Title

Testing for Non-linear Relationship in Structural Equation Modeling

Authors
Ab Mooijaart
Albert Satorra

Publication Date
2011-10-25

Testing for non-linear relationship in structural equation modeling *

Ab Mooijaart

Leiden University

&

Albert Satorra

Universitat Pompeu Fabra and Barcelona GSE

June 23, 2010

*Requests for reprints should be sent to Ab Mooijaart, Institute of Psychology, Unit Methodology and Statistics, Leiden University, P.O. Box 9555, 2300 RB, Leiden, the Netherlands. Research of the second author is supported by grant SEJ2006-13537 from the Spanish Ministry of Science and Technology. The second author's work was mainly carried out while visiting UCLA.

Testing for non-linear relationship in structural equation modeling

Abstract

Since the seminal paper of Kenny and Judd (1984) several methods have been proposed for dealing with non-linear latent variable models. In all these methods more information from the data than just means and covariances is required. In this paper we also use more than just first- and second-order moments; however, we restrict ourselves to a selection of third-order moments. The key issue in this paper is a procedure for selection of the third-order moments for estimating the parameters and testing the goodness-of-fit of a model. The procedure we propose is based on the power of the test associated to the different choices of third-order moments, where the power is defined as the probability that a model without nonlinear factors is rejected by the goodness-of-fit model test when there are in fact nonlinear factors. The main conclusion of this paper is that evaluation of power for selection of third-order moments can easily be done by multivariate analysis of third-order moments, a moment test, without reference to a structural equation model. A consequence of this result is that in practice the selection of third-order moments is conceptually and computationally simple. Examples will illustrate our method.

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

saturated model

Introduction

In Mooijaart and Satorra (2009) it has been shown that, under some general restrictions, the normal theory test statistics, which are based on means and covariances only, are not able to distinguish between models with and without nonlinear factors. One conclusion is that for analyzing nonlinear latent models more information than means and covariances has to be used for applying nonlinear models. Several methods have been proposed for analyzing models with latent nonlinear relationships. Originally, the main method was to utilize product indicators for the independent predictors. See e.g. Kenny and Judd (1984) and Jöreskog and Yang (1996), among many others. One of the key issues in these methods was the choice of the product indicators, see e.g. Marsh, Wen, and Hau (2004). In most of these methods it was assumed that the latent predictor variables are normally distributed. More recently, new methods are now been proposed, in which also the predictors are assumed to be normally distributed. The advantage of such an assumption is that well-known statistical methods can be used, like e.g., the maximum likelihood method. The reason for this is that under this normality assumption it is quite simple to derive the density function, even if there are products of factors in the model. So in principle it is not too difficult to formulate the likelihood function. However, this likelihood function contains a multivariate integral, which is not easy to deal with in practice. There are several ways to tackle this problem. To mention a few methods: Normal mixtures were used by Klein and Moosbrugger (2000) in their LMS method; the method of Muthén and Muthén (1998-2007) in their computer package MPLUS also approximates this multivariate integral, but now by numerical integration. Klein (2007) in his QML methods uses a quasi-maximum likelihood method. A different approach is, although in fact it is dealing also with finding maximum likelihood estimates, is the Bayesian approach combined with the MCMC method as discussed by Lee and Zhu (2002) and Lee (2007). In all these methods it is assumed that the predictors are normally distributed, although some method may be more robust against violation of the normality assumption. For instance, the Bayesian approach behaves better in small samples than the more traditional ML method as used in LMS/MPLUS. On the other hand, with large samples the MCMC method may become extremely slow and can hardly be recommended compared to the other methods. Moreover, for large samples the function to be optimized in the Bayesian approach becomes almost equal to the likelihood function because the influence of the prior distribution almost vanishes.

In this paper we do not use some variation of the maximum likelihood method, but concentrate on fitting the first, second, and some selection of the third order moments, using the approach of Mooijaart and Bentler (2010). It could be argued that in the analysis of models with interaction terms supplementing first- and second-order moments with a selection of higher-order moments can yield better results (in terms of stability and robustness against small samples) than methods that imply to involve the full set of higher-order moments. Just like in the more traditional approach with product indicators, here also the key issue is: what is the selection of the third order moments? An advantage of our approach is that, under rather weak assumptions and for not too small sample sizes, it is possible to define a statistical test which may be used for testing the goodness-of-fit of the model. Interesting to note is that the methods mentioned above do not have such a goodness-of-fit test. This can be explained by the fact that it is unclear under what conditions the likelihood ratio test is chi-square distributed. See also for a discussion of this point Klein and Moosbrugger. The idea of our approach is that by using such a proper goodness-of-fit test and its corresponding power for testing if a model with only linear factors have to be rejected when in fact the model is a nonlinear model, we can select third order moments which result in an optimal power.

