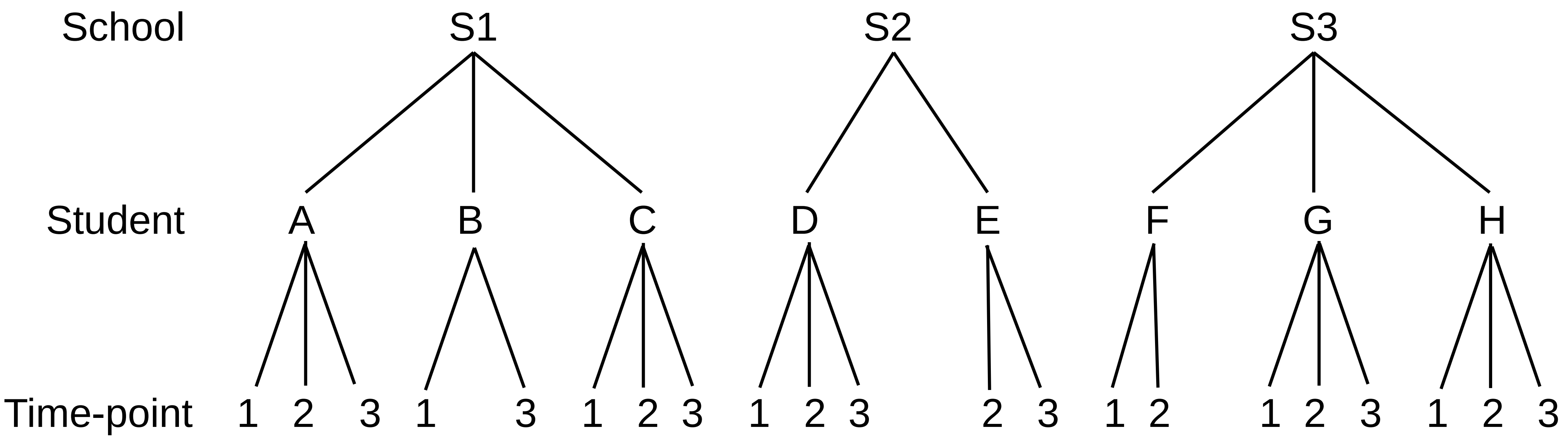
A New Way for Handling Student Mobility with Longitudinal Data in Educational Research

Congying Sun
Audrey J. Leroux
Christopher Cappelli

Georgia State University

Three-Level Growth Curve Model (GCM)

- A three-level GCM can estimate the within-individual variability and contextual effects on individual patterns of change over time.
- E.g., repeated measures nested within students nested within schools:



Baseline Three-Level GCM

Level 1: $Y_{tij} = \pi_{0ij} + \pi_{1ij} TIME_{tij} + e_{tij}, \qquad e_{tij} \sim N(0, \sigma_e^2)$

Level 2: $\begin{cases} \pi_{0ij} = \beta_{00j} + r_{0ij} \\ \pi_{1ij} = \beta_{10j} + r_{1ij} \end{cases}, \qquad \begin{bmatrix} r_{0ij} \\ r_{1ij} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{r0}^2 & \\ \sigma_{r10} & \sigma_{r1}^2 \end{bmatrix} \right)$

Level 3: $\begin{cases} \beta_{00j} = \gamma_{000} + u_{00j} \\ \beta_{10j} = \gamma_{100} + u_{10j} \end{cases}, \qquad \begin{bmatrix} u_{00j} \\ u_{10j} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u0}^2 & \\ \sigma_{u10} & \sigma_{u1}^2 \end{bmatrix} \right)$

Longitudinal Data with Multiple Membership

- Individual mobility in longitudinal studies leads to multiple membership data structure, which challenges the conventional three-level GCM.
- **Multiple membership data structure**: *Some* units of a lower-level classification (e.g., student) are members of more than one higher-level classification (e.g., school).

	School		
Student	6th Grade	7th Grade	8th Grade
A	S1	S1	S1
B	S1	S2	S2
C	S1	S1	S2
D	S1	S2	S1
E	S1	S2	S3

Handling Individual Mobility in Longitudinal Data

- Common approaches:
  - ***Delete***: delete mobile students from data, and conduct analyses only with students who have stayed in the same school throughout the study.
  - ***Single school***: keep data for mobile students, but model only one of the set of schools (e.g., first or last school) that students attended.
- Misspecification of multiple membership longitudinal data structures can lead to inaccurate estimates of between-cluster growth variance components and cluster-level fixed effects on the growth (Grady, 2010; Grady & Beretvas, 2010; Leroux, in press; Leroux & Beretvas, 2018, in press).

