

## Multilevel Factor Models

The models described in the previous chapters are all multilevel variants of the conventional multiple regression model. This is not as restrictive as it may seem, since the multiple regression model is very flexible and can be used in many different applications (for detailed examples see Cohen & Cohen, 1983). Still, there are models that cannot be analyzed with multiple regression, notably factor analysis and path analysis models.

A general approach that includes both factor and path analysis is *Structural Equation Modeling*, or *SEM*. The interest in structural equation modeling is generally on theoretical constructs, which are represented by the latent factors. The factor model, which is often called the measurement model, specifies how the latent factors are measured by the observed variables. The relationships between the theoretical constructs are represented by regression or path coefficients between the factors. The structural equation model implies a structure for the covariances between the observed variables, which explains the alternative name *Covariance Structure Analysis*. However, the model can be extended to include means of observed variables or factors in the model, which makes covariance structure analysis a less accurate name. Many researchers will simply think of these models as ‘Lisrel-models,’ which is also less accurate. LISREL is an abbreviation of LInear Structural RELations, and the name used by Jöreskog for one of the first and most popular SEM programs. Nowadays structural equation models need not be linear, and the possibilities of SEM extend well beyond the original LISREL program. Marcoulides and Schumacker (2001), for instance, discuss the possibility to fit nonlinear curves and interaction terms in SEM.

Structural equation modeling is a general and convenient framework for statistical analysis that includes as special cases several traditional multivariate procedures, such as factor analysis, multiple regression analysis, discriminant analysis, and canonical correlation. Structural equation models are often visualized by a graphical *path diagram*. The statistical model is usually represented in a set of matrix equations. In the early seventies, when this technique was first introduced in social and behavioral research, the software usually required setups that specify the model in terms of these matrices. Thus, researchers had to distill the matrix representation from the path diagram, and provide the software with a series of matrices for the different sets of parameters, such as factor loadings and regression coefficients. A recent development is software that allows the researchers to specify the model directly as a path diagram. Since the focus in this chapter is on structural equation models for multilevel data, and not on structural equation modeling itself, the models will generally be introduced using path diagrams.

Structural equation modeling has its roots in path analysis, which was invented by the geneticist Sewall Wright (Wright, 1921). It is customary to start a SEM analysis by drawing a path diagram. A path diagram consists of boxes and circles, which are connected by arrows. In Wright’s notation, observed (or measured) variables are represented by a rectangle or square box, and latent (or unmeasured) factors by a circle or ellipse. Single-headed arrows or ‘paths’ are used to define hypothesized causal relationships in the model, with the variable at the tail of the arrow being the cause of the variable at the point. Double-headed arrows indicate covariances or correlations, without a causal interpretation. Statistically, the single-headed arrows or paths represent regression coefficients, and double-headed arrows covariances.

Often a distinction is made between the measurement model and the structural model. The measurement model, which is a confirmatory factor model, specifies how the latent factors are related to the observed variables. The structural model contains the relationships between the latent factors. In this chapter, I discuss multilevel factor analysis, and introduce the techniques currently available to estimate multilevel factor models. Multilevel path models, which are structural models that may or may not include latent factors, are discussed in Chapter Fourteen. For a general introduction in SEM, I refer to the introductory article by Hox and Bechger (1998) or introductory books such as Loehlin (2004) and Schumacker and Lomax (2004). A statistical treatment is presented by Bollen (1989). An interesting collection of introductory articles focusing on SEM models for multigroup and longitudinal data is found in Little, Schnabel and Baumert (2000).

Structural equation models are often specified as models for the means and covariance matrix of multivariate normal data. The model is then

$$y_i = \mu + \Lambda \eta_i + \varepsilon, \quad (14.1)$$

which states that the observed variables  $y_i$  are predicted by a regression equation involving an intercept  $\mu$  and the regression coefficients or loadings in factor matrix  $\Lambda$  multiplying the unobserved factor scores  $\eta_i$  plus a residual measurement error  $\varepsilon$ . This can then be expressed as a model for the covariance matrix  $\Sigma$  by

$$\Sigma = \Lambda \Phi \Lambda^{\top} + \Theta, \quad (14.2)$$

where the covariance matrix  $\Sigma$  is expressed as a function of the factor matrix  $\Lambda$ , the matrix of covariances between factors  $\Phi$ , and residual measurement errors in  $\Theta$ .

This chapter discusses two different approaches to multilevel SEM. The first approach, described by Rabe-Hesketh, Skrondal and Zheng (2007) as the ‘within and between formulation’ focuses on determining separate estimates for the within (subject level) covariance matrix and the between (group level) covariance matrix. These are then modeled separately or simultaneously by an subject level (within) factor model and a group level (between) factor model, analogous to the single level equation given in 14.2. This works well, but has limitations, which will be discussed in the next section (14.1) that describes this approach in more detail. The second approach models the observed multilevel data directly with a model that includes variables at each available level and accommodates group-level variation of intercepts and slopes. It is the most accurate and versatile approach, but in some circumstances computationally demanding. It is also at the moment not widely available in standard multilevel or SEM software. This approach is described in section 14.2.

## 14.1 THE WITHIN AND BETWEEN APPROACH

The within and between approach is based on an analysis of the subject level and the group level covariance matrix. This is in turn based on a decomposition of the variables to the available levels, which is discussed in the next section.

### 14.1.1 Decomposing multilevel variables

Multilevel structural models assume that we have a population of individuals that are divided into groups. The individual data are collected in a  $p$ -variate vector  $\mathbf{Y}_{ig}$  (subscript  $i$  for

individuals,  $j$  for groups). Cronbach and Webb (1975) have proposed decomposing the individual data  $Y_{ij}$  into a between groups component  $Y_B = \bar{Y}_j$ , and a within groups component  $Y_W = Y_{ij} - \bar{Y}_j$ . In other words, for each individual we replace the observed Total score  $Y_T = Y_{ij}$  by its components: the group component  $Y_B$  (the disaggregated group mean) and the individual component  $Y_W$  (the individual deviation from the group mean.) These two components have the attractive property that they are orthogonal and additive (cf. Searle, Casella & McCulloch, 1992):

$$Y_T = Y_B + Y_W. \quad (14.3)$$

This decomposition can be used to compute a between groups covariance matrix  $\Sigma_B$  (the population covariance matrix of the disaggregated group means  $Y_B$ ) and a within groups covariance matrix  $\Sigma_W$  (the population covariance matrix of the individual deviations from the group means  $Y_W$ ). These covariance matrices are also orthogonal and additive:

$$\Sigma_T = \Sigma_B + \Sigma_W. \quad (14.4)$$

Following the same logic, we can also decompose the sample data. Suppose we have data from  $N$  individuals, divided into  $G$  groups (subscript  $i$  for individuals,  $i=1 \dots N$ ; subscript  $g$  for groups,  $g=1 \dots G$ ). If we decompose the sample data, the sample covariance matrices are also orthogonal and additive:

