HLMHCM model for the growth data

1. Introduction to HLMHCM models

In HCM2, level-1 units are nested within cells and cross-classified by two higher-level factors. HLMHCM adds a level within the cells. For example, we may have a growth model for each of a set of students, all of whom live in the same neighborhood and attend the same school. We would say that level-1 units (repeated measures) are nested within level-2 units (children); level-2 units are crossed by rows (neighborhoods) and columns (schools). Another example might involve repeated item responses at a given time for a student encountering a given teacher. The level-1 units are the item responses, nested within occasions (level-2) crossed by rows (students) and columns (teachers).

2. Description of the data

Level-1 file. The level-1 or within-cell file, GROWTH.SAV has 2008 observations collected on 567 students beginning at grade one and followed up annually thereafter for six years. The image below shows the time series data for the first three students. All of them have complete data; typically there are three or four observations per child. Following the student ID field are that student's values on two variables:

- AGE8
    The age of the child minus 8 at each testing occasion. Therefore, it is 0 at age 8, 1 at age 9, etc.
- MATH
    A math test score in an IRT metric.

	studid	age8	math
1	1	-0.420	2.100
2	1	0.580	3.000
3	1	1.580	4.300
4	1	2.580	6.200
5	1	3.580	7.300
6	1	4.580	8.100
7	2	-0.053	2.600
8	2	0.947	2.933
9	2	1.947	3.500
10	2	2.947	4.000
11	2	3.947	5.400
12	2	4.947	7.200
13	3	-0.299	2.700
14	3	0.701	3.400
15	3	1.701	4.400
16	3	2.701	5.200
17	3	3.701	7.100
18	3	4.701	8.500

We see that the first student was about seven and a half years old (AGE8 = –0.420) during the first data collection wave with a math score of 2.1.

Level-2 file. The level-2 units in the illustration are 567 students. The data are stored in the file STUDENT.SAV. The level-2 data for the first eight children are listed below. The first ID is the level-3 row-factor (*i.e.*, school) ID, the second ID is the level-3 column factor (*i.e.*, neighbor) ID, and the third ID is the level-2 (*i.e.*, student) ID. **Note that the level-2 files must be sorted in the same order of level-2 ID.**

There are three variables:

- FEMALE (1 = female, 0 = male)
- BLACK (1 = African-American, 0 = other)
- HISPANIC (1 = Hispanic, 0 = other)

	schid	neighid	studid	female	black	hispanic
1	175	68	1	0.000	1.000	0.00
2	99	68	2	0.000	1.000	0.00
3	99	68	3	0.000	1.000	0.00
4	109	23	4	1.000	0.000	0.00
5	165	72	5	0.000	1.000	0.00
6	207	72	6	1.000	1.000	0.00
7	6	72	7	1.000	1.000	0.00
8	148	72	8	0.000	0.000	0.00

We see, for example, that student 1 who attended school 175 and resided in neighborhood 68 is a African-American male (FEMALE = 0, BLACK = 1, HISPANIC = 0).

Level-3 row-factor file. The level-3 row-factor (school) level file, SCHOOL.SAV, consists of data on 1 variable for 224 schools. The variable is SCHPOV, which is an indicator of school poverty, as measured by the percentage of the total number of students enrolled in free or subsidized lunch programs.

We see that the first school, school 1, has 91% of its students enrolled in free or subsidized lunch programs.

	schid	schpov
1	1	91.400
2	2	78.900
3	3	63.200
4	4	93.000
5	5	71.800
6	6	92.600
7	7	96.200
8	8	83.400

Level-3 column-factor file. The level-3 row-factor (neighborhood) level file, NEIGH.SAV, consists of data on 1 variable for 74 neighborhoods. The variable is DISADV (a scale measuring social deprivation, which incorporates information on the poverty concentration, health, and housing stock of a local community). A measure of neighborhood disadvantage, constructed through an oblique factor analysis from the 1990 decennial census data, tapped the level of poverty and unemployment, and the percentage of families that were headed by females and percentage on welfare (Sampson & Raudenbush, 1999; Sampson, Raudenbush, & Earls, 1997).

	neighid	disadv
1	1	0.449
2	2	-0.553
3	3	-0.382
4	4	-0.563
5	5	-0.313
6	6	-0.566
7	7	0.160
8	8	1.222

In sum, there are two variables at level 1, three at level 2, and one for each of the level-3 factors.

