

On insensitivity of the chi-square model test to
non-linear misspecification in structural equation
models *

Ab Mooijaart
Leiden University, Leiden
&
Albert Satorra
Universitat Pompeu Fabra. Barcelona

October 4, 2008

*Research of the second author is supported by grant SEJ2006-13537 from the Spanish Ministry of Science and Technology.

Abstract

In this note we show that for some structural equation models (SEM), the classical chi-square goodness-of-fit test is unable to detect the presence of interaction (non-linear) terms in the model. Not only the model test has zero power against that type of misspecifications, but even the theoretical (chi-square) distribution of the test is not distorted when severe interaction term misspecification is present in the postulated model. We explain this phenomenon by exploiting results on asymptotic robustness (AR) in structural equation models. The importance of this paper is to warn against the conclusion that if a proposed linear model fits the data well according to the chi-square goodness-of-fit test, then the underlying model is linear indeed; it will be shown that the underlying model may in fact be severely nonlinear. In addition, the present paper shows that such insensitivity to interaction terms is only a particular instance of a more general problem, namely, the incapacity of the classical chi-square goodness-of-fit test to detect deviations from zero correlation among exogenous regressors (either being them observable, or latent) when the structural part of the model is just saturated.

KEYWORDS: structural equation modeling, testing model fit, nonlinear relations, interaction terms, equivalent models, asymptotic robustness, saturated model

Introduction

In this paper we deal with nonlinear relationship between the latent variables. Such nonlinearity may come from the existence of interaction or quadratic factors. It will be shown that standard chi-square goodness-of-fit test are not always capable of indicating that the underlying model is a nonlinear model. An explanation for this phenomenon will be given. The theory of asymptotic robustness (AR) (as described in Satorra 2002, and references therein) will be used to explain this insensitivity of the model test to non-linear terms misspecification. A key assumption for applying the theory of AR is that the variance matrix of non-normal constituents of the model is unrestricted by the analyzed model. Even though this condition does not hold for the model under investigation, a reparameterization will be given under which conditions for AR are fulfilled and consequently asymptotic robustness do hold for the considered model.

The structure of the paper is as follows. Next section illustrates the issue to be discussed using a small Monte Carlo study. The Kenny-Juddy's (1984) model will be used in the Monte Carlo set-up. It will be shown that in a factor model in which one of the factors is a product of two other factors, the normal theory (NT) chi-square goodness-of-fit test is still chi-square distributed despite the interaction term and non-normality of the observed variables. In Section 3 we recast a general model set-up for which the Kenny-Juddy's (1984) model is a special case, and we develop conditions under which results of AR apply on that general model set-up. In Section 4 we show that interaction terms misspecification is just a particular instance of a more general type of misspecification that is undetected by the chi-square model test when certain conditions on the model apply. Finally, Section 5 we discuss what the alternatives are if a researcher wants to take seriously the possibility that some of the factors are nonlinear.

Monte Carlo evidence

In this section a small Monte Carlo study will be used to highlight the issue to be discussed in the paper.

Consider a population for which the following regression equation holds

$$y = \bar{\beta}_0 + \bar{\beta}_1\xi_1 + \bar{\beta}_2\xi_2 + \bar{\beta}_{12}\xi_1\xi_2 + \zeta \quad (1)$$

where y is an observed variable, ξ_1 and ξ_2 are latent regressors, $\bar{\beta}_{12}\xi_1\xi_2$ is an interaction term, and ζ is the disturbance term. The $\bar{\beta}$ s are parameters with specific values, and ξ_1 , ξ_2 and ζ have a trivariate joint distribution. Let us assume that ζ is independent of the ξ s.

Consider also observable indicators for the latent regressors ξ_1 and ξ_2 , with each factor having two indicators; so, in addition to the regression equation (1), the following

measurement equations are specified

$$x_j = \bar{\alpha}_j + \bar{\lambda}_j \xi_1 + \delta_j \quad j = 1, 2 \quad (2)$$

$$x_j = \bar{\alpha}_j + \bar{\lambda}_j \xi_2 + \delta_j \quad j = 3, 4 \quad (3)$$

where δ_j are measurement errors. Assume the measurement errors δ s are independent of the ζ and of the ξ s. The $\bar{\alpha}$ s and $\bar{\lambda}$ s are parameters with specific values. For the moment, no assumption is made about the statistical distribution of the random variables in the model except for the described independence among sets of variables.

