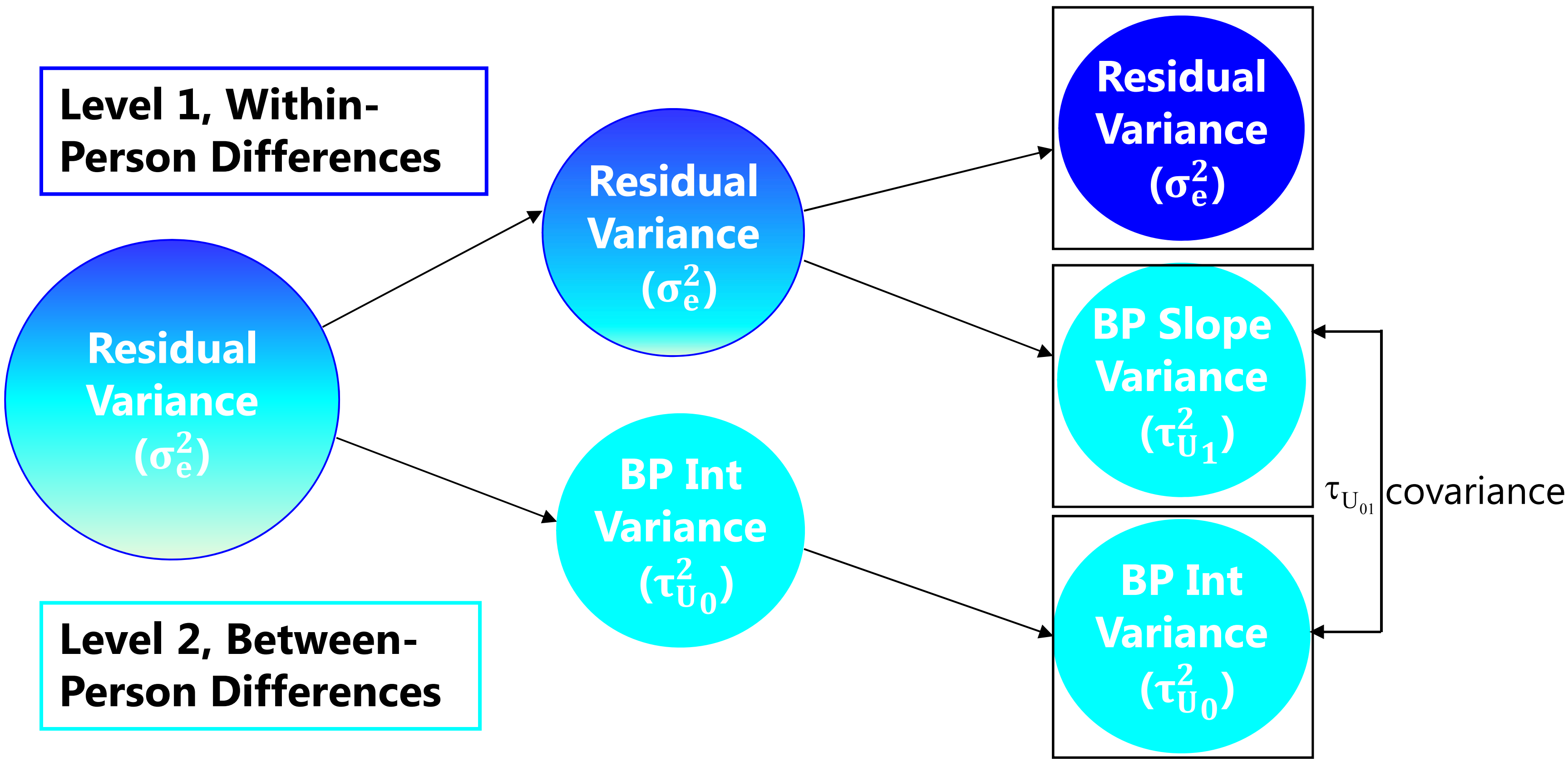
How MLM "Handles" Dependency

- Common description of the purpose of MLM is that it "addresses" or "handles" correlated (dependent) data…
- But where does this correlation come from?
  3 places (here, an example with health as an outcome):

1. *Mean differences across persons*
   - Some people are just healthier than others (at every time point)
   - This is what a random intercept is for
2. *Differences in effects of predictors across persons*
   - Does *time* (or *stress*) affect health more in some persons than others?
   - This is what random slopes are for
3. Non-constant within-person correlation for unknown reasons
   - Occasions closer together may just be more related
   - This is what alternative covariance structure models are for

MLM "Handles" Dependency

- Where does each kind of person dependency go? Into a new random effects variance component (or "pile" of variance):



