

9/24/08

Testing Model Nesting and Equivalence*

Peter M. Bentler

University of California, Los Angeles

Albert Satorra

Universitat Pompeu Fabra

*Research supported by grants DA00017 and DA01070 from the National Institute on Drug Abuse and grant SEJ2006-13537 from the Spanish Ministry of Science and Technology.

Testing Model Nesting and Equivalence

Using existing technology, it can be hard or impossible to determine whether two structural equation models that are being considered may be nested. There is also no routine technology for evaluating whether two very different structural models may be equivalent. A simple procedure is proposed that uses model-reproduced moment matrices to evaluate both model nesting and equivalence. Two standard applications are to verify whether or not two proposed models are equivalent, and whether a baseline model used in an incremental fit index is appropriately nested.

Key words: Structural equation models, covariance structure models, nested models, equivalent models.

Testing Model Nesting and Equivalence

The basic ideas of model nesting and model equivalence in structural equation modeling (SEM) are widely known, already discussed in introductory texts (Kline, 2005; Raykov & Marcoulides, 2006). Most SEM practitioners know that fixing one or more free parameters to yield a more restricted model will yield a nested model, and that changing the direction of one or more paths in a simple model may yield a structurally equivalent model. Model nesting is the easier concept. It is facilitated by SEM programs that allow a researcher to run any two models, whose test statistics output can then be used to compute a chi-square difference test. However, the difference test also can be computed when the models are not nested and such a test is meaningless. Although no researcher would do this on purpose, such a meaningless comparison often is made in the context of widely used fit indices such as the comparative fit index (CFI, Bentler, 1990). To compute such an index, a SEM program may automatically generate the standard baseline or null model of uncorrelated variables and compute the fit index by comparing the fit of the current substantive model to that of this null model. However, as noted by Widaman and Thompson (2003), the model of uncorrelated variables often is not a nested subset of the model of interest, so the resulting fit index is inappropriate and meaningless. At a minimum, it is biased. SEM programs, however, provide no information on the appropriateness of the model comparison used in computing incremental fit indices. A simple and automated method for evaluating model nesting would eliminate this problem.

Model equivalence is harder to evaluate, and tends to be overlooked in practice (e.g., Henley, Shook, & Peterson, 2006). Moment equivalent models are those “that, regardless of the data, yield identical (a) implied covariance, correlation, and other moment matrices when fit to the same data, which in turn imply identical (b) residuals and fitted moment matrices, (c) fit functions and chi-square values, and (d)

goodness-of-fit indices based on fit functions and chi-square” (Hershberger, 2006, p. 15). What is especially hard for a practitioner to do is to evaluate whether models are nested or equivalent when both models being compared are highly complicated, for example, they may contain dozens or hundreds of degrees of freedom (df). SEM programs provide no help in evaluating equivalent models, and researchers are left to study the complicated rules of path replacement most recently summarized by Hershberger (2006) to see if such models are equivalent.

We know of no SEM program that provides a researcher with evidence regarding model nesting or equivalence. This note describes simple computations that can be added to any SEM program to achieve these objectives, and that similarly allow a researcher to perform their own nesting and equivalence tests.

We use the following notation. Let two models be designated as M_1 and M_2 , where, if the models have different degrees of freedom (df), M_1 is the model with the larger df. This implies that when models are nested, M_1 is the more restricted model; or, if the df are equal, it may be equivalent to M_2 . When neither nesting or equivalence holds, M_1 and M_2 are just two models of interest.

Methodology

Bentler and Bonett (1980) introduced the concepts of *parameter nesting* and *covariance matrix nesting*. The type of nesting typically considered is parameter nesting, where a free parameter in M_2 is fixed in the more restricted model M_1 or a free parameter is added into a constraint equation that reduces the effective number of free parameters. A result is that the df in M_1 must be larger than those in M_2 . Parameter nesting is relatively easy to evaluate, especially in simple models, and hence a special technology to evaluate parameter nesting is typically not needed. However, when there are dozens or

hundreds of free and fixed parameters in a model, verifying parameter nesting also can be difficult, or at least, subject to human error, and computerized assistance may be helpful.

