

Nesting and Equivalence Testing in Mplus

Tihomir Asparouhov and Bengt Muthén

May 16, 2018

1 Introduction

In this note we describe the Mplus implementation of the nesting and equivalence testing (NET) methodology developed in Bentler and Satorra (2010). We also extend the NET methodology to multiple group models and the general structural equation model for categorical, censored and continuous dependent variables based on the weighted least squares (WLS) family of estimators. The NET technique can also be used with missing data but it does not extend to multilevel models, time-series models or mixture models.

The outline of this note is as follows. In Section 2 we discuss the implementation of the NET methodology for SEM models with continuous variables in the ML framework. In Section 3 we discuss the implementation of the NET methodology for SEM models with categorical and continuous variables in the WLS framework. In Section 4 we illustrate the methodology with several examples.

2 The NET procedure with the ML estimator

The NET implementation in Mplus consists of two steps. First, the proposed null model H0 is estimated. The H0 model-estimated mean μ_0 and variance / covariance Σ_0 are saved in a file specified with the NESTED option of the SAVEDATA command. Second, the H1 model is estimated. To perform the NET procedure, the NESTED file from the H0 model estimation is now specified in the ANALYSIS command. In this step, Mplus analyses the H1

model as if μ_0 and Σ_0 are the sample mean and sample variance/covariance of the data. That is, Mplus minimizes the log-likelihood fit function

$$F(\theta) = \frac{1}{2} \left(\ln(|\Sigma(\theta)|) - \ln(|\Sigma_0|) + p + \right. \quad (1)$$

$$\left. \text{trace} \left(\Sigma(\theta)^{-1} (\Sigma_0 + (\mu(\theta) - \mu_0)(\mu(\theta) - \mu_0)^T) \right) \right) \quad (2)$$

with respect to the H1 model parameters θ . Here $\mu(\theta)$ and $\Sigma(\theta)$ are the H1 model-implied mean and variance covariance for the particular value of H1 parameters θ and p is the number of variables in the model. If the function F is multiplied by $2n$ where n is the sample size, we would obtain the chi-square test statistic (under the assumption that the sample mean and variance of the data are μ_0 and Σ_0). It turns out that if the H1 model is nested above the H0 model the "sample statistics" μ_0 and Σ_0 can be matched precisely by the H1 model estimation, i.e., if the H1 model is indeed nested above the H0 model $F_0 = F(\hat{\theta})$ should be 0, where $\hat{\theta}$ is the value of θ that minimizes (1). The NET function value F_0 is the minimum of F and can be found in the Mplus output.

The NET check consists of verifying that F_0 is 0. Note here that $F(\theta)$ and therefore F_0 are always positive or zero. Because F_0 is computed numerically through the optimization procedure, precise zero values are unrealistic. Instead, small values are used as cutoff values. If $F_0 < 10^{-7}$ we consider such a value close enough to zero and the conclusion of the NET procedure is that the H0 model is nested within the H1 model. If $F_0 > 10^{-6}$ the NET procedure concludes that F_0 is not zero and the two models are not nested. 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 using the CONVERGENCE option of the ANALYSIS command or by utilizing a different data set. Different data sets will produce slightly different values of μ_0 and Σ_0 and that can clarify the inconclusive situation. One of the easiest way to produce a different data set is to take a small subsample from the existing data. For the purposes of the NET procedure this is easily accomplished by specifying the NOBS option in the DATA command. With this option only a subset of the data is used in the analysis. For example, specifying NOBS=50 will reduce the data set to the first 50 observations

from the original data set. To avoid any dependence on the data and to double check the result of the NET procedure one can use the NOBS command to produce different values of μ_0 and Σ_0 . Particularly valuable in this respect are small samples (i.e. small values of NOBS). That is because small samples produce the greatest variability in μ_0 and Σ_0 . Bentler and Satorra (2010) recommend various other procedures to simulate data and model to produce different values of μ_0 and Σ_0 but we have found that simply using real data sets and subsamples of it works sufficiently well. If a real data set is not available at all, simulated data can be used. In the case with simulated data, small sample sizes are preferred as well because that will sufficiently randomize the model parameter estimates. In fact, we recommend using the smallest sample size that does not produce convergence problems when estimating the H0 model in step 1.

The NET methodology in principle does not depend on the data even though the Mplus 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.

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 et al. (2007); and Crainiceanu and Ruppert (2004). In such situations the LRT test statistic distribution deviates from the standard chi-square distribution and special techniques should be applied. The NET methodology does not provide information regarding the proper LRT distribution. That said, the deviations of the LRT test are not substantial. Simply using a lower p-value cutoff such as for example 0.01 instead of 0.05 is sufficient resolution in most situations.

Alternatively, a simulation study can be conducted to determine the correct LRT distribution and p-value as follows. Generate multiple data sets according to the null hypothesis and then analyze the model with the null H0 model and the less restrictive H1 model to obtain the LRT test statistic distribution and from there the correct LRT p-value. This is known as the parameter bootstrap approach to the LRT testing.