3. Creating the command file

Once the MDM file is constructed, it can be used as input for the analysis. Model specification has three steps:

1. Specification of the level-1 model. In our case we shall model mathematics achievement (MATH) as the outcome, to be predicted by AGE8. Hence, the level-1 model will have two coefficients for each student: the intercept and the AGE slope.
2. Specification of the level-2 prediction model. Here each level-1 coefficient – the intercept and the AGE8 slope in our example – becomes an outcome variable. We may select certain student characteristics to predict each of these level-1 coefficients. In principle, the level-2 parameters then describe the distribution of growth curves cross-classified by schools and neighborhoods.
3. Specification of level-1 coefficients as random or non-random across level-two units. We shall model the intercept and the AGE8 slope as varying randomly across the students cross-classified by schools and neighborhoods.
4. Specification of the level-3 row- and/or column-factor prediction model. Here each level-2 coefficient becomes an outcome, and we can select row- and/or column-factor variables to predict school-to-school and neighbor-to-neighbor variation in these level-2 coefficients. In principle, this model specifies how schools and neighborhoods differ with respect to the distribution of growth curves within them.
5. Specification of the residual row and column as random or non-random, the effects associated with row-specific predictors as varying randomly or fixed over columns, and the effects associated with column-specific predictors as varying randomly or fixed over rows. We shall test whether the associations between neighborhood disadvantage (a column-specific predictor) and growth parameters vary over schools.

In the previous example in this set, we specified a level-1 model for the Growth data. We also opted to allow the intercept and the AGE8 slope as varying randomly across the students cross-classified by schools and neighborhoods. In this example, the second in a set of two based on these data, we set up an level-2 and a row-factor prediction model for the Growth data.

From the WHLM window, open the **File** menu. Choose **Edit/Run old command (.hlm/.mlm) file** to open an **Open Command File** dialog box. Open the command file created in the previous example (GROWTH1.MLM).
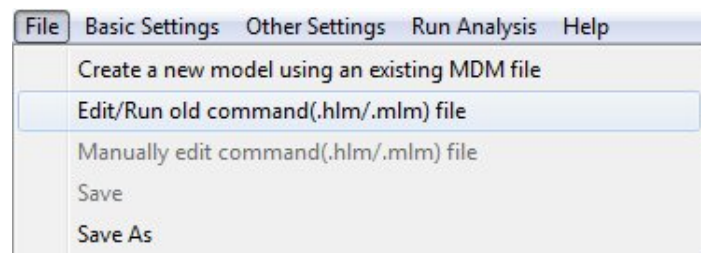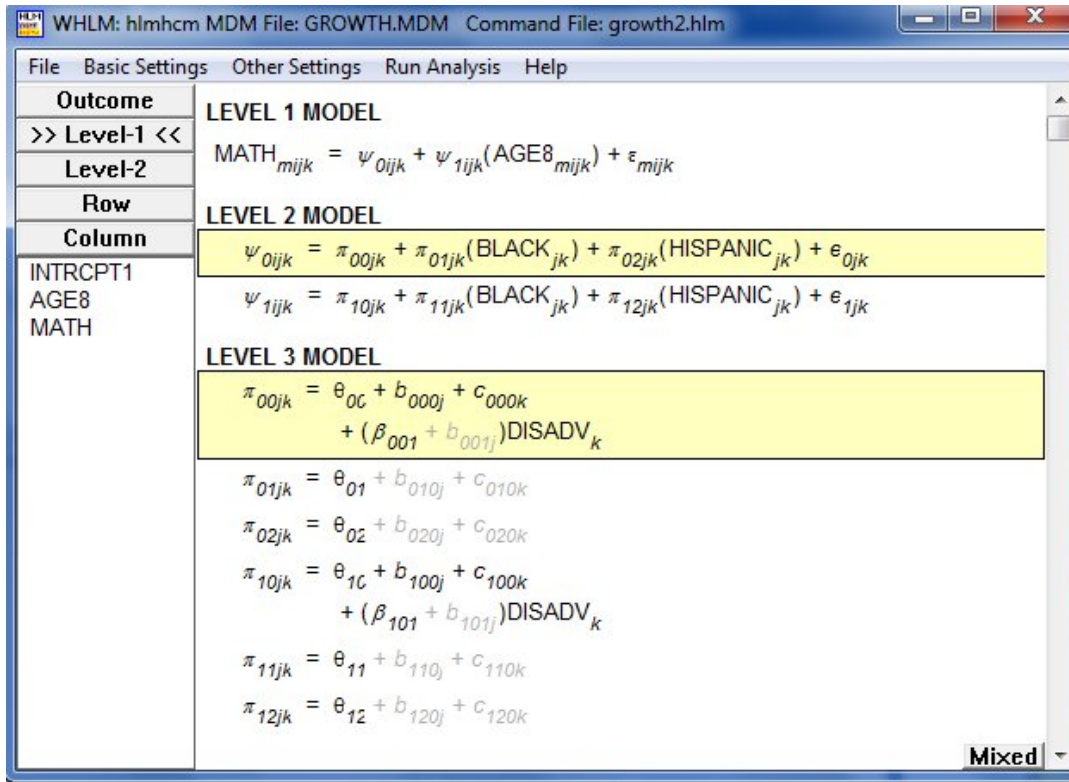
To set up the level-2 model:
Select the equation containing ψ_{pijk} to be modeled, a list box for level-2 variables (>>**Level-2**<<)
will appear. The image below shows the models with BLACK and HISPANIC as the level-2
predictors. In the interest of parsimony, all level-2 coefficients are fixed. (To specify either of them
as randomly varying, select the equation containing a specific regression coefficient, π_{pqjk}, and
click on b_{pqrj} and/or c_{pqsk}).

