Review of Longitudinal Multilevel Models

- Topics:
  - **Concepts and terminology in longitudinal models**
  - **Modeling person dependency**
  - Fixed and random intercepts
  - Fixed and random time slopes
  - Time-invariant predictors
  - Details (ML vs REML, significance tests, model comparisons)

Sources of Longitudinal Relations

- **Between-Person (BP)** Variation (at "Level 2"): People who...
  - "**INTER**-individual differences" from "**time-invariant**" **measures**
  - All longitudinal studies that begin as cross-sectional studies have this

When people have

- **Within-Person (WP)** Variation (at "Level 1"): More or less than usual...
  - "**INTRA**-individual differences" from "**time-varying**" **measures**
  - Only longitudinal studies can provide this extra type of 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 usual

- I use "person" here, but "between" units can be anything that is measured repeatedly (e.g., schools, countries, companies…)

Sources of "Time" in Longitudinal Data

- What aspects of "**time**" are relevant?
  - **WP change**: e.g., time in study, age, grade, time to/from event
  - **WP fluctuation**: e.g., time of day, day of week, day in study

- Does time vary **within persons (WP)** AND **between persons (BP)**?
  - If people differ in time at the study beginning (e.g., accelerated designs), we will need to **differentiate BP time effects from WP time effects**
  - If there is more than one kind of WP "time" (e.g., occasions within days), we will need to **differentiate distinct sources of WP time effects**

- Is time *balanced* or *unbalanced*?

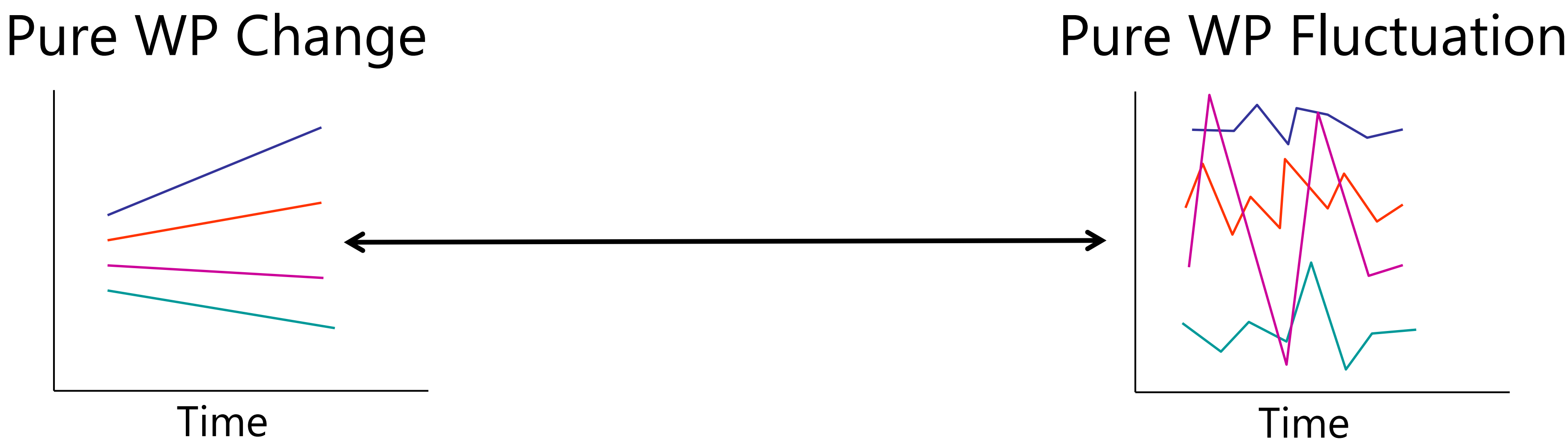
  → more common in SEM context, when each occasion has 1 box (slope)
  - **Balanced** = **shared** measurement schedule (not necessarily equal interval)

    ↳ panel design → people are measured at same time points.
    - Although some people may miss some occasions, making their data "incomplete"
  - **Unbalanced** = people have **different** possible time values
    - By definition, the possible outcomes are at least partially "incomplete" across persons
    - This may be a consequence of using a time metric that also varies between persons

A Longitudinal Data Continuum

- **Within-Person (WP) Change**: Expect systematic effect(s) of time
  - e.g., "(Latent) Growth Curve Models" → **Time is meaningfully sampled**
  - If magnitude or direction of change differs across individuals, then the outcome's variance and covariance will change over time, too!

- **Within-Person (WP) Fluctuation**: Few *expected* effects of time
  - Outcome just varies/fluctuates over time (e.g., emotion, mood, stress)
  - **Time is just a way to get lots of data per person** (e.g., EMA studies)
  - Lends itself to questions about effects of relative changes and inconsistency



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 known predictor variables
    - Fixed effects are **estimated constants** that multiply predictors

- **Model for the Variance:**
  - *Aka* **Random Effects and Residuals**, Stochastic Part of Model
    Introduce dependency among results
  - What you *were* used to **making assumptions about** instead
  - How **residuals are distributed and related** across sampling dimensions (persons, occasions) → these relationships are known as "**dependency**" and ***this is the primary way that longitudinal models differ from "regular" regression models***
    units ≠

Modeling Longitudinal Dependency

- Outcomes from the same sampling unit (i.e., person) will have one or more sources of **dependency → correlated residuals**
  - If ignored, dependency in a longitudinal outcome will result in incorrect fixed effect standard errors and *p*-values (well-known problem)
  - If ignored, dependency in a longitudinal predictor variable will result in incorrect fixed effect estimates, too (relatively less well-known problem)
    - Because time-varying predictors have both BP and WP variation—stay tuned!

- The sources of residual correlation of occasions from same person can be captured by a model in three main ways:
  1. **Fixed effects:** Add Person ID as a predictor (via *N*-1 dummy codes)
  2. **(Multivariate) alternative covariance structures (ACS):** Just allow correlation over occasions to exist (for unknown reasons)
  3. **Add a "level" (or more):** Use random effect (latent factor) variances, as possible within multilevel or structural equation modeling

1. Modeling Longitudinal Dependency

- **Fixed effects:** Add Person ID as a categorical predictor
- Estimate fixed effects of $N - 1$ dummy codes for person ID
  - Person ID **main effects** capture dependency due to mean differences
  - Interactions of Person ID with time-varying predictors (like time) capture other predictor-specific sources of person dependency

- Pro: Does adequately control for person dependency
  - Very common in econometrics, political science, sociology…
  - Does a better job in studies with "few" persons (< 15ish)
  - Useful to make individual-specific conclusions (i.e., as in aggregated N-of-1 randomized control trials)

- Con: Does not allow prediction of WHY any of those individual differences occurred ☹
  - Model would be saturated with respect to between-person differences

2. Modeling Longitudinal Dependency

- **Alternative multivariate variance–covariance structures:** Change model to allow correlation over occasions (and any residual heterogeneity) to exist
- Is only possible given **balanced data** (all people on same schedule) and conditionally normal outcomes (i.e., not when using generalized models)

- Is the basis of **repeated measures ANOVA**, of which there are **2 kinds**
  - "**Univariate approach**": residuals have equal variance and equal correlations across all repeated measures outcomes—but this "compound symmetry" pattern can only possibly hold if all people change the same!
  - "**Multivariate approach**": all residual variances and correlations are separately estimated—but this "unstructured" (MANOVA) model becomes difficult-to-impossible given many outcomes (especially with few people)
  - Estimation using ordinary least squares → listwise deletion of missing data ☹

    - It depends on the predictors → joint likelihood
    - complete case with the same predictors.

- Switching to maximum likelihood estimation uses all complete occasions AND offers more choices for patterns of residual variance and correlation
  - Btw, residual maximum likelihood = ordinary least squares given complete outcomes
  - e.g., Compound Symmetry Heterogeneous (diff variances, equal correlation)
  - Options that use time-lagged covariances also require equal-interval occasions: e.g., First-order auto-regressive, moving average, or antedependence; Toeplitz

3. Modeling Longitudinal Dependency

- **Add a "level"** → Add random effect (latent variable) variances
- Random effect = latent variable = model term that each person (can) get their own version of (in theory); directly estimate the variance of each random effect across persons → BP differences
  - Capture patterns of non-constant variance and covariance for testable reasons
  - Usable for general or generalized models (i.e., for any kind of outcome)
  - Usable in balanced or unbalanced longitudinal data (i.e., for any time structure)

- More generally, a "level" is a dimension of sampling that has unexplained outcome variability represented by 1+ random effects
  - "time" is not a level once sufficient fixed effects for its mean diffs are included
  - e.g., Randomized Control Trial (RCT) of 5 monthly occasions → 2 levels (1. within-person, 2. between-person)
  - e.g., Ecological Momentary Assessment (EMA) design of 4 observations per day for 3 weeks → 3 levels (1. within-day, 2. between-day, 3. between-person)

A Statistician's World View

- 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)

Note: OLS is only for GLM

- **General*ized* Linear Models:** **any conditional outcome distribution**, **fixed effects** through **link functions**, no random effects (one dimension)
- **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)
  - Not this week—Many of the same concepts, but with more complexity in estimation

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

Review of Longitudinal Multilevel Models

- Topics:
  - Concepts and terminology in longitudinal models
  - Modeling person dependency
  - **Fixed and random intercepts** → commonality in how people change (constant)
  - Fixed and random time slopes
  - Time-invariant predictors
  - Details (ML vs REML, significance tests, model comparisons)

The Two Sides of a General Linear Model

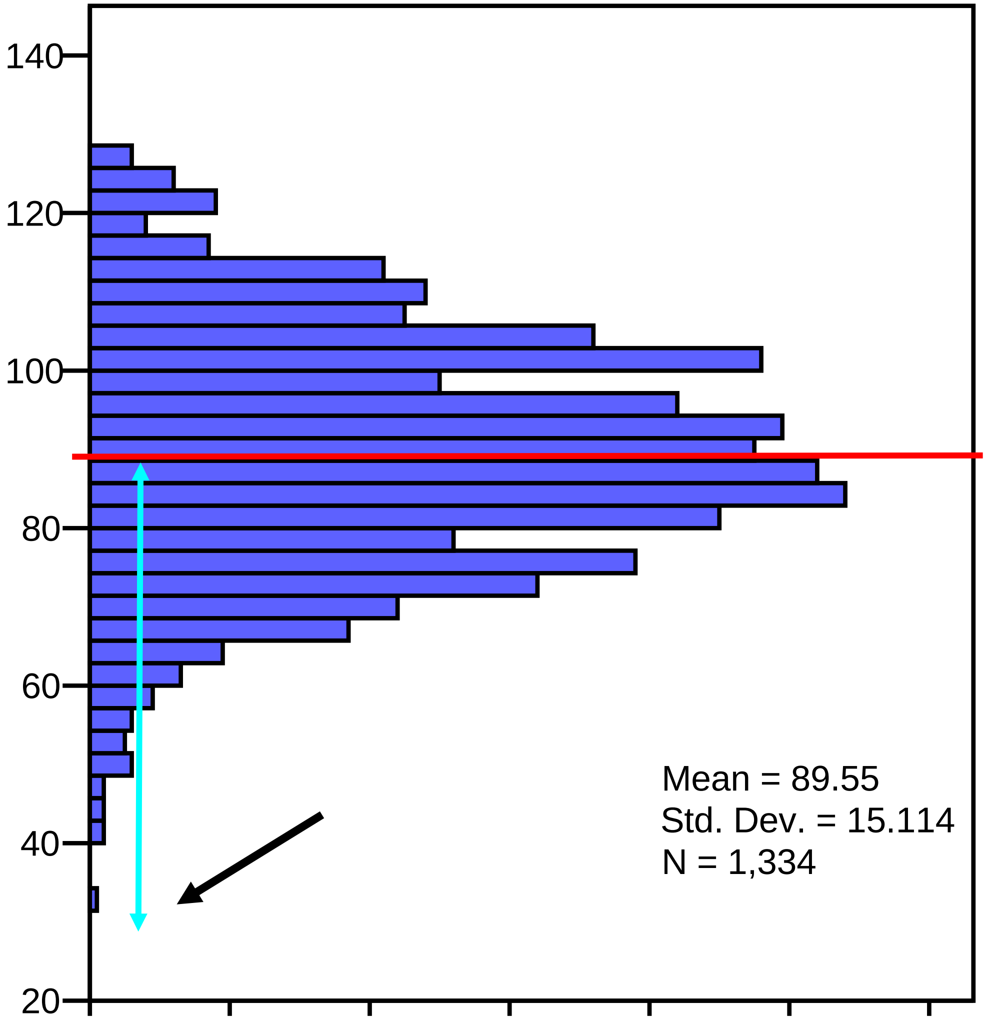
$$y_i = \boxed{\beta_0 + \beta_1(x1_i) + \beta_2(x2_i) + \cdots} + \boxed{e_i}$$

Our focus now

- **Model for the Means (→ Predicted Values):**
  - Each person's expected outcome is a weighted linear function of his/her values on $x1_i$ and $x2_i$ (and any other predictors); each variable is measured once per person (i.e., cross-sectionally)
  - **Estimated constants are called fixed effects** (here, β_0, β_1, and β_2)
  - Number of fixed effects will show up in formulas as $\boldsymbol{k}$ (so $\boldsymbol{k} = \mathbf{3}$ here)

- **Model for the Variance (→ "Piles" of Variance):**
  - $e_i \sim \mathrm{N}(0, \sigma_e^2)$→ ONE (BP) source of residual (unexplained) error
  - In GLMs, e_i has a mean of 0 with some estimated constant variance σ_e^2, is normally distributed, is unrelated to $x1_i$ and $x2_i$, and is **independent** across all observations (which is just one outcome per person here)
  - **There is only ONE source of residual variance in the above GLM because it was designed for only ONE (BP) dimension of sampling!**
    - We should change models when any of these assumptions is not plausible!