$$S_T = S_B + S_W. \quad (14.5)$$

Multilevel structural equation modeling assumes that the population covariance matrices  $\Sigma_B$  and  $\Sigma_W$  are described by distinct models for the between groups and within groups structure. To estimate the model parameters, the factor loadings, path coefficients, and residual variances, we need maximum likelihood estimates of the population between groups covariance matrix  $\Sigma_B$  and the population within groups covariance matrix  $\Sigma_W$ . What we have is the observed between groups matrix  $S_B$  and the observed within groups matrix  $S_W$ . It would be nice, if we could simply construct a within groups model for  $S_W$ , and a between groups model for  $S_B$ . Unfortunately, we cannot simply use  $S_B$  as an estimate of  $\Sigma_B$ , and  $S_W$  for  $\Sigma_W$ . The situation is more complicated. Several different approaches have been offered for estimating multilevel factor models. This section describes three approaches: the pseudobalanced (MUMML) approach, the two phase direct estimation approach, and a weighted least squares approach.

#### 14.1.2 Muthén's pseudobalanced approach

In the special case of balanced groups, meaning that all groups are the same size, estimation of a multilevel structural equation model is straightforward (Muthén, 1989). If we have  $G$  balanced groups, with  $G$  equal group sizes  $n$  and a total sample size  $N=nG$ , we define two sample covariance matrices: the pooled within covariance matrix  $S_{PW}$  and the scaled between covariance matrix  $S_B^*$ .

Muthén (1989) shows that an unbiased estimate of the population within groups covariance matrix  $\Sigma_W$  is given by the pooled within groups covariance matrix  $S_{PW}$ , calculated in the sample by:

$$\mathbf{S}_{PW} = \frac{\sum_j^G \sum_i^n (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_j)(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_j)'}{N - G} \quad (14.6)$$

Equation (14.6) corresponds to the conventional equation for the covariance matrix of the individual deviation scores, with  $N-G$  in the denominator instead of the usual  $N-1$ .

Since the pooled within groups covariance matrix  $\mathbf{S}_{PW}$  is an unbiased estimate of the population within groups covariance matrix  $\boldsymbol{\Sigma}_W$ , we can estimate the population within group structure directly by constructing and testing a model for  $\mathbf{S}_{PW}$ .

The scaled between groups covariance matrix for the disaggregated group means  $\mathbf{S}_B^*$ , calculated in the sample, is given by:

$$\mathbf{S}_B^* = \frac{\sum_j^G n(\bar{\mathbf{Y}} - \bar{\mathbf{Y}}_j)(\bar{\mathbf{Y}} - \bar{\mathbf{Y}}_j)'}{G - 1}. \quad (14.7)$$

Muthén (1989, 1990) shows that  $\mathbf{S}_{PW}$  is the maximum likelihood estimator of  $\boldsymbol{\Sigma}_W$ , with sample size  $N-G$ , and  $\mathbf{S}_B^*$  is the maximum likelihood estimator of the composite  $\boldsymbol{\Sigma}_W + c\boldsymbol{\Sigma}_B$ , with sample size  $G$ , and  $c$  equal to the common group size  $n$ :

$$\mathbf{S}_{PW} = \widehat{\boldsymbol{\Sigma}}_W \quad (14.8)$$

and

$$\mathbf{S}_B^* = \widehat{\boldsymbol{\Sigma}}_W + c\widehat{\boldsymbol{\Sigma}}_B. \quad (14.9)$$

Equations 14.8 and 14.9 suggest using the multi-group option of conventional SEM software for a simultaneous analysis at both levels. However, if we model the between groups structure, we cannot simply construct and test a simple model for  $\boldsymbol{\Sigma}_B$ , because  $\mathbf{S}_B^*$  estimates a combination of  $\boldsymbol{\Sigma}_W$  and  $\boldsymbol{\Sigma}_B$ . Instead, we have to specify for the  $\mathbf{S}_B^*$  ‘group’ a model that contains two distinct sub-models: one for the within structure and one for the between structure. The procedure is that we specify two groups, with covariance matrices  $\mathbf{S}_{PW}$  and  $\mathbf{S}_B^*$  (based on  $N-G$  and  $G$  observations). The model for  $\boldsymbol{\Sigma}_W$  must be specified for both  $\mathbf{S}_{PW}$  and  $\mathbf{S}_B^*$ , with equality restrictions between both ‘groups’ to guarantee that we are indeed estimating the same model in both covariance matrices, and the model for  $\boldsymbol{\Sigma}_B$  is specified for  $\mathbf{S}_B^*$  only, with the scale factor  $c$  built into the model.

The reasoning strictly applies only in the so-called *balanced* case, that is, if all groups have the same group size. In the balanced case, the scale factor  $c$  is equal to the common group size  $n$ . The unbalanced case, where the group sizes differ, with  $G$  groups of unequal sizes, is more complicated. In this case,  $\mathbf{S}_{PW}$  is still the maximum likelihood estimator of  $\boldsymbol{\Sigma}_W$ , but  $\mathbf{S}_B^*$  now estimates a different expression for each set of groups with distinct group size  $d$ :

$$\mathbf{S}_{Bd}^* = \widehat{\boldsymbol{\Sigma}}_W + c_d \widehat{\boldsymbol{\Sigma}}_B, \quad (14.10)$$

where equation 14.10 holds for each distinct set of groups with a common group size equal to  $n_d$ , and  $c_d = n_d$  (Muthén, 1990, 1994). Full Information Maximum Likelihood (FIML) estimation for

unbalanced groups implies specifying a separate between-group model for each distinct group size. These between groups models have different scaling parameters  $c_d$  for each distinct group size, and require equality constraints across all other parameters and inclusion of a mean structure (Muthén, 1994, p. 385). Thus, using conventional SEM software for unbalanced data requires a complicated modeling scheme that creates a different ‘group’ for each set of groups with the same group size. This results in large and complex models, with possibly groups with a sample size less than the number of elements in the corresponding covariance matrix. This makes full Maximum Likelihood estimation problematic, and therefore Muthén (1989, 1990) proposes to ignore the unbalance, and to compute a single  $\mathbf{S}_B^*$ . The model for  $\mathbf{S}_B^*$  includes an ad hoc estimator  $c^*$  for the scaling parameter, which is close to the average sample size:

$$c^* = \frac{N^2 - \sum_j n_j^2}{N(G-1)}. \tag{14.11}$$

The result is a limited information Maximum Likelihood solution, which McDonald (1994) calls a pseudobalanced solution, and Muthén (1989, 1994) the MUML (for Muthén’s ML) solution.