In the last two decades, model (1) to (3) has been studied extensively, in particular after the seminal paper by Kenny and Judd (1984), and the seminal discussion of Kenny and Judd's model by Jöreskog and Yang (1996).

In the Monte Carlo study considered here, a model \mathcal{M}_0 is specified where instead of the non-linear equation (1), we specify the linear equation

$$\mathcal{M}_0 : \quad y = \beta_0 + \beta_1 \xi_1 + \beta_2 \xi_2 + \zeta \quad (4)$$

with the β s being now parameters to be estimated. We also consider the equations of (2) and (3) with the $\bar{\alpha}$ s and $\bar{\lambda}$ s substituted by free parameter to be estimated, α and λ s respectively. For purposes of model identification, for each ξ one of the λ s is set to 1. Model \mathcal{M}_0 specifies as unrestricted parameters the variances and covariances corresponding to random constituents of the model (i.e., the variances and covariances of the ξ s, as well as the variances of ζ and of the δ s). To make the specification of the model more fully, let us even assume that \mathcal{M}_0 specifies normality for the distribution of the vector of observed variables (this to be called the NT assumption).

Note that the specification \mathcal{M}_0 ignores the interaction term $\bar{\beta}_{12} \xi_1 \xi_2$, whatever the magnitude of $\bar{\beta}$, as well as the non-normality of the distribution of y induced by the product term $\xi_1 \xi_2$ of the data generating process of (1).

The basic question we investigate in this paper is the following: when analyzing \mathcal{M}_0 using data that comes from (1) to (3) with $\beta_{12} \neq 0$, do we get any indication of misspecification of the model? Note that such an analysis can be done by the regular normal theory (NT) analysis since normality of observable variables are assumed in \mathcal{M}_0 . More specifically, we ask ourselves whether the NT chi-square model test of \mathcal{M}_0 , i.e. the likelihood ratio test (LRT) of \mathcal{M}_0 against a model that sets the mean vector and covariance matrix unconstrained, has the capacity to detect the presence of the interaction term $\bar{\beta}_{12} \xi_1 \xi_2$. This is of practical importance, because if there is not such an indication a researcher may tend to adopt the incorrect conclusion that there is just a linear relationship between the variables, and conclude that \mathcal{M}_0 is the true model.

To make the investigation more concrete, consider the population of equations (1) to (3) with the same parameter values as in Jöreskog and Yang (1996); that is, $\bar{\alpha}_1 = \dots = \bar{\alpha}_k = 0$, $\bar{\beta}_0 = 1$, $\bar{\beta}_1 = .2$, $\bar{\beta}_2 = .4$, $\bar{\beta}_{12} = .7$, $\bar{\lambda}_1 = 1$, $\bar{\lambda}_2 = .6$, $\bar{\lambda}_3 = 1$, $\bar{\lambda}_4 = .7$, $\bar{\psi} = .2$, $\bar{\phi}_{11} = .49$, $\bar{\phi}_{12} = .2352$, $\bar{\phi}_{22} = .64$, $\bar{\theta}_{\delta_1} = .51$, $\bar{\theta}_{\delta_2} = .64$, $\bar{\theta}_{\delta_3} = .36$, $\bar{\theta}_{\delta_4} = .51$, where we used the

notation $\bar{\psi} = \text{var}(\zeta)$, $\bar{\phi}_{ij} = \text{cov}(\xi_1, \xi_2)$ and $\bar{\theta}_{\delta_k} = \text{var}(\delta_k)$ for the corresponding variances and covariances. Note that in the population considered there is a substantial interaction term. The Monte Carlo illustration considers replicating 1000 times iid sampling, sample size $n = 500$, from this population.

For each of the 1000 replications, we computed the corresponding value of the chi-square model test (the LRT). This produces 1000 replicates of the LRT which distribution can be inspected. This whole process is carried out under 5 different data generating process (DGP) that just differed in the distribution of the ξ s: DGP 1 – 4 set ξ_1 and ξ_2 to be a chi-square distribution with degrees of freedom 1, 2, 5 and 10, respectively, centered to have zero mean; DGP 5 sampled the ξ s from a normal distribution. In all the cases, the ζ and the ϵ s were iid sampled from a Normal distribution. So DGP 1 – 4 do have skewed latent variables (DGP 1 is the most skewed), while the DGP 5 has Normal distribution latent variables.¹

Table 1 reports the mean and variance of the Monte Carlo distribution of the LRT for the different DGPs conditions.

