# Introduction to Random Effects of Time and Model Estimation

- Topics:
  - **The Big Picture**
  - **Multilevel model notation**
  - Fixed vs. random slopes of time
  - Effect sizes for random effects
  - Handling dependency: fixed or random effects?
  - How MLM = SEM
  - Fun with maximum likelihood estimation

Modeling Change vs. Fluctuation

Pure WP Change

Our focus for the next 7 weeks or so

Pure WP Fluctuation
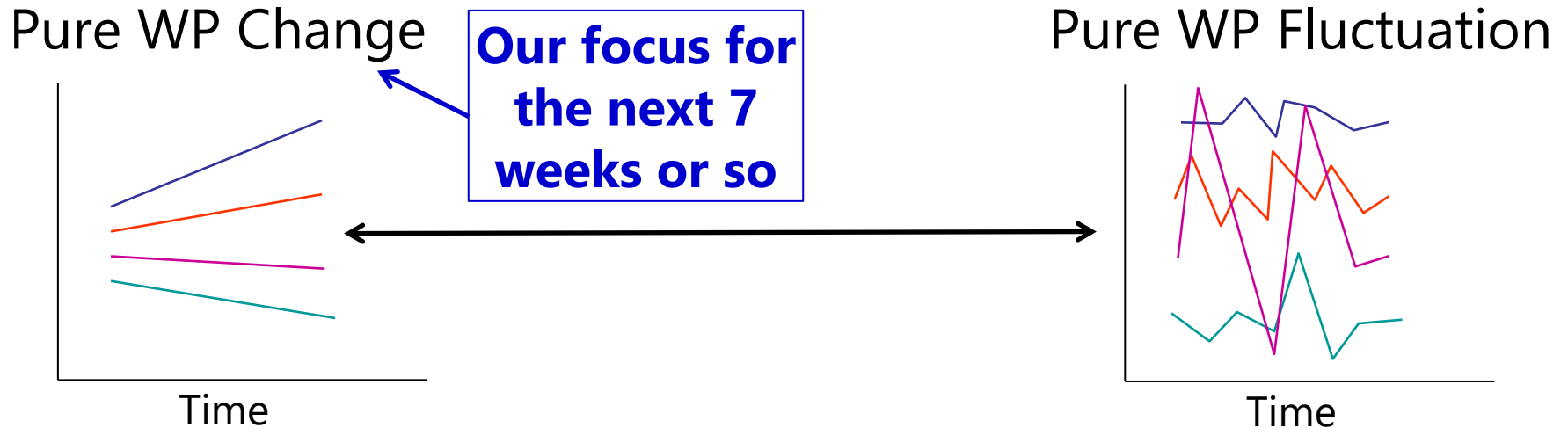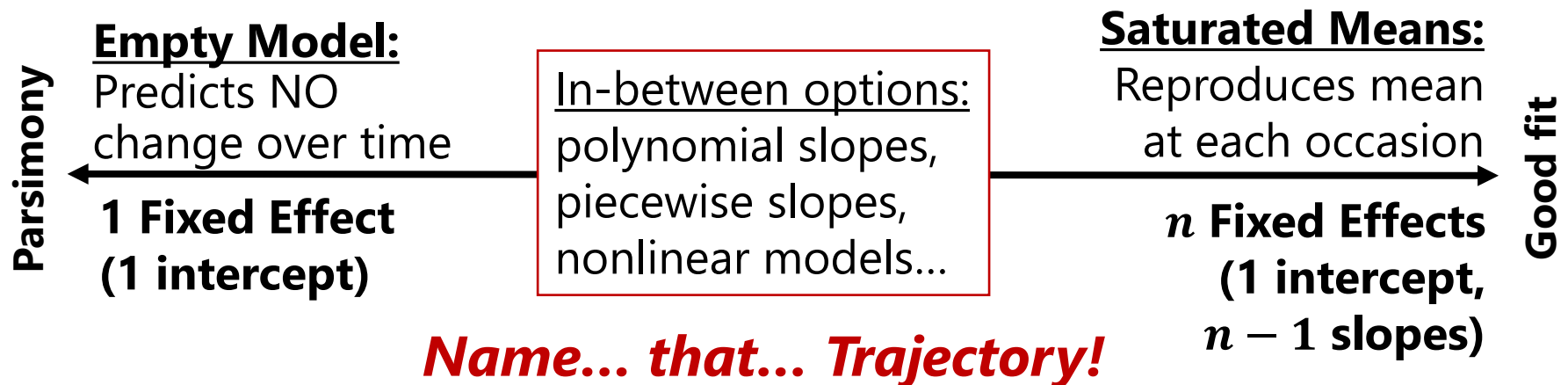

Time

Time

Model for the Means:

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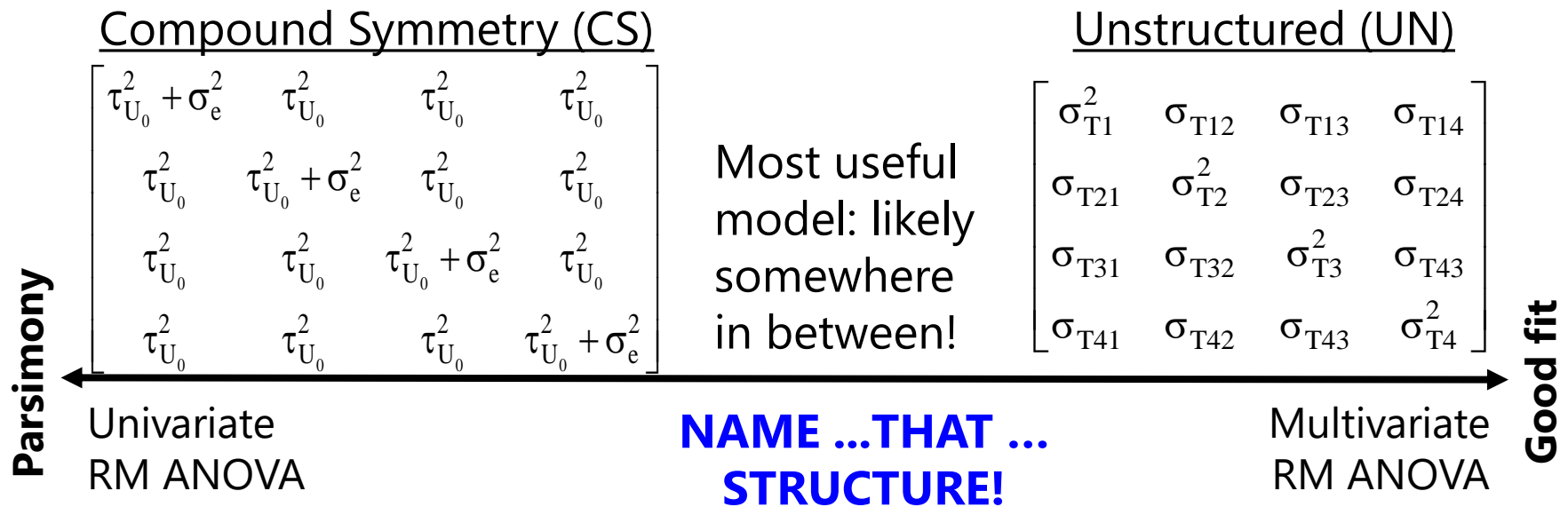
Model for the Variance:

- **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: Alternative Models for the Means

- What kind of change occurs on average over time? So far, we know of two baseline models:

  - "**Empty**" → only a fixed intercept (predicts no change)

  - "**Saturated**" → all occasion mean differences from time 0 (ANOVA model that uses n fixed effects) *** *may not be possible in unbalanced data*

Empty Model:
Predicts NO change over time

1 Fixed Effect (1 intercept)

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

Saturated Means:
Reproduces mean at each occasion

n **Fixed Effects (1 intercept, $n-1$ slopes)**

Parsimony ← → Good fit

Name... that... Trajectory!

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

Compound Symmetry (CS)

$$\begin{bmatrix} \tau_{U_0}^2 + \sigma_e^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 + \sigma_e^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 + \sigma_e^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 + \sigma_e^2 \end{bmatrix}$$

Most useful model: likely somewhere in between!

Unstructured (UN)

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

Parsimony ←————————————————————————————→ **Good fit**

Univariate
RM ANOVA

NAME …THAT … STRUCTURE!
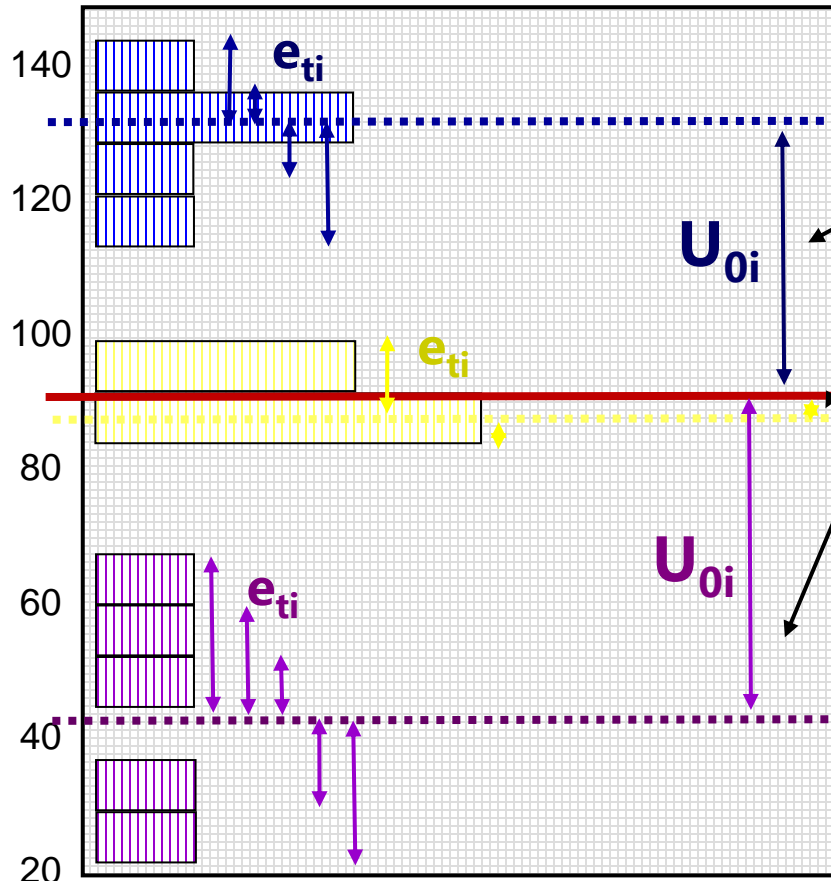
Multivariate
RM ANOVA

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, Two-Level Variance Model

y_{ti} variance → 2 sources:

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

→ **Between**-Person variance in means

→ **INTER**-Individual differences from **GRAND** mean to be explained by time-invariant predictors

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

→ **Within**-Person variance

→ **INTRA**-Individual differences from **OWN** mean to be explained by time-varying predictors

"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:

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

3 Total Parameters:

Model for the Means (1):

- Fixed Intercept γ_{00}

Model for the Variance (2):

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

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

Fixed Intercept =grand mean of person means (because no predictors yet)

Random Intercept = person-specific deviation from predicted intercept

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

"Saturated Means, Random Intercept" Model

- Although rarely shown this way, a saturated means, random intercept model would be represented as a multilevel model like this (for example $n = 4$ here, in which the time predictors are dummy codes to distinguish each occasion from time 0):

- Level 1:
$$y_{ti} = \beta_{0i} + \beta_{1i}(\text{Time1}_{ti}) + \beta_{2i}(\text{Time2}_{ti}) + \beta_{3i}(\text{Time3}_{ti}) + e_{ti}$$

- Level 2:
$$\beta_{0i} = \gamma_{00} + U_{0i}$$
$$\beta_{1i} = \gamma_{10}$$
$$\beta_{2i} = \gamma_{20}$$
$$\beta_{3i} = \gamma_{30}$$

Composite equation (6 parameters):
$$y_{ti} = \gamma_{00} + \gamma_{10}(\text{Time1}_{ti}) + \gamma_{20}(\text{Time2}_{ti}) + \gamma_{30}(\text{Time3}_{ti}) + U_{0i} + e_{ti}$$

Given the same random intercept model for the variance, the **G**, **R**, and **V** matrices would have the same form for the **empty means model** as for the **saturated means model** (but the latter would estimate remaining variance and covariance *after* controlling for all possible mean differences over time).

