

Multilevel Modeling Workshop University of Kentucky

Brandon Bartels
George Washington University

May 23-25, 2011

Introduction

- Exciting methodological toolkit
- *Multilevel modeling is not monolithic*
 - There are lots of different types of model specifications that fall under the umbrella.
 - Various specifications carry different substantive interpretations.

Outline

- I. Motivation and Core Issues
- II. Linear variance components model
- III. Random intercept model (aka, random effects model) and its alternatives (e.g., OLS, fixed effects, “between” effects)
- IV. Cluster confounding
- V. Applications to longitudinal (panel/time-series cross-sectional) data
- VI. Random coefficient model
- VII. Nonlinear models for noncontinuous dependent variables

I. Motivation and Core Issues

Multilevel Data

- Contain multiple **levels of analysis**, with each level consisting of distinct units of analysis.
- Most common form of multilevel data: *hierarchical data*.
 - Two-level structure: Units from the lowest level of analysis (level-1 units) are **nested within** units from a higher level of analysis (level-2 units)
 - Data are “clustered”
 - Level-2 units are referred to as “clusters”
 - Three-level structure: Third level is present

Multilevel Data

- Examples
 - Education: students (level-1 units) nested within schools (level-2 units)
 - Three levels: students nested within schools nested within states
 - Individuals nested within cities
 - Voters nested within congressional districts
 - Voters nested within time (or temporal contexts)
 - Panel data and time-series cross-sectional (TSCS) data
- What’s a sufficient number of *level-2 units*, or *clusters*?
 - Rough guideline: >15
- What’s a sufficient number for *cluster sizes* (number of observations per cluster/level-2 unit)?
 - Cluster sizes can vary; at least 2 and more like >5 (rough guideline)

Multilevel Data

Student	School	Y	X1	X2	X3	X4
1	1	54	2	32	1	44
2	1	64	4	25	1	44
3	1	87	9	45	1	44
4	2	24	4	44	0	36
5	2	98	7	32	0	36
6	2	65	6	22	0	36
7	3	45	9	19	0	22
8	3	32	5	15	0	22
9	3	37	2	25	0	22
10	4	84	7	30	1	45
11	4	45	4	38	1	45
12	4	65	3	36	1	45
13	5	21	8	41	1	18
14	5	65	6	22	1	18
15	5	98	1	18	1	18

- X1 and X2 are level-1 variables

- X3 and X4 are level-2 variables.

- Balanced data: cluster sizes are equal

Panel / Time-Series Cross-Sectional Data

<i>i</i>	<i>j</i>	<i>t</i>	Y	X1	X2	X3	X4
1	1	1	54	2	32	1	44
2	1	2	64	4	25	1	44
3	1	3	87	9	45	1	44
4	2	1	24	4	44	0	36
5	2	2	98	7	32	0	36
6	2	3	65	6	22	0	36
7	3	1	45	9	19	0	22
8	3	2	32	5	15	0	22
9	3	3	37	2	25	0	22
10	4	1	84	7	30	1	45
11	4	2	45	4	38	1	45
12	4	3	65	3	36	1	45
13	5	1	21	8	41	1	18
14	5	2	65	6	22	1	18
15	5	3	98	1	18	1	18

Motivation

- Types of phenomena we're interested in are multilayered and complex.
 - Incorporating these layers enhances our substantive explanations of phenomena.
- People don't make choices or behave in a vacuum; there's a *context* in which they act.
- This contextual, or situational, variation may have consequences for how people behave.
- Most simple cross-sectional data ignores this structure; "naïve pooling"

Motivation

- Parsing explained variance in the dependent variable between *individual* versus *aggregate* levels of analysis.
 - Student versus school effects on performance.

Key Topics

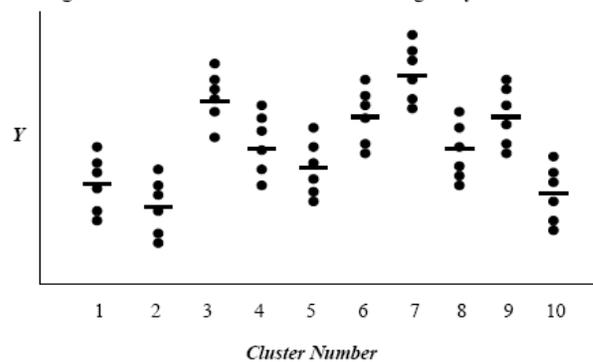
1. Unobserved heterogeneity (in the dependent variable)

- Between-cluster heterogeneity in the dependent variable
 - Unobserved factors specific to each cluster that influence the dependent variable; factors are shared by observations within each cluster.
 - Unmeasured, unobserved, and unimagined differences between clusters.
- Method: Random intercept models (aka, random effects)
- UH in a cross-sectional context:

$$y_i = b_0 + b_1x_{1i} + b_2x_{2i} + e_i$$

UH in Hierarchical Context

Figure 1: Illustration of Unobserved Heterogeneity Across Clusters



Note: Dots represent responses within a given cluster. Dashes represent the means of Y for each cluster.

