Random Slopes and Cross-Level Interactions in General Multilevel Models for Two-Level Nested Data

- Topics:
  - Random slopes of level-1 predictors
  - Fun with cross-level interactions
    - Using cluster-mean-centered level-1 predictors
    - Using constant-centered level-1 predictors
    - Hybrid models to avoid smushed random effects
    - Level-2 interactions to avoid smushed cross-level interactions
    - Systematically varying effects—a compromise between fixed and random
  - How random slopes create heterogeneity of variance and covariance
  - An overview of model estimation and its practical consequences

MLMs for Clustered Data: Review

- Multilevel models (MLMs) are used to quantify and predict how much of an outcome's total variation is due to each dimension of sampling
- Empty means, two-level model for level-1 person $\boldsymbol{p}$ in level-2 cluster $\boldsymbol{c}$:

Level-1: $y_{pc} = \beta_{0c} + e_{pc}$

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

γ_{00} = fixed intercept (mean of cluster means)
U_{0c} = level-2 random intercept (with variance $\tau^2_{U_0}$)
e_{pc} = level-1 residual (with variance σ^2_e)

- **Total** outcome variation is partitioned into **two uncorrelated sources**:
  - **Level-2 between**-cluster (BC) mean differences → random intercept $\tau^2_{U_0}$
  - **Level-1 within**-cluster (WC) cluster differences → residual σ^2_e
  - Dependency effect size via Intraclass Correlation: $\mathbf{ICC} = \tau^2_{U_0} / (\tau^2_{U_0} + \sigma^2_e)$
    - ICC = proportion of total variance due to **cluster mean differences**
    - ICC = average correlation of persons from same cluster
- Fixed slopes of level-2 predictors explain cluster mean differences, thereby reducing the level-2 random intercept variance $\tau^2_{U_0}$

MLMs for Clustered Data: Review

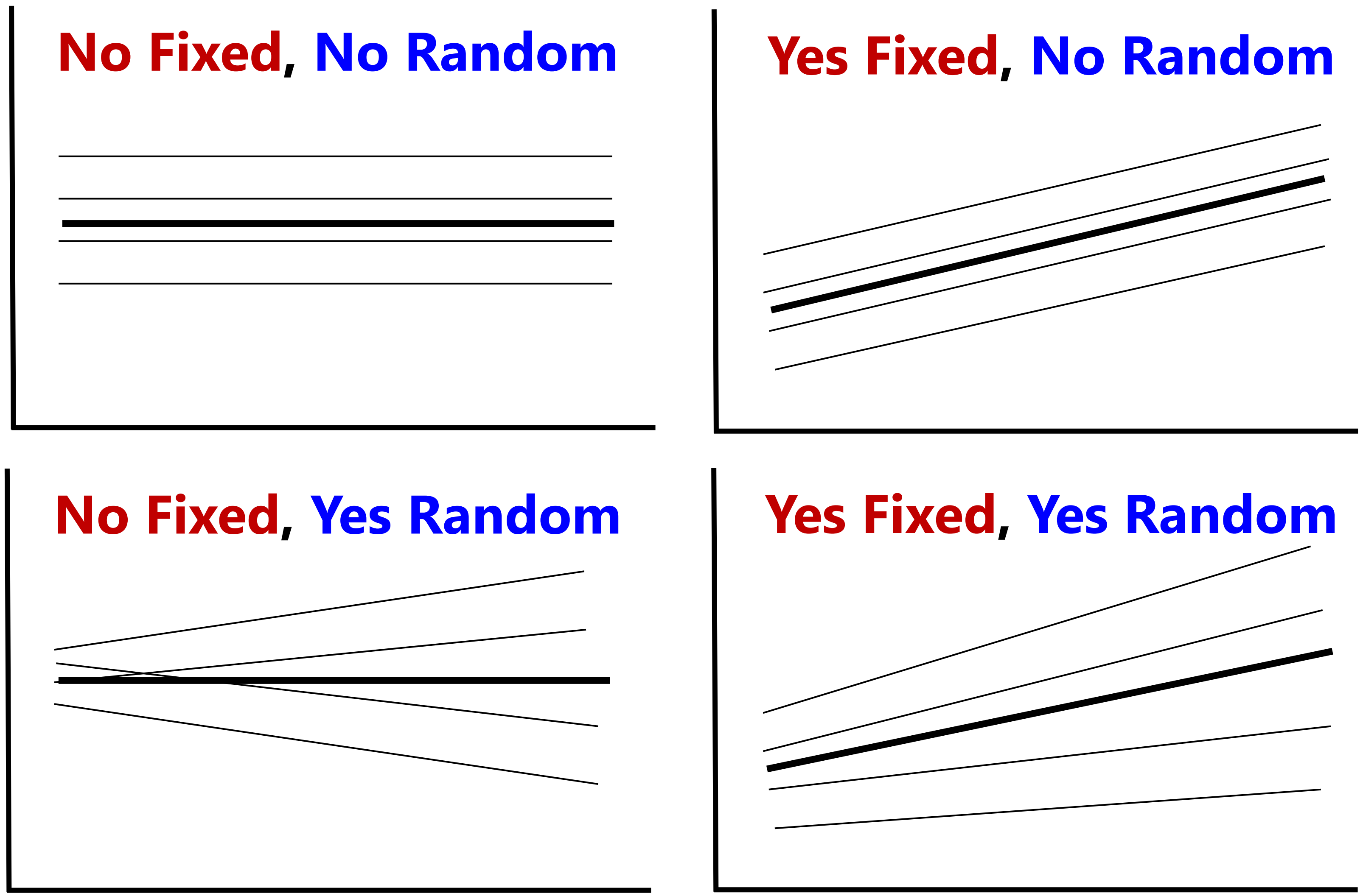
- **Level-1 predictors** are person characteristics, but they **almost always contain cluster mean differences** (level-2 variance) as well
  - **Variance** at each level → **different slope** at each level! (Yes, we must care!)

- **3 options** for specifying fixed slopes of a L1 predictor in order to distinguish its level-specific effects (i.e., **avoid smushed effects**):
  1. **Cluster-Mean-Centering** (univariate): carve up L1 pred into L2 BC (cluster mean → **L2 Between slope**) and L1 WC deviation (→ **L1 Within slope**)
  2. **Grand-Mean-Centering** (univariate): Add cluster mean to become **L2 Contextual slope**, then L1 predictor's unique effect is **L1 Within slope**
  3. **Latent-Centering** (multivariate): Let model estimate predictor's (and outcome's) L2 and L1 variance components → analogous to Cluster-MC

- But cluster-MC or latent-centering is needed instead to prevent a L1 predictor's **random slope** from being smushed…
  - **Fixed slope** → every cluster gets the **same slope** of the L1 predictor
  - **Random slope** → every cluster gets their **own slope** of the L1 predictor
    - To be explained by "**cross-level" interactions** of a L2 predictor with that L1 predictor!

Fixed and Random Slopes of L1 Predictor

(Note: The cluster intercept is random in every figure)



Cluster-MC Predictor* with Random Slope

$L1x_{pc}$ is cluster-mean-centered into WCx_{pc}, with CMx_c at L2:

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(WCx_{pc}) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

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

$WCx_{pc} = L1x_{pc} - \overline{L1x}_c$ → **only has L1 within variation**

$CMx_c = \overline{L1x}_c - C_2$ → **only has L2 between variation**

U_{1c} is a random slope for the WC effect of WCx_{pc}

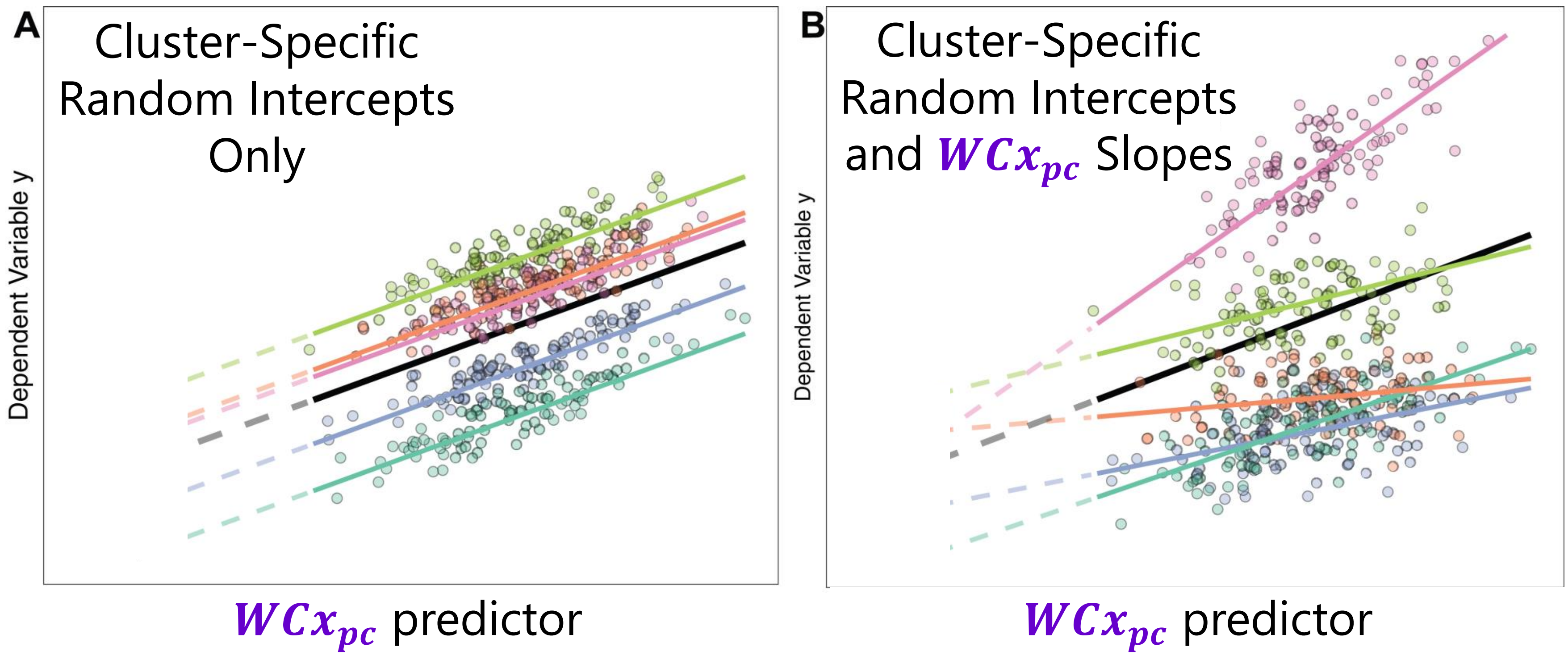
γ_{10} = within effect of having more $L1x_{pc}$ *than others in your cluster*

γ_{01} = between effect of having more $\overline{L1x}_c$ *than other clusters*

Because WCx_{pc} and CMx_c are uncorrelated, each gets the <u>total</u> effect for its level: L1 = within, L2 = between

* If a constant-centered L1 predictor were used instead, the U_{1c} random slope would also multiply its L2 between part, creating bias in the estimated random slope variance. **To avoid such a smushed random slope, we need to use either cluster-MC (in univariate MLM) or latent-centering (in multivariate MLM).**

Random Level-1 Slopes Across Clusters



- Both: the black line conveys the fixed slope for WCx_{pc}, γ_{01}
- Right: deviation for each cluster's WCx_{pc} slope is given by U_{1c}
  - Left: $\beta_{1c} = \gamma_{01}$ Right: $\beta_{1c} = \gamma_{10} + U_{1c}$

How to choose? Likelihood ratio tests: $-2\Delta LL$!

Image borrowed from: https://peerj.com/articles/4794/

When Cluster-MC ≠ Grand-MC: Random Slopes!

Cluster-MC: $WCx_{pc} = L1x_{pc} - CMx_c$

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(L1x_{pc} - CMx_c) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

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

Btw, I am using a centering constant = 0 at both levels to simplify the notation so that $\overline{L1x}_c = CMx_c$.

