Supplemental Material

A Practical Guide to Selecting and Blending Approaches for Clustered Data: Clustered Errors, Multilevel Models, and Fixed Effect Models

Table of Contents

Chapter A

Numerical Demonstration of Clustered Errors

Consider a small hypothetical data set with 3 clusters, each cluster having 3 people (9 people total):

Cluster	Person	X	Y
1	1	0.05	1.32
1	2	0.89	1.23
1	3	0.99	1.72
2	4	-1.43	-1.33
2	5	-1.23	-0.59
2	6	0.38	0.94
3	7	-1.39	0.76
3	8	1.98	0.64
3	9	1.47	0.37

Imagine fitting a simple linear regression model with *y* as the outcome and *x* as the predictor such that

$$y = \beta_0 + \beta_1 x + e \tag{A1}$$

The estimates for β_0 and β_1 could be calculated by $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Because the data are small, this can be written out entirely for completeness:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \left(\begin{bmatrix} 1 & 0.05 \\ 1 & 0.89 \\ 1 & 0.99 \\ 1 & -1.43 \\ 1 & -1.23 \\ 1 & 0.38 \\ 1 & -1.39 \\ 1 & 1.98 \\ 1 & 1.47 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1 & 0.05 \\ 1 & 0.89 \\ 1 & 0.99 \\ 1 & -1.43 \\ 1 & -1.23 \\ 1 & 0.38 \\ 1 & -1.39 \\ 1 & 1.98 \\ 1 & 1.47 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 0.05 \\ 1 & 0.89 \\ 1 & 0.99 \\ 1 & -1.43 \\ 1 & -1.23 \\ 1 & 0.38 \\ 1 & -1.39 \\ 1 & 1.98 \\ 1 & 1.47 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} 1.32 \\ 1.23 \\ 1.72 \\ -1.33 \\ -0.59 \\ 0.94 \\ 0.76 \\ 0.64 \\ 0.37 \end{bmatrix} = \begin{bmatrix} 0.481 \\ 0.429 \end{bmatrix} \tag{A2}$$

The **X** matrix consists of a column of 1s corresponding to the intercept and a column of the values of the predictor variable *x*. Carrying out these matrix computations yields an intercept estimate of 0.481 and a slope estimate of 0.429.

To perform inference on these coefficient estimates to determine if they differ from 0, first the error needs to be computed by subtracting the predicted values of *y* from the observed values of *y*:

$$\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1.32 \\ 1.23 \\ 1.72 \\ -1.33 \\ -0.59 \\ 0.94 \\ 0.76 \\ 0.64 \\ 0.37 \end{bmatrix} - \begin{bmatrix} 0.50 \\ 0.86 \\ 0.91 \\ -0.13 \\ -0.05 \\ 0.64 \\ -0.11 \\ 1.33 \\ 1.11 \end{bmatrix} = \begin{bmatrix} 0.82 \\ 0.37 \\ 0.82 \\ -1.20 \\ 0.54 \\ 0.30 \\ 0.87 \\ -0.69 \\ -0.74 \end{bmatrix} \tag{A3}$$

If assuming independence and homoskedasticity, the sampling covariance matrix of the regression coefficients would be

$$\begin{aligned} Cov(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &\overset{i.i.d}{=} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \tag{A4}$$

Where

$$\sigma^2 = \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{N-p} = \frac{5.076}{9-2} = 0.725 \tag{A5}$$

For completeness, the middle term in the first expression of Equation A4 assuming independence and homoskedasticity would be,

$$\Omega = \begin{bmatrix} 0.725 & & & & & & & & \\ & 0.725 & & & & & & & \\ & & 0.725 & & & & & & \\ & & & 0.725 & & & & & \\ & & & & 0.725 & & & & \\ & & & & & 0.725 & & & \\ & & & & & & 0.725 & & \\ & & & & & & & 0.725 & \\ & & & & & & & & 0.725 \end{bmatrix} \tag{A6}$$

This means that the sampling covariance matrix is

$$Cov(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 0.0825 & -0.0105 \\ -0.0105 & 0.0551 \end{bmatrix} \tag{A7}$$

The standard errors are then the square root of the diagonal elements of the matrix in Equation A7, so the standard error of the intercept is $\sqrt{0.08256} = 0.287$ and the standard error of the slope is $\sqrt{0.0551} = 0.235$.

To verify this result in R,

```
> summary(lm(Y ~X, data=data))
Call:
lm(formula = Y ~ X, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-1.1980 -0.6893  0.2964  0.8150  0.8748

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.4808     0.2873   1.673    0.138
X             0.4285     0.2347   1.826    0.111
```

Now, let's say I want to use clustered errors instead to yield standard errors that more accurately account for the dependence among observations. The more general formula for the sampling variability in the regression coefficients is $Cov(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ and clustered errors use $\boldsymbol{\Omega} = \sum_{j=1}^{J}\hat{\mathbf{e}}_j\hat{\mathbf{e}}_j'$. The matrix notation can make this difficult to understand, so let's walkthrough what this calculation looks like for these example data.

First, note that there is a summation sign in $\boldsymbol{\Omega} = \sum_{j=1}^{J}\hat{\mathbf{e}}_j\hat{\mathbf{e}}_j'$ and that it goes from 1 to J where J is the number of clusters (J=3 in these data). That means that the calculation will be repeated for each cluster. Note that this process requires that the cluster affiliations and cluster ID are known and included in the data; the method is not designed nor capable of identifying latent or unknown cluster within the data.

Beginning with j =1, $\hat{\mathbf{e}}_j$ corresponds to the observations belonging to Cluster 1 which would be the first 3 rows of the result in Equation A4. Using the error vector calculated in Equation A3, for j =1, this calculation would be

$$\Omega_{j=1} = \begin{bmatrix} 0.82 \\ 0.37 \\ 0.82 \end{bmatrix} \begin{bmatrix} 0.82 & 0.37 & 0.82 \end{bmatrix} = \begin{bmatrix} 0.669 & 0.301 & 0.666 \\ 0.301 & 0.135 & 0.300 \\ 0.666 & 0.300 & 0.644 \end{bmatrix} \quad \text{(A8)}$$

Repeating the same calculation but for j =2 corresponding to all observations in Cluster 2 (rows 4 through 6 of Equation A4):

$$\begin{aligned}\Omega_{j=2} &= \begin{bmatrix} -1.20 \\ 0.54 \\ 0.30 \end{bmatrix} \begin{bmatrix} -1.20 & 0.54 & 0.30 \end{bmatrix} \\ &= \begin{bmatrix} 1.435 & 0.651 & -0.355 \\ 0.651 & 0.296 & -0.161 \\ -0.355 & -0.161 & 0.088 \end{bmatrix}\end{aligned} \tag{A9}$$