Key Topics

2. Pooling

- Degree to which parameters (e.g., intercept, effects of IVs) are “pulled” toward the pooled (global) effect or reflect within-cluster variation.

Spectrum:

No *Partial Pooling* *Complete*
Pooling ----- *Pooling*

3. Distinguish within-cluster, between-cluster, and total variation

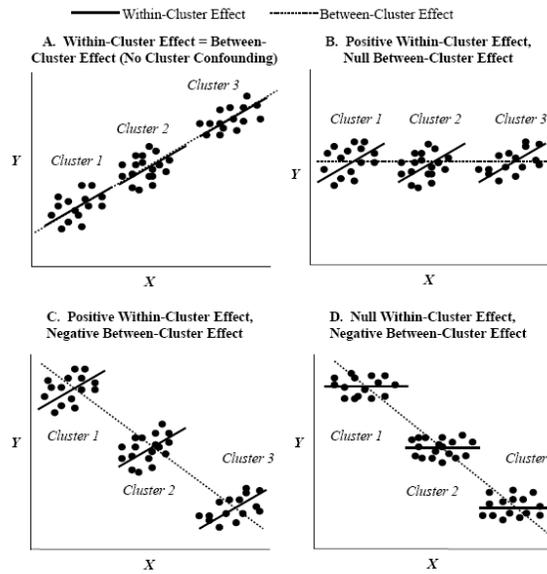
- “Cluster confounding”

Within-Cluster v. Between-Cluster Variation

Student	School	Y	X1	X2	X3	X4
1	1	54	2	32	1	44
2	1	64	4	25	1	44
3	1	87	9	45	1	44
4	2	24	4	44	0	36
5	2	98	7	32	0	36
6	2	65	6	22	0	36
7	3	45	9	19	0	22
8	3	32	5	15	0	22
9	3	37	2	25	0	22
10	4	84	7	30	1	45
11	4	45	4	38	1	45
12	4	65	3	36	1	45
13	5	21	8	41	1	18
14	5	65	6	22	1	18
15	5	98	1	18	1	18

Cluster Confounding

Figure 2: Illustration of Cluster Confounding



Key Topics

4. Causal heterogeneity

- When the relationship between X and Y varies across cluster
- How higher level variables shape lower-level relationships.
- Methods: Random coefficient models

II. Linear Variance Components Model

Modeling Clustered Data

- We'll start simple: No independent variables
- Linear variance components model
- We'll focus on making inferences about *cluster means*, i.e., mean of Y for each level-2 unit.

Pooling

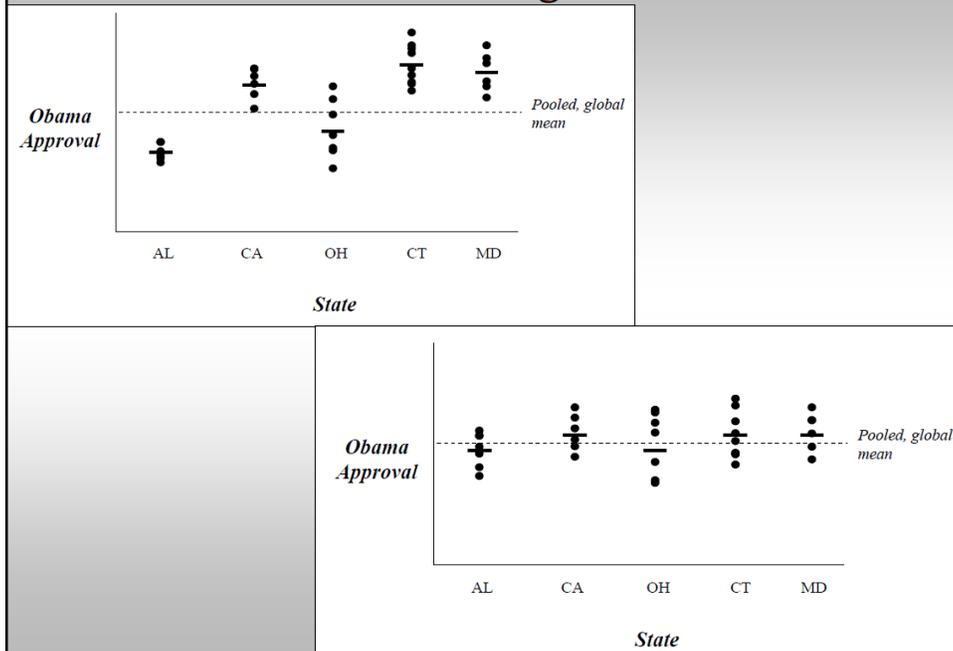
- Degree to which each cluster mean is “pulled” toward the overall, global mean.

Spectrum:

No *Partial Pooling* *Complete*
Pooling ----- *Pooling*

- Important questions:
 - What does each approach imply?
 - Under what conditions would we want to rely on each type when making inferences about cluster means?