Figure 6.1 presents a path diagram for the pseudobalanced model. Left in the figure is the within model for the pooled within matrix. Right in the diagram is the model for the scaled between matrix. It repeats the model for the within matrix, with equality constraints across all corresponding parameters. On top of this, for each variable a corresponding between variable is specified. These between level variables are latent variables representing the second or between level intercept variances.  $C$  is the scaling constant. The between level model is constructed for the latent between variables.

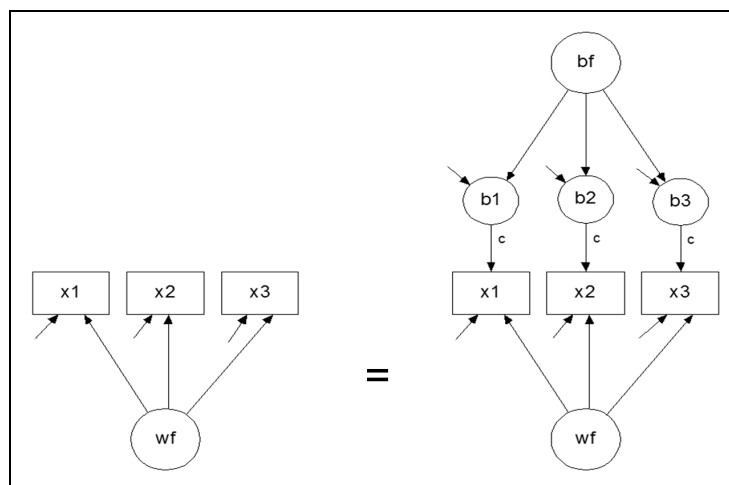


Figure 14.1 Path diagram for pseudobalanced model

Muthén (1989, 1990) shows that  $\mathbf{S}_B^*$  is a consistent and unbiased estimator of the composite  $\Sigma_W + c\Sigma_B$ . This means that with large samples (of both individuals *and* groups!)  $\mathbf{S}_B^*$  is a close estimate of  $\Sigma_B$ , and the pseudobalanced solution produces a good approximation given adequate sample sizes.

Simulation studies (Hox and Maas, 2001a; Hox, Maas & Brinkuis, 2008) find that the within groups part of the model poses no problems in any of the simulated conditions. In the between groups part of the model, the factor loadings are generally accurate. However, the

residual variances are underestimated, and the standard errors are generally too small, leading to an operating alpha level of about 8%. In addition, the chi-square model test is rejected too often, which again results in an operating alpha level of about 8%. Yuan and Hayashi (2005) show analytically that MUML standard errors and chi-square tests lead to correct inferences when the between level sample size goes to infinity and the coefficient of variation of the group sizes goes to zero. Both simulations and analytical work agree that larger sample sizes do not improve the accuracy with unbalanced data. So, with severely unbalanced data, MUML produces biased standard errors and significance tests, and this bias is not diminished when the sample size is increased.

Since the pseudobalanced approach needs the within groups model both for the pooled within groups and the scaled between groups model, and needs to incorporate the scaling factor for the between groups model, the actual model can become quite complicated. Most modern structural equation software includes the MUML approach for two-level data, generally combined with default robust estimators for the standard errors and the chi-square test statistic to correct for the remaining heterogeneity (cf. Muthén & Satorra, 1995). The complications of the setup are generally hidden for the user. In software that does not include MUML as an option, it is still possible to calculate  $S_{PW}$  and  $S_B$  separately, and constructing the complicated setup. For details I refer to Hox (1993) and the first edition of this book (Hox, 2002).

#### 14.1.3 Direct estimation of the within and between matrix

Goldstein (1987, 2003) has suggested using the multivariate multilevel model described in Chapter Ten to produce a covariance matrix at the different levels, and to input these into a standard SEM program for further analysis. For our family data, we would use a multivariate multilevel model with three separate levels for the six intelligence tests, the individual children, and the families. We create six dummy variables to indicate the six intelligence scales, and exclude the intercept from the model. Hence, at the lowest level (the variables level) we have

$$Y_{hij} = \pi_{1ij}d_{1ij} + \pi_{2ij}d_{2ij} + \dots + \pi_{6ij}d_{6ij}, \quad (14.12)$$

at the individual level we have

$$\pi_{pij} = \beta_{pj} + u_{pij}, \quad (14.13)$$

and at the family level (the third level in the multivariate model), we have

$$\beta_{pj} = \gamma_p + u_{pj}. \quad (14.14)$$

By substitution we obtain

$$Y_{hij} = \gamma_1 d_{1ij} + \gamma_2 d_{2ij} + \dots + \gamma_p d_{pij} + u_{1ij} d_{1ij} + u_{2ij} d_{2ij} + \dots + u_{pij} d_{pij} + u_{1j} d_{1ij} + u_{2j} d_{2ij} + \dots + u_{pj} d_{pij}. \quad (14.15)$$

In sum notation, we have

$$Y_{hij} = \sum_{h=1}^6 \gamma_h d_{hij} + \sum_{h=1}^6 u_{hij} d_{hij} + \sum_{h=1}^6 u_{hj} d_{hij}. \quad (14.16)$$

The model described by equation (14.16) provides us with estimates of the six test means, and of their variances and covariances at the individual and family level. Since in this application we are mostly interested in the variances and covariances, RML estimation is preferred to FML estimation. The individual level (within) covariances and the family level (between) covariances and means are direct maximum likelihood estimators of their population counterparts, which can be input to any SEM software, either separately or in a two-group analysis. Hox and Maas (2004) explain this approach in more detail.

The fact that the individual level (within families) and family level (between families) covariances are estimated directly, and consequently can be modeled directly and separately by any SEM program, is a distinct advantage of the multivariate multilevel approach. As a result, we get separate model tests and fit indices at all levels. The multivariate multilevel approach to multilevel SEM also generalizes straightforwardly to more than two levels. There are other advantages as well. First, since the multilevel multivariate model does not assume that we have a complete set of variables for each individual, incomplete data are accommodated without special effort. Second, if we have dichotomous or ordinal variables, we can use the multilevel generalized linear model to produce the covariance matrices, again without special effort.

There are some disadvantages to the multivariate multilevel approach as well. An important disadvantage is that the covariances produced by the multivariate multilevel approach are themselves estimated values. They are not directly calculated, as the pooled within groups and scaled between groups covariances are, but they are estimates produced by a complex statistical procedure. The estimation in the second step treats these estimates as data or observed covariance matrices. If the data have a multivariate normal distribution, the within groups and between groups covariances can be viewed as observed values, which have a known sampling distribution. However, when we have incomplete data it is unclear what the proper sample size is, and in the case of non-normal (e.g., dichotomous or ordinal) variables we know that the sampling distribution is not normal. Since the normal sampling distribution is used by SEM programs to calculate the chi-square model test and the standard errors of the parameter estimates, the chi-square test and standard errors cannot be trusted unless data are complete and multivariate normal.