An "Empty Means" General Linear Model → Single-Level Model for the Variance



$$y_i = \beta_0 + e_i$$

Filling in values:

$$32 = \underbrace{90}_{\hat{y}_i} + -58$$

Model for the Means

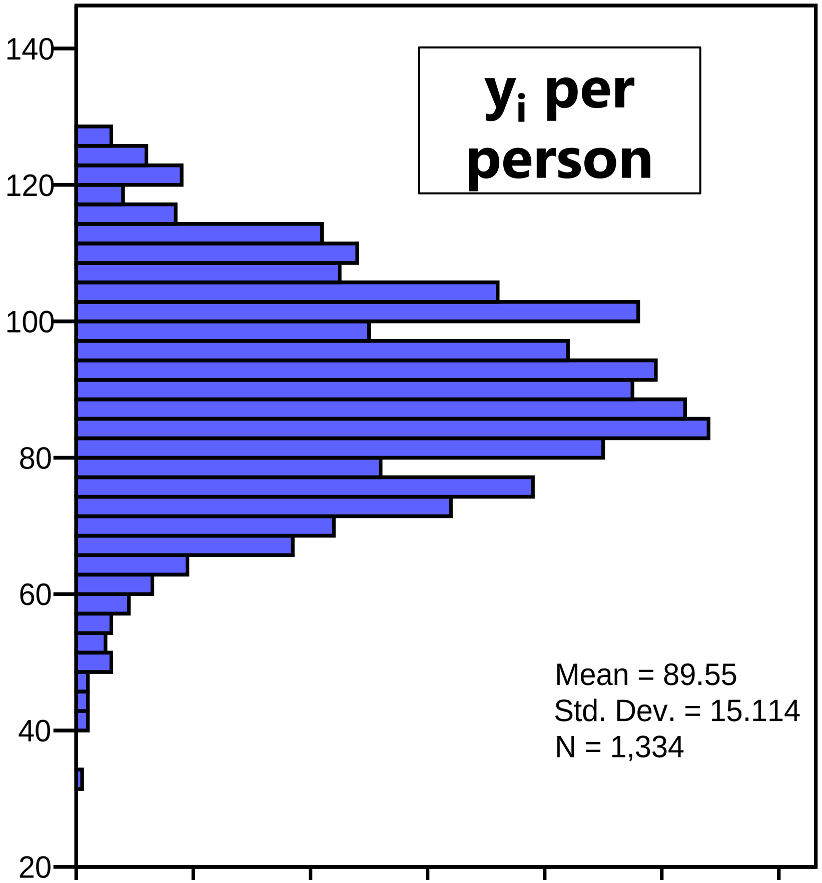
$\hat{y}_i$ **= "y-hat" model-predicted outcome**

y_i **residual ("error") variance:**

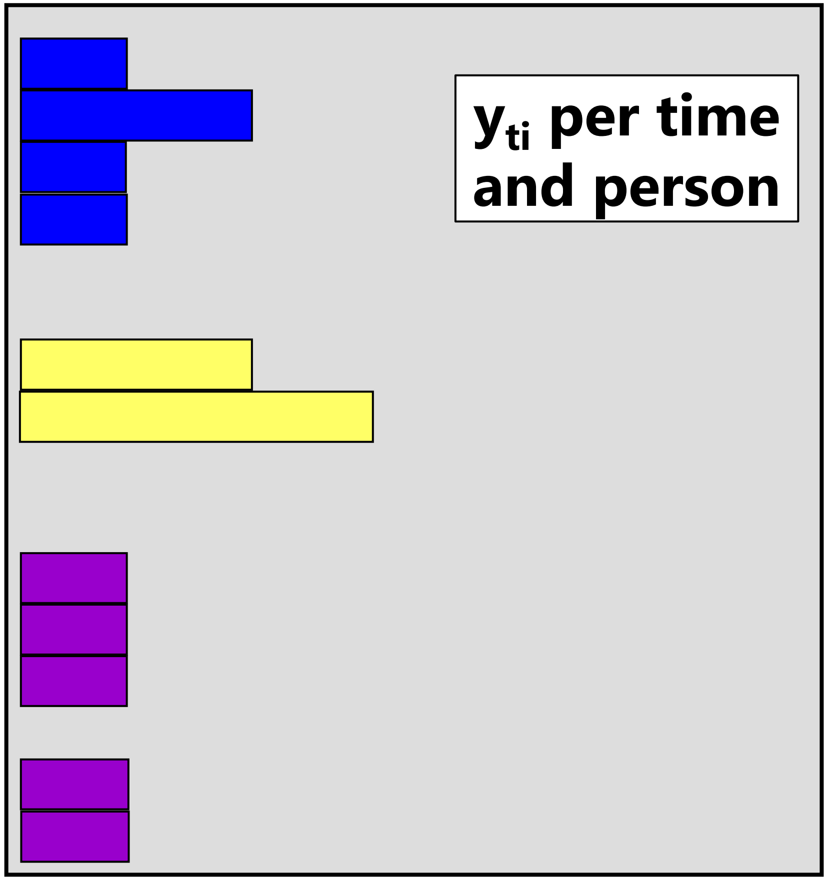
$$\frac{\sum (y_i - \hat{y}_i)^2}{N - 1}$$

Adding Repeated Occasions → Two-Level Model for the Variance

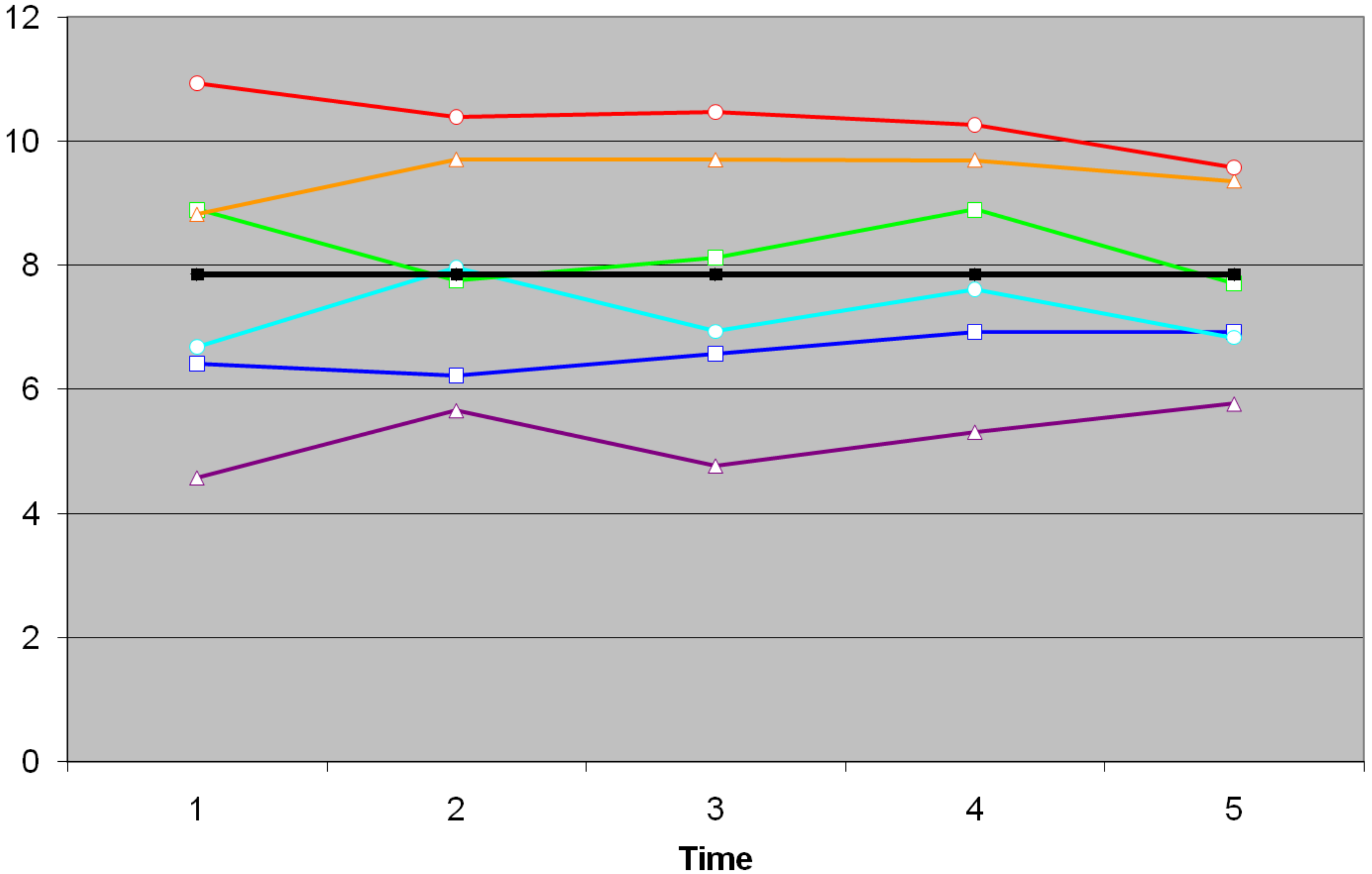
Full Sample Distribution



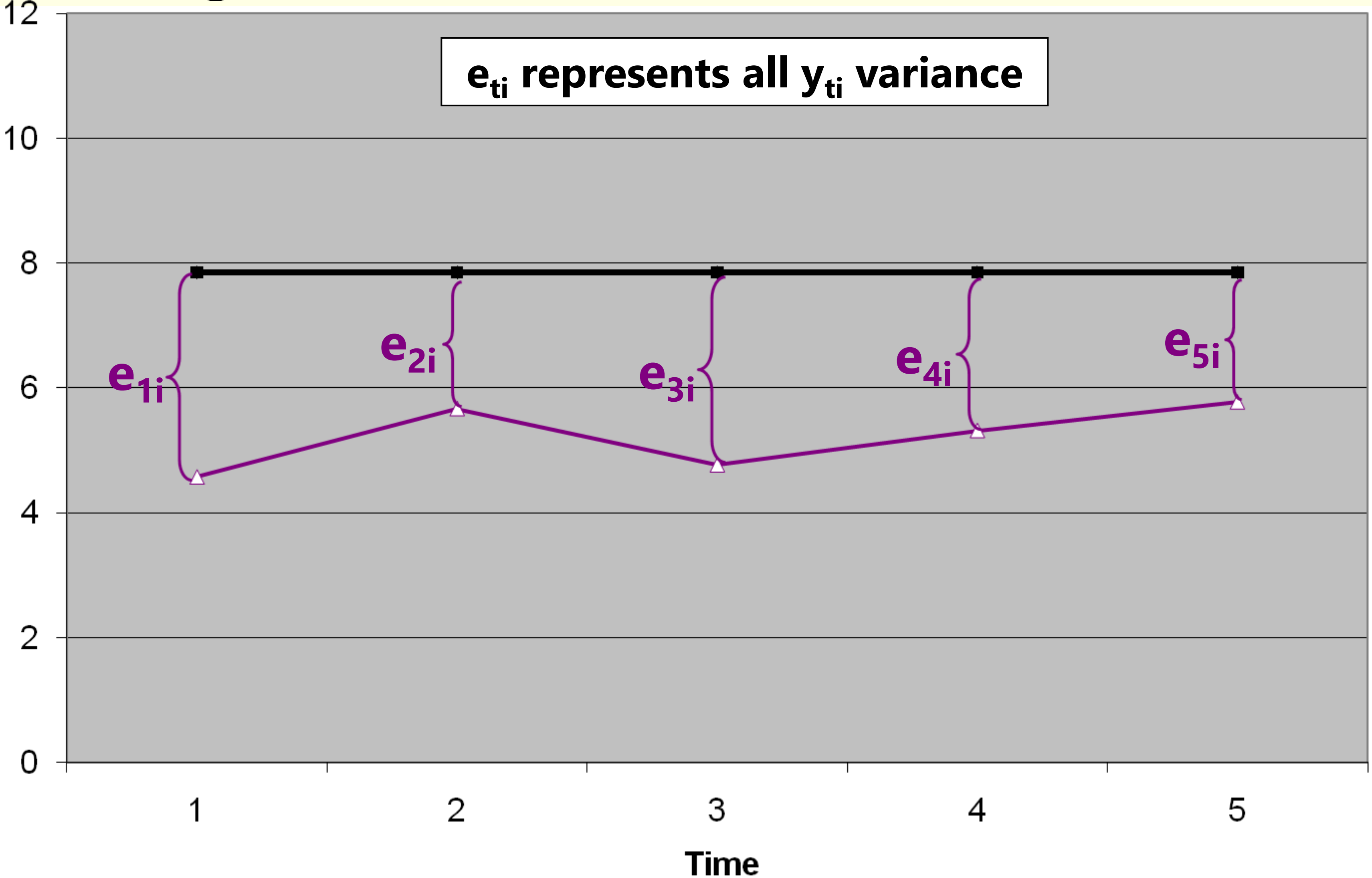
5 Occasions (t); 3 People (i)