File   Basic Settings   Other Settings   Run Analysis   Help

| Outcome |
| Level-1 |
| >> Level-2 << |
| Row |
| Column |

INTRCPT2
FEMALE
BLACK
HISPANIC

LEVEL 1 MODEL

$$MATH_{mijk} = \psi_{0ijk} + \psi_{1ijk}(AGE8_{mijk}) + \varepsilon_{mijk}$$

LEVEL 2 MODEL

$$\psi_{0ijk} = \pi_{00jk} + \pi_{01jk}(BLACK_{jk}) + \pi_{02jk}(HISPANIC_{jk}) + e_{0jk}$$
$$\psi_{1ijk} = \pi_{10jk} + \pi_{11jk}(BLACK_{jk}) + \pi_{12jk}(HISPANIC_{jk}) + e_{1jk}$$

LEVEL 3 MODEL

$$\pi_{00jk} = \theta_{00} + b_{000j} + c_{000k}$$
$$\pi_{01jk} = \theta_{01} + b_{010j} + c_{010k}$$
$$\pi_{02jk} = \theta_{02} + b_{020j} + c_{020k}$$
$$\pi_{10jk} = \theta_{10} + b_{100j} + c_{100k}$$
$$\pi_{11jk} = \theta_{11} + b_{110j} + c_{110k}$$
$$\pi_{12jk} = \theta_{12} + b_{120j} + c_{120k}$$

Mixed

To set up the level-3 row or/and column-factor prediction model:
Select the equation containing π_{pqjk} to be modeled, a list box for level-3 row-factor variables
(>>**Row**<<) will appear. To display level-3 column-factor variables, click on | Column | and
the corresponding list box of variables. Figure 17.13 shows the level-3 column-factor prediction
model with DISADV as the covariate. In the level-3 model, we treated the association between
neighborhood disadvantage and the growth parameters as fixed across all schools. Note that b_{001j}
and b_{101j} are disabled.

WHLM: hlmhcm MDM File: GROWTH.MDM   Command File: growth2.hlm

File   Basic Settings   Other Settings   Run Analysis   Help

| Outcome |
| >> Level-1 << |
| Level-2 |
| Row |
| Column |

INTRCPT1
AGE8
MATH

LEVEL 1 MODEL

$$MATH_{mijk} = \psi_{0ijk} + \psi_{1ijk}(AGE8_{mijk}) + \varepsilon_{mijk}$$

LEVEL 2 MODEL

$$\psi_{0ijk} = \pi_{00jk} + \pi_{01jk}(BLACK_{jk}) + \pi_{02jk}(HISPANIC_{jk}) + e_{0jk}$$

$$\psi_{1ijk} = \pi_{10jk} + \pi_{11jk}(BLACK_{jk}) + \pi_{12jk}(HISPANIC_{jk}) + e_{1jk}$$

LEVEL 3 MODEL

$$\pi_{00jk} = \theta_{0C} + b_{000j} + c_{000k}$$
$$+ (\beta_{001} + b_{001j})DISADV_k$$

$$\pi_{01jk} = \theta_{01} + b_{010j} + c_{010k}$$

$$\pi_{02jk} = \theta_{02} + b_{020j} + c_{020k}$$

$$\pi_{10jk} = \theta_{1C} + b_{100j} + c_{100k}$$
$$+ (\beta_{101} + b_{101j})DISADV_k$$

$$\pi_{11jk} = \theta_{11} + b_{110j} + c_{110k}$$

$$\pi_{12jk} = \theta_{12} + b_{120j} + c_{120k}$$

Mixed

Save the model before running the analysis.

4. Interpreting the output

The results of the analysis are given below.