#### 14.1.4 Analysis of the within and between matrix using weighted least squares

Asparouhov and Muthén (2007) describe an approach to multilevel SEM that uses separate estimation of the covariance matrices followed by an estimation method that overcomes the problems that are encountered with direct estimation of the within and between covariance matrix with non-normal data. In this approach, univariate maximum likelihood methods are used to estimate the vector of means  $\boldsymbol{\mu}$  at the between group level, and the diagonal elements (the variances) of  $\mathbf{S}_W$  and  $\mathbf{S}_B$ . In the case of ordinal categorical variables, thresholds are estimated as well. Next, the off-diagonal elements of  $\mathbf{S}_W$  and  $\mathbf{S}_B$  are estimated using bivariate maximum likelihood methods. Finally, the asymptotic variance-covariance matrix for these estimates is obtained, and the multilevel SEM is estimated for both levels using Weighted Least Squares (WLS). Currently, this approach is only available in Mplus.

WLS is an estimation method that uses the variance-covariance matrix of  $\mathbf{S}_W$  and  $\mathbf{S}_B$  as a weight matrix to obtain correct chi-squares and standard errors. This estimation method is developed for efficient estimation of multilevel models with non-normal variables, since for such data full maximum likelihood estimation requires high-dimensional numerical integration, which is computationally very demanding. Multilevel WLS can also be used for multilevel estimation with continuous variables, but then it does not have a real advantage.

Standard WLS uses a weight matrix based on the asymptotic covariances of all estimated parameters. For the unrestricted model, the number of parameters is large, and the asymptotic

covariance matrix is also large. Especially for the between part of the model, the number of elements in this matrix can easily become larger than the number of groups. Unless the number of groups is very large, it is preferable to use only the diagonal of this matrix, (cf. Muthén, Du Toit, & Spisic, 1997). In Mplus, choosing the diagonal weight matrix always implies using a robust chi-square (WLSM using a mean corrected (first order) and WLSMV using a mean-and-variance corrected (second order) correction).

## 14.2 FULL MAXIMUM LIKELIHOOD ESTIMATION

In two-level data, the factor structure given by 14.1 becomes

$$\begin{aligned} y_{ij} &= \mu_j + \Lambda_W \eta_{ij} + \varepsilon_W \\ \mu_j &= \mu + \Lambda_B \eta_j + \varepsilon_B \end{aligned} \quad (14.17)$$

where  $\mu_j$  are the random intercepts that vary across groups. The first equation models the within groups variation, and the second equation models the between groups variation. By substitution and rearranging terms we obtain

$$y_{ij} = \mu + \Lambda_W \eta_{ij} + \Lambda_B \eta_j + \varepsilon_B + \varepsilon_W. \quad (14.18)$$

Except for the notation, the structure of equation 14.18 follows that of a random intercept regression model, with fixed regression coefficient in the factor matrices  $\Lambda$  and a level-one and level-two error term. By allowing group level variation in the factor loadings we can generalize this to a random coefficient model. In the context of multilevel factor analysis varying loadings are problematic because they imply that the measurement model is not equivalent across the different groups. In the context of multilevel path analysis, random coefficients for relationships between variables provide information on differences between groups that have a substantive interpretation.

To provide maximum likelihood estimates for the parameters in 14.18 in the general case of unbalanced groups we need to model the observed raw data. The two-stage approaches all follow the conventional notion that structural equation models are constructed for the covariance matrix with possibly an added mean vector. When data have a multivariate normal distribution, these are the sufficient statistics, and raw data are superfluous. Thus, for a confirmatory factor model, the covariance matrix  $\Sigma$  is

$$\Sigma = \Lambda \Phi \Lambda' + \Theta, \quad (14.19)$$

Where  $\Lambda$  is the factor matrix,  $\Phi$  is the covariance matrix of the latent variables and  $\Theta$  is the vector with residual variances. The parameters are commonly estimated by maximizing the likelihood function, or equivalently minimizing the discrepancy function (Jöreskog, 1967; Browne, 1982):

$$F = \log |\Sigma| + \text{tr}(\mathbf{S}\Sigma^{-1}) - \log |\mathbf{S}| - p, \quad (14.20)$$

where “|.” indicates the determinant of a matrix, “tr” indicates the trace, and  $p$  is the number of manifest variables.



Unbalanced groups can be viewed as a form of incomplete data. For incomplete data, the maximum likelihood approach defines the model and the likelihood in terms of the raw data, which is why it is sometimes called the raw likelihood method. Raw ML minimizes the function (Arbuckle, 1996):

$$F = \sum_{i=1}^N \log |\Sigma_i| + \sum_{i=1}^N \log (\mathbf{x}_i - \boldsymbol{\mu}_i)' \Sigma_i^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_i), \quad (14.21)$$

where the subscript  $i$  refers to the observed cases,  $\mathbf{x}_i$  to the variables observed for case  $i$ , and  $\boldsymbol{\mu}_i$  and  $\Sigma_i$  contain the population means and covariances of the variables observed for case  $i$ . If the data are complete, equations 14.20 and 14.21 are equivalent. If the data are incomplete, the covariance matrix is no longer a sufficient statistic, and minimizing the discrepancy function given by 14.21 provides the maximum likelihood estimates for the incomplete data.

Mehta and Neale (2005) show that models for multilevel data, with individuals nested within groups, can be expressed as a structural equation model. The fit function given by equation (14.21) applies, with clusters as units of observation, and individuals within clusters as variables. Unbalanced data, here unequal numbers of individuals within clusters, are included the same way as incomplete data in standard SEM. While the two-stage approaches (MUMML, direct estimation and WLS) include only random intercepts in the between groups model, the ML representation can incorporate random slopes as well (Mehta & Neale, 2005). In theory, any modern SEM software that allows for incomplete data can be used to estimate multilevel SEM. In practice, specialized software is used that makes use of the specific multilevel structure in the data to simplify calculations. Full maximum likelihood multilevel SEM is currently available for two-level models in Mplus and for many-level models in GLLAMM. A recent development is to use robust standard errors and chi-squares for significance testing. With multilevel data, robust chi-squares and standard errors offer some protection against unmodeled heterogeneity, which may result from misspecifying the group level model, or by omitting a level. Finally, Skrondal and Rabe-Hesketh (2004) show how to combine this with a generalized linear model for the observed variables, which allows for non-normal variables. This is currently available only in Mplus and GLLAMM.

The next section analyzes an example data set, using all four estimation methods discussed above. When all methods are compared, the pseudobalanced MUMML approach and the multivariate multilevel two stage estimation method are the least accurate, especially for the between model estimates. These methods are clearly outdated. Simulations (Hox, Maas & Brinkhuis, 2008) have shown that both WLS and full ML estimation are more accurate than MUMML, and that the difference between WLS and ML is negligible. Our example confirms that ML and WLS are quite similar. When ML estimation is possible, it is the method of choice, but when the demands for ML estimation overtax the computer capacity, WLS is a viable alternative.