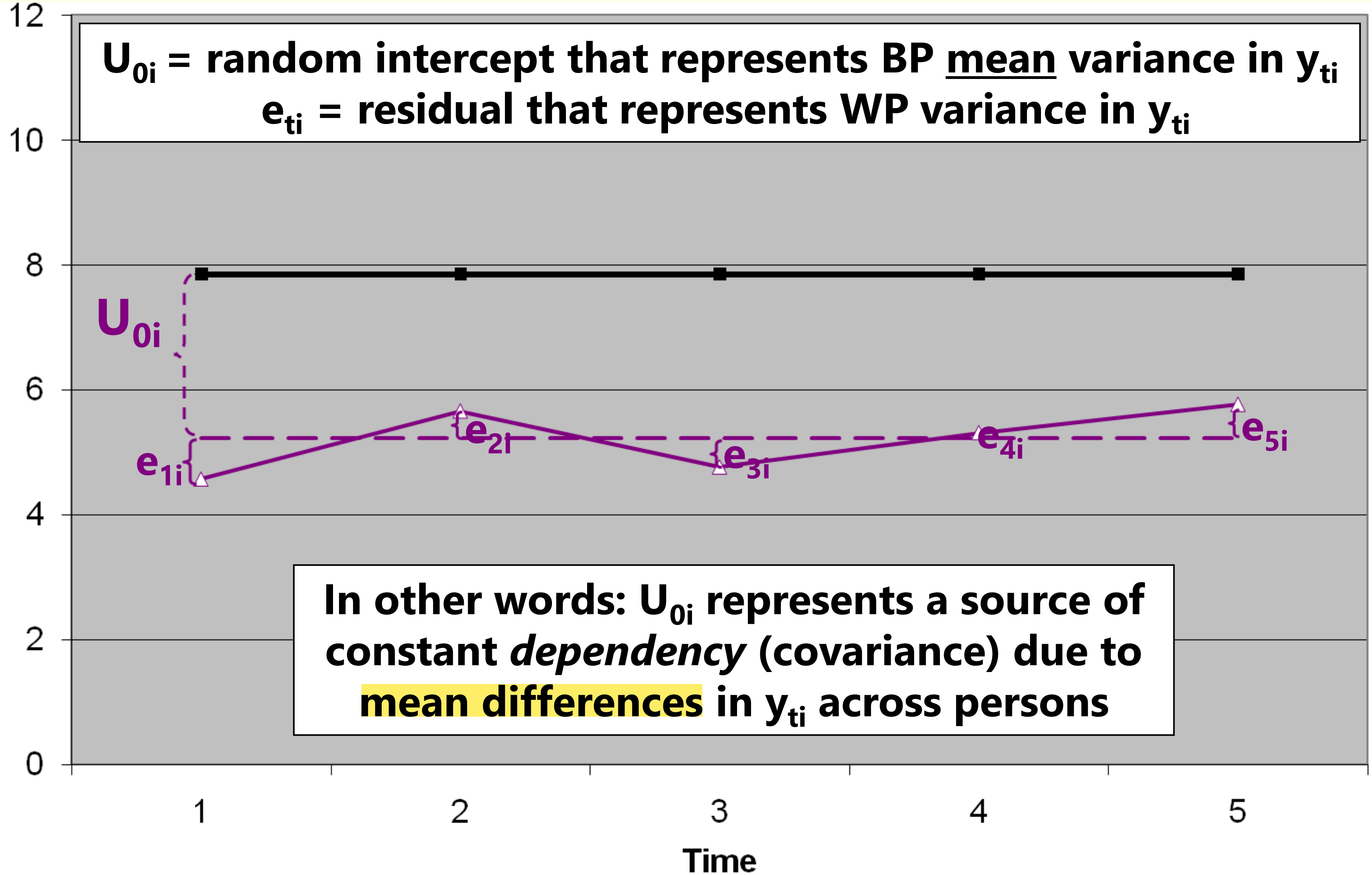
Hypothetical Longitudinal Data



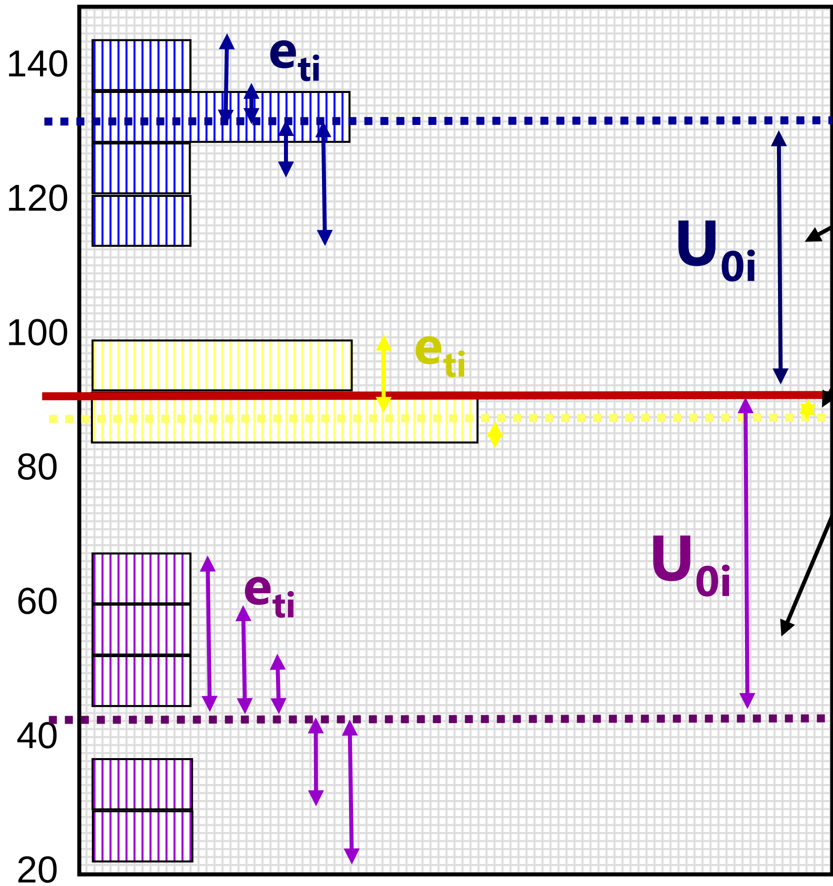
Only One Kind of "Error" in a Single-Level Model for the Variance



Two Distinct Kinds of "Error" in a Two-Level Model for the Variance



Empty Means, Two-Level Model



y_{ti} variance → 2 sources:

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

→ **Between**-Person variance in means

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

→ Cross-sectional source of variance.

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

→ **Within**-Person variance

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

Empty Means Models: Single-Level vs. Two-Level

- Empty Means, **Single-Level 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 Means, **Two-Level** Model (for 2+ occasions):

$$y_{ti} = \underbrace{\beta_0}_{\text{mean of person means}} + U_{0i} + e_{ti}$$

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

A "Random Intercept" Model for the Variance

Total Predicted Data Matrix is called V Matrix, and each person gets their own!

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

***N* = total obs**
***n* = # occasions**
(5 here)

<u>Level 2, BP Variance</u>

Unstructured **G Matrix**

(**RANDOM** statement)

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

1 Random Intercept Variance only

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

<u>Level 1, WP Variance</u>

Diagonal (VC) **R Matrix**

(**REPEATED** statement)

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

1 Residual Variance only

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

To be added to **R** in order to form **V**, **G** is pre- and post-multiplied by an ***N* x 1 Z** matrix that holds the values of the predictors with random effects (just the intercept here): $\mathbf{V}_i = \mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i^T + \mathbf{R}_i$

→ it uses the available data points.

Intraclass Correlation (ICC)

ICCs for two-level longitudinal data:

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

this in the first step.

↳ people differ from each other in more than 1 way.

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

V matrix

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

VCORR Matrix

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

- ICC = Proportion of total variance that is **between persons**
  ↳ clusters
- ICC = **Correlation of occasions** from same person (in VCORR)
- ICC is a standardized way to express *dependency due to person mean differences* → **effect size for constant person dependency**

Review of Longitudinal Multilevel Models

- Topics:
  - Concepts and terminology in longitudinal models
  - Modeling person dependency
  - Fixed and random intercepts
  - **Fixed and random time slopes**
  - Time-invariant predictors
  - Details (ML vs REML, significance tests, model comparisons)

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):

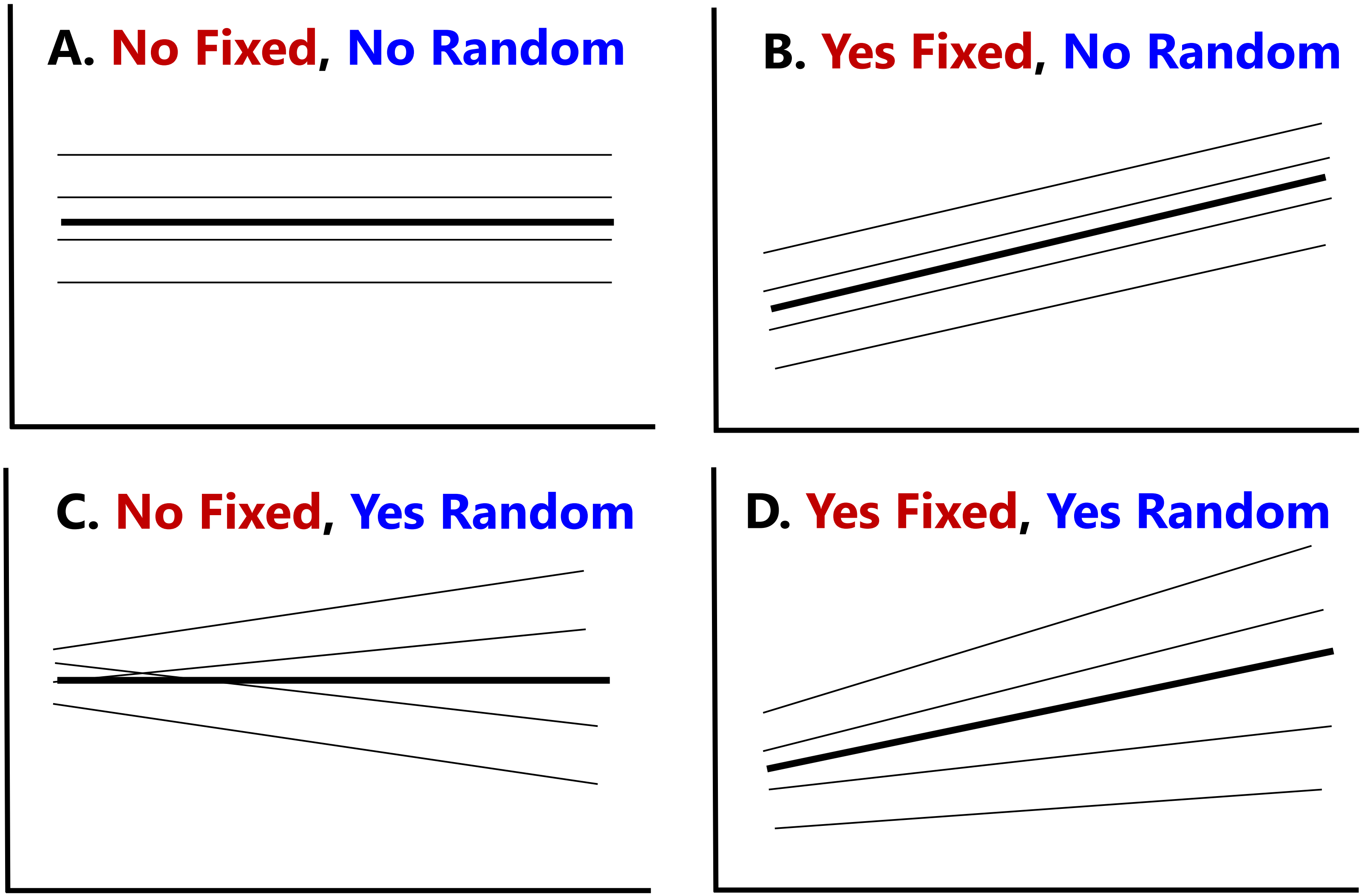
1. **Is there an effect of time on average?**
   - Is the line connecting the sample means not flat?
   - If so, you need **FIXED** effect(s) of time

2. **Does the average effect of time vary across individuals?**
   - Does each individual need their *own* version of that line?
   - If so, you need **RANDOM** effect(s) of time

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

Fixed and Random Effects of **Time**

(Note: The intercept is random in every figure)



Two-Level Model Using Multilevel Notation: A. 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 = **mean of person means (because no predictors yet)**

Random Intercept = **individual-specific deviation from predicted mean**

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

3 Parameters:

Model for the Means (1):

- Fixed Intercept γ_{00}

Model for the Variance (2):

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

Composite equation:

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

B. Fixed Linear Time, Random Intercept Model
(4 parameters: slope for change over 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 $\tau_{U_0}^2$

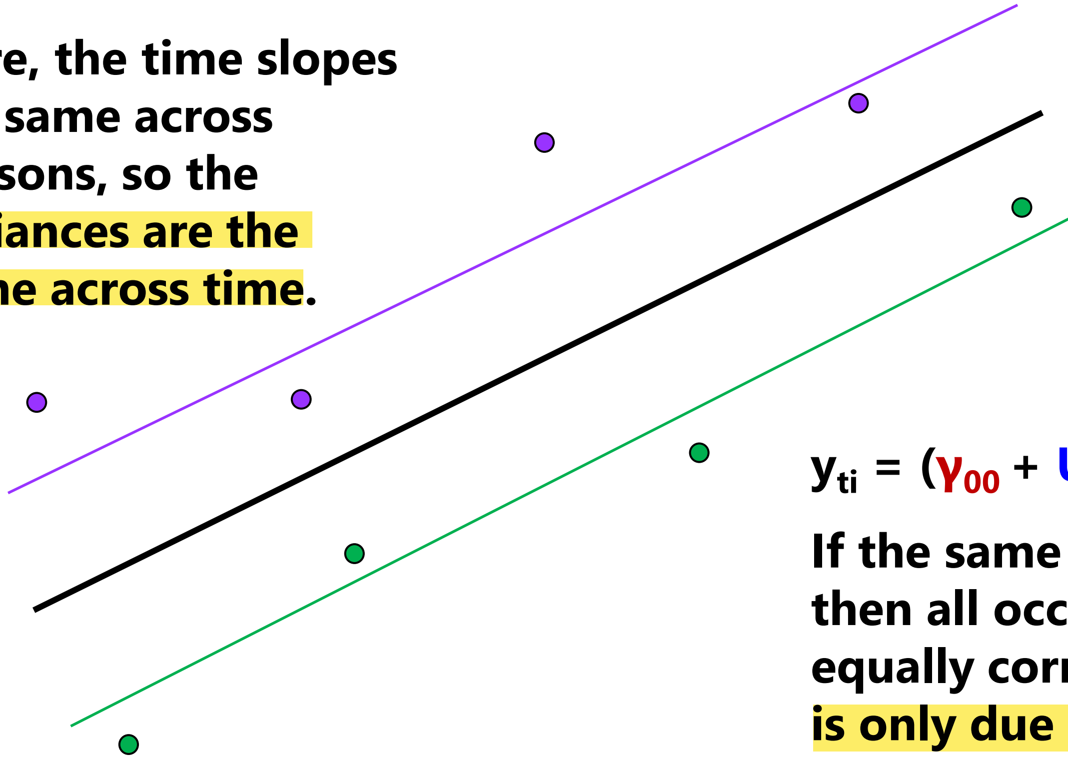
Composite Model

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

Because the effect of time is **fixed**, everyone is predicted to change at exactly the same rate → parallel lines!

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.



Combined V matrix:

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

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

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.

