Estimating a Piecewise Growth Model with Longitudinal Data that Contains Individual Mobility across Clusters

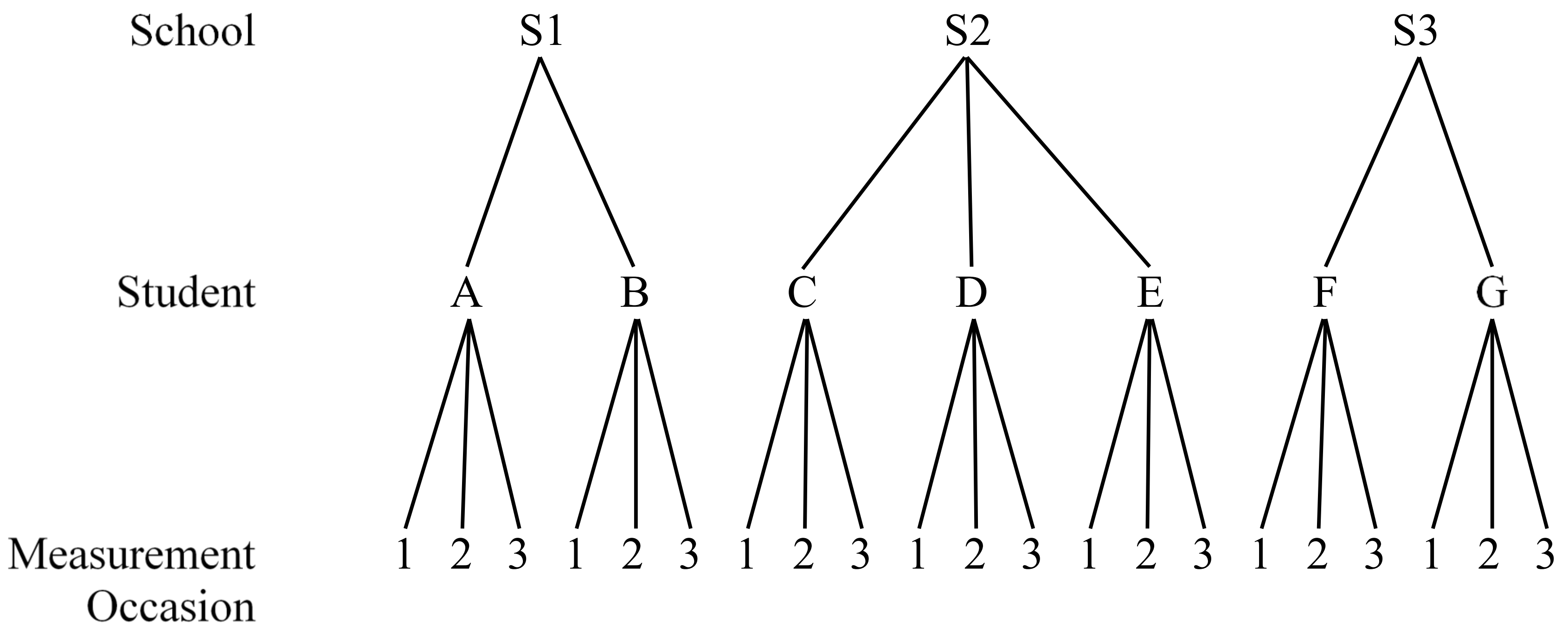
Audrey J. Leroux
Georgia State University

Piecewise Growth Model (PGM)

- PGMs are beneficial for potentially nonlinear data, because they break up curvilinear growth trajectories into separate linear components
- This modeling approach is useful when wanting to compare growth rates during two or more different time periods.
- For example:
  - Compare growth rates during two or more different time periods
  - Longitudinal data before treatment as well as during treatment
  - Longitudinal data during treatment as well as follow-up data after treatment
  - Etc.

Three-Level PGM

- Three-level PGMs are used to model the clustering of individuals, such as when students are nested within schools, classrooms, districts, etc. in educational research.



Baseline Three-Level PGM

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

where $Time_{1tij}$ and $Time_{2tij}$ are coded to represent piecewise growth

Level 2: $\begin{cases} \pi_{0ij} = \beta_{00j} + r_{0ij} \\ \pi_{1ij} = \beta_{10j} + r_{1ij}, \\ \pi_{2ij} = \beta_{20j} + r_{2ij} \end{cases}$ $\quad \begin{bmatrix} r_{0ij} \\ r_{1ij} \\ r_{2ij} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{r0}^2 & & \\ \tau_{r0,r1} & \tau_{r1}^2 & \\ \tau_{r0,r2} & \tau_{r1,r2} & \tau_{r2}^2 \end{pmatrix}\right]$