Pooling



Modeling Clustered Data

- ***No Pooling:***
 - *Within-cluster* central tendency and variation are all that matter.
 - *Between-cluster* variation ignored.
 - Generalize (in terms of means of the DV) one cluster at a time (in isolation) using cluster means
 - *Estimation technique:* Fixed-effects (within) estimator
- ***Complete Pooling:***
 - Ignores clustering/hierarchical structure.
 - Doesn't distinguish within- versus between-cluster variation
 - Generalization: "Global mean" or "grand mean"
 - Balanced data: mean of DV over entire sample
 - Unbalanced data: mean of the cluster means
 - *Estimation technique:* Plain-vanilla pooled regression (e.g., OLS)

Modeling Clustered Data

- ***Partial Pooling:***
 - *Weighted average* between no pooling and complete pooling extremes.
 - Borrows information from completely pooled mean to generate refined estimate of cluster mean (think about "uninformative" clusters)
 - *Partially-pooled estimates of cluster means* are going to be weighted averages of the "no pooling cluster means" and the "completely-pooled mean of the cluster means."
 - *Estimation technique:* random intercept (aka, random effects) model
 - Considerations affecting the degree of partial pooling?

Modeling Clustered Data

Level-1 units indexed $i=1, 2, \dots, N$. Level-2 units indexed $j=1, 2, \dots, J$.

N level-1 units nested within J level-2 units.

[Level-1 equation] $y_{ij} = \beta_{0j} + \varepsilon_{ij}$

[Level-2 equation] $\beta_{0j} = \gamma_{00} + \zeta_j$

Reduced form version:

$$y_{ij} = \gamma_{00} + \zeta_j + \varepsilon_{ij}$$

$\zeta_j =$ unobserved heterogeneity (cluster level)

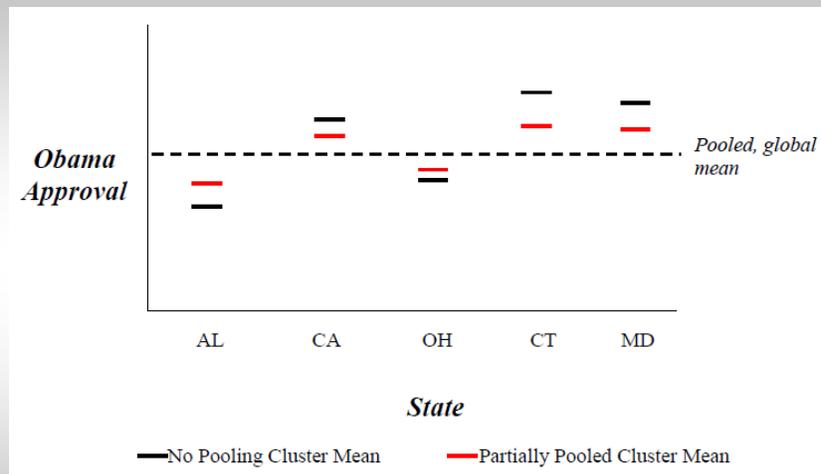
$\text{var}(\zeta_j) = \psi$: Between-cluster variance.

$\text{var}(\varepsilon_{ij}) = \theta$: Within-cluster variance.

Intraclass correlation: $\rho = \psi / (\psi + \theta)$

- Key specification decision: How we treat ζ_j is directly connected to the three approaches just discussed.
 - No pooling?
 - Complete pooling?
 - Partial pooling?

Cluster Means and Partial Pooling



Intraclass Correlation

- Intraclass correlation: $\rho = \psi / (\psi + \theta)$
- Can be thought of as:
 - Degree of cluster-level unobserved heterogeneity
 - Degree of within-cluster dependence
 - Connection to reliability
 - “Cluster differentiation” or “uniqueness”
- What makes ρ large or small?
 - Depends on changes in ψ and changes in θ

Testing for Unobserved Heterogeneity

- Is there significant between-cluster UH?
- Hypothesis test:
 - $H_0: \psi = 0$
 - $H_A: \psi > 0$
- Statistical tests

Shrinkage and Pooling in Random Intercept Model

- Shrinkage and pooling are directly related.
 - The “weight” that determines how much the within-cluster means are pulled toward the pooled mean
- *Shrinkage* is the degree to which ζ_j 's (level-2 residuals) are pulled toward zero; centers on estimating ζ_j 's.
- *Pooling* is the degree to which the cluster means gravitate toward the global mean of Y .
 - Gives us deeper insight into how much unobserved heterogeneity there is in the dependent variable.
- Recall that the variance components model (random intercept model w/no IVs) allows for *partial pooling* of the cluster means.
- We can calculate the *degree of partial pooling* using a “pooling factor.”

Generating Partially-Pooled Level-2 Residuals

- Level-2 residuals from the partially-pooled approach reflect how much each cluster deviates from the global mean.
- Partially pooled level-2 residuals are called “empirical Bayes” (EB) residuals, which uses the “prior” distribution of ζ [$\zeta \sim N(0, \psi)$], combined with the “data” (how informative the clusters are individually) to generate a “posterior” prediction of ζ (partially-pooled prediction).
- The smaller ψ is, the more informative the prior and the more it will drag the predicted ζ toward 0, which is the mean of the prior distribution (hence, “shrinkage”).