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

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

Multilevel Model

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 τ_{U01}

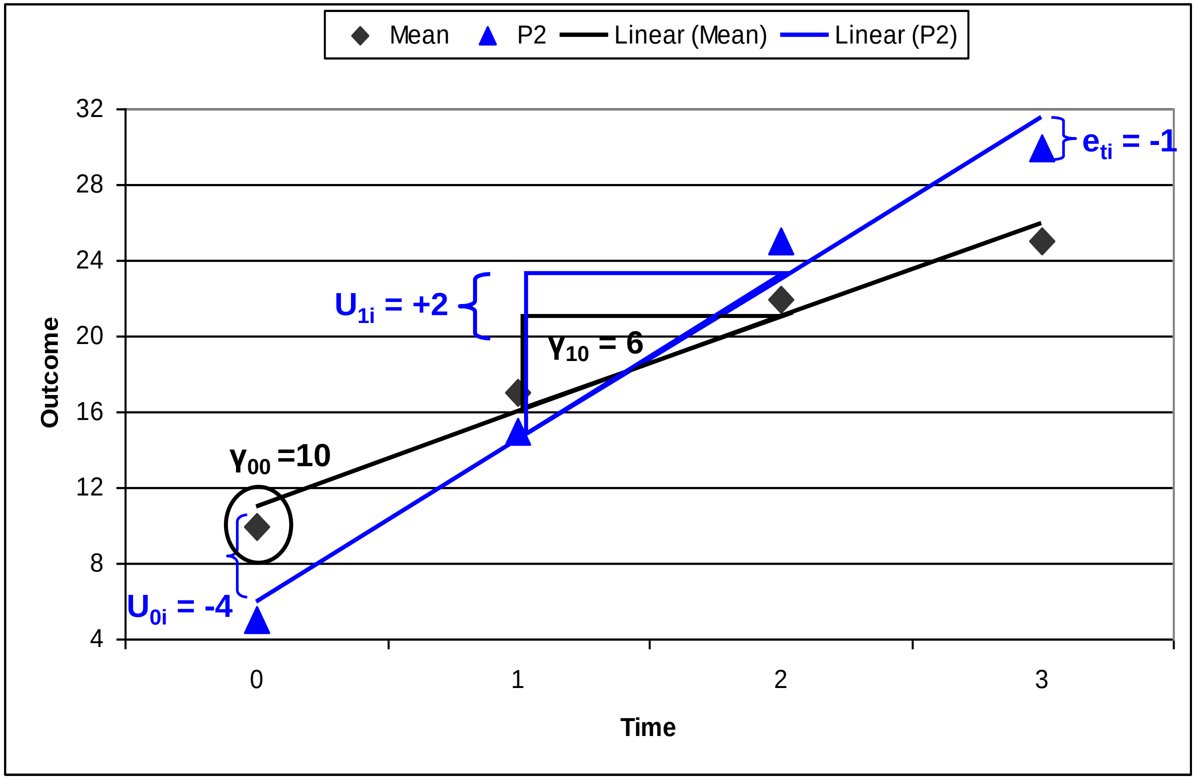
Composite Model

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

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

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 = σ^2_e

Random Linear Time Models Imply:

- **People differ from each other systematically in TWO ways**—in intercept (U_{0i}) and time 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 (or else a different **R** matrix is needed):

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

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

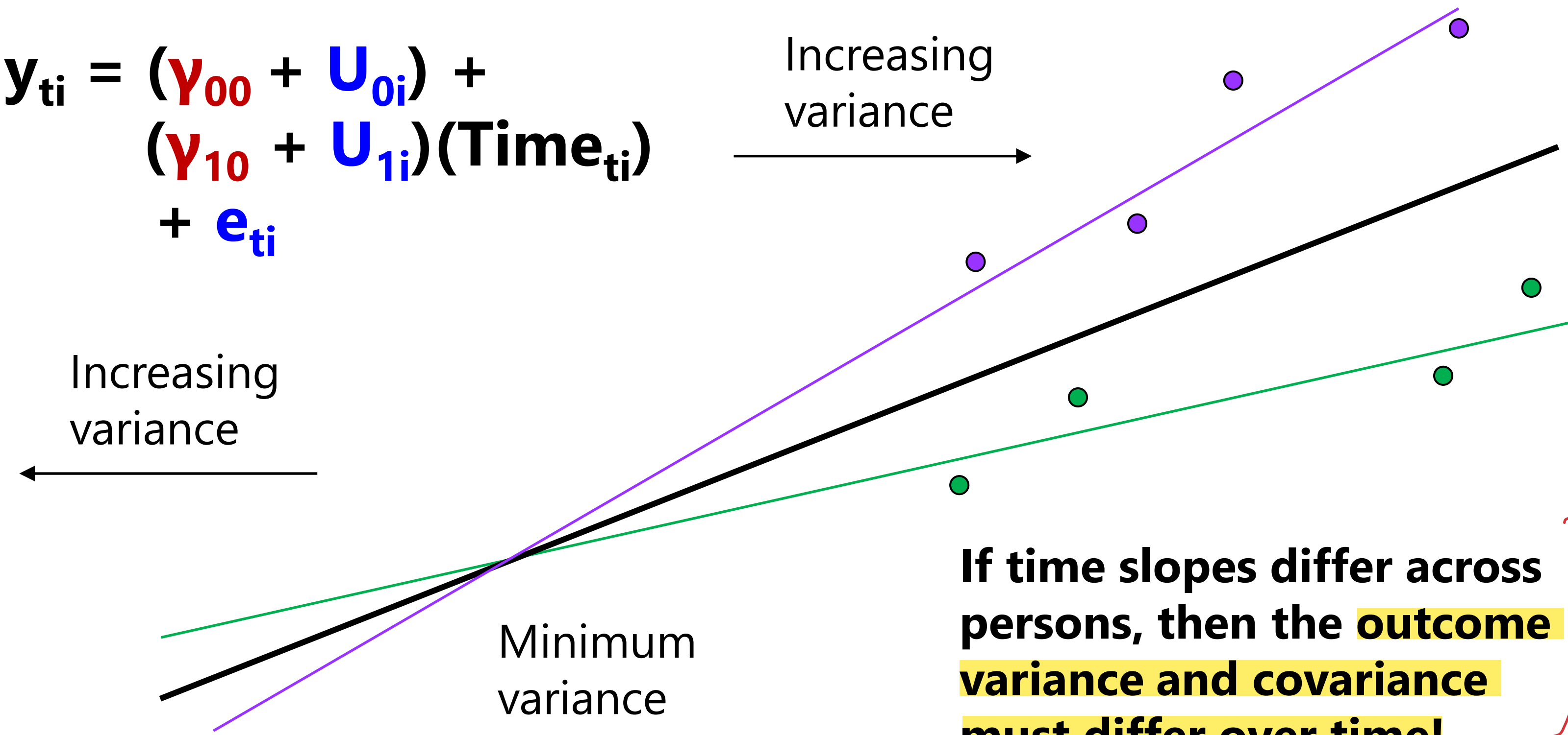
Level-1 **R** matrix: REPEATED TYPE=VC

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

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

Choices in Modeling Variance: Random Intercepts and Time Slopes

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



If time slopes differ across persons, then the outcome variance and covariance must differ over time!

We are introducing heterogeneity of variance across WP outcomes

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…

Random Linear Time Model
(6 parameters: effect of time is 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$\left[y_{time}\right]$

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

$\mathbf{V}_i$ matrix is time-specific!

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

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

It is creating time-specific variance.

$\mathbf{Z}_i$ = *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)

Var and cov become functions of time → we can use any available data point

Each person can potentially have his/her own V_i matrix.

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

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

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

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

- 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 many models based on the idea of a "lag" won't work for unequal-interval time (but AR1 can be modified to work)

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

Modeling correlation across occasions.

- 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 ← → Good fit

Univariate RM ANOVA

Multivariate RM ANOVA

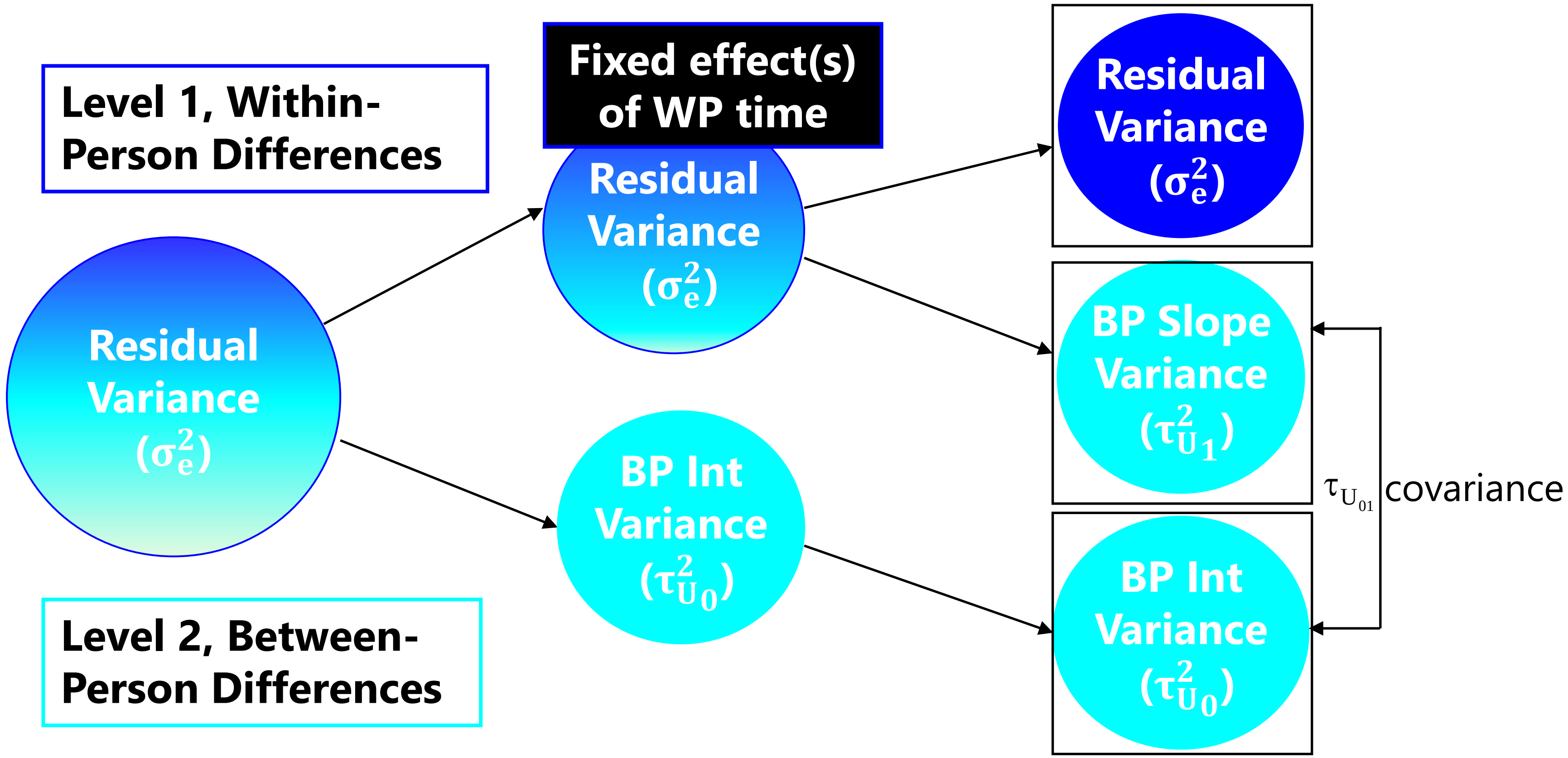
The Bigger Picture

- **Random effects** (new "piles" of variance, partitioned out of what used to be a single residual variance) are used to capture sources of **person dependency**
  - Random **intercept** → **constant** correlation over time due to person mean differences → univariate RM ANOVA
  - Random **time slope(s)** → **non-constant** correlation over time and non-constant variance over time due to between-person differences in rate(s) of change over time
  - Foreshadowing: random time-varying x_{ti} slope → heterogeneity over x_{ti}

- After accounting for BP level-2 random effects (intercepts, and any slopes for change over time), **WP level-1 residuals** are usually assumed **uncorrelated** with **constant variance**
  - But these are both testable assumptions! (fewer alternatives in unbalanced data, largely due to software inflexibility)
  - All sources of person dependency related to time should be addressed before considering other predictors!
  - Any longitudinal model not accounting for person dependency due to intercepts (at a minimum) is most likely to be WAY wrong (AR-CLPM!)