Covariance matrix nesting, or, allowing mean structures as well, moment matrix nesting, is more difficult to verify. In such models, the mapping from parameters to mean and covariances matrices may take completely different forms in the two models being compared, and yet the set of possible moment matrices under M_1 remains a subset of the set of possible mean and covariance matrices under M_2 .

Consider for example the two models in Figure 1. These models look completely different -- they have a different number of equations; they have a different number of independent variables; and only one has latent variables. Yet, under standard identification conditions, they are nested. This can be seen easily since model M_2 is a saturated model, and M_1 is just some structural model. In more complicated models it can be even more difficult to establish moment matrix nesting. Computerized assistance would definitely be helpful.

Using the idea of moment matrix nesting, equivalent models can be defined as not necessarily nested models with equal degrees of freedom in which the set of possible moment matrices under M_1 is identical to the set of possible covariance matrices under M_2 . We use these ideas to describe our method.

Nesting and Equivalence Test (NET)

Step 1. Do a normal SEM run on model M_1 . Save the df and the model reproduced covariances (and means, in a mean structure). Call these df_1 , $\hat{\Sigma}_1$ (and $\hat{\mu}_1$ if needed).

Step 2. Read in the $\hat{\Sigma}_1$ (and $\hat{\mu}_1$ if needed) as data to be analyzed in a SEM run with model M_2 using the same estimation method. The output needed from this run is the df, say df_2 , and the minimum of the fit function \hat{F} (or the basic chi-square statistic).

Step 3. Compute $d = df_1 - df_2$, and pick a small criterion value for ε , such as $\varepsilon = .001$.

Step 4. (a) If $d > 0$ and $\hat{F} < \varepsilon$, the models are nested.

(b) If $d = 0$ and $\hat{F} < \varepsilon$, the models are equivalent.

(c) If $d < 0$ or $\hat{F} \geq \varepsilon$, M_1 is not nested in, or equivalent to, M_2 .

The key idea is that when $\hat{F} < \varepsilon$, the M_1 model-reproduced means/covariance matrices can be precisely reproduced by model M_2 . With nonnegative d , as in (a), this implies the models are nested. With $d = 0$, as in (b), this implies the models are equivalent. Negative degrees of freedom, as in (c), imply non-nesting of M_1 in M_2 and non-equivalence of the models. Note that if $\hat{F} \geq \varepsilon$, i.e., if the restricted model reproduced means and covariances cannot be fit exactly by an equal or more general model, the models are neither nested or equivalent. The reason for picking some small ε as a criterion is that computations always involve some numerical approximations, terminating computations by a convergence criterion depends on the specific choice of criterion, and so on.

Illustrations

The two models M_1 and M_2 of Figure 1 will be used to illustrate our nesting test. The sample covariance matrix S is the 3 by 3 covariance matrix given by Hershberger (2006, p. 15), namely,

$$S = \begin{bmatrix} 2.00 & & \\ .94 & 1.80 & \\ 1.03 & .65 & 1.50 \end{bmatrix}.$$

Hershberger does not give the sample size, and we take it as $N=100$. Model M_1 is Hershberger's Model 1 (p. 17), which has 1 df ($df_1=1$). In Step 1, we use maximum likelihood (ML) estimation, and obtain the model-reproduced covariance matrix based on the ML estimates as

$$\hat{\Sigma}_1 = \begin{bmatrix} 2.00 & & \\ .94 & 1.80 & \\ .339 & .65 & 1.50 \end{bmatrix}.$$

This is identical to that reported by Hershberger (2nd matrix, top of p.16). Next, we do Step 2, using $\hat{\Sigma}_1$ as data to be analyzed using model M_2 of Figure 1. The output from this run provides the df ($df_2=0$) and the minimum of the ML fit function, $\hat{F} = .00000$. In step 3, we compute the difference in degrees of freedom as $d = df_1 - df_2 = 1 - 0 = 1$, and choose the default $\varepsilon = .001$. We see that the results favor part (a) of Step 4, that is, $d > 0$ and $\hat{F} < \varepsilon$, and hence we conclude that the models are nested.