Random Intercept Model In Matrices

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

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

$$\mathbf{V} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \tau_{U_0}^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix} = \begin{bmatrix} \tau_{U_0}^2 + \sigma_e^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 + \sigma_e^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 + \sigma_e^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 + \sigma_e^2 \end{bmatrix}$$

VCORR then provides the intraclass correlation, calculated as:

$$\text{ICC} = \tau_{U_0}^2 / (\tau_{U_0}^2 + \sigma_e^2)$$

$$\begin{bmatrix} 1 & \text{ICC} & \text{ICC} & \text{ICC} \\ \text{ICC} & 1 & \text{ICC} & \text{ICC} \\ \text{ICC} & \text{ICC} & 1 & \text{ICC} \\ \text{ICC} & \text{ICC} & \text{ICC} & 1 \end{bmatrix}$$ implies a constant correlation over time

For any random intercept model: VCORR provides the "unconditional" ICC when requested from an **empty means** model. When paired with any other kind of means model (e.g., **saturated means** model), **VCORR** provides a "conditional" ICC instead (after controlling for the fixed effects).

Introduction to Random Effects of Time and Model Estimation

- Topics:
  - The Big Picture
  - Multilevel model notation
  - **Fixed vs. random slopes of time**
  - Effect sizes for random effects
  - Handling dependency: fixed or random effects?
  - How MLM = SEM
  - Fun with maximum likelihood estimation

Augmenting the Empty Means, Random Intercept Model with **Time**

- 2 questions about the possible effects of **"time"** (e.g., time in study in WP change; time of day or day of week in WP fluctuation):

1.   **Is there an effect of time on average?**

   ➢ Is the line connecting the sample means not flat?

   ➢ If so, you need **FIXED** effect(s) of time

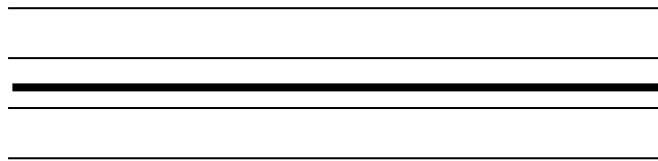2.   **Does the average effect of time vary across people?**

   ➢ Does each individual need their *own* version of that line?

   ➢ If so, you need **RANDOM** effect(s) of time

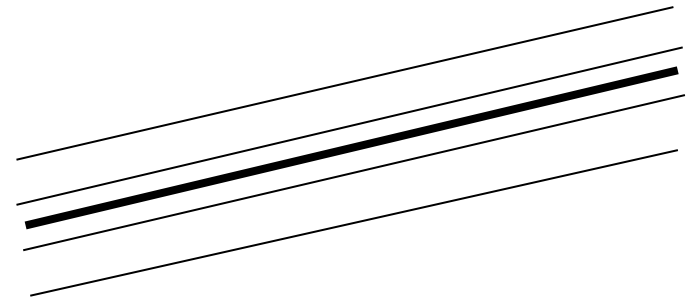- Let's look at examples using **linear time** effects to start…

Fixed and Random Effects of **Time**
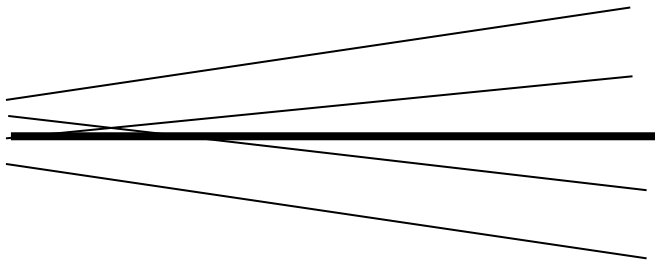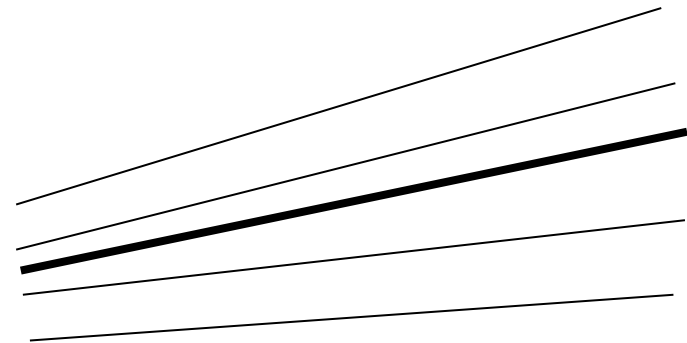## (Note: The intercept is random in every figure)

A. **No Fixed**, **No Random**

B. **Yes Fixed**, **No Random**

C. **No Fixed**, **Yes Random**

D. **Yes Fixed**, **Yes Random**

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

Multilevel Model

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

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

> **Fixed Intercept** = predicted mean outcome at time 0

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

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

> **Random Intercept** = person-specific deviation from fixed intercept → estimated variance is $\tau_{U_0}^2$

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

- Common measure of <u>effect size for fixed effects</u> is **pseudo-R²**

  - Used to assess variance accounted for by predictors, like usual, but...

  - Multiple piles of variance mean multiple possible values of pseudo R² (can be calculated *per variance component*)

  - A fixed linear effect of time will reduce level-1 residual variance $\boldsymbol{\sigma_e^2}$ in **R**

  - **By how much is the level-1 residual variance $\boldsymbol{\sigma_e^2}$ reduced?**

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

  - If time also varies between persons, then the level-2 random intercept variance $\boldsymbol{\tau_{U_0}^2}$ in **G** may also be reduced (see Hoffman 2015 ch. 10):

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

  - But you are likely to see a **(net) INCREASE in $\boldsymbol{\tau_{U_0}^2}$** 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:
    $$\text{Observed } \tau^2_{U_0} = \text{True } \tau^2_{U_0} + (\sigma^2_e/n)$$

    - With increasing n occasions, bias due to level-1 σ^2_e is minimized

  - Likelihood-based estimates of "**true**" $\tau^2_{U_0}$ use (σ^2_e/n) as correction factor:
    $$\text{True } \tau^2_{U_0} = \text{Observed } \tau^2_{U_0} - (\sigma^2_e/n)$$

- e.g., **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** −(**7.06**/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** → constant 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**

$$\begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$
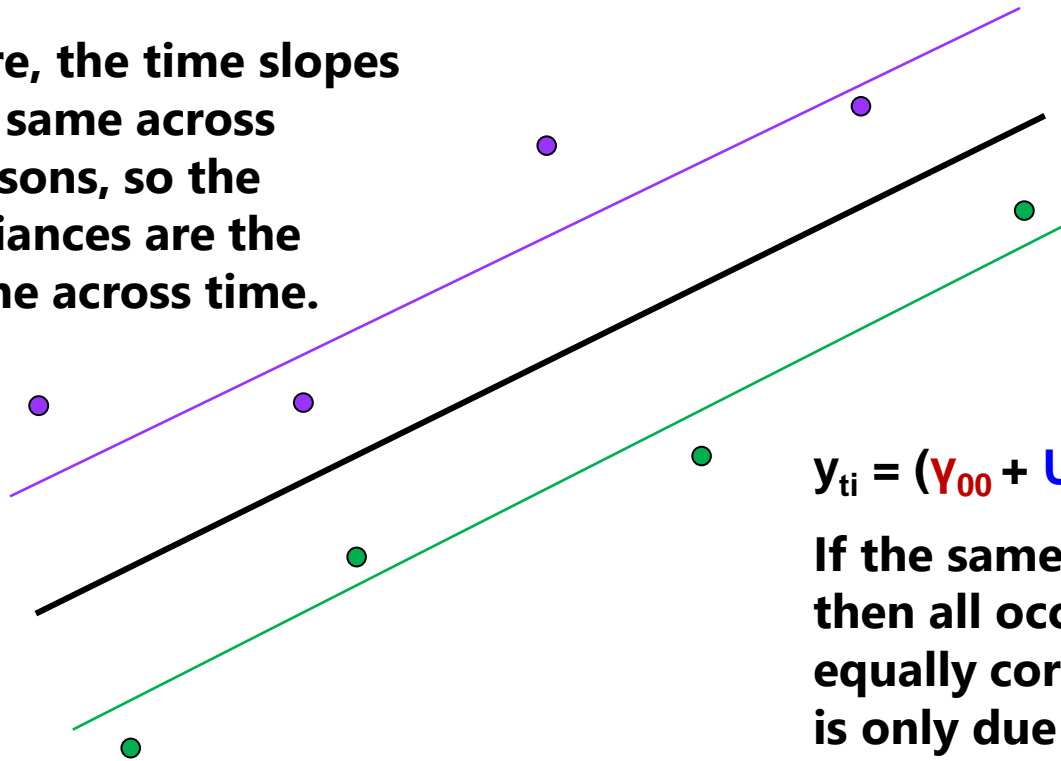
R is "conditional"

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

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

V is "marginal" (total)

Choices in Modeling Variances: Random Intercept Only (Compound Symmetry)

Here, the time slopes are same across persons, so the variances are the same across time.

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

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

If the same slope fits all persons, then all occasions should be equally correlated (correlation is only due to U_{0i} variance).

If the time slopes are the same across people, then people differ from each other systematically in only 1 way (i.e., their U_{0i} level) → THIS IS COMPOUND SYMMETRY.