Shrinkage Factor and EB Residuals

$$R_j = \frac{\psi}{\psi + \theta / n_j}$$

- The EB prediction of ζ is:

$$\zeta_j^{EB} = R_j * \zeta_j^{ML}$$

- Note that n_j represents the cluster size for cluster j .
- There will more shrinkage when (note high shrinkage is associated with small R):
 - ψ is small (informative prior)
 - θ is large (uninformative data)
 - Cluster sizes (n_j) are small (uninformative data)

Partial Pooling

- *Partial pooling* is the extent to which partially-pooled cluster means gravitate toward the pooled (global) mean of Y . The pooling factor, ω_j can be calculated as:

$$\omega_j = 1 - R_j = 1 - \left(\frac{\psi}{\psi + \theta / n_j} \right) = \frac{\theta}{\theta + n_j \psi}$$

- The same factors that increase shrinkage of the ζ_j 's will increase the degree of partial pooling. Thus, the partially-pooled cluster means will increasingly pool around the global mean when:
 - ψ is small (little differentiation in cluster means)
 - θ is large (uninformative clusters)
 - Cluster sizes are small (uninformative clusters)
- If $\omega = 0$, what happens?
- If $\omega = 1$, what happens?
- If $0 < \omega < 1$, what happens?

Pooling

- Shrinkage and pooling factors can be calculated using our model results; note that all we need are the variance estimates at each level and the cluster size(s).
- Pooling factors for balanced versus unbalanced data....
 - Balanced?
 - In unbalanced data, how will variation in cluster size influence the degree of pooling? How and why?

Calculating Partially-Pooled Cluster Means

- We can use our estimates of pooling factors to calculate partially-pooled estimates of our parameters (in this case, the random intercepts, β_{0j}).
- In a model with no independent variables, *our partially-pooled estimates of the random intercepts will be partially-pooled cluster means.*
- **First**, use the following equation:

$$\hat{\beta}_{0j} = \omega_j \mu + (1 - \omega_j) \bar{y}_j$$

- Note that μ represents the pooled mean (mean of the cluster means); \bar{y}_j represents the cluster mean for cluster j .
- Revisit: If $\omega = 0$, what happens? If $\omega = 1$, what happens?

Partially-Pooled Means

$$y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \zeta_j$$

- **Second**, to generate partially-pooled cluster means (intercepts), use:

$$\hat{\beta}_{0j} = \gamma_{00} + \zeta_j^{EB}$$

- Note the similarity to the other equation...

$$\hat{\beta}_{0j} = \omega_j \mu + (1 - \omega_j) \bar{y}_j$$

III. Random Intercept Model and Its Alternatives

Modeling Clustered Data

- Let's add independent variables!
- Four approaches (producing different inferences about the effect of X on Y):
 1. Complete pooling (OLS)
 2. No pooling (fixed effects, or within estimator)
 3. Partially pooled (random intercept model)
 - Now, we're dealing with partially-pooled *coefficients*.
 - *Effects of X's on Y are a weighted average between the complete pooling and no pooling (within) estimates.*
 4. Between estimator (which is also no pooling, but in a different way than the within approach).
- Different types of interpretations....

Fixed Effects (Within) Approach

$$y_{ij} = \gamma_{00} + \beta_1 x_{ij} + \zeta_j + \varepsilon_{ij}$$

- Two equivalent ways of thinking about this:
 1. *Dummy variable method* (LSDV): include unit-specific dummy variables (leave one as the excluded group). All of the between-unit variation is absorbed in the estimates of the fixed ζ_i 's.

Fixed Effects (Within) Approach

<i>i</i>	<i>j</i>	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
1	1	1	0	0	0	0
2	1	1	0	0	0	0
3	1	1	0	0	0	0
4	2	0	1	0	0	0
5	2	0	1	0	0	0
6	2	0	1	0	0	0
7	3	0	0	1	0	0
8	3	0	0	1	0	0
9	3	0	0	1	0	0
10	4	0	0	0	1	0
11	4	0	0	0	1	0
12	4	0	0	0	1	0
13	5	0	0	0	0	1
14	5	0	0	0	0	1
15	5	0	0	0	0	1

Fixed Effects (Within) Approach

$$y_{ij} = \gamma_{00} + \beta_1 x_{ij} + \zeta_j + \varepsilon_{ij}$$

- Two equivalent ways of thinking about this:
 1. *Dummy variable method* (LSDV): include unit-specific dummy variables (leave one as the excluded group). All of the between-unit variation is absorbed in the estimates of the fixed ζ_j 's.

Thus, the β 's are *within-cluster effects*.

- LSDV (least squares dummy variable) can be estimated via OLS with the inclusion of the unit-specific dummies (minus one).
- What happens to level-2 variables?

Fixed Effects (Within) Approach

2. *Deviations from means*

- Subtract the cluster-specific means from each value of each variable. Do this for both Y and the X's.

$$y_{ij}^W = y_{ij} - \bar{y}_j$$

$$x_{ij}^W = x_{ij} - \bar{x}_j$$

- Note: (1) ζ_j , the between-unit effect, is eliminated; and (2) how both approaches explicitly highlight that the β s are *within-unit* effects.

