Routledge
Taylor & Francis Group

Structural Equation Modeling: A Multidisciplinary Journal

ISSN: 1070-5511 (Print) 1532-8007 (Online) Journal homepage: https://www.tandfonline.com/loi/hsem20

Nesting and Equivalence Testing for Structural Equation Models

Tihomir Asparouhov & Bengt Muthén

To cite this article: Tihomir Asparouhov & Bengt Muthén (2019) Nesting and Equivalence Testing for Structural Equation Models, Structural Equation Modeling: A Multidisciplinary Journal, 26:2, 302-309, DOI: 10.1080/10705511.2018.1513795

To link to this article: https://doi.org/10.1080/10705511.2018.1513795

Structural Equation Modeling: A Multidisciplinary Journal, 26: 302–309, 2019

ISSN: 1070-5511 print / 1532-8007 online
DOI: https://doi.org/10.1080/10705511.2018.1513795

Routledge
Taylor & Francis Group

TEACHER'S CORNER

Nesting and Equivalence Testing for Structural Equation Models

Tihomir Asparouhov and Bengt Muthén
Mplus

In this article, we discuss the nesting and equivalence testing (NET) methodology developed in Bentler and Satorra (2010). We describe how the methodology is implemented in Mplus for the general structural equation model (SEM) model with continuous variables based on the maximum-likelihood (ML) estimation as well as the general SEM model with categorical, censored and continuous dependent variables based on the weighted least squares (WLS) family of estimators. We use the NET methodology to address several model nesting questions that arise in the bi-factor CFA model and the multiple group factor analysis model.

Keywords: bi-factor models, model nesting, multiple group models, structural equation models

INTRODUCTION

The nesting and equivalence testing (NET) methodology developed in Bentler and Satorra (2010) can be used to determine if two structural models are nested or if they are equivalent. In this note, we discuss how the methodology is implemented in Mplus (Muthén & Muthén, 1998-2017) and illustrate it with several examples. The NET testing can be applied to the general SEM model with continuous variables based on the ML estimation as well as the general SEM model for categorical, censored, and continuous dependent variables based on the weighted least squares (WLS) family of estimators. It can be used with multiple group and missing data but it does not extend to multilevel models, time-series models, or mixture models.

Consider first the situation with all continuous variables. Let H_0 be a SEM model with p_0 parameters that is hypothesized to be nested within an H_1 model with p_1 parameters where $p_1 \geq p_0$. The NET methodology can determine if H_0 is nested within H_1 in three steps.

- Step 1. Estimate the H_0 model and compute the model estimated mean $\hat{\mu}_0$ and model estimated variance covariance $\hat{\Sigma}_0$.
- Step 2. Read in $\hat{\mu}_0$ and $\hat{\Sigma}_0$ as data and estimate the H_1 model. Denote by F_0 the fit function value from that estimation.
- Step 3. Using a small value ϵ (e.g. 0.0000001)
  - If $F_0 > \epsilon$, the models are not equivalent or nested.
  - If $F_0 < \epsilon$ and $p_1 > p_0$, the models are nested.
  - If $F_0 < \epsilon$ and $p_1 = p_0$, the models are equivalent.

The outline of this note is as follows. First, we discuss the Mplus implementation of the NET methodology for SEM models with continuous variables in the ML framework. Next, we discuss the Mplus implementation of the NET methodology for SEM models with categorical, censored, and continuous variables in the WLS framework. Finally, we illustrate the methodology with several examples.

THE NET PROCEDURE WITH THE ML ESTIMATOR

The NET implementation in Mplus is as follows. In step 1, the H_0 model estimated mean $\hat{\mu}_0$ and variance/covariance $\hat{\Sigma}_0$ are saved in a file specified with the NESTED option of the

Correspondence should be addressed to Tihomir Asparouhov, Mplus, Salt Lake City, UT, 84108, USA. E-mail: tihomir@statmodel.com