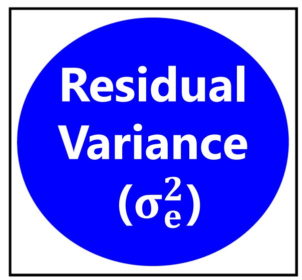
Piles of Variance

- By adding a random slope, we **carve up** our total variance into 3 piles:
  - BP (error) variance around intercept
  - BP (error) variance around slope
  - WP (error) residual variance

  These 2 piles are 1 pile of "error variance" in Univ. RM ANOVA

- **But making piles does NOT make error variance go away…**

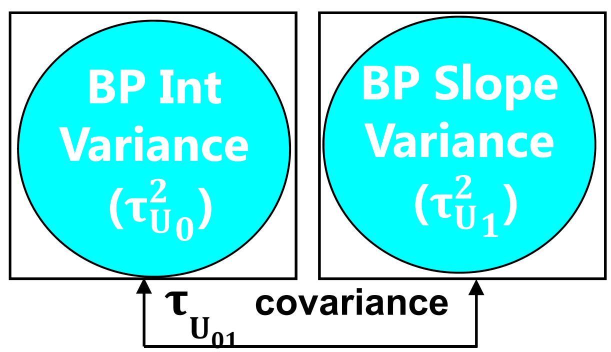
Level 1 (one source of) <u>Within-Person Variation</u>: gets accounted for by time-level predictors



FIXED effects make variance go away (explain variance).

RANDOM effects just make a new pile of variance.

Level 2 (two sources of) <u>Between-Person Variation</u>: gets accounted for by person-level predictors



Fixed vs. Random Effects of Persons

- Person dependency: via **fixed effects in the model for the means** or via **random effects in the model for the variance**?
  - Individual intercept differences can be included as:
    - **N-1 person dummy code fixed main effects OR 1 random U_{0i}**
  - Individual time slope differences can be included as:
    - **N-1*time person dummy code interactions OR 1 random U_{1i}*$time_{ti}$**
  - Either approach would appropriately control for dependency (fixed effects are used in some programs that 'control' SEs for sampling)
- Two important advantages of **random effects**:
  - Quantification: Direct measure of how much of the outcome variance is due to person differences (in intercept or in effects of predictors)
  - Prediction: Person differences (main effects and effects of time) then become predictable quantities – this can't happen using fixed effects
  - **Summary: Random effects give you *predictable* control of dependency**

Review of Multilevel Models for Longitudinal Data

- Topics:
  - Concepts in longitudinal multilevel modeling
  - Describing within-person fluctuation using ACS models
  - Describing within-person change using random effects
  - **Likelihood estimation in random effects models**
  - Describing nonlinear patterns of change
  - Time-invariant predictors

3 Decision Points for Model Comparisons

1. Are the models **nested** or **non-nested**?
   - Nested: have to add OR subtract effects to go from one to other
     - Can conduct significance tests for improvement in fit
   - Non-nested: have to add AND subtract effects
     - No significance tests available for these comparisons

2. Differ in model for the **means**, **variances**, or **both**?
   - Means? Can only use ML −2ΔLL tests (or *p*-value of each fixed effect)
   - Variances? Can use ML (or preferably REML) −2ΔLL tests, no *p*-values
   - Both sides? Can only use ML −2ΔLL tests

3. Models estimated using **ML** or **REML**?
   - ML: All model comparisons are ok
   - REML: Model comparisons are ok for the variance parameters only

Likelihood-Based Model Comparisons

- **Relative model fit** is indexed by a "**deviance**" statistic → **−2LL**
  - **−2LL indicates BADNESS of fit, so smaller values = better models**
  - **Two estimation flavors** (given as −2 log likelihood in SAS, SPSS, but given as LL instead in STATA): Maximum Likelihood (**ML**) or Restricted (Residual) ML (**REML**)

- **Nested models are compared using their deviance values: −2ΔLL Test** (i.e., Likelihood Ratio Test, Deviance Difference Test)

1. Calculate −2ΔLL: $(-2LL_{fewer}) - (-2LL_{more})$
2. Calculate Δdf: $(\#\ Parms_{more}) - (\#\ Parms_{fewer})$
3. Compare −2ΔLL to χ^2 distribution with df = Δdf
   CHIDIST in excel will give exact p-values for the difference test; so will STATA

1. & 2. must be positive values!

