Design and Analytic Implications in Modeling Student Mobility Across Correlated Schools

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

University of Maryland, College Park

Research Purpose

Longitudinal, multilevel education studies provide a wealth of information with implications for program evaluation and policy. However, these data are often quite complex, and little empirical evidence exists to guide the design and analysis of such studies. Current best-practices for analysis of these data structures make strong assumptions about the purity of hierarchical clustering and the lack of correlation among schools attended by mobile students. In the present study, we violate these assumptions and investigate the effect of student mobility across correlated schools on the parameter estimates of two modeling approaches: (1) a multiple membership random effects model (MMREM), and (2) a first-school hierarchical linear model (HLM). In addition, we examine data from a statewide longitudinal database to provide researchers with a gauge for the magnitude of the correlation among schools attended by mobile students, which may aid in the design of future studies and analysis plans.

Background

MMREMs (Beretvas, 2011; Browne, Goldstein & Rasbash, 2001; Rasbash & Browne, 2001) are a set of models that extend the HLM framework to account for impure nesting within hierarchical structures. For example, multiple membership is a form of impure clustering that occurs when a mobile individual belongs to more than one cluster at the same clustering level over a given period of observation (Browne, Goldstein, & Rabash, 2001; see Figure 1 for a depiction). The presence of multiple membership has been shown to bias the estimates of cluster-level variance components in multilevel models if a student's mobility is not properly accounted for (Chung & Beretvas, 2012; Leroux & Beretvas, 2018; Wolff Smith & Beretvas, 2015). Evidence indicates that MMREMs are robust to model misspecification, such as in the case of researcher-specified weights (Wolff Smith & Beretvas, 2014); however, little is known about the impact of non-zero correlations among the clusters of mobile individuals on the performance of either the MMREM or the HLM. The standard MMREM assumes that a random process governs the cluster membership pattern of mobile individual *i*; that is, if mobile student 2 attends school A, she is equally likely to attend School B or School C next. Nevertheless, previous longitudinal evidence indicates that this is not the case: disadvantaged students are more likely to transfer to low-performing schools while advantaged students are more likely to transfer to high-performing ones (Xu, Hannaway, & D'Souza, 2009). To that end, the current paper examines the effect of a possible correlation among level-2 units on parameter estimates in multilevel models using a simulation study and provides estimates of this correlation from a statewide longitudinal database.

Research Design

Simulation study. The following conditions were manipulated to examine the effects of correlated level-2 clustering on model parameter estimates (Table 1; replications = 100).

Cluster correlation (R0 & R1). A normally distributed random covariate was simulated to account for 80% of the correlation among level-2 unit random effects. Two level-2 unit correlation matrices were set: R0 was a J x J (where J = total number of level-2 units) diagonal matrix with 0.8 along the diagonal, and R1 was a J x J matrix with 0.8 along the diagonal, 0.5 along the first off-diagonal, and 0.2 along the second off-diagonal. All other schools were uncorrelated.

ICC (I0 & I1). Following Wolff Smith and Beretvas (2014), the intra-class correlation was set to either 15% for the "low" condition or 25% for the "high" condition. Preliminary results are presented for the low condition. The high condition will be presented in the final analysis.

Number of level-2 units (J0 & J1). The number of level-2 units was simulated to be either 50 or 100, following recommendations for no fewer than 50 level-2 units (Chung & Beretvas, 2012). Preliminary results are presented for the 100-unit condition. The 50-unit condition will be presented in the final analysis.

Percent mobility (P0 & P1). Percent mobility was set at 10% for the "low" condition and 20% for the "high" condition following methodological work by Chung & Beretvas (2012).

Estimation model (H & M). The simulated data were estimated either using the an unconditional, random intercept first-school HLM model (H) or MMREM model (M).

Empirical analysis. Data from a statewide longitudinal database were examined to estimate the correlation among schools attended by mobile high school students. A random intercept HLM model was estimated to generate level-2 random effects on purely nested students only. The correlation among schools' random effects was then calculated for students attending multiple high schools. High-stakes test scores were used as the student-level outcome variable. Results of the empirical analysis are forthcoming.

Preliminary Findings

Boxplots comparing the distribution of the parameter estimates for the uncorrelated (red) and correlated (blue) cluster conditions for the level-2 intercept fixed effect (Figure 2) and the level-2 intercept variance component (Figure 3) are presented below. In Figure 2, across correlated cluster conditions (blue), regardless of whether the model was estimated with HLM or

MMREM, the variance in the level-2 intercept fixed effect estimates is much larger than for the uncorrelated cluster conditions (red). In Figure 3, the first-school HLM approach actually outperforms the MMREM approach in terms of parameter bias when clusters are correlated and mobility rates are held constant.

