Scientific Software International

Testing heterogeneity of level-1 variance using the HSB data

1. Description of the data

High School and Beyond (**HS&B**) is a national longitudinal study originally funded by the United States Department of Education's National Center for Education Statistics (NCES) as a part of their longitudinal studies program. Its purpose was to document the "educational, vocational, and personal development of young people following them over time as they begin to take on adult roles and responsibilities". Here a subset of the data representing 160 schools and a total of 7,185 students is used to illustrate the fitting and interpretation of a two-level hierarchical linear model.

This example is the third in a set of seven examples based on these data described on the HLM Support page and used the MDM file created in the first example.

Level-1 file. For our HS&B example data, the level-1 file (HSB1.SAV) has 7,185 cases and four variables (not including the SCHOOL ID). The variables are:

- MINORITY, an indicator for student ethnicity (1 = minority, 0 = other)
- FEMALE, an indicator for student gender (1 = female, 0 = male)
- SES, a standardized scale constructed from variables measuring parental education, occupation, and income
- MATHACH, a measure of mathematics achievement

Data for the first ten cases in HSB1.SAV are shown in Fig. 1.1.

Note: level-1 cases must be grouped together by their respective level-2 unit ID. To assure this, sort the level-1 file by the level-2 unit ID field prior to entering the data into HLM2.

	id	minority	female	ses	mathach
1	1224	0	1	-1.528	5.876
2	1224	0	1	-.588	19.708
3	1224	0	0	-.528	20.349
4	1224	0	0	-.668	8.781
5	1224	0	0	-.158	17.898
6	1224	0	0	.022	4.583
7	1224	0	1	-.618	-2.832
8	1224	0	0	-.998	.523
9	1224	0	1	-.888	1.527
10	1224	0	0	-.458	21.521

Figure 1.1 First ten cases in HSB1.SAV

Level-2 file. At level 2, the illustrative data set HSB2.SAV consists of 160 schools with 6 variables per school. The variables are:

- SIZE (school enrollment)
- SECTOR (1 = Catholic, 0 = public)
- PRACAD (proportion of students in the academic track)
- DISCLIM (a scale measuring disciplinary climate)
- HIMNTY (1 = more than 40% minority enrollment, 0 = less than 40%)
- MEANSES (mean of the SES values for the students in this school who are included in the level-1 file)

The data for the first ten schools are displayed in Fig 1.2.

	id	size	sector	pracad	disclim	himinty	meanses
1	1224	842	0	.350	1.597	0	-.428
2	1288	1855	0	.270	.174	0	.128
3	1296	1719	0	.320	-.137	1	-.420
4	1308	716	1	.960	-.622	0	.534
5	1317	455	1	.950	-1.694	1	.351
6	1358	1430	0	.250	1.535	0	-.014
7	1374	2400	0	.500	2.016	0	-.007
8	1433	899	1	.960	-.321	0	.718
9	1436	185	1	1.000	-1.141	0	.569
10	1461	1672	0	.780	2.096	0	.683