These two models for the means (fixed effects side) are equivalent!

$$y_{pc} = \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(L1x_{pc} - CMx_c) + U_{0c} + U_{1c}(L1x_{pc} - CMx_c) + e_{pc}$$

$$y_{pc} = \gamma_{00} + (\gamma_{01} - \gamma_{10})(CMx_c) + \gamma_{10}(L1x_{pc}) + U_{0c} + U_{1c}(L1x_{pc} - CMx_c) + e_{pc}$$

Grand-MC:

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(L1x_{pc}) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

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

L2 predictor CMx_c is also multiplied by the random slope in Grand-MC. So these random parts cannot be made equivalent without a separate contextual L2 "random slope" for CMx_c! (Rights & Sterba, in press)

$$\rightarrow y_{pc} = \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(L1x_{pc}) + U_{0c} + U_{1c}(L1x_{pc}) + e_{pc}$$

Example Random L1 Cluster-MC Within Slope: (2b) Syntax by Univariate MLM Program

SAS:

```
PROC MIXED DATA=work.Example COVTEST NOCLPRINT IC METHOD=REML;
     CLASS schoolID;                                * GCORR = random effect correlations;
     MODEL langpost = hw2 mixgrd CMverb10 WCverb / GCORR SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT WCverb / TYPE=UN SUBJECT=schoolID;
     ESTIMATE "L2 Contextual Effect of Verbal"  CMverb10 1 WCverb -1;
RUN;
```

R lmer from lme4 package—using lmerTest package to get Satterthwaite denominator DF and contest1D:

```
name = lmer(data=Example, REML=TRUE,
            formula=langpost~1+hw2+mixgrd+CMverb10+WCverb+(1+WCverb|schoolID))
summary(name, ddf="Satterthwaite") # Shows random effect correlations already
contest1D(name, ddf="Satterthwaite", L=c(0,0,0,1,-1)) # L2 Contextual effect of verbal
```

STATA:

```
mixed langpost c.hw2 c.mixgrd c.CMverb10 c.WCverb, || schoolID: WCverb, ///
      covariance(un) reml dfmethod(satterthwaite) dftable(pvalue) nolog
estat recovariance, relevel(schoolID) correlation  // Random effect correlations
lincom c.CMverb10*1 + c.WCverb*-1, small  // L2 Contextual effect of verbal
```

SPSS:

```
MIXED langpost BY schoolID WITH hw2 mixgrd CMverb10 WCverb
      /METHOD   = REML
      /CRITERIA = DFMETHOD(SATTERTHWAITE)
      /PRINT    = SOLUTION TESTCOV
      /FIXED    = hw2 mixgrd CMverb10 WCverb
      /RANDOM   = INTERCEPT WCverb | COVTYPE(UN) SUBJECT(schoolID)
      /TEST     = "L2 Contextual effect of verbal" CMverb10 1 WCverb -1.
```

Electronic materials for this example from my 2023 APA training sessions are **here**

Example: Cluster-MC Random Slope

Level-1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(HW_c - 2) + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal}_c - 10) + U_{0c}$

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

Adding L2 random slope variance of U_{1c} (as $\tau^2_{U_1}$) and L2 random intercept–slope covariance (as $\tau_{U_{01}}$)

Results from SAS MIXED:

L1 WCverb = $Verbal_{pc} - \overline{Verbal}_c$

L2 CMverb10 = $\overline{Verbal}_c - 10$

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.5281	0.3576	177	116.14	<.0001
hw2	-0.09509	0.4464	178	-0.21	0.8316
mixgrd	-0.9337	0.5052	201	-1.85	0.0660
CMverb10	3.6212	0.2647	209	13.68	<.0001
WCverb	2.4486	0.06831	151	35.85	<.0001

Btw, L2 Contextual = 1.173, SE = 0.273, p < .0001

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolID	8.4655	1.1352	[illegible]	[illegible]
UN(2,1)	schoolID	-0.6943	0.2386	[illegible]	[illegible]
UN(2,2)	schoolID	0.2239	0.08630	[illegible]	[illegible]
Residual		39.7586	0.9910	40.12	<.0001

Estimated G Correlation Matrix

Row	Effect	schoolID	Col1	Col2
1	Intercept	1	1.0000	-0.5043
2	WCverb	1	-0.5043	1.0000

Likelihood ratio test of random slope variance (and intercept–slope covariance):
−2ΔLL(~2) = 19.29, p < .0001

Example: Cluster-MC Random Slope

Level-1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(HW_c - 2) + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal}_c - 10) + U_{0c}$

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

Adding L2 random slope variance of U_{1c} (as $\tau^2_{U_1}$) and L2 random intercept–slope covariance (as $\tau_{U_{01}}$)

Results from SAS MIXED:

With random slope U_{1c}:

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.5281	0.3576	177	116.14	<.0001
hw2	-0.09509	0.4464	178	-0.21	0.8316
mixgrd	-0.9337	0.5052	201	-1.85	0.0660
CMverb10	3.6212	0.2647	209	13.68	<.0001
WCverb	2.4486	0.06831	151	35.85	<.0001

All estimates wiggle after adding U_{1c} because they are solved for after estimating the model for the variance parameters (stay tuned!)

Without random slope U_{1c}:

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.5794	0.3624	172	114.73	<.0001
hw2	-0.05255	0.4585	179	-0.11	0.9089
mixgrd	-1.1209	0.5157	197	-2.17	0.0309
CMverb10	3.6599	0.2709	207	13.51	<.0001
WCverb	2.4227	0.05718	3373	42.37	<.0001

Without U_{1c}, for the L1 verbal slope γ_{10}, the SE is too small and the DDF are too large

Effect Size via 95% Random Slope CIs

- e.g., "I have a significant fixed $\boldsymbol{WCx_{pc}}$ effect of $\boldsymbol{\gamma_{10} = 1.72}$, so there is a **positive effect on average**. I also have a significant L2 random slope variance for $\boldsymbol{WCx_{pc}}$ of $\boldsymbol{\tau^2_{U_1} = 0.91}$, so clusters need their own $\boldsymbol{WCx_{pc}}$ slope. But how big is a variance of $\boldsymbol{0.91}$ (i.e., besides >0)?"

- **95% Random Effects Confidence Intervals**
  - Can be calculated for *each effect that is random* in your model
  - Provide range around the fixed effect within which *95% of YOUR sample* is predicted to fall based on your random effect standard deviation:

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

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

  - So although $\boldsymbol{WCx_{pc}}$ has a positive effect on average (its fixed slope), the individual cluster slopes are predicted to range from −0.15 to 3.59 → some clusters are predicted to have a negative $\boldsymbol{WCx_{pc}}$ slope instead!
  - Is NOT the same as typical CI for fixed effect using fixed effect SE!

Effect Size via Reliability Indices

How reliable is a given level-2 cluster's random effect?

Intercept Reliability (ICC2):

$\tau^2_{U_0}$ = random intercept variance
σ^2_e = residual variance
$L1n$ = L1 sample size per L2 unit

$$\text{ICC2} = \frac{\tau^2_{U_0}}{\tau^2_{U_0} + \frac{\sigma^2_e}{L1n * 1}}$$

Slope Reliability (SR):

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

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

Although slope reliability is not commonly reported, it is useful for power analyses! Choose a target slope reliability, and then work backwards to determine what the random slope variance should be →

$$\tau^2_{U_1} = \frac{\text{SR} + \frac{\sigma^2_e}{L1n * \sigma^2_{L1}}}{1 - \text{SR}}$$

Intermediate Summary: Part 1

- Presently, **level-2** predictors refer to **cluster** characteristics
  - Traditionally, can have **fixed slopes** only in a two-level model
  - e.g., Does mean school achievement differ b/t rural and urban schools?

- Presently, **level-1** predictors refer to **person** characteristics
  - Can have **fixed slopes** **AND** **random slopes** over clusters
  - e.g., Does student achievement differ by student SES?
    - **Fixed slope:** e.g., Is there an SES difference ("gap") in achievement on average?
    - **Random slope:** e.g., Does the extent of the WC SES gap differ across schools? *(specified to multiply the cluster-MC or latent-centered version of L1 SES)*
  - When a level-1 predictor has both a fixed slope and a random slope, the fixed effect is the average of the level-2 per-cluster slopes
    - The level-1 fixed slope may differ before vs. after adding a random effect when clusters have different L1n (are unbalanced) for this reason (so keep it regardless of its significance)

- Significance tests for **random slope variances** (with covariances) via $-2\Delta LL$
  - If using REML, to-be-compared models must have same fixed effects
  - Random slope CIs and reliability indices can help convey effect size

Implications of Random Slopes

- **L2 random slopes** capture a second, distinct source of cluster **dependency**—differences in **slope of a L1 person predictor**
  - Beyond the **constant** covariance for L1 persons from same L2 cluster (as created by the **L2 random intercept**), the **L2 random slope** adds **non-constant** covariance across values of its L1 predictor (e.g., WCx_{pc})
  - Also adds **quadratic heterogeneity of variance** across L1 predictor: $Var(y_{pc}) = \tau_{U_0}^2 + (WCx_{pc}^2 * \tau_{U_1}^2) + (2WCx_{pc} * \tau_{U_{01}}) + \sigma_e^2$ **Why? Stay tuned!**
- **Random slopes do NOT*** explain variance (like **fixed slopes** do) because cluster slope differences are still "**error**" conceptually
  - We know **THAT** clusters need different slopes of **L1** WCx_{pc} but not **WHY**
- Therefore, random slopes imply another role for level-2 cluster predictors—to explain cluster differences in slope of **L1** WCx_{pc}
  - To do so, we need "**cross-level interactions**" of L2 by L1 predictors!

** Hill that I will die on, but others disagree (see marginal vs. conditional R^2)*

Introducing Cross-Level Interactions

- A **cross-level interaction** is among predictors at different levels; shown here is an "**intra-variable**" cross-level interaction of the L1 within and L2 between parts of the same L1 predictor
- Cross-level interactions explain the random slope variance of L1 person predictor across L2 clusters $\tau^2_{U_1}$—here is a generic example:

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(WCx_{pc}) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{11}(CMx_c) + U_{1c}$

$WCx_{pc} = L1x_{pc} - \overline{L1x}_c$ → only has L1 within variation

$CMx_c = \overline{L1x}_c - C_2$ → only has L2 between variation

U_{1c} is a random slope for the WC effect of WCx_{pc}

γ_{10} = within effect of more $L1x_{pc}$ *than others in your cluster,* now for $CMx_c = 0$

γ_{01} = between effect of more $\overline{L1x}_c$ *than other clusters,* now for $WCx_{pc} = 0$

γ_{11} = diff in within effect of WCx_{pc} per unit CMx_c OR diff in between effect of CMx_c per unit WCx_{pc}

Example Adding Cross-Level Interactions: (2c) Syntax by Univariate MLM Program

SAS:

```
PROC MIXED DATA=work.Example COVTEST NOCLPRINT IC METHOD=REML;
     CLASS schoolID;                                  * In SAS, * creates interactions;
     MODEL langpost = hw2 mixgrd CMverb10 WCverb hw2*WCverb mixgrd*WCverb
                      CMverb10*WCverb / GCORR SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT WCverb / TYPE=UN SUBJECT=schoolID;
RUN;
```

R lmer from lme4 package—using lmerTest package to get Satterthwaite denominator DF:

```
name = lmer(data=Example, REML=TRUE,
            formula=langpost~1+hw2+mixgrd+CMverb10+WCverb+ hw2:WCverb +mixgrd:WCverb
                    +CMverb10:WCverb+(1+WCverb|schoolID))
summary(name, ddf="Satterthwaite") # In R, : creates interactions
```