Of course, can include predictors

Level 3: $\begin{cases} \beta_{00j} = \gamma_{000} + u_{00j} \\ \beta_{10j} = \gamma_{100} + u_{10j}, \\ \beta_{20j} = \gamma_{200} + u_{20j} \end{cases}$ $\quad \begin{bmatrix} u_{00j} \\ u_{10j} \\ u_{20j} \end{bmatrix} \sim MVN\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{u00}^2 & & \\ \tau_{u00,u10} & \tau_{u10}^2 & \\ \tau_{u00,u20} & \tau_{u10,u20} & \tau_{u20}^2 \end{pmatrix}\right]$

Three-Level PGM

- To estimate this particular model presented with initial status and *two* slopes varying across both individuals and clusters, a minimum of *four* time-points must be included for identification of the model.
  - One more time-point than the number of growth parameters (three).

- Otherwise, modeling the slopes as fixed or constraints on the level-1 residual variance can be placed if utilizing fewer time-points (see McCoach, O'Connell, Reis, & Levitt, 2006; Palardy, 2010).

Three-Level PGM

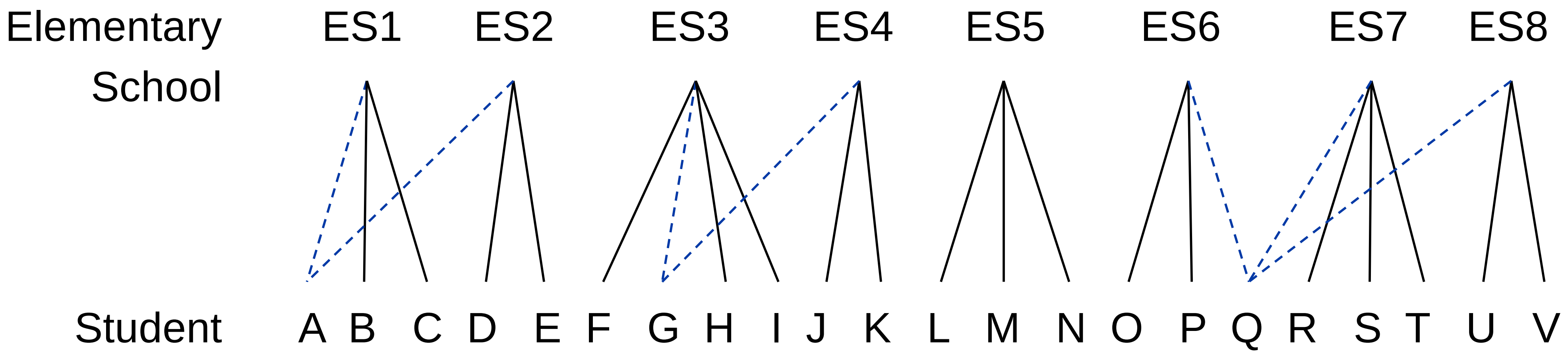
- Three-level PGMs utilized in previous research have assumed pure clustering of individuals across time or removed individuals from the analysis who changed clusters.
  - BUT…in reality, individual mobility across clusters is frequently encountered in longitudinal studies.

- Incorrect model specification in the presence of cluster mobility negatively impacts parameter estimates (Chung & Beretvas, 2012; Grady, 2010; Grady & Beretvas, 2010; Leroux, 2014; Leroux & Beretvas, 2017a, Leroux & Beretvas, 2017b; Luo & Kwok, 2009; Luo & Kwok, 2012; Meyers & Beretvas, 2006).
  - Generally, leads to inaccurate estimates of between-clusters variance components and standard errors of the fixed effects.

Longitudinal Data with Mobile Students

Student	Fall K	Spring K		Spring 1^{st}			Spring 3^{rd}				Spring 5^{th}				
	Sch. 1	Sch. 1	Sch. 2	Sch. 1	Sch. 2	Sch. 3	Sch. 1	Sch. 2	Sch. 3	Sch. 4	Sch. 1	Sch. 2	Sch. 3	Sch. 4	Sch. 5
A	✓	✓		✓			✓				✓				
B	✓	✓			✓			✓				✓			
C	✓	✓			✓			✓					✓		
D	✓		✓			✓				✓	✓				
E	✓		✓			✓				✓					✓

Multiple Membership Data



- ***Some*** units of a lower-level classification are members of more than one higher-level classification.

Multiple Membership Random Effects Model (MMREM)

- Models the contribution to the outcome, *Y*, of ***each level-2 unit*** of which the level-1 unit is a member.

