Scientific Software International

An annotated example of HLM2 analysis of incomplete data

HLM analysis of incomplete data has three major steps (Shin, 2013; Shin & Raudenbush, 2007):

1. Specify the desired models given incompletely observed multilevel data as described in the first example on this topic.

The model specified by the user is the same as that just estimated. However, both the mathematics and reading proficiency variables at level 1 and the INCOME variable at level 2 have missing data. The GRADE variable has no missing data. With the analysis of incomplete multilevel data, HLM automatically reparameterizes the models as the joint distribution of the math and reading outcomes and the variables subject to missingness conditional on the completely observed variables. HLM then efficiently estimates the joint distribution using maximum likelihood under the assumption of multivariate normality.

2. Generate multiply-imputed complete data based on the ML estimates of the joint model. The procedure consists of

a. Opening the **Other Settings** menu and selecting the **Estimation Settings** to open the **Estimation Settings – HLM2** dialog box (See Figure 4).

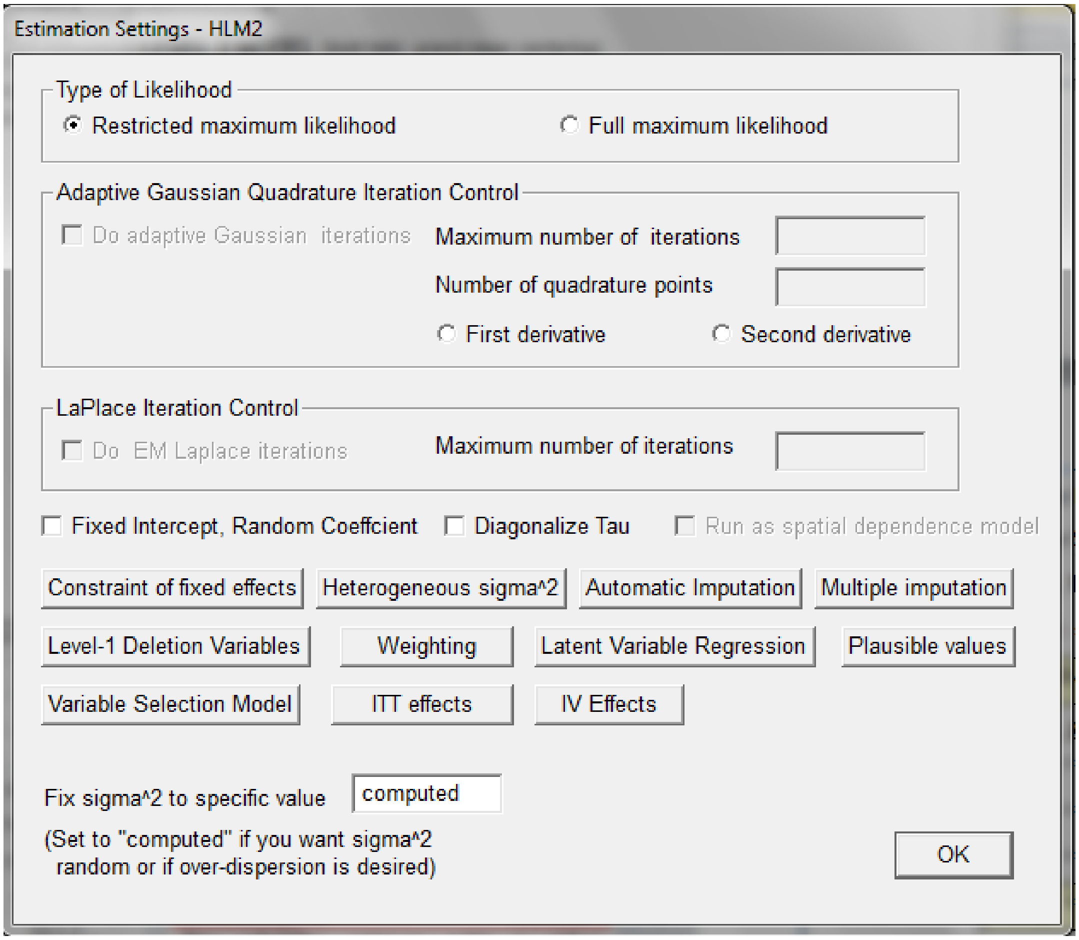

Figure 4 Estimation Settings – HLM2 dialog box

Click **Automatic Imputation** to open the **Automatic Multiple Imputation** dialog box (see Figure 5). We have selected 10 data sets.

We also choose 2 “augmentation” variables, sometimes called auxiliary variables to improve the imputations. “PARSCAR” is a measure of occupational status and should be a good predictor of income. “BLACK” is an indicator for African-American background.