Summary: "Handling" Person Dependency

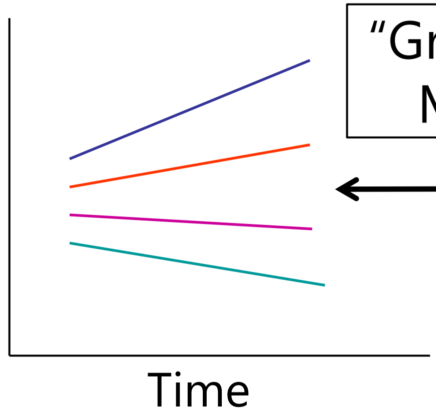
- The process of fitting "unconditional models for time" (fixed and random effects) can be depicted as follows:



Summary: Unconditional Models for Time

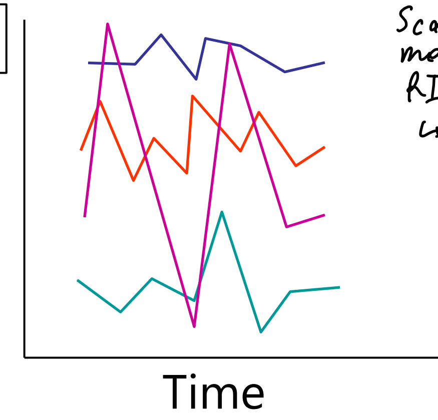
Pure WP Change

"Growth Curve Modeling"



Pure WP Fluctuation

"EMA"



Scale model RI
↳ from σ_i^2

Role of "Time" in the Model for the Means:

- WP Change → describe pattern of ***average*** change (e.g., growth curves)
- WP Fluctuation → describe ***average*** time-specific trends that may not have been expected (e.g., reactivity, day of the week, circadian/schedule effects)

Role of "Time" in the Model for the Variance:

- WP Change → describe ***individual differences*** in change (random effects)
  → this allows variances and covariances to differ over time
- WP Fluctuation → mostly describe pattern(s) of covariance over time
  (may need random effects of time for differing variances)

Families of Nonlinear Change

- Polynomial functions (e.g., $time^2$, $time^3$) → see details on next slides
  - Best suited for time slopes that should change directions (in which time is treated as continuous)

- Piecewise (linear spline) functions
  - Best suited for distinct phases of time (known "knot" points)
  - Otherwise, location of "latent" knots can be model parameters

- Linear effect of log(time) → exponential-ish
  - Good for time slopes that should level off (hit upper or lower asymptote)
  - Adding quadratic log(time) adjusts how fast the time slope levels off

- Latent basis → single slope with estimated nonlinearity
  - In SEM software, for random time slope factor: fix first loading to 0, last loading to 1, and estimate the other loadings to capture proportion of change by each occasion

- Truly nonlinear models (e.g., logistic, exponential)
  - Harder to estimate, particularly for random effects variances

Interpreting Quadratic Fixed Effects

A Quadratic time effect is a two-way interaction: time*time

- Fixed quadratic = "**half the rate of acceleration/deceleration**"
- So to interpret it as how the linear time slope changes per unit time, **you must multiply the quadratic slope coefficient by 2**
- If fixed linear time slope = 4 at time 0, with quadratic slope = 0.3?
  - Instantaneous linear rate of Δ at time 0 = 4.0, at time 1 = 4.6, at time 2 = 5.2

Intercept (Position) at Time T: $y_T = 50.0 + 4.0T + 0.3T^2$

First Derivative (Velocity) at Time T: $\frac{d\,y_T}{d(T)} = 4.0 + 0.6T$

Second Derivative (Acceleration) at Time T: $\frac{d^2 y_T}{d(T)} = 0.6$

$$y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ$$

$$\text{Effect of X} = \beta_1 + \beta_3 Z$$

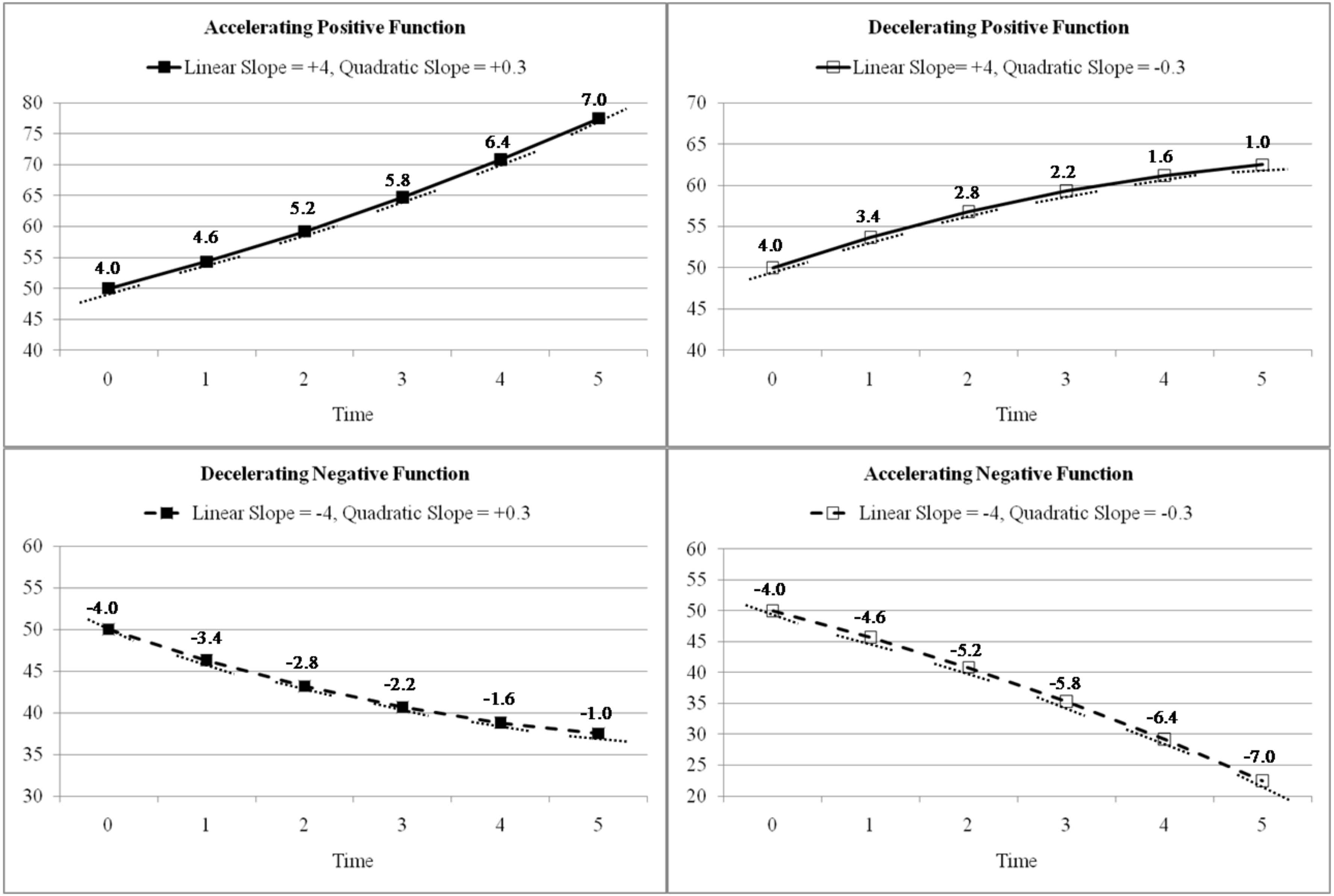
$$\text{Effect of Z} = \beta_2 + \beta_3 X$$

$$y_T = \beta_0 + \beta_1 \text{Time}_T + \_\_\_\_ + \beta_3 \text{Time}_T^2$$

$$\text{Effect of Time}_T = \beta_1 + 2\beta_3 \text{Time}_T$$

- **Why?** Left → because derivatives; Right → because time^2 is an interaction!
- Because time is interacting with itself, there is no second "main effect" in the model for the interaction to modify. So the quadratic time slope gets applied twice when added to the one (main) linear time slope

Examples of Fixed Quadratic Time Trends



Summary: Unconditional Models for Time

- Each source of correlation or dependency goes into a new variance component (or "pile" of variance) until each source meets the usual assumptions of GLM: normality, independence, constant variance
- **Example two-level longitudinal model:**

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

Multiple BP time slope variances are possible…

Soon we will add predictors to account for each pile!

Review of Longitudinal Multilevel Models

- Topics:
  - Concepts and terminology in longitudinal models
  - Modeling person dependency
  - Fixed and random intercepts
  - Fixed and random time slopes
  - **Time-invariant predictors**
  - Details (ML vs REML, significance tests, model comparisons)

Modeling Time-Invariant Predictors

- Which independent variables can be time-invariant predictors?
  - Aka, "**person-level**" or "**level-2**" or predictors ($\boldsymbol{x_i}$) in two-level models
  - Includes substantive predictors, controls, and predictors of missingness
  - Includes anything that either **does not change across time**, or that might change across time but that **you've only measured once** (you may need to argue why this is conceptually ok or limit conclusions accordingly)
  - Also includes **BP variance in time or time-varying predictors** (stay tuned)

- All predictors should be **centered** so that 0 values are meaningful:
  - This is needed to create a meaningful fixed/random intercept, and/or meaningful fixed main effects of predictors also included in interactions
    - e.g., if fixed effects of X, Z, and X*Z, the main effect of X is specifically for Z=0
  - **Quantitative** predictors can be **centered at any constant**, such as the sample mean (common, and useful if it has an unfamiliar scale) or any meaningful reference (better for translating across studies)
  - **Categorical** predictors can have their **dummy-code contrasts** created for you as "factor" variables (e.g., SAS CLASS, SPSS BY, STATA i.), but not in Mplus; I do not like ± 1 coding for group differences (because then 0 = ???)
    - I find indicator or sequential dummy-coding variants most useful

Beware of Missing Predictors

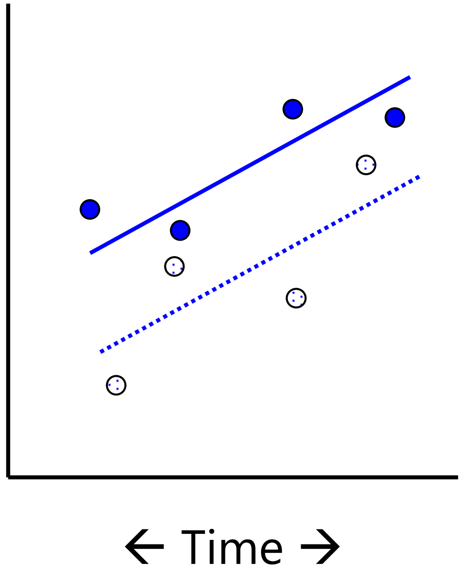
- Any cases missing model predictors that are not part of the joint likelihood* will not be used in that model
  - Not great for time or time-varying predictors (Missing At Random-ish)
  - Really bad for time-invariant predictors (listwise deletion, MCAR)

- Better options for missing predictors:
  - *Bring the predictor into the joint likelihood (only possible in software for truly multivariate MLMs, such as Mplus, or in SEM programs)
    - Its mean, variance, and covariances "get found" as model parameters
    - Predictor then has distributional assumptions (default is multivariate normal), which may not be plausible for all predictors
    - Mplus v. 8 still will not do this for non-normal "predictors" in multivariate MLM
  - Multiple imputation (and analysis of *each* imputed dataset)
    - Imputation also requires distributional assumptions for imputed variables!
    - Also requires all parameters of interest for the analysis model to be in the imputation model, too (which is problematic for interactions or random effects)

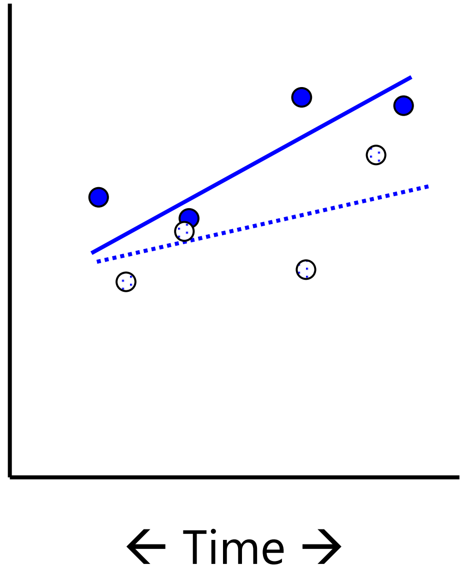
The Role of Time-Invariant Predictors in the **Model for the Means**

- **In Within-Person Change Models →** Moderate growth curve

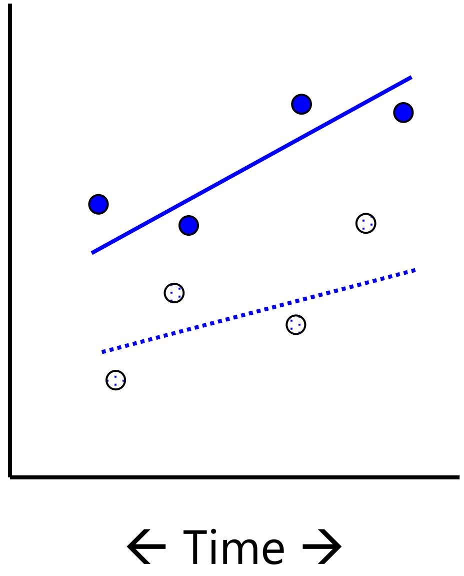
Main effect of x_i, no interaction with time



Interaction with time, main effect of x_i?



Main effect of x_i, and interaction with time

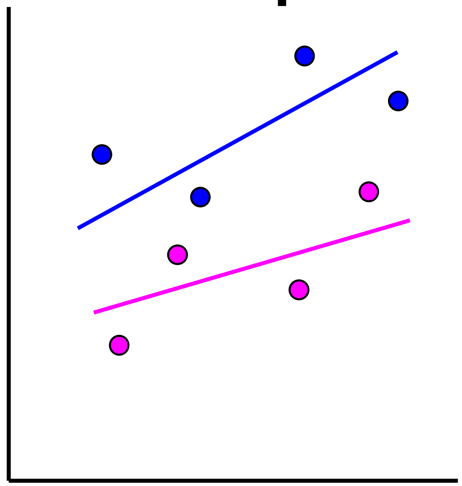


The Role of Time-Invariant Predictors in the **Model for the Variance**

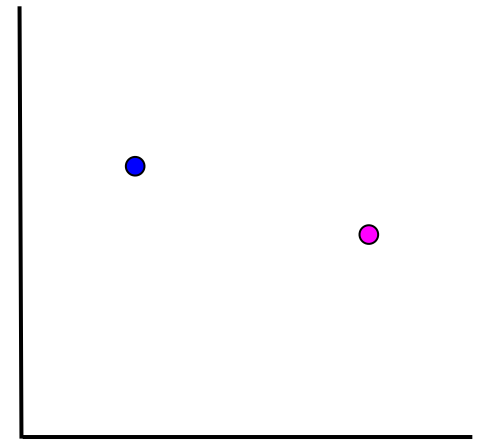
- Beyond fixed effects in the model for the means, time-invariant predictors can be used to allow **heterogeneity of variance** at their level or below in "**location–scale models**"
- e.g., Group as a predictor of heterogeneity of variance:
  - **At level 2**: *Amount* of individual differences in intercepts and/or slopes differs between control and treatment (assumed constant by default!)
  - **At level 1**: *Amount* of within-person residual variation differs between control and treatment (assumed constant by default!)
    - In within-person **fluctuation** model: differential volatility over time
    - In within-person **change** model: differential volatility/inconsistency remaining after controlling for fixed and random effects of time
- These models are harder to estimate and may require custom algorithms (e.g., SAS NLMIXED; in Mplus v 8+ using "logV")

Why Level-2 Predictors Cannot* Have Random Effects in Two-Level Models

Random Slopes for Time



Random Slopes for Group?