A more interesting case is that of model equivalence. Figure 2 shows a model that Hershberger has shown to be equivalent to model M_1 of Figure 1; it is his Model 3 (p. 17). We now use our NET procedures to evaluate the equivalence hypothesis. Step 1 gives the same results as previously. Using the output of Step 1 ($\hat{\Sigma}_1$ as shown above) as input to Step 2, we now run the model of Figure 2. The results show the df ($df_2=1$) and the minimum of the ML fit function, $\hat{F} = .00000$. In step 3, difference in df is given as $d = df_1 - df_2 = 1 - 1 = 0$. With the same default ε as before, the results are consistent with part (b) of Step 4. Since $d = 0$ and $\hat{F} < \varepsilon$, the models are equivalent.

Next we consider one of the examples provided by Widaman and Thompson (2003) on appropriate and inappropriate null models for fit indices. They analyzed several psychometric test theory models on 4 variables taken from Votaw (1948). Based on 126 subjects, the means of these variables are 14.905, 15.484, 14.444, 15.123, and the covariance matrix is

$$S = \begin{bmatrix} 25.070 & & & \\ 12.436 & 28.202 & & \\ 11.726 & 9.228 & 22.739 & \\ 20.751 & 11.973 & 12.069 & 21.871 \end{bmatrix}.$$

Among the several models they considered is a restricted one-factor model with equal factor loadings, equal unique variances, and equal means (their model 1A). The ML test statistic T_{ML} of this substantive model, based on $df=11$, is $T_{ML} = 115.266$. Their standard baseline model (model 0C) had no factors, freely estimated unique variances, and freely estimated means. This model is clearly not nested in the substantive model since it frees up parameters rather than restricting them further. It yielded $T_{ML} = 272.492$ with 6 df. In this example, the smaller rather than greater number of df should clearly warn the user about the inappropriateness of this baseline model, and hence the inappropriateness of all incremental fit indices. Widaman and Thompson suggested that an acceptable baseline model for their substantive model would be a model that has no factors, equal unique variances, and equal means (model 0A). This yielded $T_{ML} = 277.826$ with 12 df. We now show how our NET methodology verifies that model 0C is not nested in the substantive model, while model 0A is appropriately nested. We use the χ^2 statistics for this evaluation, rather than the minima of the fit function as before.

Our Step 1 model M_1 for model 0C showed $T_{ML} = 272.492$ with 6 df ($=df_1$). The model reproduced means are 14.905, 15.484, 14.444, and 15.123 and the unique variances are 25.070, 28.202, 22.739, and

21.871, i.e., these estimates correspond to the sample means and variances as would be expected. In Step 2, these are input into a new run with analysis based on M_2 , which is their model 1A (see above). The resulting $T_{ML} = 5.334$ with 11 df (=df₂). In Step 3, we compute $d = df_1 - df_2 = 6 - 11 = -5$ and choose $\varepsilon = .001$ as before. In Step 4, since $d < 0$ and $T_{ML} = 5.334 > \varepsilon$, neither conditions (a) or (b) are met, while (c) is met, attesting to the conclusion that the models are not nested as hypothesized or equivalent. Repeating this procedure for model 0A, in Step 1, as noted $T_{ML} = 277.826$ with 12 df (=df₁). The corresponding model reproduced means are 14.989, 14.989, 14.989, and 14.989, and the unique variances are 24.612, 24.612, 24.612, and 24.612. In Step 2, these are input into a new run with analysis based on model M_2 (model 1A), yielding $T_{ML} = 0.000$ with 11 df (=df₁). In Step 3, $d = df_1 - df_2 = 12 - 11 = 1$, and we choose $\varepsilon = .001$. Using the results of Steps 2 and 3 in Step 4, we have $d > 0$ and $T_{ML} < \varepsilon$, and hence we conclude, correctly, that model 0A is nested in model 1A. Hence, incremental fit indices can be based on it.

