Time-Invariant Predictors in Longitudinal Models

- Topics:
  - What happens to missing predictors
  - Effects of time-invariant predictors (in MLM and SEM)
  - Fixed vs. systematically varying vs. random effects
  - Model building strategies and assessing significance

What happens to missing predictors?

- **Incomplete data patterns in longitudinal studies**
  - Sparse missingness (within occasion)
  - Differential attrition (monotonic dropout)
  - Measurements obtained at different intervals ("unbalanced data")
  - "Planned" missing data (yes, you can do this on purpose, but carefully)
  - Often unrecognized selection bias at beginning of all studies, too

- **The goal is to make valid inferences** about population parameters despite bias introduced by attrition
  - The goal is _not_ to recover the missing data values

- Methods used to do analyses in the presence of missing data require assumptions about the causes associated with the missingness process as well as the variables' distributions

Methods of Analysis Given Missing Data

- **What NOT to do:**
  - ➢ **NEVER EVER:** Single mean replacement or regression imputation
  - ➢ **PREFERABLY NOT:** Listwise deletion (all available whole people)

- **What to do: FIML or multiple imputation (or Bayes)**
  - ➢ FIML = Full-information maximum likelihood ➔ uses all the original data in estimating model, not just a summary thereof
  - ➢ MIXED and Mplus use FIML by default for missing responses (REML and ML as we know them are both Full-Information)
  - ➢ Asymptotically equivalent results given the same missingness model, but FIML is more direct than multiple imputation (and is more readily available for non-normal variables)
  - ➢ Both of these methods assume **Missing at Random**, though…

Categorizations of Missing Data

- If data are missing from some occasions, all is not lost!

- Missingness predictors: Person-level variables, outcomes at other observed occasions:

  - <u>Missing Completely at Random (MCAR):</u> probability of missingness is unrelated to what those missing responses would have been

  - **<u>Missing at Random (MAR):</u>** probability of missingness depends on the persons' predictors or their other observed outcomes, but you can draw correct inferences after including (controlling for) their other data

  - <u>Missing Not At Random (MNAR):</u> probability of missingness is systematic but is not predictable based on the information you have (everything will be some shade of wrong)

- You will likely get different estimates from models with complete cases only… so use all the data you have if possible!

- Now, the bad news…

Missing Data in MLM Software

- Common misconceptions about how MLM "handles" missing data

- <u>Most MLM programs (e.g., MIXED) analyze only COMPLETE CASES</u>

  ➤ Does NOT require listwise deletion of *whole persons*

  ➤ DOES delete any incomplete cases (occasions within a person)

- <u>Observations missing predictors OR outcomes are not included</u>!

  ➤ **Time** is (probably) measured for **everyone**

  ➤ **Predictors may NOT be measured for everyone**

  ➤ *N* may change due to missing data for different predictors across models

- You may need to think about what predictors you want to examine PRIOR to model building, and <u>pre-select your sample accordingly</u>

  ➤ Models and model fit statistics −2LL, AIC, and BIC are only **directly comparable** if they include the **exact same observations (LL is sum of each height)**

  ➤ Will have less statistical power as a result of removing incomplete cases

Be Careful of Missing Predictors!

Multivariate (wide) data → stacked (long) data

ID	T1	T2	T3	T4	Person Pred	T1 Pred	T2 Pred	T3 Pred	T4 Pred
100	5	6	8	12	50	4	6	7	.
101	4	7	.	11	.	7	.	4	9

Row	ID	Time	DV	Person Pred	Time Pred
1	**100**	1	5	50	4
2	**100**	2	6	50	6
3	**100**	3	8	50	7
4	**100**	4	12	50	.
5	**101**	1	4	.	7
6	**101**	2	7	.	.
7	**101**	3	.	.	4
8	**101**	4	11	.	9

Only rows with complete data get used – for each model, which rows get used in univariate MLM?

Model with Time → DV: 1-6, 8

Model with Time, Time Pred → DV: 1-3, 5, 8

Model with Time, Person Pred → DV: 1-4

Model with Time, Time Pred, & Person Pred → DV: 1-3

Beware of Missing Predictors

- Any cases missing model predictors that are not part of the joint likelihood* will not be used in that model

  - ➢ Bad 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)

Modeling Time-Invariant Predictors

<u>What independent variables can be time-invariant predictors?</u>

- Also known as "person-level" or "level-2" predictors

- Include substantive predictors, controls, and predictors of missingness

- Can be anything that **does not change across time** (e.g., Biological Sex)

- Can be anything that **is not likely to change across the study,** but you may have to argue for this (e.g., Parenting Strategies, SES)

- Can be anything that **does change across the study**...

  - But you have **only measured once**
    - Limit conclusions to variable's status at time of measurement
    - e.g., "Parenting Strategies at age 10"
  - Or **is perfectly correlated with time** (age, time to event)
    - Would use Age at Baseline, or Time to Event *from Baseline* instead

Centering Time-Invariant Predictors

- Very useful to center all predictors such that 0 is a meaningful value:

  - Same significance level of main effect, different interpretation of intercept

  - Different (more interpretable) main effects within higher-order interactions

    - With interactions, main effects = simple effects when other predictor = 0

- Choices for centering **continuous** predictors:

  - At Mean: Reference point is *average level of predictor within the sample*

    - Useful if predictor is on arbitrary metric (e.g., unfamiliar test)

  - Better → At Meaningful Point: Reference point is *chosen level of predictor*

    - Useful if predictor is already on a meaningful metric (e.g., age, education)

- Choices for centering **categorical** predictors:

  - Re-code into a series of dummy-coded predictors (or let program do it)

  - I do not recommend effects-coding for categorical predictors (because who is at the mean of a categorical variable ?!?)

Coding Strategies for Categorical Predictors

Indicator coding: Each non-ref category has a 1 value in **1 predictor only** to represent its mean difference from reference (good for **nominal**)

Sequential coding: Each non-ref category can have multiple 1 values → predictors then give mean differences between sequential categories (good for **ordinal**)
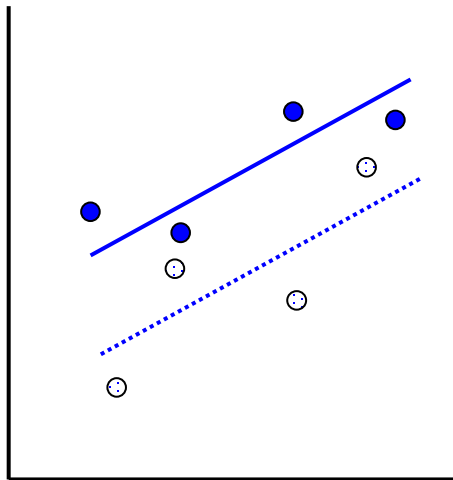
Group	(Inter-cept): A mean	AvsB: Diff for A vs B	AvsC: Diff for A vs C
A	1	0	0
B	1	1	0
C	1	0	1