And repeating the same calculation when j =3, corresponding the observations in Cluster 3 occupying rows 7 through 9 of Equation A4:

$$\begin{aligned}\Omega_{j=3} &= \begin{bmatrix} 0.87 \\ -0.69 \\ -0.74 \end{bmatrix} \begin{bmatrix} 0.87 & -0.69 & -0.74 \end{bmatrix} \\ &= \begin{bmatrix} 0.765 & -0.603 & -0.648 \\ -0.603 & 0.475 & 0.511 \\ -0.648 & 0.511 & 0.549 \end{bmatrix}\end{aligned} \tag{A10}$$

To form the full $\mathbf{\Omega}$ matrix, the results in Equation A8 through A10 are compiled into a block diagonal matrix such that

$$\mathbf{\Omega} = \begin{bmatrix} 0.669 & 0.301 & 0.666 & & & & & & \\ 0.301 & 0.135 & 0.300 & & & & & & \\ 0.666 & 0.300 & 0.644 & & & & & & \\ & & & 1.435 & 0.651 & -0.355 & & & \\ & & & 0.651 & 0.296 & -0.161 & & & \\ & & & -0.355 & -0.161 & 0.088 & & & \\ & & & & & & 0.765 & -0.603 & -0.648 \\ & & & & & & -0.603 & 0.475 & 0.511 \\ & & & & & & -0.648 & 0.511 & 0.549 \end{bmatrix} \tag{A11}$$

Compare Equation A11 (which allows observations within the same cluster to covary) to Equation A6 (which assumes independence). Some of the off-diagonal terms in Equation A11 are populated to reflect clustering. Also, note that the diagonal terms are all different, which relaxes homoskedasticity unlike Equation A6 which assumes that the diagonal terms are constant.

Equation A11 is then substituted into $Cov(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, which yields

$$Cov(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} 0.0848 & -0.0267 \\ -0.0267 & 0.1212 \end{bmatrix} \tag{A12}$$

Notice that the values in Equation A12 are larger than the corresponding values in Equation A7 because substituting Equation A10 into the standard error formula accounts for dependence among observations. From Equation A12, the standard errors can be obtained by taking the square root of the diagonal elements. The standard error of the intercept with clustered errors is $\sqrt{0.0848} = 0.291$ and the standard error of the slope with clustered errors is $\sqrt{0.1212} = 0.348$.

The `clubSandwich` package in R can verify this result:

```
> library(clubSandwich)
> m1<-lm(Y ~X, data=data)
> coef_test(m1, vcov="CR0",cluster=data$Cluster)

        Coef. Estimate    SE t-stat d.f. p-val (Satt) Sig.
1 (Intercept)    0.481 0.291   1.65 1.99        0.241
2           X    0.429 0.348   1.23 1.43        0.383
```

SAS code for manually computing this result from scratch in `PROC IML` is provided in the supplemental material

Chapter B

Model Building for Multilevel Model Example

The main text near Equation 3 presented a final multilevel model but did not show the model building steps that led up to that model. The final model was the final step of a 3-step process, so this chapter shows the first two steps that preceded the model in the main text.

Step 1: Unconditional Random Intercept Model

The first step is an unconditional random intercepts model that partitions the variance in the outcome to determine the amount of variability that is attributable to student characteristics and to school characteristics (i.e., to calculate the ICC). This does not explain any variance attributable to either source, it will merely report at which level the unexplained variance resides (i.e., the variance is partition through including an error term at each level; error terms represent unexplained variance). This model can be written in hierarchical notation as,

$$\begin{aligned}&\text{Within-School Submodel} \begin{cases} Math_{ij} = \beta_{0j} + r_{ij} \\ r_{ij} \sim N\left(0, \sigma^2\right) \end{cases} \\ &\text{Between-School Submodel} \begin{cases} \beta_{0j} = \gamma_{00} + u_{0j} \\ u_{0j} \sim N\left(0, \tau_{00}\right) \end{cases}\end{aligned} \tag{B1}$$

where β_{0j} is the school-specific intercept of math scores in school *j*, γ_{00} is the fixed effect of the intercept representing the average math score across all schools, u_{0j} is the difference in the school-specific intercept and the intercept fixed effect for school *j* (the school-level error) which follows a normal distribution $u_{0j} \sim N\left(0, \tau_{00}\right)$, and r_{ij} is the deviation of student *i*'s math score from the school-specific intercept (the student-level error), which also follows a normal distribution $r_{ij} \sim N\left(0, \sigma^2\right)$. The heterogeneity in the intercept can be seen by the β_{0j} term having a *j* subscript, indicating that its value varies across the *j* schools. Because the coefficient in the within-cluster submodel has variance, it then becomes an outcome in the between-cluster submodel so that its variability can be modeled. This is what makes the model "multilevel" or "hierarchical", parameters have variability, so coefficients in one equation simultaneously serve out outcomes in another equation.

The estimates from this model as fit in SAS `PROC MIXED` are in the left column of Table B1, which shows that 82% (39.15) of the variance in Math Scores is attributable to student characteristics and 18% (8.61) of the variance in math scores is attributable to school characteristics. The 18% value is synonymous with the ICC (calculated by $\tau_{00} / \left(\tau_{00} + \sigma^2\right)$) and shows that there is sufficient variability attributable to school sources to warrant building a submodel at that level. Again, note that this is different from the interpretation of the DEFT, which helps determine whether the standard errors are affected by clustering. A low intraclass correlation itself does not necessarily imply that clustering is ignorable.

Step 2: Disaggregated Random Coefficients Model

Next, student characteristics (non-White identity and SES) are added as predictors in the within-school submodel to assess the average achievement gap across schools and whether achievement gaps are heterogeneous across schools. Both predictors are cluster-mean centered such that 0 means the student is at their school's mean of the predictor. This is not the same as standardization because the variance of the predictor is unaffected; centering affects the location of the mean but has no impact on dispersion.