B. Fixed Linear Time, Random Intercept Model
(4 total parameters: slope of time is **FIXED** only)

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

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

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

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

Predicted Variance per *Time*:

$$Var\left[y_{Time}\right]$$

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

$$= Var\left[U_{0i} + e_{ti}\right]$$

$$= \tau_{U_0}^2 + \sigma_e^2$$

Predicted Covariance (A,B):

$$Cov\left[y_A, y_B\right]$$

$$= Cov\left[\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10}\right)\left(A\right) + e_{ti}\right],$$

$$\left[\left(\gamma_{00} + U_{0i}\right) + \left(\gamma_{10}\right)\left(B\right) + e_{ti}\right]$$

$$= Cov\left[U_{0i}\right], \left[U_{0i}\right]$$

$$= \tau_{U_0}^2$$

B. Fixed Linear Time, Random Intercept Model
(4 total parameters: slope of time is **FIXED** only)

Model equation per person using matrices:

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

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

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

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

$\mathbf{X}_i = n$ 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 ($u = 1$: intercept)

$\mathbf{U}_i = u$ x 1 estimated individual **random effects**, so can differ per person

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

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

Predicted total variances and covariances per person:

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

$$\mathbf{V}_i = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} \tau_{U_0}^2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma_e^2 & 0 & 0 & 0 \\ 0 & \sigma_e^2 & 0 & 0 \\ 0 & 0 & \sigma_e^2 & 0 \\ 0 & 0 & 0 & \sigma_e^2 \end{bmatrix}$$

$$\mathbf{V}_i = \begin{bmatrix} \tau_{U_0}^2 + \sigma_e^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 + \sigma_e^2 & \tau_{U_0}^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 + \sigma_e^2 & \tau_{U_0}^2 \\ \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 & \tau_{U_0}^2 + \sigma_e^2 \end{bmatrix}$$

$\mathbf{Z}_i = n$ x u values of **predictors with random effects**, so can differ per person ($u = 1$: intercept)

$\mathbf{Z}_i^{T} = u$ 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 ($\tau_{U_0}^2$ = intercept variance)

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

$\mathbf{V}_i$: Variance$\left[y_{time} \right] = \tau_{U_0}^2 + \sigma_e^2$, Covariance$\left[y_A, y_B \right] = \tau_{U_0}^2$

Intermediate Summary

- Regardless of what kind of model for the means you have...

  - **Empty means** = 1 fixed intercept that predicts no change

  - **Saturated means** = 1 fixed intercept + $n - 1$ fixed effects for occasion mean differences that perfectly re-create the (MAR) means over time

    - Is a description, not a model, and may not be possible with unbalanced time

  - **Fixed linear time** = 1 fixed intercept, 1 fixed linear time slope that predicts linear average change across time

    - Can be used with balanced or unbalanced time
    - Will increase level-2 random intercept variance by explaining level-1 residual variance

- A model adding a level-2 **random intercept variance ($\tau_{U_0}^2$)**:

  - Predicts **constant** total variance and covariance over time

    - Should be possible in balanced or unbalanced data

  - Still has $\mathbf{e_{ti}}$ residual variance (always there via default diagonal $\mathbf{R}$ matrix)

- Now we'll see what happens when adding other kinds of random effects, such as a **random linear effect of time**...
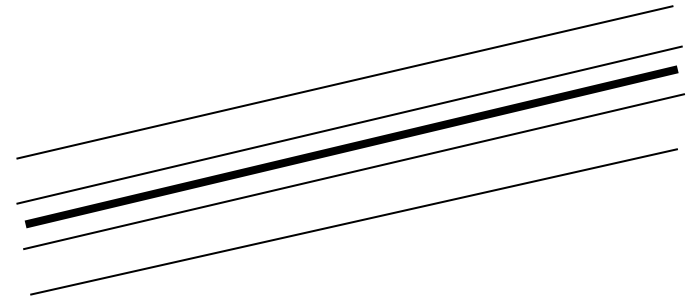
Fixed and Random Effects of **Time**
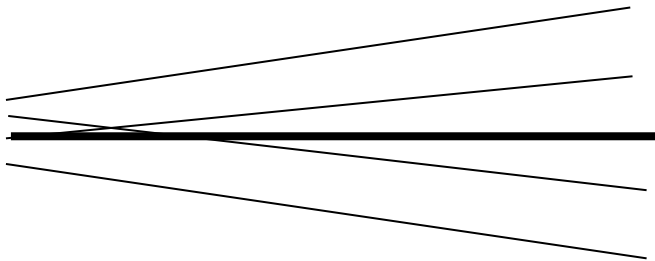## (Note: The intercept is random in every figure)
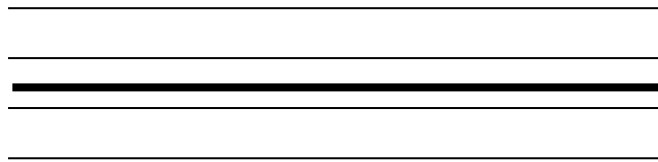
A. **No Fixed**, **No Random**

B. **Yes Fixed**, **No Random**

C. **No Fixed**, **Yes Random**

D. **Yes Fixed**, **Yes Random**
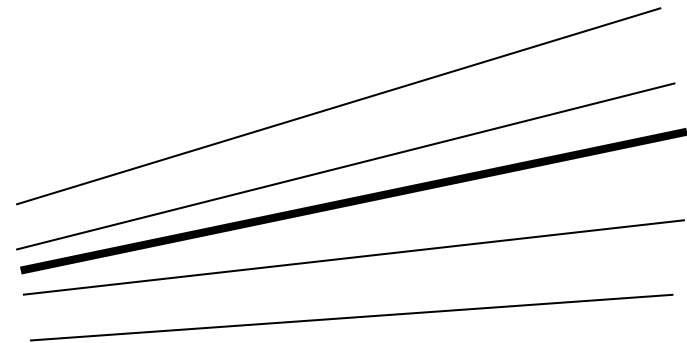
C or D: Random Linear Time Model (6 parms)

Multilevel Model

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

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

Fixed Intercept = predicted mean outcome at time 0

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

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

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

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

Also has an estimated covariance of random intercepts and slopes of $\tau_{U_{01}}$ (not shown here)

Composite Model

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

C or D: Random Linear Time Model (6 parms)

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

Fixed Intercept — Random **Intercept** Deviation — **Fixed Slope** — Random **Slope** Deviation — residual for person i at time t
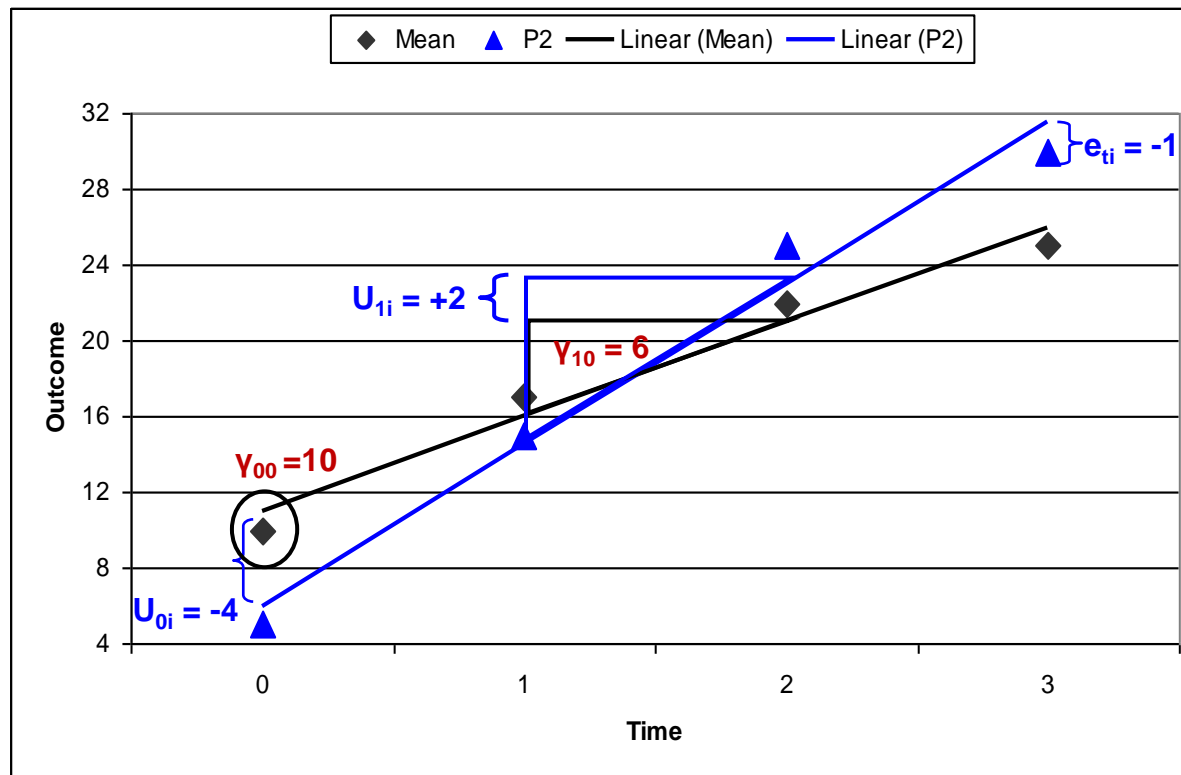
6 Parameters:

2 Fixed Effects:
γ_{00} Intercept, γ_{10} Slope

U_{0i} Random Intercept Variance $= \tau^2_{U_0}$
U_{1i} Random Slope Variance $= \tau^2_{U_1}$
Random Int-Slope Covariance $= \tau_{U_{01}}$

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

Chart labels: Mean, P2, Linear (Mean), Linear (P2)