Rationale for NET Methodology

Consider a situation in which the set of possible mean and covariance matrices under M_1 is equal to, or is a subset of, the set of possible mean and covariance matrices under M_2 . This moment matrix nesting requires that all possible sets of parameters θ_1 that could generate means $\mu_1(\theta_1)$ and covariance matrices $\Sigma_1(\theta_1)$ must generate sets of means and covariances that also are found among the means $\mu_2(\theta_2)$ and covariance matrices $\Sigma_2(\theta_2)$. As a consequence, fitting the more general model $\{\mu_2(\theta_2), \Sigma_2(\theta_2)\}$ to the restricted model $\{\mu_1(\theta_1), \Sigma_1(\theta_1)\}$ must yield a perfect fit. In practice, consistent estimators $\{\mu_1(\hat{\theta}_1), \Sigma_1(\hat{\theta}_1)\}$ are fit by the model $\{\mu_2(\theta_2), \Sigma_2(\theta_2)\}$ according to some estimation methodology. If the fit is perfect, the models are nested, and furthermore, if both models have the

same degrees of freedom, they are equivalent. The technical basis for this methodology is given in the Appendix.

Discussion

In principle, the approach outlined here for evaluating model nesting and equivalence applies to a wide variety of related modeling situations such as multiple group models or higher moment structures in SEM, log-linear models in categorical data situations, and so on. Since incremental fit indices are almost universally reported in SEM, one of the simplest yet most important areas of application is that of verifying the nestedness of a baseline or null model for the computation of incremental fit indices. Although Bentler and Bonett (1980) had emphasized that an appropriate null model for such indices must be a nested model relative to the substantive model of interest, they also noted that “there are cases when it is not a simple matter to verify that M_0 [the null model] is a special case of M_k , M_l [substantive models], or M_s [the saturated model]. Nonetheless, this is a fundamental requirement of M_0 ” (p. 596). Such a difficulty may be inevitable when the researcher uses prior results (Sobel & Bohrnstedt, 1985), an equal-correlation baseline model (Rigdon, 1998), or some untested ideas to specify the null model. The only standard default used in SEM programs that we know about is the uncorrelated variables model as the baseline model, yet, as Widaman and Thompson (2003) show, this is often an inappropriate model choice since this model may not be nested in the model of interest. The methodology provided here can be easily implemented to routinely evaluate nesting for incremental fit indices, especially when the null model is provided by a program default.

Of course, difficulties can arise in particular applications. In the context of basic SEM models considered here, an especially important situation is that of missing data. The approach developed here is difficult to implement with the popular direct maximum likelihood method. However, the two-stage maximum likelihood method described by Yuan and Bentler (2000) and Savalei and Bentler (2007) immediately

allows applying the proposed NET framework. In the two-stage approach, a saturated model (sometimes called the EM means and covariance matrix) is first estimated. These saturated means and covariances are then taken as the sample means and covariances to be modeled in the subsequent step. For the NET procedure, the saturated moments are taken as the data to be analyzed in Step 1 above, and model M_1 is fit to it. Then Steps 2-4 are completed as usual.

A more general issue is how to apply the proposed NET procedure when there are no data at all, and hence no sample covariance matrix S (and means, when relevant) is available for Step 1 of our method. When not having S from empirical data, it is possible to substitute an S obtained from simulated data, e.g., data from a hypothetical normal population with a given covariance matrix. In particular, a population specification from M_1 is used to generate the sample moment matrix S_1 which is then fitted under M_1 to obtain the fitted moment $\hat{\Sigma}_1$. Then $\hat{\Sigma}_1$ is analyzed under model M_2 to obtain \hat{F} . If $\hat{F} < \varepsilon$ and $d > 0$, then the models are nested; or if $\hat{F} < \varepsilon$ and $d = 0$, the models are equivalent.