An important result of our paper is that a test based on the sample third order moments, has the same non-centrality parameter as the test based on the least squares estimates. The importance of this result is that the non-centrality parameter can be estimated easily and so, in addition with the corresponding degrees of freedom of the test, the power can also be estimated easily.

The presentation of this paper is as follows. Section 1 discusses model formulation and estimation of the parameters and testing the model. In Section 2 a simulation study will be presented to illustrate the practical importance of the choice of the third order moments and how the alternative choices affect the power of the model test. In Section 3 selection of third order moments and the power will be discussed. Section 4 discusses the types of third-order moments on the bases of the power function. A forward selection procedure for higher order moments is discussed in Section 5. Section 6 concludes with a discussion. Lemma 1, that shows the equivalence of mean and covariance structure implied by models with and without interaction is proved in the appendix.

1 Formulation of the model and estimation and testing

In LISREL formulation we can write for a model with latent product variables:

$$\eta = \alpha + B_0\eta + \Gamma_1\xi + \Gamma_2(\xi \otimes \xi) + \zeta \quad (1)$$

$$y = \nu_y + \Lambda_y\eta + \epsilon \quad (2)$$

$$x = \nu_x + \Lambda_x\xi + \delta \quad (3)$$

where y and x are of dimensions p and q respectively, $(\xi \otimes \xi)$ defines product factors. For further use, we define $B = I - B_0$. Defining $z = (y', x')'$, the means, covariances and third order moments can be formulated as a function of the model. In a general formulation, these can be written as

$$\begin{aligned}
\sigma_1 &\equiv E[z] = A\varphi_1 \\
\sigma_2 &\equiv D^{+}E\left[(z-\sigma_1)\otimes(z-\sigma_1)\right] = D^{+}(A\otimes A)D\varphi_2 \\
\sigma_3 &\equiv T^{+}E\left[(z-\sigma_1)\otimes(z-\sigma_1)\otimes(z-\sigma_1)\right] = T^{+}(A\otimes A\otimes A)T\varphi_3
\end{aligned}$$

where the matrix A is given explicitly in (14) below and where $\sigma_1 = E(z)$, and σ_2 and σ_3 are the vectors of second- and third-order moments in deviation from the means. Furthermore, these latter vectors are in reduced form which means that all duplicated and triplicated elements have been removed. For a definition of the triplication matrix T, see Meijer (2005). The vectors φ_1, φ_2 and φ_3 are analogously defined however, not for the observed variables in z, but for the independent observed or latent variables in the model, i.e. the variables ξ, δ, ϵ and ζ. Matrix A is a function of the model parameters. The model expressed by equations (1) to (3) will be denoted by H_1. Denote $\Psi = \mathrm{cov}\,(\zeta)$ and $\Phi = \mathrm{cov}\,(\xi)$.

Now suppose s is a vector of sample estimates of vector σ, then the parameters will be estimated by minimizing the weighted least squares (WLS) fitting function:

$$f_{WLS}(s,\sigma) = (s-\sigma(\theta))'W(s-\sigma(\theta))$$

where s is a vector of all first-, second- and third-order sample moments and σ a vector of all first, second- and third-order moments as a function of the model parameters collected in vector θ. Here W is a weight matrix that possibly varies with the data but converges in probability (when sample size tends to ∞) to a positive definite matrix W_0. An often used weighted least squares function is one in which the weight matrix is an estimate of the covariance matrix of vector s. However, it has been shown in many simulation studies that estimates by using this method leads to biased estimates. This in particular holds for fitting means and covariances. In our case where we fit in addition a selection of third order moments it can be expected that the bias of the estimates will even be larger. So often the most typical fitting function is the LS one, i.e. the one where W is just the identity matrix. The result to be derived in this paper holds in general for WLS with W being block diagonal with respect to the vector s_{12} of first and second-order moments and the vector s_3 of third order moments.