Either way, all possible category means and mean differences not directly provided by the model fixed effects can be found from linear combinations of them…

Happy	(Intercept): 1 Mean	h1v2: 1→2 Diff	h2v3: 2→3 Diff	h3v4: 3→4 Diff	h4v5: 4→5 Diff
1	1	0	0	0	0
2	1	1	0	0	0
3	1	1	1	0	0
4	1	1	1	1	0
5	1	1	1	1	1

Sequential coding can be used to test whether an ordinal predictor can be treated as interval—whether it has a linear slope in predicting an outcome—by testing differences between the sequential slopes

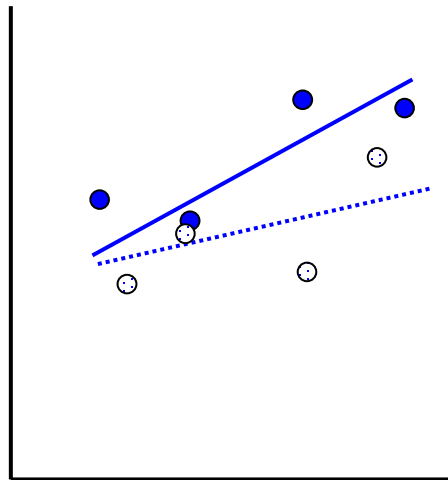# The Role of Time-Invariant Predictors in the **Model for the Means**

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

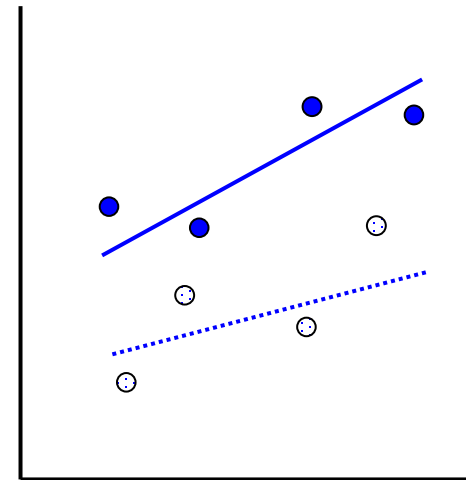Main effect of x_i, no interaction with time

Interaction with time, Main effect of x_i?

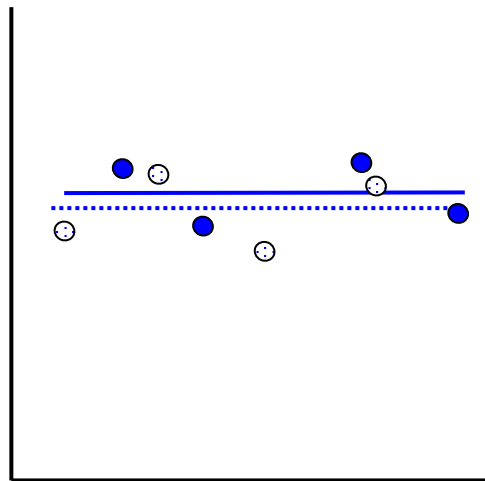Main effect of x_i, and Interaction with time

← Time →

← Time →

← Time →

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

- In **Within-Person Fluctuation Models** → Adjust mean level

No main effect of x_i

Main effect of x_i

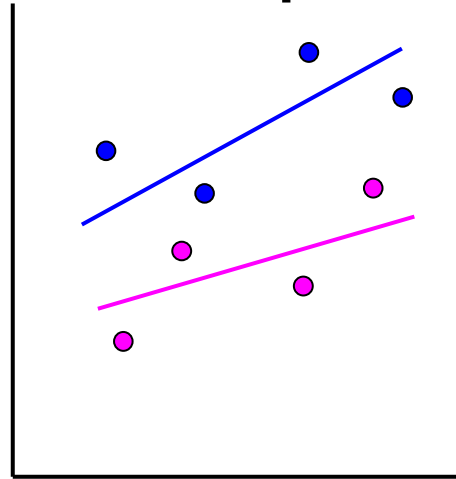$\leftarrow$ Time $\rightarrow$
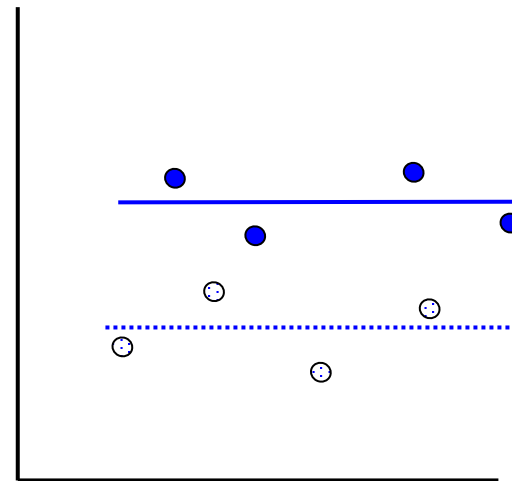
$\leftarrow$ Time $\rightarrow$

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., Sex as a predictor of heterogeneity of variance:

  - **At level 2**: *Amount* of individual differences in intercepts/slopes differs between boys and girls (i.e., one group is more variable)

  - **At level 1**: *Amount* of within-person residual variation differs between boys and girls

    - In within-person **fluctuation** model: differential fluctuation over time
    - In within-person **change** model: differential fluctuation/variation 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")

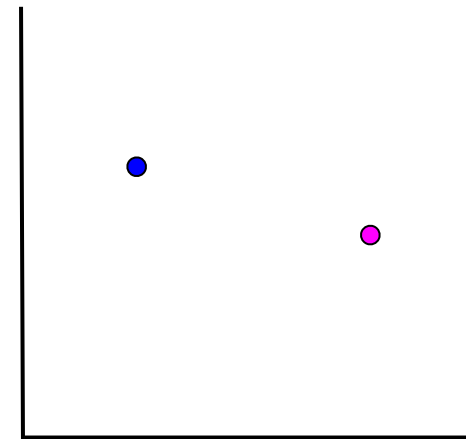 - Also described with examples in Hoffman & Walters (2022)

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

Random Slopes for Time

Random Slopes for Group?