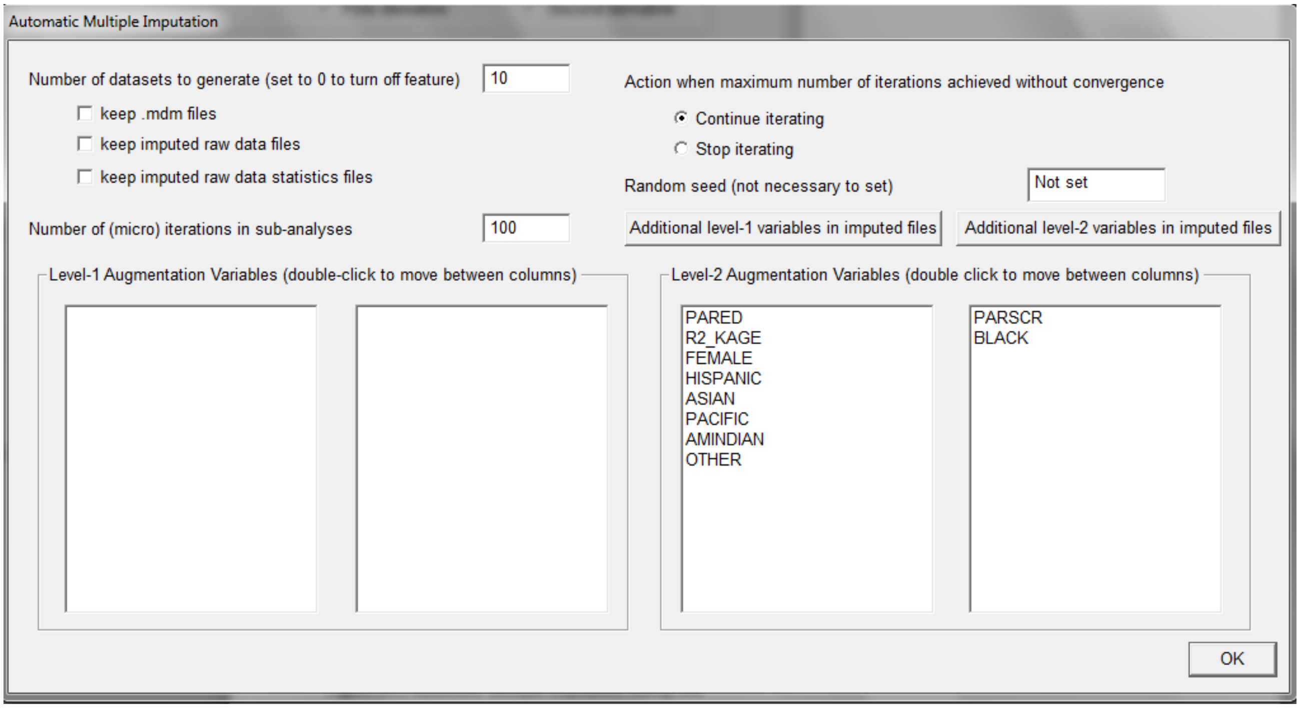

Figure 5 Automatic Multiple Imputation dialog box

c. Enter the number of datasets to generate. We enter **10** as a pilot for this example. There is an option to set a random seed number. Users have the option to save the multivariate data matrix files for each data set and keep imputed raw data files that they may like to analyze further using HLM or another program. They can also ask for a record of imputed raw data statistics (see Figure 5).

3. Analyze the desired model by complete-data analysis given the multiple imputation. Click **OK** on the **Automatic Multiple Imputation** dialog box, then click **OK** on the **Estimation Settings –HLM2** dialog box. Save and run the model.

The results of the analysis are given below.

Specifications for this HLM2 run
Problem Title: multiple imputation analysis with augmentation

The data source for this run = ecls_growth2.mdm
The command file for this run = mult_imput_aug.hlm
Output file name = mult_imp_aug_avg.html
The maximum number of level-1 units = 148470
The maximum number of level-2 units = 21210
The maximum number of iterations = 100
Method of estimation: full maximum likelihood
Automatic imputation random number seed: -1563333359

Summary of the model specified

Level-1 Model

$READING_{ti} = \pi_{0i} + \pi_{1i}*(GRADE_{ti}) + e_{ti}$
$MATH_{ti} = \pi_{2i} + \pi_{3}*(GRADE_{ti}) + e_{ti}$

Level-2 Model

$\pi_{0i} = \beta_{00} + \beta_{01}*(SQRTINC_{i}) + r_{0i}$
$\pi_{1i} = \beta_{10}$
$\pi_{2i} = \beta_{20} + \beta_{21}*(SQRTINC_{i}) + r_{2i}$
$\pi_{3i} = \beta_{30}$

Run-time deletion has reduced the number of level-2 records to 21177

Imputation Model Results - Iteration 12
NOTE: level-1 and level-2 slopes have been duplicated across all level-2 equations.