You cannot make a line out of a dot, so level-2 effects cannot vary randomly over persons.

* *Level-2 predictors can be included as predictors of heterogeneity of variance, which technically is a random slope of sorts (but interpretation is different)*

Sources of Explained Variance by Person-Level-2 Time-Invariant Predictors

- **Fixed effects of level-2 predictors *by themselves*:**
  - Level-2 (BP) main effects reduce level-2 random intercept variance
  - Level-2 (BP) interactions also reduce level-2 random intercept variance
- **Fixed effects of *cross-level interactions* (level-1* level-2):**
  - If a 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 pred1**time*, pred2**time*, and pred1*pred2**time* can each reduce the level-2 random linear time slope variance
  - If the 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 $time^2$ does not have a level-2 random slope, then pred1*$time^2$, pred2*$time^2$, and pred1*pred2*$time^2$ will reduce the level-1 residual variance → Different quadratic slopes by pred1 and pred2 create better level-1 trajectories, thus reducing level-1 residual variance around the trajectories

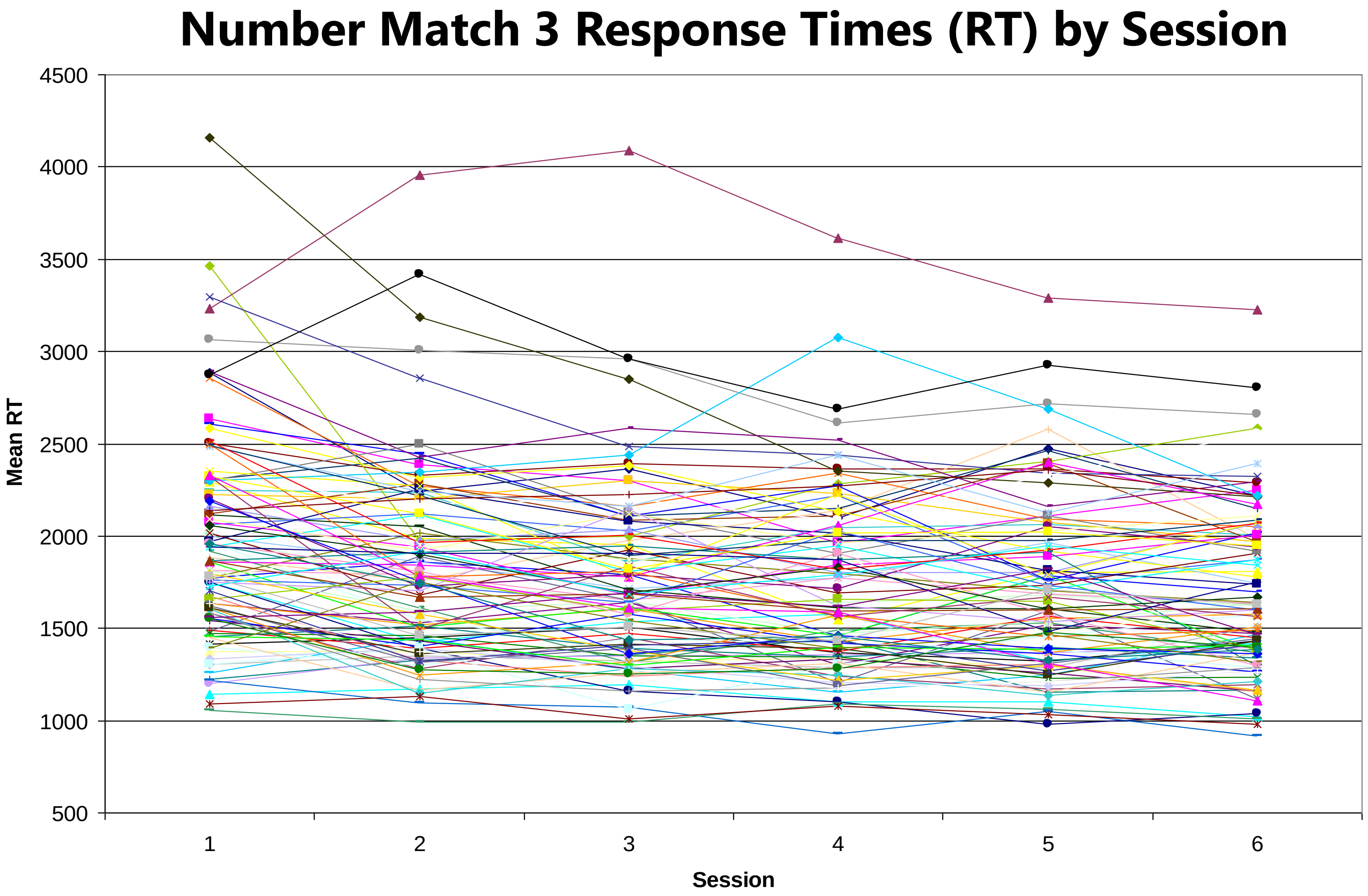
Variance Explained… Continued

- **Pseudo-R^2** is named that way for a reason… piles of variance can shift around, such that **it can actually become negative**
  - Sometimes is a sign of model mis-specification (but not always)
  - See Rights & Sterba (2019, 2020) for alternative calculation of pseudo-R^2
    - Ensure positive R^2 values, but they don't quantify R^2 for slope variances (boo)

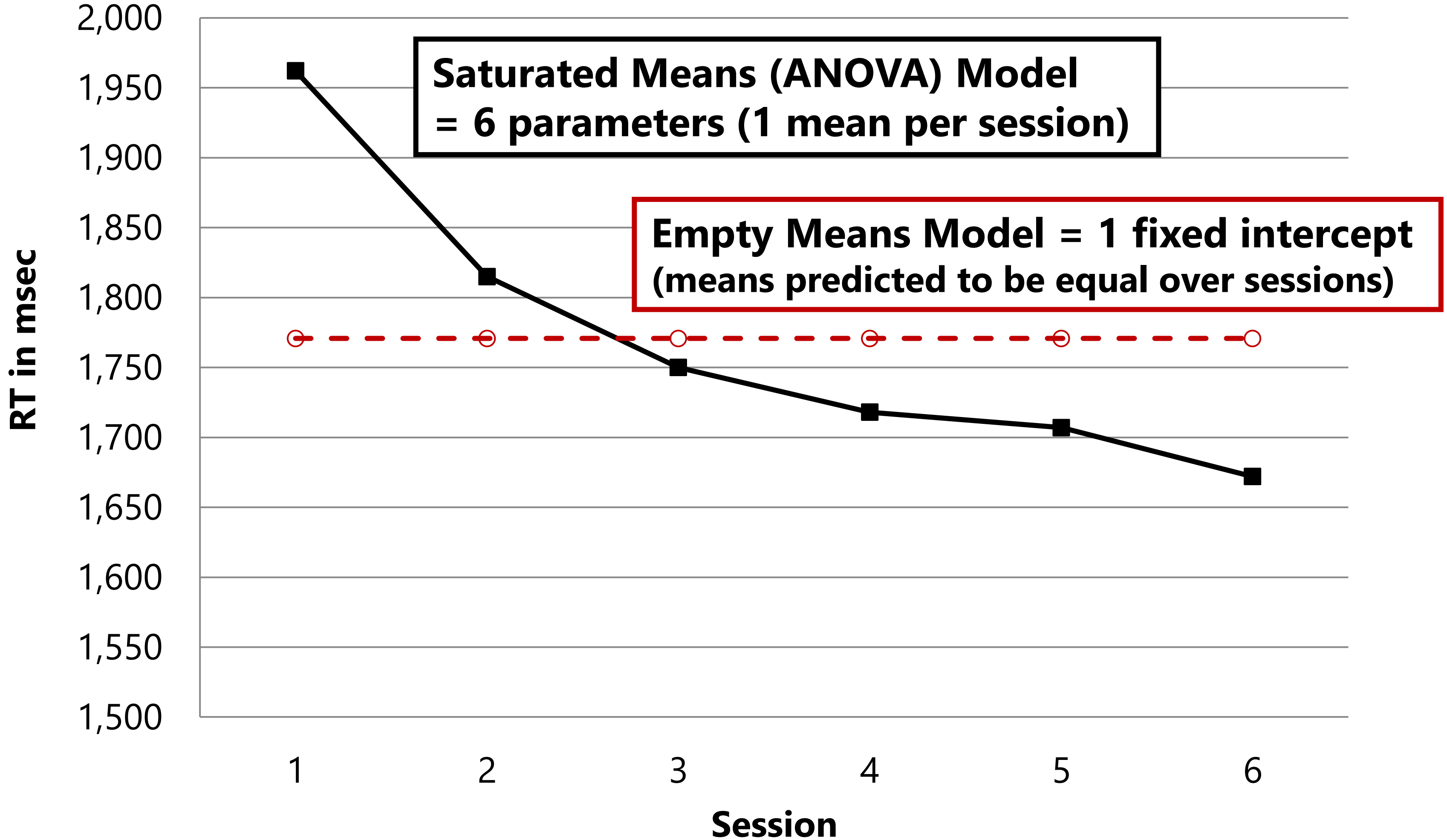
- **A simple alternative: Total R^2** (Singer & Willett, 2003)
  - Generate model-predicted $\hat{y}_{ti}$ from fixed effects only (NOT including random effects, so no cheating) and correlate it with observed y_{ti}
  - Then square that correlation → total R^2 (same as in GLM regression)
  - Total R^2 = total reduction in overall outcome variance across levels
  - Can be "unfair" in models with large unexplained sources of variance (i.e., for sampling dimensions you didn't have predictors for)

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

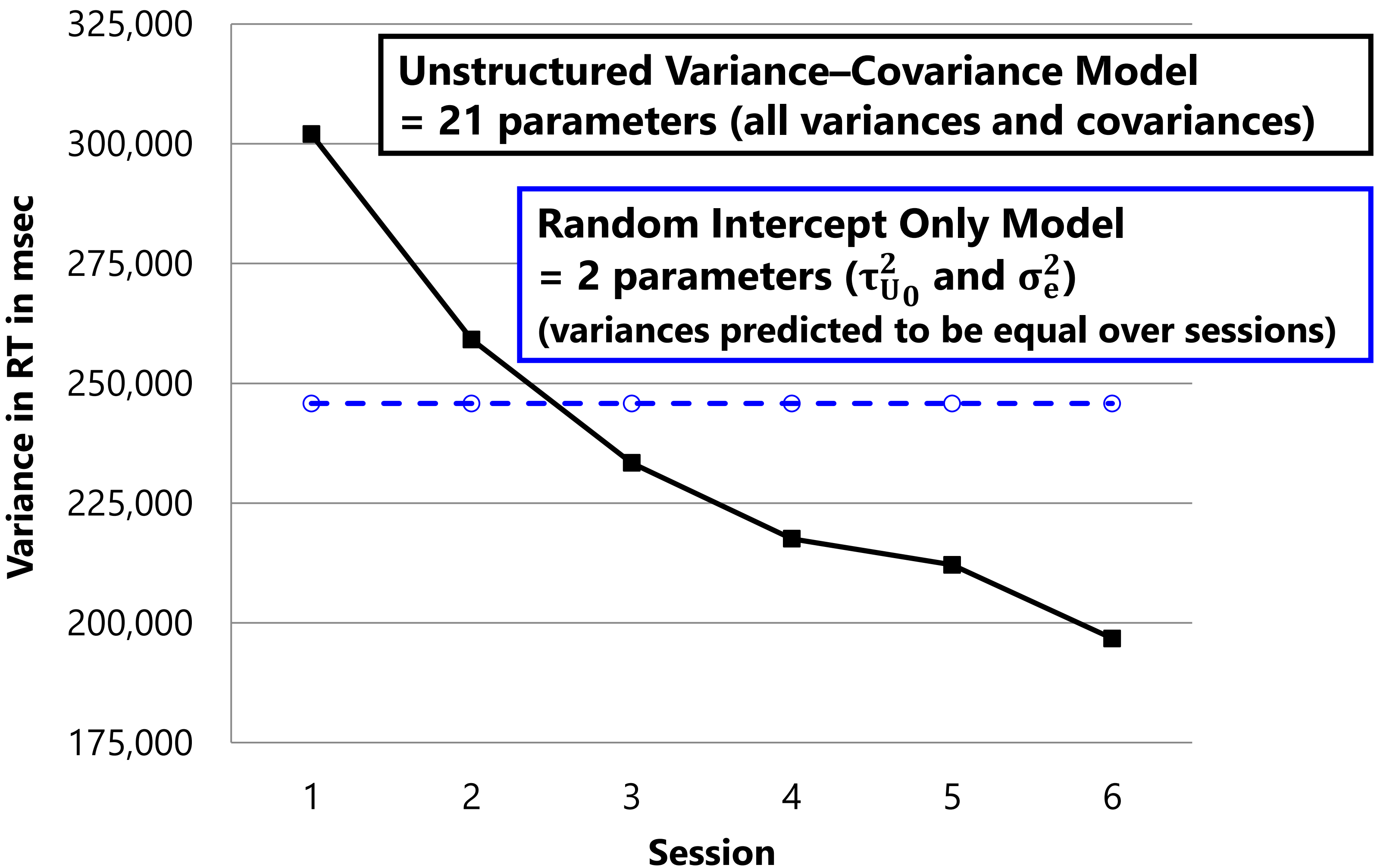
Example 1: Individual Trajectories 101 older adults, 6 occasions within 2 weeks



Example 1 Mean RT by Session: Baseline Models for the Means



Example 1 Variance in RT by Session: Baseline Models for the Variance



Random Quadratic Time Unconditional Model

Level 1: $RT_{ti} = \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti}$

Level 2 Equations (one per β):

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

Intercept for person *i* — Fixed (mean) **Intercept** — Random (Deviation) **Intercept**

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

Linear Time Slope for person *i* — Fixed (mean) **Linear Slope** — Random (Deviation) **Linear Slope**

$\beta_{2i} = \gamma_{20} + U_{2i}$

Quadratic Time Slope for person *i* — Fixed (mean) **Quad Slope** — Random (Deviation) **Quad Slope**

Time = session − 1
REML estimation using stacked data (univ MLM)
U_i covariances also estimated

Fixed Effect Subscripts:
1st = which level-1 term
2nd = which level-2 term

# of Possible Time-Related Slopes by # of Occasions (*n*):

Fixed time slopes = *n* – 1
Random time slopes = *n* – 2

Need *n* = 4 occasions to fit random quadratic time model

Adding Reasoning (0=22) as a Time-Invariant Predictor: Is RT Improvement Predicted by Fluid Intelligence?