Fixed Effects (Within) Approach

- To estimate this second approach (deviation from means), subtract cluster means, then estimate with OLS using these transformed variables.
- Note that the LSDV and “deviations from means” approaches produce *analytically equivalent* estimates of β .

$$\beta_W = W_{xx}^{-1}W_{xy}$$

Between Estimator

- Ignores within-cluster variation, focuses solely on between-cluster variation
- Regress cluster means of Y on cluster means of X.

$$\beta_B = B_{xx}^{-1} B_{xy}$$

- Note: What's the relationship between the within- and between-cluster versions of a variable?

$$y_{ij} = y_{ij}^W + \bar{y}_j$$

$$x_{ij} = x_{ij}^W + \bar{x}_j$$

Random Intercept Model (Partial Pooling)

Level-1 units indexed $i=1, 2, \dots, N$. Level-2 units indexed $j=1, 2, \dots, J$.

N level-1 units nested within J level-2 units.

[Level-1 equation] $y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij}$

[Level-2 equation] $\beta_{0j} = \gamma_{00} + \zeta_j$

Assumptions:

$$\varepsilon_{ij} | x_{ij} \sim N(0, \theta)$$

$$\zeta_j | x_{ij} \sim N(0, \psi)$$

$$\text{Cov}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0, i \neq i'$$

$$\text{Cov}(\zeta_j, \zeta_{j'}) = 0, j \neq j'$$

$$\text{Cov}(\varepsilon_{ij}, \zeta_j) = 0$$

$$\text{Cov}(\varepsilon_{ij}, x_{ij}) = 0$$

$$\text{Cov}(\zeta_j, x_{ij}) = 0$$

Two-Level Random Intercept Model

Reduced form version:

$$y_{ij} = \gamma_{00} + \zeta_j + \beta_1 x_{ij} + \varepsilon_{ij} \quad \text{Or...}$$

$$y_{ij} = \gamma_{00} + \beta_1 x_{ij} + \zeta_j + \varepsilon_{ij}$$

- *Fixed part and random part...* [note “fixed” versus “random” effects verbiage.]
- $\text{Var}(\zeta_j) = \psi$: Between-cluster (level-2) error variance.
- $\text{Var}(\varepsilon_{ij}) = \theta$: Within-cluster (level-1) error variance.
- Intraclass correlation: $\rho = \psi / (\psi + \theta)$

Two-Level Random Intercept Model

- Adding level-2 predictors

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} w_j + \zeta_j$$

- Reduced-form equation:

$$y_{ij} = \gamma_{00} + \beta_1 x_{ij} + \gamma_{01} w_j + \zeta_j + \varepsilon_{ij}$$

GLS Estimation of Linear Random Intercept Model

$$y_{ij} = \gamma_{00} + \beta_1 x_{ij} + \zeta_j + \varepsilon_{ij}$$

- Again, note that we're dealing with fixed β .
- Can be estimated via GLS and ML; both yield similar results.
- Foundation: GLS estimates of β_j are a weighted average of the pooled and within estimates of β_j .
 - *Partial pooling of coefficients.*

GLS Estimation of Linear Random Intercept Model

- Within, between, OLS, and GLS estimates:

$$\beta_W = W_{xx}^{-1} W_{xy}$$

$$\beta_B = B_{xx}^{-1} B_{xy}$$

$$\beta_{OLS} = (W_{xx} + B_{xx})^{-1} (W_{xy} + B_{xy})$$

$$\beta_{GLS} = (W_{xx} + \omega B_{xx})^{-1} (W_{xy} + \omega B_{xy})$$

$$\omega = \frac{\theta}{\theta + n\psi}$$

- n = cluster size; this equation assumes balanced structure (i.e., equal cluster sizes), though you can relax this for unbalanced structure.
- If $\omega = 0$, β_{GLS} reduces to β_W .
- If $\omega = 1$, β_{GLS} reduces to β_{OLS} .
- As cluster size increases, β_{GLS} becomes more similar to β_W .

Linear Random Intercept Model

- Goodness-of-fit measures:
 - Intraclass correlation coefficient, ρ
 - Testing RI model vs. pooled OLS ($H_0: \psi=0$)
 - Pooling factor
 - R^2 at each level
- R^2 : How much variance in the DV are we explaining at each level (R-H & S, 103)? Proportional reduction in error:

$$R_1^2 = \frac{\theta_0 - \theta_1}{\theta_0}$$

$$R_2^2 = \frac{\psi_0 - \psi_1}{\psi_0}$$

- Subscript 0: error variance (at each level) from model with no IVs.
- Subscript 1: error variance (at each level) from model with IVs.

Additional Measure of Partial Pooling

- A way to summarize the average degree of pooling, ω , for each random parameter (e.g., random intercept, random slopes) is suggested by Gelman and Hill:

$$\omega = 1 - \frac{\text{var}(\zeta_j^{EB})}{\text{var}(\zeta_j)}$$

- We can calculate this for each random parameter.
- What is the numerator and what is the denominator?
 - Numerator represents the variance of the *partially-pooled* residuals (at level 2, for each parameter)
 - Denominator: Level-2 error variance.