With regard to model equivalence, we should note that our goals have been modest. We proposed a way to evaluate whether any two models that a researcher nominates might be equivalent moment structure models. Making such a comparison is critical to ruling out potentially competing explanations of a phenomenon (MacCallum, Wegener, Uchino, & Fabrigar, 1993; Stelzl, 1986). However, our methods do not address the more complicated problem of starting with only one model and generating an entire class of models that might be equivalent to a given model. Rules for specifying some equivalent models have been developing for over two decades since Stelzl (1986) first pointed out that alternative causal hypotheses could yield identical indices of model fit (e.g., Lee & Hershberger, 1990; Luijben, 1991; Hershberger, 1994, 2006). Since there may be a lot of equivalent models (Raykov & Marcoulides, 2001, 2007), a computer algebra surely will need to be incorporated into SEM programs to help the researcher generate and evaluate such candidates. Our methods also do not address the

complicated issue of whether models might be equivalent in the broader sense of observational equivalence, implying that their generating probability distributions are equivalent. This topic may require studying individual case scores and residuals (e.g., Raykov & Penev, 2001).

References

- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238-246.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588-606.
- Henley, A. B., Shook, C. L., & Peterson, M. (2006). The presence of equivalent models in strategic management research using structural equation modeling: Assessing and addressing the problem. *Organizational Research Methods*, *9*, 516-535.
- Hershberger, S. L. (1994). The specification of equivalent models before the collection of data. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 68-108). Thousand Oaks, CA: Sage.
- Hershberger, S. L. (2006). The problem of equivalent structural models. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 13-41). Greenwich, CN: Information Age.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford.
- Lee, S., & Hershberger, S. (1990). A simple rule for generating equivalent models in covariance structure modeling. *Multivariate Behavioral Research*, *25*, 313-334.
- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185-199.
- Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling*. Mahwah, NJ: Erlbaum.
- Raykov, T., & Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling*, *8*, 142-149.
- Raykov, T., & Marcoulides, G. A. (2007). Equivalent structural equation models: A challenge and responsibility. *Structural Equation Modeling*, *14*, 695-700.

- Raykov, T., & Penev, S. (2001). The problem of equivalent structural equation models: An individual residual perspective. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 297-322). Mahwah, NJ: Erlbaum.
- Rigdon, E. E. (1998). The equal correlation baseline model for comparative fit assessment in structural equation modeling. *Structural Equation Modeling, 5*, 63-77.
- Sard, A. (1942). The measure of the critical values of differentiable maps. *Bulletin of the American Mathematical Society, 48*(12), 883-890.
- Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*, 131-151.
- Savalei, V., & Bentler, P. M. (2007). A two-stage approach to missing data: Theory and application to auxiliary variables. *UCLA Statistics Preprint #511*. <http://preprints.stat.ucla.edu/>
- Sobel, M. E., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 152-178). San Francisco: Jossey Bass.
- Votaw, D. F., Jr. (1948). Testing compound symmetry in a normal multivariate distribution. *Annals of Mathematical Statistics, 19*, 447-473.
- Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods, 8*, 16-37.
- Yuan, K.-H., & Bentler, P.M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology, 165-200*.

Appendix

Let $\sigma(\theta) \equiv [\mu', \text{vech}\{\Sigma(\theta)'\}]'$ and consider two moment structural models $M_1: \sigma = \sigma_1(\theta_1)$, $\theta_1 \in \Theta_1$, and $M_2: \sigma = \sigma_2(\theta_2)$, $\theta_2 \in \Theta_2$ defined on the space of unconstrained moments $C = \{\sigma \mid \sigma \in \Theta\}$, with Θ_1 , Θ_2 , and Θ Euclidean spaces, and $\sigma_1(\cdot)$ and $\sigma_2(\cdot)$ continuously differentiable functions. Consider the subsets of C associated to the two models $C_1 = \{\sigma \mid \sigma = \sigma_1(\theta_1), \theta_1 \in \Theta_1\}$ and $C_2 = \{\sigma \mid \sigma = \sigma_2(\theta_2), \theta_2 \in \Theta_2\}$, and the intersection $C_{12} = C_1 \cap C_2$. Then locally for any moment structure model, the following result applies (see, e.g., Satorra, 1989):

Result 1. Locally around a point $\sigma \in C$, for a moment structure model, M_j , the set of moments satisfying the model C_j can be defined as $C_j = \{\sigma \mid h_j(\sigma) = 0\}$ for a continuously differentiable (vector-valued) function $h_j(\cdot)$.