A test statistic which can be used for testing a model is defined as follows: let $\mathrm{cov}(\sqrt{n}s) = \Gamma$, and define

$$T_{WLS} = (s - \sigma(\theta))'(\hat{\Gamma}^{-1} - \hat{\Gamma}^{-1}\hat{\dot{\sigma}}(\hat{\dot{\sigma}}'\hat{\Gamma}^{-1}\hat{\dot{\sigma}})^{-1}\hat{\dot{\sigma}}'\hat{\Gamma}^{-1})(s - \sigma(\theta)) \qquad (4)$$

where $\hat{\dot{\sigma}}$ is the Jacobian of $\sigma(\theta)$ evaluated at the WLS estimator $\hat{\theta}$, and $\hat{\Gamma}$ is a consistent estimator of Γ. Under the set-up, T_{WLS} is asymptotically chi-square distributed when the analyzed model holds, and non-central chi-square distributed when there are specification errors. See for a detailed discussion of these statistics and its distribution, Browne (1984) and Satorra (1989). The noncentrality parameter plays an important role in determining the power of a test. In this paper the power will be used as a selection criterion for choosing the third-order moment to include in the vector of moments s. Therefore in a next section we will be dealing with this noncentrality parameter extensively.

An alternative test will be used also. In this test we simple test whether some set of sample third-order moments are zero or not, without involving a moment structure. We consider the following moment test (MT) statistic

$$T_{MT} = n s_3' \hat{\Gamma}_{33}^{-1} s_3 \tag{5}$$

where $\hat{\Gamma}_{33}$ is a consistent estimate of the asymptotic variance matrix of the vector of higher-order moments s_3 included in s. Note that this test does not depend on a specific model for the data and can be used easily in the sample. In a next section we will investigate the relationship between T_{WLS} and T_{MT} .

2 A motivating illustration

In this section we discuss an example and show what the influence is of selection of the third order moments on the power of the test. Note that in this section the power will be defined as the probability of rejecting the model without nonlinear factors, where in fact the true model contains nonlinear factors.

Small simulation study: In this study we reanalyse data according to the so-called Kenny and Judd (1984) model. This model was also analyzed by many others, for instance by Jöreskog and Yang (1996) and Klein and Moosbrugger (2000). In our simulation study we use the same set up as they did for their study of the LMS method. The sample size we use is $n = 600$ with number of replications 500.

The model contains two latent variables which predict an observed variable. Besides the "main" effects of the predictors there is also an "interaction" effect. The two latent predictors have each two observed indicators. The observed indicators of the predictors are $V1$ to $V4$. Variable $V5$ is the dependent variable. The parameter which plays a key role is the interaction parameter. In our study this parameter will vary from 0.0 to 0.7. See for a discussion of a Monte Carlo study in which all the model parameters are estimated by the WLS method Mooijaart and Bentler (2010).

In this paper the main idea is that the power will be used for selection of the third order moments. Therefore, we will give some information related to the size of the interaction effect. The results for the power analysis are given in tables 1 and 2. The two ncps for the moment and the least squares test are given (column 2 and 3) for different interaction sizes. The moment test and the least squares test have degrees of freedom 1 and 8, respectively.

Column 4 gives the theoretical power based on the ncp and the degree of freedom of the LS test. The theoretical test is based on a “known” matrix Γ. This matrix is not really known, however it has been computed by using a simulation with a sample of size $100,000$. The empirical power is computed as the proportion of rejections of the model without interaction factors; see column 5 of tables 1 and 2.

Table 1: Power when using $V1V3V5$ and the model with $\beta_{12} = 0$

β_{12}	$\lambda_M(1)$	$\lambda_{LS}(8)$	powTh in %	powEmp in %
.0	0	0	5.0	4.4
.1	1.472	1.472	10.7	9.2
.2	5.264	5.268	31.3	31.2
.4	14.607	14.618	78.7	75.6
.7	24.711	24.746	96.7	97.0

Table 2: Power when using $V5V5V5$ and the model with $\beta_{12} = 0$

β_{12}	$\lambda_M(1)$	$\lambda_{LS}(8)$	powTh in %	powEmp in %
.1	2.215	2.217	14.2	14.0
.2	5.752	5.777	34.3	32.4
.4	8.204	8.359	49.7	42.6*
.7	7.360	7.571	45.1	38.4*

* Estimate is outside the 95% confidence interval

From tables 1 and 2 it follows that indeed the noncentrality parameters are (about) equal for the two tests. The consequence is that the theoretical power can be computed easily from the moment test and no model parameters have to be estimated first. Further, it is remarkable that for the third order moment V5V5V5 the noncentrality parameter does not increase monotonically with the interaction parameter as it should be. An explanation for this deviation from monotonicity is given in a section below.