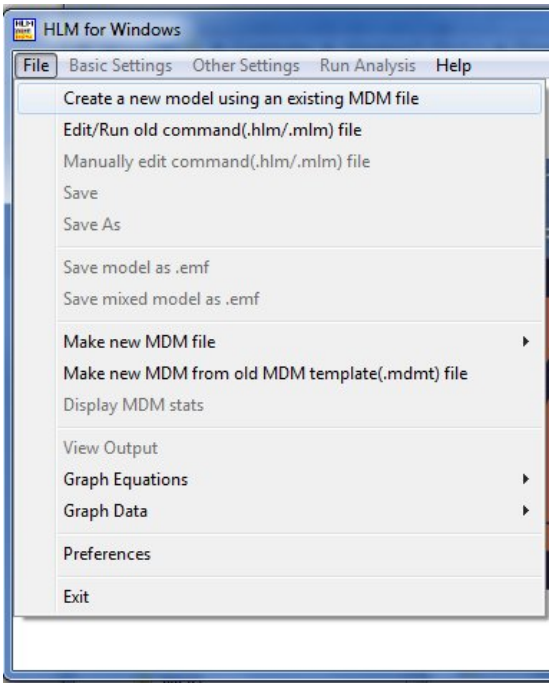
Figure 1.2 First ten cases in HSB2.SAV

2. Creating the command file

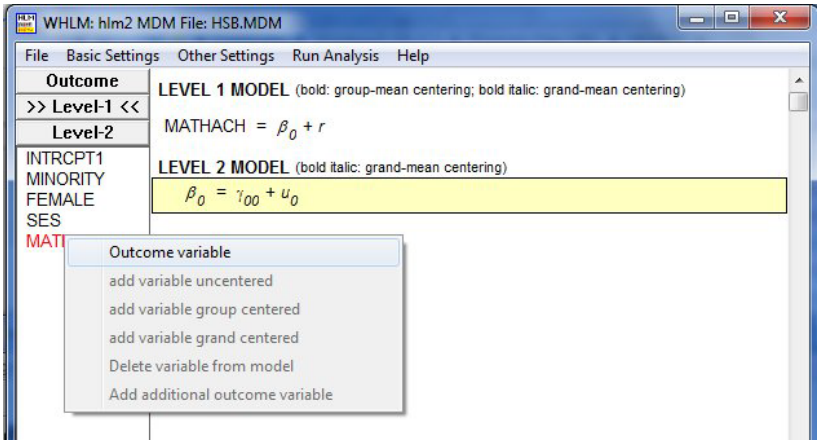
Users may wish to estimate models that allow for heterogeneous level-1 variances. A simple example (see HSB2.HLM) using the HS&B data would be a model that postulates that the two

genders have different means in and variances of math achievement scores. To specify a model that hypothesizes different central tendency and variability in math achievement for the two genders, the model must first be set up, using the existing MDM file.

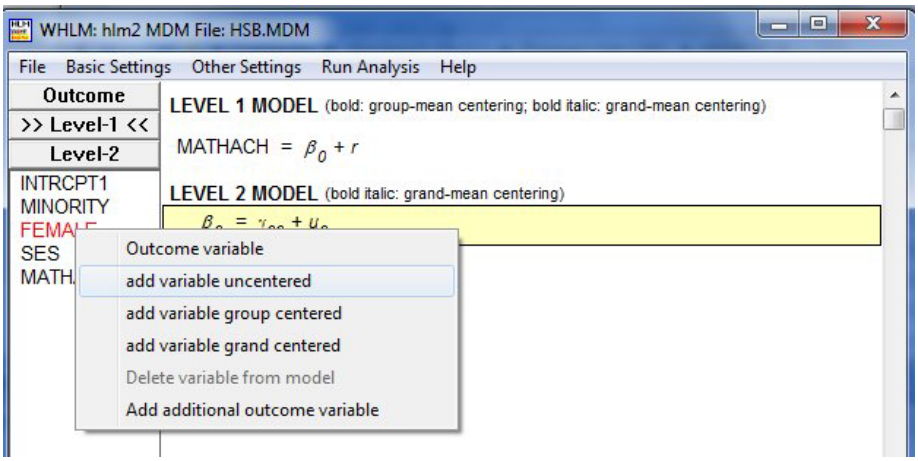
Start by selecting the **Create a new model using an existing MDM file** option from the **File** menu.



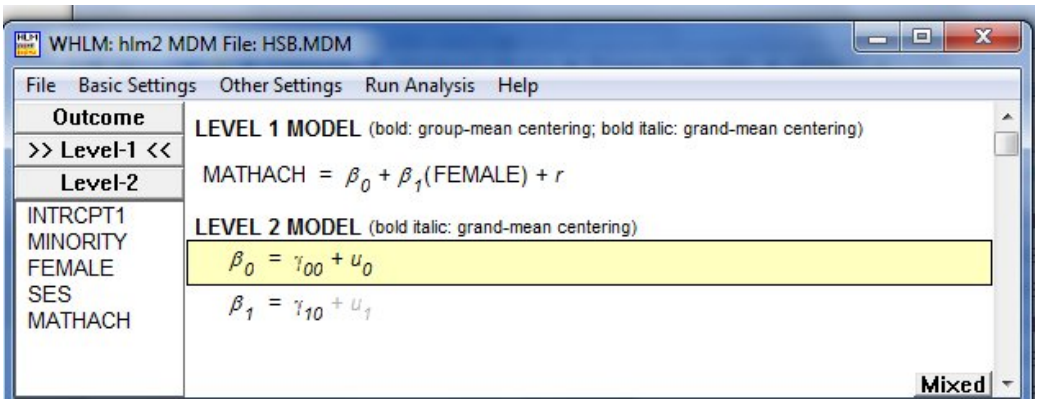
Next, browse for the MDM file **HSB.MDM**. Click **OK** to return to the main window, where the names of the variables contained in the MDM file are now displayed at the left of the window. Select the variable MATHACH as outcome variable, as shown below.