Iterations stopped due to small change in likelihood function

Σ*

READING /INTRCPT2	384.59710	212.27386
MATH /INTRCPT2	212.27386	218.08554

Standard errors of Σ*

READING /INTRCPT2	2.01342	1.32284
MATH /INTRCPT2	1.32284	1.13236

Σ (as correlations)

READING /INTRCPT2	1.000	0.733
MATH /INTRCPT2	0.733	1.000

Tau*

READING	219.61749	162.47932	18.88858	89.97725
MATH	162.47932	172.81037	15.57391	72.71537
INTRCPT1/ SQRTINC	18.88858	15.57391	8.49343	24.58023
INTRCPT1/ PARSCR	89.97725	72.71537	24.58023	271.33135

Standard errors of Tau*

READING	3.11204	2.39569	0.43120	2.16650
MATH	2.39569	2.24130	0.36548	1.83209
INTRCPT1/ SQRTINC	0.43120	0.36548	0.09854	0.41885
INTRCPT1/ PARSCR	2.16650	1.83209	0.41885	2.70453

Tau (as correlations)

READING	1.000	0.834	0.437	0.369
MATH	0.834	1.000	0.407	0.336
INTRCPT1 / SQRTINC	0.437	0.407	1.000	0.512
INTRCPT1 / PARSCR	0.369	0.336	0.512	1.000

Final estimation of fixed effects (Imputation model)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For READING /INTRCPT2					
INTRCPT2	105.186989	0.137141	767.001	106050	<0.001
BLACK	-8.957698	0.348835	-25.679	106050	<0.001
GRADE	17.364407	0.025382	684.125	21209	<0.001
For MATH /INTRCPT2					
INTRCPT2	85.276019	0.116036	734.912	106050	<0.001
BLACK	-9.810738	0.295823	-33.164	106050	<0.001
GRADE	14.585079	0.019155	761.426	21209	<0.001
For SQRTINC					
INTRCPT2	7.892670	0.024780	318.504	21208	<0.001
BLACK	-2.151829	0.068689	-31.327	21208	<0.001
For PARSCR					
INTRCPT2	42.878649	0.125462	341.766	21208	<0.001
BLACK	-9.272788	0.327137	-28.345	21208	<0.001

The value of the log-likelihood function at iteration 12 = -9.006426E+005

Note that the two outcomes, reading and math (at level 1) are in the multivariate model as outcomes. Also note that the predictor SQRTINC, which is subject to missingness, is also an outcome. In addition, the augmentation variable PARSCR is an additional outcome because HLM has found that it is subject to missingness. All outcomes are regressed on GRADE and BLACK, which are completely observed.

The results for the user specified model are below.

Final Imputation Model Results - 10 Imputations

Σ

READING /INTRCPT2	385.72825	211.59869
MATH /INTRCPT2	211.59869	217.84582

Standard errors of Σ

READING /INTRCPT2	1.29524	1.28592
MATH /INTRCPT2	0.54992	1.23742

Σ (as correlations)

READING /INTRCPT2	1.000	0.730
MATH /INTRCPT2	0.730	1.000

Tau

READING	187.73964	137.45631
MATH	137.45631	154.13887

Standard errors of Tau

READING	2.37041	1.86367
MATH	1.86367	1.80465

Tau (as correlations)

READING	1.000	0.808
MATH	0.808	1.000

Final estimation of fixed effects

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For READING /INTRCPT2					
INTRCPT2	86.624227	0.293843	294.797	4814	<0.001#
SQRTINC	2.229091	0.036345	61.331	12	<0.001#
GRADE	17.371634	0.081062	214.301	9	<0.001#
For MATH /INTRCPT2					
INTRCPT2	69.271059	0.581186	119.189	14	<0.001#
SQRTINC	1.882094	0.081095	23.209	12	<0.001#
GRADE	14.593083	0.056957	256.214	10	<0.001#

The p-vals above marked with a "#" should regarded as a rough approximation.

Note the small degrees of freedom. This reflects the large amount of missing data, particularly on income. The degrees of freedom can be increased by increasing M, the number of data sets.

Cross-Level Interactions

When working with the missing data program, cross-level interactions are presented differently than has been standard in HLM.

Consider for example the univariate model

$$Y_{ti} = \pi_{0i} + \pi_{1i} a_{ti} + e_{ti}$$

$$\pi_{0i} = \beta_{00} + \beta_{01} X_i + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} X_i$$

Suppose that the predictor X is subject to missingness. It must therefore be put on the left side of the imputation model. However, modeling X as a predictor of π_{1i} induces an interaction between a and X as we can see by inspecting the mixed model

$$Y_{ti} = \beta_{00} + \beta_{01} X_i + \beta_{10} \pi_{0i} + \beta_{11} X_i a_{ti} + r_{0i} + e_{ti} .$$

Because X is missing, so is $X_i a_{ti}$, so it must also be put on the left. Technically, many such interaction terms will not follow a normal distribution. However, the robustness of the procedure to failure of normality can typically be improved by centering both predictors *a* and *X*.

Here is how HLM will represent the model results:

Final estimation of fixed effects
(with robust standard errors)

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value	Fraction of Missing Info.
For INTRCPT1, π_0						
INTRCPT2, β_{00}	84.780324	0.381040	222.497	71	<0.001	0.282
SQRTINC, β_{01}	2.475539	0.047462	52.158	60	<0.001	0.301
For GRADE slope, π_1						
INTRCPT2, β_{10}	15.345154	0.098909	155.145	18	<0.001	0.470
For CPROD1 slope, π_2						
INTRCPT2, β_{20}	0.249093	0.012099	20.588	16	<0.001	0.483

CPROD1 is the generated cross-product of GRADE*SQRTINC.