- E.g., For student *i* who is a member of a set of ESs {*j*}, the unconditional model's L1 equation is:

$$Y_{i\{j\}} = \beta_{0\{j\}} + r_{i\{j\}}$$

- At L2:

$$\beta_{0\{j\}} = \gamma_{00} + \sum_{h \in \{j\}} w_{ih} u_{0h}$$

Georgia State University | COLLEGE OF EDUCATION & HUMAN DEVELOPMENT

MMREM

- Single equation:

$$Y_{i\{j\}} = \gamma_{00} + \sum_{h \in \{j\}} w_{ih} u_{0h} + r_{i\{j\}}$$

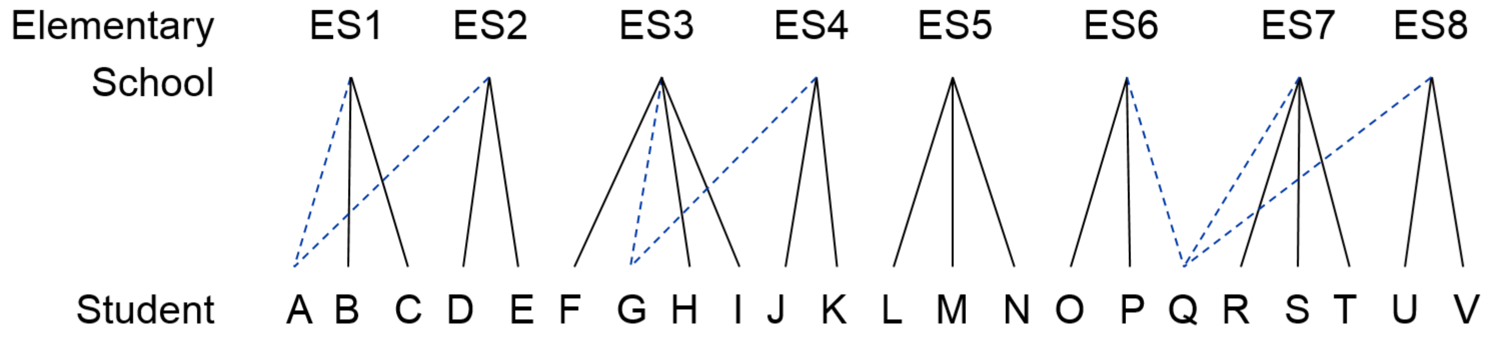
$$r_{i\{j\}} \sim N(0, \sigma^2) \text{ and } u_{0h} \sim N(0, \tau_{u00})$$

where the user specifies the weights to represent the hypothesized contribution of each L2 unit (here, elementary school)

- For each student *i*:

$$\sum_{h \in \{j\}} w_{ih} = 1$$

MMREM



- For non-mobile student B (attending ES1):

$$Y_{B\{ES1\}} = \gamma_{00} + u_{0\{ES1\}} + r_{B\{ES1\}}$$

- For mobile student A, attending ES1 and ES2:

$$Y_{A\{ES1,ES2\}} = \gamma_{00} + 0.5u_{0\{ES1\}} + 0.5u_{0\{ES2\}} + r_{A\{ES1,ES2\}}$$

- For mobile student Q, attending ES6, ES7, and ES8:

$$Y_{Q\{ES6,ES7,ES8\}} = \gamma_{00} + (1/3)u_{0\{ES6\}} + (1/3)u_{0\{ES7\}} + (1/3)u_{0\{ES8\}} + r_{Q\{ES6,ES7,ES8\}}$$

Conditional MMREM

- Can include L1 and L2 predictors.
- At L1:

$$Y_{i\{j\}} = \beta_{0\{j\}} + \beta_{1\{j\}} X_{i\{j\}} + r_{i\{j\}}$$

- And at L2:

$$\begin{cases} \beta_{0\{j\}} = \gamma_{00} + \sum_{h \in \{j\}} w_{ih}(\gamma_{01} Z_h + u_{0h}) \\ \beta_{1\{j\}} = \gamma_{10} + \sum_{h \in \{j\}} w_{ih}(\gamma_{11} Z_h + u_{1h}) \end{cases}$$