- Nested or non-nested models can also be compared by **Information Criteria** that reflect **−2LL** AND # parameters used and/or sample size
  - **AIC** = Akaike IC = **−2LL** + 2 *(#parameters)
  - **BIC** = Bayesian IC = **−2LL** + log(N)*(#parameters) → penalty for complexity
  - No significance tests or critical values, just "smaller is better"

ML vs. REML

Remember "population" vs. "sample" formulas for calculating variance?

$$\text{Population: } \sigma_e^2 = \frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N} \qquad \text{Sample: } \sigma_e^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}$$

All comparisons must have same N!!!	**ML**	**REML**
To select, type…	METHOD=ML (-2 log likelihood)	METHOD=REML *default* (-2 res log likelihood)
In estimating variances, it treats fixed effects as…	**Known** (df for having to also estimate fixed effects is not factored in)	**Unknown** (df for having to estimate fixed effects is factored in)
So, in small samples, L2 variances will be…	**Too small** (by a factor of (N – k) / N, N=# persons)	**Unbiased** (correct)
But because it indexes the fit of the…	**Entire model** (means + variances)	**Variances model only**
You can compare models differing in…	**Fixed and/or random effects** (either/both)	**Random effects only** (same fixed effects)

Rules for Comparing Multilevel Models

All observations must be the same across models!

Compare Models Differing In:

Type of Comparison:	**Means Model (Fixed) Only**	**Variance Model (Random) Only**	**Both Means and Variance Model (Fixed and Random)**
Nested? YES, can do significance tests via…	Fixed effect *p*-values from ML or REML -- *OR* -- ML −2ΔLL only (NO REML −2ΔLL)	NO *p*-values REML −2ΔLL (ML −2ΔLL is ok if big N)	ML −2ΔLL only (NO REML −2ΔLL)
Non-Nested? NO signif. tests, instead see…	ML AIC, BIC (NO REML AIC, BIC)	REML AIC, BIC (ML ok if big N)	ML AIC, BIC only (NO REML AIC, BIC)

Nested = one model is a direct subset of the other

Non-Nested = one model is not a direct subset of the other

Summary: Model Comparisons

- Significance of **fixed effects** can be tested with EITHER their ***p*-values** OR **ML** −2ΔLL (LRT, deviance difference) tests
  - *p*-value → Is EACH of these effects significant? (fine under ML or REML)
  - ML −2ΔLL test → Does this SET of predictors make my model better?
  - *REML −2ΔLL tests are WRONG for comparing models differing in fixed effects*

- Significance of **random effects** can only be tested with **−2ΔLL tests** (preferably using REML; here ML is not wrong, but results in too small variance components and fixed effect SEs in smaller samples)
  - Can get *p*-values as part of output but *shouldn't* use them
  - #parms added (df) should always include the random effect covariances

- My recommended approach to building models:
  - Stay in REML (for best estimates), test new fixed effects with their *p*-values
  - THEN add new random effects, testing −2ΔLL against previous model

Two Sides of Any Model: Estimation

- **Fixed Effects in the Model for the Means:**
  - How the expected outcome for a given observation varies as a function of values on *known* predictor variables
  - Fixed effects predict the Y values per se *but are not parameters that are solved for iteratively in maximum likelihood estimation*

- **Random Effects in the Model for the Variances:**
  - How model residuals are related across observations (persons, groups, time, etc) – *unknown* things due to sampling
  - Random effects variances and covariances are a mechanism by which complex patterns of variance and covariance among the Y residuals can be predicted (not the Y values, but their dispersion)
  - Anything besides level-1 residual variance σ_e^2 must be solved for iteratively – increases the dimensionality of estimation process
  - Estimation utilizes the predicted **V** matrix for each person
  - In the material that follows, **V** will be based on a random linear model

End Goals of Maximum Likelihood Estimation

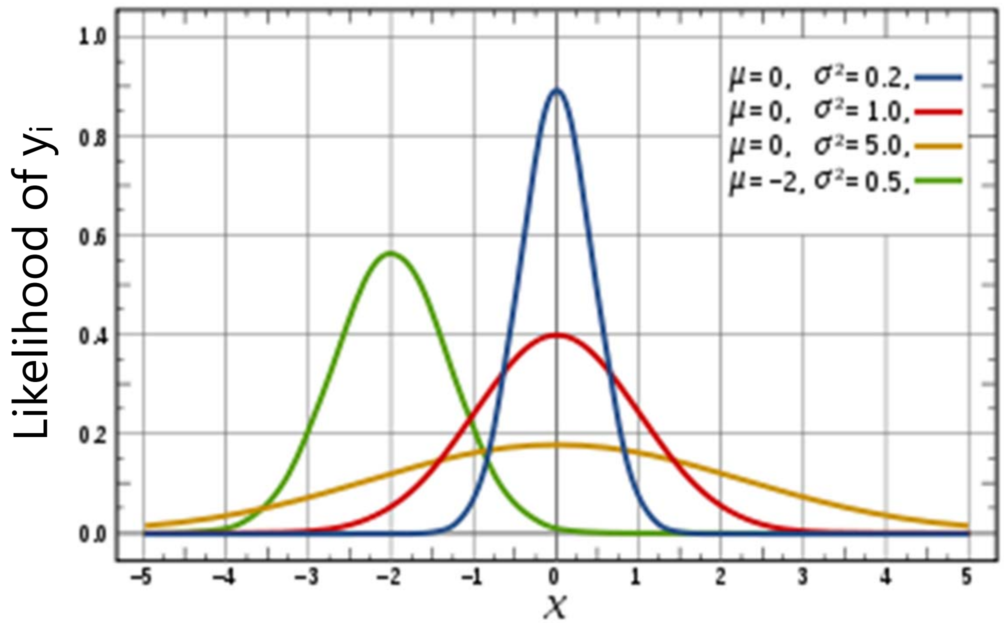
1. Obtain "most likely" values for each unknown model parameter (random effects variances and covariances, residual variances and covariances, which then are used to calculate the fixed effects) → **the estimates**