Level 1: $RT_{ti} = \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + \gamma_{01}Reas_i + U_{0i}$

- β_{0i}: **Intercept** for person *i*
- γ_{00}: **Fixed Intercept when Time=0 and Reas=22**
- γ_{01}: **Δ in Intercept per unit Δ in Reas**
- U_{0i}: **Random (Deviation) Intercept after controlling for Reas**

$\beta_{1i} = \gamma_{10} + \gamma_{11}Reas_i + U_{1i}$

- β_{1i}: **Linear Slope** for person *i*
- γ_{10}: **Fixed Linear Time Slope when Time=0 and Reas=22**
- γ_{11}: **Δ in Linear Time Slope per unit Δ in Reas (=Reas*time)**
- U_{1i}: **Random (Deviation) Linear Time Slope after controlling for Reas**

$\beta_{2i} = \gamma_{20} + \gamma_{21}Reas_i + U_{2i}$

- β_{2i}: **Quad Slope** for person *i*
- γ_{20}: **Fixed Quad Time Slope when Reas=22**
- γ_{21}: **Δ in Quad Time Slope per unit Δ in Reas (=Reas*time²)**
- U_{2i}: **Random (Deviation) Quad Time Slope after controlling for Reas**

Reasoning (0=22) as a Time-Invariant Predictor: Is RT Improvement Predicted by Fluid Intelligence?

Level 1: $RT_{ti} = \beta_{0i} + \beta_{1i}Time_{ti} + \beta_{2i}Time_{ti}^2 + e_{ti}$

Level 2 Equations (one per β):

$$\beta_{0i} = \gamma_{00} + \gamma_{01}Reas_i + U_{0i}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11}Reas_i + U_{1i}$$
$$\beta_{2i} = \gamma_{20} + \gamma_{21}Reas_i + U_{2i}$$

γ_{11} and γ_{21} are known as "**cross-level**" interactions (level-1 predictor by level-2 predictor)

Each fixed slope of reasoning will predict the random U_i variance in its level-2 equation if present, or e_{ti} residual variance otherwise. That's why random slopes should be tested **before** adding cross-level interactions!

- Composite equation:
- $y_{ti} = (\gamma_{00} + \gamma_{01}Reas_i + U_{0i}) +$
  $(\gamma_{10} + \gamma_{11}Reas_i + U_{1i})Time_{ti} +$
  $(\gamma_{20} + \gamma_{21}Reas_i + U_{2i})Time_{ti}^2 + e_{ti}$

Reasoning (0=22) as a Time-Invariant Predictor: Is RT Improvement Predicted by Fluid Intelligence?

$$\begin{aligned} RT_{ti} = &\ (1966 + -27{*}Reas_i + U_{0i}) + \\ &\ (-120 + -3.6{*}Reas_i + U_{1i})Time_{ti} + \\ &\ (13 + 1.2{*}Reas_i + U_{2i})Time_{ti}^2 + e_{ti} \end{aligned}$$

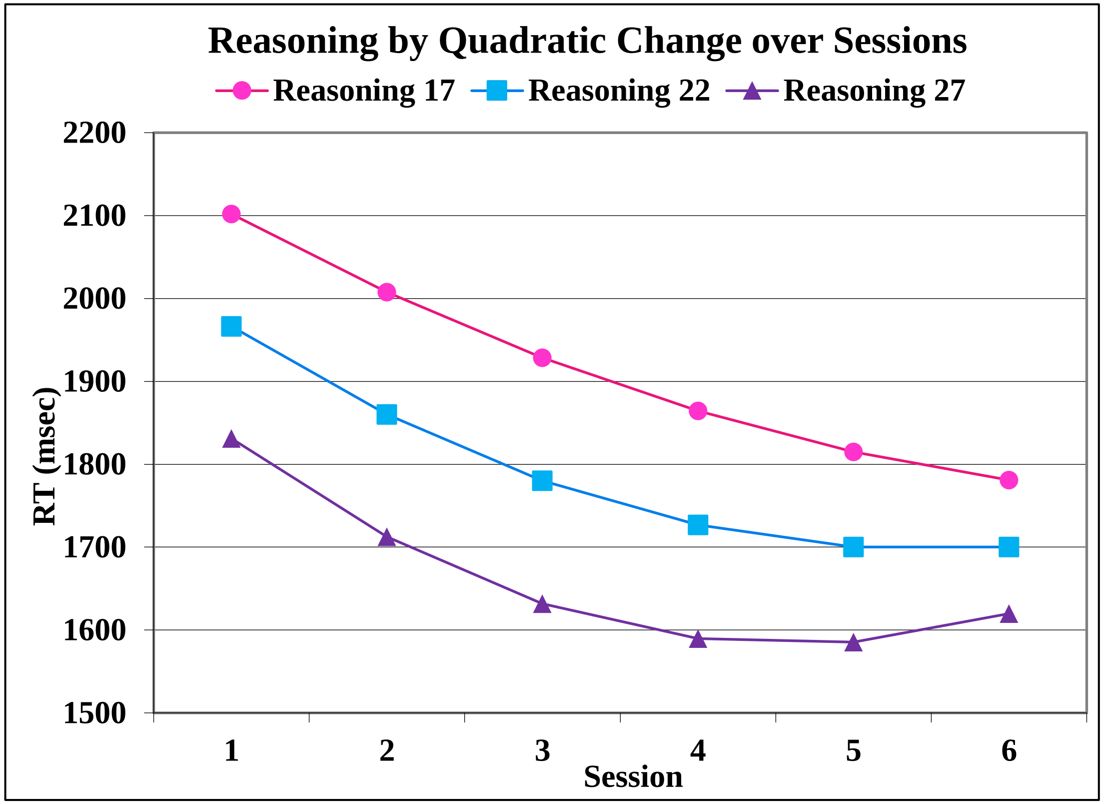
BP Pseudo-R^2 Values:

Intercept U_{0i} = .049

Linear Time U_{1i} = −.006

Quadratic Time U_{2i} = .024

WP Residual e_{ti} = **0**



People with better reasoning:

- started out faster/lower *(intercept at session 1)*,
- improved more initially *(linear slope at session 1)*,
- and had a greater rate of deceleration with practice *(quadratic slope*2!)*

Example 1: Syntax by Univariate MLM Program (in Stacked "Long" Data)

SAS:

```
PROC MIXED DATA=work.Example1 COVTEST METHOD=REML;
     CLASS ID;
     MODEL RT = time timesq reas time*reas timesq*reas / SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT time timesq / GCORR TYPE=UN SUBJECT=ID;
RUN;
```

R (lmer from lme4 package)—using lmertest package, which does provide correct denominator DF:

```
model = lmer(data=Example1, REML=TRUE,
         formula=RT~1+time+timesq+reas+reas
                     +time:reas+timesq:reas+(1+time+timesq|ID))
summary(model, ddf="Satterthwaite")
```

STATA:

```
mixed RT time timesq reas time#reas timesq#reas, || ID: time timesq, ///
      variance reml covariance(un) dfmethod(satterthwaite) dftable(pvalue)
```

SPSS:

```
MIXED RT BY ID WITH time timesq reas
      /METHOD = REML
      /PRINT = SOLUTION TESTCOV
      /FIXED = time timesq reas time*reas timesq*reas
      /RANDOM = INTERCEPT time timesq | COVTYPE(UN) SUBJECT(ID).
```

Fixed Effects of Time-Invariant Predictors

- What about predicting level-1 effects with no random variance?
  - If the random linear time slope is n.s., can I test interactions with time?

This should be ok to do…

$$\beta_{0i} = \gamma_{00} + \gamma_{01}\text{Reas}_i + U_{0i}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11}\text{Reas}_i + U_{1i}$$
$$\beta_{2i} = \gamma_{20} + \gamma_{21}\text{Reas}_i + U_{2i}$$

Is this still ok to do?

$$\beta_{0i} = \gamma_{00} + \gamma_{01}\text{Reas}_i + U_{0i}$$
$$\beta_{1i} = \gamma_{10} + \gamma_{11}\text{Reas}_i$$
$$\beta_{2i} = \gamma_{20} + \gamma_{21}\text{Reas}_i$$

  - **"NO"**: If a level-1 effect does not vary randomly over individuals, then it has "no" variance to predict (so cross-level interactions with that level-1 effect are not necessary); its SE and DDF could be inaccurate SE if $\tau^2_{U_1} \neq 0$
  - **"YES":** Because power to detect random effects is lower than power to detect fixed effects (especially with small L2n), cross-level interactions can still be significant even if there is "no" (≈0) variance to be predicted
  - Saying yes requires new vocabulary…

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

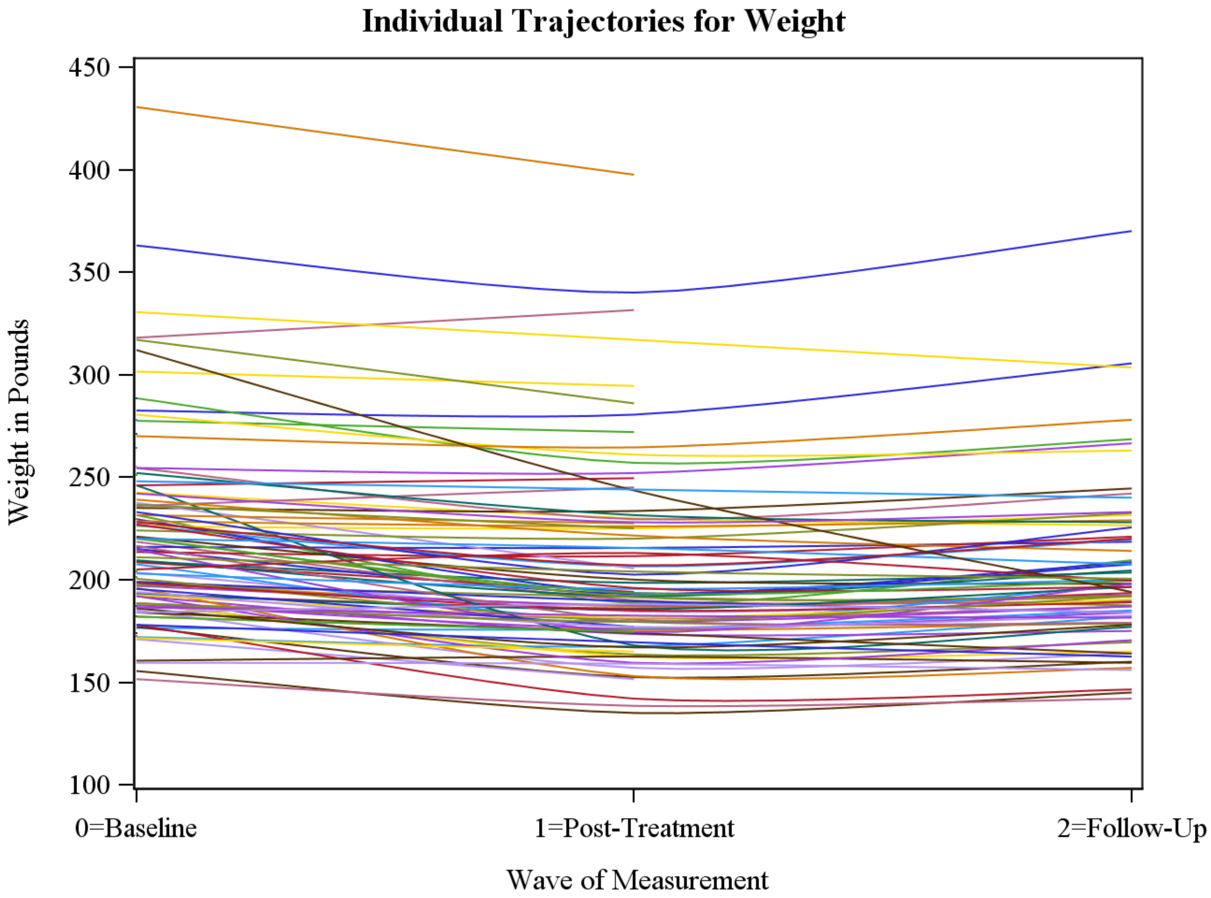
Let's say we have a significant fixed linear effect of time. What happens after we test a group*time interaction?

	Non-Significant Group*Time effect?	Significant Group*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 group*time	---	Linear effect of time is **systematically varying**
Random time initially sig, **still** sig. after group*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).