Time
(or Any Level-1 Predictor)

Group
(or any Level-2 Predictor)

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)

Random Quadratic Time Unconditional Model

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

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

Fixed Effects of Time-Invariant Education: Example using a Random Quadratic Time Model

- <u>Main Effect of Education = Education*Intercept Interaction</u>

  - Moderates the intercept → Difference in expected outcome at time 0 per year of education (per unit x_i)

- <u>Effect of Education on Linear Time = Education*Time Interaction</u>

  - Moderates the linear time slope → Difference in expected linear rate of change at time 0 per year of education (per unit x_i)

- <u>Effect of Education on Quadratic Time = Education*Time^2 Interaction</u>

  - Moderates the quadratic time slope → Difference in half the rate of expected acceleration/deceleration of linear rate of change per year of education (per unit x_i)

Education (12 years = 0) as a Time-Invariant Predictor: Example using a Random Quadratic Time Model

Level 1: $y_{ti} = \beta_{0i} + \beta_{1i}\text{Time}_{ti} + \beta_{2i}\text{Time}_{ti}^2 + e_{ti}$

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + \gamma_{01}\text{Ed}_i + U_{0i}$

Intercept for person i

Fixed Intercept when Time=0 and Ed=12

Δ in Intercept per unit Δ in Ed

Random (Deviation) Intercept after controlling for Ed

$\beta_{1i} = \gamma_{10} + \gamma_{11}\text{Ed}_i + U_{1i}$

Linear Slope for person i

Fixed Linear Time Slope when Time=0 and Ed=12

Δ in Linear Time Slope per unit Δ in Ed (=Ed*time)

Random (Deviation) Linear Time Slope after controlling for Ed

$\beta_{2i} = \gamma_{20} + \gamma_{21}\text{Ed}_i + U_{2i}$

Quad Slope for person i

Fixed Quad Time Slope when Ed = 12

Δ in Quad Time Slope per unit Δ in Ed (=Ed*time^2)

Random (Deviation) Quad Time Slope after controlling for Ed

Education (12 years = 0) as a Time-Invariant Predictor: Example using a Random Quadratic Time Model

<u>Level 1</u>: $y_{ti} = \beta_{0i} + \beta_{1i}\text{Time}_{ti} + \beta_{2i}\text{Time}_{ti}^2 + e_{ti}$

<u>Level 2 Equations (one per β)</u>:

$$\beta_{0i} = \gamma_{00} + \gamma_{01}\text{Ed}_i + U_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11}\text{Ed}_i + U_{1i}$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21}\text{Ed}_i + U_{2i}$$

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

> Each fixed slope of education 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!
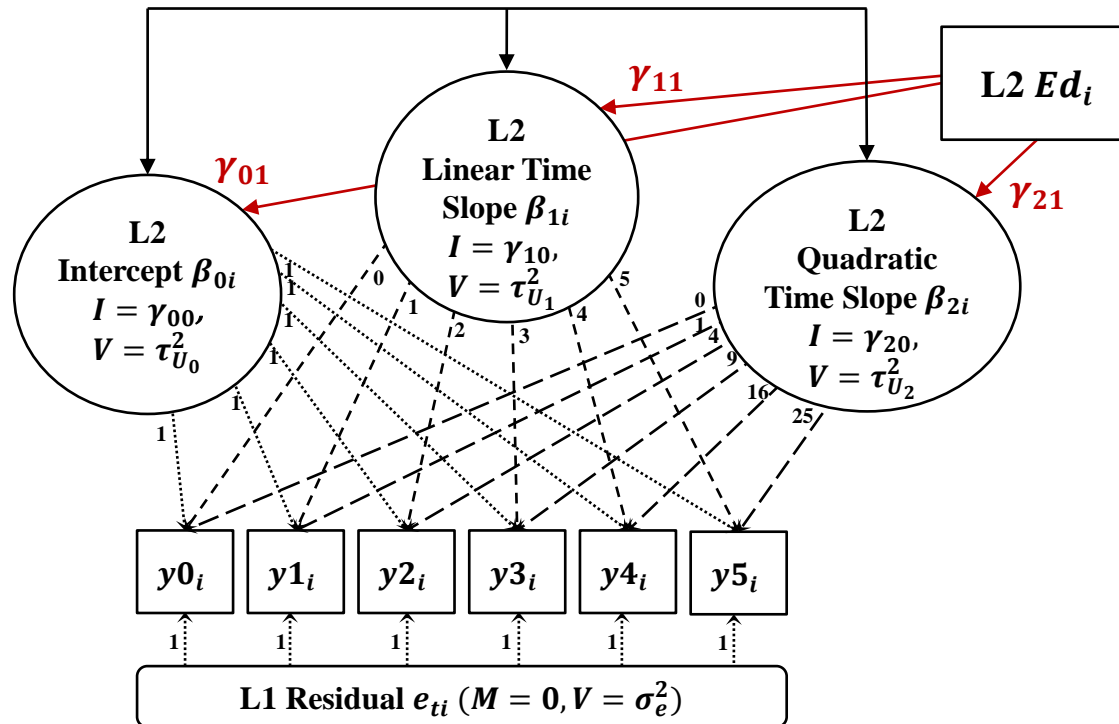
- Composite equation:

- $y_{ti} = (\gamma_{00} + \gamma_{01}\text{Ed}_i + U_{0i}) +$
  $(\gamma_{10} + \gamma_{11}\text{Ed}_i + U_{1i})\text{Time}_{ti} +$
  $(\gamma_{20} + \gamma_{21}\text{Ed}_i + U_{2i})\text{Time}_{ti}^2 + e_{ti}$

Should I have used a "latent" growth curve model (on wide data in SEM) instead?

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

Cons:
- No REML, no DDF → Type I error for small N
- Requires balanced time *(or definition variables for individual time loadings)*

Pros:
- Latent basis nonlinear change *(fix 1st loading to 0, last to 1, estimate other loadings for % change)*
- More flexibility in WP residual heterogeneity of variance and covariance
- Change in latent variables instead of observed

Example: Mplus Single-Level SEM Syntax

Just showing **MODEL** part, which would be preceded by **DATA**, **VARIABLE**, and **ANALYSIS** as usual (estimated using **wide** data)

```
!!!!! Random quadratic model of change

! Factor loadings fixed by @
  Int BY y0@1 y1@1 y2@1 y3@1 y4@1  y5@1;
  Lin BY y0@0 y1@1 y2@2 y3@3 y4@4  y5@5;
  Qua BY y0@0 y1@1 y2@4 y3@9 y4@16 y5@25;

! Factor intercepts estimated = fixed effects
  [Int Lin Qua];
! Level-2 factor variances estimated (in G)
  Int Lin Qua;
! Level-2 factor covariances estimated (in G)
  Int Lin Qua WITH Int Lin Qua;

! Per-occasion intercepts fixed to 0
  [y0@0 y1@0 y2@0 y3@0 y4@0 y5@0];

! Level-1 residual variances held equal (in R)
  y0 y1 y2 y3 y4 y5 (ResVar);

! Fixed effects of education → latent factors
  Int Lin Qua ON Ed;
```

```
!!!!! Random latent basis model of change

! Factor loadings fixed by @
  Int BY y0@1 y1@1 y2@1 y3@1 y4@1 y5@1;
  Slp BY y0@0 y1*  y2*  y3*  y4*  y5@1;
! Loadings estimated as 0.57, 0.76, 0.90, 0.97

! Factor intercepts estimated = fixed effects
  [Int Slp];
! Level-2 factor variances estimated (in G)
  Int Slp;
! Level-2 factor covariance estimated (in G)
  Int WITH Slp;

! Per-occasion intercepts fixed to 0
  [y0@0 y1@0 y2@0 y3@0 y4@0 y5@0];

! Level-1 residual variances held equal (in R)
  y0 y1 y2 y3 y4 y5 (ResVar);

! Fixed effects of education → latent factors
  Int Slp ON Ed;
```