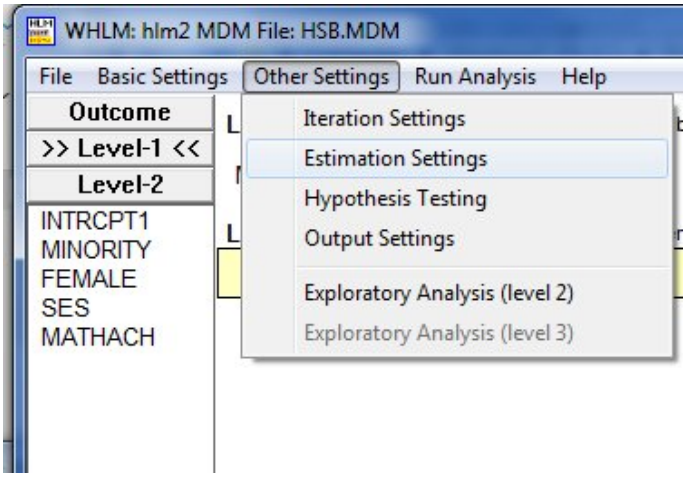
Select the variable FEMALE, representing students' gender, as level-1 predictor.



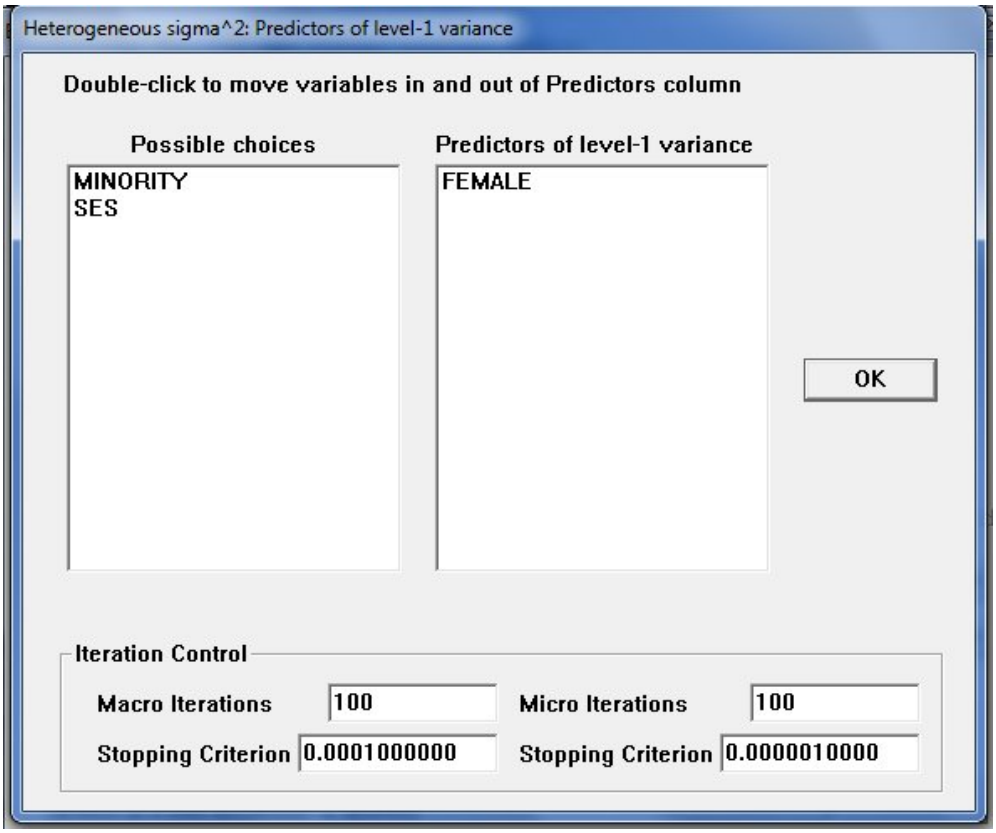
This completes the specification of the basic model of interest.