$e_{ti} = -1$
$U_{1i} = +2$
$\gamma_{10} = 6$
$\gamma_{00} = 10$
$U_{0i} = -4$

Outcome axis: 4, 8, 12, 16, 20, 24, 28, 32
Time axis: 0, 1, 2, 3

Unbalanced Time → Different time occasions across persons? No problem!
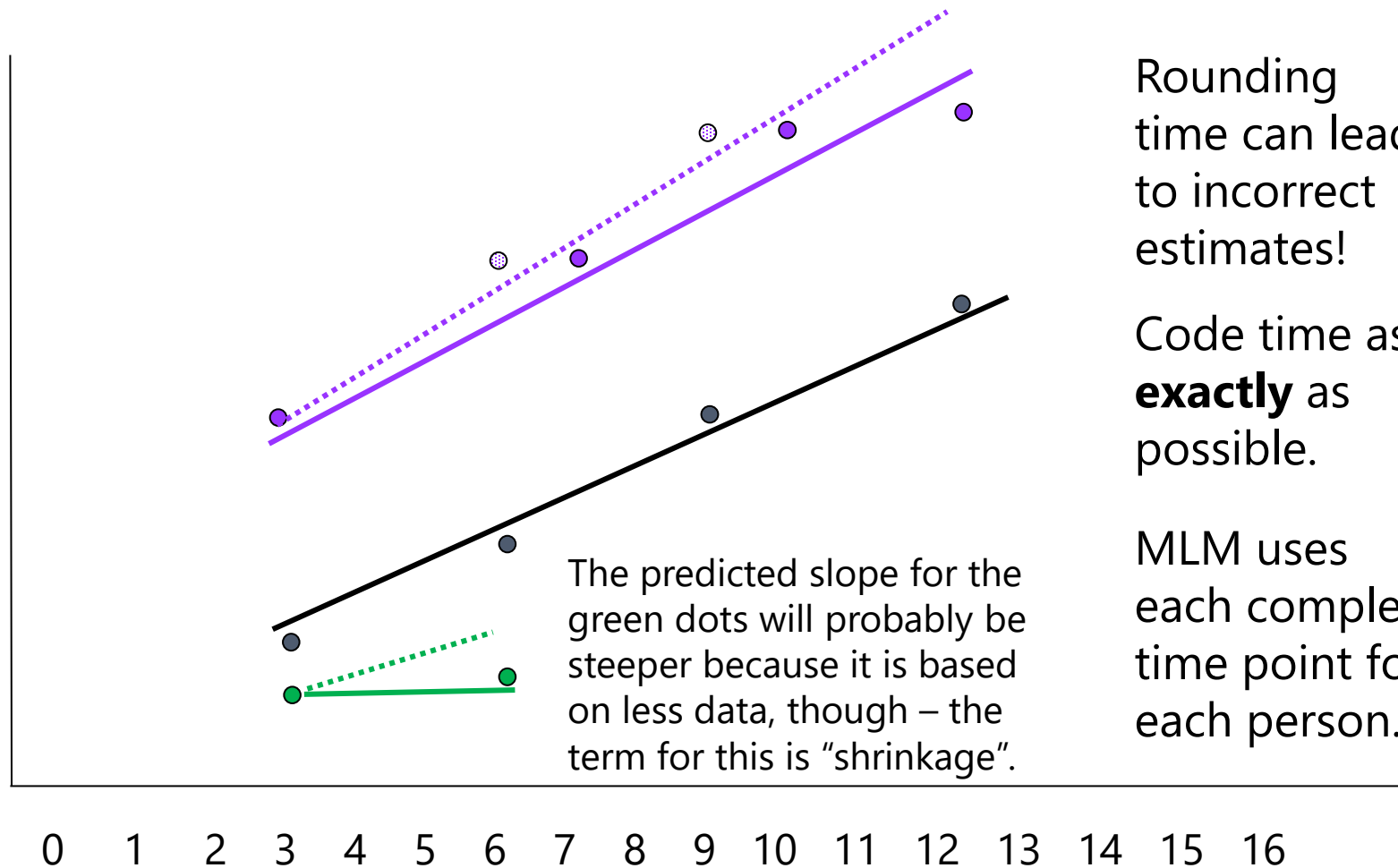
Rounding time can lead to incorrect estimates!

Code time as **exactly** as possible.

MLM uses each complete time point for each person.

The predicted slope for the green dots will probably be steeper because it is based on less data, though – the term for this is "shrinkage".

0   1   2   3   4   5   6   7   8   9   10   11   12   13   14   15   16

Summary: Sequential Models for Effects of Time

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

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

Composite: $y_{ti} = \gamma_{00} + U_{0i} + e_{ti}$

> Empty Means,
> Random Intercept Model:
> 3 parms = γ_{00}, σ_e^2, $\tau_{U_0}^2$

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

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

> Fixed Linear Time,
> Random Intercept Model:
> 4 parms = γ_{00}, γ_{10}, σ_e^2, $\tau_{U_0}^2$

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

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

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

> Random Linear Time Model:
> 6 parms = γ_{00}, γ_{10}, σ_e^2, $\tau_{U_0}^2$,
> $\tau_{U_1}^2, \tau_{U_{01}}$ ($\rightarrow$ cov of U_{0i} and U_{1i})

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

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** (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 level-2 **G** matrix), the e_{ti} **residuals** (whose variance and covariance are estimated in the level-1 **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 (marginal) **V** matrix whose per-person structure depends on the specific time occasions each person has (very flexible for unbalanced time)

Choices in Modeling Variance:
Random Intercepts and Time Slopes

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

Increasing variance →

← Increasing variance

Minimum variance

If time slopes differ across persons, then the outcome variance and covariance must differ over time!
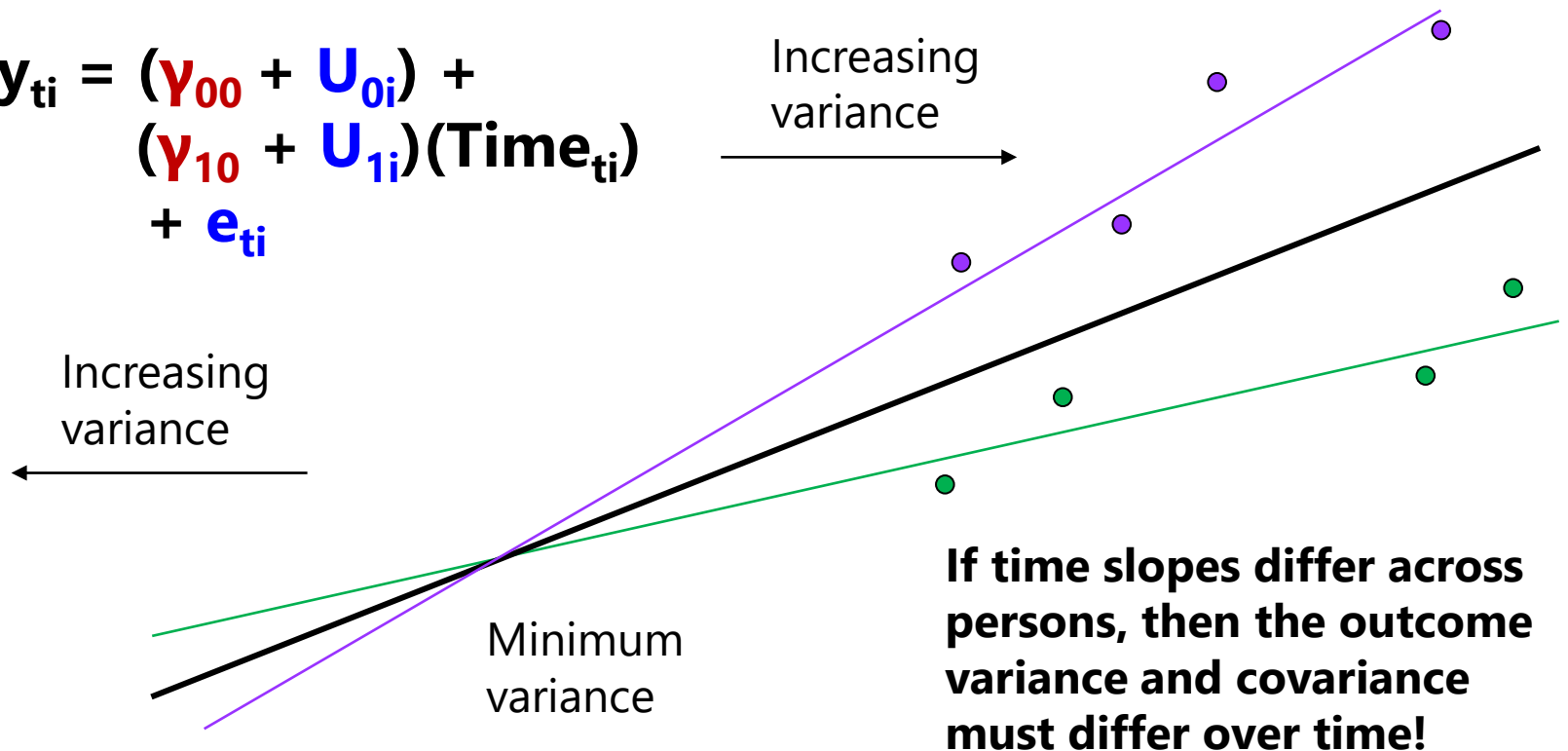
If slopes are different across people, then people differ from each other systematically in 2 ways (U_{0i} and U_{1i}) → this implies compound symmetry will NOT hold.

C or D: 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* Variance:

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

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

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

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

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

C or D: 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):

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

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

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

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

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

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

Model equation per person using matrices:

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

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

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

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

$\mathbf{X}_i = n$ 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

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

Predicted total variances and covariances per person:

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

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

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

$$= \tau_{U_0}^2 + \left[\left(\text{time}\right)^2 \tau_{U_1}^2\right] + \left[2\left(\text{time}\right)\tau_{U_{01}}\right] + \sigma_e^2$$

> $\mathbf{V}_i$ matrix = complicated ☺

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

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

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

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

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

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

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

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

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

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

- The giant marginal **V** matrix across persons is how the multilevel model is actually estimated

- Called a "**block diagonal**" structure → model-implied variances and covariances are created for each person, but 0's are shown for the elements that describe relations between persons (because level-2 persons are supposed to be independent according to this two-level model!)

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

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

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

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

- The "block diagonal" does not need to be the same size or contain the same specific time observations per person…

- **R** matrix can also include non-0 covariance patterns or differential residual variance across time (as in ACS models), although the correlation patterns 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 level-specific 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 allow the variances and covariances to vary in a time-dependent way that better matches the actual data

    - Can allow variance and covariance due to other time-varying predictors, too
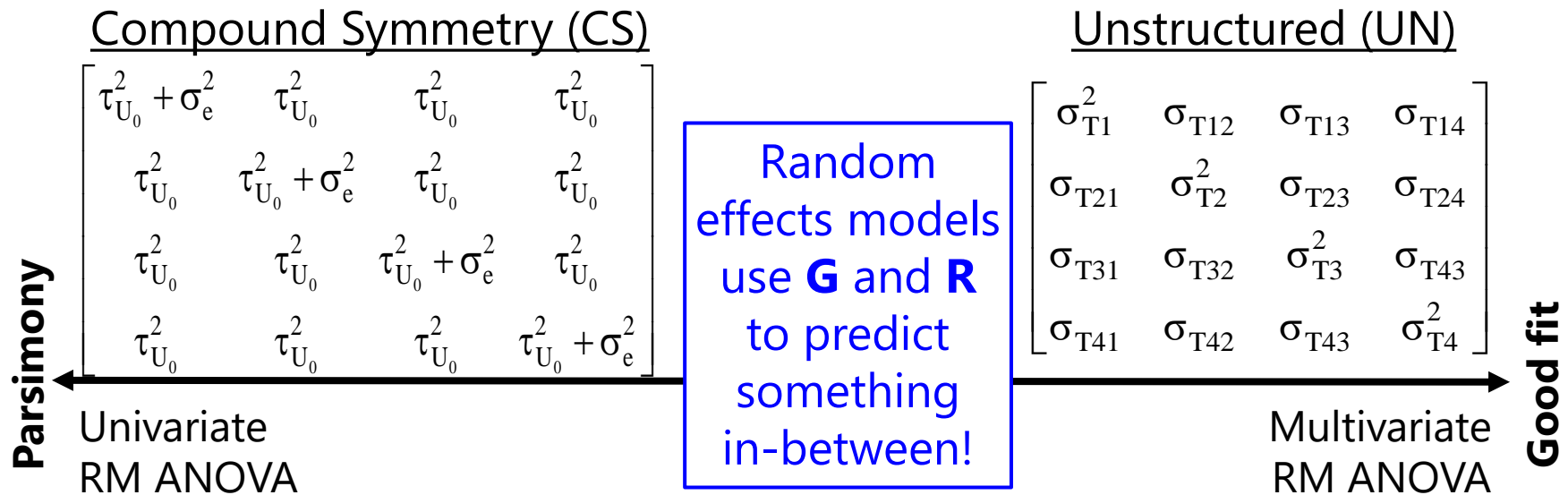
Compound Symmetry (CS)

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

Unstructured (UN)

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

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

Parsimony

Univariate RM ANOVA

Good fit

Multivariate RM ANOVA

Introduction to Random Effects of Time and Model Estimation

• Topics:

> The Big Picture

> Multilevel model notation

> Fixed vs. random slopes of time

> **Effect sizes for random effects**

> **Handling dependency: fixed or random effects?**

> How MLM = SEM

> Fun with maximum likelihood estimation

Two Ways of Conveying Effect Size for Random Effects

- $-2\Delta LL$ tests tell us if a random effect is significant, but random effects variances 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 $\tau_{U_1}^2 = 0.91$, so people need their own time slopes (people change differently). But how much is a variance of 0.91, really?"

