Design and Analytic Implications in Modeling Student Mobility Across Correlated Schools

Tessa L. Johnson, Yi Feng, & Laura M. Stapleton

Society for Research in Educational Effectiveness, March, 2019

UNIVERSITY OF MARYLAND

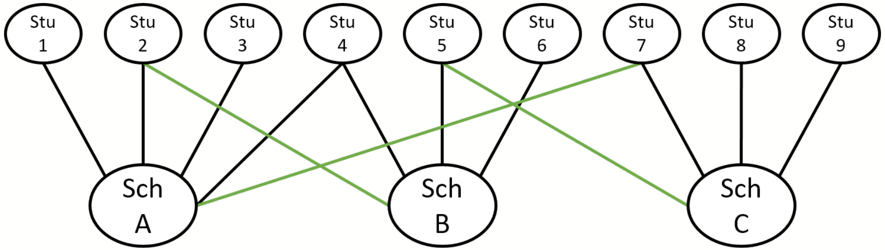
Acknowledgement

- We are grateful for the data, technical, and research support provided by the MLDS Center and its agency partners. The views and opinions expressed are those of the authors and do not necessarily represent the views of the MLDS Center or its agency partners. The contents of this presentation were developed under a grant from the Department of Education. However, these contents do not necessarily represent the policy of the Department of Education, and you should not assume endorsement by the Federal Government.

Introduction

Longitudinal, multilevel education studies provide a wealth of information with implications for program evaluation and policy.

- These data are often quite complex in terms of their nesting structures (e.g., multiple membership)

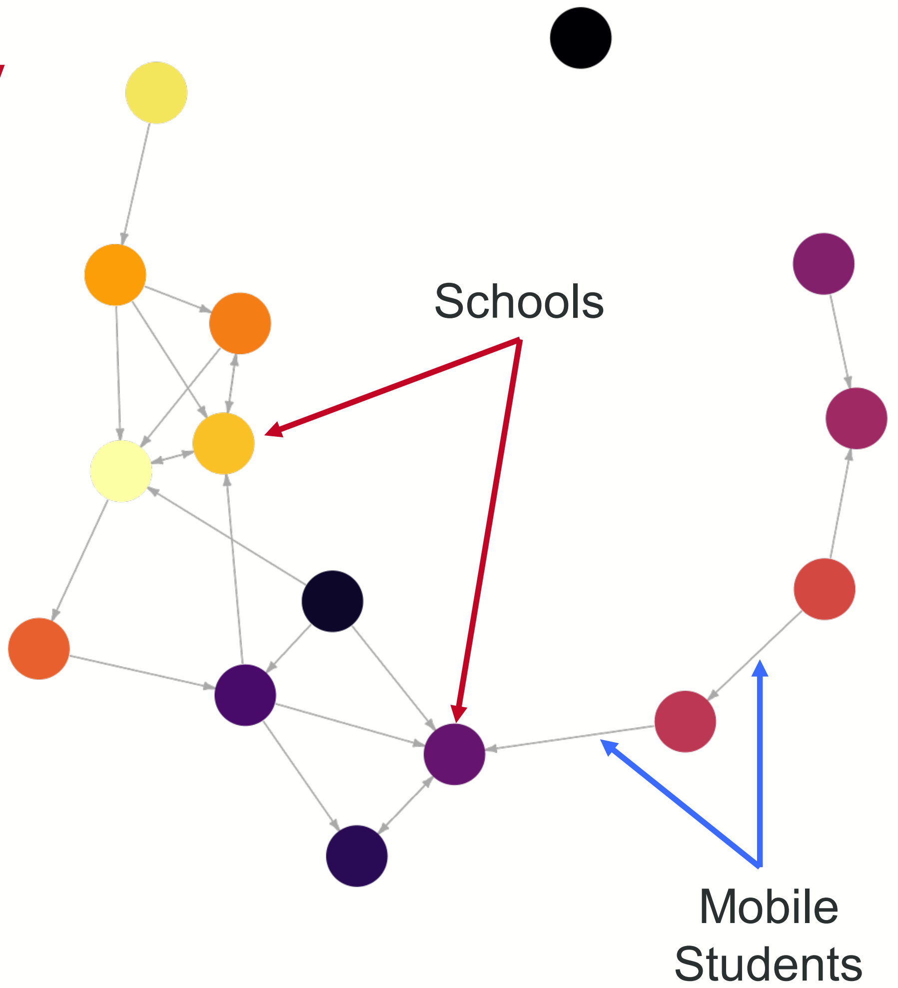


- With the improvement of administrative record systems, researchers may track students as they move across schools, either within or often even outside of the study sample

