PSYCHOMETRIKA—VOL. 74, NO. 3, 443–455
SEPTEMBER 2009
DOI: 10.1007/S11336-009-9112-5

ON INSENSITIVITY OF THE CHI-SQUARE MODEL TEST TO NONLINEAR MISSPECIFICATION IN STRUCTURAL EQUATION MODELS

AB MOOIJAART

LEIDEN UNIVERSITY, LEIDEN

ALBERT SATORRA

UNIVERSITAT POMPEU FABRA, BARCELONA

In this paper, we show that for some structural equation models (SEM), the classical chi-square goodness-of-fit test is unable to detect the presence of nonlinear terms in the model. As an example, we consider a regression model with latent variables and interactions terms. 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 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-quare 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 nonlinear 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.

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

1. Introduction

In this paper, we deal with nonlinear relationship between the latent variables. Such nonlinearity may come from the existence of, for instance, interaction or quadratic factors. It will be shown that standard chi-square goodness-of-fit tests 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 nonlinear 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 does 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. A variant of the Kenny–Judd'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 asymptotically 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–Judd (1984) model is a

Research of the second author is supported by the grants SEJ2006-13537 and PR2007-0221 from the Spanish Ministry of Science and Technology.

Requests for reprints should be sent to Ab Mooijaart, Department of Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: Mooijaart@Fsw.LeidenUniv.nl

© 2009 The Psychometric Society. This article is published with open access at Springerlink.com

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

special case, and we develop conditions under which results of AR apply in that general model set-up. In Section 4, we show that interaction-term 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 of the model apply. Finally, Section 5 concludes extending the issues discussed to more general model contexts and statistics, comments on alternative approaches that can be used to take care of nonlinear terms in the model, and proposes a decomposition of the degrees of freedom of the model test, among other final issues.

2. 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

$$\eta = \bar{\beta}_0 + \bar{\beta}_1\xi_1 + \bar{\beta}_2\xi_2 + \bar{\beta}_{12}\xi_1\xi_2 + \zeta, \tag{1}$$

where η is a latent dependent 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 of the regression equation. 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 dependent variable η and the latent regressors ξ_1 and ξ_2, with each factor having three indicators; so, in addition to the regression equation (1), the following measurement equations are specified

$$y_j = \bar{\alpha}_{yj} + \bar{\lambda}_{yj}\eta + \epsilon_j, \quad j = 1, 2, 3, \tag{2}$$

$$x_j = \bar{\alpha}_{xj} + \bar{\lambda}_{xj}\xi_1 + \delta_j, \quad j = 1, 2, 3, \tag{3}$$

$$x_j = \bar{\alpha}_{xj} + \bar{\lambda}_{xj}\xi_2 + \delta_j, \quad j = 4, 5, 6, \tag{4}$$

where ϵ_j's and δ_j's are measurement errors. Assume the ϵ_j's and δ_j's are mutually independent, and independent also of η, ξ_1, ξ_2 and ζ. 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 decades, a simple version of model (1) to (4) 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 our paper, we extended their version into a model with three indicators for either the latent predictors and the latent dependent variable.

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

$$\mathcal{M}_0: \quad \eta = \beta_0 + \beta_1\xi_1 + \beta_2\xi_2 + \zeta, \tag{5}$$

with the β's being now the parameters to be estimated. We also consider (2) to (4) with the $\bar{\alpha}$'s and $\bar{\lambda}$'s substituted by free parameters to be estimated, α's and λ's, respectively. For purposes of model identification, for η and 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 and ϵ's). To make the specification of the model more complete, we assume that $\mathcal{M}_0$ specifies normality for the distribution of the vector of observed variables (this to be called the NT assumption).

