



## Additional Topics Guide

Copyright ©2020 by Scientific Software International, Inc. All rights reserved.

Please download the examples from our website at <https://ssicentral.com/index.php/products/lisrel/lisrel-examples> and unzip them into a convenient folder location. The examples shown in the manual uses 'C:\LISREL Examples' and you are more than welcome to use the same or a different location. Please note that the actual location may be different on your machine.

## Table of Contents

<b>Additional Topics Guide</b> .....	<b>1</b>
<b>1 System Files</b> .....	<b>3</b>
1.1 The LISREL System data File.....	3
1.2 The Data System File.....	4
1.3 The Model System File.....	5
<b>2 Multiple Imputation</b> .....	<b>6</b>
2.1 Introduction .....	6
2.2 Technical Details .....	6
<b>3 Full Information Maximum Likelihood (FIML) for Continuous Variables</b> .....	<b>7</b>
<b>4 Multilevel Structural equations modeling</b> .....	<b>9</b>
4.1 Multilevel Structural Equations Models .....	9
4.2 A General Two-level Structural Equations Model.....	10
4.3 Maximum Likelihood for General Means and Covariance Structures.....	16
4.4 Fit Statistics and Hypothesis Testing.....	17
4.5 Starting Values and Convergence Issues .....	18
4.6 Practical Applications .....	22
<b>5 Multilevel non-linear regression</b> .....	<b>28</b>
5.1 Introduction to Multilevel Modeling .....	28
5.2 Multilevel Nonlinear Regression Models .....	29
5.3 Estimation Procedure for Multilevel Nonlinear Regression Models.....	31
5.4 Example .....	34
5.4.1 2-level Nonlinear Regression Model .....	34
5.4.2 Selecting the model.....	35
5.4.3 Creating the syntax .....	36
5.4.4 Selections of the output file.....	40
<b>6 References and additional reading</b> .....	<b>42</b>

# 1 System Files

PRELIS and LISREL generate several system files through which they can communicate with each other. These system files are binary files. Some of these system files can be used directly by users. Here we present these system files and their uses.

## 1.1 The LISREL System data File

The majority of software packages make use of a data file format that is unique to that package. These files are usually stored in a binary format so that reading and writing data to the hard drive is as fast as possible. Examples of file formats are

- Microsoft Excel (\*.xls)
- SPSS Data File (\*.sav)
- Minitab (\*.mtw)
- SAS Data File (\*.sas7bdat)
- STATA (\*.dta)
- STATISTICA (\*.sta)
- SYSTAT (\*.sys)

These data system files usually contain all the known information about a specific data set.

LISREL can import any of the above and many more file formats and convert them to a \*.lsf file. The \*.lsf file is analogous to, for example, a \*.sav file in the sense that one can retrieve 3 types of information from a LISREL system data file:

- **General information:**

This part of the \*.lsf file contains the number of cases in the data set, the number of variables, global missing value codes and the position (number) of the weight variable.

- **Variable information:**

The variable name (8 characters maximum); and variable type (for example, continuous or ordinal); and the variable missing value code. For an ordinal variable the following information is also stored: the number of categories, and category labels (presently 4 characters maximum) or numeric values assigned to each category.

- **Data information:**

Data values are stored in the form of a rectangular matrix, where the columns denote the variables and the rows the cases. Data are stored in double precision to ensure accuracy of all numerical calculations.

By opening a \*.Isf file, the main menu bar expands to include the **Data**, **Transformation**, **Statistics**, **Graphs** and **Multilevel** menus. Selecting any of the options from these menus will activate the corresponding dialog box. The **File** and **View** menus are also expanded. See the *Graphical User's Interface Guide* for more details.

An important feature of LISREL is that one can use \*.Isf files as part of the LISREL or SIMPLIS syntax. In doing so, one does not have to specify variable names, missing value codes, or the number of cases.

The folders **msemex**, **ls9ex**, **orfimlex**, **obsresex**, and **missingex** contain many examples to illustrate this new feature.

When building LISREL or SIMPLIS syntax from a path diagram, it is sufficient to select an appropriate \*.Isf file from the **Add/Read** option in the **Setup, Variables** dialog box. For more information and examples, see the *LISREL Examples Guide*.

## 1.2 The Data System File

A data system file, or a \*.dsf for short, is created each time PRELIS is run. Its name is the same as the PRELIS syntax file but with the suffix \*.dsf. The \*.dsf contains all the information about the variables that LISREL needs in order to analyze the data, *i.e.*, variable labels, sample size, means, standard deviations, covariance or correlation matrix, and the location of the asymptotic covariance matrix, if any. The \*.dsf file can be read by LISREL directly instead of specifying variable names, sample size, means, standard deviations, covariance or correlation matrix, and the asymptotic covariance matrix, if any, using separate commands in a LISREL or SIMPLIS syntax file.

To specify a \*.dsf in the SIMPLIS command language, write:

```
System file from File filename.DSF
```

This line replaces the following typical lines in a SIMPLIS syntax file (other variations are possible):

```
Observed Variables: A B C D E F
```

Means from File filename  
Covariance Matrix from File filename  
Asymptotic Covariance Matrix from File filename  
Sample Size: 678

To specify a \*.**dsf** in the LISREL command language, write:

```
SY=filename.DSF
```

This replaces the following typical lines in a LISREL syntax file (other variations are possible):

```
DA NI=k NO=n  
ME=filename  
CM=filename  
AC=filename
```

As the \*.**dsf** is a binary file, it can be read much faster than the syntax file. To make optimal use of this, consider the following strategy, assuming the data consists of many variables, possibly several hundreds, and a very large sample.

Use PRELIS to deal with all problems in the data, *i.e.*, missing data, variable transformation, recoding, definition of new variables, etc., and compute means and the covariance matrix or correlation matrix, say, and the asymptotic covariance matrix, if needed. Specify the \*.**dsf** file, select the variables for analysis and specify the model in a LISREL or SIMPLIS syntax file. Several different sets of variables may be analyzed in this way, each one being based on a small subset of the variables in the \*.**dsf**. The point is that there is no need to go back to PRELIS to compute new summary statistics of each LISREL model. With the SIMPLIS command language, selection of variables is automatic in the sense that only the variables included in the model will be used.

The use of the \*.**dsf** is especially important in simulations, as these will go much faster. The \*.**dsf** also facilitates the building of SIMPLIS or LISREL syntax by drawing a path diagram.

### 1.3 The Model System File

A model system file, or \*.**msf** for short, is created each time a LISREL syntax file containing a path diagram request is run. Its name is the same as the LISREL syntax file but with the suffix \*.**msf**. The \*.**msf** contains all the information about the model that LISREL needs to produce a path diagram, *i.e.*, type and form of each parameter, parameter estimates, standard errors, *t*-values, modification indices, fit statistics, etc. Usually, users do not have a direct need for the \*.**msf**.

## 2 Multiple Imputation

### 2.1 Introduction

Multivariate data sets, where missing values occur on more than one variable, are often encountered in practice. Listwise deletion may result in discarding a large proportion of the data, which in turn, tends to introduce bias.

Researchers frequently use *ad hoc* methods of imputation to obtain a complete data set. The multiple imputation procedure implemented in LISREL is described in detail in Schafer (1997) and uses the EM algorithm and the method of generating random draws from probability distributions via Markov chains.

In what follows, it is assumed that data are missing at random and that the observed data have an underlying multivariate normal distribution.

### 2.2 Technical Details

#### EM algorithm:

Suppose  $\mathbf{y} = (y_1, y_2, \dots, y_p)'$  is a vector of random variables with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$  and that  $y_1, y_2, \dots, y_n$  is a sample from  $\mathbf{y}$ .

#### Step 1: (M-Step)

Start with an estimate of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ , for example the sample means and covariances  $\bar{\mathbf{y}}$  and  $\mathbf{S}$  based on a subset of the data, which have no missing values. If each row of the data set contains a missing value, start with  $\boldsymbol{\mu} = \mathbf{0}$  and  $\boldsymbol{\Sigma} = \mathbf{I}$ .

### Step 2: (E-Step)

Calculate  $E(\mathbf{y}_{imiss} | \mathbf{y}_{iobs}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$  and  $Cov(\mathbf{y}_{imiss} | \mathbf{y}_{iobs}; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ ,  $i = 1, 2, \dots, N$ .

Use these values to obtain an update of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  (M-step) and repeat steps 1 and 2 until  $(\hat{\boldsymbol{\mu}}_{k+1}, \hat{\boldsymbol{\Sigma}}_{k+1})$  are essentially the same as  $(\hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k)$ .

### Markov chain Monte Carlo (MCMC):

In LISREL, the estimates of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  obtained from the EM-algorithm are used as initial parameters of the distributions used in Step 1 of the MCMC procedure

### Step 1: (P-Step)

Simulate an estimate  $\boldsymbol{\mu}_k$  of  $\boldsymbol{\mu}$  and an estimate  $\boldsymbol{\Sigma}_k$  of  $\boldsymbol{\Sigma}$  from a multivariate normal and an inverted Wishart distribution respectively.

### Step 2: (I-Step)

Simulate  $\mathbf{y}_{imiss} | \mathbf{y}_{iobs}$ ,  $i = 1, 2, \dots, N$  from conditional normal distributions with parameters based on  $\boldsymbol{\mu}_k$  and  $\boldsymbol{\Sigma}_k$ .

Replace the missing values with simulated values and calculate  $\boldsymbol{\mu}_{k+1} = \bar{\mathbf{y}}$  and  $\boldsymbol{\Sigma}_{k+1} = \mathbf{S}$  where  $\bar{\mathbf{y}}$  and  $\mathbf{S}$  are the sample means and covariances of the completed data set respectively. Repeat Steps 1 and 2  $m$  times. In LISREL, missing values in row  $i$  are replaced by the average of the simulated values over the  $m$  draws, after an initial burn-in period.

## 3 Full Information Maximum Likelihood (FIML) for Continuous Variables

Suppose that  $\mathbf{y} = (y_1, y_2, \dots, y_p)'$  has a multivariate normal distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$  and that  $y_1, y_2, \dots, y_n$  is a random sample of the vector  $\mathbf{y}$ .

Specific elements of the vectors  $\mathbf{y}_i, i = 1, 2, \dots, n$  may be unobserved so that the data set comprising of  $n$  rows (the different cases) and  $p$  columns (variables  $1, 2, \dots, p$ ) have missing values.