Patterns of Mobility

Students are mobile...but in a particular way

- Investigations of student mobility have found that clusters of schools form, passing students back and forth (Kerbow, 1996; Kerbow, Azcoitia, & Buell, 2003)



What if mobility isn’t of interest?

Current best-practices for modeling student outcomes when students are mobile across multiple schools indicate the use of multiple membership models (Chung & Beretvas, 2012; Wolff Smith & Beretvas, 2017).

- These models make strong assumptions about the pattern, or lack thereof, in student mobility

- Empirical analyses from a statewide longitudinal data system support previous evidence that students do not move randomly from school to school

Empirical Correlations: HS Algebra

School residuals were calculated from a null model estimated on nonmobile students only. Correlations among residuals were then calculated between first and second, second and third, and first and third schools attended by mobile students.

Correlations Among School Residuals (n=266)	**1. n = 15926**	**2. n = 15185**	**3. n = 3902**
1. First School Attended	—		
2. Second School Attended	0.432	—	
3. Third School Attended	0.359	0.375	—

Empirical Correlations: SAT Math

School residuals were calculated from a null model estimated on nonmobile students only. Correlations among residuals were then calculated between first and second, second and third, and first and third schools attended by mobile students.

Correlations Among School Residuals (n=266)	**1. n = 15926**	**2. n = 15185**	**3. n = 3902**
1. First School Attended	—		
2. Second School Attended	$\boxed{0.479}$	—	
3. Third School Attended	0.396	0.392	—

Impact of Level-2 Cluster Correlation

Findings from empirical analyses reveal relatively large inter-school correlations, which impacts relevant modeling outcomes, such as ICC and level-2 variance.

Inter-School Correlation	**Level-2 Variance**	**ICC**	**Composite ICC (across % mobility)**		
			10%	**25%**	**50%**
Nonmobile	1.00	*0.314*	—	—	—
Mobile (0.0)	0.50	0.187	0.302	0.282	0.251
Mobile (0.2)	0.60	0.216	0.305	0.290	0.265
Mobile (0.5)	0.75	0.256	0.309	0.300	0.285

20% decrease

9% decrease

Impact of Level-2 Cluster Correlation

Findings from empirical analyses reveal relatively large inter-school correlations, which impacts relevant modeling outcomes, such as ICC and level-2 variance.

Inter-School Correlation	Level-2 Variance	ICC	Composite ICC (across % mobility)			
			10%	25%	50%	
Nonmobile	1.00	*0.048*	—	—	—	
Mobile (0.0)	0.50	0.024	0.045	0.042	0.036	25% decrease
Mobile (0.2)	0.60	0.029	0.046	0.043	0.038	
Mobile (0.5)	0.75	0.036	0.046	0.045	0.042	13% decrease

Multiple Membership Model

$$\boldsymbol{\omega} \sim N\left(Z_{\mathrm{W}} \cdot \boldsymbol{\beta},\ \tau_{00} \right)$$

$$\boldsymbol{y} \sim N\left(\boldsymbol{\omega} + \boldsymbol{X} \cdot \boldsymbol{\gamma},\ \sigma^2 \right)$$

$\mathbf{Z_W}$ - weighted level-2 covariate matrix (weights sum to 1)

$\boldsymbol{\beta}$ - level-2 coefficient vector

$\boldsymbol{\tau}_{00}$ - variance of level-2 residuals

$\mathbf{X}$ - level-1 design matrix (covariates and constant)

$\boldsymbol{\gamma}$ - level-1 coefficient vector

$\boldsymbol{\sigma}^2$ - variance of level-1 residuals

Multiple Membership Model

$$\boldsymbol{\omega} \sim N\left(Z_{\mathbf{W}} \cdot \boldsymbol{\beta},\ \tau_{00} \right)$$

$$\boldsymbol{y} \sim N\left(\boldsymbol{\omega} + \boldsymbol{X} \cdot \boldsymbol{\gamma},\ \sigma^2 \right)$$