The maximum likelihood approach is the only approach that allows random slopes in the model. In a confirmatory factor analysis, this means that factor loadings are allowed to vary across groups. In our example, none of the six individual-level factor loadings has significant slope variation across families. In confirmatory factor analysis this is desirable, because finding varying factor loadings implies that the scales involved are not measuring the same thing in the same way across families. Varying slopes in a measurement model indicate lack of measurement equivalence across groups.

### 14.3 AN EXTENDED EXAMPLE OF MULTILEVEL FACTOR ANALYSIS

The example data are the scores on six intelligence measures of 400 children from 60 families, patterned after van Peet (1992). The six intelligence measures are: word list, cards, matrices, figures, animals, and occupations. The data have a multilevel structure, with children nested within families. If intelligence is strongly influenced by shared genetic and environmental influences in the families, we may expect rather strong between family effects. In this data set, the intraclass correlations of the intelligence measures range from 0.38 to 0.51.

### 14.3.1 Full Maximum Likelihood estimation

Given that full maximum likelihood estimation is the norm, we begin the analysis of the example data using this estimation method. Muthén (1994) recommends starting the analysis with an analysis of the Total scores. This may have been good advice when the complicated pseudobalanced model setups were used, but given user-friendly multilevel SEM software, this step is superfluous. Since the effective level-one sample size ( $N-G$ ) is almost always much larger than the level-two sample size ( $G$ ), it is good practice to start with the within part, either by specifying a saturated model for the between part, or by analyzing only the pooled within matrix.

In the example data, the number of observations on the individual level is  $275-50=225$ , while on the family level it is 50. Thus, it makes sense to start on the individual level by constructing a model for  $S_{PW}$  only, ignoring  $S_B^*$ .

An exploratory factor analysis on the correlations derived from  $S_{PW}$  suggests two factors, with the first three measures loading on the first factor, and the last three measures on the last. A confirmatory factor analysis on  $S_{PW}$  confirms this model ( $\chi^2=6.0$ ,  $df=8$ ,  $p=0.56$ ). A model with just one general factor for  $S_{PW}$  is rejected ( $\chi^2=207.6$ ,  $df=9$ ,  $p<0.001$ ). Figure 14.1 presents the conventional path diagram of the individual level (within families) model.

The next step is to specify a family model. For explorative purposes, we could carry out a separate analysis on the estimated between groups covariance matrix  $S_B$ . This matrix, which is a maximum likelihood estimator of  $\Sigma_B$  (not the scaled between matrix produced by 14.7), is obtained by specifying a saturated model for both the within and the between level (Mplus generates these matrices automatically). In the example data, given the good fit of the within model, we carry on with an analysis of the multilevel data, with the two-factor model retained for the within part.

We start the analysis of the between structure by estimating some ‘benchmark’ models for the group level, to test whether there is any between family structure at all. The simplest model is a null-model that completely leaves out the specification of a family level model. If the null-model holds, there is no family level structure at all; all covariances in  $S_B$  are the result of individual sampling variation. If this null-model holds, we may as well continue our analyses using simple single level analysis methods.

The next model tested is the independence model, which specifies only variances on the family level, but no covariances. The independence model estimates for the family level structure only the variances of the family level variables. If the independence model holds, there is family level variance, but no substantively interesting structural model. We can simply analyze the pooled within matrix, at some cost in losing the within groups information from  $G$  observations that is contained in the between groups covariance matrix. If the independence model is rejected, there is some kind of structure on the family level. To examine the best possible fit given the individual level model, we can estimate the saturated model; which fits a full covariance matrix to the family level observations. This places no restrictions on the family model. Table 14.3 shows the results of estimating these benchmark models on the family level.

Table 14.1. Family level benchmark models

Family model	Chi-square	df	p
Null	323.6	29	.00
Independence	177.2	23	.00
Saturated	6.7	8	.57

The null model and the independence model are both rejected. Subsequently, we specify for the family level the same one-factor and two-factor models we have used at the individual level. The one-factor model fits well ( $\chi^2=11.9, df=17, p=0.80$ ). The two-factor model is no improvement (difference chi-square 0.15,  $p=0.70$ ).

The principle of using the simplest model that fits well leads to acceptance of the one factor model on the family level. The chi-square model test is not significant, and the fit indices are fine: CFI is 1.00 and the RMSEA is 0.00. Figure 14.2 presents the within and between model in a single path diagram. Note that, consistent with the full path diagram in Figure 14.3, the between level variables that represent the family level intercept variance of the observed variables, are latent variables, depicted by circles or ellipses.

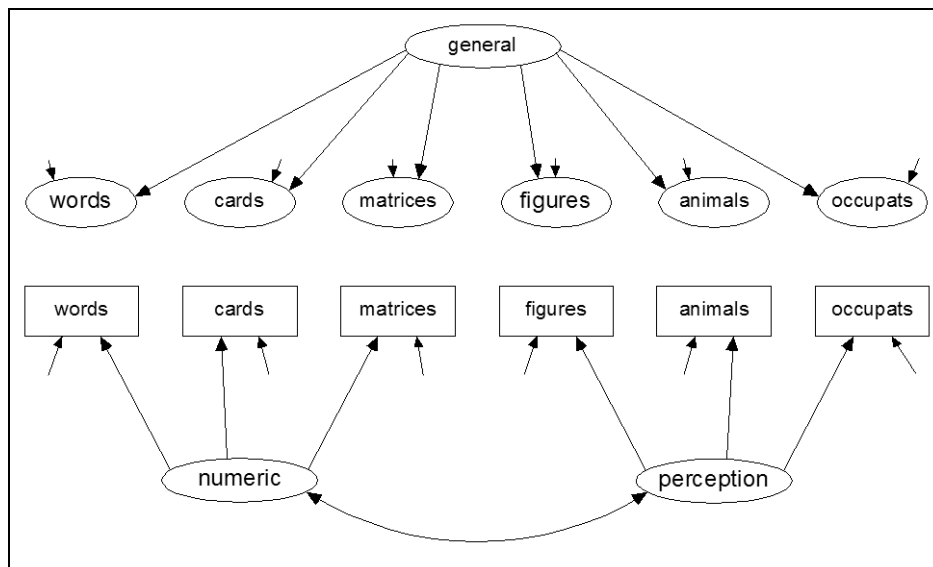


Figure 14.2 Diagram for family IQ data 1

Using the maximum likelihood method (ML) in Mplus leads to the estimates reported in Table 14.2.

Table 14.2 Individual and family level estimates, ML estimation

	Individual level			Family level	
	Numer.	Percept.	resid. var.	General	resid. var.
Wordlst	3.18 (.20)		6.19 (.74)	3.06 (.39)	1.25 (.57)
Cards	3.14 (.19)		5.40 (.69)	3.05 (.39)	1.32 (.59)
Matrix	3.05 (.20)		6.42 (.71)	2.63 (.38)	1.94 (.67)

Figures	3.10 (.20)	6.85 (.76)	2.81 (.40)	2.16 (.71)
Animals	3.19 (.19)	4.88 (.70)	3.20 (.38)	0.66 (.49)
Occupat	2.78 (.18)	5.33 (.60)	3.44 (.42)	1.58 (.62)

Standard errors in parentheses. Correlation between individual factors: 0.38.