To fully disaggregate the predictors into within-school and between-school effects (i.e., to avoid conflated coefficients for SES and non-White Identity), the cluster means of each student-level predictor are also included as predictors of the intercept coefficient in the β_{0j} equation. This model can be referred to as a *disaggregated random coefficients model*[1] and can be written as

$$\text{Within-School Submodel} \begin{cases} Math_{ij} = \beta_{0j} + \beta_{1j} SES_{ij}^{(CMC)} + \beta_{2j} \textit{Non-White}_{ij}^{(CMC)} + r_{ij} \\ r_{ij} \sim N\left(0, \sigma^2\right) \end{cases}$$

$$\text{Between-School Submodel} \begin{cases} \beta_{0j} = \gamma_{00} + \gamma_{01} \overline{SES}_j + \gamma_{02} \overline{\textit{Non-White}}_j^{(GMC)} + u_{0j} \\ \beta_{1j} = \gamma_{10} + u_{1j} \\ \beta_{2j} = \gamma_{20} + u_{2j} \\ \mathbf{u}_j \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{00} & & \\ \tau_{10} & \tau_{11} & \\ \tau_{20} & \tau_{21} & \tau_{22} \end{bmatrix} \right) \end{cases} \tag{B2}$$

where the β_j terms are random coefficients that vary across schools, $SES_{ij}^{(CMC)} = \left(SES_{ij} - \overline{SES}_j\right)$, and $\textit{Non-White}_{ij}^{(CMC)} = \left(\textit{Non-White}_{ij} - \overline{\textit{Non-White}}_j\right)$ where "CMC" is shorthand for "cluster-mean centered". This model has two additional regression coefficients (β_{1j} and β_{2j}) for the two within-cluster predictors representing student characteristics. Therefore, β_{1j} and β_{2j} become outcomes in the between-school submodel to express that the coefficients are not single values but instead are modeled as a distribution of values. The *u* terms are the between-cluster errors (a.k.a. the random

[1] Traditionally, the random coefficients model is written without disaggregating within-school predictors (e.g., Raudenbush & Bryk, 2002, p. 26). The disaggregated version is used here to avoid tacitly endorsing possibly conflated coefficients in a multilevel model.

effects) that capture the deviation in the effect of cluster *j* from the fixed effect (γ), where the fixed effect represents the average effect across all clusters.

Because there are multiple random effects in the model, the distributional assumption on the *u* terms is now a *multivariate* normal distribution that has a mean *vector* and a covariance *matrix*. The mean vector is comprised of three zeroes, one for each random effect (i.e., the average random effect value is 0, meaning that the average school's effect is equal to the fixed effect). The τ terms on the diagonal of the covariance matrix are the random effect variances, which capture the heterogeneity in each regression coefficient (β). τ_{00} corresponds to heterogeneity in the intercept (similar to the previous section), τ_{11} captures heterogeneity in the SES achievement gap (β_{1j}), and τ_{22} captures the heterogeneity in the non-White achievement gap (β_{2j}). The off-diagonal terms of this covariance matrix estimate systematic relations between the school-specific coefficients for different effects (i.e., the random effect covariances). For instance, if τ_{21} were positive, that would mean that schools with higher value of β_{1j} tend to have higher values of β_{2j} (i.e., different types of achievement gaps are systematically related within the same school).

The school means of non-White identity (i.e., the proportion of students identifying as non-White in the school) and SES are added as predictors of the intercept to fully disaggregate effects of these predictors. The school means are calculated by aggregating the student-level data.To maintain interpretability, the school mean of non-White Identity is further grand-mean centered (GMC) such that $\overline{Non\text{-}White}_j^{(GMC)} = \overline{Non\text{-}White}_j - \overline{Non\text{-}White}$ to improve interpretation such that '0' refers to a school with the sample *average* amount of students identifying as non-White (27.5% in this sample) rather than *zero* students identifying as non-White as when the school of non-White identity is included in its uncentered form.

Including the cluster mean as a school-level predictor allows effects for a one-unit change at different levels to be separately estimated because there are two coefficients associated with each student-level predictor. For instance, both γ_{10} and γ_{01} in Equation B2 capture effects of SES; the former is the within effect of SES (the effect of a one-unit change for the student) and the latter is the between effect of SES (the effect for a one-unit change in the school).

Estimates from this model as fit in SAS `PROC MIXED` with restricted maximum likelihood and Satterthwaite degrees of freedom are shown in the right column of Table B1. The fixed effect coefficients represent the average across all clusters, so, on average, the non-White achievement gap is 2.93 points (students identifying as non-White score about 3 points lower than students identifying as White; $t(105) = -11.14, p < .01$). The achievement gap between students one-unit apart on SES is 1.93 points, with higher SES students tending to have higher scores ($t(151) = 15.84, p < .01$).

The within-school residual variance in this model decreases from 39.15 to 35.66 because student-level predictors are accounting for within-school variance. Using traditional variance reduction formulas (e.g., Hox et al., 2017; Raudenbush & Bryk, 2002; sometimes referred to as pseudo-R^2, Rights & Sterba, 2020), the interpretation would be that non-White identity and SES account for about $(39.15 - 35.66)/39.15 = 8.9\%$ of student-level sources of variance in Math Scores. This corresponds to the $R_w^{2(f_1v)}$ definition of variance explained in Rights and Sterba (2019) as calculated in the `r2mlm` R package (Shaw et al., 2022). Of this variance reduction, 85% was attributable to the fixed effects ($R_w^{2(f_1)} = 7.5\%$) and 15% was attributable to variance being reassigned to the random slopes ($R_w^{2(v)} = 1.3\%$).[2]

The between effects are represented by the coefficients associated with the school means, which are school characteristics (i.e., these values only change across schools, but are constant for all students within the same school). A one-unit change in the school's average SES is associated with higher scores ($\gamma = 5.18, t(150) = 13.13, p < .01$) whereas a one-unit change in the school's average proportion of students identifying as non-White (i.e., an all-non-White versus an all-white school) is associated with lower scores ($\gamma = -2.09, t(148) = -3.87, p < .01$).

After including these predictors in the between-school submodel, the unexplained intercept variance decreased from 8.61 to 2.58. Using standard variance reduction formulas, the school means of SES and non-White identity account for $(8.61 - 2.58)/8.61 = 70.0\%$ of the between-school intercept variance, meaning that the socioeconomic and racial diversity of the school explain a large proportion of the reasons why Math Scores differ across schools. This variance reduction calculation equivalent to the $R_w^{2(f_2)}$ metric in Rights and Sterba (2019).