Interpretation of Effects for All Approaches

- How do we interpret effects from each approach?
- What do the pooled and RI approaches assume about the within- and between-cluster effects?

Within-Cluster v. Between-Cluster Variation

Student	School	Y	X1	X2	X3	X4
1	1	54	2	32	1	44
2	1	64	4	25	1	44
3	1	87	9	45	1	44
4	2	24	4	44	0	36
5	2	98	7	32	0	36
6	2	65	6	22	0	36
7	3	45	9	19	0	22
8	3	32	5	15	0	22
9	3	37	2	25	0	22
10	4	84	7	30	1	45
11	4	45	4	38	1	45
12	4	65	3	36	1	45
13	5	21	8	41	1	18
14	5	65	6	22	1	18
15	5	98	1	18	1	18

Considerations for the FE (Within) Estimator

- It's an easy way to account for unobserved heterogeneity in the response.
- Since the ζ_j are treated as fixed, instead of random, the potential for endogeneity between X and ζ_j (the controversial assumption) is eliminated.
- Consistent as N and $J \rightarrow$ infinity.
- More appropriate for inferring to clusters in sample only?

Issues:

- Overall efficiency loss by eliminating between group variation.
- FE cannot produce estimates for variables that are constant within clusters (level-2 vars; time-invariant in panel and TSCS data).
- Difficult to generate precise estimates for the effects of variables that contain small within-cluster variation.
 - This is a problem with the data, though, and not FE, per se.

Considerations for the Random Intercept Model

- More efficient than FE (minimum variance property)
- One can include variables that are constant within clusters (unlike the within estimator).
- Appropriate when inferring to *population* of clusters?
- Issue: Correlation between random effect and X at level 1.

IV. Cluster Confounding

Controversial Assumption in the RI Model

[Level-1 equation] $y_{ij} = \beta_{0j} + \beta_1 x_{ij} + \varepsilon_{ij}$

[Level-2 equation] $\beta_{0j} = \gamma_{00} + \zeta_{0j}$

Controversial Assumption

$$\text{Cov}(\zeta_{0j}, x_{ij}) = 0$$

- Issue: We need an accurate estimate of β_j . Note that this is fixed, so that β_{GLS} , β_w , β_B are all estimates of the same parameter, β .
- Note that β_{GLS} (and OLS) assumes that the between and within effects are the same (i.e., β).
 - Remember that x_{ij} varies both within and between clusters.
- But...the between effect could differ from the within effect for a variety of reasons.
- “Cluster confounding”; due to omitted variable(s) at level-2 (which are related to x_{ij}), β_1 could be confounded by conflicting between and within effects.
- **Cluster confounding** occurs when we’ve assumed the within and between effects are the same (by estimating a pooled or partially-pooled β), but they’re actually different.
 - Think about what it would take to eliminate the controversial assumption in the model above? Also, connection to ecological fallacy.

1. Hausman Test

- Tests the equality of coefficients b/w FE and RE model.
 - Essentially testing the “controversial assumption” and therefore, the existence of cluster confounding.
 - RI and FE are consistent (for β) if correctly specified. However, if we violate the “controversial assumption,” RI becomes inconsistent, while FE remains consistent.
 - If there is no cluster confounding, FE=RE.
 - Why? 2 possibilities?
 - R-H&S, p. 123

2. Accounting for Cluster Confounding

- We can solve this by estimating *both* between and within effects of β in the random intercept modeling framework (R-H & S, 113-19).
- For level-1 variables, generate a within-cluster and between-cluster operationalization.
- Recall:

$$x_{ij} = x_{ij}^W + \bar{x}_j$$

- Generating these operationalizations:

$$[\text{Between}] \quad \bar{x}_j = \frac{\sum_{i=1}^{N_j} x_{ij}}{N_j}$$

$$[\text{Within}] \quad x_{ij}^W = x_{ij} - \bar{x}_j$$

2. Accounting for Cluster Confounding

- Estimate both the within and between-cluster effects of x_{ij}

- **Method 1:**

$$y_{ij} = \beta_0 + \beta^W x_{ij}^W + \beta^B \bar{x}_j + \zeta_j + \varepsilon_{ij}$$

- What is the correlation now between the within-cluster x_{ij} and ζ_j ?

- **Method 2** (identical model, different interpretation):

$$y_{ij} = \beta_0 + \beta^W x_{ij} + \delta \bar{x}_j + \zeta_j + \varepsilon_{ij}$$

- Can perform Hausman-like test for equality of between and within estimates. δ represents the *difference* between with the within- and between-cluster effects.
- Importance: Highlights consequences for the models assuming within- and between-cluster effects are equal.

