

# Multilevel Structural Equation Modeling

## Multilevel Structural equation 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.

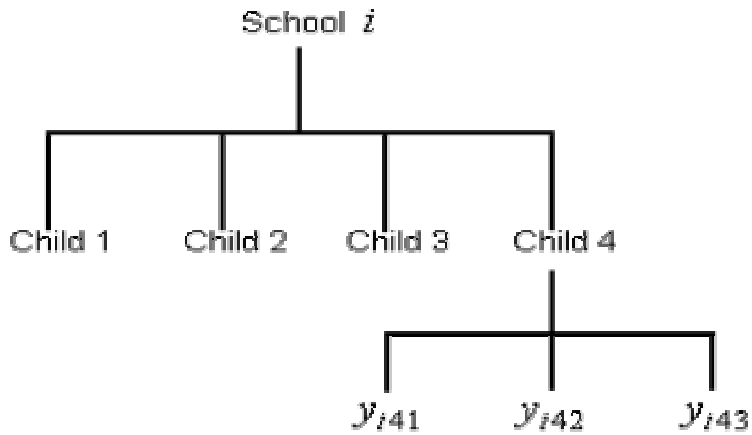
Iterative numerical procedures for the estimation of variance and covariance components for unbalanced designs were developed in the 1980s and were implemented in software packages such as MLWIN, SAS PROC MIXED and HLM. At the same time, interest in latent variables, that is, variables that cannot be 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 equation 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.

## A General Two-level Structural equation 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

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

From the distributional assumptions it follows that

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

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

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

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

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

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  $p_{ij} \times p$  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

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

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

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

where

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

and

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

From the distributional assumptions given above, it follows that

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

where

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

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

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

**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:

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

Factor analysis models typically have the covariance structures defined by (10). Consider a confirmatory factor analysis model with 2 factors and assume  $p = 6$ .

$$\Lambda_W = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix}, \Psi_W = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix},$$

and

$$D_W = \begin{bmatrix} \theta_{11} & & \\ & \ddots & \\ & & \theta_{66} \end{bmatrix}.$$

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

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

is the vector of unknown parameters.

## 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 (forthcoming).

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

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

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

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

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.

## Fit Statistics and Hypothesis Testing

The multilevel structural equation 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.

## Starting Values and Convergence Issues

In fitting a structural equation 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 equation 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.

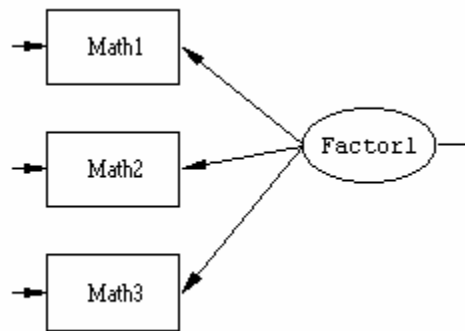
The following example illustrates the steps outlined above. The data set used in this section forms part of the data library of the Multilevel Project at the University of London, and comes from the Junior School Project (Mortimore *et al*, 1988). Mathematics

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 7.1: Confirmatory factor analysis model**