Conclusions

The primary aims of this study were (1) to provide estimates of the correlation among the random effects of schools attended by mobile students, and (2) to investigate the impact of those correlated school-level random effects on model parameter estimates using a rigorous simulation study. The results of aim 1 will be presented in the final study. Preliminary results of aim 2, discussed above, provide intriguing evidence that correlation among level-2 units may adversely impact parameter estimation of the MMREM model. By pairing these simulation results with longitudinal data from a statewide database, we provide useful criteria for both the design and analysis of studies with impure hierarchical structures caused by participant mobility.

References

Beretvas, S. N. (2011). Cross-classified and multiple-membership models. *Handbook of advanced multilevel analysis*, 313-334.

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, *1*(2), 103-124.

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,* 185-200.

Leroux, A. J. & Beretvas, S. N. (2018). Estimation of a latent variable regression growth curve model for individuals cross-classified by clusters. *Multivariate Behavioral Research, 53*(2), 231-246.

Rasbash, J., & Browne, W. J. (2001). Modelling non-hierarchical structures. In A. Leyland & H. Goldstein (Eds.), *Multilevel modelling of health statistics* (pp. 93-105). John Wiley & Sons.

Wolff Smith, L. J. & Beretvas, S. N. (2014). The impact of using incorrect weights with the multiple membership random effects model. *Methodology, 10*, 31-42.

Wolff Smith, L. J., & Beretvas, S. N. (2015). A comparison of techniques for handling and assessing the influence of mobility on student achievement. *The Journal of Experimental Education, 85*(1), 3-23.

Xu, Z., Hannaway, J., & D'Souza, S. (2009). Student Transience in North Carolina: The Effect of School Mobility on Student Outcomes Using Longitudinal Data. Working Paper 22. *National Center for Analysis of Longitudinal Data in Education Research*.

Table 1. *Simulation Conditions (replications = 100)*

Condition	Number of Levels	Levels
Cluster Correlation	2	R0 = Uncorrelated Clusters
		R1 = Correlated Clusters
ICC	2	I0 = 15%*
		I1 = 25%
Number of Level-2 Units	2	J0 = 50
		J1 = 100*
Percent Mobility	2	P0 = 10% Mobility
		P1 = 20% Mobility
Estimation Model	2	H = HLM* (first school)
		M = MMREM*

Note. HLM = hierarchical linear model. MMREM = multiple membership random effects model. Conditions marked with an * are proposed and will appear in the final analysis; preliminary results for all other conditions are presented in text.

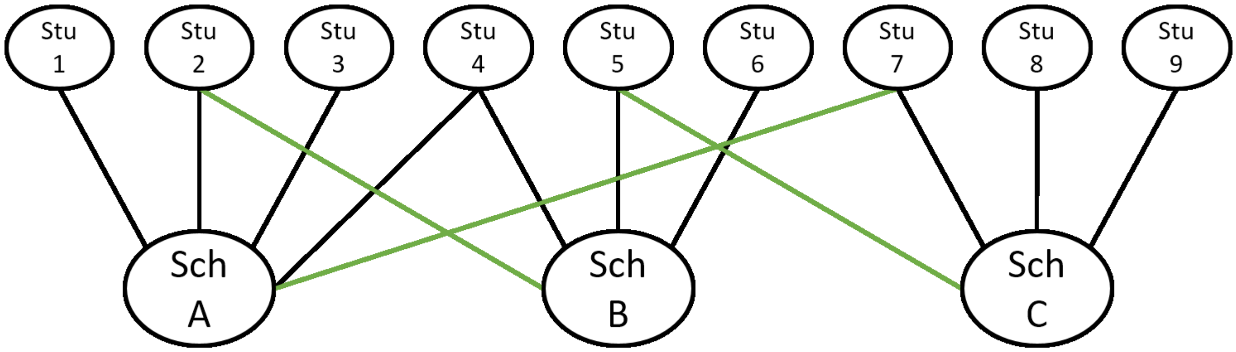

Figure 1. Depiction of a multiple membership network. Of the nine students (Stu 1 – Stu 9) presented, three (Stu 2, Stu 5, and Stu 7) are non-hierarchically nested within schools (Sch A – Sch C) such that, for example, Stu 2 belongs to both Sch A and Sch B. Non-hierarchical memberships are depicted in green.