V. Panel / Time-Series Cross-Sectional Data

Panel / Time-Series Cross-Sectional Data

<i>i</i>	<i>j</i>	<i>t</i>	Y	X1	X2	X3	X4
1	1	1	54	2	32	1	44
2	1	2	64	4	25	1	44
3	1	3	87	9	45	1	44
4	2	1	24	4	44	0	36
5	2	2	98	7	32	0	36
6	2	3	65	6	22	0	36
7	3	1	45	9	19	0	22
8	3	2	32	5	15	0	22
9	3	3	37	2	25	0	22
10	4	1	84	7	30	1	45
11	4	2	45	4	38	1	45
12	4	3	65	3	36	1	45
13	5	1	21	8	41	1	18
14	5	2	65	6	22	1	18
15	5	3	98	1	18	1	18

Issues in TSCS Data

- Unobserved heterogeneity
- Pooling
- Temporal dependence
- Efficiency – standard errors
 - Panel heteroskedasticity (panels have different error variance)
 - Contemporaneous error correlation (errors related across countries for given years)
 - Serial correlation

Beck and Katz 1995

- Recommended using OLS with panel-corrected standard errors (PCSEs)
 - Serial correlation should be eliminated before estimation.
 - Adjusts SEs for panel heterosk. and contemporaneous correlation.
 - Like robust standard errors in OLS for cross-sectional data.
$$\text{var}(\mathbf{b}) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$$
 - $\mathbf{\Omega}$ is the same as in GLS.
 - The larger T is, the better the PCSEs are.
 - In Stata, “xtpcse”
- Article did not place emphasis on UH, *just standard error correction*; they also suggest AR-1 correction.
- *What kind of an approach is this?*

Beck and Katz 1996

- Now widely accepted in political science: Beck and Katz (1996); FE with lagged DV and PCSEs.
- Same issues apply to TSCS/panel data that we have talked about
 - Modeling approaches along pooling spectrum
 - Cluster confounding
- Primary difference is dynamics.

VI. Random Coefficient Model

Two-Level Random Coefficient Model

- Motivation for random coefficient model: *Causal heterogeneity.*

Two-Level Random Coefficient Model

[Level-1 equation] $y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}$

[Level-2 equations] $\beta_{0j} = \gamma_{00} + \zeta_{1j}$

$$\beta_{1j} = \gamma_{10} + \zeta_{2j}$$

- We'll start simple; no level-2 covariates.

- Reduced form: $y_{ij} = (\gamma_{00} + \zeta_{1j}) + (\gamma_{10} + \zeta_{2j})x_{ij} + \varepsilon_{ij}$

$$y_{ij} = \gamma_{00} + \zeta_{1j} + \gamma_{10}x_{ij} + \zeta_{2j}x_{ij} + \varepsilon_{ij}$$

$$y_{ij} = \gamma_{00} + \gamma_{10}x_{ij} + \zeta_{1j} + \zeta_{2j}x_{ij} + \varepsilon_{ij}$$

$$\text{Var}(\zeta_{1j}) = \psi_{11}$$

$$\text{Var}(\zeta_{2j}) = \psi_{22}$$

$$\text{Cov}(\zeta_{1j}, \zeta_{2j}) = \psi_{21}$$

Testing the Adequacy of the Random Coefficient Model

- In ML, we can do a likelihood ratio test to test the statistical significance of the random coefficient specification.
 - Compare our full model with random intercept and random coefficient specification to a reduced model with only a random intercept specification.
 - LR test: generate a chi-square statistic, which is $2*(LL_F - LL_R)$, which is the same as $-2*(LL_R - LL_F)$.
 - $H_0: \psi_{22} = \psi_{21} = 0$ [no causal heterogeneity across clusters]
 - Specify full model (RC) first, then reduced (RI). Use Stata's "lrtest" command.
 - Or do it manually; generate chi-squared stat, then use "chprob" to get p-value (note: 2 degrees of freedom in this test.).

Random Coefficient Model

$$y_{ij} = (\gamma_{00} + \zeta_{1j}) + (\gamma_{10} + \zeta_{2j})x_{ij} + \varepsilon_{ij}$$

- Think about what this means:
 - The random effect for the intercept represents a cluster's deviation from the "mean intercept." [more specifically.....]
 - The random effect for the slope represents a cluster's deviation from the "mean slope."
- Mean-centering: The random intercept actually represents the cluster average when $x=0$. To make substantive interpretations of the random intercept part, we can mean-center the x 's at level 1.
 - Then, the intercept represents the average cluster level of the DV, since it's for a typical value of x .

Random Coefficient Model

- Generating empirical Bayes residuals at level 2 for the slope and intercept.
- Using these to calculate partially-pooled slopes and intercepts.

Estimation

- Note for linear models, differences in estimation procedures is not a huge deal; it's a bigger deal for nonlinear models.
- Estimation techniques:
 - GLS
 - Maximum likelihood
 - Restricted ML (REML)
 - All three are asymptotically equivalent

Maximum Likelihood Estimation

- Conditional distribution of the response (conditional on the random effects):
 $g^{(1)}(Y | X, \zeta; \theta)$ where $\zeta \sim \text{MVN}(0, \Sigma)$
- *Goal*: Obtain the unconditional (marginal) distribution of the response for each cluster j by integrating out the random effects:

$$f^{(2)}(Y | X; \theta) = \int h(\zeta) \prod_{i=1}^{n_j} g^{(1)}(Y | X, \zeta; \theta) d\zeta$$