The NET procedure is one directional. The two models are not interchangeable. The more restricted model (the H0 model) has to be estimated first. The less restricted model (the H1 model) has to be estimated second. This is in fact the opposite to how the DIFFTEST command works in Mplus. The DIFFTEST option is also one directional but the more restricted model is run after the less restricted model. Mplus will not allow the NET procedure to be conducted in the opposite direction when the second model (which is supposed to be the H1 model) has fewer model parameters than the first model. That said, however, the NET procedure will be conducted if the two models have the same number of parameters. In that case the NET procedure can detect if the two models are equivalent or not. If $F_0 < 10^{-7}$ the procedure concludes that the models are equivalent, if $F_0 > 10^{-6}$ the models are not equivalent and if $10^{-6} \geq F_0 \geq 10^{-7}$ the result is inconclusive.

Note here that the value F_0 does not depend on the sample size n . This makes it quite easy to compare different values of F_0 across different samples and models. The larger the value of F_0 , the larger the discrepancy between the two models. The NET procedure can be used to compare models that have the same number of variables. In addition, the name of the variables should be identical between the two models. The number of variables that are treated as covariates also needs to be identical between the two models. The number of groups in multiple group analysis also has to be identical. In multiple group analysis the sample statistics from each group are saved and analyzed. The groups must appear in the same order in the two models.

Note that with the above implementation the data sets used for the H0 and the H1 model estimation don't need to be the same. Note also that the data set used in the DATA command of the H1 model estimation does not affect the NET function value F_0 , i.e., it doesn't affect the conclusion of the NET test. In the Mplus implementation the H1 model estimation proceeds as usual after the NET function value is computed and the results reported in the output are the usual H1 model estimated results (these are not the $\hat{\theta}$ values that are used for obtaining the NET function value). In some difficult estimation problems, random starting values can be specified using the

STARTS option of the ANALYSIS command. These random starting values apply not just for the H1 model estimation but also for the minimization of the NET function.

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 H0 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 H1 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 nested as one model has equal means and the other model has equal variances. If we perform the NET procedure using the data set where Y_1 and Y_2 are identical, the H0 model estimated means will be equal and therefore the H1 model will be able to match the H0 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. Additional data sets can be obtained using the Mplus montecarlo procedure. The model population parameters used for the data generation can be taken from real data analysis or can also be randomized as in Bentler and Satorra (2010).

3 The NET procedure with the WLS family of estimators

The procedure in this case mirrors the procedure described in the previous section. The WLS family of estimators minimizes the following function to obtain the model parameter estimates

$$F(\theta) = \frac{1}{2}(s - \sigma(\theta))^T W^{-1}(s - \sigma(\theta)), \quad (3)$$

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 the H0 model using the θ parameter values. In the case of all categorical dependent variables, this unconstrained model is simply the multivariate probit model and s consists of all thresholds, regression coefficients and polychoric residual correlations.

In the more general model of the combination of continuous and categorical variables, the vector s contains the unconstrained model-estimated threshold parameters, mean parameters, 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. Suppose that we want to test if a structural model H0 is nested within another structural model H1. First, the H0 model is estimated and the model-estimated sample statistics values $s_0 = \sigma(\hat{\theta}_0)$ (where $\hat{\theta}_0$ are the H0 model parameter estimates) are saved in a file specified with the NESTED option of the SAVE-DATA command. Second, the H1 model is estimated and the NESTED file from the H0 model estimation is now specified in the ANALYSIS command. In this step Mplus analyses the H1 model as if s_0 are the sample statistics of the data. That is, Mplus minimizes the fit function

$$F(\theta) = \frac{1}{2}(s_0 - \sigma(\theta))^T W^{-1}(s_0 - \sigma(\theta)) \quad (4)$$

with respect to the H1 model parameters θ . Here we use the weight matrix W that is used in the regular H1 model estimation. If the H1 model is nested above the H0 model, the "sample statistics" s_0 can be matched precisely by the H1 model estimation, i.e., if the H1 model is indeed nested above the H0 model $F_0 = F(\hat{\theta})$ should be 0, where $\hat{\theta}$ is the value of θ that minimizes (4). Here again the NET check consists of verifying that F_0 is 0 and at this point we proceed as in the previous section.

4 Examples

Bentler and Satorra (2010) present several examples and demonstrate the use of the NET methodology. In this section we further illustrate the method with several interesting applications.

4.1 Residual correlations

It is well known that if a CFA model doesn't 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 will use the NET procedure to test this hypothesis. Suppose that we have p indicators measuring a single factor f

$$Y_i = \mu_i + \lambda_i f + \varepsilon_i \quad (5)$$

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's call this model M1. The second model we consider is the EFA model with 2 factors. Let's 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 we can 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 will 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.