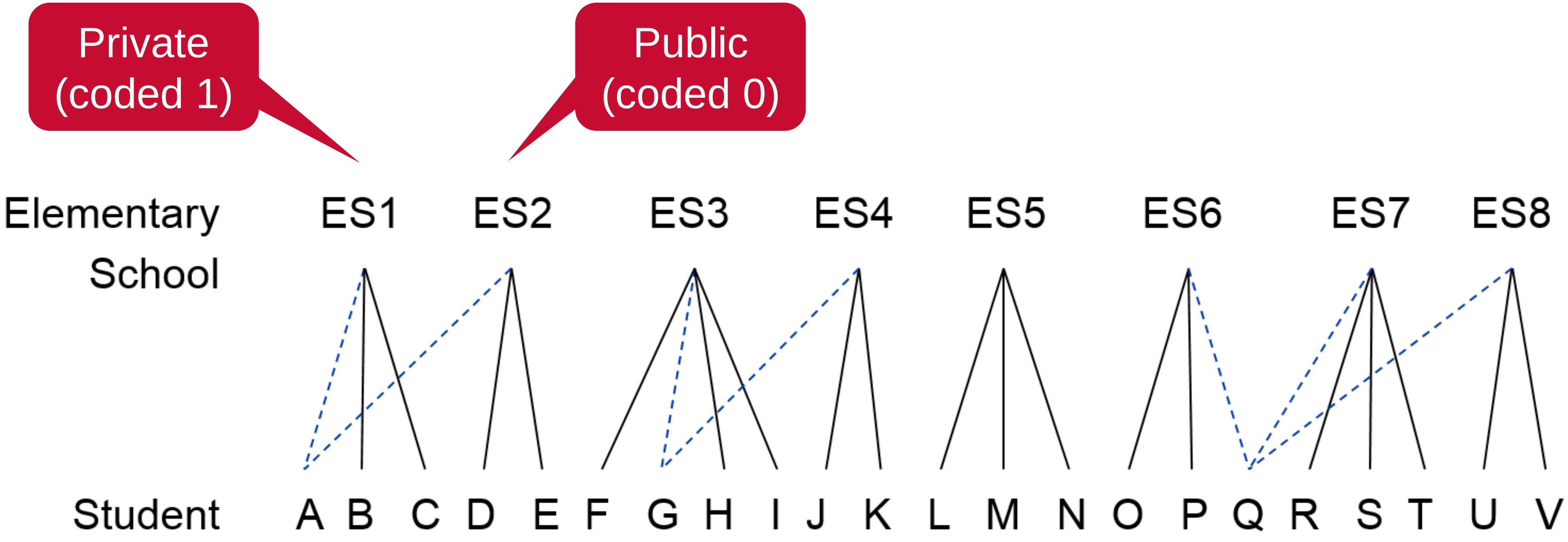
Conditional MMREM

- The following multivariate normal distribution is assumed for the level-2 residuals:

$$\begin{bmatrix} u_{0\{j\}} \\ u_{1\{j\}} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_{u00} & \\ \tau_{u10} & \tau_{u11} \end{bmatrix}\right)$$

- Note: Contribution of each ES's *Z* (for mobile students) is modeled and weighted in the same way as are schools' effects (the *u*'s).

