CHAPTER 12: MULTILEVEL MODELLING OF HUNGARIAN SCIENCE DATA

1. Data Description

The data consist of scores on four tests: a core test booklet – with components in earth science, physics and biology – plus an additional biology test taken by a random subsample of students. Each test was scored out of ten, but we analyse standardised scores (with mean zero and variance one).

The data are in files *hungary.wsz*, *hungary.sav* and *hungary.txt*. The files contain the following variables:

SCHOOL	School identifier
STUDENT	Student identifier
ES_CORE	Score in earth science test (from core booklet)
BIOL_CORE	Score in biology test (from core booklet)
PHYS_CORE	Score in physics test (from core booklet)
BIOL_R	Score in biology test (taken by random subsample of students)
ES_CORE_ST	Standardised earth science score
BIOL_CORE_ST	Standardised biology score
PHYS_CORE_ST	Standardised physics score
BIOL_R_ST	Standardised biology score (for test taken by random subsample)
FEMALE	Student's sex (1=female, 0=male)

2. MLwiN Resources

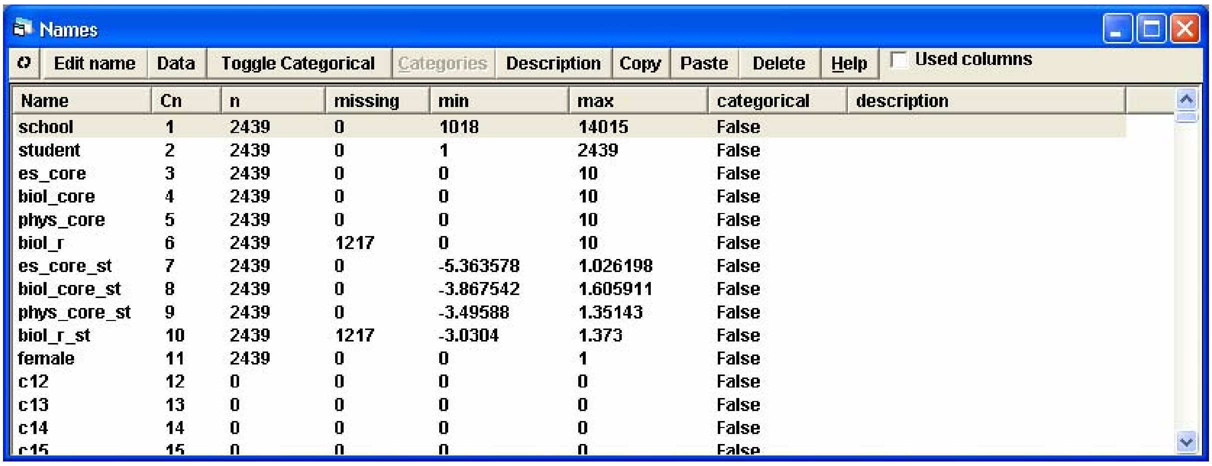
This document contains instructions for reproducing the analysis of the Hungarian data presented in Chapter 12 using MLwiN v2.10. More detailed accounts of how to use MLwiN (including how to interpret the results) can be found in the MLwiN User Guides and the online course developed by the Centre for Multilevel Modelling, University of Bristol.

- Rasbash, J., Steele, F., Browne, W.J. and Prosser, B. (2005) *A User's Guide to MLwiN*. Centre for Multilevel Modelling, University of Bristol. Download free of charge from http://www.cmm.bris.ac.uk/MLwiN/download/. Chapter 14 covers multivariate models.
- *LEMMA: Learning Environment for Multilevel Modelling and Applications*. A freely accessible online course at http://www.cmm.bris.ac.uk/learning-training/course.shtml.
- Browne, W.J. (2005) *MCMC Estimation in MLwiN*. Centre for Multilevel Modelling, University of Bristol. Download free from http://www.cmm.bris.ac.uk/MLwiN/download/. Chapter 19 covers multilevel factor analysis.