Let  $\mathbf{y}_k$  denote a vector with incomplete observations, then this vector can be replaced by  $\mathbf{y}_k^* = \mathbf{X}_k \mathbf{y}_k$  where  $\mathbf{X}_k$  is a selection matrix, and  $\mathbf{y}_k$  has typical elements  $(y_{k1}, y_{k2}, \dots, y_{kp})$  with one or more of the  $y_{kj}$ s missing,  $j = 1, 2, \dots, p$ .

**Example:**

Suppose  $p = 3$ , and that variable 2 is unobserved, then

$$\mathbf{y}_k^* = \begin{bmatrix} y_{k1} \\ y_{k3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{k1} \\ y_{k2} \\ y_{k3} \end{bmatrix}.$$

From the above example it can easily be seen that  $\mathbf{X}_k$  is based on an identity matrix with rows deleted according to missing elements of  $\mathbf{y}_k$ .

If an observed vector  $\mathbf{y}_k$  contains no unobserved values, then  $\mathbf{X}_k$  is equal to the identity matrix and hence  $\mathbf{y}_k^* = \mathbf{y}_k$ .

Without loss in generality,  $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$  can be replaced with  $(\mathbf{y}_1^*, \mathbf{y}_2^*, \dots, \mathbf{y}_n^*)$  where  $\mathbf{y}_k^*$  has a normal distribution with mean  $\mathbf{X}_k \boldsymbol{\mu}$  and covariance matrix  $\mathbf{X}_k \boldsymbol{\Sigma} \mathbf{X}_k'$ . The log-likelihood for the non-missing data is  $\sum_{k=1}^n \log f(\mathbf{y}_k^*, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ , where  $f(\mathbf{y}_k^*, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  is the pdf of  $\mathbf{X}_k \mathbf{y}_k$  given the parameters  $\boldsymbol{\mu}_k = \mathbf{X}_k \boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}_k = \mathbf{X}_k \boldsymbol{\Sigma} \mathbf{X}_k'$ .

In practice, when data are missing at random, there are usually  $M$  patterns of missingness, where  $M < n$ . When this is the case, the computational burden of evaluating  $n$  likelihood functions is considerably decreased.

It is customary to define the chi-square statistic as  $\chi^2 = F_0 - F_1$ , where  $F_0 = -2 \ln L_0$ ,  $F_1 = -2 \ln L_1$ , and where  $\ln L_1$  denotes the log-likelihood (at convergence) when no restrictions are imposed on the

parameters ( $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ ). The quantity  $\ln L_0$  denotes the log-likelihood value (at convergence) when parameters are restricted according to a postulated model. The degrees of freedom equals  $\nu$ , where  $\nu = p + p(p+1)/2 - k$  and  $k$  is the number of parameters in the model.

See the examples *Single group analysis with missing data using LISREL syntax* and *Multiple groups with missing data* in the *LISREL Examples Guide*. Additional examples are given in the **missingex** folder.

## 4 Multilevel Structural equations modeling

### 4.1 Multilevel Structural Equations Models

Social science research often entails the analysis of data with a hierarchical structure. A frequently cited example of multilevel data is a dataset containing measurements on children nested within schools, with schools nested within education departments.

The need for statistical models that take account of the sampling scheme is well recognized and it has been shown that the analysis of survey data under the assumption of a simple random sampling scheme may give rise to misleading results.

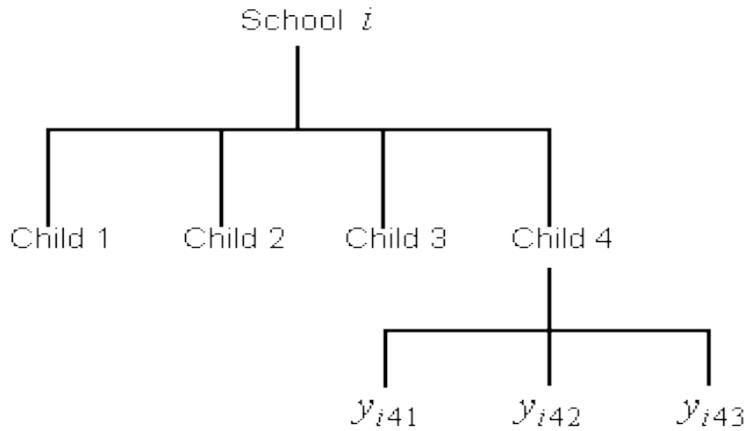
Iterative numerical procedures for the estimation of variance and covariance components for unbalanced designs were developed in the 1980s and were implemented in software packages such as MLWIN, SAS PROC MIXED and HLM. At the same time, interest in latent variables, that is, variables that cannot be directly observed or can only imperfectly be observed, led to the theory providing for the definition, fitting and testing of general models for linear structural relations for data from simple random samples.

A more general model for multilevel structural relations, accommodating latent variables and the possibility of missing data at any level of the hierarchy and providing the combination of developments in these two fields, was a logical next step. In papers by Goldstein and MacDonald (1988), MacDonald and Goldstein (1989) and McDonald (1993), such a model was proposed. Muthén (1990, 1991) proposed a partial maximum likelihood solution as simplification in the case of an unbalanced design. An overview of the latter can be found in Hox (1993).

General two-level structural equations modeling is available in LISREL. Full information maximum likelihood estimation is used, and a test for goodness of fit is given. An example, illustrating the implementation of the results for unbalanced designs with missing data at both levels of the hierarchy, is also given.

## 4.2 A General Two-level Structural Equations Model

Consider a data set consisting of 3 measurements, *math 1*, *math 2*, and *math 3*, made on each of 1000 children who are nested within  $N = 100$  schools. This data set can be schematically represented for school  $i$  as follows



For the  $i$ -th level-2 unit (school), we can write

$$\mathbf{y}_i = \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{y}_{i2} \\ \mathbf{y}_{i3} \\ \mathbf{y}_{i4} \end{pmatrix},$$

where for child 4 within school  $i$

$$\mathbf{y}'_{i4} = \begin{pmatrix} y_{i41} & y_{i42} & y_{i43} \\ \text{math 1} & \text{math 2} & \text{math 3} \end{pmatrix}.$$

A model which allows for between- and within-schools variation in math scores is the following simple variance component model

$$\begin{aligned}
\mathbf{y}_{i1} &= \mathbf{v}_i + \mathbf{u}_{i1} \\
\mathbf{y}_{i2} &= \mathbf{v}_i + \mathbf{u}_{i2} \\
\mathbf{y}_{i3} &= \mathbf{v}_i + \mathbf{u}_{i3} \\
\mathbf{y}_{i4} &= \mathbf{v}_i + \mathbf{u}_{i4},
\end{aligned}$$

or  $\mathbf{y}_{ij} = \mathbf{v}_i + \mathbf{u}_{ij}$ ,  $i=1,2,\dots,N$ , where it is assumed that  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$  are i.i.d.  $N(0, \mathbf{\Sigma}_B)$  and that  $\mathbf{u}_{i1}, \dots, \mathbf{u}_{iN}$  are i.i.d  $N(0, \mathbf{\Sigma}_W)$ . It is additionally assumed that

$$Cov(\mathbf{v}_i, \mathbf{u}_{ij}) = 0, i = 1, \dots, N; j = 1, 2, \dots, n_i.$$

From the distributional assumptions it follows that

$$Cov(\mathbf{y}_i, \mathbf{y}_i') = \begin{pmatrix} \mathbf{\Sigma}_B + \mathbf{\Sigma}_W & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B \\ \mathbf{\Sigma}_B & \mathbf{\Sigma}_B + \mathbf{\Sigma}_W & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B \\ \mathbf{\Sigma}_B & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B + \mathbf{\Sigma}_W & \mathbf{\Sigma}_B \\ \mathbf{\Sigma}_B & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B & \mathbf{\Sigma}_B + \mathbf{\Sigma}_W \end{pmatrix}$$

$$\text{or } Cov(\mathbf{y}_i, \mathbf{y}_i') = \mathbf{1}\mathbf{1}' \otimes \mathbf{\Sigma}_B + \mathbf{I} \otimes \mathbf{\Sigma}_W.$$

It also follows that

$$E(\mathbf{y}_i) = \mathbf{0}.$$

In practice, the latter assumption  $E(\mathbf{y}_i) = \mathbf{0}$  is seldom realistic, since measurements such as math scores do not have zero means. One approach to this problem is to use grand mean centering. Alternatively, one can add a fixed component to the model  $\mathbf{y}_{ij} = \mathbf{v}_i + \mathbf{u}_{ij}$ , so that

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{v}_i + \mathbf{u}_{ij}, \tag{1}$$

where  $\mathbf{X}_{ij}$  denotes a design matrix and  $\boldsymbol{\beta}$  a vector of regression coefficients.

Suppose that for the example above, the only measurements available for child 1 are *math 1* and *math 3* and for child 2 *math 2* and *math 3*.

Let  $\mathbf{S}_{i1}$  and  $\mathbf{S}_{i2}$  be selection matrices defined as follows

$$\mathbf{S}_{i1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ therefore } \mathbf{S}_{i1}\mathbf{v}_i = \begin{bmatrix} v_{i1} \\ v_{i3} \end{bmatrix}$$

and

$$\mathbf{S}_{i2} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ therefore } \mathbf{S}_{i2}\mathbf{v}_i = \begin{bmatrix} v_{i2} \\ v_{i3} \end{bmatrix}.$$

In general, if  $p$  measurements were made,  $\mathbf{S}_{ij}$  (see, for example, du Toit, 1995) consists of a subset of the rows of the  $p \times p$  identity matrix  $\mathbf{I}_p$ , where the rows of  $\mathbf{S}_{ij}$  correspond to the response measurements available for the  $(i, j)$ -th unit.

The above model can be generalized to accommodate incomplete data by the inclusion of these selection matrices. Hence

$$\mathbf{y}_{ij} = \mathbf{X}_{(y)ij}\boldsymbol{\beta} + \mathbf{S}_{ij}\mathbf{v}_i + \mathbf{S}_{ij}\mathbf{u}_{ij} \quad (2)$$

where  $\mathbf{X}_{(y)}$  is a design matrix of the appropriate dimensions.