Table 1: Monte Carlo distribution of the model test statistic

Distribution of LRT	DGP (varying the distribution of the ξ s)				
	1	2	3	4	5
mean	3.042	2.997	2.995	3.027	3.115
variance	5.755	6.391	6.371	6.299	6.041

From Table 1 we see that, although a) the model \mathcal{M}_0 is seriously misspecified (it ignores the presence of an interaction term of a substantive size) and b) the model \mathcal{M}_0 assumes normality (NT) when the data is skewed (even for DGP 5), the NT LRT has a mean and variance close to what should be expected if the test statistic were chi-square distributed. For the specified \mathcal{M}_0 , the LRT has 3 degrees of freedom (df), so were the distribution LRT chi-square, the mean and variance reported in Table 1 should be approximately 3 and 6 respectively (the theoretical values of mean and variance for a χ_3^2). This is certainly the case if we look at the numbers in the table. Inspection of the qq-plot for the fit of the empirical distribution to a χ_3^2 in the worst case scenario (the most skewed data) DGP 1, shows also an accurate fit to a χ_3^2 (points lying close to the diagonal line).

Thus, the interaction term misspecification do not seem to induce distortion of the null distribution of the LRT. From this small Monte Carlo study we thus conclude that the LRT (for all the DGP conditions) has the same distribution, χ_3^2 , as when no misspecification is present in the model, including the assumption of Normal distribution for observable variables. This is disturbing since we would like the model test to be sensi-

¹Note that even for DGP 5, that samples from a normal distribution, the distribution of y will be non-normal, due to the interaction term present in the equation (1) of the data generating process.

tive to the severe misspecifications implicit in model \mathcal{M}_0 , such as ignoring an interaction term of substantive size.

This surprising result of lack of sensitivity of the chi-square model test to substantial interaction terms in the model, will now be explained using results of the theory of asymptotic robustness of structural equation models. This will be done in a general context of structural equation models in which the the Kenny and Judd’s model, of our Monte Carlo study, is a special case.

Models and assumptions

In this section we develop analytic conditions under which the phenomenon reported for the Kenny and Judd’s model necessarily extends to general model set-up when certain conditions on the model hold. For that we will apply results of the theory of asymptotic robustness (AR). For the sake of completeness, results of AR to be used are reproduced in Appendix A.

General model set-up

Consider \mathcal{M}_0 to be the general LISREL model with mean structures (see Jöreskog and Sörbom, 1984)

$$\eta = \alpha + B_0\eta + \Gamma\xi + \zeta \quad (5)$$

$$y = \tau_y + \Lambda_y\eta + \epsilon \quad (6)$$

$$x = \tau_x + \Lambda_x\xi + \delta \quad (7)$$

where α , τ_y and τ_x are vectors of constant intercept terms, and B , Γ , Λ_y and Λ_x are matrices of parameters. Terminology in econometric would call (5) a “simultaneous equation model”; SEM terminology refers to this equation as the “structural” part of the model; equations (6) and (7) are referred as “measurement” (or factor analytic) part of model.

We assume that $B = (I_m - B_0)$ is invertible, ϵ is uncorrelated with η , and δ is uncorrelated with ξ . Typically, it is assumed that ζ is uncorrelated with ξ , so no new parameter matrix need to be introduced regarding the covariance matrix $\Phi_{\xi\zeta}$ between ξ and ζ . Unless it is said the contrary, \mathcal{M}_0 assumes that $\Phi_{\xi\zeta} = 0$.² We also assume that ζ , ϵ and δ have mean zero. The mean of ξ is a vector of parameters denoted as κ . The variance matrices of ξ , ζ , ϵ and δ are denoted respectively as Φ_ξ , Φ_ζ , Φ_ϵ and Φ_δ . Under a specific model \mathcal{M}_0 , the above vectors and matrices of parameters are expressed as a function of the vector θ of parameters of the model. Define $\Pi = B^{-1}\Gamma$, the matrix of

²Below we see that deviations from that assumption are in fact untestable within the classical SEM framework, unless we introduce sufficient restrictions on Γ .