3. A Simple Multivariate Model (Section 12.7)

- Start MLwiN and from the **File** menu select **Open worksheet**
- Open the file **hungary.wsz**

When the worksheet is opened, the filename will appear in the title bar of the main window. The **Names** window will also appear, giving a summary of the data in the worksheet:



Name	Cn	n	missing	min	max	categorical	description
school	1	2439	0	1018	14015	False	
student	2	2439	0	1	2439	False	
es_core	3	2439	0	0	10	False	
biol_core	4	2439	0	0	10	False	
phys_core	5	2439	0	0	10	False	
biol_r	6	2439	1217	0	10	False	
es_core_st	7	2439	0	-5.363578	1.026198	False	
biol_core_st	8	2439	0	-3.867542	1.605911	False	
phys_core_st	9	2439	0	-3.49588	1.35143	False	
biol_r_st	10	2439	1217	-3.0304	1.373	False	
female	11	2439	0	0	1	False	
c12	12	0	0	0	0	False	
c13	13	0	0	0	0	False	
c14	14	0	0	0	0	False	
c15	15	0	0	0	0	False	

We begin by fitting a simple multivariate model, taking the four standardised test scores as the responses. We will ignore school effects for now.

The first step is to create a ‘constant’ variable (a column of 1’s). The coefficient of this variable will be the intercept. (Note that the constant vector is automatically created when we change to simple notation, but multivariate models can only be specified using general notation.)

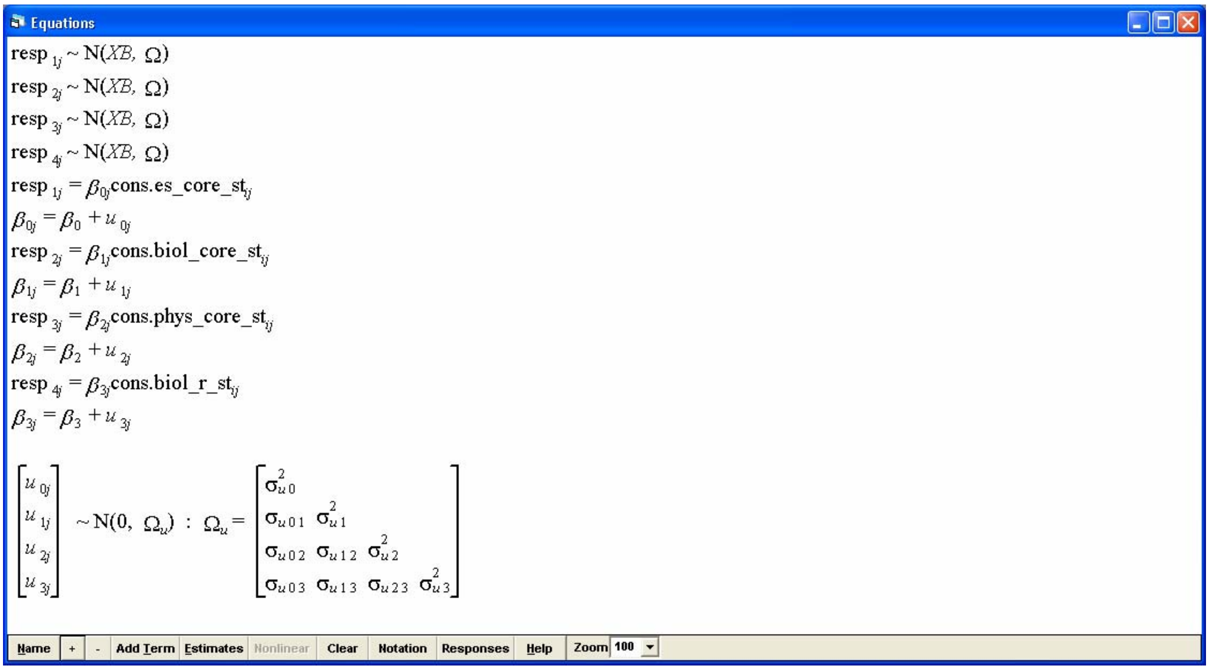
- From the **Data Manipulation** menu, select **Generate vector**
- For **Output column** select **C12**
- Next to **Number of copies** enter **2439** (the number of cases)
- Next to **Value** enter **1**
- Click **Generate**
- Go to the **Names** window and use **Edit name** to call the new variable (C12) **CONS**