Note: There are Mplus syntax shortcuts for growth models I am not using: (1) to be explicit about what the model contains, (2) to not estimate separate residual variances

Example: R Single-Level SEM Syntax

```
RandQuadSyntax = "
# Factor loadings fixed by *
Int =~ 1*y0 + 1*y1 + 1*y2 + 1*y3 +  1*y4 +  1*y5
Lin =~ 0*y0 + 1*y1 + 2*y2 + 3*y3 +  4*y4 +  5*y5
Qua =~ 0*y0 + 1*y1 + 4*y2 + 9*y3 + 16*y4 + 25*y5

# Factor intercepts estimated = fixed effects
  Int ~ 1;  Lin ~ 1;   Qua ~ 1
# Level-2 factor variances estimated (in G)
  Int ~~ Int;  Lin ~~ Lin;   Qua ~~ Qua
# Level-2 factor covariances estimated (in G)
  Int ~~ Lin + Qua;  Lin ~~ Qua

# Per-occasion intercepts fixed to 0
  y0 ~ 0; y1 ~ 0; y2 ~ 0
  y3 ~ 0; y4 ~ 0; y5 ~ 0

! Level-1 residual variances held equal (in R)
  y0 ~~ (ResVar)*y0;  y1 ~~ (ResVar)*y1
  y2 ~~ (ResVar)*y2;  y3 ~~ (ResVar)*y3
  y4 ~~ (ResVar)*y4;  y5 ~~ (ResVar)*y5

# Fixed effects of education --> latent factors
  Int + Lin + Qua ~ Ed
"
RQModel = lavaan(data=Example2wide,
                 model=RandQuadSyntax,
                 estimator="ML", mimic="mplus")
summary(RQModel, fit.measures=TRUE, rsquare=TRUE,
        standardized=TRUE)
```

```
LatentBasisSyntax = "
# Factor loadings fixed by *
Int =~ 1*y0 + 1*y1 + 1*y2 + 1*y3 +  1*y4 +  1*y5
Slp =~ 0*y0 +   y1 +   y2 +   y3 +     y4 + 1*y5
# Loadings estimated as 0.57, 0.76, 0.90, 0.97

# Factor intercepts estimated = fixed effects
  Int ~ 1;  Slp ~ 1
# Level-2 factor variances estimated (in G)
  Int ~~ Int;  Slp ~~ Slp
# Level-2 factor covariances estimated (in G)
  Int ~~ Slp

# Per-occasion intercepts fixed to 0
  y0 ~ 0; y1 ~ 0; y2 ~ 0
  y3 ~ 0; y4 ~ 0; y5 ~ 0

! Level-1 residual variances held equal (in R)
  y0 ~~ (ResVar)*y0;  y1 ~~ (ResVar)*y1
  y2 ~~ (ResVar)*y2;  y3 ~~ (ResVar)*y3
  y4 ~~ (ResVar)*y4;  y5 ~~ (ResVar)*y5

# Fixed effects of education --> latent factors
  Int + Slp ~ Ed
"
LBModel = lavaan(data=Example2wide,
                 model=LatentBasisSyntax,
                 estimator="ML", mimic="mplus")
summary(LBModel, fit.measures=TRUE, rsquare=TRUE,
        standardized=TRUE)
```

Note: There are lavaan syntax shortcuts for growth models I am not using: (1) to be explicit about what the model contains, (2) to not estimate separate residual variances

Should I have used "multilevel SEM" (on long data) instead? Not in this case…

$$y_{ti} = (\gamma_{00} + \gamma_{01}Ed_i + U_{0i}) +$$
$$(\gamma_{10} + \gamma_{11}Ed_i + U_{1i})Time_{ti} +$$
$$(\gamma_{20} + \gamma_{21}Ed_i + U_{2i})Time_{ti}^2 + e_{ti}$$
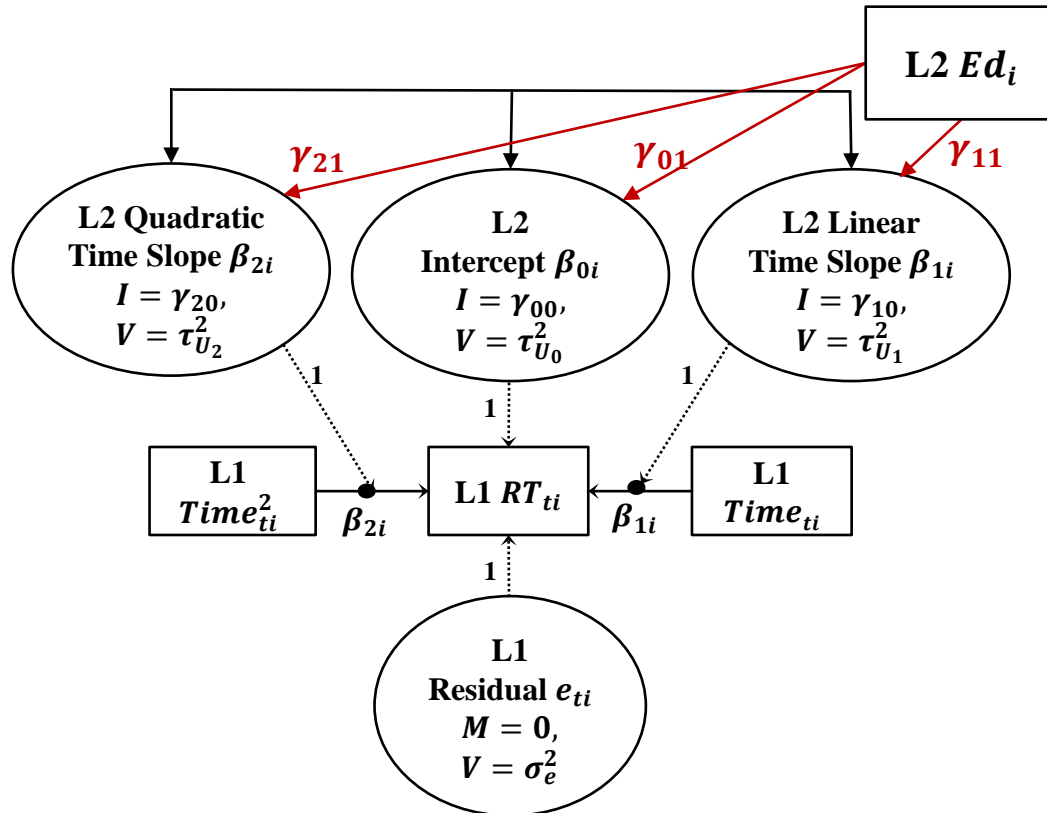
Cons:
- No REML, no DDF → Type I error for small N
- ~~Requires balanced time *(or definition variables for individual time loadings)*~~

