



## HCM2 model for the Scotland data

|                                   |   |
|-----------------------------------|---|
| 1. Introduction.....              | 1 |
| 2. Description of the data.....   | 1 |
| 3. Creating the command file..... | 2 |
| 4. Interpreting the output.....   | 4 |

### 1. Introduction

All of the applications discussed thus far have involved a strictly hierarchical data structure. Such nesting structures would occur, for example, in a study of neighborhood and school effects on child development in which all children living in the same neighborhood attended the same school, with multiple neighborhoods per school. In this case we would have children at level 1 nested within neighborhoods at level 2 and neighborhoods nested within schools at level 3. Alternatively, we might have a nested structure in which every child attending a given school lived in the same neighborhood, with multiple schools per neighborhood. In this case, we would have children nested within schools nested within neighborhoods. HLM3 can be used to accommodate such three-level nested data structures. However, we typically find, in fact, that children who reside in a specific neighborhood can enroll in one of several schools, and each school might draw students from several neighborhoods. In this case, the data gathered will no longer have a purely nested structure. Instead, a cross-classification of students by two higher-level factors, neighborhoods and schools, arises. To handle this more complex data structure while modeling the developmental influences of neighborhoods and schools requires the use of cross-classified random effects models (HCM2).

### 2. Description of the data

In this example we fit a second model (in a set of three) to the data from a study of neighborhood and school effects on educational attainment in Scotland (Garner & Raudenbush, 1991). The first example described a model that is unconditional at all levels. In this model we set up a level-1 and a row-factor prediction model, with the effect associated with the row-specific predictor specified as fixed. We use prior measures of cognitive skill, verbal reasoning quotient and reading achievement, father's employment status and occupation and father's and mother's education to predict attainment. We also extend the model by

specification of the row- or column-factor prediction model. We shall predict each student's intercept with social deprivation.

**Level-1 file.** The level-1 or within-cell file, ATTAINW.SAV has 2,310 students and 8 variables. The two IDs are NEIGHID for neighborhoods and SCHID for schools. The variables are:

- ATTAIN (a measure of educational attainment)
- P7VRQ (Primary 7 verbal reasoning quotient)
- P7READ (Primary 7 reading test scores)
- DADOCC (father's occupation scaled on the Hope-Goldthorpe scale in conjunction with the Registrar General's social-class index (Willms, 1986))
- DADUNEMP, an indicator for father's unemployment status (1 if unemployed, 0 otherwise)
- DADED, an indicator for father's educational level (1 if schooling past the age of 15, 0 otherwise)
- MOMED, an indicator for mother's educational level (1 if schooling past the age of 15, 0 otherwise)
- MALE, an indicator for student gender (1 if male, 0 if female)

#### **Level-2 row-factor file**

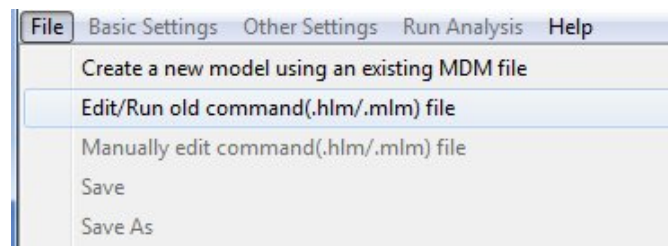
For our neighborhood example, the level-2 row-factor (neighborhood) level file, ATTAINR.SAV, consists data on 1 variable for 542 neighborhoods. The variable is DEPRIVE (a scale measuring social deprivation, which incorporates information on the poverty concentration, health, and housing stock of a local community).

#### **Level-2 column-factor file**

The level-2 column-factor (neighborhood) file, ATTAINCO.SAV, has 17 schools and 1 variable. The variable is DUMMY, a dummy variable.

### **3. Creating the command file**

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 for the unconditional model (ATTAIN.HLM in our example).



To extend our level-1 model, we select P7VCR, P7READ, DADOCC, DADUNEMP, DADED, MOMED, and MALE and grand-mean center all the predictors. In the interest of parsimony, given the small

cell sizes and within-neighborhood sizes, all level-1 coefficients are fixed. (To specify any of them as randomly varying, select the equation containing a specific regression coefficient,  $\pi_p$ , and click on  $b_{p0}$ ).

WHLM: hcm2 MDM File: ATTAIN.MDM

File Basic Settings Other Settings Run Analysis Help

**Outcome**

**Level-1**

>> Row <<

**Column**

ICPTROW  
DEPRIVE

**LEVEL 1 MODEL** (bold italic: grand-mean centering)

$$\text{ATTAIN} = \pi_0 + \pi_1(\text{PTVRQ}) + \pi_2(\text{PTREAD}) + \pi_3(\text{DADOCC}) + \pi_4(\text{DADUNEMP}) + \pi_5(\text{DADED}) + \pi_6(\text{MOMED}) + \pi_7(\text{MALE}) + e$$