To specify a multivariate model:

- From the **Model** menu, select **Equations**
- Click **Responses** at the bottom of the **Equations** window, and select **ES_CORE_ST**, **BIOL_CORE_ST**, **PHYS_CORE_ST** and **BIOL_R_ST** (using ctrl-click to make multiple selections)
- Click **Done**. Four equations should appear (with **resp**$_1$ to **resp**$_4$ as the four response variables)
- Click on **resp**$_1$. Change **N levels** to **2-ij**, and for **level 2 (j)** select **STUDENT**
- Click **Done**
- Click **Add Term** and from the **Variable** list select **CONS**

- Click **Add separate coefficients** (to allow a separate intercept for each response variable)
- Click β_0, check **j(student_long)**, then click **Done**. This causes a residual to be added to the equation for the first response.
- Repeat for the other three responses.
- Click **+** twice to see the full model specification

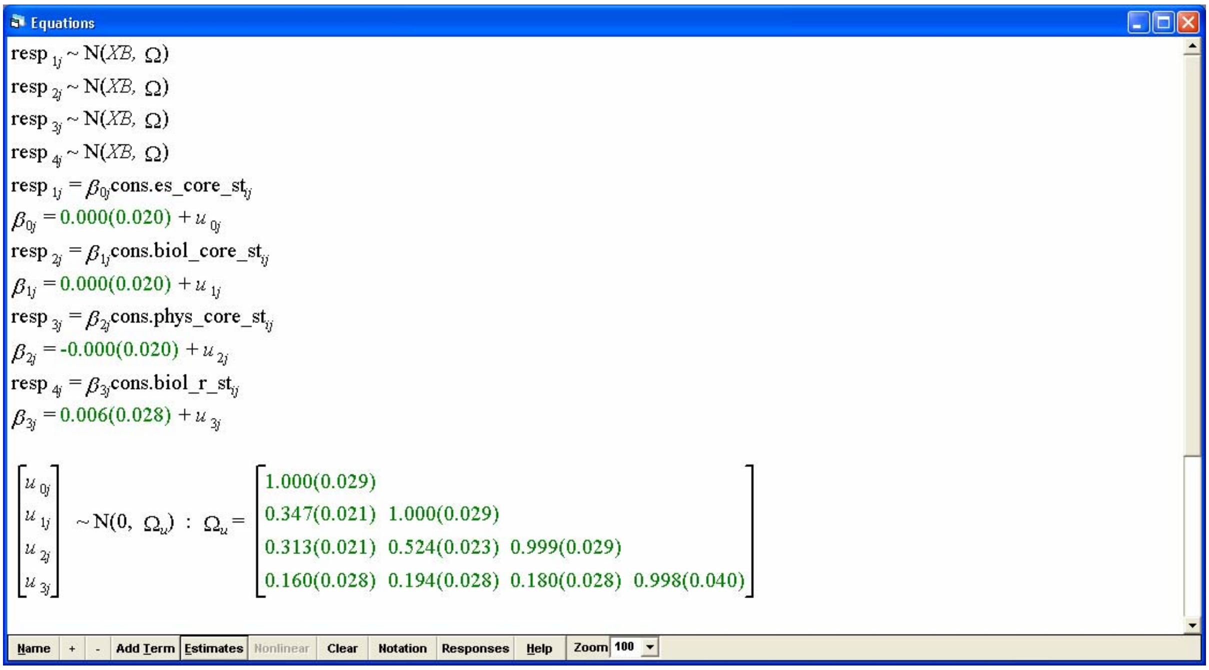
The model should look like this.



To fit the model:

- Click **Start** to fit the model
- Click **Estimates** twice to see the parameter estimates

You should get the following results (matching the correlation matrix in Table 12.6). Two of the variance estimates are a little less than one; this is due to rounding errors in the standardisation.