SAVEDATA command. In step 2, the H_1 model is estimated and the NESTED file from Step 1 is specified in the ANALYSIS command. In this step, Mplus minimizes the log-likelihood fit function

$$F(\theta) = \frac{1}{2}(Tr(\Sigma(\theta)^{-1}(\hat{\Sigma}_0 + (\mu(\theta) - \hat{\mu}_0)(\mu(\theta) - \hat{\mu}_0)^T)) + ln(|\Sigma(\theta)|/|\hat{\Sigma}_0|) - p)$$

with respect to the H_1 parameters θ. Here p is the number of variables in the model, $\mu(\theta)$ and $\Sigma(\theta)$ are the H_1 model-implied mean and variance covariance, and $\hat{\mu}_0$ and $\hat{\Sigma}_0$ take the role of the sample mean and variance/covariance. If the H_0 model is nested within the H_1 model the "sample statistics" $\hat{\mu}_0$ and $\hat{\Sigma}_0$ can be matched precisely by the H_1 model estimation and, thus, the NET function value $F_0 = min(F(\theta))$ will be 0. Note that, under the assumption that the sample mean and variance of the data are $\hat{\mu}_0$ and $\hat{\Sigma}_0$ and the sample size is n, the chi-square test statistic for the H_1 model is $2nF_0$ and, thus, will be zero when the models are nested. Note also that $F(\theta) \geq 0$ and therefore, $F_0 \geq 0$.

Because F_0 is computed numerically through the minimization procedure, precise zero values are unrealistic. Instead, small values are used as cutoff values and 10^{-7} appears to work well enough for most situations. If $F_0 < 10^{-7}$ we conclude that the models are nested (if $p_1 > p_0$) or equivalent (if $p_0 = p_1$). If F_0 is between 10^{-6} and 10^{-7} the procedure yields an inconclusive result, i.e. it is not clear if F_0 is zero or not. This inconclusive result can be resolved by sharpening (decreasing) the convergence criterion of the optimization procedure or by utilizing a different data set or a subsample of the existing data set.

The NET methodology in principle does not depend on the data even though the implementation is incorporated within a specific data analysis. It may appear that the data affect the result but that is not the case in most situations and changing the data should not affect the conclusion regarding the nesting of the two models. In certain situations, however, the nesting of the models depends on the parameter estimates. That is, the nesting of the models can change when the model estimates change. The nesting can be different in one part of the parameter space than in another. Because the parameter estimates are affected by the data we can indeed see dependence of the NET methodology on the data but this occurs only through the NET dependence on the parameters.

The NET procedure is prone to some failures due to special conditions that occur in particular data sets. We can illustrate this with the following trivial example. Suppose that the data consists of two identical variables Y_1 and Y_2. Let the H_0 model be the bivariate model where the means are unconstrained, the covariance is restricted to 0 and the variances are restricted to be equal. Let the H_1 model be the bivariate model where the means are constrained to be equal, the covariance is restricted to 0, and the variances are unconstrained. Both models have 3 parameters but are obviously not equivalent. If we perform the NET procedure using the data set where Y_1 and Y_2 are identical, the H_0 model estimated means will be equal and, therefore, the H_1 model will be able to match the H_0 estimated sample statistics. Therefore, the NET procedure will conclude that the two models are equivalent. To avoid such data specific problems, it is recommended that the NET procedure is performed several times over multiple data sets including simulated data.

It is important to note that the nesting of the models does not necessarily imply that the chi-square testing between the models is valid. That is because the models can be nested in such a way that parameters are constrained on the border of the admissible parameter space. The most common example of that is when the nesting is based on a factor variance being fixed to zero (EFA with one factor v.s. EFA with two factors). It is well-known that the log-likelihood ratio test (LRT) in such circumstances deviates from the chi-square distribution, see Self and Liang (1987); Hayashi, Bentler, and Yuan (2007); and Crainiceanu and Ruppert (2004).

THE NET PROCEDURE WITH THE WLS FAMILY OF ESTIMATORS

Bentler and Satorra (2010) point out that the NET approach for evaluating model nesting and equivalence applies to a wide variety of related modeling situations including categorical data modeling such as log-linear models. In this section, we describe how it can be implemented for the SEM modeling framework with categorical/continous/censored variables and the WLS family of estimators.