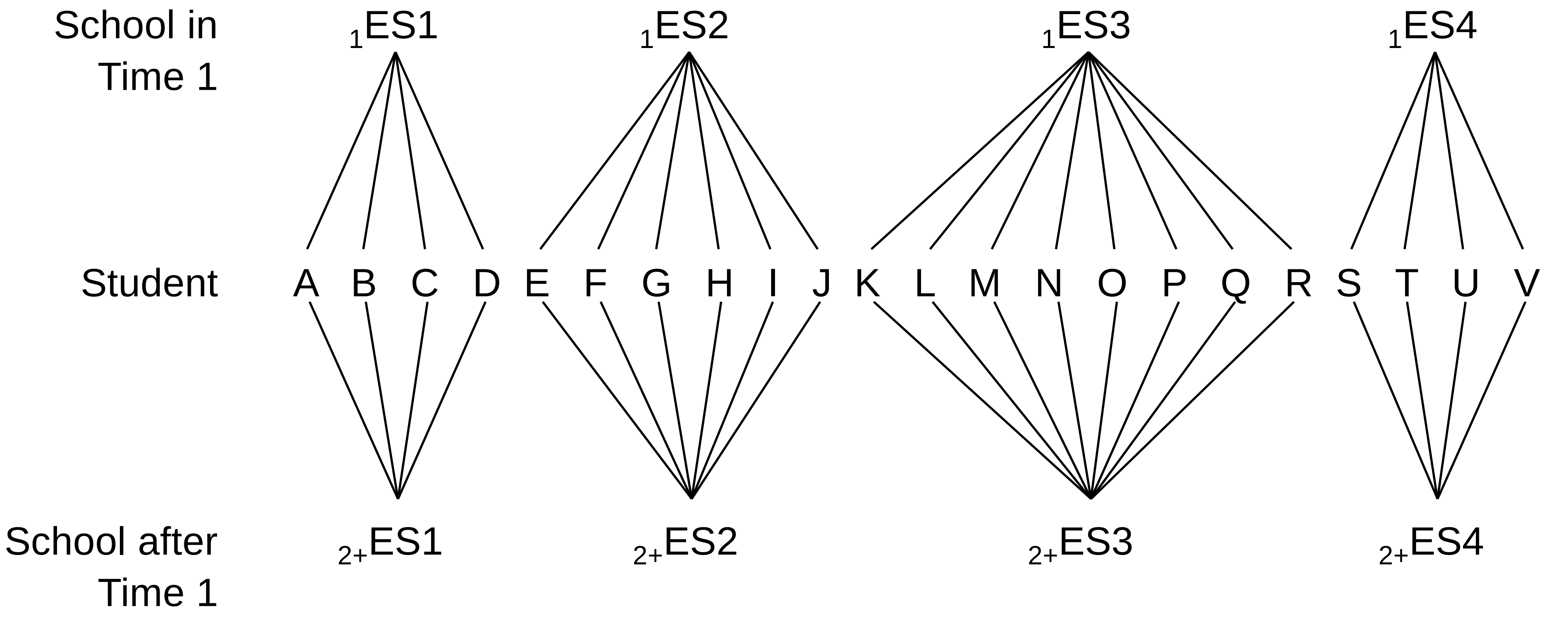
Conditional MMREM



- For mobile student A, attending <u>private</u> ES1 and <u>public</u> ES2:

$$Y_{A\{ES1,ES2\}} = \gamma_{00} + \gamma_{01}[(0.5)(1) + (0.5)(0)] + 0.5u_{0\{ES1\}} + 0.5u_{0\{ES2\}} + r_{A\{ES1,ES2\}}$$

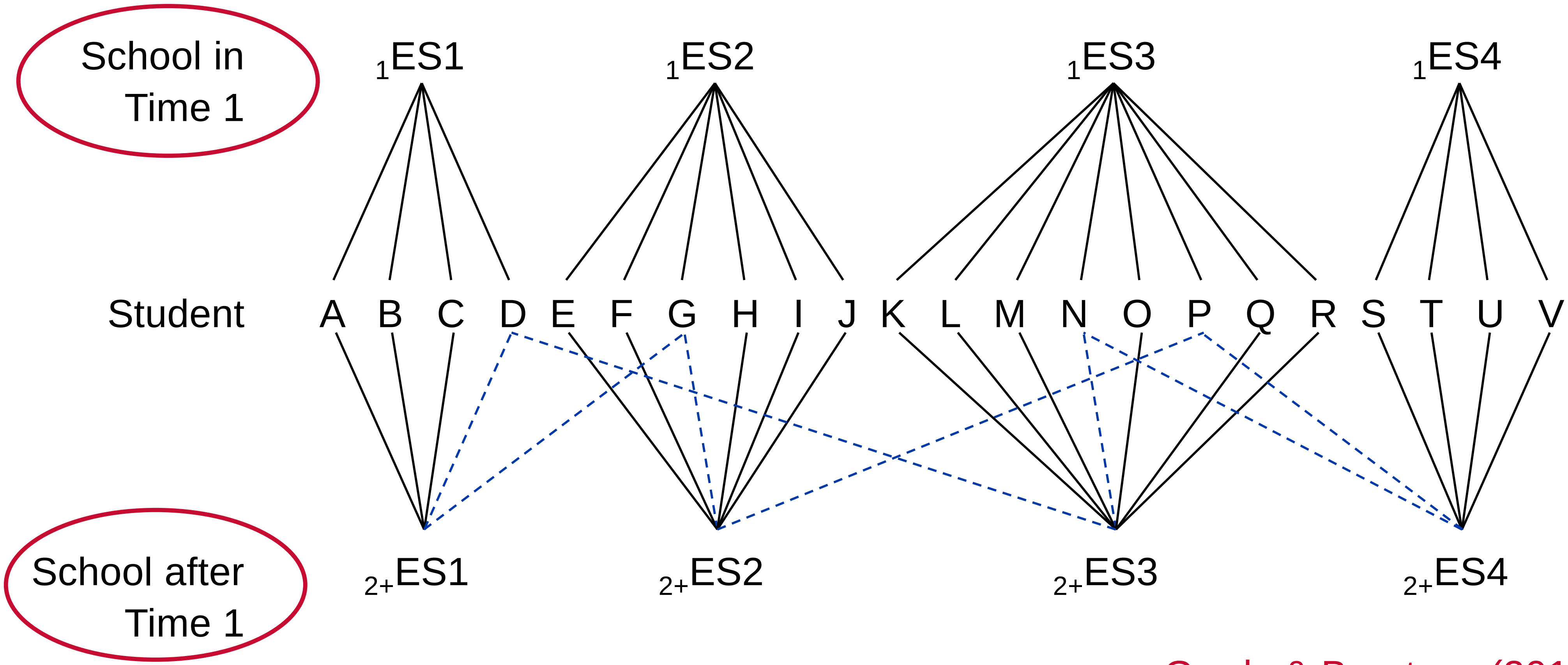
Cross-Classified Multiple Membership Longitudinal Data



Grady & Beretvas (2010)

COLLEGE OF EDUCATION & HUMAN DEVELOPMENT
Georgia State University

Cross-Classified Multiple Membership Longitudinal Data



Grady & Beretvas (2010)

Purpose of Current Study

- The current study proposes a three-level PGM to handle mobile students who change schools (clusters) during the period of data collection.