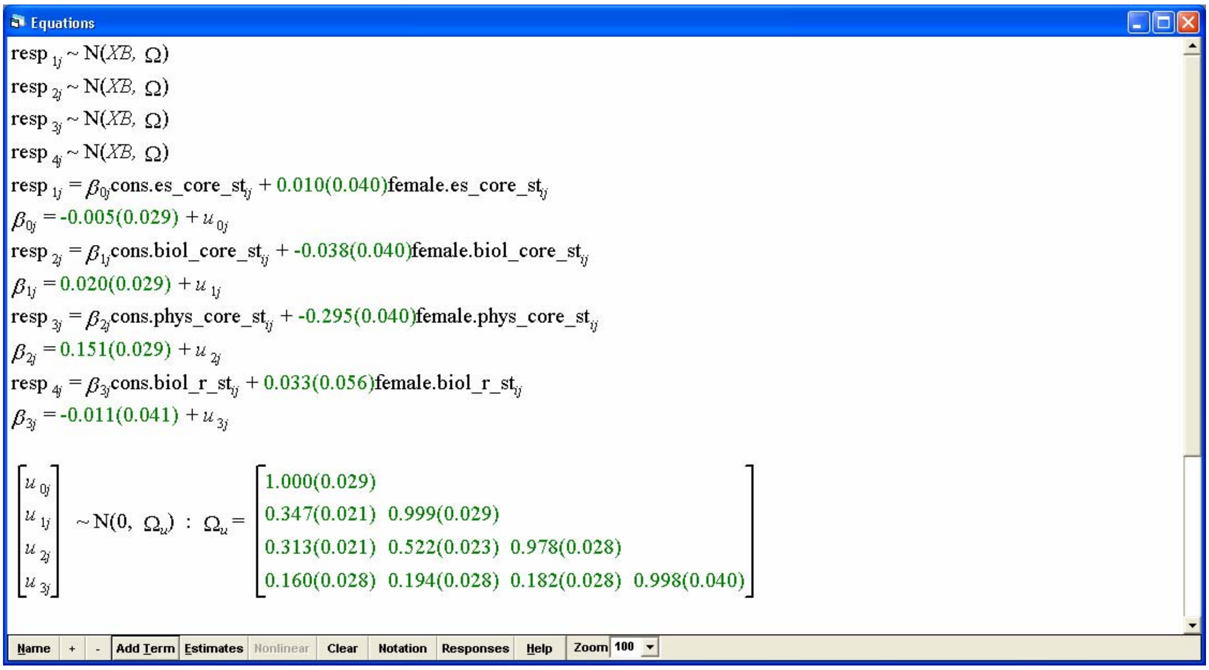


4. Multivariate Regression with Gender Effects (Section 12.7)

To allow for gender effects on each test score:

- In the **Equations** window, click **Add Term**
- Select **FEMALE** from the variable list and click **Add separate coefficients**
- Click **Start** to fit the new model

You should get the following results (given in Table 12.7).



5. Multilevel Multivariate Regression with Gender Effects (Section 12.7)

We now extend the multivariate regression model to allow for school effects. In MLwiN we specify a three-level model with responses at level 1, students at level 2, and schools at level 3.

- In the **Equations** window, click on **resp$_1$**. Change **N levels** to **3-ijk**, and for **level 3 (k)** select **SCHOOL**
- Click **Done**
- Click β_{0j}, check **k(school_long)**, then click **Done**. This causes a school-level residual to be added to the equation for the first response.
- Repeat for the other three responses
- Click **Estimates** to see the model in mathematical notation

The model should look like this:

$$\mathrm{resp}_{1jk} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{resp}_{2jk} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{resp}_{3jk} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{resp}_{4jk} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{resp}_{1jk} = \beta_{0jk}\mathrm{cons.es\_core\_st}_{ijk} + \beta_4\mathrm{female.es\_core\_st}_{ijk}$$
$$\beta_{0jk} = \beta_0 + v_{0k} + u_{0jk}$$
$$\mathrm{resp}_{2jk} = \beta_{1jk}\mathrm{cons.biol\_core\_st}_{ijk} + \beta_5\mathrm{female.biol\_core\_st}_{ijk}$$
$$\beta_{1jk} = \beta_1 + v_{1k} + u_{1jk}$$
$$\mathrm{resp}_{3jk} = \beta_{2jk}\mathrm{cons.phys\_core\_st}_{ijk} + \beta_6\mathrm{female.phys\_core\_st}_{ijk}$$
$$\beta_{2jk} = \beta_2 + v_{2k} + u_{2jk}$$
$$\mathrm{resp}_{4jk} = \beta_{3jk}\mathrm{cons.biol\_r\_st}_{ijk} + \beta_7\mathrm{female.biol\_r\_st}_{ijk}$$
$$\beta_{3jk} = \beta_3 + v_{3k} + u_{3jk}$$