The WLS family of estimators minimizes the following fit function to obtain the model parameter estimates

$$F(\theta) = \frac{1}{2}(s - \sigma(\theta))^T W^{-1}(s - \sigma(\theta)), \quad (1)$$

see Muthén and Satorra (1995) and Muthén (1998), where s represents a set of sample statistics, estimated from the unconstrained model, and $\sigma(\theta)$ represents the same quantities estimated from a structural model, where θ are the model parameters. In the case of all categorical dependent variables, this unconstrained model is simply the multivariate probit model and s consists of all thresholds and polychoric correlations. In the more general model of the combination of continuous, categorical, and censored variables, the vector s contains the unconstrained model-estimated threshold parameters, sample means, regression coefficients, and unconstrained residual variance covariance matrix with diagonal entries of 1 for all categorical variables. The weight matrix W is different for the different WLS estimators. For the ULSMV estimator this is the identity matrix. For the WLS estimator this is the asymptotic variance covariance matrix of the sample statistics, while for the WLSMV and WLSM estimators that matrix is reduced to its main diagonal.

The NET procedure in the WLS case is implemented as follows.

- Step 1. Estimate the H_0 model and compute the model estimated statistic $\hat{\sigma}_0 = \sigma(\hat{\theta}_0)$ (thresholds, polychoric correlations, etc.), where $\hat{\theta}_0$ are the H_0 model parameter estimates.
- Step 2. Read in $\hat{\sigma}_0$ as data (i.e. as the sample statistics s) and estimate the H_1 model. The weight matrix remains the same as in Step 1. Denote by F_0 the fit function value from that estimation.
- Step 3. Using a small value ϵ (e.g. 0.0000001)
  - If $F_0 > \epsilon$, the models are not equivalent or nested.
  - If $F_0 < \epsilon$ and $p_1 > p_0$, the models are nested.
  - If $F_0 < \epsilon$ and $p_1 = p_0$, the models are equivalent.

In Step 2, the fit function

$$F(\theta) = \frac{1}{2}(\hat{\sigma}_0 - \sigma(\theta))^T W^{-1}(\hat{\sigma}_0 - \sigma(\theta)) \tag{2}$$

is minimized with respect to the H_1 model parameters θ, where $\sigma(\theta)$ is the H_1 model estimated statistics. The conditions of Appendix B in Bentler and Satorra (2010) are all satisfied and, thus, we can apply the NET methodology with the WLS family of estimators.

EXAMPLES

Bentler and Satorra (2010) present several examples and demonstrate the NET methodology using an R interface to EQS (REQS;Mair, Wu, & Bentler, 2010). In this section, we further illustrate the method with several new applications encountered in our research. All model inputs and outputs for these examples can be found at statmodel.com.

Residual correlations

It is well-known that if a CFA model does not fit well due to one non-zero residual correlation, we can augment the CFA model by an additional factor and resolve the problem. The additional factor would be measured by the two variables involved in the non-zero correlation. The question we want to address here is if it is possible to resolve two residual correlations with one additional factor. We use the NET procedure to test this hypothesis. Suppose that we have p indicators measuring a single factor f

$$Y_i = \mu_i + \lambda_i f + \varepsilon_i \tag{3}$$

where $\theta_{ii} = Var(\varepsilon_i)$ and $Var(f)$ is fixed to 1 for identification purposes. We assume that all covariances between the residuals $\theta_{ij} = Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$ with the exception of two such covariances. Suppose that θ_{12} and θ_{34} are not zero. The first model we consider is the above model where θ_{12} and θ_{34} are estimated as free parameters. This model has $3p + 2$ parameters: p intercept parameters μ_i, p loading parameters λ_i, p residual variance parameters θ_{ii}, and the two residual covariance parameters θ_{12} and θ_{34}. Let us call this model M1. The second model we consider is the EFA model with 2 factors. Let us call this model M2. Model M2 has $4p$–1 parameters. The EFA model with two factors requires $p \geq 5$ and so we limit the discussion to that case. When $p \geq 5$, model M2 has more parameters than model M1.

We want to know if by adding a second factor to the one factor model can we fit those two outstanding residual covariances that the one factor model is unable to fit, i.e. we want to test if model M1 is nested within model M2. We check the nesting of M1 and M2 with two examples. The first example uses $p = 5$ and the second example uses $p = 6$. In both cases we generate a data set of size $N = 500$ using model M1 with the following parameter values $\mu_i = 0$, $\theta_{ii} = 1$, $\lambda_i = 1$, $\theta_{12} = \theta_{34} = 0.2$.

Using the NET procedure as implemented in Mplus, when $p = 5$ we obtain the NET function value of 0.00000000 and the procedure concludes that M1 is nested within M2. When $p = 6$ we obtain the NET function value of 0.00383687 and the procedure concludes that M1 is not nested within M2. Therefore, two non-zero residuals can be resolved by adding a second factor but only for the case of $p = 5$. It is not true for larger models. Note also that in the above discussion it is important for the correlations to not have a variable in common, i.e. the case where the two non-zero correlations are θ_{12} and θ_{13} is a different problem.