Now we consider a sample moment s that has an absolutely continuous distribution in the set C .¹ When fitting s using models M_1 , we obtain the fitted moments $\hat{\sigma}_1$. When fitting $\hat{\sigma}_1$ using models M_2 the minimum of the fitting function is denoted F_{12} ; we could equally consider nF_{12} , where n is sample size. With these definitions, we have the following.

Theorem. Let s be the vector of sample moments and F_{12} be the associated sample value of Step 3 of NET. Then, one, and only one, of the following outcomes apply:

- $C_{12} = C_1$, i.e. $C_1 \subset C_2$ ($M_1 < M_2$) and $F_{12} = 0$
- C_{12} is not equal to C_1 and $F_{12} > 0$ with probability 1.

Proof: Combining Result 1 with Sard's (1942) Theorem,² it follows that either 1) $C_{12} = C_1$, in which case

¹That the distribution is absolutely continuous is essential, but is not restrictive, since it encompasses SEM except for the case of degenerate distributions. It excludes the case of discrete distributions.

² Let $f: R^n \rightarrow R^m$, f continuously differentiable, with the jacobian matrix of f having rank $< m$, then the set

$F_{12} = 0$; or 2) C_{12} has Lebesgue measure zero in C_1 . Since for any absolutely continuous distribution, sets of Lebesgue measure zero have zero probability, it follows that when $C_{12} \neq C_1$ the probability that $\sigma_1 \in C_{12}$ is zero, so $F_{12} > 0$ with probability 1.³

This result suggests the procedure of the Nesting and Equivalence Test given in the text, where F_{12} is designated as \hat{F} , concluding that $M_1 < M_2$ when $F_{12} < \varepsilon$, for ε a small constant; model equivalence when in addition to $F_{12} < \varepsilon$, the two models have the same number of degrees of freedom ($d = 0$); or non-nesting and non-equivalence of M_1 in M_2 when $F_{12} > \varepsilon$.

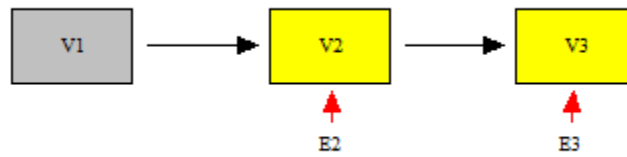
$\{x | f(x) = 0\}$ has measure zero in R^m .

³A graph can help on this. The graph would show a square which is model C_1 and a curved line that is C_{12} , and a point outside the line which represents $\hat{\sigma}_1$. In this graph, the Lebesgue measure of C_{12} is zero, unless C_{12} is the whole space C_1 , when $C_1 \subset C_2$.

Figure 1

Two Nested Models

Model M_1



Model M_2

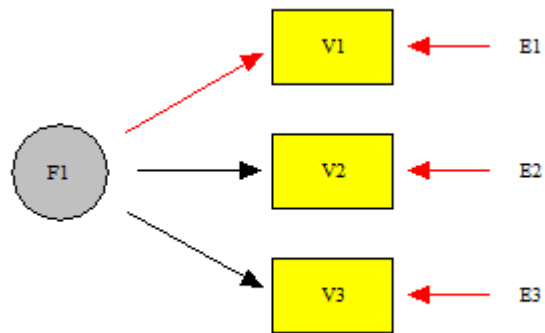


Figure 2

An Equivalent Model to Model M_1 of Figure 1