A simple confirmatory factor analysis model (see Figure 7.1) 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. These estimates may also be obtained by deleting the variable names *eng1*, *eng2*, and *eng3* in the RESPONSE command of the syntax file **jsp1.pr2** in the **mlevelx** folder.

```
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 7.1: 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 7.1 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. Another example can be found in du Toit and du Toit (forthcoming). Also see the **msemex** folder for additional examples.

## 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\_Schools.psf** data set in the **msemex** folder is as follows:

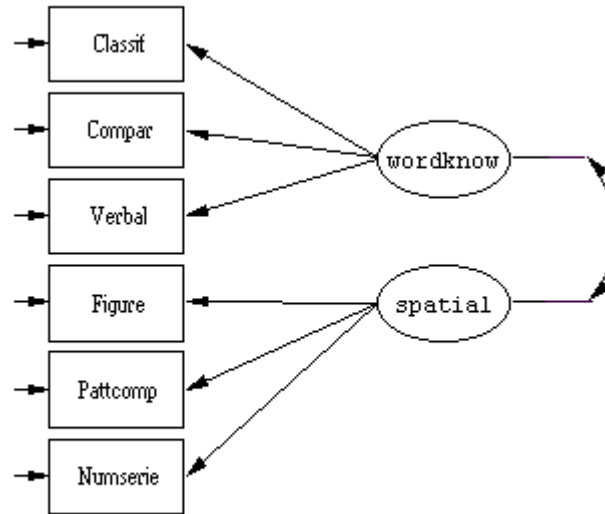
$N = 136$  schools were selected and the total number of children within schools  $\sum_{i=1}^N n_i = 6047$ , where  $n_i$  varies from 20 to 60. A description of the variables is given in Table 7.2.

**Table 7.2: Description of variables in SA\_School.psf**

Name	Description	Number missing
Pupil	Level-1 identification	0
School	Level-2 identification	0
Intcept	All values equal to 1	0
Grade	0 = Grade 2 1 = Grade 3 2 = Grade 4	0
Language	0 = Black 1 = White	0
Gender	1 = Male 2 = Female	1
Mothedu	Mother's level of education on a scale from 1 to 7	783
Fathedu	Father's level of education on a scale from 1 to 7	851
Read	Teacher's evaluation on a scale from 1 to 5: 1 = Poor 5 = Excellent	482
Speech		470
Write		467
Arithm		451
Socio	Socio-economic status indicator, scale 0 to 5 on school level	0
Classif	Classification: total correct out of 30 items	23
Compar	Comparison: total correct out of 23 items	27
Verbal	Verbal Instructions: total correct out of 50 items	20
Figure	Figure Series: total correct out of 24 items	118
Pattcomp	Pattern Completion: total correct out of 24 items	109
Knowled	Knowledge: total correct out of 32 items	112
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: word knowledge and spatial ability. The first three variables are assumed to measure *wordknow* and the last three to measure *spatial*. A path diagram of the hypothesized factor model is shown in Figure 7.2.



**Figure 7.2: Confirmatory factor analysis model for 6 variables**

The between- and within-school structural equation 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 x 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,

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

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.