A new part of this model is that the intercepts *and* slopes are allowed to vary across schools. The between-school variance in the intercept ($_{one-tail}Z = 6.58, p < .01$), SES achievement gap ($_{one-tail}Z = 1.77, p = .038$), and non-White achievement gap ($_{one-tail}Z = 2.08, p = .019$) were all

[2] These metrics require centered data in the `r2mlm` package. The version on CRAN as of April 2023 appeared to use too narrow a threshold to identify that these particular data were centered, so these metrics were not output initially. I edited the source files provided on the GitHub page for the package (https://github.com/mkshaw/r2mlm) to relax the threshold form 1E-7 to 1E-4 to allow the centering to be recognized and to output these values.

significant based on one-tail *Z*-tests.[3] The between-school variance in the non-White achievement gap was 2.14, meaning that the 95% interval for the school-specific non-White achievement gaps is $-2.93 \pm 1.96\sqrt{2.14} = [-5.79, -0.05]$. To be clear, this interval is *not* designed to inferentially assess whether the non-White achievement gap is different from 0 on average. Instead, this interval is designed to inspect the range of the non-White achievement gap across schools. Similarly, there is between-school variance in the SES achievement gap such that the 95% interval for the school-specific SES achievement gap is $1.93 \pm 1.96\sqrt{0.45} = [0.62, 3.24]$.

The school-specific intercepts do not appear to systematically correlate with the school-specific SES ($r = 0.36, Z = 1.87, p = .06$) or non-White achievement gaps ($r = -0.43, Z = -1.82, p = .07$), meaning that the school's average test scores do not seem to affect the magnitude of the achievement gap in the same school (note that the model specifies these terms as covariances, but Table B1 reports them as correlations to make them easier to interpret).[4] The school-specific SES and non-White achievement gaps appeared highly related ($r = -0.99, Z = 2.18, p = .03$)[5]; the correlation is negative due to how variables are coded because a positive value of β_{1j} indicated the presence of an SES achievement gap whereas a negative value of β_{2j} indicated a non-White achievement gap.

Clearly, there is heterogeneity in Math Score achievement gaps and the magnitude of these

[3] Inference for between-cluster variances is complicated by the fact that variances cannot be negative and are bounded below by 0. The problem is that the typical null hypothesis is that the variance equals 0, which means that the null hypothesis is testing the lower boundary of admissible values. This violates the regularity conditions upon which *p*-values are based (Self & Liang, 1987), so one-sided *Z* tests (which only assess the positive tail in which admissible value fall, Lin, 1997) and χ^2 tests are overly conservative (i.e., *p*-values are too high, Type-II error rates are inflated; Stoel et al., 2006). The proper method to compute *p*-values for variance is a 50:50 mixture test (Stram & Lee, 1994), which blends two χ^2 distributions together to arrive at the proper *p*-value. However, if an overly conservative one-sided *Z*-test is already significant, conclusions from the more formal 50:50 mixture test will not differ because the test is systematically too conservative. Because the one-sided *Z*-test was significant for all three between-cluster variances, the more formal test was not performed because it would not affect the conclusion.

[4] Alternatively, the covariance matrix can be directly parameterized with standard deviations and correlations to directly estimate the correlations such that $\begin{bmatrix} \theta_{00}^2 & & \\ \rho_{10}\theta_{00}\theta_{11} & \theta_{11}^2 & \\ \rho_{20}\theta_{00}\theta_{22} & \rho_{21}\theta_{11}\theta_{22} & \theta_{22}^2 \end{bmatrix}$ where the ρ terms are random effect correlations and θ are random effect standard deviations where $\theta = \sqrt{\tau}$. This idea is implemented in the `lme4` R package or with the `TYPE=UNR` option in SAS `PROC MIXED`.

[5] Very large correlations like this are sometimes indicative of linear dependencies or nonpositive definite covariance matrices (e.g., McNeish & Bauer, 2022). However, the covariance matrix is not singular and the random effect design matrix is full rank (minus columns for schools with no within-school variation in non-White identity), so this large estimate is more likely due to high uncertainty in the estimate or omitted school-level predictors at this intermediate step (the correlation is far more moderate in the final model presented in the main text).

achievement gaps appear to be influenced by the school context.

Figure B1 depicts the 160 school-specific achievement gap of SES on Math Scores (one line for each school). The association is generally positive (higher SES is associated with higher scores), but the strength of the association fluctuates across schools and is quite large in some schools but is essentially flat or even negative in other schools.

The model in the main text further builds upon this model. Whereas this model identified how much heterogeneity existed in each of the regression coefficients, the model in the main text aims to explain why this heterogeneity exists and what characteristics of the school are associated with higher or lower coefficients.

Table B1

Estimates from two preliminary model building steps to identify how much variance in the outcome exists at each level (left column), how much heterogeneity exists in the associations between predictors and the outcome (right column)

		Model	
Effect	Notation	Random Intercepts	Random Coefficients
Student Characteristics			
Intercept	γ_{00}	12.64	12.67
SES	γ_{10}	---	1.93
Non-White Identity	γ_{20}	---	−2.93
School Characteristics			
SES School Mean	γ_{01}	---	5.18
Non-White School Mean	γ_{02}	---	−2.09
Variances			
Intercept	τ_{00}	8.61	2.58
SES Slope	τ_{11}	---	0.45
Non-White Identity Slope	τ_{22}	---	2.14
Residual	σ^2	39.15	35.66
Correlations			
Intercept, SES Slope		---	(ns) −0.43
Intercept, Non-White Slope		---	(ns) 0.36
SES Slope, Non-White Slope		---	−0.99

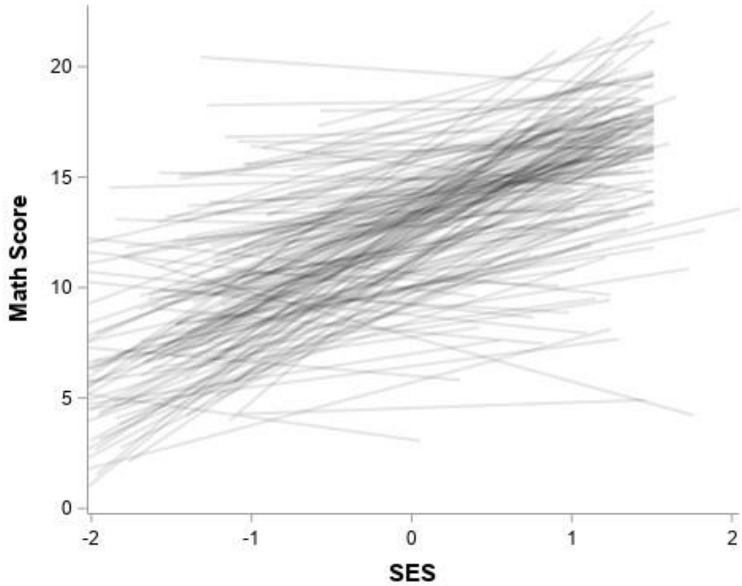

Figure B1. School-specific regression lines showing heterogeneity in the SES achievement gap in Math Scores across 160 schools in the High School Beyond data