Handling Individual Mobility in Longitudinal Data

- Models proposed in these prior studies:
  - Can only be estimated if *TIME* is coded so that the intercept represents initial status; ***and***
  - Assume noncumulative school effect on student growth; ***or***
  - Cannot be estimated using MLwiN, which is the only software package that can estimate multiple membership random effects models.

Purpose of Current Study

- Therefore, we propose a multiple membership GCM (MM-GCM) to handle student mobility in longitudinal studies.
- This model can be estimated with the **intercept representing final status**.
  - Researchers and educators might be more concerned about students' final status and the contextual effects on the final status.
- The proposed MM-GCM will be derived, justified, and explained using a large-scale longitudinal dataset.

Baseline MM-GCM

Level 1: $Y_{ti\{j\}} = \pi_{0i\{j\}} + \pi_{1i\{j\}}\boxed{TIME_{ti\{j\}}} + e_{ti\{j\}}$

Subscript $\{j\}$ indexes set of schools attended by a student

TIME coded so that intercept represents final status

Level 2: $\begin{cases} \pi_{0i\{j\}} = \beta_{00\{j\}} + r_{0i\{j\}} \\ \pi_{1i\{j\}} = \beta_{10\{j\}} + r_{1i\{j\}} \end{cases}$

Level 3: $\begin{cases} \beta_{00\{j\}} = \gamma_{000} + \boxed{\sum_{h\in\{j\}} w_{ih} u_{00h}} \\ \beta_{10\{j\}} = \gamma_{100} + \boxed{\sum_{h\in\{j\}} w_{ih} u_{10h}} \end{cases}$

Weighted random school effects on final status

Weighted random school effects on growth rate

Method

- Compared baseline and conditional results from the following approaches:
  - **MM-GCM**: took into account multiple membership structure
  - ***Final school*-GCM**: three-level GCM that ignores mobility by only modeling effect of final school attended
  - ***Delete*-GCM**: three-level GCM that deletes mobile students
- Weights for the MM-GCM were based on the proportion of time-points a student was associated with a school.

Coding Schemes for MM Weights

	School			Weights		
Student	6th Grade	7th Grade	8th Grade	1st School	2nd School	3rd School
A	S1	S1	S1	1	0	0
B	S1	S2	S2	1/3	2/3	0
C	S1	S1	S2	2/3	1/3	0
D	S1	S2	S1	2/3	1/3	0
E	S1	S2	S3	1/3	1/3	1/3

Data

- Tennessee Student/Teacher Achievement Ratio (STAR) data
  - **Repeated measures** nested within **students** nested within **schools**
  - Time-points: 6^{th}, 7^{th}, and 8^{th} grades
  - Final sample: 3,123 students from 266 schools
  - 43.7% mobile students: 1,247 attended two schools and 117 attended three schools.
- Variables
  - Outcome: social sciences achievement
  - Student-level predictor: dummy-coded gender (male referent category)
  - School-level predictor: dummy-coded urbanicity (rural referent category)
  - Coding of *TIME*: $-2 = 6^{th}$ grade, $-1 = 7^{th}$ grade, $0 = 8^{th}$ grade

Estimation Procedures

- Models were fit using MLwiN software with Bayesian estimation via the Monte Carlo Markov chain (MCMC) method.
  - Default MLwiN priors used for estimation.
  - Raftery-Lewis and Brooks-Draper indices suggested a burn-in length of 5,000 with 50,000 iterations.
- Models were compared on posterior mean and standard error values, as well as model fit using the deviance information criterion value (DIC).