Specifications for this HLM-HCM run

Problem Title: CONDITIONAL LINEAR GROWTH MODEL, WITH NEIGHBORHOOD DISADVANTAGE

The data source for this run = growth.mdm
The command file for this run = growth2.hlm
Output file name = growth1.html
The maximum number of level-1 units = 2008
The maximum number of level-2 units = 567
The maximum number of row units = 224
The maximum number of column units = 74
The maximum number of iterations = 100
Method of estimation: full maximum likelihood

The outcome variable is MATH

Summary of the model specified

Level-1 Model

$$MATH_{mijk} = \psi_{0ijk} + \psi_{1ijk}*(AGE8_{mijk}) + \varepsilon_{mijk}$$

Level-2 Model

$$\psi_{0ijk} = \pi_{00jk} + \pi_{01jk}*(BLACK_{jk}) + \pi_{02jk}*(HISPANIC_{jk}) + e_{0jk}$$
$$\psi_{1ijk} = \pi_{10jk} + \pi_{11jk}*(BLACK_{jk}) + \pi_{12jk}*(HISPANIC_{jk}) + e_{1jk}$$

Row/Column Model

$$\pi_{00jk} = \theta_{00} + b_{000j} + c_{000k}$$
$$\qquad + DISADV_k*(\beta_{001})$$
$$\pi_{01jk} = \theta_{01}$$
$$\pi_{02jk} = \theta_{02}$$
$$\pi_{10jk} = \theta_{10} + b_{100j} + c_{100k}$$
$$\qquad + DISADV_k*(\beta_{101})$$
$$\pi_{11jk} = \theta_{11}$$
$$\pi_{12jk} = \theta_{12}$$

For starting values, data from 1967 level-1, 526 level-2, 219 rows, and 74 column records were used

Final Results - iteration 1300

Iterations stopped due to small change in likelihood function

$\sigma^2 = 0.16386$

τ

INTRCPT1	AGE8
INTRCPT2,e_0	INTRCPT2,e_{1jk}
0.27546	0.08088
0.08088	0.03538

τ (as correlations)
```
 1.000  0.819
 0.819  1.000
```

Ω

INTRCPT1	AGE8
INTRCPT2	INTRCPT2
ICPTROW,b_{000}	ICPTROW,b_{100}
0.09506	-0.00711
-0.00711	0.00320

Ω (as correlations)
```
  1.000   -0.408
 -0.408    1.000
```

Δ

INTRCPT1	AGE8
INTRCPT2	INTRCPT2
ICPTCOL,c_{000}	ICPTCOL,c_{100}
0.01332	0.00656
0.00656	0.00338

Δ (as correlations)

1.000	0.979
0.979	1.000

The value of the log-likelihood function at iteration 1300 = -1.900326E+003

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, π_0					
INTRCPT2,					
INTERCEPT, θ_{00}	2.639580	0.090173	29.272	270	<0.001
DISADV, γ_{001}	-0.001726	0.050288	-0.034	222	0.973
BLACK,					
INTERCEPT, θ_{01}	-0.443355	0.103660	-4.277	270	<0.001
HISPANIC,					
INTERCEPT, θ_{02}	-0.468207	0.098680	-4.745	270	<0.001
For AGE8, π_1					
INTRCPT2,					
INTERCEPT, θ_{10}	0.933753	0.035488	26.312	270	<0.001
DISADV, γ_{101}	-0.050330	0.020853	-2.414	222	0.016
BLACK,					
INTERCEPT, θ_{11}	-0.105109	0.040518	-2.594	270	0.010
HISPANIC,					
INTERCEPT, θ_{12}	-0.036124	0.038978	-0.927	270	0.354

Final estimation of level-1 and level-2 variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1, e_0	0.52484	0.27546	268	6019.63723	<0.001
AGE8, e_{1jk}	0.18811	0.03538	268	1363.77540	<0.001
σ^2,ε	0.40480	0.16386			

Note: The chi-square statistics reported above are based on only 526 of 567 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Final estimation of row level variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/ INTRCPT2/ ICPTROW,b_{000}	0.30832	0.09506	224	79.66634	>0.500
AGE8/ INTRCPT2/ ICPTROW,b_{100}	0.05653	0.00320	224	182.46985	>0.500

Final estimation of column level variance components

Random Effect	Standard Deviation	Variance Component	d.f.	χ^2	p-value
INTRCPT1/INTRCPT2/ ICPTCOL,c_{000}	0.11543	0.01332	73	2085.34935	<0.001
AGE8/INTRCPT2/ ICPTCOL,c_{100}	0.05810	0.00338	73	1337.03181	<0.001

Statistics for the current model

Deviance = 3800.651318
Number of estimated parameters = 18

The results suggest that:

- Compared to their reference group (non-Black and non-Hispanic); African and Hispanic American students on average had a lower mathematics score at age 8 than did white students. Also, African American students had a significantly lower growth rate in mathematics achievement ($\hat{\theta}_{11}$ = -0.105, t = -2.594) than did white students.
- Neighborhood disadvantage had a negative association with the growth rate of the reference group ($\hat{\gamma}_{101}$ = -0.050, t = -2.414).
- The column level variance at level 3 of each growth parameter was substantially reduced (> 50%). The residual variation between neighborhoods in c_{000} (estimated at 0.01332) and in c_{100} (estimated at 0.00338) are less than half of those in the unconditional models (0.02840 and 0.00720).