STATA:

```
mixed langpost c.hw2 c.mixgrd c.CMverb10 c.WCverb c.hw2#c.WCverb c.mixgrd#c.WCverb ///
      c.CMverb10#c.WCverb, || schoolID: WCverb, /// In STATA, # creates interactions
      covariance(un) reml dfmethod(satterthwaite) dftable(pvalue) nolog
estat recovariance, relevel(schoolID) correlation  // Random effect correlations
```

SPSS: * In SPSS, * creates interactions.

```
MIXED langpost BY schoolID WITH hw2 mixgrd CMverb10 WCverb
      /METHOD   = REML
      /CRITERIA = DFMETHOD(SATTERTHWAITE)
      /PRINT    = SOLUTION TESTCOV
      /FIXED    = hw2 mixgrd CMverb10 WCverb hw2*WCverb mixgrd*WCverb CMverb10*WCverb
      /RANDOM   = INTERCEPT WCverb | COVTYPE(UN) SUBJECT(schoolID).
```

Electronic materials for this example from my 2023 APA training sessions are **here**

Example: Add 3 Cross-Level Interactions

Level-1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - \overline{Verbal_c}) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(HW_c - 2) + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal_c} - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{11}(HW_c - 2) + \gamma_{12}(MixGrd_c) + \gamma_{13}(\overline{Verbal_c} - 10) + U_{1c}$

Results from SAS MIXED—having more verbal ability than your peers matters more for your language score in schools with mixed grades:

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.5831	0.3629	172	114.58	<.0001
hw2	-0.04595	0.4590	179	-0.10	0.9204
mixgrd	-1.1368	0.5160	197	-2.20	0.0288
CMverb10	3.6445	0.2710	207	13.45	<.0001
WCverb	2.3903	0.1002	129	23.86	<.0001
hw2*WCverb	-0.05601	0.1305	143	-0.43	0.6683
mixgrd*WCverb	0.3210	0.1588	228	2.02	0.0444
CMverb10*WCverb	-0.04367	0.07805	182	-0.56	0.5765

L1 WCverb slope is now specifically for hw=2, mixgrd=0, and CMverb=10; those 3 slopes are now specifically for WCverb=0 (at school mean)

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolID	8.4680	1.1350	[illegible]	[illegible]
UN(2,1)	schoolID	-0.7095	0.2379	[illegible]	[illegible]
UN(2,2)	schoolID	0.2231	0.08640	[illegible]	[illegible]
Residual		39.7407	0.9903	[illegible]	[illegible]

Relative to the previous model, the 3 new cross-level interactions explained **0.04%** of the **L2 random WCverb slope** variance

L1 WCverb slope is significantly more positive (stronger) in schools with mixed grades (and nonsignificantly weaker in schools with more homework and higher mean verbal ability).

Cross-Level Interactions: Danger Ahead!

- To continue, let's use a simplified version of the prior example without the 3 nonsignificant slopes (γ_{01}, γ_{11}, and γ_{13}) in L2 model:
  - **Level-1:** $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$
  - **Level-2:** $\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}\left(\overline{Verbal}_c - 10\right) + U_{0c}$

    $\beta_{1c} = \gamma_{10} + \gamma_{12}(MixGrd_c) + U_{1c}$
  - Because we had **cluster-mean-centered** L1 verbal (→ within info only), the cross-level interaction γ_{12} gives difference of **L1 within-school** verbal slope for L2 schools with mixed grades (versus ref = not mixed)

- What if we had **constant-centered** L1 verbal (→ info for both levels still) to get L2 contextual slopes directly as L2 fixed effects instead?
  - L1 fixed "main" verbal effect would still be unsmushed **L1 within slope** controlling for **L2 contextual** fixed "main" effect of school mean verbal
  - The **random slope of L1 verbal** would still be smushed, though! To fix it, we need a "**hybrid**" model, where the **fixed slopes** and **random slopes** multiply **different level-1 predictors…!**

Hybrid Model: Fixed Main Effects Only

- Goal: Provide **L2 contextual effects** directly on the fixed side of the model **without smushing the random slope** of L1 predictor
- The solution is known as a "**hybrid**" model (see below): **Fixed slope** → constant-centered; **random slope** → cluster-MC

L1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - 10) + \beta_{2c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

L2: $\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal}_c - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10}$ → ***No random slope!***

$\beta_{2c} = U_{2c}$ → ***No fixed slope!***

Composite:

$$Lang_{pc} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal}_c - 10) + \gamma_{10}(Verbal_{pc} - 10) + U_{0c} + U_{2c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$$

Hybrid Model with Fixed Main Effects Only: (4a) Syntax by Univariate MLM Program

Oops! Predictor hw2 should not be included.

SAS:

```
PROC MIXED DATA=work.Example COVTEST NOCLPRINT IC METHOD=REML;
     CLASS schoolID;
     MODEL langpost = hw2 mixgrd CMverb10 verb10 / GCORR SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT WCverb / TYPE=UN SUBJECT=schoolID;
RUN;
```

R lmer from lme4 package—using lmerTest package to get Satterthwaite denominator DF:

```
name = lmer(data=Example, REML=TRUE,
            formula=langpost~1+hw2+mixgrd+CMverb10+verb10+(1+WCverb|schoolID))
summary(name, ddf="Satterthwaite")
```

STATA:

```
mixed langpost c.hw2 c.mixgrd c.CMverb10 c.verb10, || schoolID: WCverb, ///
      covariance(un) reml dfmethod(satterthwaite) dftable(pvalue) nolog
estat recovariance, relevel(schoolID) correlation  // Random effect correlations
```

SPSS: `* In SPSS, * creates interactions.`

```
MIXED langpost BY schoolID WITH hw2 mixgrd CMverb10 verb10 WCverb
      /METHOD    = REML
      /CRITERIA  = DFMETHOD(SATTERTHWAITE)
      /PRINT     = SOLUTION TESTCOV
      /FIXED     = hw2 mixgrd CMverb10 verb10
      /RANDOM    = INTERCEPT WCverb | COVTYPE(UN) SUBJECT(schoolID).
```

Electronic materials for this example **added** to my 2023 APA training sessions are **here**

Comparing Models for the Variance

L1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - 10) + \beta_{2c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

L2: $\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal}_c - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10} + (U_{1c})$; $\beta_{2c} = (U_{2c})$

(U) indicates only one or the other

Results from SAS MIXED—different results!

With random slope U_{2c} for **WCverb** (cluster-MC version):

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolID	8.4655	1.1352	[illegible]	[illegible]
UN(2,1)	schoolID	-0.6943	0.2386	[illegible]	[illegible]
UN(2,2)	schoolID	0.2239	0.08630	[illegible]	[illegible]
Residual		39.7586	0.9910	[illegible]	[illegible]

Covariance Parameter Estimates

Row	Effect	schoolID	Col1	Col2
1	Intercept	1	1.0000	-0.5043
2	WCverb	1	-0.5043	1.0000

Estimated G Correlation Matrix

With random slope U_{1c} for **verb10** (constant-C version):

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	schoolID	8.5134	1.1444	[illegible]	[illegible]
UN(2,1)	schoolID	-0.6962	0.2370	[illegible]	[illegible]
UN(2,2)	schoolID	0.1996	0.08092	[illegible]	[illegible]
Residual		39.8104	0.9917	[illegible]	[illegible]

Covariance Parameter Estimates

Row	Effect	schoolID	Col1	Col2
1	Intercept	1	1.0000	-0.5341
2	verb10	1	-0.5341	1.0000

Estimated G Correlation Matrix

Hybrid Model: Fixed Main Effects Only

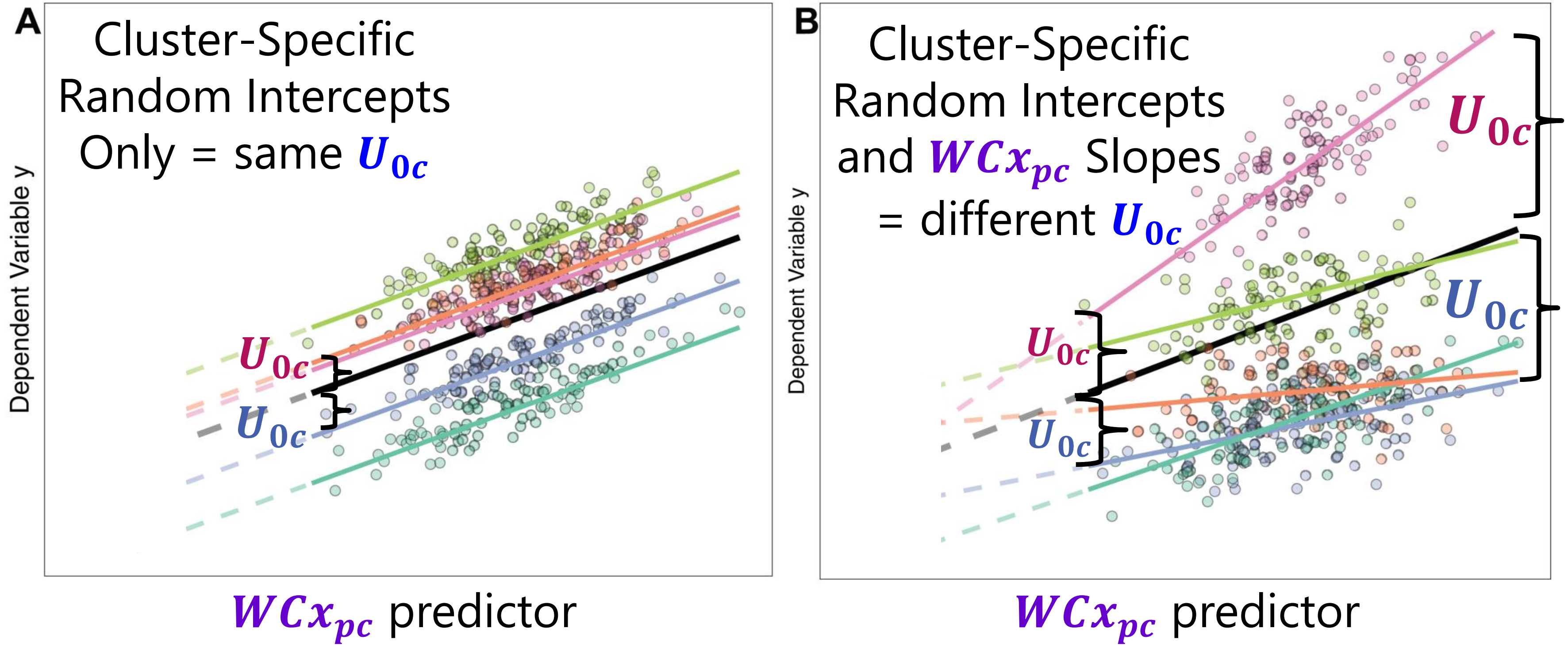
- Goal: Provide **L2 contextual effects** directly on the fixed side of the model **without smushing the random slope** of L1 predictor

Composite: $Lang_{pc} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}\left(\overline{Verbal}_c - 10\right)$
$+ \gamma_{10}\left(Verbal_{pc} - 10\right) + U_{0c} + U_{2c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

Interpreting the Model for the Variance:

- U_{0c} = **level-2 random intercept** → deviation of original from predicted mean language for school c (where "original" is from an empty means, random intercept model), now **specifically where student verbal = their school mean** (with variance = $\tau^2_{U_0}$)
- U_{2c} = **level-2 random slope** → deviation of **L1 within** verbal slope for school c from γ_{10}, its average slope across all schools (with variance = $\tau^2_{U_2}$ and U_{0c} covariance = $\tau_{U_{02}}$)
  - *If applied to constant-centered student verbal instead, it would reflect both school differences in the L1 within verbal slope AND intercept heteroscedasticity (bad)*
- e_{pc} = **level-1 residual** = deviation of the observed outcome for student p from their outcome predicted by all fixed and random effects

Random Level-1 Slopes Across Clusters



- Both: the black line conveys the fixed slope for WCx_{pc}, γ_{01}
- After adding a L1 predictor's random slope, the random intercept no longer applies equally along that predictor—**the random intercept is then specifically at a predictor value = 0** (and will differ at a new "0")

Image borrowed from: https://peerj.com/articles/4794/

Hybrid Model: Add a Cross-Level Interaction

- Goal: explain school differences in **L1 within-school** verbal slope (random variance $\tau^2_{U_2}$) using cross-level interaction with **L2** $MixGrd_c$
  - Effect size would be found using pseudo-R^2 for the random slope variance, so **ALWAYS test for L2 random slope variance of the L1 fixed slope first** before examining any of its cross-level interactions—otherwise you'll have high Type I errors for the cross-level interaction if you omit a necessary L2 random slope!