“reduced form” coefficients of the “simultaneous equation” part of the model. For the results to be developed, we require the model \mathcal{M}_0 satisfies the following condition below.

CONDITION U: *The matrices Π , Φ_ξ and $B^{-1}\Phi_\zeta B^{-T}$ are unrestricted, except for Φ_ξ and Φ_ζ to be symmetric and positive definite.*

The above model \mathcal{M}_0 can be re-written in the following vector equation form

$$z = \mathcal{A}_0 + \begin{pmatrix} \Lambda_y B^{-1} \Gamma \\ \Lambda_x \end{pmatrix} \xi + \begin{pmatrix} \Lambda_y B^{-1} \\ 0 \end{pmatrix} \zeta + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix} \quad (8)$$

where z is the vector $z = (y', x)'$ of observed variables and

$$\mathcal{A}_0 = \begin{pmatrix} \tau_y + \Lambda_y B^{-1} \alpha \\ \tau_x \end{pmatrix}.$$

Associated to \mathcal{M}_0 there is a specific structure (i.e., vector- and matrix-valued functions of a parameter θ) for \mathcal{A}_0 , κ , B , Γ , and the variance matrices Φ_ξ , Φ_ζ , Φ_ϵ and Φ_δ .

Re-write (8) as

$$z = \mathcal{A}_0 + \begin{pmatrix} \Lambda_y & 0 \\ 0 & \Lambda_x \end{pmatrix} \begin{pmatrix} B^{-1} \Gamma & B^{-1} \\ I & 0 \end{pmatrix} \begin{pmatrix} \xi \\ \zeta \end{pmatrix} + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix}$$

so that

$$z = \mathcal{A}_0 + \mathcal{A}_1 \begin{pmatrix} \Pi \xi + B^{-1} \zeta \\ \xi \end{pmatrix} + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix}$$

with

$$\mathcal{A}_1 = \begin{pmatrix} \Lambda_y & 0 \\ 0 & \Lambda_x \end{pmatrix}.$$

Set $\xi^* = \Pi \xi + B^{-1} \zeta$ ($= \eta - B^{-1} \alpha$), and consider the alternative factor-analysis representation

$$\mathcal{M}_0^* \quad z = \mathcal{A}_0 + \mathcal{A}_1 v + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix}. \quad (9)$$

where

$$v = \begin{pmatrix} \xi^* \\ \xi \end{pmatrix}.$$

Letting

$$\Phi_v (= \text{cov}(v)) = \begin{pmatrix} \Phi_{11}^* & \Phi_{12}^* \\ \Phi_{21}^* & \Phi_{22}^* \end{pmatrix},$$

we obtain

$$\begin{aligned} \Phi_v &= \begin{pmatrix} \Pi & B^{-1} \\ I & 0 \end{pmatrix} \begin{pmatrix} \Phi_\xi & \Phi_{\xi\zeta} \\ \Phi_{\zeta\xi} & \Phi_\zeta \end{pmatrix} \begin{pmatrix} \Pi & B^{-1} \\ I & 0 \end{pmatrix}' \\ &= \begin{pmatrix} \Pi \Phi_\xi \Pi' + B^{-1} \Phi_\zeta B^{-T} + \Pi \Phi_{\xi\zeta} B^{-T} + B^{-1} \Phi_{\zeta\xi} \Pi' & \Pi \Phi_\xi + B^{-1} \Phi_{\zeta\xi} \\ \Phi_\xi \Pi' + \Phi_{\xi\zeta} B^{-T} & \Phi_\xi \end{pmatrix} \end{aligned} \quad (10)$$

in terms of the parameters of \mathcal{M}_0 . We added the parameter matrix $\Phi_{\xi\zeta} = \text{cov}(\xi, \zeta)$ to represent the (possible) covariance among ξ and ζ in the population.³

In \mathcal{M}_0^* we assume that the vector κ , the matrices $\mathcal{A}_0, \mathcal{A}_1$ and the variance matrices of Φ_ξ, Φ_ϵ and Φ_δ have the same model structure as under \mathcal{M}_0 . We also let the variance matrix Φ_{ξ^*} of ξ^* , and the $(m \times n)$ covariance matrix among ξ^* and ξ , $\Phi_{\xi^*\xi}$, to be parameter matrices of the model \mathcal{M}_0^* . In parallel to Condition U above, we define

CONDITION U^* : *The matrices Φ_ξ, Φ_{ξ^*} and $\Phi_{\xi^*\xi}$ are unrestricted parameters of the model, except for Φ_ξ, Φ_{ξ^*} to be symmetric and positive definite.*

The following Lemma 1 will be needed.