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

Note that the specification $\mathcal{M}_0$ ignores the interaction term $\bar{\beta}_{12}\xi_1\xi_2$, whatever the magnitude of $\bar{\beta}_{12}$, 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 (4) with $\bar{\beta}_{12} \neq 0$, do we get any indication of misspecification of the model? Note that such an analysis can be done by the regular 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 no 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 model equations (1) to (4) with the structural parameters set to the same values as in Jöreskog and Yang (1996); that is, $\bar{\beta}_0 = 1$, $\bar{\beta}_1 = 0.2$, $\bar{\beta}_2 = 0.4$, $\bar{\beta}_{12} = 0.7$, $\bar{\phi}_{11} = 0.49$, $\bar{\phi}_{22} = 0.64$, $\bar{\phi}_{12} = 0.2352$ and $\bar{\psi} = 0.2$, where we used the notation $\bar{\psi} = \mathrm{var}(\zeta)$ and $\bar{\phi}_{ij} = \mathrm{cov}(\xi_i, \xi_j)$ $(i, j = 1, 2)$. Note that in the population considered here the interaction effect is substantial. A measure of the size of this effect is the proportion of variance of the dependent variable, η, that is explained by this interaction. Under the normality assumption for the factors, this relative interaction effect is $\bar{\beta}_{12}^2(\bar{\phi}_{11}\bar{\phi}_{22} + \bar{\phi}_{12}^2)/\mathrm{var}(\eta)$, where $\mathrm{var}(\eta) = \bar{\beta}_1^2\bar{\phi}_{11} + \bar{\beta}_2^2\bar{\phi}_{22} + 2\bar{\beta}_1\bar{\beta}_2\bar{\phi}_{12} + \bar{\beta}_{12}^2(\bar{\phi}_{11}\bar{\phi}_{22} + \bar{\phi}_{12}^2) + \bar{\psi}$, which equals 0.3345, while the squared multiple regression coefficient is 0.6299; so more than 33% of the variance of the dependent factor is explained by the interaction factor. In the well-known study of Marsh, Wen and Hau (2004) the relative interaction effect is 0.0436 and the squared multiple regression is 0.4596, that is, an interaction effect much less substantial than in our population. The parameters of the measurement part are differently defined for our population as is done in Jöreskog and Yang (1996). Our model has three indicators for each factor; the α's are set to zero; the three factor loadings of the Y variables are set to 0.8, and of the X variables are set to 1, 0.8, and 0.6. The error variances corresponding to the Y variables are all equal to 0.36; the error variances for the first three X variables are 0.51, 0.64, and 0.64, and for the other three X variables are 0.36, 0.51, and 0.51.

The Monte Carlo illustration considers replicating 1000 times independent and identically distributed (iid) sampling from this population, samples of size $n = 500$.[1]

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. Simulations are carried out under 5 different data generating processes (DGP) that differ by the distribution of the ξ's: DGP1 to DGP4 set ξ_1 and ξ_2 to be chi-square distributed with 1, 2, 5, and 10 degrees of freedom, respectively, centered to have zero mean; DGP5 sets the ξ's to be normally distributed. In all the cases, the ζ and the ϵ's are iid, sampled from a normal distribution. So DGP1 to DGP4 do have skewed latent variables (DGP1 is the most skewed), while DGP5 has normally distributed ξ's.[2]

Regarding the skewness and kurtosis of the simulated data, the univariate skewness of the Y variables for the data generating process DGP5 (the case where ξ_1 and ξ_2 are normally distributed) was close to 0.40, a value that deviates significantly from 0, its value under normality (for a sample size $n = 500$, the standard error of the estimate of skewness was close to 0.11).

[1] All computations were carried out by MATLAB, version R2008a. MATLAB is a registered trademark of The MathWorks, Inc. The functions `mvnrnd` and `chi2rnd` were used for generating scores according to a normal and a chi-square distribution, respectively. The negative of the likelihood function was minimized using the function `fminunc` from MATLAB's 'Optimization Toolbox'. The 'qqplot' was created by using the standard plot function in MATLAB.

[2] Note that even for DGP5 the distribution of y will deviate from normality due to the interaction term that is present in (1).

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

TABLE 1.
For each of the sampling distributions considered, DGP1 to DGP5, the table reports the mean and variance of the LRT (df = 32) and the Kolmogorov–Smirnov test (and p-values) for the fit of the LRT to a χ^2_{32}.