$$\begin{bmatrix} v_{0k} \\ v_{1k} \\ v_{2k} \\ v_{3k} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_v)\ :\ \Omega_v = \begin{bmatrix} \sigma^2_{v0} & & & \\ \sigma_{v01} & \sigma^2_{v1} & & \\ \sigma_{v02} & \sigma_{v12} & \sigma^2_{v2} & \\ \sigma_{v03} & \sigma_{v13} & \sigma_{v23} & \sigma^2_{v3} \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} \sigma^2_{u0} & & & \\ \sigma_{u01} & \sigma^2_{u1} & & \\ \sigma_{u02} & \sigma_{u12} & \sigma^2_{u2} & \\ \sigma_{u03} & \sigma_{u13} & \sigma_{u23} & \sigma^2_{u3} \end{bmatrix}$$

*-2*loglikelihood(IGLS Deviance)* = 22924.670(8539 of 9756 cases in use)

To fit the model:

- Click **Start** to fit the model
- Click **Estimates** twice to see the parameter estimates

You should get the following results (given in Table 12.8).

$$
\begin{aligned}
&\text{resp}_{1jk} \sim \mathrm{N}(XB,\ \Omega)\\
&\text{resp}_{2jk} \sim \mathrm{N}(XB,\ \Omega)\\
&\text{resp}_{3jk} \sim \mathrm{N}(XB,\ \Omega)\\
&\text{resp}_{4jk} \sim \mathrm{N}(XB,\ \Omega)\\
&\text{resp}_{1jk} = \beta_{0jk}\text{cons.es\_core\_st}_{ijk} + \text{-}0.019(0.038)\text{female.es\_core\_st}_{ijk}\\
&\beta_{0jk} = \text{-}0.009(0.049) + v_{0k} + u_{0jk}\\
&\text{resp}_{2jk} = \beta_{1jk}\text{cons.biol\_core\_st}_{ijk} + \text{-}0.081(0.036)\text{female.biol\_core\_st}_{ijk}\\
&\beta_{1jk} = 0.021(0.055) + v_{1k} + u_{1jk}\\
&\text{resp}_{3jk} = \beta_{2jk}\text{cons.phys\_core\_st}_{ijk} + \text{-}0.336(0.035)\text{female.phys\_core\_st}_{ijk}\\
&\beta_{2jk} = 0.149(0.056) + v_{2k} + u_{2jk}\\
&\text{resp}_{4jk} = \beta_{3jk}\text{cons.biol\_r\_st}_{ijk} + 0.023(0.055)\text{female.biol\_r\_st}_{ijk}\\
&\beta_{3jk} = \text{-}0.019(0.048) + v_{3k} + u_{3jk}
\end{aligned}
$$

$$
\begin{bmatrix} v_{0k} \\ v_{1k} \\ v_{2k} \\ v_{3k} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_v)\ :\ \Omega_v = \begin{bmatrix} 0.166(0.029) & & & \\ 0.133(0.027) & 0.228(0.037) & & \\ 0.116(0.027) & 0.212(0.035) & 0.245(0.039) & \\ 0.054(0.018) & 0.085(0.022) & 0.099(0.023) & 0.074(0.021) \end{bmatrix}
$$

$$
\begin{bmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0.840(0.025) & & & \\ 0.222(0.017) & 0.780(0.023) & & \\ 0.204(0.017) & 0.317(0.017) & 0.738(0.022) & \\ 0.110(0.026) & 0.114(0.025) & 0.088(0.024) & 0.926(0.039) \end{bmatrix}
$$

$$
\textit{-2*loglikelihood(IGLS Deviance)} = 22106.385(8539 \text{ of } 9756 \text{ cases in use})
$$

6. Multilevel Factor Analysis (Section 12.8)

Factor models can be fitted in MLwiN using Markov chain Monte Carlo estimation (see Browne, 2005; Chapter 19 for full details).