2. Obtain an index as to how likely each parameter value actually is (i.e., "really likely" or pretty much just a guess?) → **the standard error (SE) of the estimates**

3. Obtain an index as to how well the model we've specified actually describes the data → **the model fit indices**

How does all this happen? The magic of multivariate normal…(but let's start with univariate normal first)

Univariate Normal



- This function tells us how **likely** any value of y_i is given two pieces of info:
  - predicted value $\hat{y}_i$
  - residual variance σ_e^2

Univariate Normal PDF (two ways):

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma_e^2}} * \exp\left[-\frac{1}{2} * \frac{\left(y_i - \hat{y}_i\right)^2}{\sigma_e^2}\right]$$

$$f(y_i) = \left(2\pi\sigma_e^2\right)^{-1/2} * \exp\left[-\frac{1}{2} * \left(y_i - \hat{y}_i\right)\left(\sigma_e^2\right)^{-1}\left(y_i - \hat{y}_i\right)\right]$$

- Example: regression

$$y_i = \beta_0 + \beta_1 X_i + e_i$$

$$\hat{y}_i = \beta_0 + \beta_1 X_i$$

$$e_i = y_i - \hat{y}_i \qquad \sigma_e^2 = \frac{\sum_{i=1}^{N} e_i^2}{N-2}$$

Multivariate Normal for Y_i (height for all *n* outcomes for person *i*)

$$\text{Univariate Normal PDF: } f(y_i) = \left(2\pi\sigma_e^2\right)^{-1/2} * \exp\left[-\frac{1}{2} * \left(y_i - \hat{y}_i\right)\left(\sigma_e^2\right)^{-1}\left(y_i - \hat{y}_i\right)\right]$$

$$\text{Multivariate Normal PDF: } f(\mathbf{Y}_i) = \left(2\pi\right)^{-n/2} * \left|\mathbf{V}_i\right|^{-1/2} * \exp\left[-\frac{1}{2} * \left(\mathbf{Y}_i - \mathbf{X}_i\gamma\right)^T \left(\mathbf{V}_i\right)^{-1} \left(\mathbf{Y}_i - \mathbf{X}_i\gamma\right)\right]$$

- In a random linear time model, the only fixed effects (in $\boldsymbol{\gamma}$) that predict the $\mathbf{Y}_i$ outcome values are the fixed intercept and fixed linear time slope
- The model also gives us $\mathbf{V}_i$ → the model-predicted total variance and covariance matrix across the occasions, taking into account the time values
- Uses $|\mathbf{V}_i|$ = determinant of $\mathbf{V}_i$ = summary of *non-redundant* info
  - Reflects sum of variances across occasions controlling for covariances
- $(\mathbf{V}_i)^{-1}$ → matrix inverse → like dividing (so can't be 0 or negative)
  - $(\mathbf{V}_i)^{-1}$ must be "positive definite", which in practice means no 0 random variances and no out-of-bound correlations between random effects
  - Otherwise, software uses "generalized inverse" → questionable results

Now Try Some Possible Answers...
(e.g., for the 4 **V** parameters in this random linear model example)

- Plug $\mathbf{V}_i$ predictions into log-likelihood function, sum over persons:

$$L = \prod_{i=1}^{N} \left\{ (2\pi)^{-n/2} * \left|\mathbf{V}_i\right|^{-1/2} * \exp\left[-\frac{1}{2} \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)^T \left(\mathbf{V}_i\right)^{-1} \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right) \right] \right\}$$

$$LL = \sum_{i=1}^{N} \left\{ \left[-\frac{n}{2} \log(2\pi) \right] + \left[-\frac{1}{2} \log\left|\mathbf{V}_i\right| \right] + \left[-\frac{1}{2} \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)^T \left(\mathbf{V}_i\right)^{-1} \left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right) \right] \right\}$$