Bi-factor CFA

Here, we consider the bi-factor CFA models discussed in Reise (2012). The question we want to address here is whether and when a regular factor analysis model is nested within a bi-factor model. Let Y be the vector of dependent variables of size p, and let the variables be grouped in m equal size groups of size l, i.e. $p = ml$. Each of these groups will be used to measure one specific factor. Denote by Y_{ij} the $i-$th variable in the $j-$th group. The groups need not be of the same size in general but we assume this here for simplicity. The first model M1 we are interested in is the standard CFA with m correlated factors

$$Y_{ij} = \mu_{ij} + \lambda_{ij}\eta_j + \varepsilon_{ij}. \tag{4}$$

The factor variances are fixed to 1 but the correlations between the factors are freely estimated. The number of parameters in the model is $q_1 = 3p + m(m-1)/2$. There are p intercept parameters, p loading parameters, and p

residual variance parameters plus the $m(m-1)/2$ factor correlation parameters. We want to compare this model to the following bi-factor model M2,

$$Y_{ij} = \mu_{ij} + \lambda_{ij}\eta_j + \lambda_{0,ij}\eta_0 + \varepsilon_{ij}, \tag{5}$$

where η_0 is the general factor measured by all indicators. In this bi-factor model, the correlations between the m specific factors η_j are fixed to 0. In addition, the correlations between the general factor η_0 and the specific factors η_j are also fixed to 0. The number of parameters in this model is $q_2 = 4p$. There are p intercept parameters, $2p$ loading parameters, and p residual variance parameters. Model M1 has more parameters than model M2 when $m > 2l + 1$. The two models have equal number of parameters when $m = 2l + 1$. Model M2 has more parameters when $m < 2l + 1$. The nesting of these models will now be investigated algebraically and with the NET procedure.

It is fairly easy to see that if $m \leq 3$, the M1 model can be nested within the M2 model. That is because the variance covariance matrix of η_j can be represented by a one factor analysis model. Let us consider first the case of $m = 3$. It is well-known that a three-by-three correlation matrix is equivalent to a one factor model if all three correlations are positive (or two are negative and one is positive which by reversing the sign of one of the factors can be converted to all positive). Denote the correlation between η_i and η_j by ρ_{ij}. Then the unconstrained correlation matrix is equivalent to the factor analysis model

$$\eta_j = \lambda_{0,j}\eta_0 + \xi_j \tag{6}$$

where $Var(\eta_0) = 1$ and

$$\lambda_{0,1} = \sqrt{\frac{\rho_{12}\rho_{13}}{\rho_{23}}} \tag{7}$$

$$\lambda_{0,2} = \sqrt{\frac{\rho_{12}\rho_{23}}{\rho_{13}}} \tag{8}$$

$$\lambda_{0,3} = \sqrt{\frac{\rho_{13}\rho_{23}}{\rho_{12}}}. \tag{9}$$

To keep the variances of ξ_j positive we also need the additional constraints $\lambda_{0,j} < 1$. If this is the case model M1 can be written as

$$Y_{ij} = \mu_{ij} + \lambda_{ij}\xi_j + \lambda_{ij}\lambda_{0,j}\eta_0 + \varepsilon_{ij} \tag{10}$$

which clearly is a model nested within M2, after rescaling ξ_j to have a unit variance (and take the place of the specific factors). Note here that the rescaling requires $Var(\xi_j) > 0$, i.e. $\lambda_{0,j} < 1$ is a requirement for M1 to be nested within M2. This requirement can be reformulated as follows: the product of the two larger correlations should be smaller than the smallest of the three correlations.