If we further suppose that we have a  $q \times 1$  vector of variables  $\mathbf{x}_i$  characterizing the level-2 units (schools), then we can write the observed data for the  $i$ -th level-2 unit as

$$\mathbf{y}'_i = [\mathbf{y}'_{i1}, \mathbf{y}'_{i2}, \dots, \mathbf{y}'_{in_i}, \mathbf{x}'_i],$$

where

$$\mathbf{y}'_{ij} = [y_{ij1}, y_{ij2}, \dots, y_{ijp}]$$

and

$$\mathbf{x}'_i = [x_{i1}, x_{i2}, \dots, x_{iq}]. \quad (3)$$

We assume that  $\mathbf{y}_{ij}$  and  $\mathbf{x}_i$  can be written as

$$\mathbf{y}_{ij} = \mathbf{X}_{(y)ij}\boldsymbol{\beta}_y + \mathbf{S}_{ij}\mathbf{v}_i + \mathbf{S}_{ij}\mathbf{u}_{ij}, j = 1, 2, \dots, n_i \quad (4)$$

$$\mathbf{x}_i = \mathbf{X}_{(x)i}\boldsymbol{\beta}_x + \mathbf{R}_i\mathbf{w}_i, i = 1, 2, \dots, N \quad (5)$$

where  $\mathbf{X}_{(y)}$  and  $\mathbf{X}_{(x)}$  are design matrices for fixed effects, and  $\mathbf{S}_{ij}$  and  $\mathbf{R}_i$  are selection matrices for random effects of order  $p_j \times p$  and  $q_i \times q$  respectively. Note that (4) defines two types of random effects, where  $\mathbf{v}_i$  is common to level-3 units and  $\mathbf{u}_{ij}$  is common to level-1 units nested within a specific level-2 unit.

Additional distributional assumptions are

$$\begin{aligned} \text{Cov}(\mathbf{w}_i) &= \boldsymbol{\Sigma}_{xx}, \quad i = 1, 2, \dots, N \\ \text{Cov}(\mathbf{y}_{ij}, \mathbf{w}_i) &= \boldsymbol{\Sigma}_{xy}, \quad i = 1, 2, \dots, N; j = 1, 2, \dots, n_i \\ \text{Cov}(\mathbf{u}_{ij}, \mathbf{w}_i) &= \mathbf{0}. \end{aligned} \quad (6)$$

From (4) and (5), it follows that

$$\mathbf{y}_i = \begin{bmatrix} \mathbf{X}_{(y)i}\boldsymbol{\beta}_y + \mathbf{S}_i\mathbf{v}_i + \sum_{j=1}^{n_i} \mathbf{Z}_{ij}\mathbf{u}_{ij} \\ \mathbf{X}_{(x)i}\boldsymbol{\beta}_x + \mathbf{R}_i\mathbf{r}_i \end{bmatrix} \quad (7)$$

where

$$\mathbf{X}_{(y)i} = \begin{bmatrix} \mathbf{X}_{(y)i1} \\ \mathbf{M} \\ \mathbf{X}_{(y)in_i} \end{bmatrix}, \mathbf{S}_i = \begin{bmatrix} \mathbf{S}_{i1} \\ \mathbf{M} \\ \mathbf{S}_{in_i} \end{bmatrix}, \mathbf{R}_i = \begin{bmatrix} \mathbf{R}_{i1} \\ \mathbf{M} \\ \mathbf{R}_{in_i} \end{bmatrix},$$

and

$$\mathbf{Z}_{ij} = \begin{bmatrix} \mathbf{0} \\ \mathbf{M} \\ \mathbf{0} \\ \mathbf{S}_{ij} \\ \mathbf{0} \\ \mathbf{M} \\ \mathbf{0} \end{bmatrix}.$$

From the distributional assumptions given above, it follows that

$$\mathbf{y}_i : N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i),$$

where

$$\boldsymbol{\mu}_i = \begin{bmatrix} \mathbf{X}_{(y)i} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{(x)i} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_y \\ \boldsymbol{\beta}_x \end{bmatrix} = \mathbf{X}_i \boldsymbol{\beta}, \quad (8)$$

and

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \mathbf{V}_i & \mathbf{S}_i \boldsymbol{\Sigma}_{yx} \mathbf{R}_i' \\ \mathbf{R}_i \boldsymbol{\Sigma}_{xy} \mathbf{S}_i' & \mathbf{R}_i \boldsymbol{\Sigma}_{xx} \mathbf{R}_i' \end{bmatrix} \quad (9)$$

where

$$\mathbf{V}_i = \text{Cov} \begin{pmatrix} \mathbf{y}_{i1} \\ \mathbf{M} \\ \mathbf{y}_{in_i} \end{pmatrix} = \mathbf{S}_i \boldsymbol{\Sigma}_B \mathbf{S}_i' + \sum_{j=1}^{n_i} \mathbf{Z}_{ij} \boldsymbol{\Sigma}_W \mathbf{Z}_{ij}'.$$

#### Remark

If  $\mathbf{R}_i = \mathbf{I}_q$  and  $\mathbf{S}_{ij} = \mathbf{I}_p$ , corresponding to the case of no missing  $y$  or  $x$  variables, then  $\mathbf{S}_i \boldsymbol{\Sigma}_{yx} \mathbf{R}_i' = \mathbf{1} \otimes \boldsymbol{\Sigma}_{yx}$  where  $\mathbf{1}'$  is a  $n_i \times 1$  row vector  $(1, 1, \dots, 1)$ .

Furthermore, for  $\mathbf{S}_{ij} = \mathbf{I}_p$ ,  $j = 1, 2, \dots, n_i$

$$\mathbf{V}_i = \mathbf{I}_{n_i} \otimes \boldsymbol{\Sigma}_W + \mathbf{1}\mathbf{1}' \otimes \boldsymbol{\Sigma}_B$$

(see, for example, MacDonald and Goldstein, 1989). The unknown parameters in (8) and (9) are  $\boldsymbol{\beta}$ ,  $\text{vecs}\boldsymbol{\Sigma}_B$ ,  $\text{vecs}\boldsymbol{\Sigma}_W$ ,  $\text{vecs}\boldsymbol{\Sigma}_{xy}$  and  $\text{vecs}\boldsymbol{\Sigma}_{xx}$ .

Structural models for the type of data described above may be defined by restricting the elements of  $\boldsymbol{\beta}$ ,  $\boldsymbol{\Sigma}_B$ ,  $\boldsymbol{\Sigma}_W$ ,  $\boldsymbol{\Sigma}_{xy}$ , and  $\boldsymbol{\Sigma}_{xx}$  to be some basic set of parameters  $\boldsymbol{\gamma}' = (\gamma_1, \gamma_2, \dots, \gamma_k)$ .

For example, assume the following pattern for the matrices  $\boldsymbol{\Sigma}_W$  and  $\boldsymbol{\Sigma}_B$ , where  $\boldsymbol{\Sigma}_W$  refers to the within (level-1) covariance matrix and  $\boldsymbol{\Sigma}_B$  to the between (level-2) covariance matrix:

$$\begin{aligned} \boldsymbol{\Sigma}_W &= \boldsymbol{\Lambda}_W \boldsymbol{\Psi}_W \boldsymbol{\Lambda}_W' + \mathbf{D}_W \\ \boldsymbol{\Sigma}_B &= \boldsymbol{\Lambda}_B \boldsymbol{\Psi}_B \boldsymbol{\Lambda}_B' + \mathbf{D}_B. \end{aligned} \tag{10}$$

Factor analysis models typically have the covariance structures defined by (10).

Consider a confirmatory factor analysis model with 2 factors and assume  $p = 6$ .

$$\boldsymbol{\Lambda}_W = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}, \quad \boldsymbol{\Psi}_W = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix},$$

and

$$D_W = \begin{bmatrix} \theta_{11} & & \\ & \mathbf{O} & \\ & & \theta_{11} \end{bmatrix}.$$

If we restrict all the parameters across the level-1 and level-2 units to be equal, then

$$\boldsymbol{\gamma}' = [\lambda_{11}, \lambda_{21}, \dots, \lambda_{44}, \psi_{11}, \psi_{22}, \theta_{11}, \dots, \theta_{66}]$$

is the vector of unknown parameters.

### 4.3 Maximum Likelihood for General Means and Covariance Structures

In this section, we give a general framework for normal maximum likelihood estimation of the unknown parameters. In practice, the number of variables ( $p + q$ ) and the number of level-1 units within a specific level-2 unit may be quite large, which leads to  $\Sigma_i$  matrices of very high order. It is therefore apparent that further simplification of the likelihood function derivatives and Hessian is required if the goal is to implement the theoretical results in a computer program. These aspects are addressed in du Toit and du Toit (2008).

Denote the expected value and covariance matrix of  $\mathbf{y}_i$  by  $\boldsymbol{\mu}_i$  and  $\Sigma_i$  respectively (see (8) and (9)). The log-likelihood function of  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N$  may then be expressed as

$$\ln L = -\frac{1}{2} \sum_{i=1}^N \{n_i \ln 2\pi + \ln |\Sigma_i| + \text{tr} \Sigma_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'\} \quad (11)$$

Instead of maximizing  $\ln L$ , maximum likelihood estimates of the unknown parameters are obtained by minimizing  $-\ln L$  with the constant term omitted, *i.e.*, by minimizing the following function

$$F(\boldsymbol{\gamma}) = \frac{1}{2} \sum_{i=1}^N \{\ln |\Sigma_i| + \text{tr} \Sigma_i^{-1} \mathbf{G}_{y_i}\}, \quad (12)$$

where

$$\mathbf{G}_{y_i} = (\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'. \quad (13)$$

Its minimum  $\frac{\partial F(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \mathbf{0}$  yields the normal maximum likelihood estimator  $\hat{\boldsymbol{\gamma}}$  of the unknown vector of parameters  $\boldsymbol{\gamma}$ .

Unless the model yields maximum likelihood estimators in closed form, it will be necessary to make use of an iterative procedure to minimize the discrepancy function. The optimization procedure (Browne and du Toit, 1992) is based on the so-called Fisher scoring algorithm, which in the case of structured means and covariances may be regarded as a sequence of Gauss-Newton steps with quantities to be fitted as

well as the weight matrix changing at each step. Fisher scoring algorithms require the gradient vector and an approximation to the Hessian matrix.