- Marginal likelihood:

$$L = \prod_{j=1}^J f^{(2)}(Y | X; \theta)$$

- Or:

$$L = \prod_{j=1}^J \int h(\zeta) \prod_{i=1}^{n_j} g^{(1)}(Y | X, \zeta; \theta) d\zeta$$

Maximum Likelihood Estimation

- For linear model: there's a closed form solution to integral.
 - For nonlinear models, integral is approximated using quadrature
- Use EM algorithm to maximize the likelihood
 - Mutual dependence of estimates for fixed effects and variance components
 - Iterative procedure; alternate between “expectation” and “maximization” steps

Differences b/w ML and REML

- Differences are minor, and primarily center on calculating the variance components.
- In REML, likelihood function not directly applied to the response, Y . Instead, the restricted likelihood is the full likelihood with the variance components only and the fixed effects swept out. Fixed effects (coefficients) estimated in second step.
- REML accounts for the loss of degrees of freedom due to estimation of parameters; generates unbiased estimates of the variance components.
- ML estimates do not account for this loss of df, and they are consistent. There'll be a downward bias of ψ in small samples (particularly small number of clusters).
- This is analogous to OLS versus ML estimates of error variance in linear regression.

Summarizing the Degree of Pooling

- A way to summarize the average degree of pooling, λ , for each random parameter (e.g., random intercept, random slopes) is suggested by Gelman and Hill:

$$\omega = 1 - \frac{\text{var}(\zeta_j^{EB})}{\text{var}(\zeta_j)}$$

- We can calculate this for each random parameter.
- What is the numerator and what is the denominator?
 - Numerator represents the variance of the *partially-pooled* residuals (at level 2, for each parameter)
 - Denominator: Level-2 error variance.
- Measures of R^2

VII. Multilevel Models for Binary Dependent Variables

Binary Responses: GLM Specification

- Hierarchical generalized linear models (HGLM):
 1. Specify the sampling model for the dependent variable.
 2. Specify link function (first, conditional expectation of response; then, the link is the inverse of that)
 3. Specify structural model; model link as a linear function of independent variables.
- For linear models, we don't really need to think in a GLM format:
$$Y_i = b_0 + b_1 X_i + e_i$$
- But we could: Normal sampling model, identity link
[$\mu_i = E(Y_i | X_i) = x' \beta$]

GLM Specification

- Utility of HGLM: For nonlinear models.
- Example for binary DVs
 1. Bernoulli sampling model
 2. Logit link:
 - a. Conditional expectation of response:
$$E(Y_i | X_i) = \mu_i = \Pr(Y_i=1 | X_i) = \exp(x' \beta) / [1 + \exp(x' \beta)]$$
 - b. Link is the inverse of μ_i
$$\eta_i = \log[\mu_i / (1 - \mu_i)] \quad [\text{log-odds}]$$
 3. Structural model: write η_i as a linear function of level-1

Estimation

- See handout
- `gllamm` or `xtmelogit` (uses quadrature)

Estimating Multilevel Models with Binary Responses in Stata

- **xtlogit** and **xtprobit**: random intercept models only (default is adaptive quadrature, 12 points)
- **xtmelogit**: RI and RC logit model (default is adaptive quadrature, 7 points)
- **gllamm**: add-on package to Stata (created by Rabe-Hesketh and Skrondal); estimates RI and RC models for all types of DVs (continuous, binary, ordinal, count, duration, nominal); uses quadrature and adaptive quadrature.
 - To install, type (in Stata): `ssc install gllamm`
- See handout on `gllamm`.
- Using good start values and increasing number of quadrature points.

Generating Quantities of Interest

- In gllamm, use the “**gllapred**” command (after specifying a gllamm model)
- To retrieve empirical Bayes residuals:
gllapred eb, u
- For an RI model, this will generate two variables: ebm1 and ebs1.
 - ebm1 is the empirical Bayes residuals (like what we get with “reflects” in the canned xtmixed and xtmelogit).
 - ebs1 is the s.e. of the EB residuals
- For an RC model, the command will generate four variables: ebm1, ebs1, ebm2, and ebs2. “1” is for random intercept, “2” is for random slope.

Probabilities in Binary Response Models

- Two different brands of predicted probabilities:
 - **Cluster-specific probabilities**
 - Takes into account the clustering, hierarchical structure in the data.
 - RI and RC models fit cluster specific probabilities: $\Pr(Y=1 | \zeta, x)$
 - **Marginal, or “population-averaged,” probabilities**
 - Marginal with respect to the random effects; plain-vanilla logit and probit produces marginal, PA probabilities: $\Pr(Y=1 | x)$. They don’t depend on ζ , because we’re not modeling it.
 - To generate marginal probabilities from an RI or RC model, need to integrate out ζ , as on p. 254 (eq. 6.7).
- See page 255 in RH&S; difference between cluster-specific and marginal probabilities.

Probabilities in Binary Response Models

- Generating **cluster-specific probabilities** after running a gllamm model; use the “**gllapred**” command.

```
gllapred cs_prob, mu
```

- This will generate a predicted prob for observations in the sample, using the cluster’s particular ζ . See pp. 269-70.
- Generating **marginal, or PA, probabilities** after running a gllamm model:

```
gllapred marg_prob, mu marginal
```