- We need to convey effect size for random slopes, but pseudo-R^2 is not appropriate* because variance has not been explained

  - Fixed effects reduce variance; random effects make new variances (piles)

  - *There are "conditional" R^2 measures including random effects, but ugh ☹

- Two ways of conveying effect size for random effects:

  - 95% random effects confidence intervals (CI)—not a typical fixed effect CI!

  - Indices of random effect reliability (less common; useful for power analyses)

Effect Size via 95% Random Effect CIs

- $-2\Delta LL$ tests tell us if a random effect is significant, but random effects variances do not have inherent meaning…

  - e.g., "I have a significant fixed linear time effect of **γ_{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 time slopes (people change differently). But how much is a variance of **0.91**?"

- **(1) 95% Random Effects Confidence Intervals**

  - Can be calculated for each effect **that is random** in your model

  - Provide **range around the fixed effect** in which 95% of your sample *is predicted to fall* given 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 time slopes are predicted to range from -0.15 to 3.59 (so some people may decline)

  - This is NOT the same as a fixed effect CI ($\rightarrow$ inconsistency of fixed effect)!

Effect Size via Reliability Indices

(2): How reliable is a given level-2 unit's random effect?

**Intercept Reliability (IR);
also known as "ICC2":**
$\tau_{U_0}^2$ = random intercept variance
σ_e^2 = residual variance
$L1n$ = L1 sample size per L2 unit

$$IR = \frac{\tau_{U_0}^2}{\tau_{U_0}^2 + \dfrac{\sigma_e^2}{L1n * 1}}$$

Slope Reliability (SR):
$\tau_{U_1}^2$ = random slope variance
σ_e^2 = residual variance
$L1n$ = L1 sample size per L2 unit
σ_{L1}^2 = variance of L1 predictor

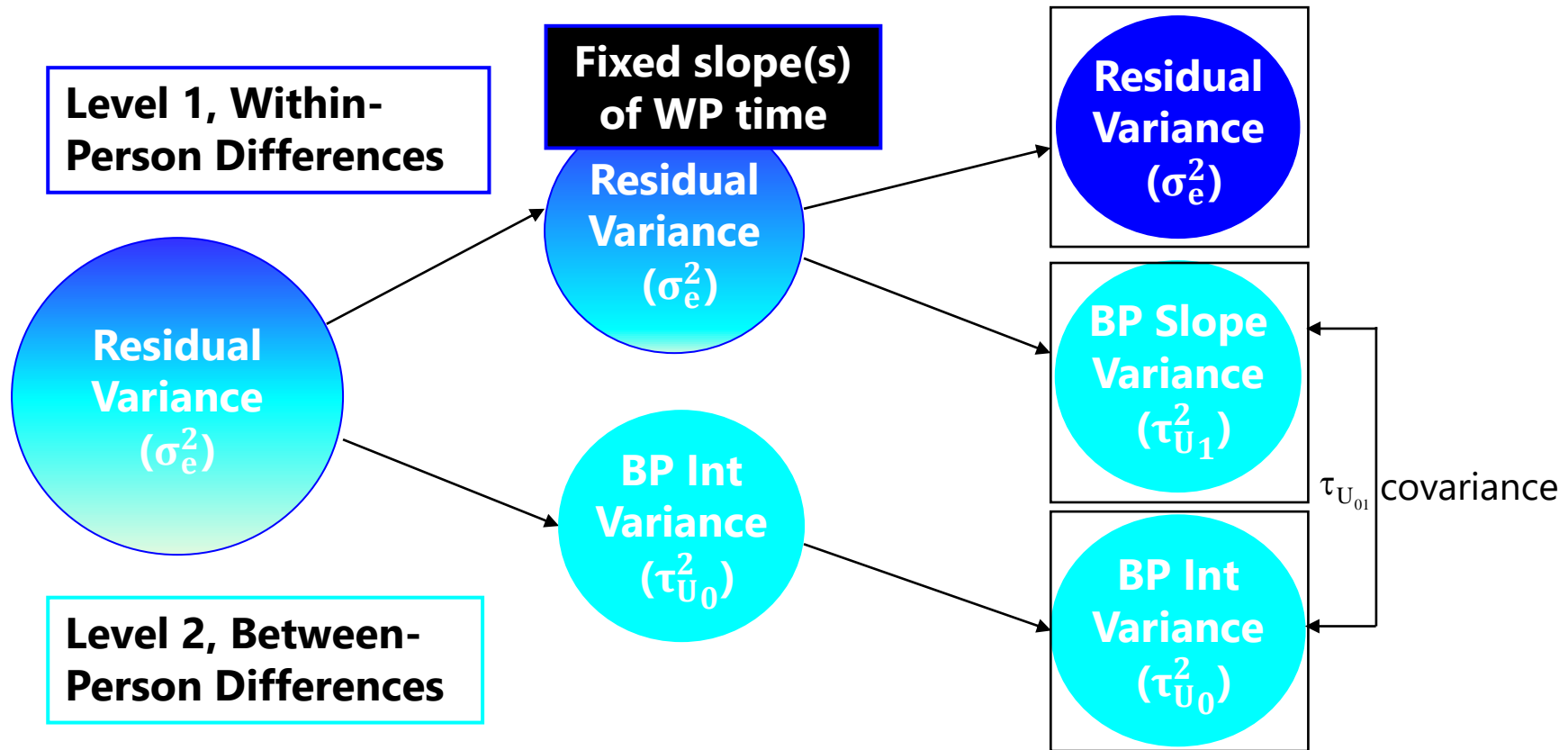$$SR = \frac{\tau_{U_1}^2}{\tau_{U_1}^2 + \dfrac{\sigma_e^2}{L1n * \sigma_{L1}^2}}$$

Although these reliability indices are not commonly reported in many fields (especially SR), they can be very useful in doing power analyses.

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 occasion)
   - This is what a random intercept is for (i.e., a main effect of person)

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 (i.e., a person*predictor interaction)

3. Non-constant within-person correlation for unknown reasons
   - Occasions closer together in time may just be more related
   - This is what ACS models are for (add a residual correlation pattern to R)

Summary: "Handling" Person Dependency

- The process of fitting "unconditional models for time" (fixed and random time slopes) can be depicted as follows:

Level 1, Within-Person Differences

Fixed slope(s) of WP time

Residual Variance (σ_e^2)

Residual Variance (σ_e^2)

Residual Variance (σ_e^2)

BP Int Variance ($\tau_{U_0}^2$)

BP Slope Variance ($\tau_{U_1}^2$)

Level 2, Between-Person Differences

BP Int Variance ($\tau_{U_0}^2$)

$\tau_{U_{01}}$ covariance

Piles of Variance for Handling Dependency

- By adding a random slope, we **carve up** our total variance into 3 piles:

  - BP (error) variance around intercept

  - BP (error) variance around time slope ⎤ These 2 piles are 1 confounded pile of

  - WP (error) residual variance   ⎦ "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	**Residual Variance** (σ_e^2)	**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	**BP Int Variance** $(\tau_{U_0}^2)$ **BP Slope Variance** $(\tau_{U_1}^2)$	

$\tau_{U_{01}}$ covariance

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?**

  - Between-person intercept differences can be included as:

    - $N - 1$ **person dummy code fixed main effects** **OR** **1 random U_{0i} variance**

  - Between-person time slope differences can be included as:

    - $N - 1$***time person dummy interactions** **OR** **1 random U_{1i}*time$_{ti}$ variance**

  - 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**:

  - <u>Quantification</u>: Direct measure of how much of the outcome variance is due to person differences (in intercept or in slopes of predictors)

  - <u>Prediction</u>: Person differences (main effects and slopes of time) then become predictable quantities—can't happen with $N - 1$ fixed effects

  - **Summary: Random effects give you *predictable* control of dependency**

Introduction to Random Effects of Time and Model Estimation

- Topics:
  - The Big Picture
  - Multilevel model notation
  - Fixed vs. random slopes of time
  - Effect sizes for random effects
  - Handling dependency: fixed or random effects?
  - **How MLM = SEM**
  - Fun with maximum likelihood estimation

Translating MLMs into SEMs…

- "**Random effects**" = "pile of variance" = "variance components"
  - ➤ Random effects represent "person*predictor" interaction terms
  - ➤ Random intercept ➔ person*intercept (person "main effect")
  - ➤ Random linear slope ➔ person*time interaction
  - ➤ Capture **specific patterns of covariation** of unknown origin…
    - ▪ *Why do people need their own random intercepts and slopes?*
      We can add person-level predictors to answer these questions

- Random effects can also be seen as **latent variables**
  - ➤ Latent variable = unobservable construct (ability or trait)
    - ▪ In longitudinal data, the latent variables can be thought of as "general tendency" and "propensity to change" as created by measuring the same outcome over time (occasions ➔ indicators)
  - ➤ Let's see how MLMs can be estimated as single-level SEMs using **wide-format data** (one row per person, occasions as variables)…

Structural Equation Models (SEMs)

- **CFA measurement model: $y_{is} = \mu_i + \lambda_i F_s + e_{is}$**

  Btw, there is NO REML estimation in SEM software!

  - Observed response for item i and subject s
    = intercept of item i (μ_i)
      + subject s's latent trait/factor (F_s), item-weighted by λ_i
   + residual error (e_{is}) of item i and subject s

- Differences for two factors modeling longitudinal data instead:

  - Usually two factors for "level" and "change" (intercept and slope):
    $y_{ti} = (\gamma_{00} + U_{0i}) + (\gamma_{10} + U_{1i})time_{ti} + e_{ti}$ → **so each U → "F" in SEM**

  - Fixed effects → factor means; random effects → factor variances

  - The **occasion-specific intercepts μ_i** cannot be separately identified from the "intercept" latent factor and therefore must be fixed to 0

  - **Factor loadings λ_i** for how each outcome relates to the latent factor are (usually) pre-determined by how much time has passed → **fixed to what the "time" predictor would be for each occasion**

  - Unbalanced time requires "definition variables" → use variables for person-specific time loadings rather than fixing loadings to same values for all

    - In Mplus, is TSCORES option; could not find an equivalent option in R lavaan

Random Effects as Latent Variables

- **Single-level model for the variance → σ_e^2 only**
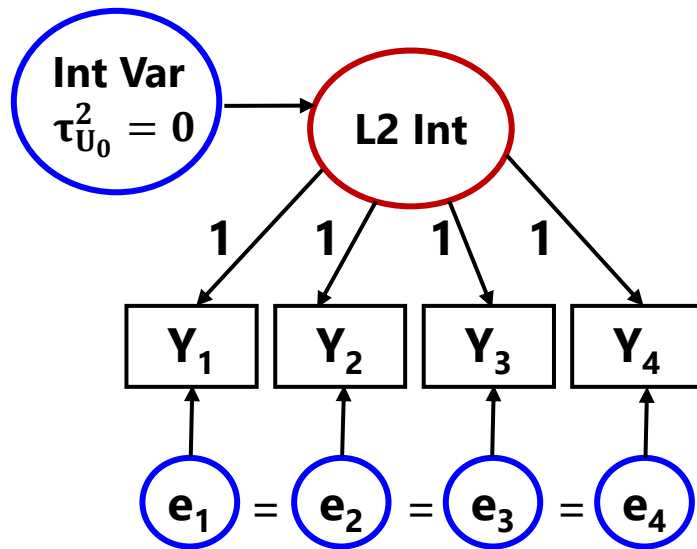
  ➢ $y_{ti} = \gamma_{00} + e_{ti}$

Mean of the intercept factor = fixed intercept γ_{00}

Loadings of intercept factor = 1 (all occasions contribute equally)

Indicator intercepts = 0 (always)

L2 variance of intercept factor $\tau_{U_0}^2$ = 0 so far

L1 residual variance (σ_e^2) is predicted to be equal across occasions (e=e)

Int Var $\tau_{U_0}^2 = 0$ → L2 Int

1   1   1   1

Y_1 Y_2 Y_3 Y_4

e_1 = e_2 = e_3 = e_4

➢ After controlling for the *fixed* intercept (factor mean), level-1 residuals are predicted to be uncorrelated with constant variance (not the default in SEM software!)