- Try one set of possible parameter values for $\mathbf{V}_i$, compute LL
- Try another possible set for $\mathbf{V}_i$, compute LL….
  - Different algorithms are used to decide which values to try given that each parameter has its own distribution → like an uncharted mountain
  - Calculus helps the program scale this multidimensional mountain
    - At the top, all first partial derivatives (linear slopes at that point) ≈ 0
    - Positive first partial derivative? Too *low*, try again. Negative? Too *high*, try again.
    - Matrix of partial first derivatives = "score function" = "gradient"
      (as in NLMIXED output for models with truly nonlinear effects)

End Goals 1 and 2: Model Estimates and SEs

- Process terminates (the model "converges") when the next set of tried values for Vi don't improve the LL very much…
  - e.g., SAS default convergence criteria = .00000001
  - Those are the values for the parameters that, relative to the other possible values tried, are "most likely" → the variance estimates

- But we need to know how trustworthy those estimates are…
  - Precision is indexed by the steepness of the multidimensional mountain, where steepness → more negative partial second derivatives
  - Matrix of partial second derivatives = "Hessian matrix"
  - Hessian matrix * -1 = "information matrix"
  - So steeper function = more information = more precision = smaller SE

$$\text{Each parameter SE} = \frac{1}{\sqrt{\text{information}}}$$

What about the Fixed Effects?

- Likelihood mountain does NOT include fixed effects as additional search dimensions (only variances and covariances that make $\mathbf{V}_i$)
- **Fixed effects are determined** given the parameters for $\mathbf{V}_i$:

$$\boldsymbol{\gamma} = \left\{ \sum_{i=1}^{N} \left(\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right) \right\}^{-1} \sum_{i=1}^{N} \left(\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{Y}_i \right), \qquad \mathrm{Cov}(\boldsymbol{\gamma}) = \left\{ \sum_{i=1}^{N} \left(\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right) \right\}^{-1}$$

All we need is $\mathbf{V}_i$ and the data: **X**, **Y**

γ = fixed effect estimates

Cov(γ) = γ sampling variance (SQRT of diagonal = SE)

- This is actually what happens in regular regression (GLM), too:

$$\text{GLM matrix solution: } \boldsymbol{\beta} = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \left(\mathbf{X}^T \mathbf{Y} \right), \qquad \mathrm{Cov}(\boldsymbol{\beta}) = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \sigma_e^2$$

$$\text{GLM scalar solution: } \beta = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}, \quad \mathrm{Cov}(\beta) = \frac{\sigma_e^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

- **Implication: fixed effects don't cause estimation problems**… (at least in general models with normally distributed residuals)

What about ML vs. REML?

$$\text{ML:}\quad LL = \left[-\frac{T-0}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\log\left|\mathbf{V}_i\right|\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)^T \mathbf{V}_i^{-1}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)\right]$$

$$\text{REML:}\ LL = \left[-\frac{T-k}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\log\left|\mathbf{V}_i\right|\right] + \left[-\frac{1}{2}\sum_{i=1}^{N}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)^T \mathbf{V}_i^{-1}\left(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma}\right)\right]$$

$$+\left[-\frac{1}{2}\log\left|\sum_{i=1}^{N}\mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i\right|\right]$$

where:

$$= \left[\frac{1}{2}\log\left|\left(\sum_{i=1}^{N}\mathbf{X}_i^T\mathbf{V}_i^{-1}\mathbf{X}_i\right)^{-1}\right|\right] \quad = \left[\frac{1}{2}\log\left|\text{Cov}(\boldsymbol{\gamma})\right|\right]$$

- Extra part in REML is the sampling variance of the fixed effects… it is added back in to account for uncertainty in estimating fixed effects
- REML maximizes the likelihood of the residuals specifically, so models with different fixed effects are not on the same scale and are not comparable
  - This is why you can't do -2ΔLL tests in REML when the models to be compared have different fixed effects → the model residuals are defined differently

End Goal #3: How well do the model predictions match the data?

- End up with ML or REML LL from predicting $\mathbf{V}_i$ → so how good is it?
- Absolute model fit assessment is only possible when the $\mathbf{V}_i$ matrix is organized the same for everyone – in other words, balanced data
  - Items are usually fixed, so can get absolute fit in CFA and SEM → χ^2 test is based on match between actual and predicted data matrix
  - Time is often a continuous variable, so no absolute fit provided in MLM (or in SEM when using random slopes or T-scores for unbalanced time)
    - Can compute absolute fit when the saturated means, unstructured variance model is estimable in ML → is -2ΔLL versus "perfect" model for time

- Relative model fit is given as **−2LL** in SAS, in which smaller is better
  - −2* needed to conduct "likelihood ratio" or "deviance difference" tests
  - Also information criteria:
    - **AIC**: -2LL + 2*(#parms) ; **BIC**: -2LL + log(N)*(#parms)
    - ML #parms = all parameters; REML #parms = variance model parameters only

What about testing variances > 0?