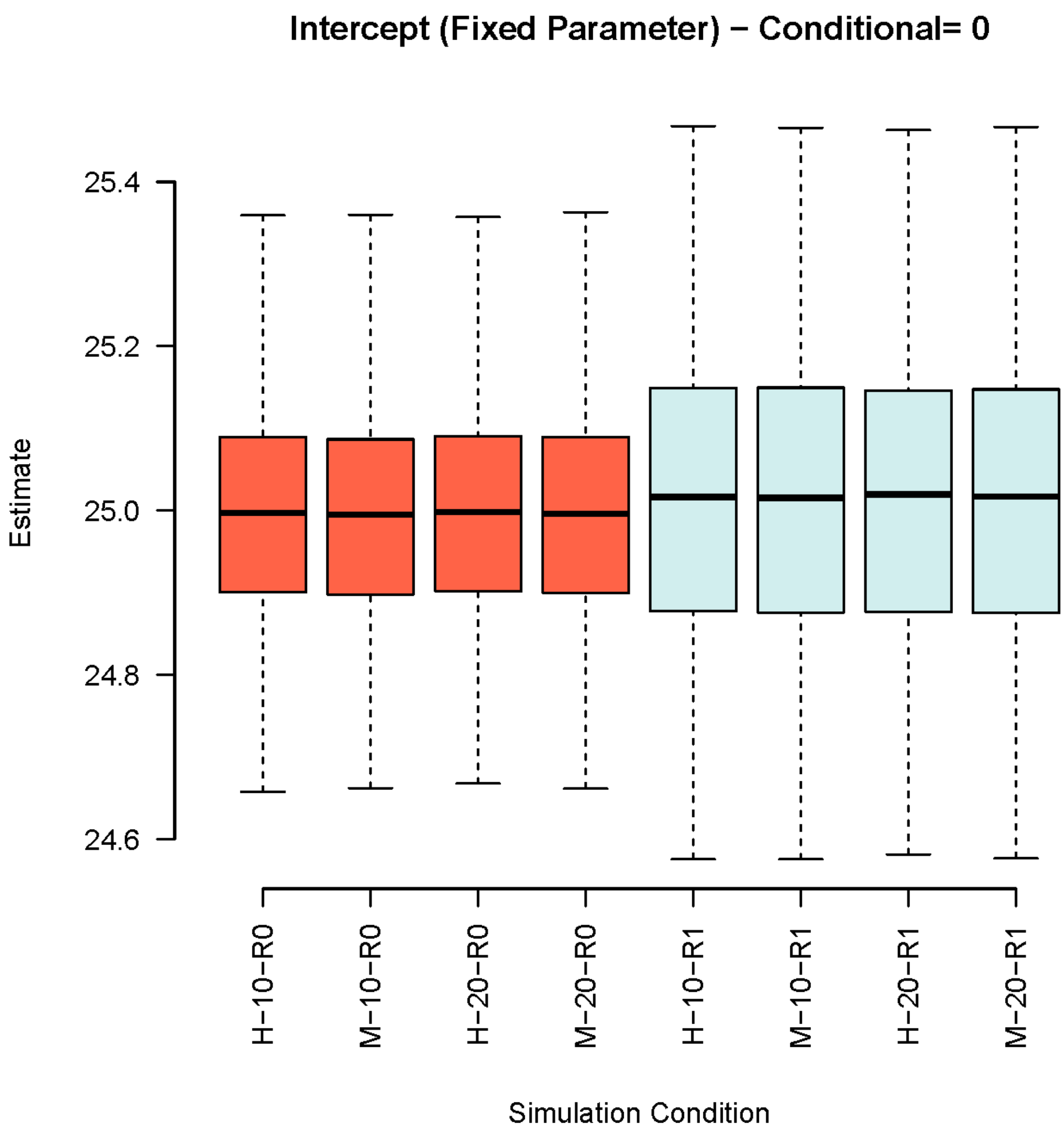

Figure 2. Preliminary results of the Monte Carlo simulation for the unconditional estimation of the level-2 intercept fixed effect. Uncorrelated cluster conditions are depicted in red; correlated cluster conditions are depicted in blue. (*Note.* H=HLM (first school), M=MMREM, 10=10% mobility, 20=20% mobility, R0=uncorrelated clusters, R1=correlated clusters.)

