When People Move: Using Cross-Classified and Multiple Membership Growth Curve Modeling in Non-Hierarchical Multilevel Data Structures

Bess A. Rose
AEA Conference: Evaluation 2013
Demonstration Session 548
October 18, 2013

Mobility
• Evaluations often look at change over time
• Family, employee, and student mobility is the norm in the U.S. today
• So how do you analyze data?

BACKGROUND
Terms and Definitions

Multilevel Data
• Units “nested” within units
• Examples:
  – Students in classrooms
  – Employees in job sites
  – Measurement occasions in students in schools
• Outcomes within groups are likely correlated, so use multilevel modeling, not regression

Unconditional HLM Growth Models
The reading score at time $t$ for student $i$ who attended school $j$:
At Level 1 (measurement time):

$$Rdg_{ij} = \pi_{0ij} + \pi_{1ij}Yr_{ij} + e_{ij}$$

Intercept:
- (Starting Point)
Slope:
- (Annual Growth)

At Level 2 (student):

Intercept:
- $\pi_{0ij} = \beta_{00j} + r_{0ij}$
Slope:
- $\pi_{1ij} = \beta_{10j} + r_{1ij}$
Unconditional HLM Growth Models

The reading score at time $t$ for student $i$ who attended school $j$:

At Level 3 (school):

Intercept:

$$\beta_{00j} = \gamma_{000} + u_{00j}$$

Slope:

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

Cross-Classification

- Lower-level units belong to more than 1 higher-level classification
- Examples:
  - Students may attend the same school but live in different neighborhoods (e.g., Scotland Neighbourhood Study, Garner & Raudenbush, 1991)

Multiple Membership

- Lower-level units belong to more than 1 higher-level unit within the same classification
- Examples:
  - Patients served by multiple nurses
  - Doctors practicing in multiple hospitals
  - Students taking multiple classes
  - Students attending more than one high school

Hierarchical

- Usually multilevel = hierarchical
- Each unit belongs to one (and only one) higher-level unit
- When this isn’t true, we have non-hierarchical multilevel data

(Fielding & Goldstein, 2006)
Growth Models With Mobility

The reading score at time t for student i who attended (the set of) school(s) j1 in the first year and (the set of) school(s) j2 in subsequent years:

At Level 1 (measurement time):

\[ R_{dyt(j_1,j_2)} = \pi_{dyt(j_1,j_2)} + \beta_{dyt(j_1,j_2)}Y_{j1,j2} + e_{dyt(j_1,j_2)} \]

Intercept (Starting Point)  Slope (Annual Growth)

Year has to start at 0

(Adapted from Grady & Beretvas, 2010, pp. 405-407)

At Level 2 (student):

- Intercept:
  \[ \pi_{0j(i,j_1,j_2)} = \beta_{0j(i,j_1,j_2)} + r_{0j(i,j_1,j_2)} \]
  Intercept:
  Starting point takes into account all first-year schools

- Slope:
  \[ \pi_{1j(i,j_1,j_2)} = \beta_{1j(i,j_1,j_2)} + r_{1j(i,j_1,j_2)} \]
  Slope:
  Growth curve also takes into account all subsequent schools

(Adapted from Grady & Beretvas, 2010, pp. 405-407)

At Level 3 (school):

- Intercept:
  \[ \beta_{00(j_1,j_2)} = \gamma_{000} + \sum_{h \in \{j_1\}} w_{ih} \mu_{00h} \]
  Intercept:
  Starting point takes into account all first-year schools

- Slope:
  \[ \beta_{10(j_1,j_2)} = \gamma_{100} + \sum_{h \in \{j_1\}} w_{ih} \mu_{10h} + \sum_{h \in \{j_2\}} w_{ih} \mu_{10h} \]
  Slope:
  Growth curve also takes into account all subsequent schools

(Adapted from Grady & Beretvas, 2010, pp. 405-407)

Can ignoring mobility change your study’s findings?