Chapter C

Details of Combining Clustered Errors with Multilevel Models

When quantifying sampling variability of fixed effects in a multilevel model, the model-based formula is

$$Cov(\hat{\boldsymbol{\gamma}}) = \left[\sum_{j=1}^{J}\left(\mathbf{X}_j' \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_j\right)\right]^{-1} \tag{C1}$$

Where $\mathbf{X}$ is a design matrix of predictor values in cluster j (i.e., each column is a different predictor, each row is a person in cluster j, and each element is a person's value of that variable) and $\hat{\boldsymbol{\Sigma}}$ is the model-implied covariance matrix, which is based on both the random effect covariance matrix and the within-cluster residual covariance matrix. This compact form of the formula in Equation C1 can be expanded to a more general form,

$$Cov(\hat{\boldsymbol{\gamma}}) = \left[\sum_{j=1}^{J}\left(\mathbf{X}_j' \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_j\right)\right]^{-1}\left[\sum_{j=1}^{J}\left(\mathbf{X}_j' \hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\Sigma}}\mathbf{X}_j\hat{\boldsymbol{\Sigma}}^{-1}\right)\right]\left[\sum_{j=1}^{J}\left(\mathbf{X}_j' \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_j\right)\right]^{-1} \tag{C2}$$

which will be useful shortly. Note that the general form for quantifying fixed effect sampling variability in multilevel models in Equation C2 looks similar to the general form with the single-level models presented in the "More Technical Explanation" subsection of the "Clustered Errors" section except that additional subscripts are added to denote that data are clustered and the $\hat{\boldsymbol{\Sigma}}$ matrix appears in more locations to account for dependence among observations in the same cluster.

To the extent that the multilevel model assumptions are met (e.g., normality is reasonable, all necessary random effects are in the model, the random effect covariance was modeled correctly), model-based standard errors based on the model-implied covariance matrix $\hat{\boldsymbol{\Sigma}}$ will accurately quantify the sampling variability in the fixed effects. However, to the extent that these assumptions are unmet, the sampling variability will be incorrect.

Similar to single-level models, when assumptions are unmet, it is possible to substitute an *empirical* covariance matrix into the standard error calculation to make standard errors robust to potential misspecifications or assumption violations. In multilevel models, this involves substituting the middle term of the middle expression in Equation C2 (the $\hat{\boldsymbol{\Sigma}}$ term) with the empirical covariance $\mathbf{r}_j\mathbf{r}_j'$ where $\mathbf{r}_j = \mathbf{y}_j - \hat{\mathbf{y}}_j$ (Raudenbush & Bryk, 2002, p. 278). After making this substitution, the standard errors for the fixed effects would be calculated as

$$Cov(\hat{\boldsymbol{\gamma}}) = \left[\sum_{j=1}^{J}\left(\mathbf{X}_j' \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_j\right)\right]^{-1}\left[\sum_{j=1}^{J}\left(\mathbf{X}_j' \hat{\boldsymbol{\Sigma}}^{-1}\underline{\mathbf{r}_j\mathbf{r}_j'}\,\mathbf{X}_j\hat{\boldsymbol{\Sigma}}^{-1}\right)\right]\left[\sum_{j=1}^{J}\left(\mathbf{X}_j' \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_j\right)\right]^{-1} \quad \text{(C3)}$$

where the empirical covariance substitution is underlined and in red text to make it easy to locate.

The main point is that the idea and mechanism of clustered errors is applicable to any model for clustered data. If inference is sought from a single sample, assumptions are necessary to quantify the theorical variability of the sampling distribution. This quantification will be correct to the extent that the assumptions are correct, but replacing the model-based matrix for the covariance between observations in the same cluster with the empirical covariance can provide flexibility that permits inferences to be robust to moderate assumption violations.

Chapter D

Interactions and Moderation in Fixed Effect Models

Near Equation 4 in the main text, it was mentioned that there are specification considerations for fixed effect models that have interaction terms (e.g., for testing moderation) because the fixed effect model does not properly disaggregate products of predictor variables. This section expands the reasoning for this statement by extending the empirical example in the main text and adding a simulated example to verify the result.

Extended Empirical Example

This section adds a within-school interaction between SES and non-White identity to the example used in Table 6 main text using the High School Beyond data. The intent is to highlight some of the challenges when including interactions in fixed effect models.

The multilevel model for this example with a within-school interaction would be written as (with the new interaction term noted in red),

$$\begin{aligned}
&\textit{Within-Cluster}\begin{cases} Math_{ij} = \beta_{0j} + \beta_{1j}\left(SES_{ij} - \overline{SES}_j\right) + \beta_{2j}\left(\textit{Non-White}_{ij} - \overline{\textit{Non-White}}_j\right) + \\ \qquad \underline{\beta_{3j}\left(SES_{ij} - \overline{SES}_j\right) \times \left(\textit{Non-White}_{ij} - \overline{\textit{Non-White}}_j\right)} + e_{ij} \\ \mathbf{e} \sim N\left(0, \sigma^2\right) \end{cases} \\
&\textit{Between-Cluster}\begin{cases} \beta_{0j} = \gamma_{00} + u_{0j} \\ \beta_{1j} = \gamma_{10} \\ \beta_{2j} = \gamma_{20} \\ \beta_{3j} = \gamma_{30} \\ \mathbf{u} \sim N\left(0, \tau_{00}\right) \end{cases}
\end{aligned} \tag{D1}$$

In this model, the moderation effect (γ_{30}) is estimated to be −0.39. The within-school main effects for SES and non-White identity are close to the values presented in Table 6 and are 1.96 and −2.92, respectively (compared to 1.95 and −2.90 in Table 6). This gives some preliminary confidence that the model is properly disaggregated. This disaggregation can be assessed with the fixed effect correlation matrix from the `lme4` output in R; if the effects are disaggregated, the correlations between the within effects and the cluster means should ideally be 0, or at least very close to 0 (they may not be exactly 0 in this example because one of the predictors is binary). This appears to be upheld for this model (where any slight deviation occurs when the binary predictor non-White identity is involved; R code and output are on the next page).

```
D1<-lme4::lmer(math~
                  sesCMC + nwCMC + sesCMC:nwCMC +
                  meanSES + meanNW +
                  (1|school),
                data=HSB)
summary(D1)
```

```
Correlation of Fixed Effects:
            (Intr) sesCMC nwCMC  menSES meanNW
sesCMC      -0.003
nwCMC        0.003  0.158
meanSES     -0.336  0.000  0.000
meanNW      -0.716 -0.002  0.002  0.478
sesCMC:nwCMC 0.036 -0.085  0.089  0.003  0.028
```