Let us illustrate how this analytical argument can also be confirmed with the NET procedure. We generate two data sets one of size $n = 100$ and one of size $n = 300$. We use $l = 3$, $m = 3$, and $p = 9$ in this simulation. Model M1 is used to generate the data where all loadings and residual variances are set to 1 and all means are set to 0. The three factor correlations are set to 0.5, 0.4, and 0.3. Using the smaller data set, we estimate the M1 model and the three factor correlations are estimated to 0.519, 0.206, 0.090. In this case the condition $\lambda_{0,j} < 1$ is violated since the product of the two larger correlations is bigger than the smallest. Using the larger data set of $n = 300$, the three factor correlations are estimated as 0.455, 0.247, 0.255 and the conditions for the nested models are satisfied. This result is also confirmed by the NET procedure. The NET value in the case of $n = 100$ is 0.00008274 and the models are identified as not nested, while in the case of $n = 300$ the value is 0.00000000 and the models are identified as nested. This example illustrates the complexities that can occur in determining if two models are nested. In a large portion of the parameter space when $m = 3$, the M1 model is nested within the M2 model, but in another part of the parameter space, the models are not nested. In our example with $n = 100$, M2 has more parameters than M1 and a higher log-likelihood value. Nevertheless, since the models are not nested, formal chi-square testing should not be performed.

Let us also consider the implications of the second requirement for M1 to be nested in M2, that is, all three correlations need to be positive. We generate a data set with $n = 1000$ using the M1 model and the three correlations 0.2, 0.2, and −0.2. Estimating the M1 model is not a problem at all, however, estimating the M2 model on such a data set is very problematic. The result is non-convergence even with many random starting values. This illustrates how the parameter spaces of M1 and M2 are not aligned at all.

The situation for the case of $m = 2$ and $m = 1$ does not have such complications and in these cases M1 is always nested within M2. For the case of $M \geq 4$ the models are not nested. We illustrate this with another simulation study, using a sample size of $n = 300$, $p = 12$, $m = 4$, and $l = 3$. All means again are set to 0 and all residual variances and loadings are set to 1. All factor correlations are set to 0.5. When we estimate the M1 model the factor correlation estimates are between 0.43 and 0.55. In this case the NET value is 0.00969020 and the procedure concludes that the models are not nested. Here it is interesting to point out that again the M2 model has more parameters and a better log-likelihood value. Since the models are not nested, however, a chi-square test would be invalid. In addition, note that if the M1 model had produced the exact values that we used for data generation purposes the models would have been nested since the equal correlation matrix is also equivalent to a one factor model. That, however, is not the case for the M1 model estimates in this finite population and the correlations were not estimated to the same value. Another interesting perspective is the following. If in the above setup we change just one of the factor correlations to 0.1 instead of 0.5 (making it further away from a one factor model) and we estimate the models M1 and M2, we get a NET value of 0.06604214 and again the models are identified as not nested. But in this case the NET procedure was not needed: the M1 model produced a better log-likelihood value than the M2 model and it has fewer parameters, clearly revealing the fact that the models are not nested.

Multiple group CFA

In this section, we consider the multiple group CFA model and in particular we discuss conditions for the scalar invariance CFA model to be nested within the configural CFA model. This topic has been discussed in Raykov, Marcoulides, and Li (2012) from an interpretive point of view where here we use a statistical approach based on the NET methodology. First, we discuss the models with continuous variables, then we discuss the models for the combination of categorical and continuous variables. Finally, we discuss the special case where a factor is measured by only two indicators and consider the implications that has on the nesting of the models. We analyze the nesting of the models algebraically and with the NET procedure.

Continuous variables

The scalar invariance model, which we refer to as the M1 model, is defined as follows:

$$Y_{ig} = \mu + \Lambda\eta_{ig} + \varepsilon_{ig}. \tag{11}$$

$$\eta_{ig} \sim N(\alpha_g, \Psi_g), \varepsilon_{ig} \sim N(0, \Theta_g), \tag{12}$$

where the indices i and g refer to individual i in group g. For identification purposes, one factor loading in Λ is fixed to 1 for each factor and $\alpha_1 = 0$. The configural CFA model, which we refer to as the M2 model, is defined as follows:

$$Y_{ig} = \mu_g + \Lambda_g\eta_{ig} + \varepsilon_{ig}. \tag{13}$$

$$\eta_{ig} \sim N(0, \Psi_g), \varepsilon_{ig} \sim N(0, \Theta_g). \tag{14}$$

For identification purposes one factor loading in Λ_g is fixed to 1 for each factor. Model M1 is nested within model M2 and that can be seen algebraically as follows. If we add the following constraints to the parameters of model M2 for $g>1$

$$\Lambda_g = \Lambda_1 \tag{15}$$

$$\mu_g = \Lambda_1\alpha_g + \mu_1 \tag{16}$$

the model becomes equivalent to model M1, where Λ_1 takes the role of Λ and μ_1 takes the role of μ. Under the above constraints the log-likelihood values of M2 and M1 would be identical. Therefore, M1 is nested within M2, even though the factor means α_g are fixed to zero in M2 while they are estimated as free parameters in M1. This is an example where the more restricted model has new parameters that are not present in the less restrictive model without compromising the nesting of the models.