YES

Goldstein, Burgess, & McConnell (2007)
Chung (2009)
Grady & Beretvas (2010)
Luo & Kwok (2012)

Data for MLwiN

- Prepare your data file in another stats program
- Single data file (unlike HLM)
- Each row is a measurement occasion
- Student and school info repeated within student
- Student and school IDs must start at 1
- Some data manipulation can be done in MLwiN (sort, rename, select cases)
Data for MLwiN

- **Columns:**
  - Measurement Occasion or Time (Level 1)
  - Year (for growth model, starts at 0)
  - ID (Level 2)
  - Rdg (Dependent Var or Outcome)
  - First_School_1, First_School_2, etc and weights
  - Subsequent_School_1, Subs_Sch_2, etc and weights
  - Student covars
- **Let’s look at the data!**

<table>
<thead>
<tr>
<th>T</th>
<th>Yr</th>
<th>id</th>
<th>Rdg</th>
<th>FS1</th>
<th>FS2</th>
<th>FS1 wt</th>
<th>FS2 wt</th>
<th>Sub Sch1</th>
<th>Sub Sch2</th>
<th>Sub Sch3</th>
<th>SS1 wt</th>
<th>SS2 wt</th>
<th>SS3 wt</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1.04</td>
<td>25</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>20</td>
<td>0</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>1</td>
<td>1.04</td>
<td>25</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>25</td>
<td>20</td>
<td>0</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>.80</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>15</td>
<td>20</td>
<td>0</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>2</td>
<td>1.84</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>15</td>
<td>20</td>
<td>0</td>
<td>.5</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>81</td>
<td>1.84</td>
<td>68</td>
<td>153</td>
<td>.33</td>
<td>.33</td>
<td>68</td>
<td>65</td>
<td>62</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>81</td>
<td>0.28</td>
<td>68</td>
<td>153</td>
<td>.33</td>
<td>.33</td>
<td>68</td>
<td>65</td>
<td>62</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>81</td>
<td>0.25</td>
<td>68</td>
<td>153</td>
<td>.33</td>
<td>.33</td>
<td>68</td>
<td>65</td>
<td>62</td>
<td>.33</td>
<td>.33</td>
<td>.33</td>
</tr>
</tbody>
</table>

**USING MLWIN TO MODEL STUDENT GROWTH WITH MOBILITY**

Growth Models or Repeated Measures

**Setting Up Models in MLwiN**

- **Sort by:**
  - First_Sch_1
  - Subsequent_Sch_1
  - Student
  - Time

**Setting Up Models in MLwiN**

- **Run “naïve” model using hierarchical nesting:**
  - First_Sch_1 (Level 4)
  - Subsequent_Sch_1 (Level 3)
  - Student (Level 2)
  - Time (Level 1)
- **This gives starting values for actual modeling using Monte Carlo**
Add Year for Growth Model

Set Intercept Random at Levels 2 & 4

Set Year Random at All Levels

Run Naïve Model

Not for interpreting! Just starting values for Monte Carlo.

Adding Cross-Classification & Multiple Membership to the Model

- The “real” model using CCMM:
  - First_Sch_1 – First_Sch_4 (CC Level 3 – MM)
  - Subs_Sch_1 – Subs_Sch_12 (CC Level 3 – MM)
  - Student (Level 2)
  - Time (Level 1)
- First switch to Monte Carlo
- Then set cross-classifications and multiple memberships

Estimation Control - MCMC
MCMC - Classifications

Cross-Classification & Multiple Membership

CCMM Model

Monte Carlo Results

Variance Components

Variance Components

Variance in intercept among 1st-year schools

Variance in intercept among students
Variance Components

Variance in growth among 1st-year schools

Variance in growth among subsequent schools

Variance in growth among students

Added Covars

<table>
<thead>
<tr>
<th>Intercept Estimate (SE)</th>
<th>Growth Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>.479 (.029)</td>
</tr>
<tr>
<td>First schools</td>
<td>.135 (.023)</td>
</tr>
<tr>
<td>Subsequent schools</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Most of the variation in initial reading scores is due to variation among students.

Students’ growth is due largely to influence of schools, not students.

Small estimates of growth likely due to use of standardized z scores.

Required Reading:

- Leckie & Bell (2013): MLwiN Practical on Cross-Classified Multilevel Models (MLwiN course)
- Leckie & Owen (2013): MLwiN Practical on Multiple Membership Multilevel Models (MLwiN course)

My contact info: Bess Rose brose6@jhu.edu

Required Reading:

MLwiN online course at Center for Multilevel Modelling [www.bristol.ac.uk/cmm/]

MLwiN online course at Center for Multilevel Modelling [www.bristol.ac.uk/cmm/]

MLwiN online course at Center for Multilevel Modelling [www.bristol.ac.uk/cmm/]
References


References


References


References