#### 4.4 Fit Statistics and Hypothesis Testing

The multilevel structural equations model,  $M(\boldsymbol{\gamma})$ , and its assumptions imply a covariance structure  $\boldsymbol{\Sigma}_B(\boldsymbol{\gamma})$ ,  $\boldsymbol{\Sigma}_W(\boldsymbol{\gamma})$ ,  $\boldsymbol{\Sigma}_{xy}(\boldsymbol{\gamma})$ ,  $\boldsymbol{\Sigma}_{xx}(\boldsymbol{\gamma})$  and mean structure  $\boldsymbol{\mu}(\boldsymbol{\gamma})$  for the observable random variables where  $\boldsymbol{\gamma}$  is a  $k \times 1$  vector of parameters in the statistical model. It is assumed that the empirical data are a random sample of  $N$  level-2 units and  $\sum_{i=1}^N n_i$  level-1 units, where  $n_i$  denotes the number of level-1 units within the  $i$ -th level-2 unit. From this data, we can compute estimates of  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}_B$ , ...,  $\boldsymbol{\Sigma}_{xx}$  if no restrictions are imposed on their elements. The number of parameters for the unrestricted model is

$$k^* = m + 2 \left[ \frac{1}{2} p(p+1) \right] + pq + \frac{1}{2} q(q+1)$$

and is summarized in the  $k^* \times 1$  vector  $\boldsymbol{\pi}$ . The unrestricted model  $M(\boldsymbol{\pi})$  can be regarded as the "baseline" model.

To test the model  $M(\boldsymbol{\gamma})$ , we use the likelihood ratio test statistic

$$c = -2 \ln L(\hat{\boldsymbol{\gamma}}) - 2 \ln L(\hat{\boldsymbol{\pi}}) \quad (14)$$

If the model  $M(\boldsymbol{\gamma})$  holds,  $c$  has a  $\chi^2$ -distribution with  $d = k^* - k$  degrees of freedom. If the model does not hold,  $c$  has a non-central  $\chi^2$ -distribution with  $d$  degrees of freedom and non-centrality parameter  $\lambda$  that may be estimated as (see Browne and Cudeck, 1993):

$$\hat{\lambda} = \max\{(c - d), 0\} \quad (15)$$

These authors also show how to set up a confidence interval for  $\lambda$ .

It is possible that the researcher has specified a number of competing models  $M_1(\boldsymbol{\gamma}_1), M_2(\boldsymbol{\gamma}_2), \dots, M_k(\boldsymbol{\gamma}_k)$ . If the models are nested in the sense that  $\boldsymbol{\gamma}_j : k_j \times 1$  is a subset of  $\boldsymbol{\gamma}_i : k_i \times 1$ , then one may use the likelihood ratio test with degrees of freedom  $k_i - k_j$  to test  $M(\boldsymbol{\gamma}_j)$  against  $M(\boldsymbol{\gamma}_i)$ .

Another approach is to compare models on the basis of some criteria that take parsimony as well as fit into account. This approach can be used regardless of whether or not the models can be ordered in a nested sequence. Two strongly related criteria are the AIC measure of Akaike (1987) and the CAIC of Bozdogan (1987).

$$AIC = c + 2d \quad (16)$$

$$CAIC = c + (1 + \ln \sum_{i=1}^N n_i) d \quad (17)$$

The use of  $c$  as a central  $\chi^2$ -statistic is based on the assumption that the model holds exactly in the population. A consequence of this assumption is that models that hold approximately in the population will be rejected in large samples.

Steiger (1990) proposed the root mean square error of approximation (RMSEA) statistic that takes particular account of the error of approximation in the population

$$RMSEA = \sqrt{\frac{\hat{F}_0}{d}}, \quad (18)$$

where  $\hat{F}_0$  is a function of the sample size, degrees of freedom and the fit function. To use the RMSEA as a fit measure in multilevel SEM, we propose

$$\hat{F}_0 = \max \left\{ \left[ \frac{c-d}{N} \right], 0 \right\} \quad (19)$$

Browne and Cudeck (1993) suggest that an RMSEA value of 0.05 indicates a close fit and that values of up to 0.08 represent reasonable errors of approximation in the population.

## 4.5 Starting Values and Convergence Issues

In fitting a structural equations model to a hierarchical data set, one may encounter convergence problems unless good starting values are provided. A procedure that appears to work well in practice is to start the estimation procedure by fitting the unrestricted model to the data. The first step is therefore to obtain estimates of the fixed components ( $\beta$ ) and the variance components ( $\Sigma_B, \Sigma_{xy}, \Sigma_{xx}$  and  $\Sigma_W$ ). Our experience with the Gauss-Newton algorithm (see, for example, Browne and du Toit, 1992) is that

convergence is usually obtained within less than 15 iterations, using initial estimates  $\beta = \mathbf{0}$ ,  $\Sigma_B = \mathbf{I}_p$ ,  $\Sigma_{xy} = \mathbf{0}$ ,  $\Sigma_{xx} = \mathbf{I}_q$  and  $\Sigma_W = \mathbf{I}_p$ . At convergence, the value of  $-2 \ln L$  is computed.

Next, we treat

$$\mathbf{S}_B = \begin{bmatrix} \hat{\Sigma}_B & \hat{\Sigma}_{yx} \\ \hat{\Sigma}_{xy} & \hat{\Sigma}_{xx} \end{bmatrix} \text{ and } \mathbf{S}_W = \begin{bmatrix} \hat{\Sigma}_W & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

as sample covariance matrices and fit a two group structural equations model to the between- and within-groups. Parameter estimates obtained in this manner are used as the elements of the initial parameter vector  $\gamma_0$ .

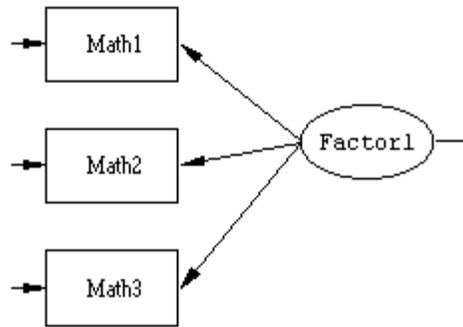
In the third step, the iterative procedure is restarted and  $\gamma_k$  updated from  $\gamma_{k-1}$ ,  $k = 1, 2, \dots$  until convergence is reached.

The following example illustrates the steps outlined above. The data set used in this section forms part of the data library of the Multilevel Project at the University of London and comes from the Junior School Project (Mortimore *et al*, 1988). Mathematics and language tests were administered in three consecutive years to more than 1000 students from 49 primary schools, which were randomly selected from primary schools maintained by the Inner London Education Authority.

The following variables were selected from the data file:

- *School*            School code (1 to 49)
- *Math1*            Score on mathematics test in year 1 (score 1 - 40)
- *Math2*            Score on mathematics test in year 2 (score 1 - 40)
- *Math3*            Score on mathematics test in year 3 (score 1 - 40)

The school number (*School*) is used as the level-2 identification.



**Figure 1: Confirmatory factor analysis model**

A simple confirmatory factor analysis model (see Figure 1) is fitted to the data:

$$\Sigma_B = \lambda \Psi \lambda' + \mathbf{D}_B,$$

$$\Sigma_W = \lambda \Psi \lambda' + \mathbf{D}_W,$$

where

$$\lambda' = (1, \lambda_{21}, \lambda_{31})$$

and  $\mathbf{D}_B$  and  $\mathbf{D}_W$  are diagonal matrices with diagonal elements equal to the unique (error) variances of *Math1*, *Math2* and *Math3*. The variance of the factor is denoted by  $\Psi$ . Note that we assume equal factor loadings and factor variances across the between- and within-groups, leading to a model with 3 degrees of freedom. The SIMPLIS (see Jöreskog and Sörbom, 1993) syntax file to fit the factor analysis model is shown below. Note that the between- and within-groups covariance matrices are the estimated  $\Sigma_B$  and  $\Sigma_W$  obtained in the first step by fitting the unrestricted model.

```

Group 1: Between Schools JSP data (Level 2)
Observed Variables: Math1 Math2 Math3
Covariance matrix
3.38885
2.29824  5.19791
2.31881  3.00273  4.69663
Sample Size=24 ! Taken as (n1+n2+...nN)/N rounded to
! nearest integer
Latent Variables: Factor1
Relationships
Math1=1*Factor1

```

Math2-Math3=Factor1  
 Group 2: Within Schools JSP data (Level 1)  
 Covariance matrix  
 47.04658  
 38.56798 55.37006  
 30.81049 36.04099 40.71862  
 Sample Size=1192 ! Total number of pupils  
 ! Set the Variance of Factor1 Free ! Remove comment to  
 !free parameter  
 Set the Error Variance of Math1 Free  
 Set the Error Variance of Math2 Free  
 Set the Error Variance of Math3 Free  
 Path Diagram  
 LISREL OUTPUT ND=3  
 End of Problem

**Table 1: Parameter estimates and standard errors for factor analysis model**

	SIMPLIS		Multilevel SEM	
	Estimate	Standard error	Estimate	Standard error
<b>Factor loadings</b>				
$\lambda_{11}$	1.000	-	1.000	-
$\lambda_{21}$	1.173	0.031	1.177	0.032
$\lambda_{31}$	0.939	0.026	0.947	0.028
<b>Factor variance</b>				
$\Psi$	32.109	1.821	31.235	1.808
<b>Error variances</b> (between)				
Math1	1.640	0.787	1.656	0.741
Math2	2.123	1.059	2.035	0.942
Math3	1.868	0.779	1.840	0.734
<b>Error variances</b> (within)				
Math1	14.114	0.810	14.209	0.890
Math2	10.274	0.884	10.256	0.993
Math3	11.910	0.699	11.837	0.806
<b>Chi-square</b>	36.233		46.56	
<b>Degrees of freedom</b>	3		3	

Table 1 shows the parameter estimates, estimated standard errors and  $\chi^2$ -statistic values obtained from the SIMPLIS output and from the multilevel SEM output respectively.

## Remarks:

1. The between-groups sample size of 26 used in the SIMPLIS syntax file was computed as  $\frac{1}{N} \sum_{i=1}^N n_i$ , where  $N$  is the number of schools and  $n_i$  the number of children within school  $i$ . Since this value is only used to obtain starting values, it is not really crucial how the between-group sample size is computed. See, for example, Muthén (1990,1991) for an alternative formula.
2. The within-group sample size of 1192 used in the SIMPLIS file syntax is equal to the total number of school children.
3. The number of missing values per variable is as follows:

Math1: 38

Math2: 63

Math3: 239

The large percentage missing for the Math3 variable may partially explain the relatively large difference in  $\chi^2$ -values from the SIMPLIS and multilevel SEM outputs.

4. If one allows for the factor variance parameter to be free over groups, the  $\chi^2$  fit statistic becomes 1.087 at 2 degrees of freedom. The total number of multilevel SEM iterations required to obtain convergence equals eight.