We illustrate the nesting of the above models using the NET methodology applied to a two-group two-factor CFA model where each factor is measured by three different indicators and there are no cross-loadings. We generate the data according to model M1 using 100 observations in each group and the following parameter values: the loading parameters are set to 1, the intercept parameters are set to 0, the residual variance parameters are set to 1, the factor variance covariance matrix is set to the identity matrix, the factor means are set to 0 in the first group and to 1 in the second. Using these data, we estimate model M1 and M2 and apply the NET procedure to verify that the models are nested. Model M1 uses scalar invariance across the groups and has a total of 30 parameters: 6 intercept parameters, 4 loading parameters, 6 residual variance parameters in each group, 3 parameters in the factor variance covariance matrix in each group, and 2 factor mean parameters in the second group. Model M2 uses configural invariance across the groups and has 38 parameters, that is, 19 parameters in each group: 6 intercept parameters, 4 loading parameters, 6 residual variance parameters, and 3 parameters in the factor variance covariance matrix. Model M1 has 2 factor mean parameters not present in model M2. Nevertheless, model M1 is nested within model M2. This is confirmed by the NET procedure which produces a NET value of 0.00000000.

Combination of categorical and continuous variables

In this situation the models can be estimated with the WLS family of estimators. Here, we use the "theta" parametrization, see Muthén and Asparouhov (2002), but the conclusions apply to the "delta" parametrization as well. Several changes apply to the above models when categorical variables are involved. First, the above equations apply to the underlying continuous variables Y^*_{ig}. Second the parameters μ and μ_g are zero for every categorical variable. Third, in the M2 model the diagonal entries in Θ_g are fixed to 1 for every categorical variable, while in the M1 model the diagonal entries in Θ_1 are fixed to 1 for every categorical variable while for all other groups these parameters are not fixed to 1 but are free to be estimated. The fourth change is regarding the thresholds for the categorical variables. In the M1 model the thresholds are group invariant τ_{pj} while in the M2 model they are group specific τ_{pjg}, where p refers to the variable the $p-$ th variable in the vector Y_{ig} which we denote by Y_{pig}. Thus, for model M1 we have

$$Y_{pig} = j \Leftrightarrow \tau_{p,j-1} < Y^*_{pig} \leq \tau_{pj} \tag{17}$$

while for model M2

$$Y_{pig} = j \Leftrightarrow \tau_{p,j-1,g} < Y^*_{pig} \leq \tau_{pjg}. \tag{18}$$

We can also express this in probability scale. For model M1

$$\begin{aligned} P(Y_{pig} = j|\eta_{ig}) = \Phi(\frac{\tau_{pj} - \Lambda_p \eta_{ig}}{\sqrt{\theta_{gpp}}}) \\ - \Phi(\frac{\tau_{p,j-1} - \Lambda_p \eta_{ig}}{\sqrt{\theta_{gpp}}}) \end{aligned} \tag{19}$$

and for model M2

$$\begin{aligned} P(Y_{pig} = j|\eta_{ig}) = \Phi(\tau_{pjg} - \Lambda_{gp}\eta_{ig}) \\ - \Phi(\tau_{p,j-1,g} - \Lambda_{pg}\eta_{ig}) \end{aligned} \tag{20}$$

where Λ_p and Λ_{gp} refer to the $p-$ th row of Λ and Λ_g while θ_{gpp} is the $p-$ th diagonal entry of Θ_g for model M1. If the following constraints are imposed on the parameters of model M2 for $g>1$

$$\Lambda_{gp} = \frac{\Lambda_{1p}}{\sqrt{\theta_{gpp}}} \tag{21}$$

$$\tau_{pjg} = \frac{\tau_{pj1}}{\sqrt{\theta_{gpp}}} - \Lambda_{gp}\alpha_g \tag{22}$$

the model becomes equivalent to model M1 where Λ_{1p} takes the role of Λ_p and τ_{pj1} takes the role of τ_{pj}. The parameter constraints for the continuous variables in the model that are needed to reduce model M2 to model M1 are the same as in the previous section. We conclude that for the combination of categorical and continuous variables, model M1 is nested within model M2.