	DGP				
	1	2	3	4	5
Mean LRT	32.432	32.403	32.272	31.994	32.386
Variance LRT	68.560	65.890	61.688	62.235	63.253
	KS test of fit to a χ^2_{32}				
KS test	0.027	0.032	0.037	0.024	0.037
p-value	0.463	0.224	0.130	0.585	0.121

Mardia's multivariate test for skewness (see Mardia, 1970, 1974) for the 9 variables computed for each sample gives a mean value over the 1000 replications of 3.28 (the mean of the p-values is 0.0001, over the 1000 replications). Thus, in the case of DGP5, a normality test based on skewness would reject the normality assumption. Since DGP5 is the case closest to normality of the ones considered, we conclude that standard skewness-based test of normality applied to sample data would clearly reject the normality assumption in all the data generating processes considered. We should report, however, that a test of normality based on kurtosis would not lead to rejection of the normality assumption in the case of DGP5, but it would reject the normality assumption in all other data generating processes, DGP1 to DGP4. For the DGP5, over the 1000 replications, the mean of the univariate kurtosis of the Y variables was 3.8 and of the Mardia's multivariate kurtosis was 100.86, with a mean of the p-value of 0.31 (note that the expected value under normality of Mardia's multivariate kurtosis is $p(p+2) = 99$). Thus, for all the data generating processes considered (DGP1 to DGP5), standard test of normality applied to sample data would reject the NT assumption. Samples arising from DGP1 are the most extreme regarding deviation from normality.[3]

The fit of the Monte Carlo distribution of LRT to a χ^2_{32} is assessed by a Kolmogorov–Smirnov tests and a qq-plot. Table 1 reports the mean and variance of the Monte Carlo distribution of LRT for the different DGPs conditions and the results for the Kolmogorov–Smirnov test.

From Table 1 we see that, although the model $\mathcal{M}_0$ is seriously misspecified (it ignores the presence of an interaction term of substantial size) and assumes NT when in fact the observable variables are skewed (a standard normality test would reject the normality assumption), for all the data conditions, the Monte Carlo distribution of LRT has mean and variance close to what is expected under a chi-square distribution. For the specified $\mathcal{M}_0$, the LRT has 32 degrees of freedom (df); so, if the distribution of the LRT were chi-square, the mean and variance reported in Table 1 should be approximately 32 and 64, respectively (i.e., the mean and variance for a χ^2_{32}). The means and variances reported in the table are very close to those theoretical values. Further, the qq-plot (see Fig. 1) of the fit of the Monte Carlo distribution of LRT to a χ^2_{32} in the case of DGP1 (the case of highest deviation from normality) shows an striking accurate fit to a χ^2_{32} (points lying close to the diagonal line), that is, misspecification of an interaction term does not seem to induce any distortion on the distribution of the LRT. From this small Monte Carlo study, we thus conclude that in all the DGPs conditions considered the LRT has the same distribution (a χ^2_{32}) as if no misspecification were present in the model and the observable variables were normally distributed. This is disturbing since we would like the model test to be sensitive to the severe misspecification of model $\mathcal{M}_0$, that ignores an interaction term of substantive size.

[3] A full table with all univariate and multivariate tests for the skewness and kurtosis of the variables under DGP1 to DGP5 is available upon request from the first author.

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

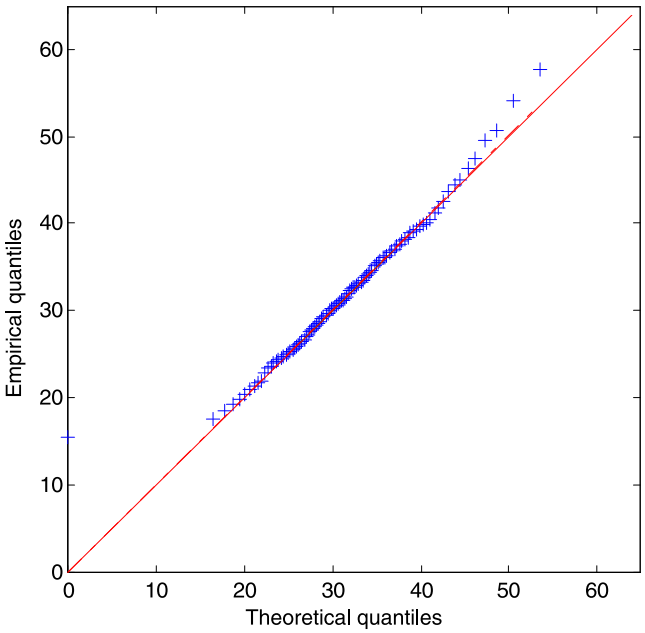


FIGURE 1.
A qq-plot of the Monte Carlo distribution of LRT versus χ^2_{32} for DGP1.

This surprising lack of sensitivity of the chi-square model test to a substantial misspecification of the model, such as an interaction term of a substantial size, 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 where the model considered in our Monte Carlo study is a special case.