- Weights are often assigned (not estimated) as 1/*H*, where *H* is the number of schools attended by student *i*

- A naive, first-school approach is a special case of this model where the first school is given a weight of 1 and subsequent school weights are set at 0

- $\mathbf{Z_W}$ is constructed as $w_{i,1} * z_{p,1} + \ldots + w_{i,H} * z_{p,H}$ - assumes 0 correlation between schools

Assessing Model Robustness

Research Question:

Is the multiple membership random effects model (MMREM) robust to violations of the independence assumption among mobile students' schools?

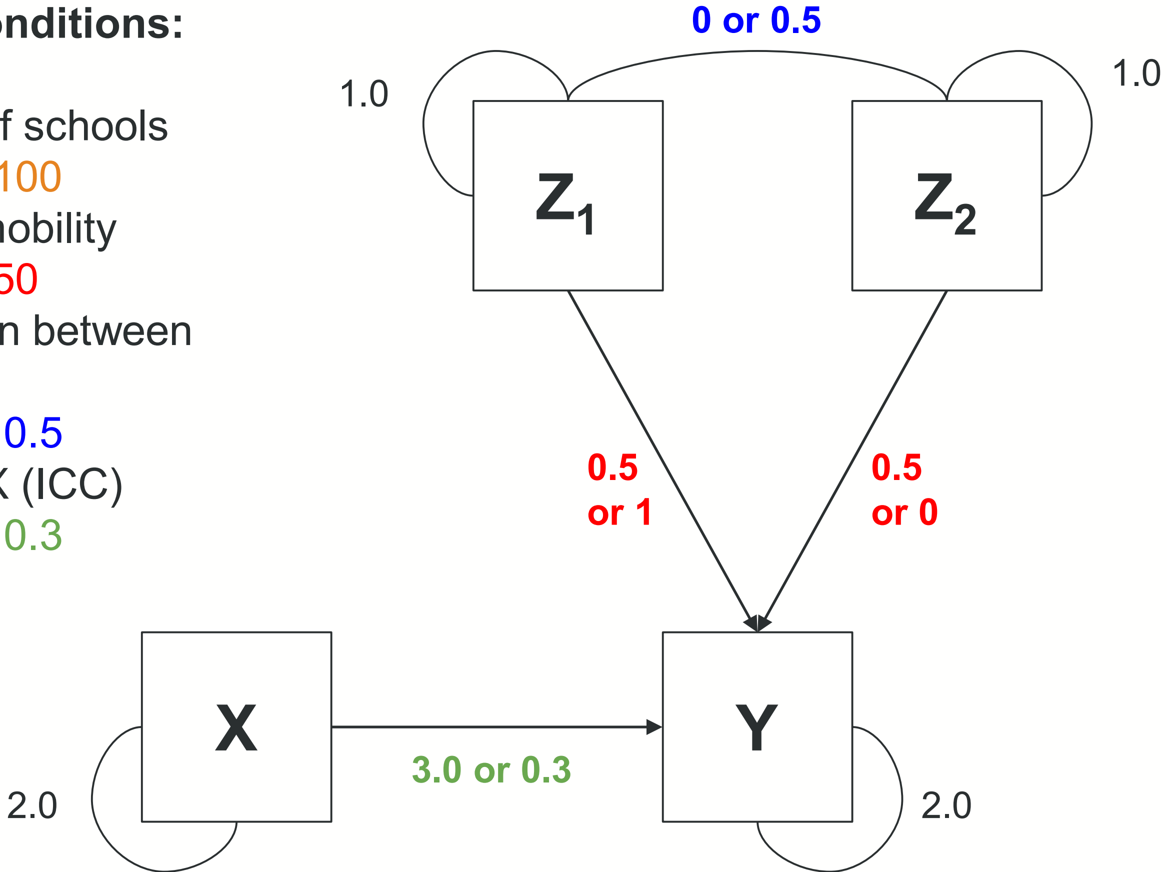
Secondary Question:

If the model is not robust, what alternate strategies can researchers employ in the design or modeling phase to improve parameter coverage?