The special case of two-indicator factors

In this section, we limit the discussion to a loading matrix of complexity 1, i.e. there are no cross-loadings and each dependent variable measures exactly one factor. We are particularly interested in the special situation when one of the factors is measured by just two indicator variables. We discuss the issues that occur in this case in regard to the nesting of M1 and M2. We also illustrate the NET methodology with several examples.

Using only two indicators to measure a factor is by no means recommended here. It is of course preferable to have more than two indicators. However, in many practical situations this is not an option for the data analysts. First, note that if the model consists of just one factor, measured by two variables, even in the case of only one group, there is an identification problem. In the continuous case, the number of sufficient statistics is five; two means, two variances, and one covariance. Any SEM model that can be estimated in this context can have no more than five parameters and the above CFA models would have six: two means, two residual variances, one loading, and one factor variance. In the binary case, we have only three degrees of freedom/sufficient statistics which are the three cell probabilities in the joint distribution of the two binary variables. The above CFA model, however, would estimate four parameters: two thresholds, one loading, and one factor variance. In both cases the identifiability problem is often resolved by fixing the second loading to 1 (the first is already fixed to 1). These considerations, however, do not apply to the more general SEM model where there are other factors and other indicator variables. The second loading is no longer unidentified and can be estimated because it reflects information regarding how the second indicator correlates to other variables in the model as compared to the first indicator. We illustrate this in the simulation study below. Very often, however, the identifiability of the second loading, even though it is possible, it is still fairly poor. The model estimation may fail to converge or even if it converges the standard errors of many of the parameter estimates (particularly in the categorical case) maybe so large that it makes inference quite unsatisfactory (poor power). For these reasons, the second loading is often fixed to 1 even if in principle the loading can be identified. This, however, has important implication for the configural v.s. scalar testing. We illustrate that with a simulation study, using the NET procedure.

Consider the following two-group CFA model with two factors. The first factor is measured by five indicators and the second factor is measured by two. The model has a total of seven dependent variables. We generate the data using 10,000 observations for each group. The data is generated from the M1 model where all 7 loadings are set to 1, all intercepts μ are set to 0. The residual variances in the first group are set to 1 and in the second group to 0.8. The factor variances in the first group are set to 1 and in the second group to 1.2. The factor covariance is 0.6 in the first group and 0.4 in the second group. The factor means are set to 0 in the first group and 0.3 in the second group. We estimate models M1 and M2 using this data set. Both models converged and the two-indicator factor did not cause any problems. The NET procedure confirms that the models are nested and the NET value is 0.00000000. Next we estimate the models M1 and M2 where the second loading is also fixed to 1. As expected the models converged as well and the NET value again confirmed that the models are nested.

Let us now repeat this simulation in the context of binary indicator variables. We use the same setup and parameter values. The threshold values we use for data generation purposes are as follows. The thresholds for the five indicators measuring one factor are 0.1,0.2,0.3,0.4,0.5, while the thresholds for the two indicators for the second factor are set to 0.5 and 0.8. Again we estimate the M1 and M2 models using this data set. Both models converged and the two-indicator factor did not cause any problems. The second loading was estimated near its true value. The NET procedure concluded that the two models are nested as we would expect and the NET value is again 0.00000000. Next we estimate the M1 and the M2 models with the second loading also fixed to 1 (which is the true value used for data generation purposes). Again the two models converged, however the NET value in this case is 0.00000704 and the NET procedure concludes surprisingly that the two models are not nested. Model M1 has 26 parameters and a chi-square value of 27.4. Model M2 has 28 parameters and a chi-square value of 21.1. Still the NET procedure implies that the models are NOT nested. This invalidates also the chi-square difference testing. In this situation, the chi-square difference testing is done in Mplus with the DIFFTEST command, when using the WLSMV default estimator. It is important to note here that the DIFFTEST command does not perform a complete check to verify the nesting of the two models. It verifies that the M2 model has more parameters than the M1 model and a smaller fit function value, however, these two conditions are only necessary but not sufficient conditions for nesting. The DIFFTEST command generally assumes that the models are nested. In certain situations, when the models are not nested the DIFFTEST command encounters computational problems due to the violated nesting assumption. However, in some examples, the DIFFTEST command will complete the computation of the chi-square difference testing even when the models are not nested, i.e. will produce an invalid result. Thus, the NET procedure complements the DIFFTEST command and allows us to verify the nesting and avoid such invalid difference testing results.