3. Models and Assumptions

In this section, we present a general model set-up and analytic conditions under which the same phenomenon as the one reported in our Monte Carlo study necessarily arises. For that we will apply results of the theory of asymptotic robustness (AR). For the sake of completeness of this paper, results of AR to be used in the paper are given in Appendix.

3.1. General Model Set-up

Consider $\mathcal{M}_0$ to be the general LISREL model with mean structures (see Jöreskog & Sörbom, 1996)

$$\eta = \alpha + B_0\eta + \Gamma\xi + \zeta, \tag{6}$$

$$y = \tau_y + \Lambda_y\eta + \epsilon, \tag{7}$$

$$x = \tau_x + \Lambda_x\xi + \delta, \tag{8}$$

where α, τ_y and τ_x are vectors of constant and intercept terms, and B_0, Γ, Λ_y and Λ_x are matrices of parameters. In the econometric and SEM terminology, (6) is referred as "simultaneous equations" or "structural" part of the model, while (7) and (8) are the "measurement" (or factor analytic) part of the model.

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

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 said otherwise, $\mathcal{M}_0$ assumes that $\Phi_{\xi\zeta} = 0$.[4] 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 as Φ_ξ, Φ_ζ, Φ_ϵ, and Φ_δ, respectively. Under a specific model $\mathcal{M}_0$, the above vectors and matrices of parameters are expressed as a function of a vector θ of parameters of the model. Define $\Pi = B^{-1}\Gamma$, which is the matrix of "reduced form" coefficients of the "simultaneous equation". For the results to be developed, we require the model $\mathcal{M}_0$ to satisfy 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 rewritten in the following vector equation form

$$z = \mathcal{A}_0 + \begin{pmatrix} \Lambda_y B^{-1}\Gamma \\ \Lambda_x \end{pmatrix}(\xi - \kappa) + \begin{pmatrix} \Lambda_y B^{-1} \\ 0 \end{pmatrix}\zeta + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix}, \tag{9}$$

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

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

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.

Rewrite (9) as

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

so that

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

with

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

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

$$\mathcal{M}_0^*: \quad z = \mathcal{A}_0 + \mathcal{A}_1\,\upsilon + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix}, \tag{10}$$

where

$$\upsilon = \begin{pmatrix} \xi^* \\ \xi - \kappa \end{pmatrix}.$$

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

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

Letting

$$\Phi_{\upsilon}\big(=\mathrm{cov}(\upsilon)\big)=\begin{pmatrix}\Phi^{*}_{11} & \Phi^{*}_{12}\\ \Phi^{*}_{21} & \Phi^{*}_{22}\end{pmatrix},$$

we obtain

$$\begin{aligned}\Phi_{\upsilon}&=\begin{pmatrix}\Pi & B^{-1}\\ I & 0\end{pmatrix}\begin{pmatrix}\Phi_{\xi} & \Phi_{\xi\zeta}\\ \Phi_{\zeta\xi} & \Phi_{\zeta}\end{pmatrix}\begin{pmatrix}\Pi & B^{-1}\\ I & 0\end{pmatrix}'\\ &=\begin{pmatrix}\Pi\Phi_{\xi}\Pi'+B^{-1}\Phi_{\zeta}B^{-T}+\Pi\Phi_{\xi\zeta}B^{-T}+B^{-1}\Phi_{\zeta\xi}\Pi' & \Pi\Phi_{\xi}+B^{-1}\Phi_{\zeta\xi}\\ \Phi_{\xi}\Pi'+\Phi_{\xi\zeta}B^{-T} & \Phi_{\xi}\end{pmatrix}. \quad (11)\end{aligned}$$

We added the parameter matrix $\Phi_{\xi\zeta}=\mathrm{cov}(\xi,\zeta)$ to represent the (possible) covariance among ξ and ζ in the population.[5]

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 of 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}, $\Phi_{\xi^\star}$ and $\Phi_{\xi^*\xi}$ are unrestricted parameters of the model, except for Φ_{ξ}, $\Phi_{\xi^\star}$ to be symmetric and positive definite.*

The following Lemma 1 will be needed.