Example 2: Differences in Weight Loss

- Randomized Control Trial for differences in weight loss (and then maintenance of weight loss) between control and treatment groups
  - 105 persons measured up to 3 occasions → balanced but incomplete data
  - Univariate MLM on stacked data (REML estimation, Kenward-Roger DDF)



Example 2: Piecewise Model for the Means

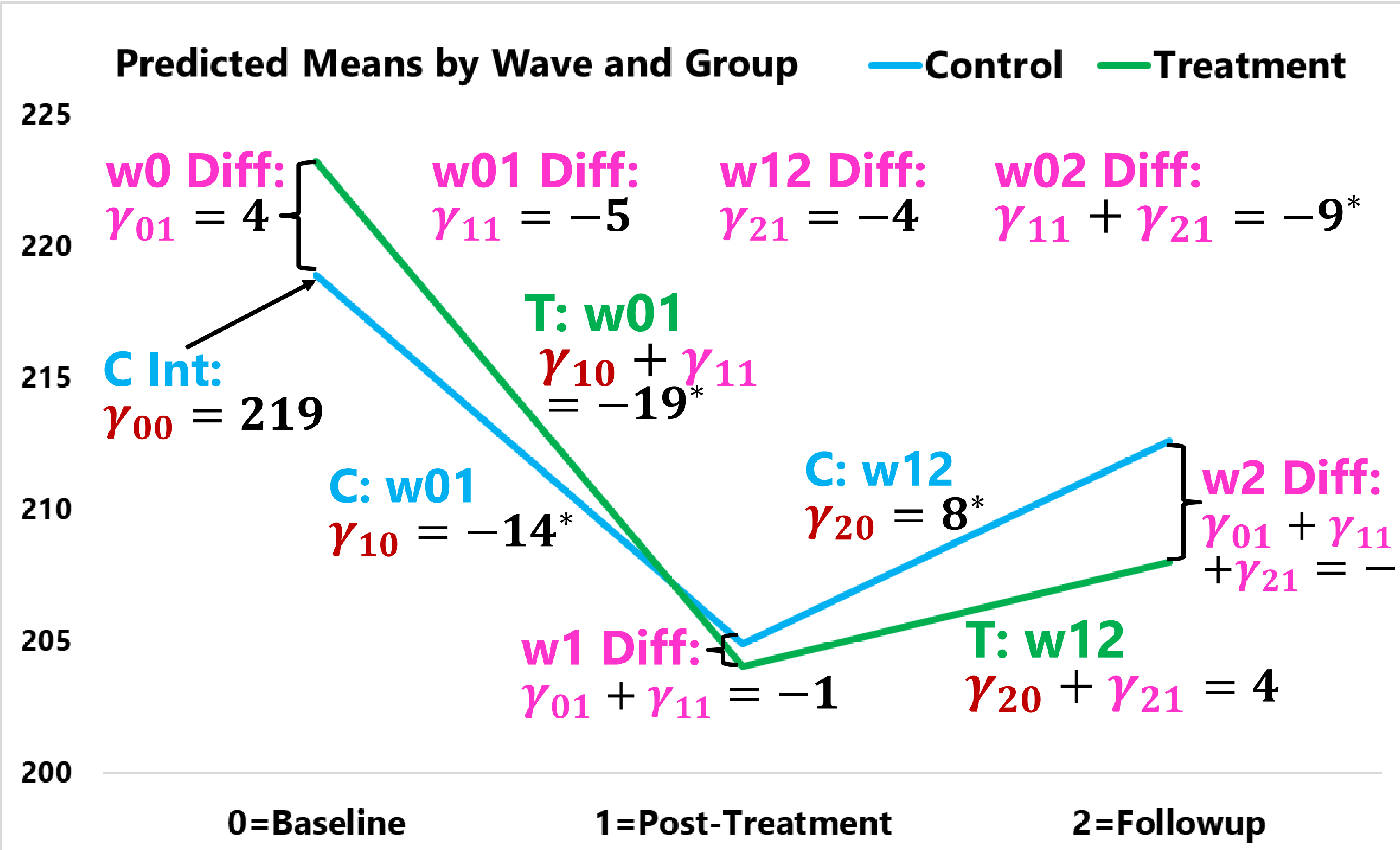
$$\widehat{weight}_{ti} = \gamma_{00} + \gamma_{10}(w01_{ti}) + \gamma_{20}(w12_{ti}) + \gamma_{01}(Tx_i) + \gamma_{11}(w01_{ti})(Tx_i) + \gamma_{21}(w12_{ti})(Tx_i)$$

Wave	w01	w12
0. Baseline	0	0
1. Post-Treat	**1**	0
2. Follow-Up	**1**	**1**

- Estimated fixed effects for control group ($Tx_i = 0$):
  - γ_{00} for intercept: Control weight at baseline
  - γ_{10} for w01 slope: Control change from baseline to post-treatment
  - γ_{20} for w12 slope: Control change from post-treatment to follow-up
- Estimated fixed effects for differences by treatment ($Tx_i = 0\ vs\ 1$):
  - γ_{01} for Tx slope: difference in weight at baseline
  - γ_{11} for w01*Tx slope: difference in change from baseline to post-treatment
  - γ_{21} for w12*Tx slope: differences in change from post-treatment to follow-up
- Model-implied fixed effects for treatment group ($Tx_i = 1$):
  - $\gamma_{00} + \gamma_{01}$: Tx weight at baseline
  - $\gamma_{10} + \gamma_{11}$: Tx change from baseline to post-treatment
  - $\gamma_{20} + \gamma_{21}$: Tx change from post-treatment to follow-up

Example 2: Piecewise Model for the Means

$$\widehat{weight}_{ti} = \gamma_{00} + \gamma_{10}(w01_{ti}) + \gamma_{20}(w12_{ti}) + \gamma_{01}(Tx_i) + \gamma_{11}(w01_{ti})(Tx_i) + \gamma_{21}(w12_{ti})(Tx_i)$$

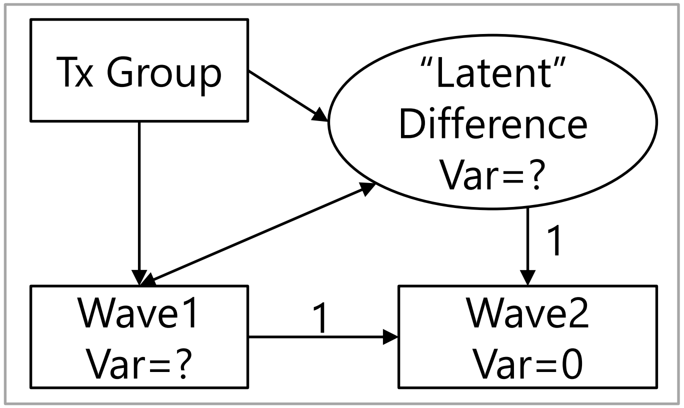


Example 2: Model for the Variance

- Random effects were not used to model dependency
  - Why? Given 2 slopes across 3 occasions, only 1 of the slopes could have a random effect → assumes parallel change for other time period!
  - The only way to get 2 random slopes from 3 occasions is to remove the WP residual variance, which assumes BP differences in change are measured perfectly!

    perfect measure

    - Btw, this is the basis of two-occasion "latent" difference score models (but latent variables are not magic)



- Instead, an unstructured variance–covariance matrix was estimated across occasions, separately for each group
  - Unstructured = different variances (and covariances) by occasion
  - Still implies differences in individual change (they just can't be quantified separately from individual intercept differences)
  - Separate matrices by group allows treatment differences in individual *heterogeneity* of mean levels and change

Example 2: Syntax by Univariate MLM Program (Stacked Data)

SAS:

```
PROC MIXED DATA=work.Example2 COVTEST METHOD=REML;
     CLASS ID wave GroupTx;
     MODEL weight = w01 w12 Tx w01*Tx w12*Tx / SOLUTION DDFM=KR;
     REPEATED wave / RCORR TYPE=UN SUBJECT=ID GROUP=GroupTx;
RUN;
```

R (gls from nlme package)—not sure how to get a different unstructured R matrix by group, described here, and denominator DF will not be correct without follow-up commands:

```
gls(data=Example2, method="REML",model=weight~1+w01+w12+Tx+w01:Tx+w12:Tx,
    correlation=corSymm(form=~as.numeric(wave)|ID),weights=varIdent(form=~1|wave))
```

STATA:

```
mixed weight w01 w12 Tx w01*Tx w12*Tx, || ID: , noconstant variance reml ///
      residuals(unstructured,t(wave) by(GroupTx)) dfmethod(kroger) dftable(pvalue)
```

SPSS—I don't think you can get fully different unstructured R matrices by group, but I could be wrong, and only Satterthwaite DF are available:

```
MIXED weight BY ID wave WITH w01 w12 Tx
      /METHOD = REML
      /PRINT = SOLUTION TESTCOV R
      /FIXED = w01 w12 Tx w01*Tx w12*Tx
      /REPEATED = wave | COVTYPE(UN) SUBJECT(ID).
```

Review of Longitudinal Multilevel Models

- Topics:
  - Concepts and terminology in longitudinal models
  - Modeling person dependency
  - Fixed and random intercepts
  - Fixed and random time slopes
  - Time-invariant predictors
  - **Details** (ML vs REML, significance tests, model comparisons)

Details: ML vs. REML Estimation

- What are REML and ML? Two flavors of likelihood estimation:
- **REML = "Restricted (or residual) maximum likelihood"**
  - Only available for general linear models or general linear mixed models (that assume normally distributed residuals); not in any SEM software
  - Is same as OLS given complete outcomes, but it doesn't require them
  - Estimates variances the same way as in OLS (accurate) → $\frac{\sum(y_{ti} - \hat{\boldsymbol{y}}_{ti})^2}{N - k}$
- **ML = "Maximum likelihood" (also called FIML*)**
  - Is more general, is available for the above plus for non-normal outcomes and latent variable models (CFA/SEM/IRT; multilevel SEM)
  - Is NOT the same as LS: it under-estimates variances by not accounting for the # of estimated fixed effects → $\frac{\sum(y_{ti} - \hat{\boldsymbol{y}}_{ti})^2}{N}$

**FI = Full information → it uses all the original data (they both do)*

Details: ML vs. REML Estimation

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

"Population"

$$\frac{\sum(y_i - \hat{y}_{ti})^2}{N}$$

"Sample"

$$\frac{\sum(y_i - \hat{y}_{ti})^2}{N - k}$$

All comparisons must have same *N*!!!	**ML**	**REML**
In software:	Only choice in SEM or M-SEM; available in MLM	Default in univariate general MLM programs
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** (less difference after *N*=30-50 or so)	**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)

Details: Assessing Significance

- **Model for the Means** → which **fixed effects** of predictors should be included in the model (e.g., main effects, interactions)
  - **Significance tests** do not require assessment of relative model fit using −2ΔLL (can always use univariate or multivariate Wald tests)
  - **Effect sizes** can come from the significance tests (e.g., *t* → Cohen's *d* or partial *r*), or from reductions in variance (pseudo-R^2 or total-R^2)

- **Model for the Variance** → what pattern(s) of variance and covariance the residuals from the same unit have; what **random effects** are needed to describe these pattern(s)
  - **Significance tests** DO require assessing relative model fit via −2ΔLL
    - Cannot use the Wald test *p*-values for variances because those *p*-values use a two-sided sampling distribution, but variances cannot be negative
  - **Effect sizes** (less commonly provided) can come from random effects confidence intervals (CI) or random effects reliability measures
    - Random Effect 95% CI = fixed effect $\pm \left(1.96 * \sqrt{\text{Random Variance}}\right)$

Pseudo-R^2 Effect Size of Fixed Effects

- Pseudo-R^2 = proportion of variance accounted for by fixed effects of predictors **in each pile of variance →** multiple pseudo-R^2 values
- For example, a fixed linear effect of WP time will reduce level-1 residual variance σ_e^2 in **R** by this much:

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

$$\text{More generally, Pseudo } R^2 = \frac{\text{was} - \text{is}}{\text{was}}$$

"**fewer**" = "**was**" = from model with fewer parameters
"**more**" = "**is**" = from model with more parameters

- But whenever <u>only</u> level-1 residual variance σ_e^2 is reduced, the level-2 random intercept variance $\tau_{U_0}^2$ will <u>INCREASE</u> as a result. Why?
  - Likelihood-based estimates of "true" $\tau_{U_0}^2$ use (σ_e^2 / level-1 n) as correction factor for the amount of between-person difference attributable to chance:
    True $\tau_{U_0}^2$ = Observed $\tau_{U_0}^2$ − (σ_e^2 / level-1 n)
  - For example: observed level-2 $\tau_{U_0}^2$=4.65, level-1 σ_e^2=7.06, n=4
    - True $\tau_{U_0}^2$= 4.65 −(**7.60**/4) = **2.88** in empty means model
    - Add fixed linear time slope → reduce σ_e^2 from 7.06 to 2.17 (Pseudo-R^2 = .69)
    - But now True $\tau_{U_0}^2$= 4.65 −(**2.17**/4) = **4.10** in fixed linear time model

Details: Significance of Fixed Effects in MLM

	Denominator DF is infinite (Proper Wald test)	**Denominator DF is estimated instead** ("Modified" Wald test)
Numerator DF = 1 *(test one fixed effect)* is **Univariate Wald Test**	use z distribution (all of SEM; Mplus MLM, STATA MIXED default)	use t distribution (R nlme or lme4; SAS; SPSS; STATA MIXED with *dfmethod* and *small*)
Numerator DF > 1 *(test 2+ fixed effects)* is **Multivariate Wald Test**	use χ^2 distribution (Mplus, STATA default)	use F distribution (R glht; SAS, SPSS; STATA MIXED with *dfmethod* and *small*)
Options for estimating Denominator DF (DDF)	not applicable	R, SAS, STATA: Kenward-Roger R, SAS, STATA, SPSS: **Satterthwaite**

Details: Comparing Models for the Variance

- **Two strategies for choosing a model for the variance:**
  - Does the more complex model fit better (than a simpler model)?
  - Does the simpler model fit worse (than a more complex model)?

- Nested models are compared using a **"likelihood ratio test": −2ΔLL test** (aka, "χ^2 test" in SEM; "deviance difference test" in MLM)

"fewer" = from model with fewer parameters "more" = from model with more parameters	Results of 1. & 2. must be positive values!

1. Calculate **−2ΔLL:** if given −2LL, do $-2\Delta LL = (-2LL_{fewer}) - (-2LL_{more})$
   if given LL, do $-2\Delta LL = -2 * (LL_{fewer} - LL_{more})$
2. Calculate **Δdf =** $(\# \text{ Parms}_{more}) - (\# \text{ Parms}_{fewer})$
3. **Compare −2ΔLL to χ^2 distribution with df = Δdf**
4. Get *p*-value (e.g., from Excel CHIDIST, R anova, or STATA LRTEST)

Details: Comparing Models for the Variance

- <u>What your *p*-value for the −2ΔLL test means:</u>
  - If you **ADD** parameters, then your model can get **better** (if −2ΔLL test is significant) or **not better** (not significant)
  - If you **REMOVE** parameters, then your model can get **worse** (if −2ΔLL test is significant) or **not worse** (not significant)

- Nested or non-nested models can also be compared by **Information Criteria** that also reflect model parsimony
  - No significance tests or critical values, just "smaller is better"
  - **AIC** = Akaike IC = −2LL + 2 *(#parameters)
  - **BIC** = Bayesian IC = −2LL + log(N)*(#parameters)
  - What "parameters" means depends on flavor (except in STATA):
    - ML = ALL parameters; REML = variance model parameters only