**LEVEL 2 MODEL** (bold italic: grand-mean centering)

$$\pi_0 = \theta_0 + b_{00} + c_{00}$$

$$\pi_1 = \theta_1 + b_{10} + c_{10}$$

$$\pi_2 = \theta_2 + b_{20} + c_{20}$$

$$\pi_3 = \theta_3 + b_{30} + c_{30}$$

$$\pi_4 = \theta_4 + b_{40} + c_{40}$$

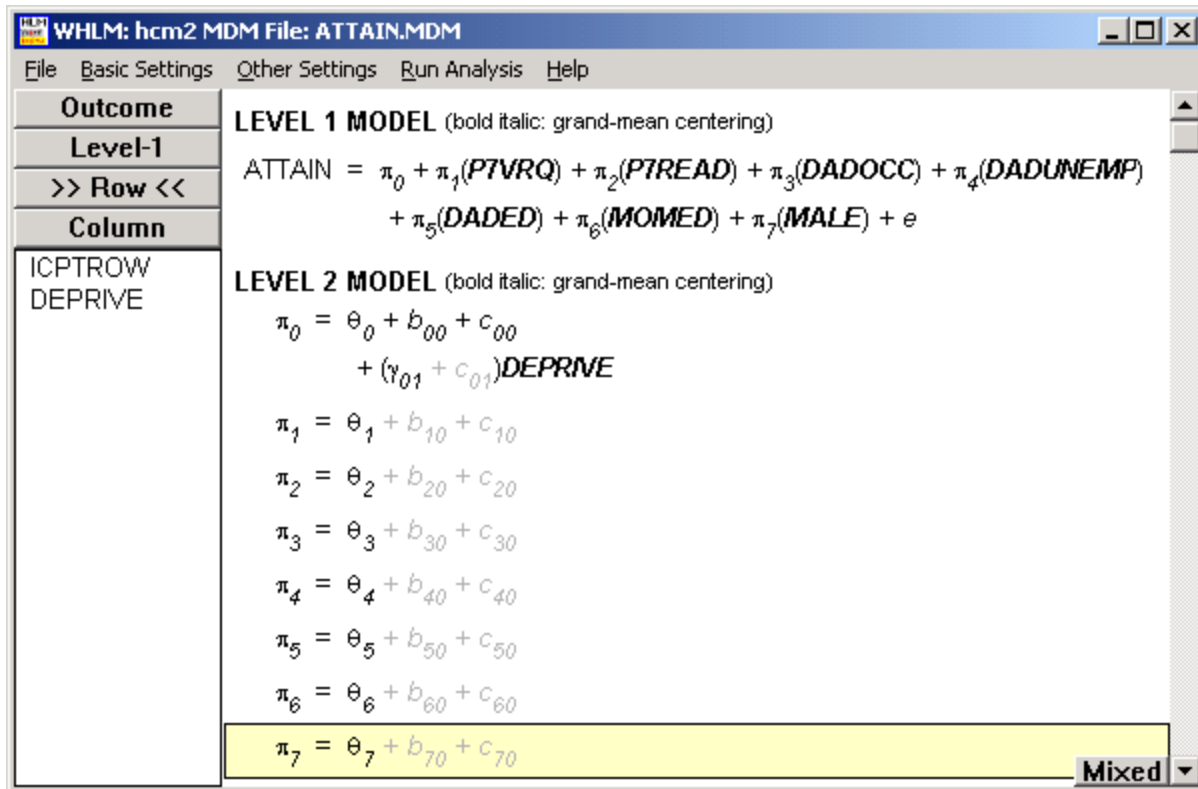
$$\pi_5 = \theta_5 + b_{50} + c_{50}$$

$$\pi_6 = \theta_6 + b_{60} + c_{60}$$

$$\pi_7 = \theta_7 + b_{70} + c_{70}$$

Mixed ▾

To set up the level-2 row-factor prediction model, we select the equation containing  $\pi_0$ . A list box for row-factor variables (>>**Row**<<)) will appear. Click DEPRIVE and apply the grand-mean centering scheme. In the level-2 model, we treated the association between social deprivation and educational attainment as fixed across all schools. We relax this assumption in our next model. The conditional model is displayed below. Note that  $c_{01}$  is disabled.



Save the model before running the analysis.

## 4. Interpreting the output

The results of the analysis are given below.

Problem Title: Conditional Model, with social deprivation effect fixed  
 The data source for this run = ATTAIN.MDM  
 The command file for this run = ATTAIN2.hlm  
 Output file name = hcm2.html  
 The maximum number of level-1 units = 2310  
 The maximum number of row-level units = 524  
 The maximum number of column-level units = 17  
 The maximum number of iterations = 100  
 Method of estimation: full maximum likelihood  
 The maximum number of iterations = 100  
 Z-structure: independent

The outcome variable is ATTAIN

### Summary of the model specified

#### Level-1 Model

$$ATTAIN_{ijk} = \pi_{0jk} + \pi_{1jk}*(P7VRQ_{ijk}) + \pi_{2jk}*(P7READ_{ijk}) + \pi_{3jk}*(DADOCC_{ijk}) + \pi_{4jk}*(DADUNEMP_{ijk}) \\ + \pi_{5jk}*(DADED_{ijk}) + \pi_{6jk}*(MOMED_{ijk}) + \pi_{7jk}*(MALE_{ijk}) + e_{ijk}$$