Lemma 1. *Consider the matrix Φ_{υ} defined in (11). Assume that Φ_{ξ} is non-singular, then*

(i) *Conditional to any arbitrary matrix $\Phi_{\xi\zeta}$, Φ_{υ} is unrestricted iff the matrices Π, Φ_{ξ} and $B^{-1}\Phi_{\zeta}B^{-T}$ are unrestricted.*
(ii) *Conditional to any arbitrary matrix Π, Φ_{υ} is unrestricted iff the matrices $\Phi_{\xi\zeta}$, Φ_{ξ} 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 (11), 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. For a detailed discussion of equivalent models, see 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. Lemma 2 follows as an immediate consequence of Lemma 1(i).

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.*

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

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

Even though $\mathcal{M}_0$ under Condition U implies $\mathcal{M}_0^*$ under Condition U*, the reverse is not true. In fact, there is a general class of models that is equivalent to the factor analysis model $\mathcal{M}_0^*$ under Condition U*. By Lemma 1(ii), 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.

The models $\mathcal{M}_0$ and $\mathcal{M}_0^*$ are said to be equivalent with respect to 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 will now see that when, in addition to the equivalence at the moment structure level, the random components υ and $(\epsilon', \delta')'$ are independent, further properties for the chi-square model test arise.

4. 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 nonlinear terms as the one reported in Section 2.

Let r be the number of 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 (A.2) of Appendix). Let the model be identified with $r > 0$ degrees of freedom. Assume further that*

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

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

Proof: It follows by combining Lemma 2 with Theorem A.1 of Appendix. □

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 zero non-centrality shift of the distribution), but also the chi-square distribution itself does not get distorted. This explains Table 1 and the qq-plot showing a close fit of the empirical distribution of the LRT to the χ_{32}^2 despite the gross misspecification in the model.

We should stress the implication of Corollary 1 when Condition U holds: (a) the asymptotic distribution will be chi-square; (b) the test will have no power, i.e., the non-centrality parameter for the power of the test (cf., Satorra & Saris, 1985) will be zero, regardless the size of the coefficients of the nonlinear terms.

The phenomenon of the model test being insensitive to the interaction term is just one example of an omitted correlation among regressors and disturbance terms not being detected by

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

the model test. In fact, this insensitivity of the model test to fundamental misspecifications arises 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 number of 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* (6) *to* (8). *Assume the model is identified and there are 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* (6) (*supposing* η *and* ξ *observable*) *has the number of 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 (6) has necessarily 0 degrees of freedom (since it is an identified model, $\text{df} \geq 0$, but $\text{df} = 0$, otherwise some of the matrices would have an excess of restrictions, so Condition U would not hold). Conversely, it follows easily that when the model is identified and $\text{df} = 0$ then Condition U is necessarily satisfied. □

When Condition U holds, we will say that 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 with Γ also fully unrestricted. There are other models for which Condition U holds, 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 Corollary 1(ii)). 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 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 υ, even though in each group such a matrix Φ_υ is left unrestricted (see Satorra, 2002 for details on AR for multiple group models).

5. Discussion

It has been shown that when using only means and covariances the standard chi-square goodness-fit-test of SEM can give a good fit of the model despite substantial misspecification of nonlinear terms in the model. This was illustrated in a small Monte Carlo study by using an extended version of the well-know Kenny and Judd (1984). The conclusion of our finding is that if we are concerned with misspecification on nonlinear relationships, e.g., interaction terms among latent regressors, it is not sufficient to use a test which is based on means and variances only, such as the LRT. Therefore, it is recommended to use other tests. Additional tests have been in development during the last decades to assess nonlinear components of the model. Roughly,

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

several approaches can be distinguished. In the first approach, products of observed variables are used as indicators for the interaction factors. This idea is applied in the method 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 et al. (2004).

Another approach of dealing with nonlinear relationships in SEM is to use the maximum likelihood (ML) approach under the assumption 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, see Klein and Moosbrugger (2000), Lee and Zhu (2002), and Klein (2007), to mention just a few publications. The SEM software Mplus of Muthén and Muthén (1998–2007) has also an option to deal with these kinds of interaction models under the normality assumption for the regressors.