Baseline Results – Fixed Effects

	MM-GCM		*Final school*-GCM		*Delete*-GCM	
Fixed effect Parameter	Estimate	(*SE*)	Estimate	(*SE*)	Estimate	(*SE*)
Model for final status						
Intercept (γ_{000})	763.38	(1.89)	764.33	(1.78)	773.40	(2.00)
Model for growth rate						
Intercept (γ_{100})	8.08	(0.59)	8.46	(0.56)	11.14	(0.65)

Baseline Results – Random Effects

Random Effect Parameter	MM-GCM Estimate	MM-GCM (*SE*)	*Final school*-GCM Estimate	*Final school*-GCM (*SE*)	*Delete*-GCM Estimate	*Delete*-GCM (*SE*)
Level-1 variance						
Measures (σ_e^2)	572.28	(14.19)	572.75	(14.22)	553.13	(18.63)
Final status variance						
Students (σ_{r0}^2)	981.15	(39.97)	992.30	(39.94)	1,136.69	(57.75)
Schools (σ_{u0}^2)	535.28	(81.89)	349.70	(52.91)	205.95	(53.12)
Growth rate variance						
Students (σ_{r1}^2)	36.47	(10.08)	36.89	(10.38)	54.40	(14.52)
Schools (σ_{u1}^2)	35.08	(7.42)	24.79	(5.05)	12.63	(4.53)
Final status/growth covariance						
Students (σ_{r10})	37.34	(14.99)	37.67	(15.18)	30.89	(21.55)
Schools (σ_{u10})	96.63	(20.95)	65.26	(13.72)	23.93	(12.38)
Model fit						
DIC	87,810		87,826		49,581	

Conditional Results – Fixed Effects

Fixed Effect Parameter	MM-GCM Estimate	MM-GCM (*SE*)	*Final school-GCM* Estimate	*Final school-GCM* (*SE*)	*Delete-GCM* Estimate	*Delete-GCM* (*SE*)
Model for final status						
Intercept (γ_{000})	772.25	(2.38)	772.00	(2.20)	774.69	(2.48)
FEMALE (γ_{010})	2.57	(1.41)	2.56	(1.43)	1.59	(1.97)
INNER_CITY (γ_{001})	−36.18	(4.35)	−32.33	(4.16)	−31.08	(9.31)
SUBURBAN (γ_{002})	−4.66	(3.80)	−5.09	(3.50)	3.18	(4.59)
URBAN (γ_{003})	−3.70	(5.78)	−4.45	(5.25)	−5.59	(6.34)
Model for growth rate						
Intercept (γ_{100})	11.15	(0.75)	11.11	(0.72)	11.85	(0.80)
FEMALE (γ_{110})	1.78	(0.66)	1.72	(0.66)	1.44	(0.88)
INNER_CITY (γ_{101})	−9.82	(1.45)	−9.32	(1.41)	−9.70	(3.20)
SUBURBAN (γ_{102})	−3.27	(1.26)	−2.96	(1.20)	−0.30	(1.56)
URBAN (γ_{103})	−2.88	(1.89)	−2.17	(1.76)	−2.09	(2.08)

Conditional Results – Random Effects

	MM-GCM		*Final school*-GCM		*Delete*-GCM	
Random Effect Parameter	Estimate	(*SE*)	Estimate	(*SE*)	Estimate	(*SE*)
Level-1 variance						
Measures (σ_e^2)	572.23	(14.46)	571.61	(14.14)	553.56	(18.52)
Final status variance						
Students (σ_{r0}^2)	984.31	(40.10)	995.11	(40.26)	1,136.13	(57.26)
Schools (σ_{u0}^2)	297.61	(55.45)	206.49	(37.00)	167.98	(47.55)
Growth rate variance						
Students (σ_{r1}^2)	36.21	(11.01)	37.70	(10.48)	52.89	(14.15)
Schools (σ_{u1}^2)	19.21	(5.29)	14.17	(3.76)	10.90	(4.16)
Final status/growth covariance						
Students (σ_{r10})	37.70	(5.84)	38.51	(15.20)	29.44	(20.83)
Schools (σ_{u10})	33.98	(5.29)	24.98	(9.44)	12.61	(11.03)
Model fit						
DIC	87,798		87,805		49,580	

Discussion

- Advantages of MM-GCM:
  - Explicitly incorporates mobile data in GCM.
  - Helpful for researchers and educators who are more concerned about students' final status instead of their initial status.
  - Assumes and estimates the cumulative school effect on student growth.
  - Can be estimated using MLwiN.

Implications

- Ignoring mobility seemed to only alter the school-level fixed effects and school-level variability when compared to the MM-GCM, which corresponds to prior simulation studies.
- Most of the fixed and random effects as well as their associated *SE*s differed under the *delete-GCM* when compared to the MM-GCM and *final school*-GCM.

Thanks you !