In conclusion, a small number of variables and a single factor SEM model were used to illustrate the starting values procedure that we adopted. The next section contains additional examples, also based on a schools data set. See also du Toit and du Toit (forthcoming) and the `\msemex` folder for additional examples.

## 4.6 Practical Applications

The example discussed in this section is based on school data that were collected during a 1994 survey in South Africa.

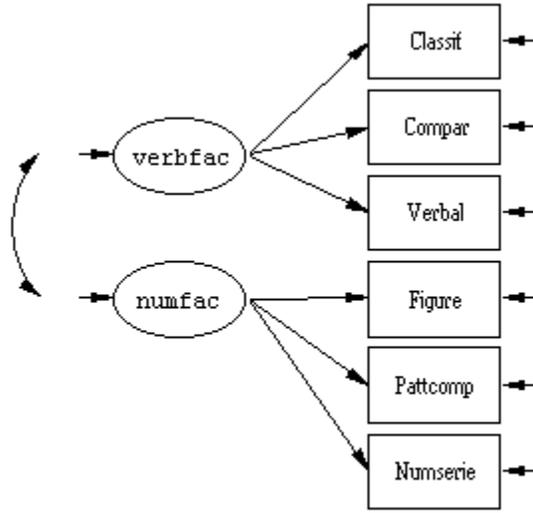
A brief description of the **SA\_Schools94.dat** data set is as follows:  $N = 136$  schools were selected and the total number of children within schools  $\sum_{i=1}^N n_i = 6047$ , where  $n_i$  varies from 20 to 60. The data set contains 20 variables as shown in Table 2.

**Table 2: Description of variables in SA\_School94.dat**

Variable	Name	Description	Number missing
1	Student	Level-1 identification	0
2	School	Level-2 identification	0
3	Constant	All values equal to 1	0
4	Grade	0 = Grade 2 1 = Grade 3 2 = Grade 4	0
5	Language	0 = White 1 = Black	0
6	Gender	1 = Male 2 = Female	1
7	Mothedu	Mother's level of education on a scale from 1 to 7	783
8	Fathede	Father's level of education on a scale from 1 to 7	851
9	Read	Teacher's evaluation on a scale from 1 to 5: 1 = Poor 5 = Excellent	482
10	Speech		470
11	Write		467
12	Arithm		451
13	Socio	Socio-economic status indicator, scale 0 to 5 on school level	0
14	Classif	Classification: total correct out of 30 items	23
15	Compar	Comparison: total correct out of 23 items	27
16	Verbal	Verbal Instructions: total correct out of 50 items	20
17	Figure	Figure Series: total correct out of 24 items	118
18	Pattcomp	Pattern Completion: total correct out of 24 items	109
19	Knowled	Knowledge: total correct out of 32 items	112
20	Numserie	Number Series: total correct out of 15 items	2305

The variables Language and Socio are school-level variables and their values do not vary within schools. Listwise deletion of missing cases results in a data set containing only 2691 of the original 6047 cases.

For this example, we use the variables Classif, Compar, Verbal, Figure, Pattcomp and Numserie from the schools data set discussed in the previous section. Two common factors are hypothesized: verbal and numerical ability. The first three variables are assumed to measure Verbfac and the last three to measure Numfac. A path diagram of the hypothesized factor model is shown in Figure 2.



**Figure 2: Confirmatory factor analysis model for 6 variables**

The between- and within-school structural equations models are

$$\begin{aligned}\Sigma_W &= \Lambda_W \Psi_W \Lambda_W' + \mathbf{D}_W \\ \Sigma_B &= \Lambda_B \Psi_B \Lambda_B' + \mathbf{D}_B.\end{aligned}\tag{20}$$

where

$$\Lambda_W = \Lambda_B = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix},$$

and where factor loadings are assumed to be equal on the between (schools) and within (children) levels. The  $2 \times 2$  matrices  $\Psi_B$  and  $\Psi_W$  denote unconstrained factor covariance matrices. Diagonal elements of  $\mathbf{D}_B$  and  $\mathbf{D}_W$  are the unique (error) variances.

Gender and Grade differences were accounted for in the means part of the model,

$$E(y_{ijk}) = \beta_{k0} + \beta_{k0} \text{Gender} + \beta_{k2} \text{Grade},$$

where the subscripts  $i, j$  and  $k$  denote schools, students and variables  $k$ , respectively.

From the description of the school data set, we note that the variable Numserie has 2505 missing values. An inspection of the data set reveals that the pattern of missingness can hardly be described as missing at random. To establish how well the proposed algorithm perform in terms of the handling of missing cases, we have decided to retain this variable in this example. The appropriate LISREL syntax file for this example is given below.

```

Group1: Between Schools HSRC School Project
DA NI=6 NO=0 NG=2 MA=CM MI=-9.0
LA
Classif Compar Verbal Figure Pattcomp Numserie
RA = SA_Schools94.dat
$CLuster= School
SE
1 2 3 4 5 6 /
MO NY=6 NE=2 LY=FU,FI PS=SY,FR TE=DI,FR
LE
verbfac numfac
FR LY(2,1) LY(3,1) LY(5,2) LY(6,2)
VA 1.00 LY(1,1) LY(4,2)
PD
OU ME=ML
Group2: Within Schools HSRC School Project
LA
Classif Compar Verbal Figure Pattcomp Numserie
DA NI=6 NO=0 NG=2 MA=CM MI=-9.0
RA = SA_Schools94.dat
SE
1 2 3 4 5 6 /
MO NY=6 NE=2 LY=IN PS=IN TE=IN
LK
verbfac numfac
FR PS(1,1) PS(2,1) PS(2,2) TE(1,1) TE(2,2)
FR TE(3,3) TE(4,4) TE(5,5) TE(6,6)
! FR LY(2,1) LY(3,1) LY(5,2) LY(6,2)
OU

```

Table 3 shows the estimated between-schools covariance matrix  $\hat{\Sigma}_B$  when no restrictions are imposed on its elements, and the fitted covariance matrix  $\Sigma_B(\hat{\gamma})$  where  $\gamma$  is the vector of parameters of the CFA models given in (20).

**Table 3: Estimated between-schools covariance matrix,  $\Sigma_B$**

(i)  $\hat{\Sigma}_B$  unrestricted

	Classif	Compar	Verbal	Figure	Pattcomp	Numserie
Classif	1.29					
Compar	1.27	2.66				
Verbal	2.83	3.54	10.42			
Figure	2.06	2.70	6.89	5.53		
Pattcomp	2.17	2.60	6.58	5.09	5.34	
Numserie	1.46	1.85	4.93	3.85	3.81	3.16

(ii)  $\Sigma_B(\hat{\gamma})$  for the CFA model

	Classif	Compar	Verbal	Figure	Pattcomp	Numserie
Classif	1.61					
Compar	1.74	3.86				
Verbal	2.21	3.22	6.76			
Figure	2.52	3.67	4.66	5.82		
Pattcomp	2.29	3.33	4.22	5.03	4.93	
Numserie	1.79	2.60	3.30	3.93	3.56	3.11

Likewise, Table 4 shows  $\hat{\Sigma}_W$  for the unrestricted model and  $\Sigma_W(\hat{\gamma})$  for the CFA model.

**Table 4: Estimated within-schools covariance matrix,  $\Sigma_W$**

(i)  $\hat{\Sigma}_W$  unrestricted

	Classif	Compar	Verbal	Figure	Pattcomp	Numserie
Classif	8.49					
Compar	4.59	18.77				
Verbal	5.52	7.64	17.26			
Figure	4.45	7.21	8.49	16.27		

Pattcomp	4.30	7.21	8.45	9.55	16.19	
Numserie	2.69	4.05	5.28	7.31	5.80	7.31

(ii)  $\hat{\Sigma}_W(\gamma)$  for the CFA model

	Classif	Compar	Verbal	Figure	Pattcomp	Numserie
Classif	7.81					
Compar	3.31	17.12				
Verbal	4.19	6.11	15.07			
Figure	3.72	5.42	6.87	14.58		
Pattcomp	3.37	4.91	6.23	8.13	14.61	
Numserie	2.64	3.84	4.87	6.36	5.76	7.21

The goodness of fit statistics for the CFA model are shown in Table 5.

**Table 5: Goodness-of-fit statistics**

(6047 students, 136 schools)

$\chi^2 = 159.87$	degrees of freedom = 20
RMSEA=0.061	

Parameter estimates and estimated standard errors are given in Table 6.

It is typical of SEM models to produce large  $\chi^2$ -values when sample sizes are large, as in the present case. The RMSEA may be a more meaningful measure of goodness of fit and the value of 0.061 indicates that the assumption of equal factor loadings between and within schools is reasonable.

**Table 6: Parameter estimates and standard errors**

	Estimate	Standard error
<b>Factor loadings</b>		
$\lambda_{11}$	1.0	
$\lambda_{21}$	1.456	0.048
$\lambda_{31}$	1.846	0.054
$\lambda_{42}$	1.0	
$\lambda_{52}$	0.906	0.017
$\lambda_{62}$	0.708	0.014
<b>Factor covariances</b>		

(between schools)		
$\Psi_{11}$	1.196	0.185
$\Psi_{21}$	2.524	0.342
$\Psi_{22}$	5.546	0.729

<b>Error variances</b> (between schools)		
Classif	0.413	0.081
Compar	1.327	0.223
Verbal	2.673	0.388
Figure	0.279	0.090
Pattcomp	0.377	0.092
Numserie	0.325	0.069
<b>Factor covariances</b> (within schools)		
$\Psi_{11}$	2.271	0.114
$\Psi_{21}$	3.722	0.128
$\Psi_{22}$	8.976	0.272
<b>Error variances</b> (within schools)		
Classif	5.538	0.119
Compar	12.305	0.262
Verbal	7.328	0.222
Figure	5.606	0.169
Pattcomp	7.239	0.178
Numserie	2.710	0.098

## 5 Multilevel non-linear regression

### 5.1 Introduction to Multilevel Modeling

The analysis of data with a hierarchical structure has been described in the literature under various names. It is known as hierarchical modeling, random coefficient modeling, latent curve modeling, growth curve modeling or multilevel modeling. The basic underlying structure of measurements nested within units at a higher level of the hierarchy is, however, common to all. In a repeated measurements growth model, for example, the measurements or outcomes are nested within the experimental units (second level units) of the hierarchy.