LEMMA 1 *Consider the matrix Φ_ν defined in equation (10). Assume that Φ_ξ is non-singular, then*

1. *Conditional to any arbitrary matrix $\Phi_{\xi\zeta}$, Φ_ν is unrestricted iff the matrices Π, Φ_ξ and $B^{-1}\Phi_\zeta B^{-T}$ are unrestricted.*
2. *Conditional to any arbitrary matrix Π , Φ_ν is unrestricted iff the matrices $\Phi_{\xi\zeta}, \Phi_\xi$ and $B^{-1}\Phi_\zeta B^{-T}$ are unrestricted.*

PROOF Due to the identity, Φ_{22}^* is unrestricted iff Φ_ξ is unrestricted. Now, given the expression of Φ_{12}^* in the last equality of (10), it is clear that given $\Phi_{\xi\zeta}$ (idem, given Π), Φ_{21}^* is unrestricted iff Π (idem Φ_{21}^*) is unrestricted. Now, given $\Phi_{\xi\zeta}$, Π and Φ_ξ , Φ_{11}^* is (symmetric) and unrestricted iff $B^{-1}\Phi_\zeta B^{-T}$ is (symmetric) unrestricted. ■

Two models are equivalent at the moment structure level, if they reproduce the same set of moments matrices. When two models are equivalent, the chi-square model test coincide. So the distribution of the model test for one model is the same as the distribution of the chi-square model test of an equivalent model. See for a detailed discussion of equivalent models Luijben (1991). In his paper a sharper definition of equivalent models than in our paper is given. The reason of this sharper definition is that Luijben is also dealing with non-identified and locally-identified models. These points are not an issue of our paper. The following Lemma 2 follows as immediate consequence of 1. of Lemma 1.

LEMMA 2. *The model \mathcal{M}_0 under Condition U, and the model \mathcal{M}_0^* under condition U^* are equivalent at the mean and covariance structure level.*

Even though \mathcal{M}_0 Condition U implies \mathcal{M}_0^* under condition U^* , the reverse is not true. In fact, there are a general class of models that are equivalent to the factor analysis model \mathcal{M}_0^* under Condition U^* . In fact, by 2. of Lemma 1, we see that restricting Π can be interchanged with allowing free covariances among ξ and ζ , something which is not contemplated within the frame of the LISREL model \mathcal{M}_0 specified above.

³Note that \mathcal{M}_0 assumes $\Phi_{\xi\zeta} = 0$.

The models \mathcal{M}_0 and \mathcal{M}_0^* are said to be equivalent on their covariance matrix structure, since they parameterize the same set of covariance matrices for z , and therefore they have the same value of the chi-square goodness-of-fit test statistic. So the distribution of the model test for \mathcal{M}_0 is the same as the distribution of the model test of \mathcal{M}_0^* .

We now will see that when, in addition to equivalence at the moment structure level, the random components v and $(\epsilon', \delta)'$ are independent, further properties for the chi-square model test arises.

Insensitivity of the model test

In this section we will show that when Condition U holds, the model test for the LISREL model posited above has a fundamental insensitivity to severe misspecifications of the model, in particular to non-linear interactions terms as the one reported in Section 2.

Let r be the degrees of freedom associated to the LISREL model \mathcal{M}_0 above (i.e., that is, r is the number of independent moments minus the number of independent parameters). The following Corollary applies.

COROLLARY 1. *Consider the LISREL model formulation as above. Let T be any of the NT chi-square model test mentioned above (for a precise definition, see (12) of the Appendix A1). Let the model be identified with degrees of freedom $r > 0$. Assume further*

1. *The random vector $(\xi', \zeta)'$ is independent of $(\epsilon', \delta)'$, not only uncorrelated.*
2. *The vector $(\epsilon', \delta)'$ is normally distributed*
3. *Condition U holds.*

Then, $T \xrightarrow{L} \chi_r^2$, as $n \rightarrow \infty$.