- $-2\Delta LL$ between two nested models is χ^2-distributed only when the added parameters do not have a boundary (like 0 or 1)
  - Ok for fixed effects (could be any positive or negative value)
  - NOT ok for tests of random effects variances (must be > 0)
  - Ok for tests of heterogeneous variances and covariances (extra parameters can be phrased as unbounded deviations)

- When testing addition of parameters that have a boundary, $-2\Delta LL$ will follow a **mixture** of χ^2 distributions instead
  - e.g., when adding random intercept variance (test > 0?)
    - When estimated as positive, will follow χ^2 with df=1
    - When estimated as negative… can't happen, will follow χ^2 with df=0

  - End result: **$-2\Delta LL$ will be too conservative in boundary cases**

Solutions for Boundary Problems when using −2ΔLL tests

- If adding random intercept variance only, use $p < .10$; $\chi^2(1) > 2.71$
  - Because $\chi^2(0) = 0$, can just cut *p*-value in half to get correct *p*-value

- If adding ONE random slope variance (and covariance with random intercept), can use mixture p-value from $\chi^2(1)$ and $\chi^2(2)$

$$\text{Mixture } p\text{-value} = 0.5 * \text{prob}(\chi_1^2 > -2\Delta\text{LL}) + 0.5 * \text{prob}(\chi_2^2 > -2\Delta\text{LL})$$

so critical χ^2 = 5.14, not 5.99

- However—using a 50/50 mixture assumes a diagonal information matrix for the random effects variances (assumes the estimated values for each are arrived at independently, which pry isn't the case)
- Two options for more complex cases:
  - Simulate data to determine actual mixture for calculating *p*-value
  - Accept that −2ΔLL is conservative in these cases, and use it anyway
    → I'm using ~ to acknowledge this: e.g., −2ΔLL(~2) > 5.99, $p < .05$

Estimation: The Grand Finale

- Estimation in MLM is all about the random effects variances and covariances
  - The more there are, the harder it is to find them (the more dimensions of the likelihood mountain there are to scale)
  - "Non-positive-definite" **G** matrix means "broken model"
  - Fixed effects are solved for after-the-fact, so they rarely cause estimation problems (at least in general models)
  - Individual random effects are not model parameters, but can be predicted after-the-fact (with some problems in doing so)

- Estimation comes in two flavors:
  - ML → maximize the data; compare any nested models
  - REML → maximize the residuals; compare models that differ in their model for the variance only

Review of Multilevel Models for Longitudinal Data

- Topics:
  - Concepts in longitudinal multilevel modeling
  - Describing within-person fluctuation using ACS models
  - Describing within-person change using random effects
  - Likelihood estimation in random effects models
  - **Describing nonlinear patterns of change**
  - Time-invariant predictors

Summary: Modeling Means and Variances

- We have two tasks in describing within-person change:

- **Choose a Model for the Means**
  - What kind of change in the outcome do we have **on average**?
  - What kind and how many **fixed effects** do we need to predict that mean change as parsimoniously but accurately as possible?

- **Choose a Model for the Variances**
  - What pattern do the variances and covariances of the outcome show over time because of **individual differences** in change?
  - What kind and how many **random effects** do we need to predict that pattern as parsimoniously but accurately as possible?

The Big Picture of Longitudinal Data: Model for the Means (Fixed Effects)

- What kind of change occurs on average over "time"?
  - What is the most appropriate **metric of time**?
    - Time in study (with predictors for BP differences in time)?
    - Time since birth (age)? Time to event (time since diagnosis)?
    - Measurement occasions need not be the same across persons or equally spaced (code time as exactly as possible)
  - What kind of **theoretical process** generated the observed trajectories, and thus what kind of model do we need?
    - Linear or nonlinear? Continuous or discontinuous? Does change keep happening or does it eventually stop?
    - Many options: polynomial, piecewise, and nonlinear families