L1: $$Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - 10) + \beta_{2c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$$

L2: $$\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal}_c - 10) + U_{0c}$$

$$\beta_{1c} = \gamma_{10} + \gamma_{12}(MixGrd_c)$$

$$\beta_{2c} = U_{2c}$$

- All good, right? Nope—many researchers may mistakenly think so, but this model is now VERY LIKELY to be mis-specified at L2
  - Same problem as when adding the fixed main effect of a constant-centered L1 predictor by itself without a fixed main effect of its L2 cluster mean!

Hybrid: Smushed Cross-Level Interaction

L1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - 10) + \beta_{2c}(Verbal_{pc} - \overline{Verbal_c}) + e_{pc}$

L2: $\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal_c} - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{12}(MixGrd_c)$; $\beta_{2c} = U_{2c}$

Interpreting Fixed Effects:

- γ_{00} = **intercept**: expected language for a student with verbal = 10 from a school with school mean verbal = 10 and no mixed grades
- γ_{10} = ***simple* L1 within slope**: difference in student language per unit higher verbal than school mean, *specifically for schools without mixed grades*
- γ_{03} = **L2 contextual slope**: extra difference in school language per unit higher school mean verbal than other schools (controlling for student verbal; not explicitly conditional on mixed grade)
- $\gamma_{10} + \gamma_{03}$ = **L2 between slope**: difference in school language per unit higher school mean verbal than other schools (NOT controlling for student verbal; not explicitly conditional on mixed grade)
- γ_{12} = **smushed cross-level interaction**: how the **L1 within slope** AND the **L2 between slope** each differ in schools with mixed grades → **assumes equal moderation by mixed grade** of L1 within and L2 between slopes!

Hybrid: Unsmushed Cross-Level Interaction

L1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - 10) + \beta_{2c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

L2: $\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal}_c - 10)$

$+ \gamma_{04}(MixGrd_c)(\overline{Verbal}_c - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{12}(MixGrd_c)$; $\beta_{2c} = U_{2c}$

Interpreting Fixed Effects:

- γ_{10} = ***simple* L1 within slope**: difference in student language per unit higher verbal than school mean, *specifically for schools without mixed grades*
- γ_{03} = ***simple* L2 contextual slope**: extra difference in school language per unit higher school mean verbal than other schools (controlling for student verbal), *now specifically for schools without mixed grades*
- $\gamma_{10} + \gamma_{03}$ = ***simple* L2 between slope**: difference in school language per unit higher school mean verbal than other schools (NOT controlling for student verbal), *now specifically for schools without mixed grades*
- γ_{12} = **unsmushed cross-level interaction**: how the **L1 within** verbal slope differs in schools with mixed grades
- γ_{04} = **new level-2 interaction**: how the **L2 contextual** verbal slope differs in schools with mixed grades (added to unsmush cross-level interaction γ_{12})
- $\gamma_{12} + \gamma_{04}$ = **implied level-2 interaction**: how the **L2 between** verbal slope differs in schools with mixed grades

Hybrid with Unsmushed Cross-Level Int: (4d) Syntax by Univariate MLM Program

Oops! Predictor hw2 should not be included.

SAS:

```
PROC MIXED DATA=work.Example COVTEST NOCLPRINT IC METHOD=REML;
     CLASS schoolID;          * In SAS, * creates interactions;
     MODEL langpost = hw2 mixgrd CMverb10 verb10 mixgrd*verb10 mixgrd*CMverb10
                    / GCORR SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT WCverb / TYPE=UN SUBJECT=schoolID;
RUN;
```

R lmer from lme4 package—using lmerTest package to get Satterthwaite denominator DF:

```
name = lmer(data=Example, REML=TRUE,
            formula=langpost~1+hw2+mixgrd+CMverb10+verb10+ mixgrd:verb10
                             +mixgrd:CMverb10+(1+WCverb|schoolID))
summary(name, ddf="Satterthwaite") # In R, : creates interactions
```

STATA:

```
mixed langpost c.hw2 c.mixgrd c.CMverb10 c.verb10 c.mixgrd#c.verb10 ///
      c.mixgrd#c.CMverb10, || schoolID: WCverb, /// In STATA, # creates interactions
      covariance(un) reml dfmethod(satterthwaite) dftable(pvalue) nolog
estat recovariance, relevel(schoolID) correlation  // Random effect correlations
```

SPSS: `* In SPSS, * creates interactions.`

```
MIXED langpost BY schoolID WITH hw2 mixgrd CMverb10 verb10 WCverb
      /METHOD   = REML
      /CRITERIA = DFMETHOD(SATTERTHWAITE)
      /PRINT    = SOLUTION TESTCOV
      /FIXED    = hw2 mixgrd CMverb10 verb10 mixgrd*verb10 mixgrd*CMverb10
      /RANDOM   = INTERCEPT WCverb | COVTYPE(UN) SUBJECT(schoolID).
```

Electronic materials for this example **added** to my 2023 APA training sessions are **here**

Unsmushed vs Smushed Cross-Level Int

L1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - 10) + \beta_{2c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

L2: $\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal}_c - 10)$

$+ \gamma_{04}(MixGrd_c)(\overline{Verbal}_c - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{12}(MixGrd_c)$; $\beta_{2c} = U_{2c}$

Oops! Predictor hw2 should not be included.

With L2 interaction γ_{04}:

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.6217	0.3592	176	115.86	<.0001
~~hw2~~	~~-0.06631~~	~~0.4457~~	~~178~~	~~-0.15~~	~~0.8819~~
mixgrd	-1.1165	0.5161	197	-2.16	0.0317
CMverb10	0.8255	0.4053	182	2.04	0.0431
verb10	2.3613	0.07856	124	30.06	<.0001
mixgrd*verb10	0.3362	0.1567	239	2.15	0.0329
mixgrd*CMverb10	0.4136	0.5536	235	0.75	0.4557

L2 context interaction γ_{04} for mixgrd*CMverb10 is also the **difference in moderation** by mixgrd of the L1 within and L2 between verbal slopes

Without L2 interaction γ_{04}:

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.6228	0.3585	177	116.11	<.0001
~~hw2~~	~~-0.08069~~	~~0.4446~~	~~178~~	~~-0.18~~	~~0.8562~~
mixgrd	-1.1575	0.5120	199	-2.26	0.0249
CMverb10	1.0471	0.2766	245	3.79	0.0002
verb10	2.3540	0.07791	128	30.21	<.0001
mixgrd*verb10	0.3692	0.1502	281	2.46	0.0146

Cross-level interaction γ_{12} for mixgrd*verb10 **assumes equal moderation** by mixgrd of the L1 within and L2 between verbal slopes (and here is positively biased by missing γ_{04})

Non-Hybrid: All Cluster-MC Version

L1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

L2: $\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal}_c - 10)$
$+ \gamma_{04}(MixGrd_c)(\overline{Verbal}_c - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{12}(MixGrd_c) + U_{1c}$

Interpreting Fixed Effects:

- γ_{10} = ***simple*** **L1 within slope**: difference in student language per unit higher verbal than school mean, *specifically for schools without mixed grades*
- γ_{03} = ***simple*** **L2 between slope**: difference in school language per unit higher school mean verbal than other schools (NOT controlling for student verbal), *now specifically for schools without mixed grades*
- $\gamma_{03} - \gamma_{10}$ = ***simple*** **L2 contextual slope**: extra difference in school language per unit higher school mean verbal than other schools (controlling for student verbal), *now specifically for schools without mixed grades*
- γ_{12} = **guaranteed-to-be-unsmushed cross-level interaction**: how the **L1 within** verbal slope differs in schools with mixed grades
- γ_{04} = **level-2 interaction**: how the **L2 between** verbal slope differs in schools with mixed grades
- $\gamma_{04} - \gamma_{12}$ = **implied level-2 interaction**: how the **L2 contextual** verbal slope differs in schools with mixed grades (or how moderation differs: BC − WC)

Cluster-MC with Unsmushed Cross-Level Int: (4e) Syntax by Univariate MLM Program

SAS:

Oops! Predictor hw2 should not be included.

```
PROC MIXED DATA=work.Example COVTEST NOCLPRINT IC METHOD=REML;
     CLASS schoolID;          * In SAS, * creates interactions;
     MODEL langpost = hw2 mixgrd CMverb10 WCverb mixgrd*WCverb mixgrd*CMverb10
                    / GCORR SOLUTION DDFM=Satterthwaite;
     RANDOM INTERCEPT WCverb / TYPE=UN SUBJECT=schoolID;
RUN;
```

R lmer from lme4 package—using lmerTest package to get Satterthwaite denominator DF and contest1D:

```
name = lmer(data=Example, REML=TRUE,
            formula=langpost~1+hw2+mixgrd+CMverb10+WCverb+ mixgrd:WCverb
                             +mixgrd:CMverb10+(1+WCverb|schoolID))
summary(name, ddf="Satterthwaite") # In R, : creates interactions
```

STATA:

```
mixed langpost c.hw2 c.mixgrd c.CMverb10 c.WCverb c.mixgrd#c.WCverb ///
      c.mixgrd#c.CMverb10, || schoolID: WCverb, /// In STATA, # creates interactions
      covariance(un) reml dfmethod(satterthwaite) dftable(pvalue) nolog
estat recovariance, relevel(schoolID) correlation  // random effect correlations
```

SPSS: `* In SPSS, * creates interactions.`

```
MIXED langpost BY schoolID WITH hw2 mixgrd CMverb10 WCverb
      /METHOD    = REML
      /CRITERIA  = DFMETHOD(SATTERTHWAITE)
      /PRINT     = SOLUTION TESTCOV
      /FIXED     = hw2 mixgrd CMverb10 WCverb mixgrd*WCverb mixgrd*CMverb10
      /RANDOM    = INTERCEPT WCverb | COVTYPE(UN) SUBJECT(schoolID).
```

Electronic materials for this example **added** to my 2023 APA training sessions are here

Hybrid vs. Cluster-MC: Different L2 Slopes!