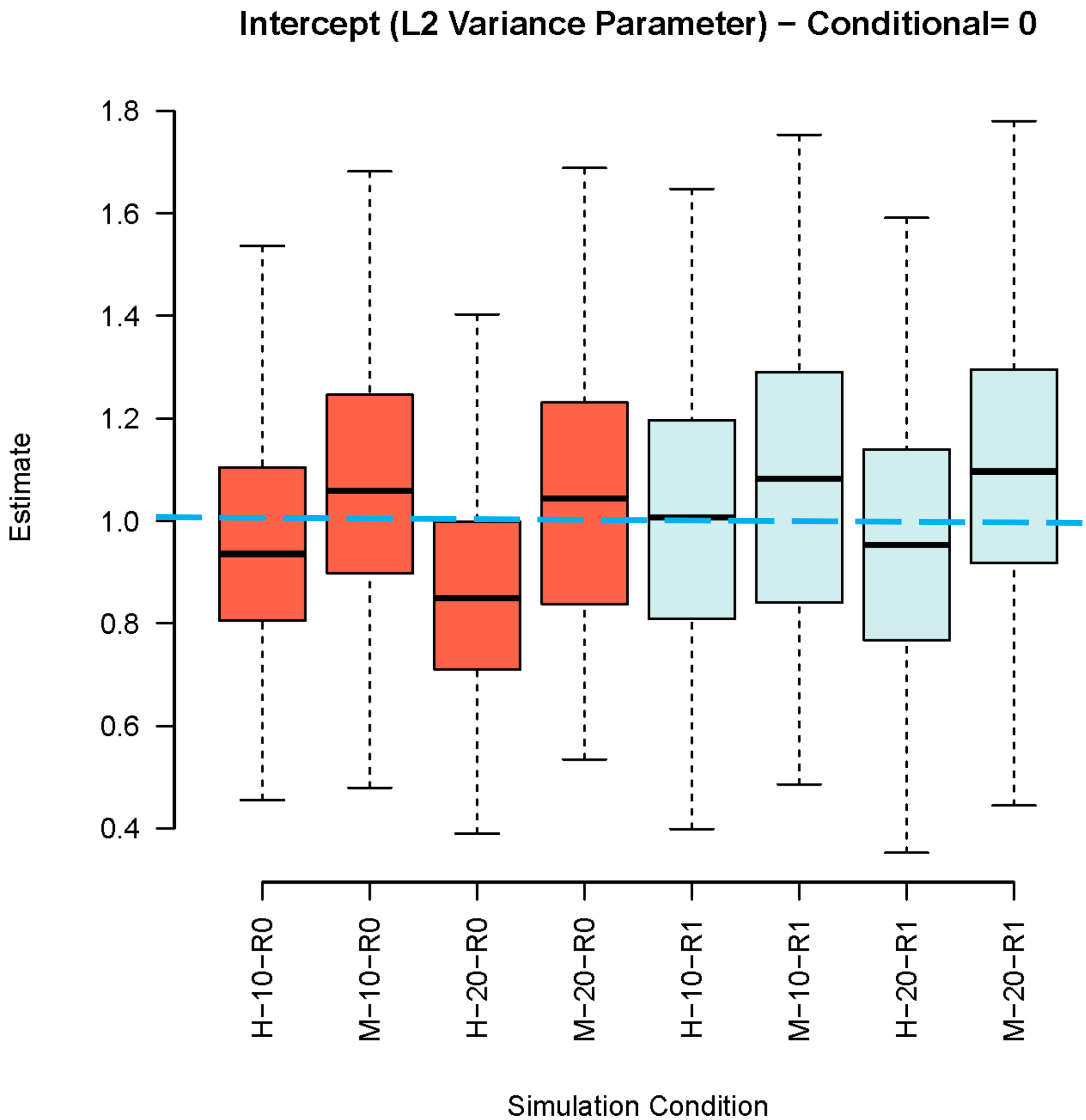

Figure 3. Preliminary results of the Monte Carlo simulation for the unconditional estimation of the level-2 intercept variance component. Uncorrelated cluster conditions are depicted in red; correlated cluster conditions are depicted in blue. True value of the estimate is 1.0, marked with a blue dashed line. (*Note.* H=HLM (first school), M=MMREM, 10=10% mobility, 20=20% mobility, R0=uncorrelated clusters, R1=correlated clusters.)