4.2 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

$$Y_{ij} = \mu_{ij} + \lambda_{ij}\eta_j + \varepsilon_{ij}. \quad (6)$$

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

$$Y_{ij} = \mu_{ij} + \lambda_{ij}\eta_j + \lambda_{0,ij}\eta_0 + \varepsilon_{ij}. \quad (7)$$

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's consider first the case of $m = 3$. It is well known that a 3 by 3 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

$$\eta_j = \lambda_{0,j}\eta_0 + \xi_j \quad (8)$$

where $Var(\eta_0) = 1$ and

$$\lambda_{0,1} = \sqrt{\frac{\rho_{12}\rho_{13}}{\rho_{23}}} \quad (9)$$

$$\lambda_{0,2} = \sqrt{\frac{\rho_{12}\rho_{23}}{\rho_{13}}} \quad (10)$$

$$\lambda_{0,3} = \sqrt{\frac{\rho_{13}\rho_{23}}{\rho_{12}}}. \quad (11)$$

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

$$Y_{ij} = \mu_{ij} + \lambda_{ij}\xi_j + \lambda_{ij}\lambda_{0,j}\eta_0 + \varepsilon_{ij} \quad (12)$$

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's 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's 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 will 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.

4.3 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 et al.

(2012) from interpretive point of view where here we use 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.

4.3.1 Continuous variables

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

$$Y_{ig} = \mu + \Lambda\eta_{ig} + \varepsilon_{ig}. \quad (13)$$

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

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 will refer to as the M2 model, is defined as follows

$$Y_{ig} = \mu_g + \Lambda_g\eta_{ig} + \varepsilon_{ig}. \quad (15)$$

$$\eta_{ig} \sim N(0, \Psi_g), \varepsilon_{ig} \sim N(0, \Theta_g). \quad (16)$$

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

$$\Lambda_g = \Lambda_1 \quad (17)$$

$$\mu_g = \Lambda_1\alpha_g + \mu_1 \quad (18)$$

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 M2 is nested above M1, 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 3 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.

4.3.2 Combination of categorical and continuous variables

In this situation the models can be estimated with the WLS family of estimators. Here we will 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

$$Y_{pig} = j \iff \tau_{p,j-1} < Y_{pig}^* \leq \tau_{pj} \quad (19)$$

while for model M2

$$Y_{pig} = j \iff \tau_{p,j-1,g} < Y_{pig}^* \leq \tau_{pjg}. \quad (20)$$

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

$$P(Y_{pig} = j | \eta_{ig}) = \Phi\left(\frac{\tau_{pj} - \Lambda_p \eta_{ig}}{\sqrt{\theta_{gpp}}}\right) - \Phi\left(\frac{\tau_{p,j-1} - \Lambda_p \eta_{ig}}{\sqrt{\theta_{gpp}}}\right) \quad (21)$$

and for model M2

$$P(Y_{pig} = j | \eta_{ig}) = \Phi\left(\tau_{pjg} - \Lambda_{gp} \eta_{ig}\right) - \Phi\left(\tau_{p,j-1,g} - \Lambda_{pg} \eta_{ig}\right) \quad (22)$$

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 $g > 1$

$$\Lambda_{gp} = \frac{\Lambda_{1p}}{\sqrt{\theta_{gpp}}} \quad (23)$$

$$\tau_{pjg} = \frac{\tau_{pj1}}{\sqrt{\theta_{gpp}}} - \Lambda_{gp} \alpha_g \quad (24)$$

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 M2 is nested above model M1.

4.3.3 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 will 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 5, two means, two variances and one covariance. Any SEM model that can be estimated in this context can have no more than 5 parameters and the above CFA models would have 6: two means, two residual variances, one loading and one factor variance. In the binary case we have only 3 degrees of freedom/sufficient statistics which are the 3 cell probabilities in the joint distribution of the two binary variables. The above CFA model however would estimate 4 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 5 indicators and the second factor is measured by 2. The model has a total of 7 dependent variables. We generate the data using 10000 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. 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's 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 5 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 this modified configural model is nested above the scalar 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 (13-16) 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's summarize our findings. We showed here that the configural model is generally nested above the scalar model. If, however, non-standard constraints are added in the model, the nesting could be broken. We illustrated this with the example of two-indicator factor model for categorical indicators when 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

- [1] Bentler, P. M., & Satorra, A. (2010). Testing model nesting and equivalence. *Psychological Methods*, 5, 111-123.
- [2] Crainiceanu, C., and 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.
- [3] 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.
- [4] Muthén, B. O. (1998). *Mplus technical appendices*. Los Angeles: Muthén & Muthén.
- [5] Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: Multiple-group and growth modeling in Mplus. *Mplus Web Notes*, 4. <https://www.statmodel.com/download/webnotes/CatMGLong.pdf>
- [6] Muthén, B. and Satorra, A. (1995) Technical Aspects of Muthens Liscomp Approach to Estimation of Latent Variable Relations With a Comprehensive Measurement Model, *Psychometrika*, 60, 489-503.
- [7] 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.
- [8] Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667-696.
- [9] Self, S.G. and 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.