### 14.3.2 Weighted Least Squares estimation

Using the separate estimation/WLS method with robust chi-square (WLSMV) in Mplus leads to the estimates reported in Table 14.7.<sup>1</sup> The chi-square model test accepts the model ( $\chi^2=5.91$ ,  $df=7$ ,  $p=0.55$ ), the fit indices are good: CFI is 1.00 and the RMSEA is 0.00. The parameter estimates in Table 14.3 are similar to the full Maximum Likelihood estimates, but not identical. The robust standard errors lead to the same conclusions as the asymptotic standard errors used with full Maximum Likelihood estimation.

Table 14.3 Individual and family level estimates, WLS estimation

	Individual level			Family level	
	<b>Numer.</b>	<b>Percept.</b>	<b>resid. var.</b>	<b>General</b>	<b>resid. var.</b>
Wordlst	3.25 (.15)		5.67 (.84)	3.01 (.48)	1.51 (.62)
Cards	3.14 (.18)		5.44 (.68)	3.03 (.38)	1.25 (.71)
Matrix	2.96 (.22)		6.91 (.92)	2.62 (.45)	2.02 (.69)
Figures		2.96 (.22)	7.67 (.92)	2.80 (.46)	2.03 (.72)
Animals		3.35 (.21)	3.79 (.99)	3.15 (.41)	0.96 (.61)
Occupat		2.75 (.24)	5.49 (.94)	3.43 (.44)	1.67 (.63)

Standard errors in parentheses. Correlation between individual factors: 0.38.

### 14.3.3 Direct estimation of the within and between matrix followed by separate modeling

Since the family example data are multivariate normal and complete, the direct estimation method discussed in section 14.2.3 is a viable option. Covariance matrices were estimated for the within and the between part using the multivariate multilevel model presented in Chapter Ten. Estimation was carried out in MLwiN Restricted Maximum Likelihood; subsequent SEM modeling was done in Amos. The within model does fits well:  $\chi^2=6.9$ ,  $df=8$ ,  $p=.55$ , CFI is 1.00 and the RMSEA is 0.00. Analysis of the between matrix produces inconsistent results:  $\chi^2=15.8$ ,  $df=9$ ,  $p<.07$ , CFI is 0.99 but the RMSEA is 0.11. The parameter estimates for the model are presented in Table 14.4.

Table 14.4 Individual and family level estimates, MUML estimation

<sup>1</sup> The asymptotic chi-square is available by specifying full WLS estimation. As explained earlier, this leads to a very large weight matrix. With a group level sample size of only 60 this is a recipe for disaster, hence the choice for a robust WLS estimation.

	Individual level			Family level	
	<b>Numer.</b>	<b>Percept.</b>	<b>resid. var.</b>	<b>General</b>	<b>resid. var.</b>
Wordlst	3.14 (.20)		6.21 (.73)	3.04 (.32)	1.52 (.33)
Cards	3.15 (.20)		5.32 (.69)	3.07 (.32)	1.13 (.26)
Matrix	3.03 (.20)		6.40 (.71)	2.64 (.31)	1.16 (.43)
Figures		3.11 (.21)	6.80 (.75)	2.81 (.32)	2.03 (.41)
Animals		3.20 (.19)	4.84 (.68)	3.18 (.32)	0.82 (.22)
Occupat		2.81 (.18)	5.34 (.60)	3.44 (.36)	1.64 (.39)

Standard errors in parentheses. Correlation between individual factors: 0.38.

#### 14.3.4 Pseudobalanced (MUML) estimation

Using the pseudobalanced (MUML) estimation in Mplus leads to the estimates reported in Table 14.8. The chi-square model test accepts the model ( $\chi^2=11.3$ ,  $df=17$ ,  $p=0.8455$ ), the fit indices are good: CFI is 1.00 and the RMSEA is 0.00. The parameter estimates in Table 14.5 are similar to the full Maximum Likelihood estimates, but not identical. The standard errors lead to the same conclusions as the asymptotic standard errors used with full Maximum Likelihood estimation.

Table 14.5 Individual and family level estimates, MUML estimation

	Individual level			Family level	
	<b>Numer.</b>	<b>Percept.</b>	<b>resid. var.</b>	<b>General</b>	<b>Resid. var.</b>
Wordlst	3.16 (.20)		6.24 (.74)	2.95 (.38)	1.32 (.58)
Cards	3.15 (.19)		5.36 (.69)	2.97 (.38)	1.35 (.57)
Matrix	3.06 (.20)		6.37 (.72)	2.60 (.37)	1.95 (.66)
Figures		3.10 (.20)	6.85 (.76)	2.77 (.39)	2.19 (.71)
Animals		3.19 (.19)	4.81 (.68)	3.12 (.37)	0.70 (.46)
Occupat		2.78 (.18)	5.33 (.60)	3.31 (.40)	1.59 (.61)

Standard errors in parentheses. Correlation between individual factors: 0.38.

## 14.5 STANDARDIZING ESTIMATES IN MULTILEVEL STRUCTURAL EQUATION MODELING

The estimates reported are all unstandardized estimates. For interpretation, it is often useful to inspect the standardized estimates as well, because these can be used to compare the loadings and residual variances for variables that are measured in a different metric. A convenient standardization is to standardize both the latent factors and the observed variables on each level separately. Table 14.6 presents the standardized estimates for the ML estimates. It shows that the factor structure at the family level is stronger than at the individual level. This is typical; one

reason is that measurement errors accumulate at the individual level.

Table 14.6 Individual and family level estimates, standardized estimates

	Individual level			Family level	
	<b>Numer.</b>	<b>Percept.</b>	<b>resid. var.</b>	<b>General</b>	<b>Resid. var.</b>
Wordlst	0.79 (.03)		0.38 (.05)	0.94 (.03)	0.12 (.06)
Cards	0.80 (.03)		0.35 (.05)	0.94 (.03)	0.12 (.06)
Matrix	0.77 (.03)		0.41 (.05)	0.88 (.05)	0.22 (.08)
Figures		0.76 (.03)	0.42 (.05)	0.89 (.04)	0.22 (.08)
Animals		0.82 (.03)	0.32 (.05)	0.97 (.02)	0.06 (.05)
Occupat		0.77 (.03)	0.40 (.05)	0.94 (.03)	0.12 (.05)

Standard errors in parentheses. Correlation between individual factors: 0.38.