Ignoring the hierarchical structure of data can have serious implications, as the use of alternatives such as aggregation and disaggregation of information to another level can induce high collinearity among predictors and large or biased standard errors for the estimates. Standard fixed parameter regression models do not allow for the exploration of variation between groups, which may be of interest in its own right. For a discussion of the effects of these alternatives, see Bryk and Raudenbush (1992), Longford (1987) and Rasbash (1993).

Multilevel or hierarchical modeling provides the opportunity to study variation at different levels of the hierarchy. Such a model can also include separate regression coefficients at different levels of the hierarchy that have no meaning without recognition of the hierarchical structure of the population. The dependence of repeated measurements belonging to one experimental unit in a typical growth curve analysis, for example, is taken into account with this approach. In addition, the data to be analyzed need not be balanced in nature. This has the advantage that estimates can also be units for which a very limited amount of information is available.

## 5.2 Multilevel Nonlinear Regression Models

It was pointed out by Pinheiro and Bates (2000) that one would want to use nonlinear latent coefficient models for reasons of *interpretability*, *parsimony*, and more importantly, *validity* beyond the observed range of the data.

By increasing the order of a polynomial model, one can get increasingly accurate approximations to the true, usually nonlinear, regression function, within the range of the observed data. High order polynomial models often result in multicollinearity problems and provide no theoretical considerations about the underlying mechanism producing the data.

There are many possible nonlinear regression models to select from. Examples are given by Gallant (1987) and Pinheiro and Bates (2000). The Richards function (Richards, 1959) is a generalization of a family of nonlinear functions and is used to describe growth curves (Koops, 1988). Three special cases of the Richards function are the logistic, Gompertz and Monomolecular functions, respectively. Presently, one can select curves of the form

$$y = f_1(x) + f_2(x) + e$$

where the first component,  $f_1(x)$  may have the form:

- logistic:  $\frac{b_1}{(1 + s \exp(b_2 - b_3 x))}$
- Gompertz:  $b_1 \exp(-b_2 \exp(-b_3 x))$

- Monomolecular:  $b_1(1 + s \exp(b_2 - b_3x))$
- power:  $b_1x^{b_2}$
- exponential:  $b_1 \exp(-b_2x)$

The second component,  $f_2(x)$ , may have the form:

- logistic:  $\frac{c_1}{(1 + s \exp(c_2 - c_3x))}$
- Gompertz:  $c_1 \exp(-c_2 \exp(-c_3x))$
- Monomolecular:  $c_1(1 + s \exp(c_2 - c_3x))$
- power:  $c_1x^{c_2}$
- exponential:  $c_1 \exp(-c_2x)$

In the curves above  $s$  denotes the sign of the term  $\exp(b_2 - b_3x)$  and is equal to 1 or -1.

Since the parameters in the first three functions above have definite physical meanings, a curve from this family is preferred to a polynomial curve, which may often be fitted to a set of responses with the same degree of accuracy. The parameter  $b_1$  represents the time asymptotic value of the characteristic that has been measured, the parameter  $b_2$  represents the potential increase (or decrease) in the value of the function during the course of time  $t_1$  to  $t_p$ , and the parameter  $b_3$  characterizes the rate of growth.

The coefficients  $b_1, b_2, \dots, c_3$  are assumed to be random, and, as in linear hierarchical models, can be written as level-2 outcome variables where

$$\begin{aligned}
 b_1 &= \beta_1 + u_1 \\
 b_2 &= \beta_2 + u_2 \\
 b_3 &= \beta_3 + u_3 \\
 c_1 &= \beta_4 + u_4 \\
 c_2 &= \beta_5 + u_5 \\
 c_3 &= \beta_6 + u_6.
 \end{aligned}$$

It is assumed that the level-2 residuals  $u_1, u_2, \dots, u_6$  have a normal distribution with zero means and covariance matrix  $\Phi$ .

In LISREL, it may further be assumed that the values of any of the random coefficients are affected by some level-2 covariate so that, in general,

$$\begin{aligned} b_1 &= \beta_1 + \gamma_1 z_1 + u_1 \\ b_2 &= \beta_2 + \gamma_2 z_2 + u_2 \\ b_3 &= \beta_3 + \gamma_3 z_3 + u_3 \\ c_1 &= \beta_4 + \gamma_4 z_4 + u_4 \\ c_2 &= \beta_5 + \gamma_5 z_5 + u_5 \\ c_3 &= \beta_6 + \gamma_6 z_6 + u_6. \end{aligned}$$

where  $z_i$  denotes the value of a covariate and  $\gamma_i$  the corresponding coefficient.

It is usually sufficient to select a single component ( $f_1(x)$ ) to describe a large number of monotone increasing (or decreasing) growth patterns.

To describe more complex patterns, use can be made of two-component regression models. LISREL allows the user to select any of the 5 curve types as component 1 and to combine it with any one of the 5 curve types for the second component.

Valid choices are, for example,

- Monomolecular + Gompertz
- logistic
- exponential + logistic
- logistic + logistic

The unknown model parameters are the vector of fixed coefficients ( $\beta$ ), the vector of covariate coefficients ( $\gamma$ ), the covariance matrix ( $\Phi$ ) of the level-2 residuals and the variance ( $\sigma^2$ ) of the level-1 measurement errors.

### 5.3 Estimation Procedure for Multilevel Nonlinear Regression Models

In linear multilevel models,  $y$  has a normal distribution, since  $y$  is a linear combination of the random coefficients. For example, the intercept-and-slopes-as-outcomes model is

$$y = b_1 + b_2 x + e$$

where  $b_1 = u_1$ ,  $b_2 = u_2$  and  $(u_1, u_2)$  is assumed to be normally distributed.

A multilevel nonlinear model is a regression model which cannot be expressed as a linear combination of its coefficients and therefore  $y$  is no longer normally distributed. The probability density function of  $y$  can be evaluated as the multiple integral

$$f(y) = \int_{b_1} \dots \int_{c_3} f(y, b_1, \dots, c_3) db_1 \dots dc_3$$

that, in general, cannot be solved in closed form.

To evaluate the likelihood function

$$L = \prod_{i=1}^n f(y_i),$$

one has to use a numerical integration technique. We assume that  $e$  has a  $N(0, \sigma^2)$  distribution and that  $(b_1, b_2, b_3, c_1, c_2, c_3)$  has a  $N(\mathbf{0}, \mathbf{\Phi}^2)$  distribution.

In the multilevel procedure, use is made of a Gauss quadrature procedure to evaluate the integrals numerically. The ML method requires good starting values for the unknown parameters. The estimation procedure is described in detail by Du Toit and Cudeck, R. (2009). See also Cudeck and du Toit (2003).

### Starting values

Once a model is selected to describe the nonlinear pattern in the data, for example as revealed by a plot of  $y$  on  $x$ , a curve is fitted to each individual using ordinary nonlinear least squares.

In step 1 of the fitting procedure, these OLS parameter estimates are written to a file and estimates of  $\beta$  and  $\Phi$  are obtained by using the sample means and covariances of the set of fitted parameters. Since observed values from some individual cases may not be adequately described by the selected model, these cases can have excessively large residuals, and it may not be advisable to include them in the calculation of the  $\beta$  and  $\Phi$  estimates.

In step 2 of the model fitting procedure, use is made of the MAP (Maximum Aposterior) estimator of the unknown parameters.

Suppose that a single component  $f_1(b, x)$  is fitted to the data. Since

$$f(b|y) = f(b)f(y|b)/f(y),$$

where  $f(b|y)$  is the conditional probability density function of the random coefficients  $b$  given the observations, it follows that

$$\ln f(b|y) = \ln f(b) + \ln f(y|b) + k,$$

where  $k = -\ln f(y)$ .

The MAP procedure can briefly be described as follows:

**Step a:**

Given starting values of  $\boldsymbol{\beta}$ ,  $\boldsymbol{\Phi}$  and  $\sigma^2$ , obtain estimates of the random coefficients  $\hat{b}_i$  from

$$\frac{\partial}{\partial b_i} \ln f(b_i | y_i) = 0, \quad i = 1, 2, \dots, n.$$

**Step b:**

Use the estimates  $\hat{b}_1, \hat{b}_2, \dots, \hat{b}_n$  and  $Cov(\hat{b}_1), \dots, Cov(\hat{b}_n)$  to obtain new estimates of  $\boldsymbol{\beta}$ ,  $\boldsymbol{\Phi}$  and  $\sigma^2$  (see Herbst, A. (1993) for a detailed discussion).

Repeat steps a and b until convergence is attained.

For further information, see Bock and du Toit (2004).

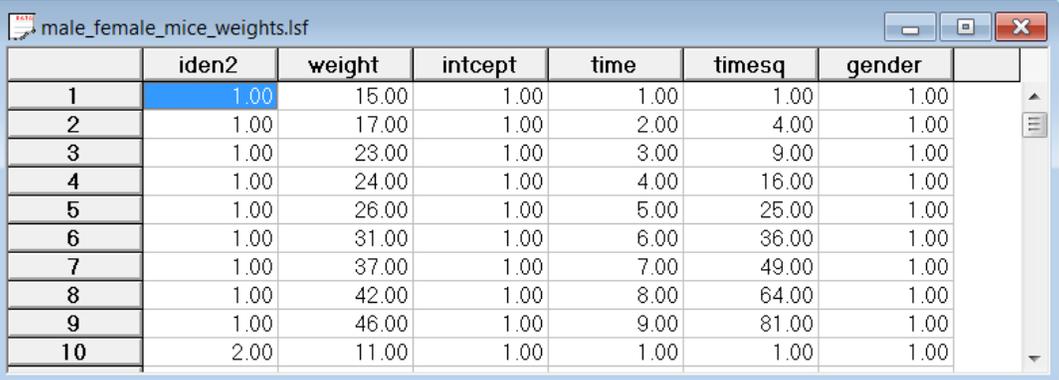
For many practical purposes, results obtained from the MAP procedure may be sufficient. However, if covariates are included in the model, parameter estimates are only available via the ML option, which uses the MAP estimates of  $\boldsymbol{\beta}$ ,  $\boldsymbol{\Phi}$  and  $\sigma^2$  as starting values.

## 5.4 Example

### 5.4.1 2-level Nonlinear Regression Model

The data set used contain repeated measurements on 82 striped mice and was obtained from the Department of Zoology at the University of Pretoria, South Africa (see du Toit, 1979). A number of male and female mice were released in an outdoor enclosure with nest boxes and sufficient food and water. They were allowed to multiply freely. Occurrence of birth was recorded daily and newborn mice were weighed weekly, from the end of the second week after birth until physical maturity was reached. The data set consists of the weights of 42 male and 40 female mice. For male mice, 9 repeated weight measurements are available and for the female mice 8 repeated measurements.