PROOF: It follows as direct consequence of Theorem 1 of Appendix and Lemma 2. ■

This Theorem explains the Monte Carlo results of Section 2. Indeed, for the DGPs being considered, the presence of the interaction term $\bar{\beta}_{12}\xi_1\xi_2$ is an omitted variable in the structural part of the model that induces a non-null correlation of ζ with the ξ s. Despite that correlation, however, the model estimated in each replication of the Monte Carlo study is a model \mathcal{M}_0 with the conditions of the Corollary 1 being verified, so the NT chi-square test is indeed asymptotically chi-square distributed. That is, the model test statistic is not only insensitive to the correlation among ξ and ζ (i.e., there is no noncentrality shift on the distribution), but neither the chi-square distribution itself gets distorted. This explains Table 1 and the qq-plot showing a close fit of the empirical distribution of the LRT to the χ_3^2 despite the gross misspecification in the model.

We should stress the implication of Corollary 1 when Condition U holds: a) the asymptotic distribution is chi-square; b) the test will have no power, i.e. the non-centrality

parameter for the power of the test (cf., Satorra and Saris, 1985) will be zero, regardless the size of the the coefficients of the non-linear terms.

The phenomenon of the model test being insensitive to the the interaction term is just one example of an omitted correlation among regressors and disturbance terms not being detected by the model test. In fact, this insensitive of the model test to fundamental misspecifications arise when the model ignores variables that can cause spurious correlation among the disturbance terms of regression equations and exogenous variables (variables whose distribution is not being explained by the model). In practice this is a serious limitation of the chi-square goodness-of-fit test in SEM, when Condition U holds.

Note that Condition U ensures that the structural part of the model is saturated (i.e. we have a set of unrestricted regression equations). We see that under Condition U, the degrees of freedom (df) of the model test do not assess any restriction on the structural part of the model. In fact, we could express Condition U saying that the degrees of freedom associated to the structural part of the model is just zero. This is made precise in the following lemma.

LEMMA 3. *Consider a model specification \mathcal{M}_0 as in equations (5) to (7). Assume the model is identified and there is no cross-restrictions among the sets of parameters B , Γ , Φ_ξ and Φ_ζ with the other parameter vectors and matrices. Then, Condition U holds iff the model restricted to equation (5) (supposing η and ξ observable) has degrees of freedom (df) equal to zero.*

PROOF: If there are no cross restrictions as the one mentioned in the conditions of the theorem, the model is identified, and Condition U holds, the model reduced to equation (5) has necessarily 0 degrees of freedom (since it is an identified model, $df \geq 0$, but $df = 0$, otherwise some of the matrices would have an excess of restriction, so Condition U would not hold). Conversely, it follows easily that when the model is identified and $df = 0$ then Condition U is necessarily satisfied. ■

When Condition U holds, we will say the structural part of the model is saturated. Note that a particular instance where Condition U holds is when $B_0 = 0$, i.e. $B = I_m$ with Φ_ζ unrestricted and when Γ is fully unrestricted also. There are other models for which Condition U holds, as for example, when B_0 is lower triangular, Φ_ζ diagonal and Γ fully unrestricted. In all the cases, however, we have the situation where the model at the level of the structural equations is just saturated. Of course, a trivial model where Condition U holds, is the case of a regression model with all the variables observable. Clearly, another way of reading the consequences of Condition U, is that \mathcal{M}_0^* is a factor model with the variance matrix of the common factors unrestricted.

For the sake of simplicity, Corollary 1 has been formulated for the single-group case only, and with the restriction of the ϵ s and δ s to be normally distributed (see 2 of the Corollary). Exactly the same type of results extend in the case of multiple-group models, and also when the ϵ s and δ s are possibly non-normal, provided their components are independent (not just uncorrelated) with unrestricted, and variances. Note that in

a multiple group set-up it is essential not to have restrictions that may restrict across groups the matrix of variances and covariances of the vector v , even though in each group such a matrix Φ_v is left unrestricted (see Satorra (2002) for details on AR for multiple group models).