Hybrid: $\beta_{1c}(Verbal_{pc} - 10)$
→ Direct **L2 Context** Effects

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	41.6217	0.3592	176	115.86	<.0001
~~hw2~~	~~-0.06631~~	~~0.4457~~	~~178~~	~~-0.15~~	~~0.8819~~
mixgrd	-1.1165	0.5161	197	-2.16	0.0317
CMverb10	0.8255	0.4053	182	2.04	0.0431
verb10	2.3613	0.07856	124	30.06	<.0001
mixgrd*verb10	0.3362	0.1567	239	2.15	0.0329
mixgrd*CMverb10	0.4136	0.5536	235	0.75	0.4557

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Simple L2 between	3.1868	0.3992	172	7.98	<.0001
L2 between*mixgrd	0.7498	0.5308	199	1.41	0.1593

CMC: $\beta_{1c}(Verbal_{pc} - \overline{Verbal_c})$
→ Direct **L2 Between** Effects

Solution for Fixed Effects					
Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	41.6217	0.3592	176	115.86	<.0001
~~hw2~~	~~0.06631~~	~~0.4457~~	~~178~~	~~0.15~~	~~0.8819~~
mixgrd	-1.1165	0.5161	197	-2.16	0.0317
CMverb10	3.1868	0.3992	172	7.98	<.0001
WCverb	2.3613	0.07856	124	30.06	<.0001
mixgrd*WCverb	0.3362	0.1567	239	2.15	0.0329
mixgrd*CMverb10	0.7498	0.5308	199	1.41	0.1593

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Simple L2 context	0.8255	0.4053	182	2.04	0.0431
L2 context*mixgrd	0.4136	0.5536	235	0.75	0.4557

L1 within verbal slope is signif more positive (stronger) by **0.3362** in mixed-grade schools
L2 between verbal slope is n.s. more positive (stronger) by **0.7498** in mixed-grade schools
L2 contextual verbal slope is n.s. more positive (stronger) by **0.4136** in mixed-grade schools

Same Model for the Variance Either Way

L1: $Lang_{pc} = \beta_{0c} + \beta_{1c}(Verbal_{pc} - 10) + \beta_{2c}(Verbal_{pc} - \overline{Verbal_c}) + e_{pc}$

L2: $\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}(\overline{Verbal_c} - 10)$
$+ \gamma_{04}(MixGrd_c)(\overline{Verbal_c} - 10) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{12}(MixGrd_c)$; $\beta_{2c} = U_{2c}$

Hybrid →
$\beta_{1c}(Verbal_{pc} - 10)$
Cluster-MC →
$\beta_{1c}(Verbal_{pc} - \overline{Verbal_c})$

Interpreting the Model for the Variance:

- U_{0c} = **level-2 random intercept →** deviation of original from predicted mean language for school *c* (where "original" is from an empty means, random intercept model), now **specifically where student verbal = their school mean** (with variance = $\tau^2_{U_0}$)
- U_{2c} = **level-2 random slope →** deviation of original from predicted **L1 within** verbal slope for school *c* (where "original" is from a model without cross-level interactions for β_{1c}), (with variance = $\tau^2_{U_2}$ and U_{0c} covariance = $\tau_{U_{02}}$)
  - *If applied to constant-centered student verbal instead, it would reflect both school differences in the L1 within verbal slope AND intercept heteroscedasticity (bad)*
- e_{pc} = **level-1 residual** = deviation of the observed outcome for student *p* from their outcome predicted by all fixed and random effects

Intra-Variable Cross-Level Interactions

- What if we wanted to see if the **L1 within effect** *(of more verbal ability than your peers on student math)* depends on how much verbal ability your peers have on average *(school mean verbal)*?
  - Back to the **hybrid model** to illustrate:

L1: $Lang_{pc} = \beta_{0c} + \beta_{1c}\left(Verbal_{pc} - 10\right) + \beta_{2c}(Verbal_{pc} - \overline{Verbal}_c) + e_{pc}$

L2: $\beta_{0c} = \gamma_{00} + \gamma_{02}(MixGrd_c) + \gamma_{03}\left(\overline{Verbal}_c - 10\right)$

$+ \gamma_{04}(MixGrd_c)\left(\overline{Verbal}_c - 10\right) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{12}(MixGrd_c) + \gamma_{13}\left(\overline{Verbal}_c - 10\right); \beta_{2c} = U_{2c}$

- Same potential for a smushed cross-level interaction when using a constant-centered L1 predictor in the intra-variable interaction
  - Slope γ_{13} says the L1 within and L2 between verbal slopes are moderated to the same extent by school mean verbal
  - The solution is the same as before, but it looks strange at first…!

Intra-Variable Cross-Level Interactions

- To unsmush the cross-level interaction, we add the corresponding L2 interaction with the L2 moderator, just as we did before…

L1: $Lang_{pc} = \beta_{0c} + \beta_{1c}\left(Verbal_{pc} - 10\right) + \beta_{2c}\left(Verbal_{pc} - \overline{Verbal}_c\right) + e_{pc}$

L2: $\beta_{0c} = \gamma_{00} + \gamma_{02}\left(MixGrd_c\right) + \gamma_{03}\left(\overline{Verbal}_c - 10\right)$

$+ \gamma_{04}\left(MixGrd_c\right)\left(\overline{Verbal}_c - 10\right)$

$+ \gamma_{05}\left(\overline{Verbal}_c - 10\right)\left(\overline{Verbal}_c - 10\right) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{12}\left(MixGrd_c\right) + \gamma_{13}\left(\overline{Verbal}_c - 10\right)$; $\beta_{2c} = U_{2c}$

- …the solution is a **quadratic slope** for L2 school mean verbal!
  - γ_{13} = how the **L1 within** verbal slope differs by school mean verbal
  - γ_{05} = how the **L2 contextual** verbal slope differs by school mean verbal
  - $\gamma_{13} + \gamma_{05}$ = how the **L2 between** verbal slope differs by school mean verbal

Hybrid with Unsmushed Intra-Variable Int: (5a) Syntax by Univariate MLM Program

SAS:

Oops! Predictor hw2 should not be included.

```
PROC MIXED DATA=work.Example COVTEST NOCLPRINT IC METHOD=REML;
     CLASS schoolID;           * In SAS, * creates interactions;
     MODEL langpost = hw2 mixgrd CMverb10 verb10 mixgrd*verb10 mixgrd*CMverb10
                      CMverb10*verb10 CMverb10*Cmverb10/ GCORR SOLUTION DDFM=Satterth;
     RANDOM INTERCEPT WCverb / TYPE=UN SUBJECT=schoolID;
RUN;
```

R lmer from lme4 package—using lmerTest package to get Satterthwaite denominator DF and contest1D:

```
name = lmer(data=Example, REML=TRUE,
            formula=langpost~1+hw2+mixgrd+CMverb10+verb10+I(CMverb10^2)
                  +mixgrd:verb10+mixgrd:CMverb10+CMverb10:verb10+(1+WCverb|schoolID))
summary(name, ddf="Satterthwaite") # In R, : creates interactions, I(^2) creates quad
```

STATA:

```
                                                     //In STATA, # creates interactions
mixed langpost c.hw2 c.mixgrd c.CMverb10 c.verb10 c.mixgrd#c.verb10  ///
      c.mixgrd#c.CMverb10 c.CMverb10#c.verb10 c.CMverb10#c.CMverb10, ///
      || schoolID: WCverb, covariance(un) reml dfmethod(satterthwaite) dftable(pvalue)
estat recovariance, relevel(schoolID) correlation  // random effect correlations
```

SPSS:

```
* In SPSS, * creates interactions.
MIXED langpost BY schoolID WITH hw2 mixgrd CMverb10 verb10 WCverb
      /METHOD   = REML
      /CRITERIA = DFMETHOD(SATTERTHWAITE)
      /PRINT    = SOLUTION TESTCOV
      /FIXED    = hw2 mixgrd CMverb10 verb10 mixgrd*verb10 mixgrd*CMverb10
                  CMverb10*verb10 CMverb10*CMverb10
      /RANDOM   = INTERCEPT WCverb | COVTYPE(UN) SUBJECT(schoolID).
```

Electronic materials for this example **added** to my 2023 APA training sessions are **here**

Cluster-MC with Intra-Variable Interaction: (5b) Syntax by Univariate MLM Program

Oops! Predictor hw2 should not be included.

SAS:

```
PROC MIXED DATA=work.Example COVTEST NOCLPRINT IC METHOD=REML;
     CLASS schoolID;          * In SAS, * creates interactions;
     MODEL langpost = hw2 mixgrd CMverb10 WCverb mixgrd*WCverb mixgrd*CMverb10
                      CMverb10*WCverb CMverb10*CMverb10 / GCORR SOLUTION DDFM=Satterth;
     RANDOM INTERCEPT WCverb / TYPE=UN SUBJECT=schoolID;
RUN;
```

R lmer from lme4 package—using lmerTest package to get Satterthwaite denominator DF:

```
name = lmer(data=Example, REML=TRUE,
            formula=langpost~1+hw2+mixgrd+CMverb10+WCverb+I(CMverb10^2)
            +mixgrd:WCverb+mixgrd:CMverb10+CMverb10:WCverb+(1+WCverb|schoolID))
summary(name, ddf="Satterthwaite") # In R, : creates interactions, I(^2) creates quad
```

STATA:

```
                                                  /// In STATA, # creates interactions
mixed langpost c.hw2 c.mixgrd c.CMverb10 c.WCverb c.mixgrd#c.WCverb  ///
      c.mixgrd#c.CMverb10 c.CMverb10#c.WCverb c.CMverb10#c.CMverb10, ///
      || schoolID: WCverb, covariance(un) reml dfmethod(satterthwaite) dftable(pvalue)
estat recovariance, relevel(schoolID) correlation  // Random effect correlations
```

SPSS:

```
                                                   * In SPSS, * creates interactions.
MIXED langpost BY schoolID WITH hw2 mixgrd CMverb10 WCverb
      /METHOD   = REML
      /CRITERIA = DFMETHOD(SATTERTHWAITE)
      /PRINT    = SOLUTION TESTCOV
      /FIXED    = hw2 mixgrd CMverb10 WCverb mixgrd*WCverb mixgrd*CMverb10
                  CMverb10*WCverb CMverb10*CMverb10
      /RANDOM   = INTERCEPT WCverb | COVTYPE(UN) SUBJECT(schoolID).
```

Electronic materials for this example **added** to my 2023 APA training sessions are here

Hybrid vs. Cluster-MC: Different L2 Slopes!

Hybrid: $\beta_{1c}(Verbal_{pc} - 10)$ → Direct **L2 Context** Effects	CMC: $\beta_{1c}(Verbal_{pc} - \overline{Verbal_c})$ → Direct **L2 Between** Effects

Hybrid — Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.8383	0.3628	178	115.32	<.0001
~~hw2~~	~~-0.06701~~	~~0.4367~~	~~176~~	~~-0.15~~	~~0.8782~~
mixgrd	-0.8419	0.5213	196	-1.61	0.1079
CMverb10	0.8592	0.4061	192	2.12	0.0357
verb10	2.3589	0.07905	123	29.84	<.0001
mixgrd*verb10	0.3394	0.1578	231	2.15	0.0325
mixgrd*CMverb10	-0.1281	0.5772	231	-0.22	0.8246
CMverb10*verb10	-0.04328	0.07779	179	-0.56	0.5787
CMverb10*CMverb10	-0.3817	0.1671	344	-2.28	0.0229

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Simple L2 between	3.2181	0.4004	181	8.04	<.0001
L2 between*mixgrd	0.2113	0.5549	198	0.38	0.7038
L2 between*CMverbal	-0.4250	0.1486	233	-2.86	0.0046

CMC — Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	41.8383	0.3628	178	115.32	<.0001
~~hw2~~	~~-0.06701~~	~~0.4367~~	~~176~~	~~-0.15~~	~~0.8782~~
mixgrd	-0.8419	0.5213	196	-1.61	0.1079
CMverb10	3.2181	0.4004	181	8.04	<.0001
WCverb	2.3589	0.07905	123	29.84	<.0001
mixgrd*WCverb	0.3394	0.1578	231	2.15	0.0325
mixgrd*CMverb10	0.2113	0.5549	198	0.38	0.7038
CMverb10*WCverb	-0.04328	0.07779	179	-0.56	0.5787
CMverb10*CMverb10	-0.4250	0.1486	233	-2.86	0.0046

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
Simple L2 context	0.8592	0.4061	192	2.12	0.0357
L2 context*mixgrd	-0.1281	0.5772	231	-0.22	0.8246
L2 context*CMverbal	-0.3817	0.1671	344	-2.28	0.0229

L1 within verbal slope is n.s. less positive (weaker) by **0.0433** per unit school mean verbal
L2 between verbal slope is n.s. less positive by **0.4250** per unit school mean verbal
L2 contextual verbal slope is n.s. less positive by **0.3817** per unit school mean verbal

Prerequisites for Cross-Level Interactions?

- Let's go back to this generic cluster-MC model for a moment:

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(WCx_{pc}) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$

$\beta_{1c} = \gamma_{10} + \gamma_{11}(CMx_c) + U_{1c}$

Can I still include γ_{11} without U_{1c}?