## Level-2 Model

$$\pi_{0jk} = \theta_0 + b_{00j} + c_{00k} + (\gamma_{01}) * DEPRIVE_j$$

$$\pi_{1jk} = \theta_1$$

$$\pi_{2jk} = \theta_2$$

$$\pi_{3jk} = \theta_3$$

$$\pi_{4jk} = \theta_4$$

$$\pi_{5jk} = \theta_5$$

$$\pi_{6jk} = \theta_6$$

$$\pi_{7jk} = \theta_7$$

P7VRQ P7READ DADOCC DADUNEMP DADED MOMED MALE have been centered around the grand mean.

DEPRIVE has been centered around the grand mean.

For starting values, data from 2310 level-1, 524 row-level and 17 column-level records were used

## Final Results - iteration 34

Iterations stopped due to small change in likelihood function

$$\sigma^2 = 0.45891$$

Trows

INTRCPT1

ICPTROW, $b_{00j}$

0.00014

Tcolumns

INTRCPT1

ICPTCOL, $c_{00k}$

0.00389

The value of the log-likelihood function at iteration 34 = -2.384802E+003

## Final estimation of fixed effects:

| Fixed Effect           | Coefficient | Standard error | t-ratio | Approx. d.f. | p-value |
|------------------------|-------------|----------------|---------|--------------|---------|
| For INTRCPT1, $\pi_0$  |             |                |         |              |         |
| INTERCEPT, $\theta_0$  | 0.094740    | 0.021133       | 4.483   | 1769         | <0.001  |
| DEPRIVE, $\gamma_{01}$ | -0.156676   | 0.025178       | -6.223  | 522          | <0.001  |
| For P7VRQ, $\pi_1$     |             |                |         |              |         |
| INTERCEPT, $\theta_1$  | 0.027556    | 0.002263       | 12.176  | 1769         | <0.001  |
| For P7READ, $\pi_2$    |             |                |         |              |         |
| INTERCEPT, $\theta_2$  | 0.026291    | 0.001749       | 15.028  | 1769         | <0.001  |
| For DADOCC, $\pi_3$    |             |                |         |              |         |
| INTERCEPT, $\theta_3$  | 0.008165    | 0.001359       | 6.008   | 1769         | <0.001  |

|                       |           |          |        |      |        |
|-----------------------|-----------|----------|--------|------|--------|
| For DADUNEMP, $\pi_4$ |           |          |        |      |        |
| INTERCEPT, $\theta_4$ | -0.120771 | 0.046779 | -2.582 | 1769 | 0.010  |
| For DADED, $\pi_5$    |           |          |        |      |        |
| INTERCEPT, $\theta_5$ | 0.144426  | 0.040782 | 3.541  | 1769 | <0.001 |
| For MOMED, $\pi_6$    |           |          |        |      |        |
| INTERCEPT, $\theta_6$ | 0.059440  | 0.037381 | 1.590  | 1769 | 0.112  |
| For MALE, $\pi_7$     |           |          |        |      |        |
| INTERCEPT, $\theta_7$ | -0.056058 | 0.028401 | -1.974 | 1769 | 0.049  |

**Final estimation of row and level-1 variance components:**

| Random Effect                | Standard Deviation | Variance Component | d.f. | $\chi^2$  | p-value |
|------------------------------|--------------------|--------------------|------|-----------|---------|
| INTRCPT1/ ICPTROW, $b_{00j}$ | 0.01184            | 0.00014            | 522  | 548.81015 | 0.202   |
| level-1, e                   | 0.67743            | 0.45891            |      |           |         |

**Final estimation of column level variance components:**

| Random Effect                | Standard Deviation | Variance Component | d.f. | $\chi^2$ | p-value |
|------------------------------|--------------------|--------------------|------|----------|---------|
| INTRCPT1/ ICPTCOL, $c_{00k}$ | 0.06239            | 0.00389            | 15   | 36.38151 | 0.002   |

**Statistics for the current model**

Deviance = 4769.604659

Number of estimated parameters = 12

Several features of the results are remarkable:

- Several level-1 covariates are significantly related to educational attainment, with especially large effects for P7READ and P7VRQ.
- The residual level-1 variance is estimated to be 0.459, implying that 43% of the unconditional level-1 variance (estimated at 0.799) is accounted for by the covariates.
- Controlling these level-1 effects, a highly significant negative effect of social deprivation appears ( $\hat{\gamma}_{01} = -0.157$ ,  $t = -6.22$ ).
- The residual variation between neighborhoods,  $\tau_{b00}$ , (estimated at 0.0001), and between schools,  $\tau_{c00}$  (estimated at 0.004) are close to zero; compare to the unconditional variance estimates (0.141 and 0.075). The level-2 neighborhood variance component was substantially reduced.