Pros:
- ~~Latent basis nonlinear change *(fix 1st loading to 0, last to 1, estimate other loadings for % change)*~~
- ~~More flexibility in WP residual heterogeneity of variance and covariance~~
- Change in latent variables instead of observed

Example: Mplus M-SEM Syntax

Just showing **MODEL** part, which would be preceded by **DATA**, **VARIABLE**, and **ANALYSIS** as usual (estimated using **long** data)

```
%WITHIN%
  y;                          ! Level-1 residual variance
  Lin | y ON time;           ! Create beta1i placeholder
  Qua | y ON timesq;         ! Create beta2i placeholder

%BETWEEN%
  [y Lin Qua];                          ! Intercepts
   y Lin Qua;                           ! Level-2 random effect variances
   y Lin Qua WITH y Lin Qua;            ! Level-2 random effect covariances
   y Lin Qua ON Ed;                     ! Fixed effects of reasoning
```

• Note: R's lavaan package does have M-SEM capability, but it is much more limited than M-SEM in Mplus:

  ➢ Listwise deletion for any rows (occasions) with missing values

  ➢ No random slopes!

Fixed Effects of Time-Invariant Predictors

- Question of interest: <u>Why do people change differently?</u>

  - We're trying to predict individual differences in intercepts and slopes (i.e., reduce or explain the variances of the level-2 random effects)

  - So level-2 random effects variances are then conditional on predictors → actually random effects variances *left over* (aka "level-2 residuals")

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

  - Can calculate pseudo-R^2 for each level-2 random effect variance between models with *fewer* versus *more* parameters as:

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

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…	**Is this still ok to do?**
$\beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + U_{0i}$	$\beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + U_{0i}$
$\beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i + U_{1i}$	$\beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i$
$\beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i + U_{2i}$	$\beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i$

  - **"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).

Are Systematically Varying Effects ok?

- **YES**, so long as you haven't accidentally omitted a "sizeable" random slope variance (i.e., made a Type II error)

- How to know? Consider significance of slope variance AND **Slope Reliability** (see Hoffman & Templin, under revision)

$\tau^2_{U_1}$ = random slope variance
σ^2_e = residual variance
$L1n$ = L1 sample size per L2 unit
σ^2_{L1} = variance of L1 predictor

$$SR = \frac{\tau^2_{U_1}}{\tau^2_{U_1} + \dfrac{\sigma^2_e}{L1n * \sigma^2_{L1}}}$$

- Simulation examining L2n = 10 to 50 and L1n = 3 to 10 suggests keeping nonsignificant random slope variances with SR > .20 when using REML or SR > .15 when using ML maintains acceptable Type I errors for cross-level interactions

Adding **Group** as a Second Time-Invariant Predictor: Example using a Random Quadratic Time Model

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

Level 2 Equations (one per β):

$\beta_{0i} = \gamma_{00} + \gamma_{01}Ed_i + \gamma_{02}Group_i + U_{0i}$

$\beta_{1i} = \gamma_{10} + \gamma_{11}Ed_i + \gamma_{12}Group_i + U_{1i}$

$\beta_{2i} = \gamma_{20} + \gamma_{21}Ed_i + \gamma_{22}Group_i + U_{2i}$

- Composite equation:

- $y_{ti} = (\gamma_{00} + \gamma_{01}Ed_i + \gamma_{02}Group_i + U_{0i}) + (\gamma_{10} + \gamma_{11}Ed_i + \gamma_{12}Group_i + U_{1i})Time_{ti} + (\gamma_{20} + \gamma_{21}Ed_i + \gamma_{22}Group_i + U_{2i})Time_{ti}^2 + e_{ti}$

Demonstrating Pseudo-R^2

Start with Empty Means,
Random Intercept Model

Level-2 BP
Random
Intercept
$\tau^2_{U_0}$

Level-1 WP
Residual σ^2_e

L1 Time

First, we introduce a
**fixed effect of the level-1
time predictor** to explain
level-1 WP residual variance:

Is now a **Fixed Linear Time,
Random Intercept Model**

Demonstrating Pseudo-R^2

Level-2 BP Random Intercept $\tau^2_{U_0}$

Level-2 BP Random Linear Time Slope $\tau^2_{U_1}$

Level-1 WP Residual σ^2_e

L1 Time

Next, we introduce a **level-2 random slope for the level-1 time predictor**, which re-allocates some of the previous level-1 residual variance to the new level-2 variance:

Is now a **Random Linear Time Model**

Demonstrating Pseudo-R^2

Level-2 BP Random Intercept $\boldsymbol{\tau^2_{U_0}}$

Level-2 BP Random Linear Time Slope $\boldsymbol{\tau^2_{U_1}}$
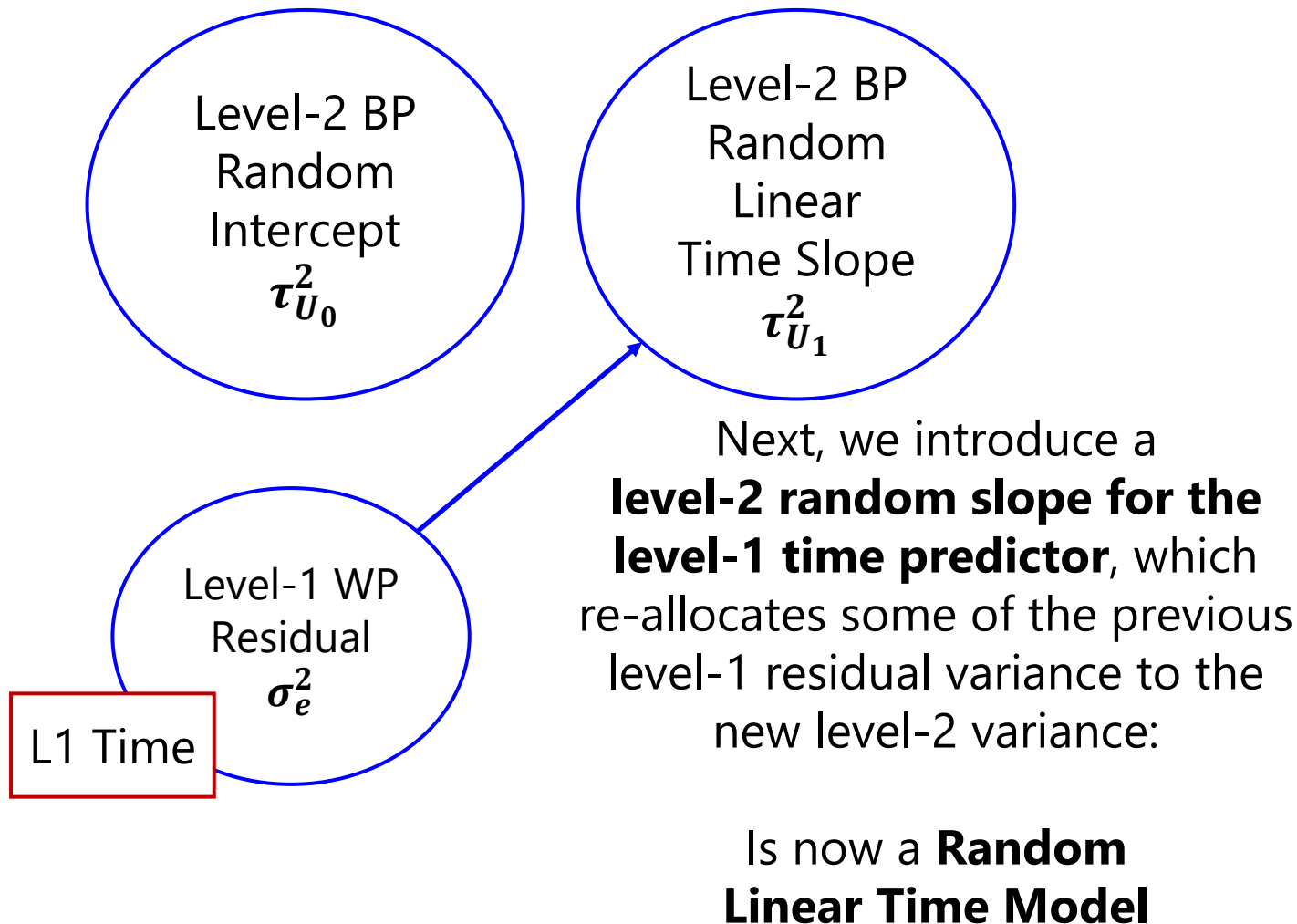
L1 Time^2