Second, an interaction is added to the fixed effect model such that the fitted model is

$$\begin{aligned} &y_{ij} = \sum_{j=1}^{J} C_j \alpha_j + \beta_1 \times SES_{ij} + \beta_2 \times Non\text{-}White_{ij} + \beta_3 \left(SES_{ij} \times Non\text{-}White_{ij} \right) + e_{ij} \\ &e_{ij} \sim \left(0, \sigma^2\right) \end{aligned} \tag{D2}$$

The SES and non-White identity estimates in this model are 2.24 and −3.01, respectively, and stray further from the within effects estimated in Table 6 (which were 1.95 and −2.90, respectively). Furthermore, the moderation effect is −0.89 and is quite far from the moderation effect from the multilevel model in Equation D1 (which was −0.39). This suggests one of the models is no longer fully disaggregated, likely the fixed effect model because the main effects are more discrepant from Table 6.

To verify, the equivalent multilevel model to the fixed effect model in Equation D2 is specified with cluster-mean centered main effects but a *raw* interaction such that the within-cluster submodel is,

$$Within\text{-}Cluster \begin{cases} Math_{ij} = \beta_{0j} + \beta_{1j} \left(SES_{ij} - \overline{SES}_j \right) + \beta_{2j} \left(Non\text{-}White_{ij} - \overline{Non\text{-}White}_j \right) + \\ \qquad \beta_{3j} \underline{\left(SES_{ij} \times Non\text{-}White_{ij} \right)} + e_{ij} \\ \mathbf{e} \sim N\left(0, \sigma^2\right) \end{cases} \tag{D3}$$

Where the difference is underlined in red text. This model yields main effect of 2.23 and −3.00 for SES and non-White identity, respectively, and the moderation estimate is close to the model in Equation D2 at −0.86. However, the fixed effect correlation matrix shows that the within-cluster effects are not uncorrelated with the cluster means, which suggests that the model is not properly disaggregated and that between-cluster characteristics may permeate into the within-cluster submodel (R code and output are on the next page).

```
MLM<-lme4::lmer(math~
                  sesCMC+nwCMC+ses:nonwhite+
                  meanSES+meanNW+
                  (1|school),
                data=HSB)

summary(MLM)

Correlation of Fixed Effects:
             (Intr) sesCMC nwCMC  menSES meanNW
sesCMC       -0.014
nwCMC         0.003  0.075
meanSES      -0.336  0.097 -0.022
meanNW       -0.712 -0.053  0.012  0.449
ses:nonwhite 0.027 -0.532  0.122 -0.182  0.099
```

If using the suggested specification in Giesselmann and Schmidt-Catran (2022) where the terms in the interaction are manually cluster-mean centered,

$$\begin{aligned} y_{ij} &= \sum_{j=1}^{J} C_j \alpha_j + \beta_1 \times SES_{ij} + \beta_2 \times Non\text{-}White_{ij} + \\ &\quad \underline{\beta_3 \left[\left(SES_{ij} - \overline{SES}_j \right) \times \left(Non\text{-}White_{ij} - \overline{Non\text{-}White}_j \right) \right]} + e_{ij} \\ e_{ij} &\sim \left(0, \sigma^2\right) \end{aligned} \tag{D4}$$

Then the main effect estimate for SES is 1.96 and the main effect estimate for non-White identity is −2.92, which matches the multilevel model estimates. The interaction effect is −0.47, which is still a little distant from the estimate of the within-school moderation effect in Equation D1 (which was −0.39).

This may be attributable to two causes. First, these data are unbalanced and the number of students in each cluster varies fairly widely (14 to 67 students, depending on the cluster), so there may be some differences in estimates because the model in Equation D1 partially pools estimates whereas the model in Equation D3 does not pool estimates. Second, the non-White identity predictor is a binary variable, which sometimes can be nuanced to properly center (Yaremych et al., 2021). Therefore, the two estimates may not be expected to be identical even under ideal circumstances.

The next section provides a simulation to verify that the idea of manually centering the product results in the true within-cluster moderation effect and that the within-cluster moderation is equal with multilevel models and fixed effect models. To do this, data with equally sized clusters are simulated with all continuous predictors so that the estimates from the two models are identical and have the same population values.

Simulated Example

Clustered data with 5000 clusters and 500 observations per cluster (2.5 million total observations) are generated from the following model,

$$
\begin{aligned}
y_{ij} &= \beta_{0j} + \beta_{1j}\left(x_{1ij} - \bar{x}_{1j}\right) + \beta_{2j}\left(x_{2ij} - \bar{x}_{2j}\right) + \beta_{3j}\left(x_{1ij} - \bar{x}_{1j}\right)\left(x_{2ij} - \bar{x}_{2j}\right) + e_{ij} \\
\beta_{0j} &= 1 + 0.15 \times z_{1j} + 0.20 \times z_{2j} - 0.15 \times \bar{x}_{1j} - 0.075 \times \bar{x}_{2j} + u_{0j} \\
\beta_{1j} &= 0.25 \\
\beta_{2j} &= 0.15 \\
\beta_{3j} &= 0.20 \\
\mathbf{u} &\sim N(0,1) \\
\mathbf{e} &\sim N(0,1) \\
\bar{\mathbf{x}} &\sim MVN\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 & \\ 0 & 2 \end{bmatrix}\right) \\
\left(\mathbf{x} - \bar{\mathbf{x}}\right) &\sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \\ .45 & 1 \end{bmatrix}\right) \\
\mathbf{z} &\sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \\ .30 & 1 \end{bmatrix}\right)
\end{aligned} \tag{D5}
$$

This model has two continuous within-cluster predictors (x), two continuous between-cluster predictors (z), and a continuous outcome (y). The within and between effects of the x variables are in opposite directions. For x_1, the within effect is 0.25 (from the β_1 equation) and the between effect is -0.15 (the third term from the β_0 equation). For x_2, the within effect is 0.15 (from the β_2 equation) and the between effect is -0.075 (the fourth term from the β_0 equation). The within-cluster moderation effect is 0.20 (from the β_3 equation). Code for simulating data from this model is shown at the end of this chapter.