  ▪ **Complete** independence is not likely to hold in longitudinal data!

Random Effects as Latent Variables

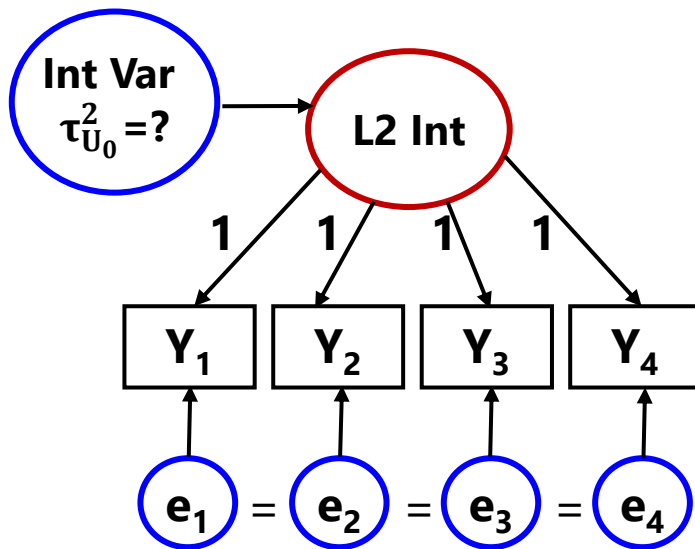- **Two-level model for the variance → add $\tau_{U_0}^2$**

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

Mean of the intercept factor = fixed intercept γ_{00}

Loadings of intercept factor = 1 (all occasions contribute equally)

L2 variance of intercept factor $\tau_{U_0}^2$ = random intercept variance

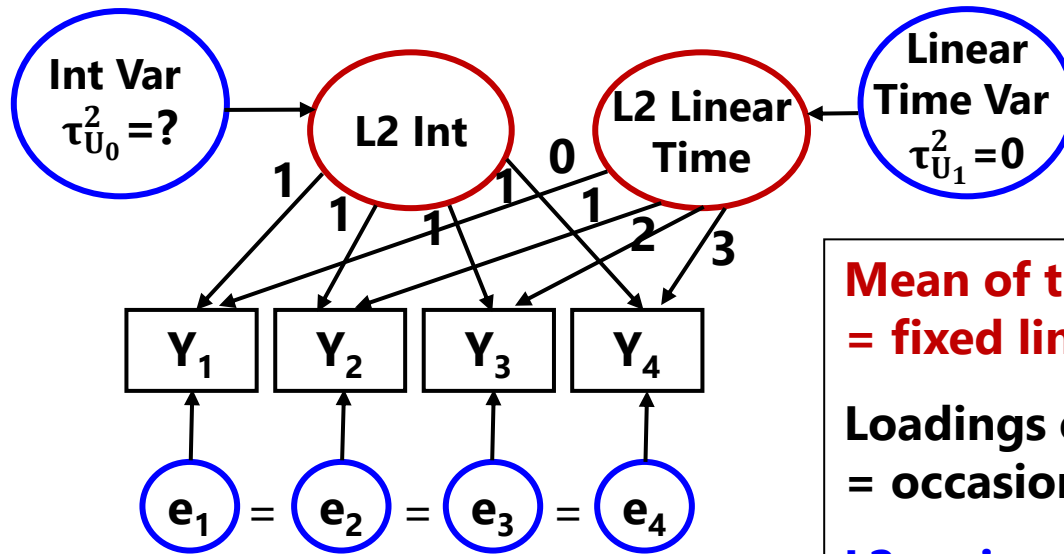**L1 residual variance (σ_e^2) is predicted to be equal across occasions**

  - After controlling for the *random* intercept (factor mean and variance), level-1 residuals are predicted to be uncorrelated with constant variance (not default in SEM software!)

    - Same "**local** independence" idea as discussed in CFA context

Random Effects as Latent Variables

- **Fixed linear time, random intercept model:**

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

  Int Var $\tau_{U_0}^2 = ?$ → L2 Int

  L2 Linear Time ← Linear Time Var $\tau_{U_1}^2 = 0$

  1, 1, 1, 1, 0, 1, 2, 3

  Y_1, Y_2, Y_3, Y_4

  $e_1 = e_2 = e_3 = e_4$

  Mean of the linear time factor = fixed linear slope γ_{10}

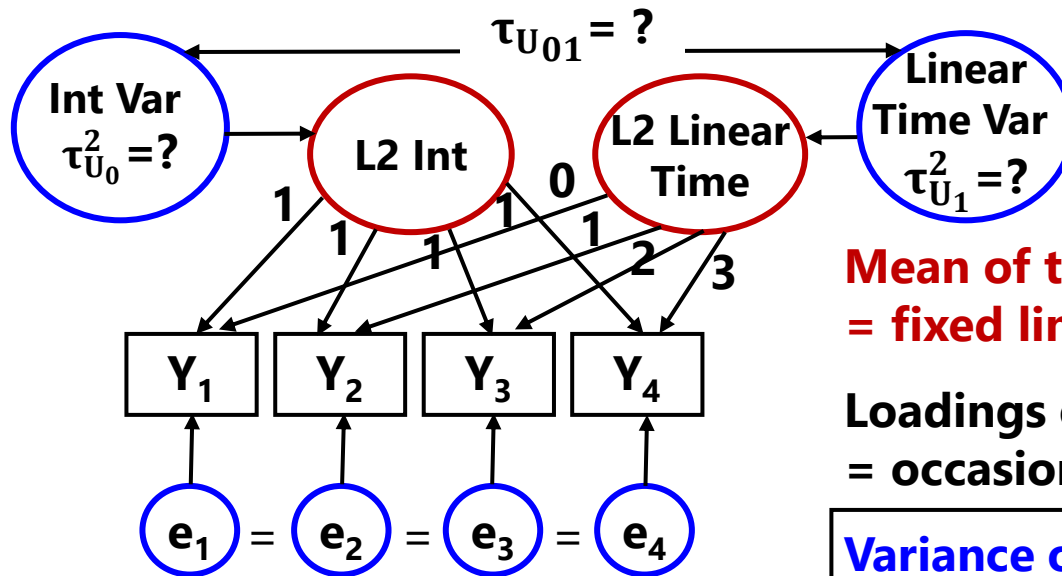 **Loadings of linear time factor = occasions (keep real time)**

  L2 variance of linear time factor $\tau_{U_1}^2 = 0$

  - After controlling for the *fixed linear time slope* (factor mean) and *random* intercept (factor mean and variance), level-1 residuals are (still) predicted to be uncorrelated with constant variance

Random Effects as Latent Variables

- **Random linear time model:**

  ➢ $y_{ti} = \textcolor{red}{\gamma_{00}} + (\textcolor{red}{\gamma_{10}}\text{Time}_{ti}) + \textcolor{blue}{U_{0i}} + (\textcolor{blue}{U_{1i}}\text{Time}_{ti}) + \textcolor{blue}{e_{ti}}$

$\tau_{U_{01}} = ?$

Int Var $\tau_{U_0}^2 = ?$

L2 Int

L2 Linear Time

Linear Time Var $\tau_{U_1}^2 = ?$

1, 1, 1, 1, 0, 1, 2, 3

Y_1, Y_2, Y_3, Y_4

$e_1 = e_2 = e_3 = e_4$

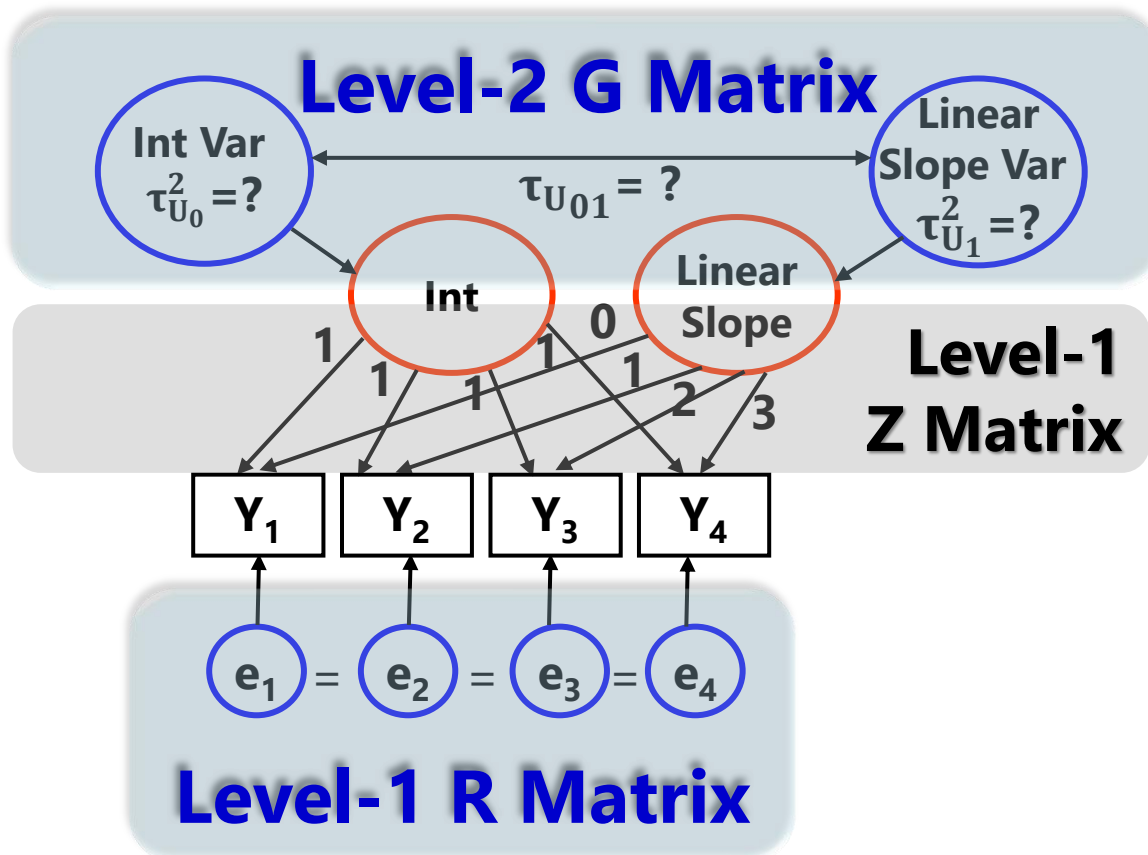**Mean of the linear slope factor = fixed linear slope γ_{10}**

Loadings of linear slope factor = occasions (keep real time)

Variance of linear time factor $\tau_{U_1}^2$ = random slope variance

  ➢ After controlling for the *random* linear time slope and *random* intercept (both factor means and variances), level-1 residuals are (still) predicted to be uncorrelated with constant variance (random slope → het var!)