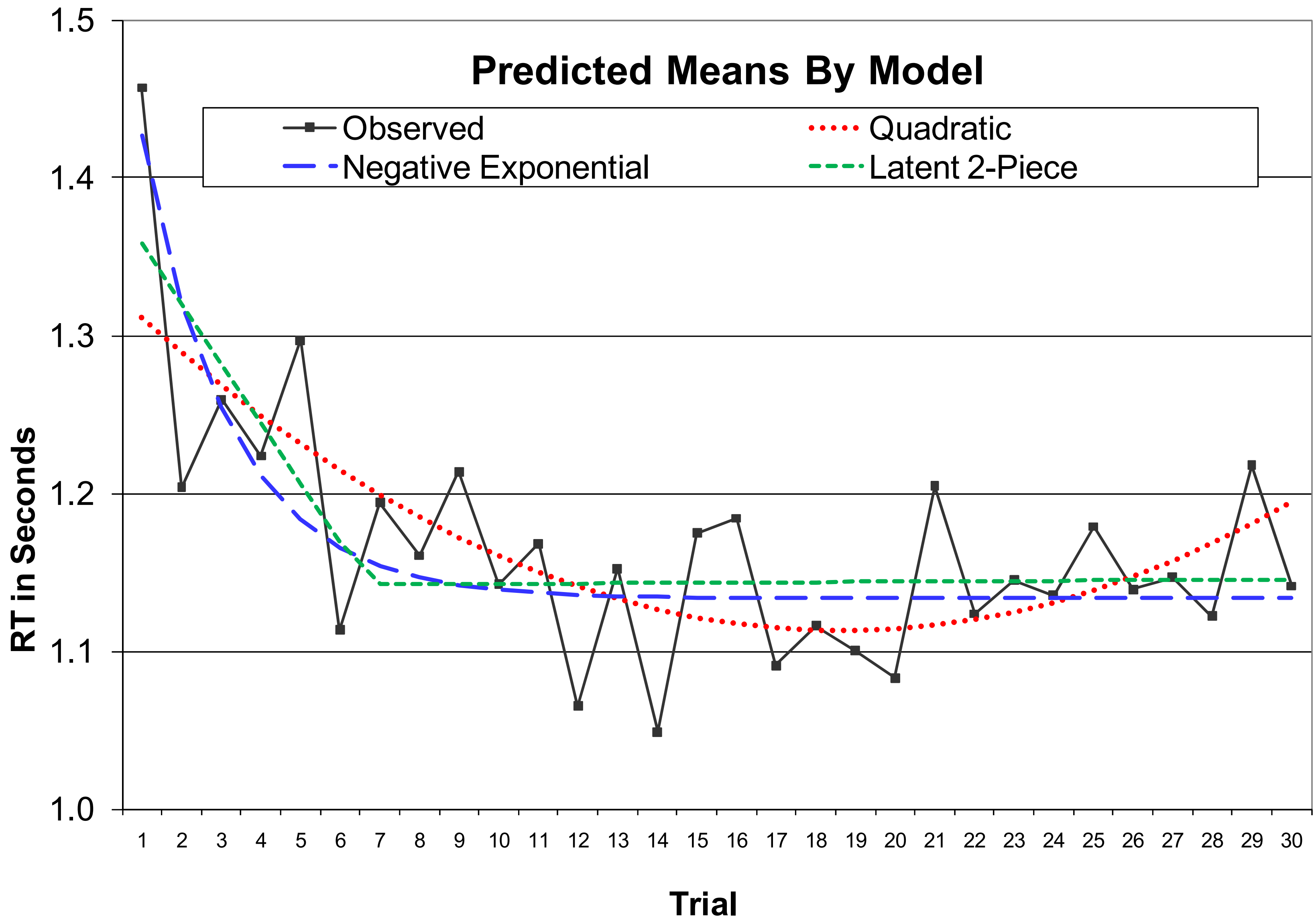
Options for Nonlinear Change*

- Polynomial models (linear, quadratic, cubic…)
  - Twice the quadratic coefficient is how the linear rate of change *changes* per unit time (becomes more/less positive/negative)
- Piecewise models (*aka*, spline models)
  - Use discrete slopes to describe sections of time (e.g., before or after an event), each of which can be linear or nonlinear
- Truly nonlinear models (with asymptotes)
  - e.g., exponential or logistic change for learning)

- Complexity possible depends on your (balanced) data:
  - Fixed slopes for time → can have up to $n – 1$
  - Random slopes for time → can have up to $n – 2$

* These models are useful for predictors other than time…

Competing Models of Change



Which family should I choose?

- Within a given family of models of change, nested models can usually be compared to judge the need for each term
  - e.g., linear vs. quadratic? one slope vs. two slopes?
  - Usual nested model comparison rules apply (*p*-values for fixed effects, −2ΔLL tests for assessing random effects)

- Between families, however, alternative models of change may not be nested, so deciding among them can be tricky
  - e.g., quadratic vs. two-slope vs. exponential?
  - Use ML AIC and BIC to see what is "preferred" among the families
  - In balanced data, can also compare each alternative to a saturated means, UN model using ML as test of exact fit
  - Also consider plausibility of alternative models in terms of both data predictions and theoretical predictions in deciding

Review of Multilevel Models for Longitudinal Data

- Topics:
  - Concepts in longitudinal multilevel modeling
  - Describing within-person fluctuation using ACS models
  - Describing within-person change using random effects
  - Likelihood estimation in random effects models
  - Describing nonlinear patterns of change
  - **Time-invariant predictors**