Discussion

It has been shown that in SEM assuming linear relationships between the latent variables and testing the fit of the model by NT likelihood ratio test, may lead to the conclusion that the model fits the data and to the wrong conclusion that the relationship is linear. This was illustrated in a small Monte Carlo study by using a well-known example which has been studied by several authors, e.g. Kenny and Judd (1984) and Jöreskog and Yang (1996). The conclusion of our finding is that if there may be a nonlinear relationship between the latent variables, it is not sufficient to use a test which is based on means and variances only, like the Normal Theory Likelihood ratio test does. Therefore it is recommended to use other tests. The last decades such tests are in development. Roughly, two groups of approaches can be distinguished. In the first approach product of observed variables are used as indicators for the interaction factors. This method is applied in the method as suggested by Kenny and Judd (1984) and after that by Jöreskog and Yang (1996), and many others. In this approach the choice of the product indicators is an important issue. We refer to a discussion of this approach to Marsh, Wen, and Hau (2004).

Another approach of dealing with nonlinear relationships in SEM is to use a maximum likelihood method in which it is assumed that the observed predictors are normally distributed. Defining the joint density of the latent and the observed variables is not very difficult, however, integrating out the latent variables to get the proper likelihood function, results in a rather unattractive multivariate integral. There are several ways to tackle this problem. To mention just a few publications: Klein and Moosbrugger (2000), Lee and Zhu (2002), and Klein (2007). Also the computer package Mplus, Muthén and Muthén (1998-2007) has a possibility to deal with these kinds of interaction models, however it is unclear to the present authors which method is actually used in Mplus.

We have also seen that the problem with the non-linear terms is a specific instance of a more general problem of the model test in SEM, namely the insensitivity of the model test to correlation among the disturbance terms in the equation and the regressors when the structural equation part of the model has been specified so that Condition U holds. If we were dealing with regression with observed variables, the degrees of freedom of the model test would just be zero under an standard (“saturated”) regression model. It is well known however that in that instance moment structure analysis (standard OLS analysis) is unable to detect omitted variables that cause spurious correlations of disturbance terms and regressors.

In the case we have considered of model \mathcal{M}_0 under Condition U, all the degrees of freedom r of the model test correspond to restrictions on other vector/matrices of parameters (such as α , τ_x , τ_y , B , the Λ s, the Φ_ϵ or Φ_δ), none to restrictions of zero correlation among regressors and the disturbance term. By saturating the structural part of the model, the model test gets free of non-normality and covariances misspecification. The model test do not carry any restriction on that part of the model. This is more general than just misspecification of interaction terms, and may have larger implications for general SEM analysis.

We restricted the formulation to the single group case. It can easily be seen (attending to the general formulation of the conditions of AR that we describe in the Appendix) that also the case of multiple group analysis could be considered. Also the case where the ϵ and δ are not normally distributed, but their components are independent with unconstrained variances. We did not discuss those generalities for the sake of keeping the presentation simple.

We have just considered conditions of the model that make the test insensitive to model misspecification, such as not accounting for the interaction terms. The key condition was the structural part of the model to be saturated, i.e. Condition U. When Condition U does not hold, the model test is likely to loose such insensitivity property. Research on the extend of the distortion of the distribution of the standard chi-square goodness-of-fit test, when there are non-accounted interaction terms (or other forms of correlation among regressors and disturbance terms) and Condition U does not hold, is a matter of extreme interest. We feel however that such additional issues would get us beyond the intended scope of the present paper.

References

- Jöreskog, K.G. & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In: R.E. Marcoulides & G.A. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57-88). Mahwah, NJ: Erlbaum.
- Kenny, D.A. & Judd, C.M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201-210.
- Klein, A.G. (2007). QuasiML (Version 3.10) [Computer software]. Champaign, IL: Author.
- Klein, A.G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457-474.
- Lee, S.Y., & Zhu, H.T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika*, *67*, 189-210.
- Luijben, T.C.W. (1991). Equivalent models in covariance structure analysis. *Psychome-*

trika, 56, 653-665.

Marsh, H.W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, 9, 275-300.

Muthén, L.K., & Muthén, B.O. (1998-2007). *Mplus users guide* (5th ed.). Los Angeles, CA: Muthn & Muthn.

Satorra, A. (2002). Asymptotic Robustness in Multiple Group Linear-Latent Variable Models, *Econometric Theory*, 18, 297-312.