To these data, four models are fit:

1. A random intercepts multilevel that cluster-mean centers the within-cluster main effects and within-cluster interaction
2. A traditional fixed effect model that includes the raw main effects and raw interaction
3. A random intercepts model that cluster-mean centers the within-cluster main effects but includes the raw interaction
4. A fixed effect model that includes the raw within-cluster main effects but a manually cluster-mean centered interaction

Model 1. Model 1 can be written out as

$$\begin{aligned}
&\textit{Within-Cluster} \begin{cases} y_{ij} = \beta_{0j} + \beta_{1j}\left(x_{1ij} - \bar{x}_{1j}\right) + \beta_{2j}\left(x_{2ij} - \bar{x}_{2j}\right) + \beta_{3j}\left(x_{1ij} - \bar{x}_{1j}\right) \times \left(x_{2ij} - \bar{x}_{2j}\right) + e_{ij} \\ \mathbf{e} \sim N\left(0, \sigma^2\right) \end{cases} \\
&\textit{Between-Cluster} \begin{cases} \beta_{0j} = \gamma_{00} + \gamma_{01} \times z_{1j} + \gamma_{02} \times z_{2j} + \gamma_{03} \times \bar{x}_{1j} + \gamma_{04} \times \bar{x}_{2j} + u_{0j} \\ \beta_{1j} = \gamma_{10} \\ \beta_{2j} = \gamma_{20} \\ \beta_{3j} = \gamma_{30} \\ \mathbf{u} \sim N\left(0, \tau_{00}\right) \end{cases}
\end{aligned} \tag{D6}$$

Which matches the data generation model in Equation D5 and unsurprisingly estimates the within-cluster effects accurately: the estimated main effect of x_1 is 0.247, the estimated main effect of x_2 is 0.151, and the estimated within-cluster interaction effect is 0.200 (compared to population values of 0.25, 0.15, and 0.20). Table D1 compares these estimates to the population models and to models that are fit in subsequent subsections.

Model 2. Model 2 can be written out as

$$\begin{aligned}
&y_{ij} = \sum_{j=1}^{J} C_j \alpha_j + \beta_1 \times x_{1ij} + \beta_2 \times x_{2ij} + \beta_3\left(x_{1ij} \times x_{2ij}\right) + e_{ij} \\
&e_{ij} \sim \left(0, \sigma^2\right)
\end{aligned} \tag{D7}$$

which is the traditional fixed effect model specification. Though fixed effect models are often claimed to provide within-cluster effects that account for all between-cluster characteristics, this property is not upheld when there are within-cluster interaction terms. The estimated main effect of x_1 is 0.177, the estimated main effect of x_2 is 0.108, and the estimated within-cluster interaction effect is 0.038. All of these estimates are quite far from the true within-cluster effects of 0.25, 0.15, and 0.20, respectively, which should be more closely recovered with 2.5 million observations.

As noted in the main text, this occurs because the interaction is not disaggregated appropriately in the traditional fixed effect model because the β_3 corresponds to $\left(x_1 x_2 - \overline{x_1 x_2}\right)$ rather than the actual disaggregated interaction term of $\left(x_{1ij} - \bar{x}_{1j}\right)\left(x_{2ij} - \bar{x}_{2j}\right)$. The $\left(x_1 x_2 - \overline{x_1 x_2}\right)$ term is not orthogonal to the between-cluster submodel, so the 4 between-cluster effects in the simulated data are not completely controlled for and can still permeate to the β coefficients in Equation D7. Correspondingly, estimates from Model 2 are not actually within-cluster effects as is often assumed from fixed effect models but instead are conflated effects that include a mix of within-cluster and between-cluster information.

Model 3. To show how the traditional fixed effect model produces conflated coefficients, Model 3 is a multilevel model similar to Model 1 except that the interaction term is specified as the product of the *raw* within-cluster predictors rather than the *centered* within-cluster predictors. This model is shown in Equation D6 with the major difference from Model 1 underlined and in red text.

$$\begin{array}{ll} \textit{Within-Cluster} & \begin{cases} y_{ij} = \beta_{0j} + \beta_{1j}\left(x_{1ij} - \bar{x}_{1j}\right) + \beta_{2j}\left(x_{2ij} - \bar{x}_{2j}\right) + \underline{\beta_{3j}\left(x_{1ij} \times x_{2ij}\right)} + e_{ij} \\ \mathbf{e} \sim N\left(0, \sigma^2\right) \end{cases} \\ \textit{Between-Cluster} & \begin{cases} \beta_{0j} = \gamma_{00} + \gamma_{01} \times z_{1j} + \gamma_{02} \times z_{2j} + \gamma_{03} \times \bar{x}_{1j} + \gamma_{04} \times \bar{x}_{2j} + u_{0j} \\ \beta_{1j} = \gamma_{10} \\ \beta_{2j} = \gamma_{20} \\ \beta_{3j} = \gamma_{30} \\ \mathbf{u} \sim N\left(0, \tau_{00}\right) \end{cases} \end{array} \tag{D8}$$

This model is conflated because x_1 and x_2 include both within-cluster and between-cluster information, meaning that the within-cluster submodel is not orthogonal to the between-cluster submodel. Identical to Model 2, the estimated main effect of x_1 is 0.177, the estimated main effect of x_2 is 0.108, and the estimated within-cluster interaction effect is 0.038. Despite fixed effect models' desire to isolate within-cluster variance and yield completely disaggregated within-cluster effects free of between-cluster influence, a traditional fixed effect model with interaction terms is equivalent to the *conflated* multilevel model in Model 3 rather than the disaggregated multilevel model in Model 1 that accurately estimates pure within-cluster effects.

Model 4. To fully disaggregate a fixed effect model with within-cluster interactions, the interaction term must be manually centered because the cluster affiliation dummy variables will not properly disaggregate the effects. This model would be written out as,

$$
\begin{aligned}
y_{ij} &= \sum_{j=1}^{J} C_j \alpha_j + \beta_1 \times x_{1ij} + \beta_2 \times x_{2ij} + \underline{\beta_3 \left(x_{1ij} - \bar{x}_{1j}\right)\left(x_{2ij} - \bar{x}_{2j}\right)} + e_{ij} \\
e_{ij} &\sim \left(0, \sigma^2\right)
\end{aligned}
\tag{D9}
$$

Where the main difference between Model 2 and Model 4 is underlined and in red text. Estimates from this model accurately reflect the population values: the estimated main effect of x_1 is 0.247, the estimated main effect of x_2 is 0.151, and the estimated within-cluster moderation effect is 0.200 (where the population values are 0.25, 0.15, and 0.20).