Missing Data in MLM

- **Missing outcomes are assumed MAR**
  - Because the likelihood function is for predicted Y, just estimated on whatever Y responses a person does have (can be incomplete)
- **Missing time-varying predictors are MAR-to-MCAR ish**
  - Would be MCAR because X is not in the likelihood function (is Y given X instead), but other occasions may have predictors (so MAR-ish)
- **Missing time-invariant predictors are assumed MCAR**
  - Because the predictor would be missing for all occasions, whole people will be deleted (may lead to bias)

- Missingness on predictors can be accommodated:
  - In Multilevel SEM with certain assumptions (≈ outcomes then)
  - Via multilevel multiple imputation in Mplus v 6.0+ (but careful!)
    - Must preserve all effects of potential interest in imputation model, including random effects; $-2\Delta LL$ tests are not done in same way

Centering Predictors

- Very useful to center all predictors such that 0 is a meaningful value:
  - Same significance level of main effect, different interpretation of intercept
  - Different (more interpretable) main effects within higher-order interactions
    - With interactions, main effects = simple effects when other predictor = 0

- Choices for centering **continuous** predictors:
  - At Mean: Reference point is *average level of predictor within the sample*
    - Useful if predictor is on arbitrary metric (e.g., unfamiliar test)
  - Better → At Meaningful Point: Reference point is *chosen level of predictor*
    - Useful if predictor is already on a meaningful metric (e.g., age, education)

- Choices for centering **categorical** predictors:
  - Re-code group so that your chosen reference group = **reference (0) category!** (highest is the default in SAS and SPSS; lowest is default in STATA)
  - I do not recommend mean-centering categorical predictors (because who is at the mean of a categorical variable ?!?)

What Level-2 Predictors Do…

- The purpose of level-2 predictors in the model for the means is to moderate the effects of level-1 predictors
  - Main effects moderate the intercept
  - Interactions with time moderate the time slopes
  - Interactions with other time-varying predictors moderate their effect between persons

- In addition, level-2 predictors can be used to allow heterogeneity of variance (stay tuned for next week)
  - At level 2: different **G** matrix variances and covariances
  - At level 1: different **R** matrix variances and covariances

Variance Accounted For By Level-2 Time-Invariant Predictors

- **Fixed effects of level 2 predictors *by themselves*:**
  - L2 (BP) main effects (e.g., sex) reduce L2 (BP) random intercept variance
  - L2 (BP) interactions (e.g., sex by ed) also reduce L2 (BP) random intercept variance

- **Fixed effects of *cross-level interactions* (level 1* level 2):**
  - If the interacting level-1 predictor is <u>random</u>, any cross-level interaction with it will reduce its corresponding level-2 BP random slope variance
    - e.g., if *time* is random, then sex**time*, ed**time*, and sex*ed**time* can each reduce the random linear time slope variance
  - If the interacting level-1 predictor <u>not random</u>, any cross-level interaction with it will reduce the level-1 WP residual variance instead
    - e.g., if $time^2$ is fixed, then sex*$time^2$, ed*$time^2$, and sex*ed*$time^2$ will reduce the L1 (WP) residual variance → Different quadratic slopes from sex and ed will allow better trajectories, reduce the variance around trajectories

3 Types of Effects: Fixed, Random, and Systematically (Non-Randomly) Varying

Let's say we have a significant fixed linear effect of time. What happens after we test a sex*time interaction?

	Non-Significant Sex*Time effect?	Significant Sex*Time effect?
Random time slope initially **not** significant	Linear effect of time is **FIXED**	Linear effect of time is **systematically varying**
Random time initially sig, **not** sig. after sex*time	---	Linear effect of time is **systematically varying**
Random time initially sig, **still** sig. after sex*time	Linear effect of time is **RANDOM**	Linear effect of time is **RANDOM**

The effects of level-1 predictors (time-level) can be fixed, random, or systematically varying. The effects of level-2 predictors (person-level) can only be fixed or systematically varying (nothing to be random over…yet).

Variance Accounted for… For Real

- **Pseudo-R^2** is named that way for a reason… piles of variance can shift around, such that pseudo-R^2 can actually be negative
  - Sometimes a sign of model mis-specification
  - Hard to explain to readers when it happens!

- **One last simple alternative: Total R^2**
  - Generate model-predicted y's from fixed effects only (NOT including random effects) and correlate with observed y's
  - Then square correlation → total R^2
  - Total R^2 = total reduction in overall variance of y across levels
  - Can be "unfair" in models with large unexplained sources of variance

- MORAL OF THE STORY: Specify EXACTLY which kind of pseudo-R^2 you used—give the formula and the reference!!