L1 Time

Level-1 WP Residual $\boldsymbol{\sigma^2_e}$

Next, we introduce a **fixed effect of the level-1 time^2 predictor** to explain level-1 WP residual variance:

Is now a **Fixed Quadratic, Random Linear Time Model**

Demonstrating Pseudo-R²

Level-2 BP
Random
Intercept
$\tau^2_{U_0}$

Level-2 BP
Random
Linear
Time Slope
$\tau^2_{U_1}$

Level-2 BP
Random
Quadratic
Time² Slope
$\tau^2_{U_2}$

Level-1
WP
Residual
σ^2_e

L1Time²

L1 Time

Next, we introduce a **level-2 random slope for the level-1 time² predictor**, which re-allocates some of the previous level-1 residual variance to the new level-2 variance:

Is now a **Random Quadratic Time Model**
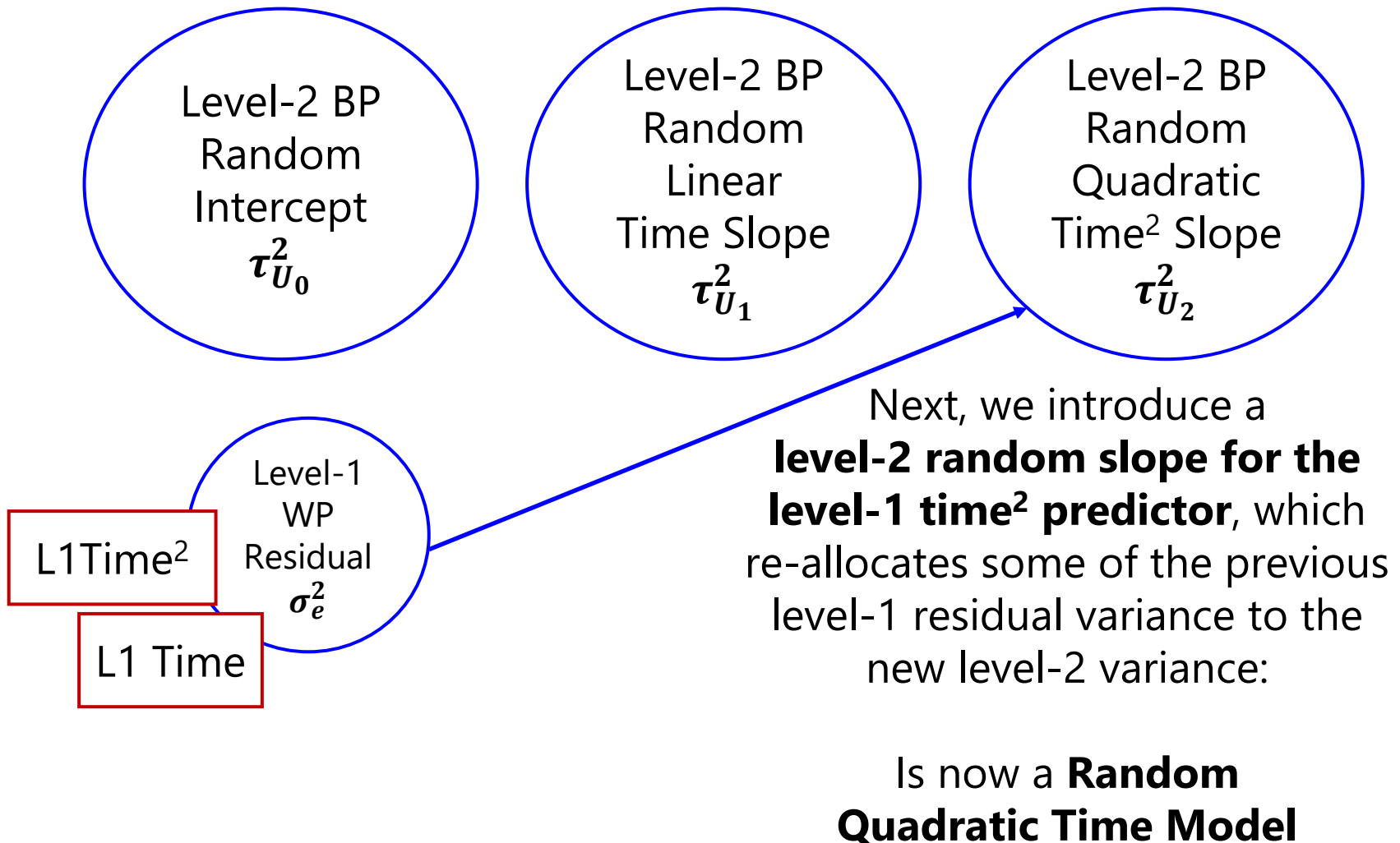
So which pile of variance does each fixed effect (in red boxes) explain?