- The proposed cross-classified multiple membership PGM (CCMM-PGM) will be derived, justified, and explained using a real dataset.

Purpose of Current Study

- This extension is of particular importance when repeated measures over time are captured for students within schools because there is a high probability that at least some substantial proportion of students change schools during a study's time period.
  - **38.5%** of people aged 5-17 years moved within 2005 to 2010 (Ihrke & Faber, 2012)
    - **25%** of those between the ages 5-17 relocated within the same county
  - From 2012 to 2013, **12%** of people between the ages 5-17 years old moved
    - **69%** of those moves occurred within the same county (U.S. Census Bureau, 2013)
  - **13%** of students changed schools **4 or more** times between kindergarten and 8th grade (U.S. Government accounting office, 2010)

Baseline CCMM-PGM

Level 1: $Y_{ti(j_1,\{j_2\})} = \pi_{0i(j_1,\{j_2\})} + \pi_{1i(j_1,\{j_2\})} TIME_{1ti(j_1,\{j_2\})} + \pi_{2i(j_1,\{j_2\})} TIME_{2ti(j_1,\{j_2\})} + e_{ti(j_1,\{j_2\})}$

Level 2: $$\begin{cases} \pi_{0i(j_1,\{j_2\})} = \beta_{00(j_1,\{j_2\})} + r_{0i(j_1,\{j_2\})} \\ \pi_{1i(j_1,\{j_2\})} = \beta_{10(j_1,\{j_2\})} + r_{1i(j_1,\{j_2\})} \\ \pi_{2i(j_1,\{j_2\})} = \beta_{20(j_1,\{j_2\})} + r_{2i(j_1,\{j_2\})} \end{cases}$$

Subscripts j_1 and $\{j_2\}$ index the first and set of subsequent schools attended by a student.

Level 3: $$\begin{cases} \beta_{00(j_1,\{j_2\})} = \gamma_{0000} + u_{00j_10} \\ \beta_{10(j_1,\{j_2\})} = \gamma_{1000} + u_{10j_10} + \sum_{h\in\{j_2\}} w_{1tih} u_{100h} \\ \beta_{20(j_1,\{j_2\})} = \gamma_{2000} + u_{20j_10} + \sum_{h\in\{j_2\}} w_{2tih} u_{200h} \end{cases}$$

Intercept (initial status)

1st Slope

2nd Slope

No subsequent school residual for initial status

Note two different weights because each slope is associated with different time-points

Cross-classification of first and subsequent schools

COLLEGE OF EDUCATION & HUMAN DEVELOPMENT
Georgia State University

Data

- ECLS-K data were used with **time** nested within **students** nested within **schools**.
- Multiple membership structure due to some students' switching elementary schools across the course of data collection
- Time-points: Fall of kindergarten and springs of kindergarten, 1^{st}, 3^{rd}, and 5^{th} grade
- Outcome: Math IRT-scaled scores
- Growth rates from K – 1^{st} grade appeared faster than those from 1^{st} – 5^{th} grade
- Gender (1 = female; 0 = male) and school type (1 = private; 0 = public)
- 10,906 students (29.8% mobile) from 970 schools

Descriptive Statistics

	Variable Name	*M*	*SD*	*N*
Outcome				
Math achievement in Fall Kindergarten	Y_{1ij}	26.69	9.20	9,724
Math achievement in Spring Kindergarten	Y_{2ij}	37.17	11.95	10,664
Math achievement in Spring 1st Grade	Y_{3ij}	62.26	17.96	10,803
Math achievement in Spring 3rd Grade	Y_{4ij}	99.73	24.47	10,764
Math achievement in Spring 5th Grade	Y_{5ij}	124.05	24.66	10,801

	Variable Name	Percentage	*N*
Level-2 variable			
Female student	$FEMALE_{ij}$	49.78%	5,429
Male student		50.22%	5,477
Level-3 variable			
Private school	$PRIVATE_j$	23.51%	228
Public school		76.49%	742

Georgia State University | COLLEGE OF EDUCATION & HUMAN DEVELOPMENT

Coding of *Time* Variables

	Grades					
	Fall K	*Spring K*	*Spring 1st*	*Spring 3rd*	*Spring 5th*	*Interpretation of πs*
						π_{0ij} status in Fall K
$Time_{1tij}$	0	0.5	1.5	1.5	1.5	π_{1ij} growth rate period 1
$Time_{2tij}$	0	0	0	2	4	π_{2ij} growth rate period 2

- Exploratory analyses suggested a two-piece growth model because growth rates from kindergarten through 1st grade appeared faster than those from 1st through 5th grade

Analyses

- Baseline and conditional versions of the following models were estimated:
  - **<u>CCMM-PGM</u>**: appropriately took into account student mobility
  - ***<u>First school</u>*****<u>-PGM</u>**: ignored mobility by only modeling effect of the first school attended
  - ***<u>Delete</u>*****<u>-PGM</u>**: ignored mobility by deleting students who changed schools

- Weights are based on the proportion of time-points a student was associated with a school.