Satorra, A. and W.E. Saris (1985), Power of the likelihood ratio test in covariance structure analysis, *Psychometrika* 50, 1, 83-89

Appendix

Asymptotic robustness in multiple group analysis

In this section we review briefly asymptotic robustness (AR) results for the NT chi-square goodness-of-fit test as derived in Satorra (2002). We will only give the implications for the distribution of the test statistics and not for the standard errors of the estimates of the model parameters. AR shows that the basic result that the likelihood ratio test statistic of the model is chi-square distributed holds under more general assumptions than just normality of the observable variables.

Consider a multiple group data set-up of $\{z_{gi}\}$, where z_{gi} is a $p_g \times 1$ vector of observable variables, $i = 1, \dots, n_g$, indexes individuals and $g = 1, \dots, G$, indexes groups (the dimension of z_{gi} is allowed to vary with g). When clear from the context, the subindex i is suppressed.

We define $n = \sum_{g=1}^G n_g$ and we assume that $\lim_{n \rightarrow \infty} \frac{n_g}{n} = \pi_g > 0$. Consider the $p_g \times p_g$ matrix of (uncentered) sample cross-product moments $S_g = \frac{1}{n_g} \sum_{i=1}^{n_g} z_{gi} z_{gi}'$, and let $s = (s'_1, \dots, s'_G)'$, where $s_g = \text{vech}(S_g)$ be the overall $p^* \times 1$ vector of (non-redundant) sample moments, where $p^* = \sum_{g=1}^G p_g^*$ and $p_g^* = p_g(p_g + 1)/2$. Let Σ_g be the probability limit of S_g as $n_g \rightarrow \infty$, and the $p^* \times 1$ vector $\sigma = \{(\text{vech}(\Sigma_1))', \dots, (\text{vech}(\Sigma_G))'\}'$ be the probability limit of s .

A structural equation model corresponds to the a particular moment structure $\mathcal{M}_0 : \sigma = \sigma(\vartheta)$, where $\sigma(\cdot)$ is continuously differentiable and ϑ is an unconstrained vector of parameters that vary in an open set Θ .

Consider the NT minimum distance (MD) estimator

$$\hat{\vartheta} = \underset{\vartheta \in \Theta}{\text{argmin}} \{s - \sigma(\vartheta)\}' V \{s - \sigma(\vartheta)\}, \quad (11)$$

where

$$V = \bigoplus_{g=1}^G \pi_g V_g^*, \quad \text{with} \quad V_g^* = \frac{1}{2} D'(S_g^{-1} \otimes S_g^{-1}) D.$$

For this NT-MD analysis, the NT chi-square goodness-of-fit test statistic is

$$T = n \left\{ s - \sigma(\hat{\vartheta}) \right\}' V \left\{ s - \sigma(\hat{\vartheta}) \right\}, \quad (12)$$

An alternative approach to NT-MD estimation is pseudo-maximum likelihood (PML), where the function to be minimized is an affine transformation of the log-likelihood function (under NT). A likelihood ratio test statistic for the postulated model is defined that is asymptotically equivalent to the test statistic defined in (12).

The following theorem gives the results that ensure the robustness of the statistic T and the NT LRT statistic (see Satorra (2002, p. 306-307) for a more general version of this theorem).

THEOREM 1A. *Suppose that \mathcal{M}_0 is a model specification with parameter vector ϑ , degrees of freedom r , and estimable by NT-MD. Assume*

1. $z_g = \mu_g + \sum_{j=1}^{J_g} A_{gj} \xi_{jg}$ where the μ_g are vectors, the A_{gj} are matrices, and the ξ_{jg} are random vectors with finite variance matrices, $\Phi_{jg} = \text{cov}(\xi_{jg})$
2. $\mu_g = \mu_g(\vartheta)$, $A_{jg} = A_{jg}(\vartheta)$, $\Phi_{jg} = \Phi_{jg}(\vartheta)$, with $\mu_g(\cdot)$, $A_{jg}(\cdot)$ and $\Phi_{jg}(\cdot)$ continuously differentiable
3. The ξ_{jg} 's have finite fourth-order moment and are independent across j and g
4. Either (for any g and j):
 - a) ξ_{jg} is normally distributed, or/and
 - b) $\phi_j = \text{vech} \Phi_j$ is a sub-vector of the parameter vector ϑ

Then, T of (12) is asymptotically χ_r^2 .