An alternative method is to use two-stage instrumental variables, a method borrowed from econometrics. This method was proposed by Bollen and Paxton (1998a, 1998b). An alternative to the ML approach is based on Bayesian methods, which have a growing interest these days. See Chap. 4 of the book of Lee (2007) that is devoted to this problem; see also the paper of Arminger and Muthén (1998). To mention a few other methods, Wall and Amemiya (2003) propose a two-stage method in which at the first stage the measurement parameters are estimated and the estimates are used for estimating the factor scores. On the basis of these factor scores, errors are estimated and then estimation of model parameters is carried out by using higher-order moments of these errors. Finally, we mention Bauer (2005) where mixture models are applied to obtain approximations to nonlinear models.

Note that our results apply both to just covariance structures and to the case of mean and covariance structures where degrees of freedom are gained also throughout restrictions on the first-order moments. In fact, when the means are included, Condition U remains the same, as only the conditions on the covariances of the variables play a role in the results discussed in our paper. This is a consequence of the fact that the results of AR apply without changing the conditions to the case where the means are restricted (Satorra, 2002). Note that the Monte Carlo illustration of Section 2 was, in fact, an example where the analysis included restrictions on the means.

We concentrated on the behavior of the LRT. The same results hold for the asymptotic distribution-free approach of Browne (1984) or the scaled or adjusted test statistics of Satorra and Bentler (1994). This is so since, when the LRT is asymptotically chi-squared, the mentioned chi-square goodness-of-fit are equivalent test statistics (see Satorra, 1989), so any test (whether the NT, or the asymptotic distribution-free test, or the scaled test statistic) that is based in the analysis of the means and covariances would suffer of the same insensitivity.

In this paper, 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, that is, Condition U. When Condition U does not hold, the model test is likely to loose this insensitivity property. An issue for further research is to explore how much distortion of the chi-square distribution arises under misspecification of interaction terms (or other forms of nonlinear misspecification) when Condition U does not hold. Such additional issues, however, would take us beyond the intended scope of the present paper.

We have also seen that the problem with the nonlinear 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 Condition U holds. If we were dealing with regression with observed variables, the number of degrees of freedom of the model test would just be zero under a standard (“saturated”) regression model. It is well known, however, that in that instance moment structure analysis (standard OLS analysis)

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

is unable to detect omitted variables that cause spurious correlations of disturbance terms and regressors. Under Condition U, all the "active" degrees of freedom r of the model test correspond to restrictions on 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 is unaffected by both the non-normality of the variables and the misspecification on the covariance among regressors and disturbance term. This is more general than just misspecification of interaction terms, and may have larger implications for general SEM analysis.

Condition U ensures zero restrictions on the structural part of the model. Given a SEM model, a distinction could be drawn between the degrees of freedom of the structural part of the model, say df_s, and the degrees of freedom of the measurement part, say df_m. In particular, when Condition U holds, the structural part of the model is saturated, so $\mathrm{df}_s = 0$ and the result of insensitivity of the model test to nonlinear terms in the model applies. Computation of df_s could be done by evaluating the rank of the Jacobian of the covariance matrix of $(\eta', \xi')'$ with respect to the free parameters in the matrices B, Γ, Φ_ξ, and Φ_ζ. Although, more experience and research is needed on this issue of a distinction between structural and measurement degrees of freedom, SEM software could easily provide the df_m and df_s values associated to the chi-square goodness-of-fit test of the analyzed model.

Appendix

For the sake of completeness, this appendix gives a summary of the theoretical result on asymptotic robustness (AR) needed for proving Corollary 1 of Section 3. Only asymptotic robustness results for the test statistics are reviewed, not including results on AR for NT standard errors. Basically, conditions are stated under which the NT chi-square goodness-of-fit test statistics retain their asymptotic chi-squaredness even when the observable variables deviate severely from normality. We adhere to the formulation of Satorra (2002) restricted, though, to the single group case. For further references on AR, see Satorra (2002) and references therein. For a simple description of AR in SEM, see Satorra (1990).

Consider the data $\{z_i\}$, where z_i is a $p \times 1$ vector of observable variables, $i\,(= 1, \ldots, n)$ indexes cases, n is the sample size. The subindex i will be suppressed when it is clear from the context.

Consider the $p \times p$ matrix of (uncentered) sample cross-product moments $S = \frac{1}{n-1} \times \sum_{i=1}^{n} z_i z_i'$, and let $s = \mathrm{vech}(S)$ be the $p^\star \times 1$ vector of (non-redundant) sample moments, where $p^\star = p_g(p_g + 1)/2$. Let Σ and σ be the respective probability limits of S and s as $n \to \infty$. A structural equation model corresponds to 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 and compact set $\Theta \subset R^q$, q being the dimension of ϑ.