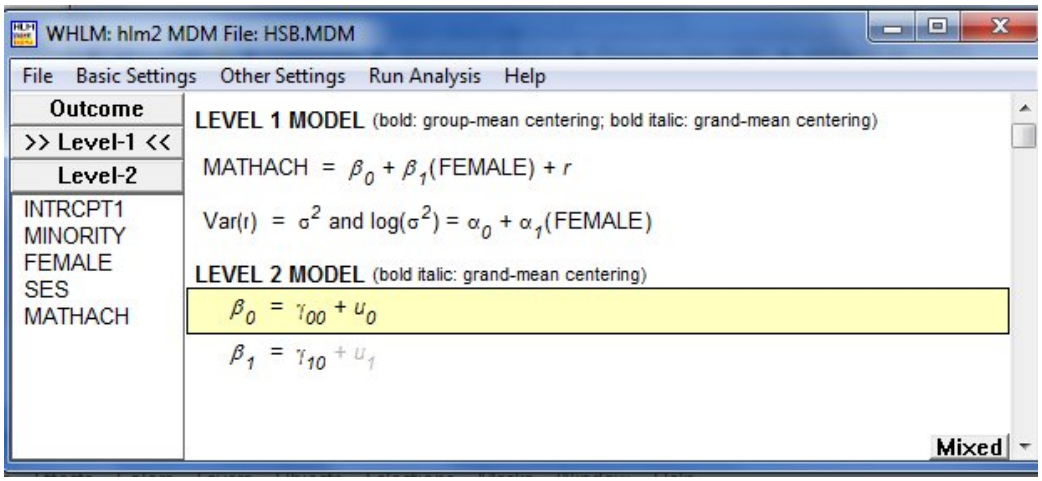
Select the **Estimation Settings** option from the **Other Settings** menu on the main menu bar.



On the **Estimation Settings** dialog box, click the **Heterogeneous sigma^2** button to open the **Heterogeneous sigma^2: Predictors of level-1 variance** dialog box. Double-click on the variable FEMALE (in the left column) to move it to the column of **Predictors of level-1 variance**. Click **OK** to return to the **Estimation Settings** dialog box, and **OK** to return to the main window.



The final model is now displayed, as shown below. Save this model using the **File**, **Save As** option prior to clicking **Run**.



The model estimated is a log-linear model for the level-1 variances, which can be generally stated as

$$\sigma_{ij}^2 = \exp\left\{\alpha_0 + \alpha_1 FEMALE_{ij}\right\}$$

Once the iterative procedure has converged, the output will automatically be displayed in the format set on the **Preferences** dialog box.

3. Interpreting the output

Partial output for this model is shown below. The first set of results is for the homogeneous model:

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	13.345654	0.261011	51.131	159	<0.001
For FEMALE slope, β_1					
INTRCPT2, γ_{10}	-1.381355	0.185852	-7.433	159	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	13.345654	0.261027	51.127	159	<0.001
For FEMALE slope, β_1					
INTRCPT2, γ_{10}	-1.381355	0.185897	-7.431	159	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	2.93070	8.58901	122	589.08609	<0.001
FEMALE slope, u_1	0.80442	0.64709	122	139.99921	0.127
level-1, *r*	6.22361	38.73330			

Note: The chi-square statistics reported above are based on only 123 of 160 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Statistics for the current model
Deviance = 47049.462733
Number of estimated parameters = 6

This is followed by the results for the heterogeneous model.

$\text{Var}(R) = \sigma^2$ and $\log(\sigma^2) = \alpha_0 + \alpha_1(\text{FEMALE})$

Model for level-1 variance

Parameter	Coefficient	Standard Error	*Z*-ratio	*p*-value
INTRCPT1 ,α_0	3.70388	0.024798	149.364	0.000
FEMALE ,α_1	-0.09127	0.034089	-2.677	0.008

Summary of Model Fit

Model	Number of Parameters	Deviance
1. Homogeneous σ^2	6	47049.46273
2. Heterogeneous σ^2	7	47042.27724

Model Comparison	χ^2	d.f.	*p*-value
Model 1 vs Model 2	7.18549	1	0.007

The Z-ratio for α_1 (Z =-2.677) for FEMALE indicate that the math achievement scores of males are on average higher than those for females. Furthermore, a comparison of the fits of the models suggests that the model with heterogeneous within-school variances appears appropriate (χ^2 = 7.45604, df = 1).

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	13.347288	0.261178	51.104	159	<0.001
For FEMALE slope, β_1					
INTRCPT2, γ_{10}	-1.384269	0.186401	-7.426	159	<0.001

Final estimation of fixed effects (with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	13.347288	0.261028	51.134	159	<0.001
For FEMALE slope, β_1					
INTRCPT2, γ_{10}	-1.384269	0.185836	-7.449	159	<0.001

Final estimation of variance components

Random Effect	Standard Deviation	Variance Component	*d.f.*	χ^2	*p*-value
INTRCPT1, u_0	2.91733	8.51079	122	561.99707	<0.001
FEMALE slope, u_1	0.80941	0.65515	122	139.31550	0.135

Note: The chi-square statistics reported above are based on only 123 of 160 units that had sufficient data for computation. Fixed effects and variance components are based on all the data.

Statistics for the current model
Deviance = 47042.277245
Number of estimated parameters = 7