Random Linear Time Model: From MLM to Single-Level SEM

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

Level-2 G Matrix

Int Var $\tau^2_{U_0} = ?$

$\tau_{U01} = ?$

Linear Slope Var $\tau^2_{U_1} = ?$

Int

Linear Slope

Level-1 Z Matrix

1 1 1 0
1 1 1
1 2 3

Y_1 Y_2 Y_3 Y_4

e_1 = e_2 = e_3 = e_4

Level-1 R Matrix

For unbalanced time, you need "definition variables" (like Mplus TSCORES) that allow different loadings ($\rightarrow$ occasions) per person

Btw, allowing different residual variances for every occasion may be redundant with the random slope factor—it already predicts variance to change over time!

Random Linear Time Model:
From MLM to "Multilevel SEM"

$$y_{ti} = \gamma_{00} + (\gamma_{10}\text{Time}_{ti}) + U_{0i} + (U_{1i}\text{Time}_{ti}) + e_{ti}$$
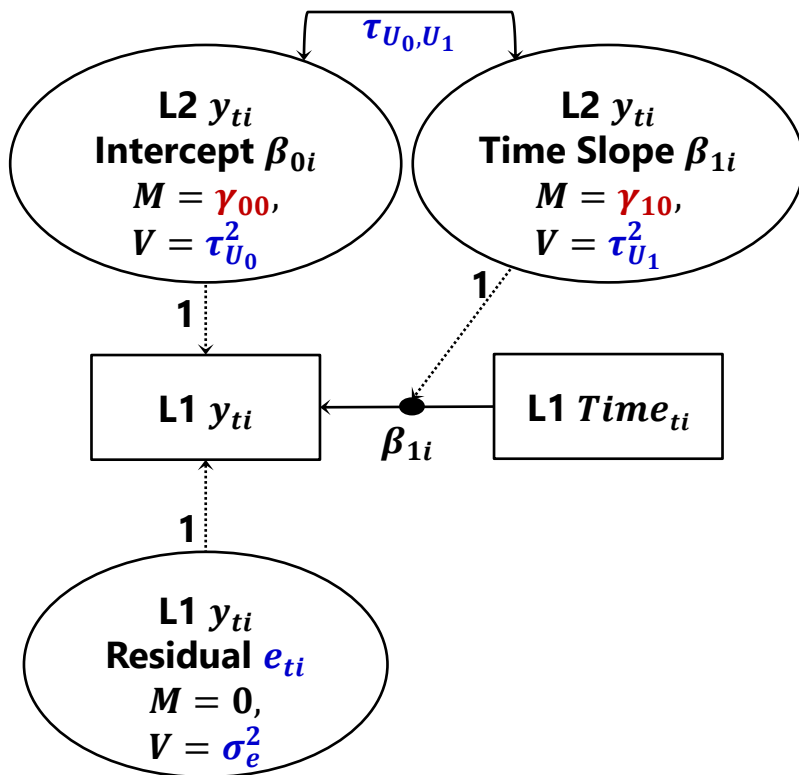
Multilevel SEM (or what I prefer to call "multivariate MLM" in absence of a true measurement model) uses a **long-format** (stacked) data structure with one row per level-1 unit (so per occasion per person), just like univariate MLM.

The difference is that in M-SEM, multiple variables (predictors or outcomes) can have their variance partitioned into BP intercepts, BP slopes, and WP residuals at the same time (with additional features for autoregressive relations possible in "dynamic SEM", which is still M-SEM).

Introduction to Random Effects of Time and Model Estimation

- Topics:
  - The Big Picture
  - Multilevel model notation
  - Fixed vs. random slopes of time
  - Effect sizes for random effects
  - Handling dependency: fixed or random effects?
  - How MLM = SEM
  - **Fun with maximum likelihood estimation**

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_{ti} values per se *but are not parameters that are solved for iteratively in maximum likelihood estimation****
    - ▪ ****Unless you have a generalized MLM, in which case they are*

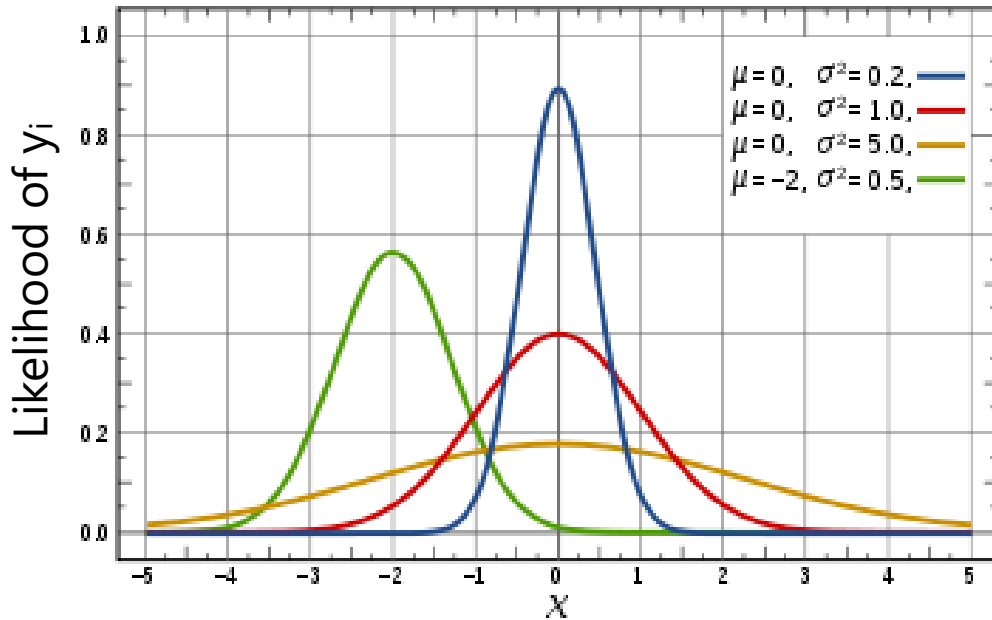- ## **Random Effects in the Model for the Variance:**

  - ➢ How model residuals are distributed and related across observations (persons, time, etc) – *leftover variation* due to dimensions of sampling

  - ➢ Random effects variances and covariances are a mechanism by which complex patterns of variance and covariance among the y_{ti} residuals can be built into the model (not the y_{ti} values, but their dispersion)

  - ➢ Anything besides level-1 residual variance σ_e^2 must be solved for iteratively, which increases the dimensionality of estimation process

  - ➢ Estimation utilizes the predicted **V** matrix for each person

  - ➢ In the material to follow, **V** will be based on a <u>random linear time model</u>

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)

Remember Univariate Normal?

Likelihood of y_i vs X

Legend:
$\mu=0, \quad \sigma^2=0.2$ (blue)
$\mu=0, \quad \sigma^2=1.0$ (red)
$\mu=0, \quad \sigma^2=5.0$ (orange)
$\mu=-2, \quad \sigma^2=0.5$ (green)

- 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

- Example: regression

$$y_i = \beta_0 + \beta_1 x_i + e_i$$
$$y_i = \beta_0 + \beta_1 x_i$$
$$e_i = y_i - y_i \qquad \sigma_e^2 = \frac{\sum_{i=1}^{N} e_i^2}{N-2}$$

Univariate Normal PDF (two ways):

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

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

Multivariate Normal for $\mathbf{Y}_i$
(height for all *n* outcomes for person *i*)

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

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

- In a random linear time model, the only fixed effects (in **γ**) 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 marginal 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-bounds correlations between random effects

  - Otherwise, SAS 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:

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

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

- 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 to build **V**$_i$ 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 $\mathbf{V_i}$:

$$\gamma = \left\{ \sum_{i=1}^{N} \left(\mathbf{X_i^T V_i^{-1} X_i} \right) \right\}^{-1} \sum_{i=1}^{N} \left(\mathbf{X_i^T V_i^{-1} Y_i} \right), \qquad \text{Cov}(\gamma) = \left\{ \sum_{i=1}^{N} \left(\mathbf{X_i^T V_i^{-1} X_i} \right) \right\}^{-1}$$

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

γ = **fixed effect estimates** $\qquad$ $\text{Cov}(\gamma) = \gamma$ **sampling variance (SQRT of diagonal = SE)**

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

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

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

- **Implication: fixed effects don't cause estimation problems**... (***at least in general linear mixed models with normal residuals)

What about ML vs. REML?

- **REML** estimates of random effects variances and covariances are **unbiased** because they account for the uncertainty that results from simultaneously also estimating fixed effects (whereas ML estimates do not, so they are too small)

- Say what??? Think of it this way: Remember "population" vs. "sample" formulas for computing variance?

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

  - *N* – 1 is used because the mean had to be estimated from the data (i.e., the mean is the fixed intercept)…

- Similar idea: ML estimates of random effects variances will be too small by a factor of (*N* – *k*) / *N*, where *N* = # persons and *k* = #fixed effects… it just looks way more complicated…

What about ML vs. REML?