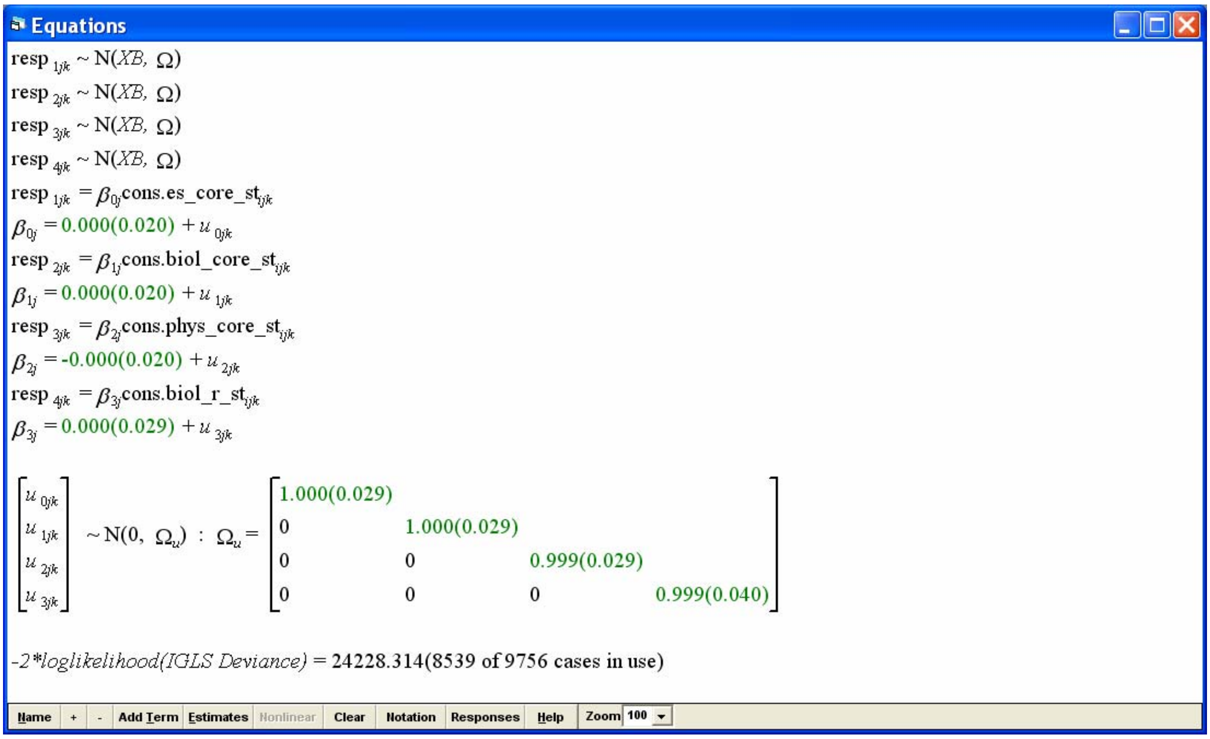
Single-level factor model (one factor)

To fit a factor model in MLwiN, we first need to fit the corresponding multivariate model with correlations set to zero. The estimates for this model are used as starting values in the

MCMC estimation. We will begin with a single-level factor model, which can be viewed as an extension of the simple multivariate model fitted at the start of this document (Section 12.7). We therefore need to remove the gender and school effects from the multivariate regression fitted above. We also need to constrain the pairwise correlations between the responses to zero.

- In the **Equations** window, click on **female.es_core_st** and click **Delete Term**
- Repeat for **female.biol_core_st**, **female.phys_core_st** and **female_biol_r_st**
- Now click β_{0jk} , uncheck **k(school_long)**, then click **Done**. This causes the school-level residual to be removed from the equation for the first response.
- Repeat for the other three responses
- Click on the covariance matrix Ω_u and select **set diagonal matrix**. The off-diagonal elements (the correlations) should be replaced by zeros.
- Click **Start**

You should get the following results (i.e. the first set of results in this document).