- If the U_{1c} random slope for WCx_{pc} was not initially significant (via $-2\Delta LL$), can I still test cross-level interactions with WCx_{pc}?
    - **"NO"**: If a level-1 slope does not vary randomly over clusters, then it has ~0 variance to predict (so cross-level interactions with that level-1 slope are not necessary); its SE and DDF could be inaccurate SE if $\tau^2_{U_1} > 0$ at all
    - **"YES":** Because power to detect random slope variances 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 L1 fixed slope of WCx. What can happen if we test a L2group*WCx cross-level interaction?

	Non-Significant L2group*WCx effect?	Significant L2group*WCx effect?
Random WCx slope initially **not** significant	Effect of WCx is **FIXED**	Effect of WCx is **systematically varying**
Random WCx initially sig, **not** sig after L2group*WCx	---	Effect of WCx is **systematically varying**
Random WCx initially sig, **still** sig after L2group*WCx	Effect of WCx is **RANDOM**	Effect of WCx is **RANDOM**

The effects of level-1 predictors (person-level) can be fixed, random, or systematically varying. The effects of level-2 predictors (cluster-level) can only be fixed or systematically varying (not random, *at least in the traditional sense that is not creating intercept heteroscedasticity*).

Explained Variance by Fixed Slopes

- **Fixed slopes of level-2 cluster predictors *by themselves*:**
  - L2 BC main effects or interactions reduce L2 random intercept variance
- **Fixed slopes of *cross-level interactions* (level-1 * level-2):**
  - If the **L1 person predictor also has a random slope**, its cross-level interaction will reduce its corresponding **L2 random slope variance**
    - So make sure you test the L2 random slope before any cross-level interactions!
  - If the **L1 person predictor does NOT have a random slope**, its cross-level interaction will reduce the **L1 residual variance** instead
    - This condition creates a "systematically varying" L1 slope instead, in which the slope varies only by interacting predictors (but not randomly otherwise)
- **Fixed slopes of level-1 person predictors *without L2 variance*:**
  - L1 WC main effects or interactions reduce L1 residual variance
- **Fixed slopes of level-1 person predictors *with L2 variance*:**
  - L1 WC main effects or interactions can reduce both L1 residual variance and L2 random intercept variance; need to add corresponding L2 main effects, L2 interactions, or cross-level interactions in order to prevent smushing!
    - See Hoffman & Walters (2022) and Hoffman (2019) for elaboration

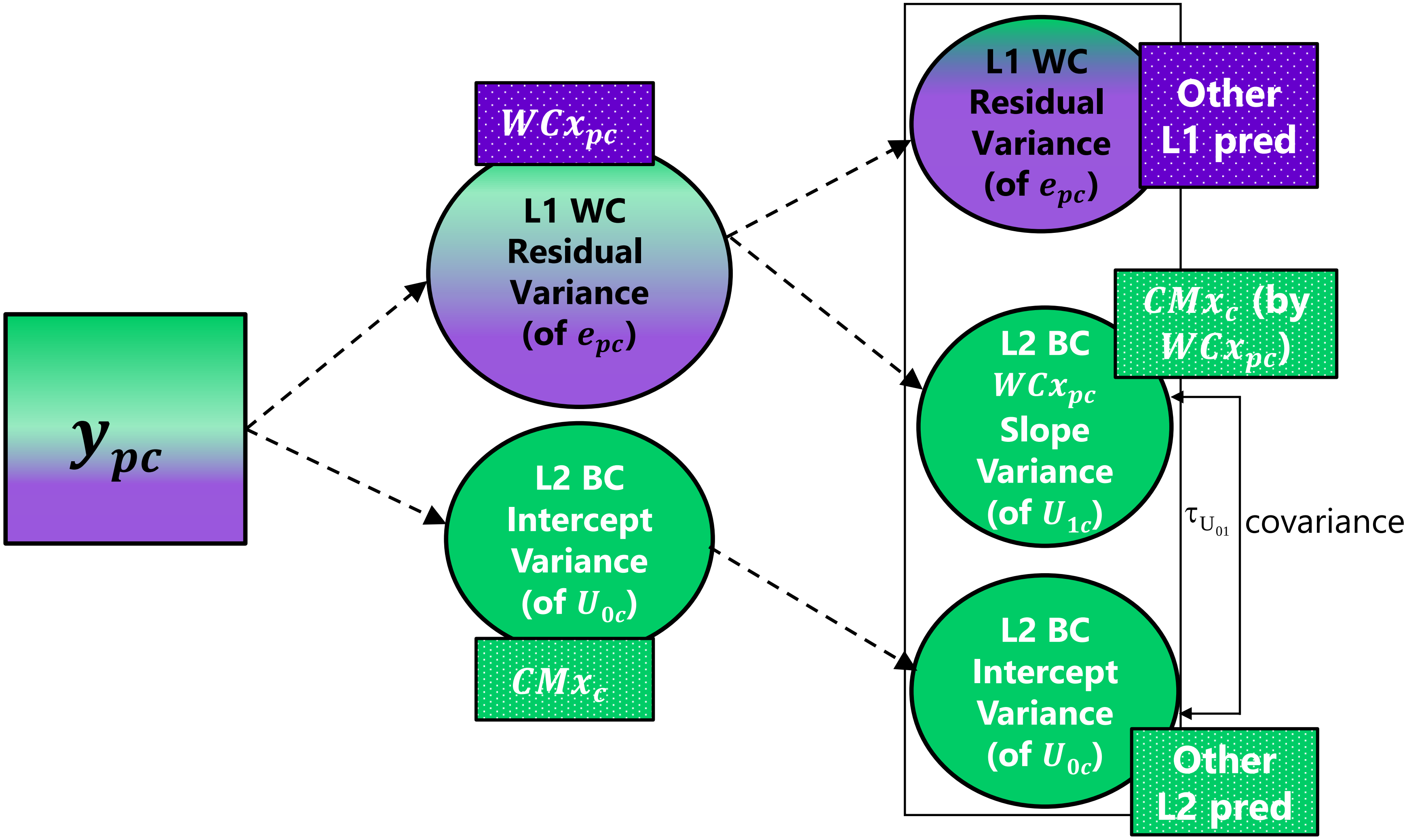
Intermediate Summary: Part 2

- A **level-2 random slope** variance allows cluster differences in the within-cluster effect of a L1 person predictor
  - Should be specified to multiply the cluster-MC or latent-centered version of the L1 predictor, otherwise the random slope will be a new kind of smushed!
  - Implies **quadratic heterogeneity of variance** and covariance across the within part of the L1 predictor (and L2 mean part if random slope multiplies both)
  - Implies **another way that clusters differ from each other** (to be explained by **cross-level interactions** between that L1 predictor and L2 predictors)

- **Meaning of cross-level interactions vary by type of level-1 predictor:**
  - Cluster-MC: WCx*L2z → L1 within x slope only moderated by L2z
  - Constant-C: L1x*L2z only → L1 within x slope AND L2 between x slope moderated by L2z the same (smushed)
- **After adding the corresponding L2 interaction of CMx*L2z:**
  - Cluster-MC: CMx*L2z → How L2 between x slope is moderated by L2z (was 0)
  - Constant-C: CMx*L2z → How L2 contextual x slope is moderated by L2z (was 0); also difference in moderation of L1 within x slope and L2 between x slope by L2z

How MLM "Handles" Dependency

- How does MLM "handle" dependency? By forming a new random effect variance component (or "pile" of variance) for each source of dependency



Model-Implied Variance and Covariance

- So far we've only used scalar equations to describe how the model predicts each person's outcome, but to understand the model-implied pattern of variance and covariance across persons and clusters, we need to show the model using matrices instead!

- Example cluster-MC model with a random intercept only:

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(WCx_{pc}) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_{c}) + U_{0c}$

$\beta_{1c} = \gamma_{10}$

Composite: $y_{pc} = \gamma_{00} + \gamma_{01}(CMx_{c}) + \gamma_{10}(WCx_{pc}) + U_{0c} + e_{pc}$

Matrices per Cluster: $\mathbf{Y}_c = \mathbf{X}_c\boldsymbol{\gamma} + \mathbf{Z}_c\mathbf{U}_c + \mathbf{E}_c$

Btw—this equation is where the terms "**columns in X**" and "**columns in Z**" on the SAS MIXED output come from

Example Model for $L1n = 4$ in One Cluster

Random Int Model: $\gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{pc}) + U_{0c} + e_{pc}$

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

$$\begin{bmatrix} y_{1c} \\ y_{2c} \\ y_{3c} \\ y_{4c} \end{bmatrix} = \begin{bmatrix} 1 & CMx_c & WCx_{1c} \\ 1 & CMx_c & WCx_{2c} \\ 1 & CMx_c & WCx_{3c} \\ 1 & CMx_c & WCx_{4c} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} U_{0c} \end{bmatrix} + \begin{bmatrix} e_{1c} \\ e_{2c} \\ e_{3c} \\ e_{4c} \end{bmatrix}$$

$$\begin{bmatrix} y_{1c} \\ y_{2c} \\ y_{3c} \\ y_{4c} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{1c}) \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{2c}) \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{3c}) \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{4c}) \end{bmatrix} + \begin{bmatrix} U_{0c} \\ U_{0c} \\ U_{0c} \\ U_{0c} \end{bmatrix} + \begin{bmatrix} e_{1c} \\ e_{2c} \\ e_{3c} \\ e_{4c} \end{bmatrix}$$

$$\begin{bmatrix} y_{1c} \\ y_{2c} \\ y_{3c} \\ y_{4c} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{1c}) + U_{0c} + e_{1c} \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{2c}) + U_{0c} + e_{2c} \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{3c}) + U_{0c} + e_{3c} \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{4c}) + U_{0c} + e_{4c} \end{bmatrix}$$

$\mathbf{X}_c$ = $L1n \times k$ values of **predictors with fixed effects**, so can differ by cluster ($k = 3$ here)

$\boldsymbol{\gamma}$ = $k \times 1$ estimated **fixed effects** → same for all clusters ($k = 3$ here)

$\mathbf{Z}_c$ = $L1n \times u$ values of **predictors with random effects**, so can differ by cluster ($u = 1$ here)

$\mathbf{U}_c$ = $u \times 1$ estimated cluster-specific **random effects** (here, just U_{0c})

$\mathbf{E}_c$ = $L1n \times L1n$ person-specific cluster residuals

Same Random Intercept Model: Predicted Marginal Variance–Covariance V Matrix per Cluster

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

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

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

$\mathbf{Z}_c$ = $L1n \times u$ values of **predictors with random effects**, so can differ by cluster ($u = 1$ here)

$\mathbf{Z}_c^T$ = $\mathbf{Z}_c$ transposed → on its side

$\mathbf{G}_c$ = $u \times u$ estimated **random effects variances and covariances**, so will be the same for all clusters (here, just $\tau_{U_0}^2$ = intercept variance)

$\mathbf{R}_c$ = $L1n \times L1n$ **person residual variances and covariances**, so will be same for all clusters (here, same σ_e^2 on the diagonal because persons are exchangeable; all 0 values on the off-diagonals because persons are conditionally independent)

Adding a Random Slope Implies…

- **Clusters differ** from each other randomly **in TWO ways**—in intercept (U_{0c}) and the slope of a person predictor (U_{1c}), which implies **TWO kinds of between-cluster variance**, which translates to **TWO sources of cluster dependency** → TWO reasons for the correlation of outcomes from persons in the same cluster

- Example cluster-MC model adding a **random slope** for WCx_{pc}:

Level-1: $y_{pc} = \beta_{0c} + \beta_{1c}(WCx_{pc}) + e_{pc}$

Level-2: $\beta_{0c} = \gamma_{00} + \gamma_{01}(CMx_c) + U_{0c}$
$\beta_{1c} = \gamma_{10} + U_{1c}$

Composite: $\gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{pc}) + U_{0c} + U_{1c}(WCx_{pc}) + e_{pc}$

Matrices per Cluster: $\mathbf{Y}_c = \mathbf{X}_c\boldsymbol{\gamma} + \mathbf{Z}_c\mathbf{U}_c + \mathbf{E}_c$