The separate standardization presented in Table 14.8 is called the within groups completely standardized solution. Standardization takes place separately in the within part and in the between part. Some software also produces a ‘common metric solution,’ which standardizes the variables using a common metric across the groups or levels. In multilevel SEM, this solution produces standardized estimates that have no meaningful interpretation. Gustafsson and Stahl (1999) propose a different standardization, built into their preprocessor STREAMS. This is a multilevel standardization which standardizes the latent variables at each separate level, but uses the total standard deviation of the observed variables to standardize both the within groups and the between groups level. This provides a better insight into how much variance each factor explains at the different levels. This multilevel standardization can be accomplished by hand calculation using other SEM software.<sup>1</sup> Summarizing: some careful thought is needed in choosing the correct standardization method.

## 14.6 GOODNESS OF FIT IN MULTILEVEL STRUCTURAL EQUATION MODELING

SEM programs produce in addition to the chi-square test a number of goodness-of-fit indices that indicate how well the model fits the data. Statistical tests for model fit have the problem that their power varies with the sample size. If we have a very large sample, the statistical test will almost certainly be significant. Thus, with large samples, we will always reject our model, even if the model actually describes the data quite well. Conversely, with a very small sample, the model will always be accepted, even if it fits rather badly.

Given the sensitivity of the chi-square statistic to the sample size, researchers have proposed a variety of alternative fit indices to assess model fit. All goodness-of-fit measures are some function of the chi-square and the degrees of freedom. Most of these fit indices do not only consider the fit of the model, but also its simplicity. A saturated model, that specifies all possible paths between all variables, always fits the data perfectly, but it is just as complex as the observed data. In general, there is a trade-off between the fit of a model and the simplicity of a model. Several goodness-of-fit indices assess simultaneously both the fit and

<sup>1</sup>Loadings are standardized by dividing by the standard deviation of the variables, variances by dividing by the square of the standard deviation.

the simplicity of a model. The goal is to produce a goodness-of-fit index that does not depend on the sample size or the distribution of the data. In fact, simulations have shown that most goodness-of-fit indices still depend on sample size and distribution, but the dependency is much smaller than that of the routine chi-square test.

Most SEM software computes a bewildering array of goodness-of-fit indices. All of them are functions of the chi-square statistic, but some include a second function that penalizes complex models. For instance, Akaike's information criterion (AIC) is twice the chi-square statistic minus the degrees of freedom for the model. For a detailed review and evaluation of a large number of fit indices, including those mentioned here, I refer to Gerbing and Anderson (1993).

Jöreskog and Sörbom (1989) have introduced two goodness-of-fit indices called GFI (Goodness of Fit) and AGFI (Adjusted GFI). The GFI indicates goodness-of-fit, and the AGFI attempts to adjust the GFI for the complexity of the model. Bentler (1990) has introduced a similar index called the Comparative Fit Index CFI. Two other well-known fit measures are the Tucker-Lewis Index TLI (Tucker & Lewis, 1973), also known as the Non-Normed Fit Index or NNFI, and the Normed Fit Index NFI (Bentler & Bonett, 1980). Both the NNFI and the NFI adjust for complexity of the model. Simulation research shows that all these indices still depend on sample size and estimation method (e.g., ML or GLS), with the CFI and the TLI/NNFI showing the best overall performance (Chou & Bentler, 1995; Kaplan, 1995). If the model fits perfectly, these fit indices should have the value 1. Usually, a value of at least 0.90 is required to accept a model, while a value of at least 0.95 is required to judge the model fit as 'good.' However, these are just rules of thumb.

A different approach to model fit is to accept that models are only approximations, and that perfect fit may be too much to ask for. Instead, the problem is to assess how well a given model approximates the true model. This view led to the development of an index called RMSEA, for Root Mean Square Error of Approximation (Browne & Cudeck, 1992). If the approximation is good, the RMSEA should be small. Typically, a RMSEA of less than 0.08 is required (Kline, 2004), with RMSEA less than 0.05 is required to judge the model fit as 'good'. Statistical tests or confidence intervals can be computed to test if the RMSEA is significantly larger than this lower bound.

Given the many possible goodness-of-fit indices, the customary advice is to assess fit by inspecting several fit indices that derive from different principles. Therefore, for the confirmatory factor model for the family data, I have reported the chi-square test, and the fit indices CFI and RMSEA.

A general problem with these goodness-of-fit indices in multilevel SEM is that they apply to the entire model. Therefore, the goodness-of-fit indices reflect both the degree of fit in the within model and in the between model. Since the sample size for the within 'group' is generally the largest, this part of the model dominates the value of the fit indices. Clearly, it makes sense to assess the fit for both parts of the model separately.

Since the within groups sample size is usually much larger than the between groups sample size, we do not lose much information if we model the within groups matrix separately, and interpret the fit indices produced in this analysis separately.

A simple way to obtain goodness-of-fit indices for the between model is to specify a saturated model for the within groups level. The saturated model estimates all covariances between all variables. It has no degrees of freedom, and always fits the data perfectly. As a result, the degree of fit indicated by the goodness-of-fit indices, represents the (lack of) fit of the between model. This is not the best way to assess the fit of the between model, because the perfect fit of the within model also influences the value of the fit index. Fit indices that are mostly sensitive to the degree of fit will show a spuriously good fit, while fit indices that also reflect the parsimony of the model may show a spurious lack of fit.

A better way to indicate the fit of the within and between model separately is to calculate these by hand. Most fit indices are a simple function of the chi-square, sample size  $N$ , and degrees of freedom  $df$ . Some consider only the current model, the target model  $M_t$ , others also consider a baseline model, usually the independence model  $M_I$ . By estimating the independence and the target model for the within matrix, with a saturated model for the between matrix, we can assess how large the contribution to the overall chi-square is for the various within models. In the same way, by estimating the independence and the target model for the between matrix, with a saturated model for the within matrix, we can assess how large the contribution to the overall chi-square is for the various between models. Using this information, we can calculate the most common goodness-of-fit indices. Most SEM software produces the needed information, and the references and formulas are in the user manuals and in the general literature (e.g., Gerbing & Anderson, 1992).

Table 14.7 gives the separate chi-squares, degrees of freedom, and sample sizes for the independence model and the final model for the family intelligence example.