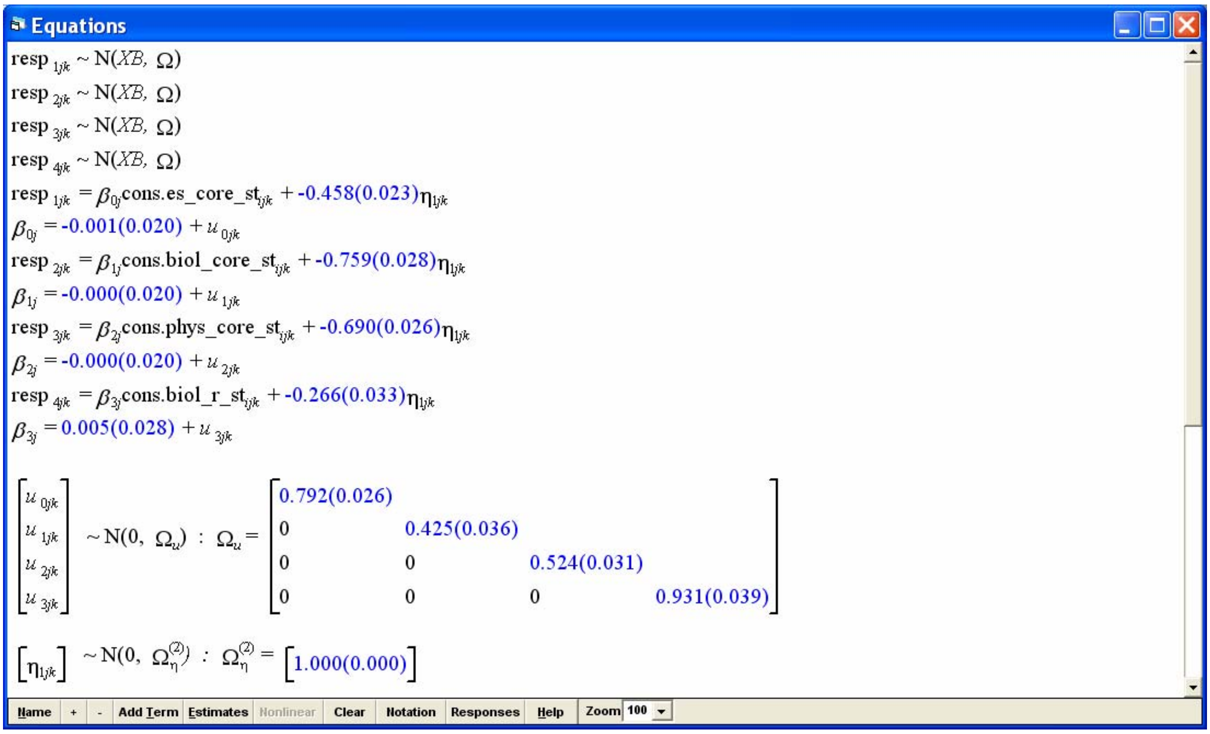


To extend to a factor model:

- Click on the **Estimation control** tab (on the toolbar) and then **MCMC**
- We will retain the default settings for **Burn-in length** and **Monitoring chain length**, so click **Done**
- From the **Model** menu, select **MCMC** then **Factor Analysis**
- Increase the **Number of factors** to **1** using the **+** button
- By default a factor is defined at level 2 (student, here) and the factor variance is constrained to 1

- Click **Set factors**. If you have the **Equations** window open you will see the factor, denoted by η_{1jk}, added to the equation for each response. Click **Done**
- Now go to the **Equations** window, and click **Start** to fit the model

The results are as follows (shown in Table 12.10).



Note that some of the estimates differ very slightly from those given in Table 12.10, and you may see small differences in your results. This is normal because MCMC methods are simulation based and there will be random fluctuations. Note also that the factor loadings are all negative; they were all switched to positive in Table 12.10 so that high values of the factor denote higher levels of science ability.