The first 10 observations from this data set, contained in `\nonlinex\mouse.lsf`, and the variable names to be used are shown below:



	iden2	weight	intcept	time	timesq	gender
1	1.00	15.00	1.00	1.00	1.00	1.00
2	1.00	17.00	1.00	2.00	4.00	1.00
3	1.00	23.00	1.00	3.00	9.00	1.00
4	1.00	24.00	1.00	4.00	16.00	1.00
5	1.00	26.00	1.00	5.00	25.00	1.00
6	1.00	31.00	1.00	6.00	36.00	1.00
7	1.00	37.00	1.00	7.00	49.00	1.00
8	1.00	42.00	1.00	8.00	64.00	1.00
9	1.00	46.00	1.00	9.00	81.00	1.00
10	2.00	11.00	1.00	1.00	1.00	1.00

A logistic regression model is fitted to the data:

$$y_{ij} = b_{1i} / (1 + \exp(b_{2i} - b_{3i}x_{ij})) + e_{ij},$$

where

$$b_{1i} = \beta_1 + \gamma_1 \text{gender} + u_{1i}$$

$$b_{2i} = \beta_2 + \gamma_2 \text{gender} + u_{2i}$$

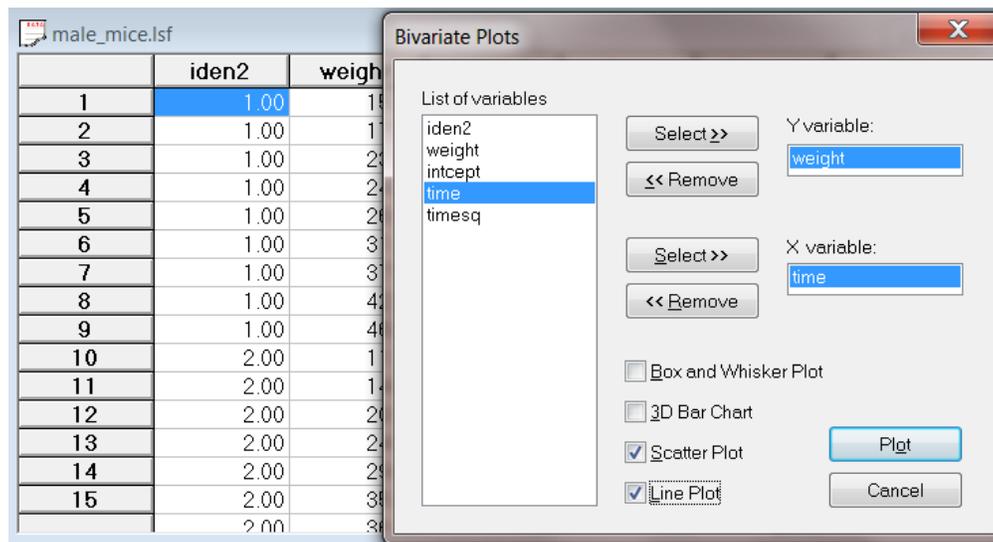
$$b_{3i} = \beta_3 + \gamma_3 \text{gender} + u_{3i}.$$

## 5.4.2 Selecting the model

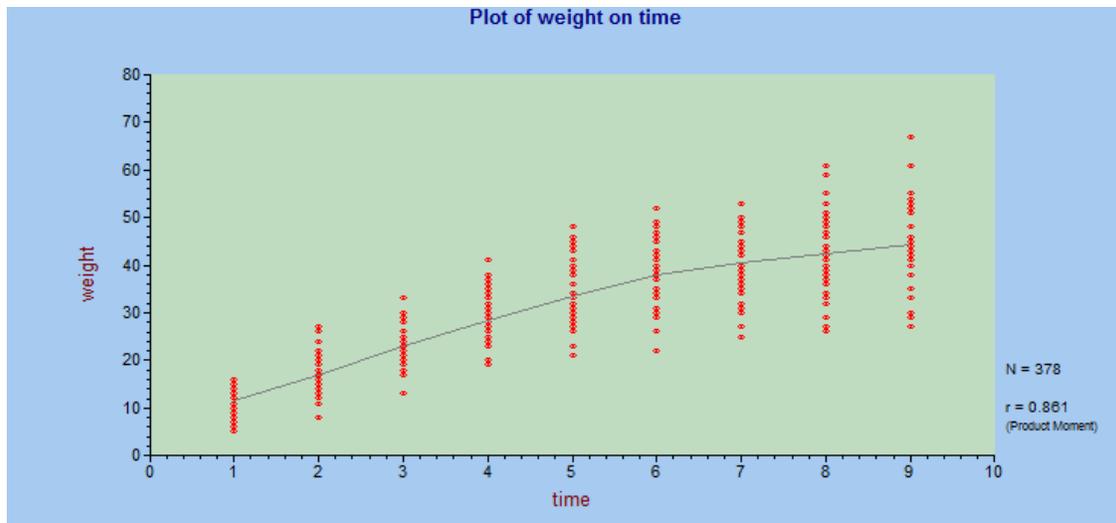
Use the **File, Open** option to locate **male\_female\_mice.lsf** in the **nonlinex** folder. A portion of the spreadsheet is shown above. The variable **iden2** (level 2 ID) is the number of a specific mouse.

A preliminary inspection of the data reveals that the weight measurements increase *monotonically* over time, but reaches an asymptote value at maturity. It is apparent that a polynomial will describe the data quite accurately within the range of 1 to 9 weeks, but predictions will be unreliable outside the observed range of the data as is illustrated on page 275 of Pinheiro and Bates (2000).

To obtain some idea of the pattern of weight measurements over time, we used the **Data, Select Cases** option and selected only those cases where **gender = 1**. Using the **Output Options** dialog box, the male data set is saved as **malemice.lsf**. From the **Graphs** menu, select **Bivariate Plot** and enter **weight** and **time** as the **Y** and **X** variables respectively.

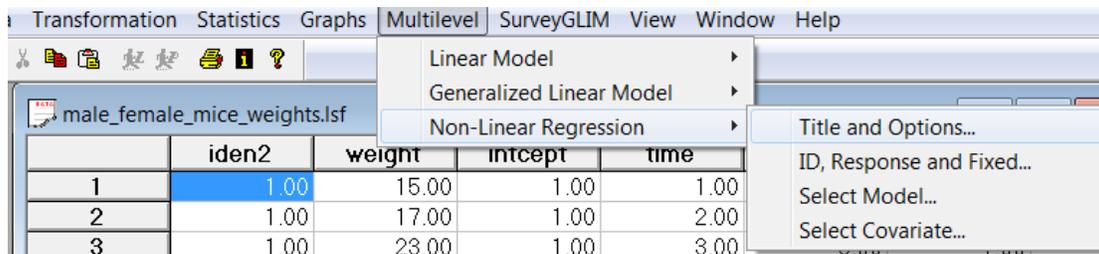


The figure below displays the gain in weight pattern over time. This curve follows a pattern more like a logistic than a polynomial model.



### 5.4.3 Creating the syntax

From the main menu bar, select the **Multilevel, Non-Linear Model** option. To fit a non-linear model to repeated measurements data, click on the **Title and Options** option.



The **Titles and Options** dialog box enables the researcher to assign a title, and to specify the maximum number of iterations, the convergence criterion, the missing data code and the number of quadrature points. The quadrature points determine the accuracy of the numerical integration procedure. Since numerical integration turns out to be computationally intensive, one would generally choose a large number of points for 2 to 3 coefficient functions (30 – 80) but much fewer points (6 – 12) for non-linear curves with more than 3 coefficients, e.g. the logistic + logistic (6 coefficients) function.

Finally, we select the **Use Full Maximum Likelihood** option.

Title and Options

Title (Maximum 70 characters):  
Weights of 42 male and 40 female mice. Use Gender as covariate

Maximum Number of Iterations: 30

Convergence Criterion: 0.001

Missing Data Value: -999999

Number of Quadrature Points: 15

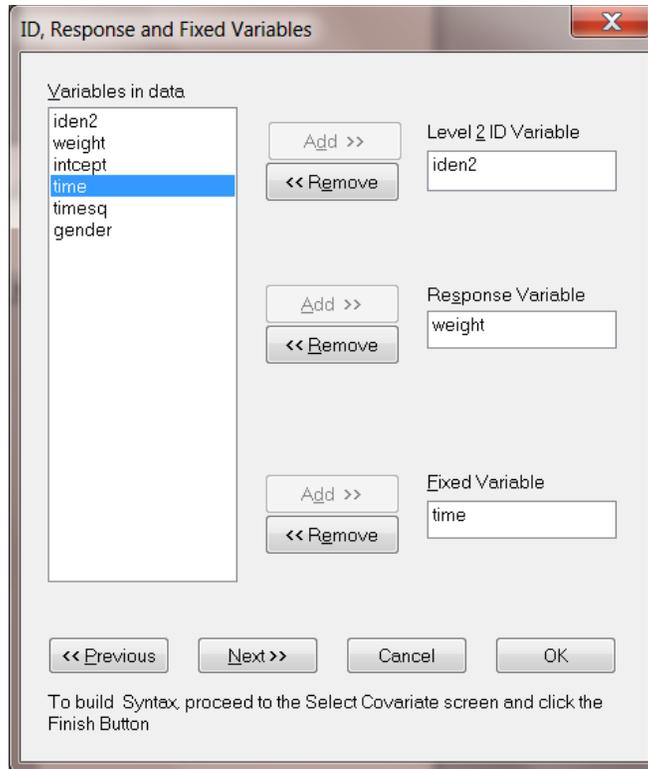
Use Maximum A priori

Use Full Maximum Likelihood

Next >> Cancel OK

To build Syntax, proceed to the Select Covariate screen and click the Finish Button

The **ID, Response and Fixed Variables** dialog box is used to assign a level-2 identification variable and to select the response and fixed variables as shown below.



The next dialog box enables the researcher to specify an appropriate model for the data, see the **Select Model** dialog box below.

It is realistic to assume that gender differences will result in different asymptote, potential growth and growth rate parameters. We therefore select gender as a level-2 covariate influencing all 3 coefficients of the logistic curve

$$\frac{b_1}{1 + \exp(b_2 - b_3 t)}$$

that is

$$b_i = \beta_i + \gamma_i \text{ gender} + u_i, i = 1, 2, 3.$$

From the **Select Model** dialog box, select the single component logistic function and click **Next** to go to the **Covariates** dialog box.