Table D1

Comparison of within-cluster main effects and moderation effect for applying Models 1 through 4 to simulated data

	Population	Model 1	Model 2	Model 3	Model 4
x_1 Within Effect	0.25	0.247	0.177	0.177	0.247
x_2 Within Effect	0.15	0.151	0.108	0.108	0.151
Within-Cluster Moderation	0.20	0.200	0.038	0.038	0.200

Note: Model 1 and Model 4 properly disaggregate such that the within-cluster effect estimates correctly reproduce the population value. Model 2 and Model 3 are conflated and do not produce proper within cluster estimates

Simulated Example: Data Generation R Code

```
################
#Data Structure#
################

#number of clusters
N<-5000
#observations per cluster
nj<-500
#empty matrix to be populated
dat<-matrix(, nrow=N*nj,ncol=9)
#eventual column names
colnames(dat)<-c("cluster", "person", "u0","x1", "x2", "y", "z1" ,"z2","e")
#index used later for looping through clusters
index<-1
```

```
#####################
#Generate Predictors#
#####################

#set seed value for reproducibility
set.seed(9718)
#random intercepts
u<-rnorm(n=N,0,sd=1)
#between cluster predictors
z<-MASS::mvrnorm(n=N, mu=c(0,0), Sigma=matrix(c(1,.3,.3,1), ncol=2))
#within-cluster predictor cluster means
xmean<-MASS::mvrnorm(n=N, mu=c(1,2), Sigma=matrix(c(1,0,0,4), ncol=2))
#centered within-cluster predictors
x<-MASS::mvrnorm(n=N*nj, mu=c(0,0), Sigma=matrix(c(1,.45,.45,1), ncol=2))

###################
#Population Values#
###################

#fixed intercept
gam00<-1
#fixed within effect, x1
gam10<-.25
#fixed within effect, x2
gam20<-.15
#fixed within moderation effect
gam30<-.20

#effect of Z1
gam01<-.15
#effect of Z2
gam02<-.20
#between effect of x1
gam03<--.15
#between effect of x2
gam04<--.075

############################
#Loop to Generate Full Data#
############################

for (cluster in 1:N){

  u0<-u[cluster]
  z1<-z[cluster,1]
  z2<-z[cluster,2]
  xmean1<-xmean[cluster,1]
  xmean2<-xmean[cluster,2]

  for (person in 1: nj){

    x1<-x[(cluster+person-1),1]+xmean1
    x2<-x[(cluster+person-1),2]+xmean2
    e<-rnorm(1,0,1)

    #define each regression equation
```

```
    b0<-gam00+gam01*z1+gam02*z2+gam03*xmean1+gam04*xmean2+u0
    b1<-gam10
    b2<-gam20
    b3<-gam30

    #build model for outcome
    y<-b0+b1*(x1-xmean1)+b2*(x2-xmean2)+b3*(x1-xmean1)*(x2-xmean2)+e

    dat[index,1:9]<-c(cluster, person, u0,x1,x2,y,z1,z2,e)
    #update cluster ID for simulation
    index<-index+1
  }
}

#create data frame
d<-as.data.frame(dat)

#cluster mean center x1 and x2
d1<-misty::center(d[,c("x1","x2")],type="CWC", cluster=d$cluster)

#cluster means for x1 and x2
d2<-misty::cluster.scores(d[,c("x1","x2")],fun="mean", cluster=d$cluster)

#new dataset with centered variables
d3<-cbind(d,d1,d2)

#calculate within-cluster interaction of centered variables
d3$int<-d3$x1.c*d3$x2.c

#d3 is the final data set
```

Model Fitting R Code and Output

Model 1

```
library(lme4)

MLM_CMC<-lme4::lmer(
      y~x1.c+x2.c+x1.c:x2.c+z1+z2+x1.a+x2.a+
      (1|cluster),
      data=d3)
summary(MLM_CMC)

      Fixed effects:
                   Estimate Std. Error t value
      (Intercept)  0.9858746  0.0244747   40.28
      x1.c         0.2466733  0.0006958  354.51
      x2.c         0.1508103  0.0007062  213.55
      z1           0.1523827  0.0149384   10.20
      z2           0.1946677  0.0149467   13.02
      x1.a        -0.1414010  0.0141392  -10.00
      x2.a        -0.0729191  0.0070666  -10.32
      x1.c:x2.c    0.1997313  0.0005744  347.70
```

Model 2

```
library(plm)
FEM_trad<-plm::plm (y ~ x1+ x2+x1:x2,
               data = d3,
               index= "cluster",
               model="within")
summary(FEM_trad)
```

```
      Coefficients:
              Estimate Std. Error t-value  Pr(>|t|)
      x1    0.17723983 0.00087535  202.48 < 2.2e-16 ***
      x2    0.10783302 0.00076183  141.55 < 2.2e-16 ***
      x1:x2 0.03811953 0.00025423  149.94 < 2.2e-16 ***
```

Model 3

```
MLM_raw<-lme4::lmer(
      y~x1.c+x2.c+x1:x2+z1+z2+x1.a+x2.a+
      (1|cluster),
      data=d3)
summary(MLM_raw)
```

```
      Fixed effects:
                    Estimate Std. Error t value
      (Intercept)  1.1350918  0.0245593   46.22
      x1.c         0.1773479  0.0008752  202.65
      x2.c         0.1078856  0.0007618  141.62
      z1           0.1528164  0.0149893   10.20
      z2           0.1958584  0.0149975   13.06
      x1.a        -0.2147907  0.0141958  -15.13
      x2.a        -0.1111273  0.0070952  -15.66
      x1:x2        0.0380660  0.0002541  149.83
```

Model 4

```
FEM_cent<-plm::plm (y ~ x1+ x2+x1.c:x2.c,
               data = d3,
               index= "cluster",
               model="within")
summary(FEM_cent)
```

```
      Coefficients:
                  Estimate Std. Error t-value  Pr(>|t|)
      x1        0.24667324 0.00069582  354.51 < 2.2e-16 ***
      x2        0.15081035 0.00070622  213.55 < 2.2e-16 ***
      x1.c:x2.c 0.19973165 0.00057444  347.70 < 2.2e-16 ***
```