Multilevel factor model (one factor)

We next fit a multilevel factor model which includes a school-level factor and school-level residuals in the equations for the four test scores. To fit this model in MLwiN, we first need to fit the corresponding multilevel multivariate model to obtain starting values for MCMC.

- Click on the **Estimation control** tab (on the toolbar) and then **IGLS/RIGLS**
- Click **Done**
- In the **Equations** window, click on β_{0j} and check **k(school-long)**
- Repeat for β_{1j}, β_{2j} and β_{3j}
- Click on the school-level covariance matrix Ω_v and select **set diagonal matrix**. The off-diagonal elements (the correlations) should be replaced by zeros.
- Click **Start**

To extend to a factor model:

- Click on the **Estimation control** tab (on the toolbar) and then **MCMC**
- We will retain the default settings for **Burn-in length** and **Monitoring chain length**, so click **Done**
- From the **Model** menu, select **MCMC** then **Factor Analysis**
- Note that a level 2 factor η_{1jk} has already been added to the model
- Increase the **Number of factors** to **2** using the **+** button
- Change **Show factor** to **2** using the **+** button
- Change **Random level for factor** to **3**, then click **Set Factors**
- You should find that a second (school-level) factor, η_{2k}, has been added to the equation for each response.
- Click **Start** to fit the model

The results are as follows (shown in Table 12.11).

$$\mathrm{resp}_{1jk} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{resp}_{2jk} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{resp}_{3jk} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{resp}_{4jk} \sim \mathrm{N}(XB,\ \Omega)$$
$$\mathrm{resp}_{1jk} = \beta_{0jk}\mathrm{cons.es\_core\_st}_{ijk} + -0.379(0.024)\eta_{1jk} + -0.286(0.045)\eta_{2k}$$
$$\beta_{0jk} = -0.018(0.046) + v_{0k} + u_{0jk}$$
$$\mathrm{resp}_{2jk} = \beta_{1jk}\mathrm{cons.biol\_core\_st}_{ijk} + -0.596(0.028)\eta_{1jk} + -0.479(0.044)\eta_{2k}$$
$$\beta_{1jk} = -0.018(0.052) + v_{1k} + u_{1jk}$$
$$\mathrm{resp}_{3jk} = \beta_{2jk}\mathrm{cons.phys\_core\_st}_{ijk} + -0.543(0.026)\eta_{1jk} + -0.466(0.045)\eta_{2k}$$
$$\beta_{2jk} = -0.020(0.052) + v_{2k} + u_{2jk}$$
$$\mathrm{resp}_{4jk} = \beta_{3jk}\mathrm{cons.biol\_r\_st}_{ijk} + -0.193(0.035)\eta_{1jk} + -0.202(0.040)\eta_{2k}$$
$$\beta_{3jk} = -0.006(0.039) + v_{3k} + u_{3jk}$$

$$\begin{bmatrix} v_{0k} \\ v_{1k} \\ v_{2k} \\ v_{3k} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_v)\ :\ \Omega_v = \begin{bmatrix} 0.095(0.020) & & & \\ 0 & 0.014(0.011) & & \\ 0 & 0 & 0.038(0.015) & \\ 0 & 0 & 0 & 0.033(0.017) \end{bmatrix}$$

$$\begin{bmatrix} u_{0jk} \\ u_{1jk} \\ u_{2jk} \\ u_{3jk} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_u)\ :\ \Omega_u = \begin{bmatrix} 0.698(0.023) & & & \\ 0 & 0.428(0.030) & & \\ 0 & 0 & 0.474(0.027) & \\ 0 & 0 & 0 & 0.897(0.040) \end{bmatrix}$$

$$\begin{bmatrix} \eta_{2k} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_\eta^{(3)})\ :\ \Omega_\eta^{(3)} = \begin{bmatrix} 1.000(0.000) \end{bmatrix}$$

$$\begin{bmatrix} \eta_{1jk} \end{bmatrix} \sim \mathrm{N}(0,\ \Omega_\eta^{(2)})\ :\ \Omega_\eta^{(2)} = \begin{bmatrix} 1.000(0.000) \end{bmatrix}$$

$Deviance(MCMC)$ = 22637.719(8539 of 9756 cases in use)