Level-2 BP
Random
Intercept
$\tau^2_{U_0}$

Level-2 BP
Random
Linear
Time Slope
$\tau^2_{U_1}$

Level-2 BP
Random
Quadratic
Time^2 Slope
$\tau^2_{U_2}$

Level-1
WP
Residual
σ^2_e

L1Time^2

L1 Time

L2 Ed (* Intercept)	L2 Group (* Intercept)
L2 Ed * L1 Time	L2 Group * L1 Time
L2 Ed * L1 Time^2	L2 Group * L1 Time^2
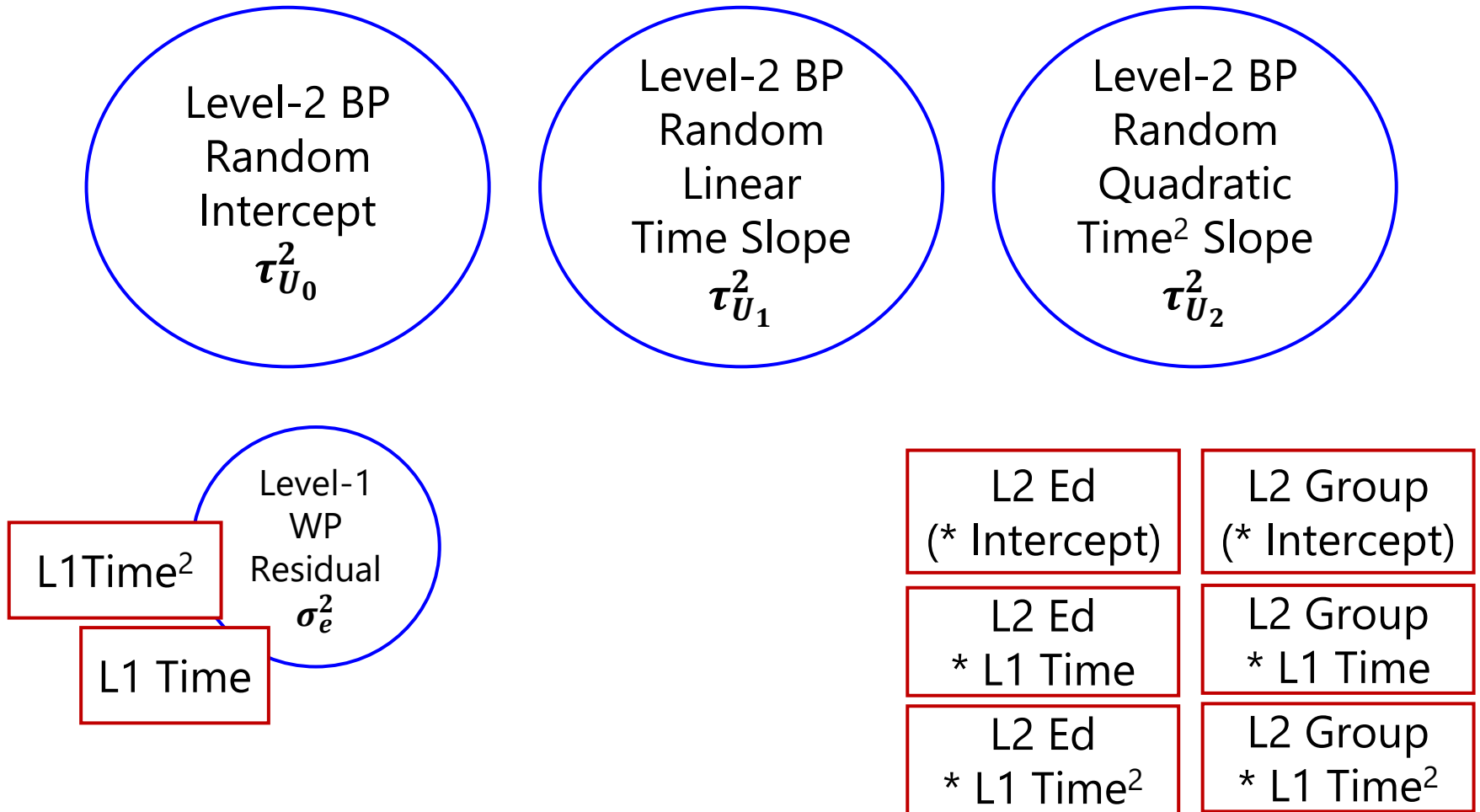
What about Group*Ed interactions?