Coding Schemes for Weights

Student	Schools					Weights (K – 1st)		Weights (1st – 5th)			
	Fall K	Spring K	Spring 1st	Spring 3rd	Spring 5th	1st School	2nd School	1st School	2nd School	3rd School	4th School
A	S1	S1	S1	S1	S1	1	0	1	0	0	0
B	S1	S1	S1	S1	S2	1	0	3/4	1/4	0	0
C	S1	S1	S1	S2	S2	1	0	1/2	1/2	0	0
D	S1	S1	S2	S2	S2	1/2	1/2	1/4	3/4	0	0
E	S1	S2	S2	S2	S2	1	0	1	0	0	0
F	S1	S2	S2	S2	S3	1	0	3/4	1/4	0	0
G	S1	S2	S2	S3	S3	1	0	1/2	1/2	0	0
H	S1	S2	S3	S3	S3	1/2	1/2	1/4	3/4	0	0
I	S1	S2	S3	S3	S4	1/2	1/2	1/4	1/2	1/4	0
J	S1	S2	S3	S4	S4	1/2	1/2	1/4	1/4	1/2	0
K	S1	S2	S3	S4	S5	1/2	1/2	1/4	1/4	1/4	1/4

Estimation

- Models were fit using R with MCMC estimation using R2jags to interface with Just Another Gibbs Sampler (JAGS).
  - Non-informative normal priors were used for fixed effects parameters and inverse-Wishart distributions for the covariance matrices.
  - Burn-in period of 5,000 iterations and an additional 50,000 iterations

- Parameter and *SE* estimates were compared, as well as model fit using the deviance information criterion value (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002).

Baseline Fixed Effects

	Estimating Model								
	CCMM-PGM[1]			*First School*-PGM[1]			*Delete*-PGM[2]		
Parameter	Coeff.	Est.	(*SE*)	Coeff.	Est.	(*SE*)	Coeff.	Est.	(*SE*)
Model for initial status									
Intercept	γ_{0000}	25.313	(0.171)	γ_{000}	25.307	(0.171)	γ_{000}	25.441	(0.194)
Model for 1st slope									
Intercept	γ_{1000}	25.480	(0.150)	γ_{100}	25.488	(0.146)	γ_{100}	25.442	(0.171)
Model for 2nd slope									
Intercept	γ_{2000}	15.499	(0.066)	γ_{200}	15.500	(0.066)	γ_{200}	15.544	(0.078)
DIC		451,380.3			453,361.6			305,458.7	

Baseline Random Effects

	Estimating Model								
	CCMM-PGM			*First School*-PGM			*Delete*-PGM		
Parameter	Coeff.	Est.	(*SE*)	Coeff.	Est.	(*SE*)	Coeff.	Est.	(*SE*)
Level-1 variance between									
Measures	σ^2	62.453	(0.631)	σ^2	62.479	(0.626)	σ^2	63.345	(0.706)
Initial status variance between									
Students	τ^2_{r0}	32.443	(1.057)	τ^2_{r0}	32.376	(1.020)	τ^2_{r0}	32.911	(1.273)
1st schools	$\tau^2_{u0_{j1}}$	19.386	(1.269)	τ^2_{u0}	19.356	(1.288)	τ^2_{u0}	19.844	(1.495)
1st slope variance between									
Students	τ^2_{r1}	39.281	(1.132)	τ^2_{r1}	39.339	(1.113)	τ^2_{r1}	37.989	(1.279)
1st schools	$\tau^2_{u1_{j1}}$	8.663	(1.139)	τ^2_{u1}	11.626	(0.927)	τ^2_{u1}	12.457	(1.132)
Subsequent schools	$\tau^2_{u1_{\{j2\}}}$	3.567	(0.990)	—	—	—	—	—	—
2nd slope variance between									
Students	τ^2_{r2}	7.761	(0.214)	τ^2_{r2}	7.826	(0.224)	τ^2_{r2}	7.043	(0.254)
1st schools	$\tau^2_{u2_{j1}}$	1.861	(0.349)	τ^2_{u2}	2.585	(0.184)	τ^2_{u2}	2.870	(0.249)
Subsequent schools	$\tau^2_{u2_{\{j2\}}}$	0.935	(0.414)	—	—	—	—	—	—