Simulation: Data-Generating Model

Simulation Conditions:

- Number of schools
  - 50 / 100
- Percent mobility
  - 25 / 50
- Correlation between schools
  - 0.0 / 0.5
- Effect of X (ICC)
  - 3.0 / 0.3



Relative Parameter Bias

Where do the models fail?

- Low ICC
- High percent mobility
- High correlation between schools

Level-2 Variance Component

N Sch	% Mob	Sch Corr	ICC	HLM-First	MMREM
50	25%	0.0	5%	**-0.260**	-0.014
50	25%	0.0	30%	**-0.196**	0.042
50	25%	0.5	5%	**-0.178**	0.014
50	25%	0.5	30%	**-0.120**	0.012
50	50%	0.0	5%	**-0.457**	-0.018
50	50%	0.0	30%	**-0.379**	**0.056**
50	50%	0.5	5%	**-0.264**	**0.127**
50	50%	0.5	30%	**-0.197**	**0.052**
100	25%	0.0	5%	**-0.235**	-0.009
100	25%	0.0	30%	**-0.215**	0.018
100	25%	0.5	5%	**-0.124**	**0.060**
100	25%	0.5	30%	**-0.102**	0.026
100	50%	0.0	5%	**-0.424**	0.010
100	50%	0.0	30%	**-0.398**	0.020
100	50%	0.5	5%	**-0.226**	**0.156**
100	50%	0.5	30%	**-0.219**	0.020

Relative Std. Error Bias

Where do the models fail?

- High ICC
- Mid & high percent mobility
- High correlation between schools
- High level-2 sample size

Level-2 Variance Component

N Sch	% Mob	Sch Corr	ICC	HLM-First	MMREM
50	25%	0.0	5%	-0.056	-0.027
50	25%	0.0	30%	-0.010	-0.005
50	25%	0.5	5%	-0.098	-0.042
50	25%	0.5	30%	**-0.229**	**-0.173**
50	50%	0.0	5%	-0.075	-0.051
50	50%	0.0	30%	0.003	0.075
50	50%	0.5	5%	**-0.171**	-0.060
50	50%	0.5	30%	**-0.236**	**-0.123**
100	25%	0.0	5%	-0.094	**-0.102**
100	25%	0.0	30%	-0.044	-0.024
100	25%	0.5	5%	**-0.161**	**-0.152**
100	25%	0.5	30%	**-0.242**	**-0.194**
100	50%	0.0	5%	**-0.104**	-0.065
100	50%	0.0	30%	-0.057	-0.009
100	50%	0.5	5%	**-0.146**	-0.087
100	50%	0.5	30%	**-0.264**	**-0.157**

Results Summary

- Consistent with previous findings, fixed effects parameters and level-1 variance components were not impacted by multiple membership, even when inter-cluster correlations were high
- MMREMs over-correct level-2 variance estimates when level-2 units are correlated (positive parameter bias), while naïve HLMs under-correct
- Including level-2 covariates that are highly predictive of school residual correlations doesn’t reduce the parameter bias, though it does improve standard error estimation bias

The Path Forward

- Even if mobility is not a variable of interest, it still has impacts on student outcomes

- Further, the correlations between mobile students' schools will positively bias variance component estimates, even when estimated using MMREMs

- Future research will explore explicitly accounting for inter-school correlations in MMREM formulation

- Large-scale studies should make every effort to track students across schools; studies with large numbers of schools are not immune

References

- Chung, H., & Beretvas, S. N. (2012). The impact of ignoring multiple membership data structures in multilevel models. *British Journal of Mathematical and Statistical Psychology*, *65*(2), 185-200.
- Kerbow, D. (1996). Patterns of urban student mobility and local school reform. *Journal of Education for Students Placed at Risk*, *1*(2), 147-169.
- Kerbow, D., Azcoitia, C., & Buell, B. (2003). Student mobility and local school improvement in Chicago. *Journal of Negro education*, 158-164.
- Smith, L. J. W., & Beretvas, S. N. (2017). A comparison of techniques for handling and assessing the influence of mobility on student achievement. *The Journal of Experimental Education*, *85*(1), 3-23.

UNIVERSITY OF MARYLAND

Tessa L. Johnson
3942 Campus Drive, College Park, MD 20742
johnsont@umd.edu
@tessajolee