Consider the NT minimum distance (MD) estimator

$$\hat{\vartheta} = \underset{\vartheta \in \Theta}{\operatorname{argmin}} \big\{ s - \sigma(\vartheta) \big\}' V \big\{ s - \sigma(\vartheta) \big\}, \tag{A.1}$$

where

$$V = \frac{1}{2} D' \big(S^{-1} \otimes S^{-1} \big) D.$$

(D is the so-called "duplication" matrix for which $\mathrm{vec}\, S = Ds$, where vec is the vectorization operator.)

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

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

$$T = n\{s - \sigma(\hat{\vartheta})\}' V \{s - \sigma(\hat{\vartheta})\}. \tag{A.2}$$

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 (LRT) for $\mathcal{M}_0$ is defined, and it can be shown that it is asymptotically equivalent to the test statistic of (A.2).

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

Theorem A.1. *Suppose the model $\mathcal{M}_0$ with r degrees of freedom. Suppose estimation is carried out using NT–MD. Consider T of (A.2), and assume*

1. $z = \mu + \sum_{j=1}^{J} A_j \xi_j$, *where the μ are vectors, the A_j are matrices, and the ξ_j are random vectors with finite variance matrices, $\Phi_j = \mathrm{cov}(\xi_j)$.*
2. $\mu = \mu(\vartheta)$, $A_j = A_j(\vartheta)$, $\Phi_j = \Phi_j(\vartheta)$, *with $\mu(\cdot)$, $A_j(\cdot)$ and $\Phi_j(\cdot)$ continuously differentiable vector- and matrix-valued functions.*
3. *The ξ_j's have finite fourth-order moment and are independent across j.*
4. *For any j, either*:
   (a) *ξ_j is normally distributed, or/and*
   (b) *$\phi_j = \mathrm{vech}\,\Phi_j$ is a sub-vector of the parameter vector ϑ.*

Then $T \xrightarrow{L} \chi^2_r$ as $n \to \infty$.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Arminger, G., & Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis–Hastings algorithm. *Psychometrika*, *63*(3), 271–300.

Bauer, D.J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*, *4*, 513–535.

Bollen, K.A., & Paxton, P. (1998a). Interactions of latent variables in structural equation models. *Structural Equation Modeling*, *5*, 267–293.

Bollen, K.A., & Paxton, P. (1998b). Two-stage least squares estimation of interaction effects. In R.E. Schumacker & G.A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling*. Hillsdale: Erlbaum.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83.

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). Hillsdale: Erlbaum.

Jöreskog, K.J., & Sörbom, D. (1996). *LISREL 8: Users Reference Guide*. Chicago: SSI Scientific Software International, Inc.

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)*. Champaign: Author. [Computer software].

Klein, A.G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.

Lee, S.-Y. (2007). *Structural equation modelling: a Bayesian approach*. New York: Wiley.

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. *Psychometrika*, *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.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530.

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press

Mardia, K.V. (1974). Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhya, Series B*, *36*, 115–128.
Muthén, L.K., & Muthén, B.O. (1998–2007). *Mplus user's guide* (5th ed.). Los Angeles: Muthén & Muthén.
Satorra, A. (1989). Alternative test criteria in covariance structure analysis: a unified approach. *Psychometrika*, *54*, 131–151.
Satorra, A. (1990). Robustness issues in structural equation modeling: a review of recent developments. *Quality & Quantity*, *24*, 367–386.
Satorra, A. (2002). Asymptotic robustness in multiple group linear-latent variable models. *Econometric Theory*, *18*, 297–312.
Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. van Eye & C.C. Clogg (Eds.), *Latent variable analysis in developmental research* (pp. 285–305). Thousand Oaks: Sage.
Satorra, A., & Saris, W.E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*(1), 83–89.
Wall, M.M., & Amemiya, Y. (2003). A method of moment technique for fitting interaction effects in structural equation models. *The British Journal of Mathematical and Statistical Psychology*, *56*, 47–63.

Manuscript Received: 6 OCT 2008
Final Version Received: 26 JAN 2009
Published Online Date: 17 MAR 2009

https://doi.org/10.1007/s11336-009-9112-5 Published online by Cambridge University Press