Btw—this equation is where the terms "**columns in X**" and "**columns in Z**" on the SAS MIXED output come from

Example Model for $L1n = 4$ in One Cluster

Random Slope Model: $\gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{pc}) + U_{0c} + U_{1c}(WCx_{pc}) + e_{pc}$

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

$$\begin{bmatrix} y_{1c} \\ y_{2c} \\ y_{3c} \\ y_{4c} \end{bmatrix} = \begin{bmatrix} 1 & CMx_c & WCx_{1c} \\ 1 & CMx_c & WCx_{2c} \\ 1 & CMx_c & WCx_{3c} \\ 1 & CMx_c & WCx_{4c} \end{bmatrix} \begin{bmatrix} \gamma_{00} \\ \gamma_{01} \\ \gamma_{10} \end{bmatrix} + \begin{bmatrix} 1 & WCx_{1c} \\ 1 & WCx_{2c} \\ 1 & WCx_{3c} \\ 1 & WCx_{4c} \end{bmatrix} \begin{bmatrix} U_{0c} \\ U_{1c} \end{bmatrix} + \begin{bmatrix} e_{1c} \\ e_{2c} \\ e_{3c} \\ e_{4c} \end{bmatrix}$$

$$\begin{bmatrix} y_{1c} \\ y_{2c} \\ y_{3c} \\ y_{4c} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{1c}) \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{2c}) \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{3c}) \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{4c}) \end{bmatrix} + \begin{bmatrix} U_{0c} + U_{1c}(WCx_{1c}) \\ U_{0c} + U_{2c}(WCx_{2c}) \\ U_{0c} + U_{3c}(WCx_{3c}) \\ U_{0c} + U_{4c}(WCx_{4c}) \end{bmatrix} + \begin{bmatrix} e_{1c} \\ e_{2c} \\ e_{3c} \\ e_{4c} \end{bmatrix}$$

$$\begin{bmatrix} y_{1c} \\ y_{2c} \\ y_{3c} \\ y_{4c} \end{bmatrix} = \begin{bmatrix} \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{1c}) + U_{0c} + U_{1c}(WCx_{1c}) + e_{1c} \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{2c}) + U_{0c} + U_{2c}(WCx_{2c}) + e_{2c} \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{3c}) + U_{0c} + U_{3c}(WCx_{3c}) + e_{3c} \\ \gamma_{00} + \gamma_{01}(CMx_c) + \gamma_{10}(WCx_{4c}) + U_{0c} + U_{4c}(WCx_{4c}) + e_{4c} \end{bmatrix}$$

$\mathbf{X}_c$ = $L1n \times k$ values of **predictors with fixed effects**, so can differ by cluster ($k = 3$ here)

$\boldsymbol{\gamma}$ = $k \times 1$ estimated **fixed effects** → same for all clusters ($k = 3$ here)

$\mathbf{Z}_c$ = $L1n \times u$ values of **predictors with random effects**, so can differ by cluster ($u = 2$ here)

$\mathbf{U}_c$ = $u \times 2$ estimated cluster-specific **random effects** (here, U_{0c} and U_{1c})

$\mathbf{E}_c$ = $L1n \times L1n$ person-specific cluster residuals

Same Random Slope Model: Predicted Marginal Variance–Covariance **V** Matrix per Cluster

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

$$\mathbf{V}_c = \begin{bmatrix} 1 & WCx_{1c} \\ 1 & WCx_{2c} \\ 1 & WCx_{3c} \\ 1 & WCx_{4c} \end{bmatrix} \begin{bmatrix} \tau^2_{U_0} & \tau_{U_{01}} \\ \tau_{U_{01}} & \tau^2_{U_1} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ WCx_{1c} & WCx_{2c} & WCx_{3c} & WCx_{4c} \end{bmatrix} + \begin{bmatrix} \sigma^2_e & 0 & 0 & 0 \\ 0 & \sigma^2_e & 0 & 0 \\ 0 & 0 & \sigma^2_e & 0 \\ 0 & 0 & 0 & \sigma^2_e \end{bmatrix}$$

$\mathbf{V}_c$ matrix = complicated, but summarized below

$\mathbf{V}_c$ matrix: Marginal Variance at a Given $[WCx]$

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

$\mathbf{V}_c$ matrix: Marginal Covariance at a Given $[WCx_A, WCx_B]$

$$= \tau^2_{U_0} + \left[\left(WCx_A + WCx_B\right)\tau_{U_{01}}\right] + \left[\left(WCx_A * WCx_B\right)\tau^2_{U_1}\right]$$

$\mathbf{Z}_c$ = $L1n \times u$ values of **predictors with random effects**, so can differ by cluster (u = 2 here)

$\mathbf{Z}_c^T$ = $\mathbf{Z}_c$ transposed

$\mathbf{G}_c$ = $u \times u$ estimated **random effects variances and covariances**, so will be same for all clusters (here, $\boldsymbol{\tau}^2_{U_0}$, $\boldsymbol{\tau}^2_{U_1}$ and $\boldsymbol{\tau}_{U_{01}}$)

$\mathbf{R}_c$ = $L1n \times L1n$ **person residual variances and covariances**, so will be same for all clusters (same $\boldsymbol{\sigma}^2_e$ on the diagonal and 0 values on off-diagonals)

Building a Combined **V** across Clusters: Same Random Slope Model

V for two clusters, both of size $L1n = 4$:

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

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

- The combined **V** matrix across all clusters is used in estimation
- It has a "**block diagonal**" structure → predictions are given for each cluster, but 0 values are given for the elements that describe relationships across clusters (because clusters are supposed to be independent in a two-level model!)

Building a Combined V across Clusters: Same Random Slope Model

V for a cluster with $L1n = 4$ and a cluster with $L1n = 3$:

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

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

- Take home message: Partitioning variance into piles…
  - **Level 2 = Between →** **G** matrix of random effects variances/covariances
  - **Level 1 = Within →** **R** matrix of residual variances/covariances
  - **G** and **R** combine via **Z** into **V** matrix of marginal variances/covariances
  - These flexible options allow the outcome variances and covariances to vary in a predictor-dependent way to better match the actual data

Two Sides of Any Model: Estimation

- **Fixed Effects in the Model for the Means:**
  - How the expected outcome for a given observation varies as a function of values on *known* predictor variables
  - Fixed effects parameters do NOT need to be solved for iteratively in (residual) maximum likelihood estimation for *general* MLMs

- **Random Effects in the Model for the Variance:**
  - How model residuals are related across observations (dependency across persons, clusters, time, etc)—*unknown* things due to sampling
  - Random effects variances and covariances can predict complex patterns of variance and covariance among the outcome residuals
  - Anything besides level-1 residual variance $\boldsymbol{\sigma_e^2}$ must be solved for iteratively—this increases the dimensionality of estimation process
  - Estimation utilizes the predicted $\mathbf{V}$ matrix for each cluster
  - In what follows, $\mathbf{V}$ will be based on the ***previous random slope model***

End Goals of Maximum Likelihood Estimation

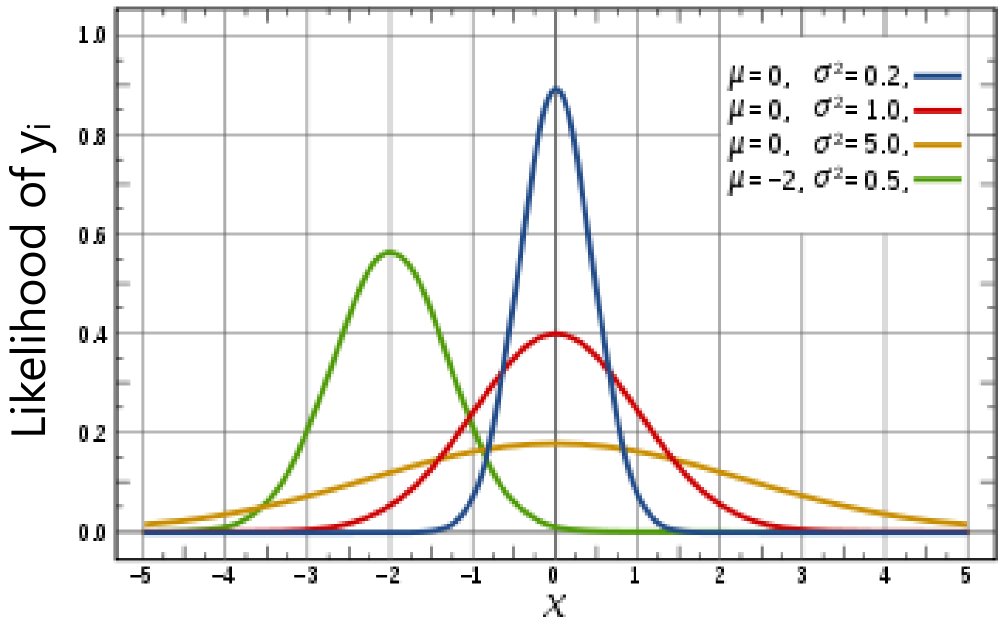
1. Obtain "most likely" values for each unknown model parameter (random effects variances and covariances, residual variances and covariances, which then are used to calculate the fixed effects) → **the estimates**

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

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

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

Univariate Normal Probability Distribution Function



Univariate Normal PDF (two ways):

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

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

- This PDF tells us how **likely** (i.e., **tall**) any value of y_i is given two things:
  - Conditional mean $\hat{y}_i$
  - Residual variance σ_e^2

- We can see this work using the NORMDIST function in excel!
  - Easiest for **empty** model:

$$y_i = \beta_0 + e_i$$

$$\hat{y}_i = \beta_0$$

From Univariate to Multivariate Normal: Joint Height for All $L1n$ Outcomes for Cluster c

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

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

- In our example random slope model, three are **three fixed effects** (in γ) that predict the $\mathbf{Y}_c$ outcomes: intercept $\boldsymbol{\gamma_{00}}$, L2 slope $\boldsymbol{\gamma_{01}}$, and L1 slope $\boldsymbol{\gamma_{10}}$
- Model also gives us $\mathbf{V}_c$ → the model-predicted marginal variance and covariance matrix across persons, taking into account their $\boldsymbol{WPx_{pc}}$ values
- Uses $|\mathbf{V}_c|$ = determinant of $\mathbf{V}_c$ = summary of *non-redundant* info in $\mathbf{V}_c$

- $(\mathbf{V}_c)^{-1}$ → matrix inverse → analogous to dividing (so can't be 0 or negative)
  - $(\mathbf{V}_c)^{-1}$ must be "positive definite", which in practice means no 0 random variances or covariances that cause out-of-bound correlations between random effects
  - Otherwise, program uses "generalized inverse" → questionable results

Now Try Some Possible Answers...
(e.g., for the 4 parameters in example random slope model)

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

$$L = \prod_{c=1}^{L2n} \left\{ (2\pi)^{-n/2} * \left|\mathbf{V}_c\right|^{-1/2} * \exp\left[-\frac{1}{2} (\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\gamma})^T (\mathbf{V}_c)^{-1} (\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\gamma}) \right] \right\}$$

$$LL = \sum_{c=1}^{L2n} \left\{ \left[-\frac{n}{2} \log(2\pi) \right] + \left[-\frac{1}{2} \log\left|\mathbf{V}_c\right| \right] + \left[-\frac{1}{2} (\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\gamma})^T (\mathbf{V}_c)^{-1} (\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\gamma}) \right] \right\}$$

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

End Goals 1 and 2: Model Estimates and SEs

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

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

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

What about the Fixed Effects?