```

Msem6a.ls8
Group1: Between Schools, S.A. Schools Project
DA NI=23 NO=0 NG=2 MA=CM MI=-9.0
RA = sa_schools.psf
$CLUSTER School
! $PREDICT Grade
SE
16 17 18 20 21 22/
M0 NX=6 NK=2 LX=FU,FI PH=SY,FR TD=DI,FR
LK
wordknow spatial
FR LX(2,1) LX(3,1) LX(5,2) LX(6,2)
VA 1.00 LX(1,1) LX(4,2)
PD
OU ME=ML ND=4
Group2: Within Schools, S.A. Schools Project
DA NI=23 NO=0 NG=2 MA=CM MI=-9.0
RA = sa_schools.psf
SE
16 17 18 20 21 22/
M0 NX=6 NK=2 LX=IN PH=IN TD=IN
LK
wordknow spatial
FR PH(1,1) PH(2,1) PH(2,2) TD(1,1) TD(2,2)
FR TD(3,3) TD(4,4) TD(5,5) TD(6,6)
! FR LX(2,1) LX(3,1) LX(5,2) LX(6,2)
VA 1.00 LX(1,1) LX(4,2)
OU

```

Table 7.3 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 7.3: Estimated between-schools covariance matrix,  $\Sigma_B$**

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

	Classif	Compar	Verbal	Figure	Pattcomp	Numserie
Classif	1.32					
Compar	1.32	2.75				
Verbal	2.98	3.76	10.93			
Figure	2.17	2.88	7.29	5.85		
Pattcomp	2.26	2.75	6.94	5.37	5.57	
Numserie	1.56	2.03	5.25	4.12	4.01	3.32

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

	Classif	Compar	Verbal	Figure	Pattcomp	Numserie
Classif	1.61					
Compar	1.79	3.97				
Verbal	2.26	3.35	7.03			
Figure	2.64	3.91	4.93	6.24		
Pattcomp	2.41	3.57	4.50	5.42	5.28	
Numserie	1.83	2.71	3.42	4.12	3.75	3.17

Likewise, Table 7.4 shows  $\hat{\Sigma}_W$  for the unrestricted model and  $\hat{\Sigma}_W(\hat{\gamma})$  for the CFA model.

**Table 7.4: Estimated within-schools covariance matrix,  $\hat{\Sigma}_W$**

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

	Classif	Compar	Verbal	Figure	Pattcomp	Numserie
Classif	8.47					
Compar	4.56	18.72				
Verbal	5.45	7.54	17.01			
Figure	4.39	7.12	8.28	16.11		
Pattcomp	4.26	7.14	8.28	9.41	16.08	
Numserie	3.03	4.47	5.64	7.55	6.04	7.63

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

	Classif	Compar	Verbal	Figure	Pattcomp	Numserie
Classif	8.40					
Compar	4.21	18.50				
Verbal	5.31	7.85	17.23			
Figure	4.61	6.83	8.61	15.99		
Pattcomp	4.21	6.23	7.85	9.52	16.18	
Numserie	3.20	4.73	5.97	7.23	6.60	7.77

The goodness of fit statistics for the CFA model are shown in Table 7.5.

**Table 7.5: Goodness-of-fit statistics**

(6047 students, 136 schools)

$\chi^2 = 243.408$	degrees of freedom = 20
RMSEA=0.061	

Parameter estimates and estimated standard errors are given in Table 7.6.

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 7.6: Parameter estimates and standard errors**

	Estimate	Standard error
<b>Factor loadings</b>		
$\lambda_{11}$	1.0	
$\lambda_{21}$	1.480	0.042
$\lambda_{31}$	1.866	0.047
$\lambda_{42}$	1.0	
$\lambda_{52}$	0.912	0.015
$\lambda_{62}$	0.693	0.012
<b>Factor covariances</b> (between schools)		
$\Phi_{11}$	1.212	0.186
$\Phi_{21}$	2.644	0.356
$\Phi_{22}$	5.940	0.777

**Table 7.6: Parameter estimates and standard errors (continued)**

	<b>Estimate</b>	<b>Standard error</b>
<b>Error variances</b> (between schools)		
Classif	0.399	0.078
Compar	1.321	0.221
Verbal	2.811	0.400
Figure	0.304	0.091
Pattcomp	0.333	0.089
Numserie	0.314	0.068
<b>Factor covariances</b> (within schools)		
$\Phi_{11}$	2.844	0.127
$\Phi_{21}$	4.615	0.146
$\Phi_{22}$	10.437	0.298
<b>Error variances</b> (within schools)		
Classif	5.554	0.119
Compar	12.276	0.263
Verbal	7.328	0.225
Figure	5.553	0.168
Pattcomp	7.496	0.183
Numserie	2.760	0.098

## References

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, **52**, 317-332.
- 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., 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.
- Bozdogan, H. (1987). Model selection and Akaike's information criteria (AIC). *Psychometrika*, **52**, 345-370.
- Du Toit, M. (1995). *The Analysis of Hierarchical and Unbalanced Complex Survey Data using Multilevel Models*. Unpublished Ph.D. dissertation, University of Pretoria: Pretoria.
- Goldstein, H., and McDonald, R.P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, **53**, 455-467.
- Hox, J.J. (1993). Factor analysis of multilevel data: Gauging the Muthén model. In: Oud, J.H.L. and van Blokland-Vogeleang, R.A.W. (Eds.). *Advances in longitudinal and multivariate analysis in the behavioral sciences*. Nijmegen, NL, 141-156.
- Jöreskog, K.G., and Sörbom, D. (1996). *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language*. Scientific Software International: Chicago.
- 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.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D., and Ecob, R. (1988). *School Matters, the Junior Years*. Wells, Open Books.
- 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.
- Steiger, J.H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, **25**, 173-180.