## Select Model

Select Model

Function

First Component

Logistic

Gompertz

Monomolecular

Power

Exponential

Second Component

Logistic

Gompertz

Monomolecular

Power

Exponential

Current Model

Logistic

<< Previous   Next >>   Cancel   OK

To build Syntax, proceed to the Select Covariate screen and click the Finish Button

## Select covariates

Select Covariate for First Component (Logistic)

Variables in data

iden2

weight

intcept

time

timesq

gender

Add >>

<< Remove

Coefficient 1 (b1)

gender

Add >>

<< Remove

Coefficient 2 (b2)

gender

Add >>

<< Remove

Coefficient 3 (b3)

gender

<< Previous   Finish   Cancel   OK

To build Syntax, proceed to the Select Covariate screen and click the Finish Button

To generate the required syntax, click **Finish** in this dialog box. The syntax generated is shown below (see also **mouse1.prl** in the **nonlinex** folder)

```

male_female_mice_weights.PRL
OPTIONS METHOD = ML CONVERGE = 0.001000 MAXITER = 30 QUADPTS = 15;
TITLE = Weights of 42 male and 40 female mice. Use Gender as covariate;
SY='C:\LISREL9 Examples\NONLINEX\male_female_mice_weights.lsf';
ID2 = iden2;
RESPONSE = weight;
FIXED = time;
MODEL = Logistic;
COVARIATES b1 = gender
            b2 = gender
            b3 = gender;

```

### 5.4.4 Selections of the output file

Portions of the output are shown below.

#### (i) Data Summary

```

male_female_mice_weights.OUT
+-----+
| DATA SUMMARY |
+-----+

NUMBER OF LEVEL 2 UNITS :      82
NUMBER OF LEVEL 1 UNITS :     698

N2 :    1    2    3    4    5    6    7    8
N1 :    9    9    9    9    9    9    9    9

N2 :    9   10   11   12   13   14   15   16
N1 :    9    9    9    9    9    9    9    9

```

There are 82 level 2 units (42 male and 40 female mice). For each of the male mice there are 9 repeated measurements and for each of the female mice (numbers 43 – 82) there are 8 repeated measurements.

**(ii) ML estimates, fixed part of the model**

Maximum Likelihood Solution

Coefficients	Beta	Std.Err.	Z-value	P >  z
b1	45.57666	0.86649	52.59929	0.00000
b2	1.59382	0.03202	49.76858	0.00000
b3	0.52430	0.01310	40.02862	0.00000

Covariate Names	Gamma	Std.Err.	Z-value	P >  z
gender	2.44737	0.86647	2.82454	0.00473
gender	0.09677	0.03202	3.02230	0.00251
gender	0.03747	0.01310	2.86089	0.00422

**(iii) Maximum Likelihood estimates, random part**

Variance estimate

Level 1	Std.Err.	Z-value	P >  z	
Sigma**2	2.73565	0.10453	26.17006	0.00000

Covariances

Level 2	Std.Err.	Z-value	P >  z	
u1,u1	102.84576	13.12483	7.83597	0.00000
u2,u1	2.12285	0.38836	5.46622	0.00000
u2,u2	0.13572	0.01854	7.32032	0.00000
u3,u1	-0.62441	0.15767	-3.96030	0.00007
u3,u2	0.00464	0.00539	0.86016	0.38970
u3,u3	0.01961	0.00298	6.57006	0.00000

Note: ML estimates of individual coefficients written to file THETA1.EST

Additional examples, based on an assortment of models, are given in the **nonlinex** folder

## 6 References and additional reading

- Agresti, A. (1990). *Categorical Data Analysis*. Wiley: New York.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, **52**, 317-332.
- Bock, R.D. & du Toit, S.H.C. (2004). Parameter estimation in the context of non-linear longitudinal growth models. In R.C. Hauspie, N. Cameron & L. Molinari (Eds.), *Methods of Human growth Research*, pp. 234-257. Cambridge, UK: Cambridge University Press.
- Bollen, K.A. (1995). Structural equation models that are nonlinear in latent variables: A least squares estimator. In: P.M. Marsden (Ed.) *Sociological Methodology*. Blackwell: Cambridge, MA.
- Bollen, K.A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, **61**, 109-121.
- Bozdogan, H. (1987). Model selection and Akaike's information criteria (AIC). *Psychometrika*, **52**, 345-370.
- Browne, M.W. and Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen and J.S. Long (Eds.) *Testing Structural Equation Models*. Sage: Newbury Park, CA.
- Browne, M. W., & Du Toit, S. H. C. (1991). *Models for learning data*. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 47-68). Washington, DC: American Psychological Association.
- Browne, M.W. and du Toit, S.H.C. (1992). Automated fitting of nonstandard models for mean vectors and covariance matrices. *Multivariate Behavioral Research*, **27**(2), 269-300.
- Bryk, A.S. and Raudenbush, S.W. (1992). *Hierarchical Linear Models*. Sage: Newbury Park, CA.
- Cudeck, R. & du Toit, S. H. C. (2003). *Nonlinear multilevel models for repeated measures data*. In Duan, N. & Reise, S. P. (Eds.), *Multilevel modeling: Methodological advances, issues and applications* (pp. 1-24). Mahwah, NJ: Erlbaum
- Cudeck, R., Harring, J.R., & du Toit, S.H.C. (2009). Marginal Maximum Likelihood Estimation of a Latent Variable Model with Interaction. *Journal of Educational and Behavioral Statistics*, March 2009, Vol 34, No1: pages 131-134.

- Du Toit, S.H.C. (1979). *The analysis of growth curves*. Unpublished PhD dissertation, University of South Africa: Pretoria.
- Du Toit, M. (1995). *The Analysis of Hierarchical and Unbalanced Complex Survey Data using Multilevel Models*. Unpublished Ph.D. dissertation, University of Pretoria: Pretoria.
- Du Toit, S.H.C. & Cudeck, R. (2009). Estimation of the nonlinear random coefficient model when some random effects are separable. *Psychometrika*, March 2009, Vol 74, No 1: pages 65-82.
- Du Toit, S.H.C. & du Toit, M. (2008). Multilevel Structural Equation Modeling. In *The analysis of multilevel models*. DeLeeuw and Meijer (Eds.), Springer-Verlag.
- Gallant, A.R. (1987). *Nonlinear Statistical Models*. Wiley: New York.
- Goldstein, H. and McDonald, R.P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, **53**, 455-467.
- Harring, J.R., Cudeck, R., & du Toit, S.H.C. (2006). Fitting partially nonlinear random coefficient models as SEMs. *Multivariate Behavioral Research*, Vol 41, No 4: pages 579-596.
- Herbst, A. (1993). *The Statistical Modelling of Growth*. Unpublished Ph.D. dissertation, University of Pretoria: Pretoria.
- Heymann, C. (1981). *XAID—an Extended Automatic Interaction Detector*. Internal Report SWISK 28, Council for Scientific and Industrial Research: Pretoria.
- Hox, J.J. (1993). Factor analysis of multilevel data: Gauging the Muthén model. In: Oud, J.H.L. and van Blokland-Vogeleesang, R.A.W. (Eds.). *Advances in longitudinal and multivariate analysis in the behavioral sciences*. Nijmegen, NL, 141-156.
- Jöreskog, K.G. (1979). Basic ideas of factor and component analysis. In: Jöreskog, K.G. and Sörbom, D. (Eds.): *Advances in factor analysis and structural equation models*. Abt Books: Cambridge, MA., 5-20.
- Jöreskog, K.G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, **24**, 387-404.
- Jöreskog, K.G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, **59**, 381-389.
- Jöreskog, K.G., and Sörbom, D. (1989). *LISREL 7-A guide to the program and applications*. SPSS Publications: Chicago.

- Jöreskog, K.G. and Sörbom, D. (1993a). *New features in PRELIS 2*. Scientific Software International: Chicago.
- Jöreskog, K.G. and Sörbom, D. (1993b). *New features in LISREL 8*. Scientific Software International: Chicago.
- Jöreskog, K.G. and Sörbom, D. (1996a). *PRELIS 2: User's Reference Guide*. Scientific Software International: Chicago.
- Jöreskog, K.G. and Sörbom, D. (1996b). *LISREL 8: User's Reference Guide*. Scientific Software International: Chicago.
- Jöreskog, K.G. and Sörbom, D. (1996c). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International: Chicago.
- Jöreskog, K.G., and Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In: Marcoulides, G.A. and Schumacker, R.E. (Eds.) *Advanced structural equation modeling: Issues and techniques*. Lawrence Erlbaum Associates, Publishers: Mahwah, NJ.
- Kanfer, R. and Ackerman, P.L. (1989). Motivation and cognitive abilities: an integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, Monograph*, **74**, 657-690.
- Klein, L.R. (1950). *Economic fluctuations in the United States 1921-1941*. Cowles Commission Monograph No. 11.
- Koops, W.J. (1988). Multiphasic Analysis of Growth Curves in Chickens. *Poultry Science*, **67**, 33-42.
- Kreft, I. and de Leeuw, J. (1988). *Introducing multilevel modeling*. Sage: Newbury Park, CA.
- Longford, N.T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. *Biometrika*, **74**(4), 817-827.
- McDonald, R.P. (1993). A general model for two-level data with responses missing at random. *Psychometrika*, **58**(4), 575-585.
- McDonald, R.P. and Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *Br. J. Math. Stat. Psych.*, **42**, 215-232.
- McLachlan, G.J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley: New York.

- Muthén, B. (1990). *Means and Covariance Structure Analysis of Hierarchical Data*. UCLA Statistics series, no 62: Los Angeles.
- Muthén, B. (1991). Multilevel Factor Analysis of Class and Student Achievement Components. *Journal of Educational Measurement*, **28**, 338-354.
- Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-Plus*. Springer-Verlag: New York.
- Richards, F.J. (1959). A flexible growth function for empirical use. *Journal of Experimental Botany*, **10**, 29(290-300).
- Schafer (1997). *Analysis of Incomplete Multivariate Data*. Monographs on Statistics and Applied Probability 72, Chapman and Hall/CRC.
- Sörbom, D. (1981). Structural equation models with structured means. In: Jöreskog, K.G. and Wold, H. (Eds.) *Systems under indirect observation: Causality, structure and prediction*. North-Holland Publishing Co: Amsterdam.
- Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, **25**, 173-180.