The above problem with the nesting of the scalar and the configural model is actually quite easy to explain. Suppose that Y_6 and Y_7 are the indicators for the second factor. For model M2 in group 2, the model implies that $Cor(Y_1^*, Y_6^*) = Cor(Y_1^*, Y_7^*)$ because that covariance is channelled through the factor covariance and the loadings as well as the residual variances for Y_6^* and Y_7^* are identical. In model M1, this constraint is not present (i.e. the model is less restricted in that part of the model) because the residual variances are free to be estimated. It is interesting to note that if in the configural model instead of fixing the loadings to 1 in both groups, we fix the loadings to 1 in group 1 but we free the second loading in group 2, the nesting is restored, i.e. in that case the NET procedure concludes that the scalar model is nested within this modified configural model.

The above logic brings to the spotlight an additional concept: the order of the groups. For the original scalar and configural models given in Equations (11–14) the order of the groups is irrelevant. All reordering of the groups yield equivalent models and the same chi-square values. This, however, is no longer the case for the M1 model when two of the loadings are fixed to 1. In group 1 we have the constraint $Cor(Y_1^*, Y_6^*) = Cor(Y_1^*, Y_7^*)$, while in group 2 that constraint is relaxed and, therefore, the group modeling is not equivalent and the group order matters. Indeed, when the group order is reversed we obtain a different chi-square value. In addition, we can use the NET procedure to test if reversing the order of the groups produces equivalent models. Note that the NET procedure in its current implementation requires that the groups come in the same order for the two models that are to be tested. Therefore, to test the equivalence between the two models with reversed group order we have to override the Mplus defaults of factor variance means being fixed to 0 in the first group and the residual variances being fixed to 1 in the first group, i.e. to reverse the order of the groups we have to specify the model for the second group to be the reference group model. Using the NET procedure we obtain the NET value of 0.00000704 and we conclude that the models with reversed order of the groups are not equivalent.

Let us summarize our findings. We showed here that the scalar model is generally nested within the configural model. If, however, non-standard constraints are added in the model, the nesting could be broken. We illustrated this with the example of a two-indicator factor model for categorical indicators where both loadings are fixed to 1. The NET procedure should be utilized with non-standard parameter constraints to verify proper nesting of the models. The multiple group scalar modeling that we discussed above also applies to longitudinal studies where the same measurement model occurs at different time points, see Muthén and Asparouhov (2002). Therefore, the NET procedure should be used to verify model nesting for longitudinal models in the presence of non-standard measurement model constraints.

REFERENCES

Bentler, P. M., & Satorra, A. (2010). Testing model nesting and equivalence. *Psychological Methods*, *5*, 111–123. doi:10.1037/a0019625

Crainiceanu, C., & Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B*, *66*, 165–185. doi:10.1111/j.1467-9868.2004.00438.x

Hayashi, K., Bentler, P. M., & Yuan, K. H. (2007). On the likelihood ratio test for the number of factors in exploratory factor analysis. *Structural Equation Modeling*, *14*, 505–526. doi:10.1080/10705510701301891

Mair, P., Wu, E., & Bentler, P. M. (2010). EQS goes R: Simulations for SEM using the package REQS. *Structural Equation Modeling*, *17*, 333–349. doi:10.1080/10705511003661587

Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. Mplus Web Notes, 4. Retrieved from https://www.statmodel.com/download/webnotes/CatMGLong.pdf doi:10.1044/1059-0889(2002/er01)

Muthén, B., & Satorra, A. (1995). Technical aspects of Muthén's Liscomp approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika*, *60*, 489–503. doi:10.1007/BF02294325

Muthén, B. O. (1998). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Raykov, T., Marcoulides, G. A., & Li, C. H. (2012). Measurement invariance for latent constructs in multiple populations: A critical view and refocus. *Educational and Psychological Measurement*, *72*, 954–974. doi:10.1177/0013164412441607

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*, 667–696. doi:10.1080/00273171.2012.715555

Self, S. G., & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*, 605–610. doi:10.1080/01621459.1987.10478472