Georgia State University | COLLEGE OF EDUCATION & HUMAN DEVELOPMENT

Conditional Fixed Effects

	Estimating Model								
	CCMM-PGM[1]			*First School*-PGM[1]			*Delete*-PGM[2]		
Parameter	Coeff.	Est.	(*SE*)	Coeff.	Est.	(*SE*)	Coeff.	Est.	(*SE*)
Model for initial status									
Intercept	γ_{0000}	24.255	(0.182)	γ_{000}	24.259	(0.168)	γ_{000}	24.281	(0.204)
FEMALE	γ_{0100}	−0.207	(0.176)	γ_{010}	−0.208	(0.187)	γ_{010}	−0.340	(0.216)
Sch1_PRIVATE	γ_{0010}	5.057	(0.385)	γ_{001}	5.082	(0.374)	γ_{001}	5.459	(0.423)
Model for 1st slope									
Intercept	γ_{1000}	25.237	(0.163)	γ_{100}	25.246	(0.156)	γ_{100}	25.140	(0.197)
FEMALE	γ_{1100}	−1.556	(0.186)	γ_{110}	−1.562	(0.185)	γ_{110}	−1.651	(0.227)
Sch1_PRIVATE	γ_{1010}	1.743	(1.529)	γ_{101}	0.955	(0.340)	γ_{101}	1.285	(0.403)
SubSch_PRIVATE	γ_{1001}	−0.800	(1.569)	—	—	—	—	—	—
Model for 2nd slope									
Intercept	γ_{2000}	15.446	(0.072)	γ_{200}	15.447	(0.073)	γ_{200}	15.511	(0.086)
FEMALE	γ_{2100}	−0.699	(0.078)	γ_{210}	−0.707	(0.077)	γ_{210}	−0.635	(0.096)
Sch1_PRIVATE	γ_{2010}	1.601	(0.407)	γ_{201}	0.284	(0.168)	γ_{201}	0.116	(0.180)
SubSch_PRIVATE	γ_{2001}	−1.466	(0.425)	—	—	—	—	—	—
DIC		454,180.8			456,478.6			306,754.3	

Conditional Random Effects

	Estimating Model								
	CCMM-PGM			*First School*-PGM			*Delete*-PGM		
Parameter	Coeff.	Est.	(*SE*)	Coeff.	Est.	(*SE*)	Coeff.	Est.	(*SE*)
Level-1 variance between									
Measures	σ^2	62.407	(0.657)	σ^2	62.413	(0.657)	σ^2	63.287	(0.711)
Initial status variance between									
Students	τ^2_{r0}	32.443	(0.972)	τ^2_{r0}	32.482	(1.043)	τ^2_{r0}	33.049	(1.295)
1st schools	$\tau^2_{u0_{j1}}$	15.222	(1.041)	τ^2_{u0}	15.165	(1.082)	τ^2_{u0}	14.809	(1.215)
1st slope variance between									
Students	τ^2_{r1}	38.769	(1.107)	τ^2_{r1}	38.861	(1.082)	τ^2_{r1}	37.232	(1.337)
1st schools	$\tau^2_{u1_{j1}}$	8.958	(1.089)	τ^2_{u1}	11.388	(0.908)	τ^2_{u1}	12.205	(1.111)
Subsequent schools	$\tau^2_{u1_{\{j2\}}}$	2.974	(0.932)	—	—	—	—	—	—
2nd slope variance between									
Students	τ^2_{r2}	7.638	(0.217)	τ^2_{r2}	7.699	(0.220)	τ^2_{r2}	6.960	(0.255)
1st schools	$\tau^2_{u2_{j1}}$	1.826	(0.353)	τ^2_{u2}	2.562	(0.190)	τ^2_{u2}	2.851	(0.240)
Subsequent schools	$\tau^2_{u2_{\{j2\}}}$	0.927	(0.395)	—	—	—	—	—	—

Implications

- Ignoring mobility could lead to inaccurate conclusions about:
  - The intercept and slope estimates in a three-level PGM if one were to delete mobile cases
  - The impact of cluster-level predictors (regardless if you delete or ignore mobile individuals)
  - The impact of both level-2 and level-3 predictors if one were to delete mobile cases.

Implications

- Researchers using a PGM ignoring multiple membership data should be careful when making inferences about the nature of variability in growth rates.
  - For the *delete*-PGM, the *SE* estimates of the other variances were larger, which could then lead to erroneous conclusions about random effects if mobile individuals were removed from analysis.
- The CCMM-PGM fit to the data better than the *first school*-PGM.
- Because of these findings, a simulation study will be conducted this summer, so stay tuned…

Thank you!

aleroux@gsu.edu