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

$$\gamma = \left\{ \sum_{c=1}^{L2n} \left(\mathbf{X}_c^T \mathbf{V}_c^{-1} \mathbf{X}_c \right) \right\}^{-1} \sum_{c=1}^{L2n} \left(\mathbf{X}_c^T \mathbf{V}_c^{-1} \mathbf{Y}_c \right), \qquad \text{Cov}(\gamma) = \left\{ \sum_{c=1}^{L2n} \left(\mathbf{X}_c^T \mathbf{V}_c^{-1} \mathbf{X}_c \right) \right\}^{-1}$$

All we need is $\mathbf{V_c}$ and the data: $\mathbf{X_c}$, $\mathbf{Y_c}$

γ = fixed effect estimates

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

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

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

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

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

What about ML vs. REML?

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

- What does this mean? Remember "population" vs. "sample" formulas for computing variance?

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

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

- Same idea: ML estimates of random effects variances will be downwardly biased by a factor of $(L2n - k)$ / $L2n$, where k = #fixed effects… it just looks way more complicated

What about ML vs. REML? (N = # obs)

$$\text{ML:}\quad LL = \left[-\frac{N-0}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{c=1}^{L2n}\log\left|\mathbf{V}_c\right|\right] + \left[-\frac{1}{2}\sum_{c=1}^{L2n}\left(\mathbf{Y}_c-\mathbf{X}_c\boldsymbol{\gamma}\right)^T\mathbf{V}_c^{-1}\left(\mathbf{Y}_c-\mathbf{X}_c\boldsymbol{\gamma}\right)\right]$$

$$\text{REML:}\ LL = \left[-\frac{N-k}{2}\log(2\pi)\right] + \left[-\frac{1}{2}\sum_{c=1}^{L2n}\log\left|\mathbf{V}_c\right|\right] + \left[-\frac{1}{2}\sum_{c=1}^{L2n}\left(\mathbf{Y}_c-\mathbf{X}_c\boldsymbol{\gamma}\right)^T\mathbf{V}_c^{-1}\left(\mathbf{Y}_g-\mathbf{X}_c\boldsymbol{\gamma}\right)\right]$$

$$+\left[-\frac{1}{2}\log\left|\sum_{c=1}^{L2n}\mathbf{X}_c^T\mathbf{V}_c^{-1}\mathbf{X}_c\right|\right]$$

$$\text{where:}\quad \left[-\frac{1}{2}\log\left|\sum_{c=1}^{L2n}\mathbf{X}_c^T\mathbf{V}_c^{-1}\mathbf{X}_c\right|\right] = \left[\frac{1}{2}\log\left|\left(\sum_{c=1}^{L2n}\mathbf{X}_c^T\mathbf{V}_c^{-1}\mathbf{X}_c\right)^{-1}\right|\right] = \left[\frac{1}{2}\log\left|\mathrm{Cov}(\boldsymbol{\gamma})\right|\right]$$

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

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

- End up with "best" LL from predicting $\mathbf{V_c}$ → so how good is it?
- Absolute model fit assessment is only possible when the $\mathbf{V_c}$ matrix is organized the same for all L2 units and there are no random slopes
  - If items are treated as fixed, we can get absolute fit in CFA and SEM → χ^2 test is based on match between actual and predicted data matrix
  - No absolute fit provided by default in univariate MLM programs (or in SEM or multilevel SEM when using random slopes), as a saturated model for the answer key of person dependency is not really possible

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

What about testing variances > 0?

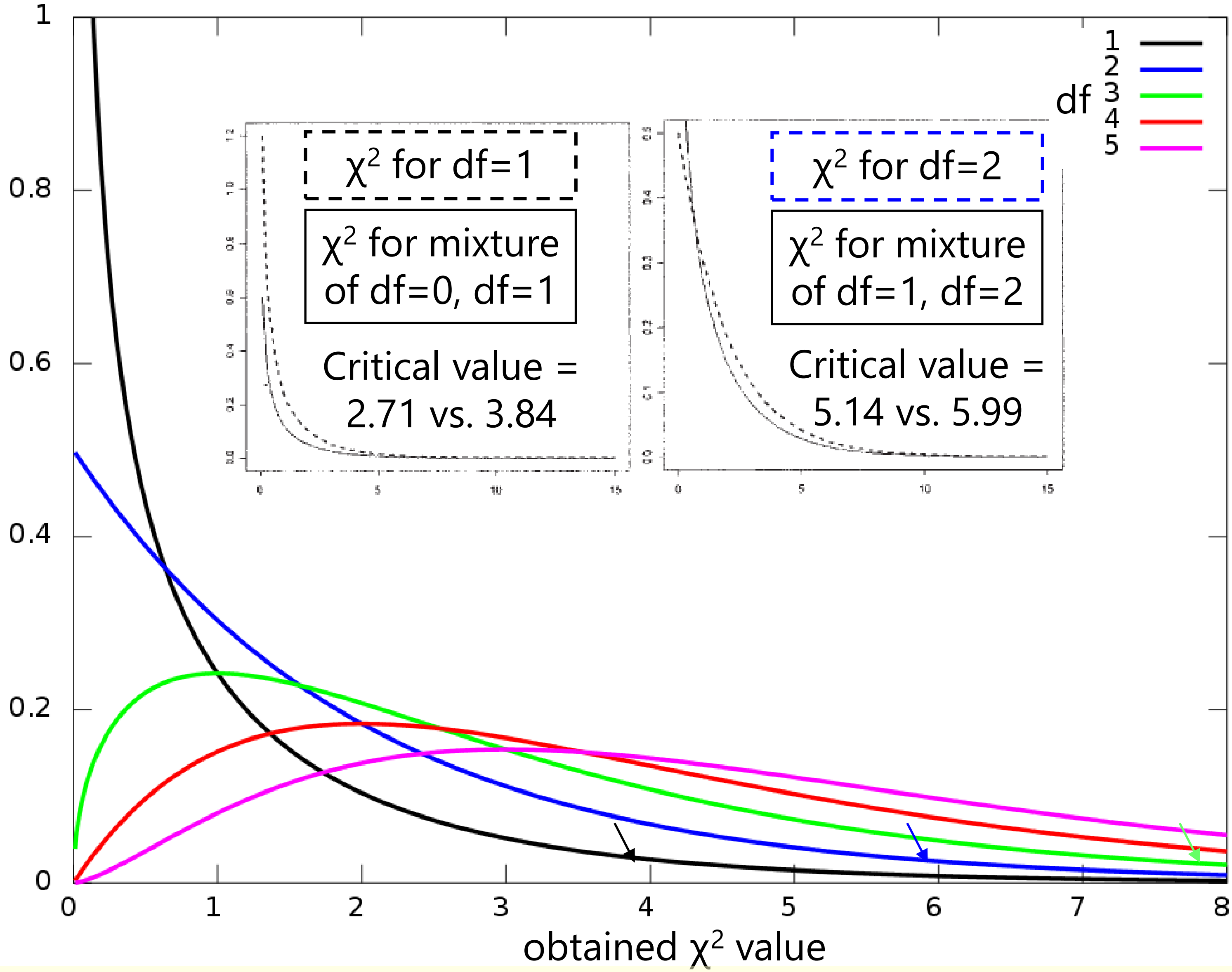
- $-2\Delta LL$ between two nested models is distributed as χ^2 only when added parameters do not have a boundary (like 0 or 1)
  - Is ok for fixed effects using ML (could be any positive or negative value)
  - Is NOT ok for ML or REML tests of random variances (must be > 0)
  - Is ok for ML or REML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

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

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

χ^2 Distributions

small pictures from Stoel et al., 2006



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

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

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

+ Random Intercept
df=1: 2.71 vs. 3.84

+ Random Slope #1
df=2: 5.14 vs. 5.99

+ Random Slope #2
df=3: 7.05 vs. 7.82

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

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

Solutions for Boundary Problems when using −2ΔLL tests

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

$$\text{Mixture } p\text{-value} = 0.5 * \text{prob}(\chi^2_1 > -2\Delta\text{LL}) + 0.5 * \text{prob}(\chi^2_2 > -2\Delta\text{LL})$$

so critical χ^2 = 5.14, not 5.99

- However—using a 50/50 mixture assumes a diagonal information matrix for the random effects variances (i.e., it assumes the values for each are arrived at independently, which is not likely to be true)
- Two options for more complex cases:
  - Simulate data to determine actual mixture for calculating *p*-value
  - Accept that −2ΔLL is conservative in these cases, and use it anyway
    → I use ~ to acknowledge this: e.g., $-2\Delta\text{LL}(\sim 2) > 5.99$, $p < .05$

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

- Level-2 $\mathbf{U}_g$ random effects also require further explanation…
  - Empty two-level model: $y_{pc} = \gamma_{00} + U_{0c} + e_{pc}$
  - U_{0c} values are deviated cluster means, right? Well, not exactly…
- 3 ways of representing size of individual differences in individual intercepts and slopes across level-2 clusters:
  - Get each level-2 unit's OLS intercepts and slopes, save them to a dataset, and calculate their observed variances
  - Estimate variance of the $\mathbf{U}_c$ values (what we do in MLM)
  - Predict $\mathbf{U}_c$ cluster values; calculate their variance (2-stage MLM)
- Expected order of magnitude of variance estimates:
  - OLS variance > MLM variance > Predicted $\mathbf{U}_c$ variance
  - Why are these different? "**Shrinkage**"

What about the U random effects?

- Level-2 unit $\mathbf{U}_c$ values are NOT estimated in the likelihood function
  - $\mathbf{G}$ matrix variances and covariances are sufficient statistics for the estimation process assuming multivariate normality of $\mathbf{U}_c$ values
  - Level-2 $\mathbf{U}_c$ random effects are **predicted** (SOLUTION on SAS RANDOM, pred without xb in STATA, predict in R) as: $\mathbf{U}_c = \mathbf{G}_c \mathbf{Z}_c^T \mathbf{V}_c^{-1} (\mathbf{Y}_c - \mathbf{X}_c \boldsymbol{\gamma})$
    - Which then create cluster estimates as: $\beta_{0c} = \gamma_{00} + U_{0c}$ and $\beta_{1c} = \gamma_{10} + U_{1c}$

- What isn't obvious: the composite $\boldsymbol{\beta}_c$ values are weighted combos of the fixed effects ($\boldsymbol{\gamma}$) and their level-2 OLS estimates ($\boldsymbol{\beta}_{\mathbf{OLS_c}}$) :

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

  - The more "true" variation in intercepts and slopes in the data (in $\mathbf{G}$), the more the $\boldsymbol{\beta}_c$ estimates are based on level-2 unit OLS estimates
  - The more "unexplained" residual variation around the level-2 slopes (in $\mathbf{R}$), the more the fixed effects are heavily weighted instead
    - = **SHRINKAGE** (more so for clusters with fewer persons, too)

What about the U random effects?

- Point of the story: $\mathbf{U_c}$ values are NOT single scores!
  - They are the mean of a distribution of possible values for each person (i.e., as given by the SE for each $\mathbf{U_c}$)
  - These "best estimates" of the $\mathbf{U_c}$ values are shrunken anyway

- Good news: you don't need those $\mathbf{U_c}$ values in the first place!
  - Goal of MLM is to estimate and predict the variance of the $\mathbf{U_c}$ values (in $\mathbf{G}$) with cluster-level characteristics directly in the model
  - If you want your $\mathbf{U_c}$ values to be predictors instead, then you need to estimate your model using multivariate MLM ("M-SEM")
  - You could use the predicted $\mathbf{U_c}$ values to examine potential violations of model assumptions, though…

Estimation: The Grand Finale

- Estimation in MLM is all about finding the most likely estimates for the random effects variances and covariances
  - The more of them there are, the harder it is to find them (the more dimensions of the likelihood mountain there are to scale)
  - "Non-positive-definite" **G** matrix means "broken model" (usually because a variance went to 0 or a correlation went out of bounds)
  - Fixed effects are solved for given **V** in general MLMs, so they rarely cause estimation problems
  - Individual random effects are not model parameters, but can be predicted after-the-fact (but try never to use these as data)

- Estimation comes in two flavors:
  - ML → maximize the data; use $-2\Delta LL$ to compare any nested models
  - REML → maximize the residuals; use $-2\Delta LL$ to compare models that differ in their model for the variance ONLY