Table 14.7 Chi-squares and  $df$  separate for individual and family level models

	Individual level, Between model saturated		Family level, Within model saturated	
	independence	2 factors	independence	1 factor
chi-square	805.51	6.72	168.88	4.74
Df	30	8	15	9
N	340	340	60	60

The comparative fit index CFI (Bentler, 1990) is given by

$$CFI = 1 - \frac{\chi_t^2 - df_t}{\chi_I^2 - df_I} \quad (14.22)$$

In equation (14.18),  $\chi_t^2$  is the chi-square of the target model,  $\chi_I^2$  is the chi-square for the independence model, and  $df_t$  and  $df_I$  are the degrees of freedom for the target and the independence model. If the difference of the chi-square and the degrees of freedom is negative, it is replaced by zero. So, for example, the CFI for the family level model is given by

$$CFI = 1 - (4.74 - 9) / (168.88 - 15) = 1 - 0 / 153.88 = 1.00 .$$

The Tucker-Lewis index, TLI, which is also known as the Non-Normed Fit Index, NNFI, is given by

$$TLI = \frac{\frac{\chi_t^2}{df_t} - \frac{\chi_I^2}{df_I}}{\frac{\chi_I^2}{df_I} - 1} \quad (14.23)$$



Finally, the Root Mean Square Error of Approximation RMSEA is given by

$$RMSEA = \sqrt{\left(\frac{\chi^2 - df_t}{Ndf_t}\right)} \quad (14.24)$$

where  $N$  is the total sample size. If RMSEA is negative, it is replaced by zero. Using equations 14.18 to 14.20 and the values in Table 14.10, we can calculate the CFI, TLI and RMSEA separately for the within and between models. The results are in Table 14.11

Table 14.11 Fit indices for individual and family level models separately

	Individual level, 2 factors	Family level, 1 factor
<b>CFI</b>	1.00	1.00
<b>TLI</b>	1.01	1.05
<b>RMSEA</b>	0.00	0.00

The goodness-of-fit indices in Table 14.10 all indicate excellent fit, for both the within and between model.

## 14.7 NOTATION AND SOFTWARE

Most modern SEM software includes routines for two-level SEM. Having only two levels may seem an important limitation, but one must appreciate that SEM is an inherently multivariate technique, and multilevel regression is univariate. Consequently, for multivariate analysis or including a measurement model, multilevel regression needs an extra ‘variable’ level, and for longitudinal analysis it needs an ‘occasion’ level. Multilevel SEM does not need this.

Nevertheless, having only two levels can put strong limits on the analysis. At the time, only the software Gllamm (Rabe-Hesketh & Skrondal, 2008) can analyze multiple-level SEM. Using direct estimation as described above, more than two levels can be accommodated, but this is restricted to multivariate normal variables, and the example shows that the estimates and standard errors are not very accurate.

The two-stage approaches are simpler than the general random coefficient model. They are comparable to the multilevel regression model with random variation only for the intercepts. There is no provision for randomly varying slopes (factor loadings and path coefficients). Although it would be possible to include cross-level interactions, introducing interaction variables of any kind in structural equation models is complicated (cf. Bollen, 1989; Marcoulides & Schumacker, 2001). An interesting approach is allowing different within groups covariance matrices in different subsamples, by combining two-level and multigroup models.

When maximum likelihood estimation is used, multilevel SEM can include varying slopes. At the time, only Mplus and Gllamm support this. Muthén and Muthén (1998-2007) have extended the standard path diagram by using a black dot on an arrow in the level-1 model to indicate a random slope. This slope appears in the level-2 model as a latent variable. This is consistent with the use of latent variables for the level-2 intercept variances. This highlights an important link between multilevel regression and multilevel SEM: random coefficients are

latent variables, and many multilevel regression models can also be specified in the SEM context (Curran, 2003; Mehta & Neale, 2005).

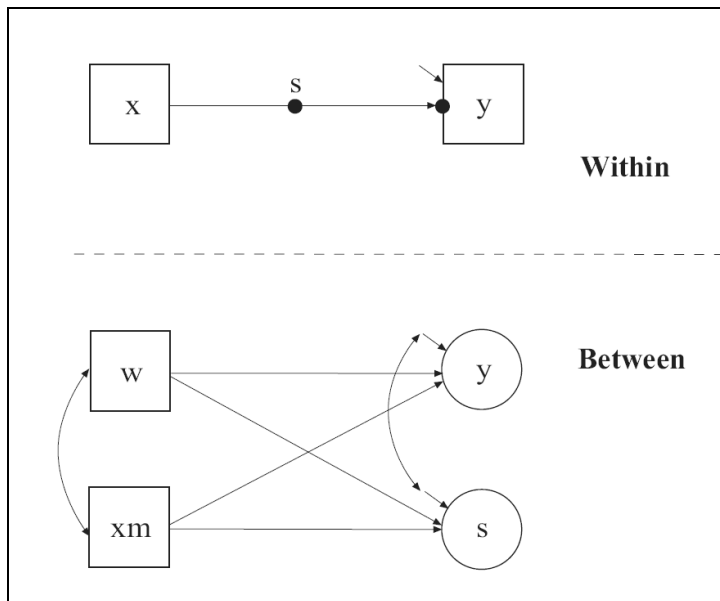


Figure 14.3 Example of path model with random slope and intercept

Figure 14.3 shows an example of a path diagram from the Mplus manual (Muthén & Muthén, 2007, p232). The within model depicts a simple regression of the outcome variable  $y$  on the predictor variable  $x$ . The black dot on  $y$  indicates a random intercept for  $y$ , which is referred to as  $y$  in the between part of the model. The black dot on the arrow from  $x$  to  $y$  indicates a random slope, which is referred to as  $s$  in the between part of the model. In the between part of the model, there are two predictors which are measured only at the between level: the group level variable  $w$  and the group mean on the variable  $x$  which is referred to as  $xm$ .

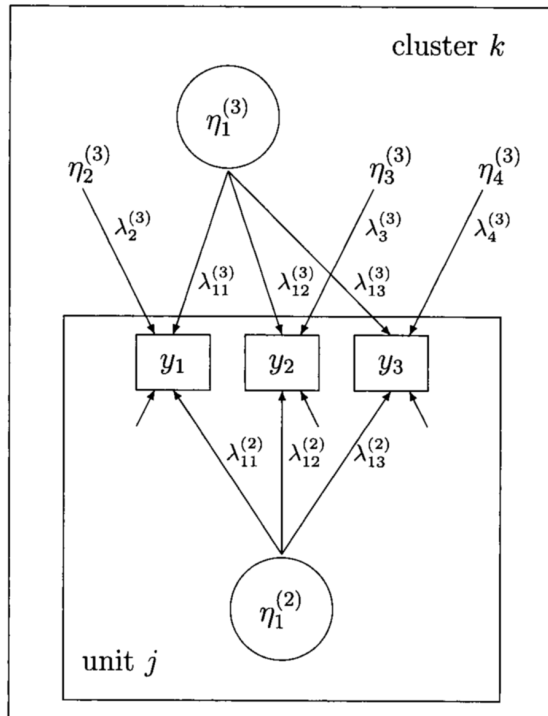


Figure 14.4 Example of multilevel factor model

Figure 14.4 is an example of a multilevel path diagram from Skrondal and Rabe-Hesketh (2004, p104). This is a three level model, with items at the lowest level. The path diagram shows the unit (2<sup>nd</sup>) and cluster (3<sup>rd</sup>) level. This notation is a little more complicated, but is more easily extended to difficult models, e.g. with a partial nesting structure.