Level-2 BP
Random
Intercept
$\tau^2_{U_0}$

Level-2 BP
Random
Linear
Time Slope
$\tau^2_{U_1}$

Level-2 BP
Random
Quadratic
Time^2 Slope
$\tau^2_{U_2}$

Level-1
WP
Residual
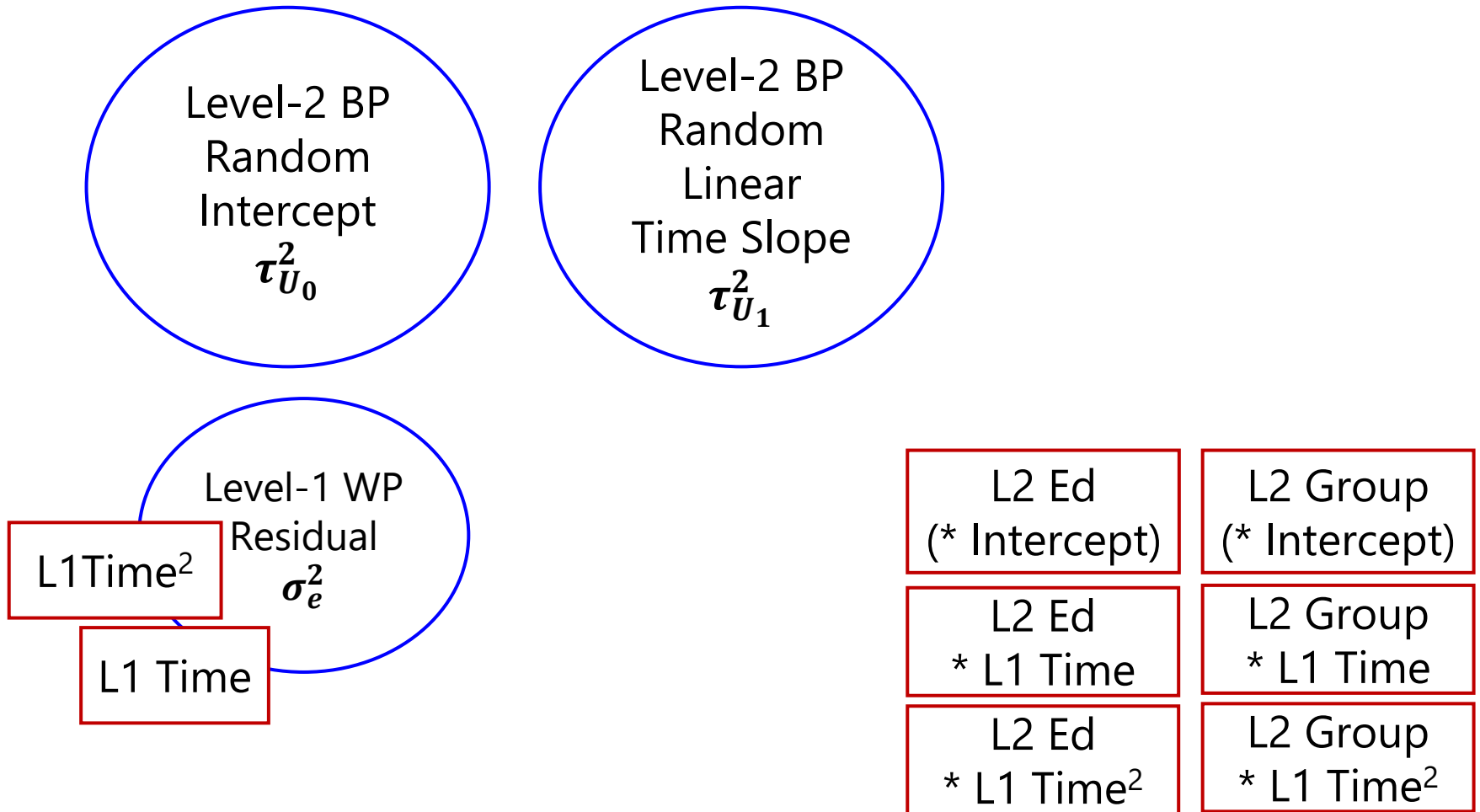σ^2_e

L1Time^2

L1 Time

L2 Group * L2 Ed
(* Intercept)

L2 Group * L2 Ed
* L1 Time

L2 Group * L2 Ed
* L1 Time^2

What if the random quadratic time slope was not initially needed?

Level-2 BP Random Intercept $\tau^2_{U_0}$

Level-2 BP Random Linear Time Slope $\tau^2_{U_1}$

Level-1 WP Residual σ^2_e

L1Time^2

L1 Time

L2 Ed (* Intercept)

L2 Group (* Intercept)

L2 Ed * L1 Time

L2 Group * L1 Time

L2 Ed * L1 Time^2

L2 Group * L1 Time^2

Variance Accounted For By 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 Accounted for… Continued

- **Pseudo-R²** 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 marginal versions of R^2
    - ▪ Ensures positive R^2 values, but they don't quantify R^2 for slope variances (boo)

- **A simple alternative: Total R²** (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 that 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!!

Model-Building Strategies

- It may be helpful to examine predictor effects in **separate** models at first, including interactions with all growth terms to see the **total** bivariate relations for a single predictor

  - Question: Does age matter at all in predicting change over time?

  - e.g., in random quadratic time model + age, age*time, age*time^2

- Then predictor effects can be **combined** in layers in order to examine **unique** contributions (and interactions) of each

  - Question: Does age *still* matter after considering reasoning?

  - random quadratic + age, age*time, age*time^2,
                                + reason, reason*time, reason*time^2

  - Potentially also   + age*reason, age*reason*time, age*reason*time^2

- Sequence of predictors should be guided by theory and research questions—there may not be a single "best model"

  - One person's "control" is another person's "question", so you may not end up in the same place given different orders of predictor inclusion

Evaluating Statistical Significance of Multiple New Fixed Effects at Once

- Can always do **multivariate Wald test** in REML or ML (using SAS CONTRAST, STATA TEST, or contestMD in R)

- Can only compare nested models via −2ΔLL test **in ML**

- Either is useful for 'borderline' cases—for example:

  - Ed*time^2 interaction at p = .04, with nonsignificant ed*time and ed*Intercept (main effect of ed) terms?

  - Is it worth keeping a marginal higher-order interaction that requires two (possibly non-significant) lower-order terms?

- **−2ΔLL, AIC, or BIC comparisons for models with different fixed effects are wrong when using REML estimation**

  - Because of this, many books (including my first edition) switch to ML when focusing on modeling fixed effects of predictors

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 (Mplus, STATA default)	use t distribution (SAS, SPSS, STATA with dfmethod option)
Numerator DF > 1 *(test 2+ fixed effects)* is **Multivariate Wald Test**	use χ^2 distribution (Mplus, STATA default)	use F distribution (SAS, SPSS, STATA with dfmethod option)
Options for estimating Denominator DF (DDF)	not applicable	SAS, STATA 14+: Kenward-Roger SAS, STATA 14+, SPSS: **Satterthwaite**

* Btw, I am still figuring this out in R… default and optional DDF behavior seem to vary by package (like everything else)

Denominator DF (DDF) Methods

- **Between-Within** (DDFM=BW in SAS, REPEATED in STATA):
  - Total DDF (T) comes from total number of observations, separated into level-2 for N persons and level-1 for n occasions (like in RM ANOVA)
    - **Level-2 DDF** = N – #level-2 fixed effects
    - **Level-1 DDF** = Total DDF – Level-2 DDF – #level-1 fixed effects
    - Level-1 effects with random slopes still get level-1 DDF

- **Satterthwaite** (DDFM=Satterthwaite in SAS and STATA, available in LME and LMER in R, default in SPSS):
  - More complicated, but analogous to two-group t-test given unequal residual variances and unequal group sizes
  - Incorporates contribution of variance components at each level
    - Level-2 DDF will resemble Level-2 DDF from BW
    - Level-1 DDF will resemble Level-1 DDF from BW if the level-1 effect is not random, but will resemble level-2 DDF if it is random

Denominator DF (DDF) Methods

- **Kenward-Roger** (DDFM=KR in SAS, KROGER in STATA, available in LME and LMER in R, not available in SPSS):

  - Adjusts the asymptotic covariance matrix of the fixed effects to reflect the uncertainty introduced by using large-sample techniques of REML in small N samples

  - This creates different (larger) SEs for the fixed effects

  - Then uses Satterthwaite DDF, new SEs, and t to get p-values

- In an unstructured variance model, all effects use level-2 DDF

- Differences in inference not likely to matter often in practice

  - e.g., critical t-value at DDF=20 is 2.086, at infinite DDF is 1.960

- When in doubt, use KR (is overkill at worst, becomes Satterthwaite)

  - I used Satterthwaite in the book to maintain comparability across programs

Wrapping Up

- MLM uses ONLY the rows of data that are COMPLETE—all predictors AND outcomes must be there!

  - Using whatever data you do have for each person will likely lead to better inferences and more statistical power than using only complete persons (so avoid listwise deletion if you can)

  - Missing predictors can be addressed through multiple imputation or by bringing them into the likelihood in multivariate software

- Time-invariant predictors modify the level-1 created growth curve → predict individual intercepts and slopes

  - They account for random effect variances (the predictors are the reasons WHY people need their own intercepts and slopes)

  - If a level-1 predictor does not have a level-2 random slope, it can still be moderated by a cross-level interaction with a time-invariant predictor…

    - … but then it will predict level-1 residual variance instead
    - Make sure you test the level-2 random slope first!!