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

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

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

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

- Extra part in REML is the sampling variance of the fixed effects... it is added back in as a way to account for uncertainty in estimating fixed effects

- **REML** maximizes the likelihood of the residuals specifically, so LL values for models with **different fixed effects are not on the same scale**

  > This is why you can't do -2ΔLL (LRTs) in REML when the models to be compared have different fixed effects → the model residuals are defined differently

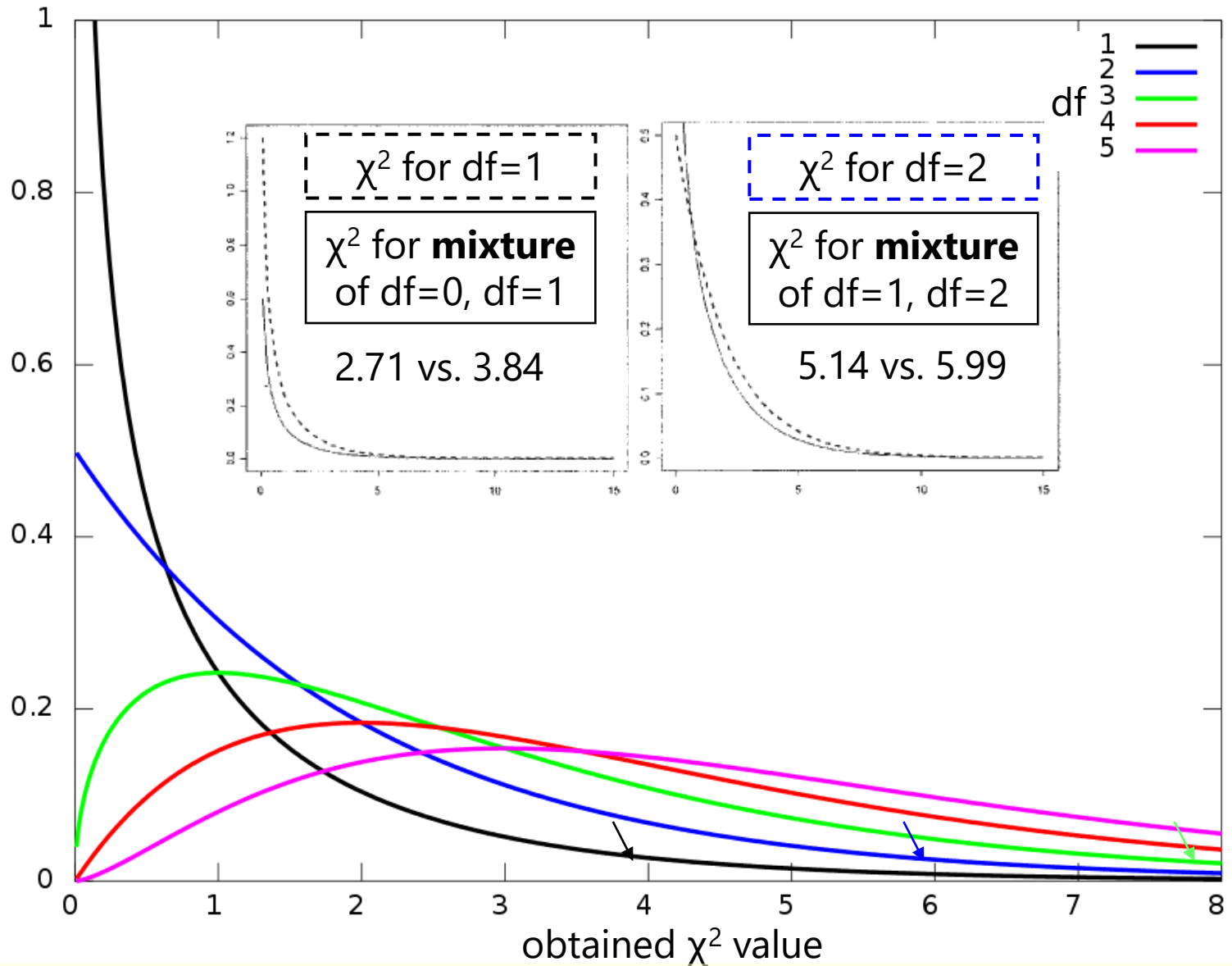# End Goal #3: How well do the model predictions match the data?

- 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

  - ➤ Btw, indicators are always balanced, so can get absolute fit in CFA and SEM → model χ^2 test reflects match between actual and model-implied data

  - ➤ Time is often a quantitative variable, so no absolute fit is provided in MLM (or in SEM when using random slopes or TSCORES for unbalanced time)

    - ▪ In ML, can compute*** absolute fit when the saturated means, unstructured variance model is estimable → is −2ΔLL for your model versus "perfect" model for time

    - ▪ For absolute fit tests given balanced time using REML instead, stay tuned!

- Relative model fit is given as **−2LL** in some programs, in which smaller is better (as opposed to LL, in which bigger is better)

  - ➤ −2* needed to conduct "likelihood ratio" or "deviance difference" tests

  - ➤ Also information criteria:

    - ▪ **AIC**: −2LL + 2*(#parms)   **BIC**: −2LL + log(N)*(#parms)

    - ▪ #parms = all parameters in ML; #parms = variance model parameters only in REML

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 null hypothesis on a boundary (e.g., variance=0 or factor correlation=1)

  - ➢ Ok for fixed effects (could be any positive or negative value)

  - ➢ NOT ok for tests of random variances (must be > 0)

  - ➢ Ok for tests of *heterogeneity* of 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 a little conservative in boundary cases**

Regular vs Mixture χ² Distributions
small interior pictures from Stoel et al., 2006

χ² for df=1

χ² for **mixture** of df=0, df=1

2.71 vs. 3.84

χ² for df=2

χ² for **mixture** of df=1, df=2

5.14 vs. 5.99

df 1 2 3 4 5

obtained χ² value

Critical Values for 50:50 Mixture of Chi-Square Distributions

Significance Level

df (q)	0.10	0.05	0.025	0.01	0.005
0 vs. 1	1.64	2.71	3.84	5.41	6.63
1 vs. 2	3.81	5.14	6.48	8.27	9.63
2 vs. 3	5.53	7.05	8.54	10.50	11.97
3 vs. 4	7.09	8.76	10.38	12.48	14.04
4 vs. 5	8.57	10.37	12.10	14.32	15.97
5 vs. 6	10.00	11.91	13.74	16.07	17.79
6 vs. 7	11.38	13.40	15.32	17.76	19.54
7 vs. 8	12.74	14.85	16.86	19.38	21.23
8 vs. 9	14.07	16.27	18.35	20.97	22.88
9 vs. 10	15.38	17.67	19.82	22.52	24.49
10 vs. 11	16.67	19.04	21.27	24.05	26.07

This may work ok if only one new parameter is bounded ... for example:

+ Random Intercept
df=1: 2.71 vs. 3.84

+ Random Linear
df=2: 5.14 vs. 5.99

+ Random Quad
df=3: 7.05 vs. 7.82

Critical values such that the right-hand tail probability =
$$0.5 \times \Pr(\chi^2_q > c) + 0.5 \times \Pr(\chi^2_{q+1} > c)$$

Source: Appendix C (p. 484) from Fitzmaurice, Laird, & Ware (2004). *Applied Longitudinal Analysis*. Hoboken, NJ: Wiley

Solutions for Boundary Problems when using −2ΔLL tests (LRTs)

- 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 $\chi^2(2)$

$$\text{Mixture } p\text{-value} = 0.5 * \text{prob}(\chi_1^2 > -2\Delta LL) + 0.5 * \text{prob}(\chi_2^2 > -2\Delta LL)$$ 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 values for each are arrived at independently, which pry isn't the case)

- Two options for more complex cases involving LRTs:

  - Simulate data to determine actual mixture for calculating p-value

  - Accept that −2ΔLL is a little conservative in these cases, and use it anyway → In the book I used ~ to acknowledge this: e.g., −2ΔLL(~2) > 5.99, $p < .05$

Predicted Level-2 $\mathbf{U}_i$ Random Effects
(aka Empirical Bayes or BLUP Estimates)

- Level-2 $\mathbf{U}_i$ random effects require further explanation…

  - Empty two-level model: $\mathbf{y_{ti}} = \mathbf{\gamma_{00}} + \mathbf{U_{0i}} + \mathbf{e_{ti}}$

  - U_{0i} is a person-mean deviation, right? Well, not exactly…

- 3 ways of representing size of individual differences in individual intercepts and slopes across people:

  - Get people's OLS intercepts and slopes; calculate their variance

  - Estimate variance of the person $\mathbf{U}_i$'s (what we do in MLM)

  - Predict person $\mathbf{U}_i$'s; calculate their variance (2-stage MLM)

- Expected order of magnitude of variance estimates:

  - OLS variance > MLM variance > Predicted $\mathbf{U}_i$'s variance

  - Why are these different? **Shrinkage!**

What about the random effect U_i values?

- Person $\mathbf{U}_i$ values are NOT estimated in the ML process
  - $\mathbf{G}$ matrix variances and covariances are sufficient statistics for the estimation process assuming multivariate normality of $\mathbf{U}_i$ values
  - Person $\mathbf{U}_i$ random effects are **predicted** (e.g., via SOLUTION on the SAS RANDOM statement as: $\mathbf{U}_i = \mathbf{G}_i\mathbf{Z}_i^{\mathrm{T}}\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mathbf{X}_i\boldsymbol{\gamma})$
    - Which then create individual estimates as $\boldsymbol{\beta}_{0i} = \boldsymbol{\gamma}_{00} + \mathbf{U}_{0i}$ and $\boldsymbol{\beta}_{1i} = \boldsymbol{\gamma}_{10} + \mathbf{U}_{1i}$

- What isn't obvious: the composite $\boldsymbol{\beta}_i$ values are weighted combinations of the fixed effects ($\boldsymbol{\gamma}$) and individual OLS estimates (β_{OLSi}) :

  $$\text{Random Effects: } \boldsymbol{\beta}_i = \mathbf{W}_i\boldsymbol{\beta}_{OLSi} + \left(\mathbf{I} - \mathbf{W}_i\right)\boldsymbol{\gamma} \qquad \text{where: } \mathbf{W}_i = \mathbf{G}_i\left[\mathbf{G}_i + \sigma_e^2\left(\mathbf{Z}_i^{\mathrm{T}}\mathbf{Z}_i\right)^{-1}\right]^{-1}$$

  - The more "true" variation in intercepts and slopes there is in the data (in $\mathbf{G}$), the more the $\boldsymbol{\beta}_i$ estimates are based on individual OLS estimates
  - But the more "unexplained" residual variation there is around the individual trajectories (in $\mathbf{R}$), the more the fixed effects are heavily weighted instead
    - = **SHRINKAGE** (more so for people with relatively fewer occasions, too)

What about the random effect U_i values?

- Point of the story: **U_i** values are NOT single scores:
  - ➢ They are the mean of a distribution of possible values for each person (i.e., as given by the SE for each **U_i**, which is also provided)
  - ➢ These "best estimates" of the **U_i** values are shrunken anyway
  - ➢ In other words, **U_i** values are FACTOR SCORES!

- Good news: you don't need those **U_i** values in the first place!
  - ➢ Goal of MLM is to estimate and predict the variance of the **U_i** values (in **G**) with person-level characteristics directly within the same model
  - ➢ If you want your **U_i** values to be predictors instead, then you need to buy your growth curve model at the SEM store instead of the MLM store
  - ➢ We could use the predicted **U_i** values to examine violations of model MVN assumptions (but research suggests this doesn't matter much)
    - ▪ SAS: Get **U_i** values by adding: ODS OUTPUT SolutionR=dataset;
    - ▪ SAS: Get e_{ti} residuals by adding OUTP=dataset after / on MODEL statement
    - ▪ SAS: Add RESIDUAL option after / on MODEL statement to make plots

Estimation: The Grand Finale

- Estimation in MLM is all about the finding the model for the variance (random effects variances and covariances)

  - The more parameters there are in the model for the variance, the harder it is to find them (the more dimensions of the likelihood mountain there are to scale simultaneously)

  - "Non-positive-definite" **G** matrix means "broken model"

  - Fixed effects are solved for after-the-fact, so they rarely cause estimation problems (except in generalized MLM variants)

  - Person random effects are not model parameters, but can be predicted afterwards (but try never to use these as new data)

- Estimation comes in two flavors:

  - ML → maximize the data; −2ΔLL to compare any nested models

  - REML → maximize the residuals; −2ΔLL to compare models that differ in their model for the variance ONLY