3 Selection of third-order moments and the power

Consider the expression (4) of the WLS model test with $\sigma = (\sigma_{12}', \sigma_3')'$ where σ_{12} contains the first-and second-order moments and σ_3 the third-order moments only. By considering the vector of parameters partitioned as $\theta' = (\theta_1', \theta_3')$, where θ_3 is the vector of parameters representing interaction terms, we have

$$\dot{\sigma} = \begin{pmatrix} \dot{\sigma}_{12,1} & \dot{\sigma}_{12,3} \\ \dot{\sigma}_{3,1} & \dot{\sigma}_{3,3} \end{pmatrix}$$

where $\dot{\sigma}_{12,1}$ and $\dot{\sigma}_{3,1}$ represent the Jacobian of σ_{12} and σ_3 respectively with respect to θ_1 and $\dot{\sigma}_{12,3}$ and $\dot{\sigma}_{3,3}$ represent the Jacobian of σ_{12} and σ_3 respectively with respect to the interaction term parameters θ_3.

Note that model H_0 does not specify an interaction parameters and so $\theta = \theta_1$ thus θ_3 is vacuous. In spite of that, H_0 uses higher-order moments in the analysis, and so σ_3 is present in the moment vector to be modeled. Thus, the Jacobian matrix associated to the model H_0 is

$$\dot{\sigma}_{|H_0} = \begin{pmatrix} \dot{\sigma}_{12,1} \\ \dot{\sigma}_{3,1} \end{pmatrix}$$

Furthermore, since under H_0 it holds $\sigma_3(\theta) = 0$, whatever the combination of the values of the parameter vector θ_1, we get $\dot{\sigma}_{3,1} = 0$, and thus

$$\dot{\sigma}_{|H_0} = \begin{pmatrix} \dot{\sigma}_{12,1} \\ 0 \end{pmatrix} \tag{6}$$

Note that we require this matrix to be of full column rank for the model to be identified.

Let the alternative model H_1 be the true model with interaction parameter $\theta_3 \neq 0$. Still consider the analysis using the specification by H_0. Then the non-centrality parameter will be

$$\lambda_{WLS}(\sigma_a \mid H_0) = n(\sigma_a - \hat{\sigma}_0)'(\hat{\Gamma}^{-1} - \hat{\Gamma}^{-1}\hat{\dot{\sigma}}(\hat{\dot{\sigma}}'\hat{\Gamma}^{-1}\hat{\dot{\sigma}})^{-1}\hat{\dot{\sigma}}'\hat{\Gamma}^{-1})(\sigma_a - \hat{\sigma}_0) \tag{7}$$

where $\hat{\sigma}_0$ denote the model fitted by H_0 when analyzing the moment vector σ_a. Now it holds in general, provided $\dot{\sigma}_{|H_0}$ is of full column rank,

$$\lambda_{WLS}(\sigma_a \mid H_0) = n(\sigma_a - \hat{\sigma}_0)'F(F'\Gamma F)^{-1}F'(\sigma_a - \hat{\sigma}_0) \tag{8}$$

where F is an orthogonal complement of the matrix $\dot{\sigma}_{|H_0}$, that is $F'\dot{\sigma}_{|H_0} = 0$.

We will set up conditions under which the models H_0 and H_1 are equivalent for first- and second-order moments. More specifically, when considering a moment vector σ_a generated according to H_1, when H_0 is fitted to σ_a producing a fitted moment $\hat{\sigma}$, we will show that $(\hat{\sigma})_{12} = (\sigma_a)_{12}$, the subscript "$_{12}$" denoting first- and second-order moments. The next lemma shows that this is the case for the class of models defined by (1) to (3) satisfying the following Condition 1.

CONDITION 1: *The analyzed moment vector σ_a corresponds to a population satisfying (1) to (3) and the analyzed model H_0 sets unconstrained the intercept vector α, the matrix Γ_1, the product $B^{-1}\Psi B^{-T}$ and Φ (aside from its symmetry).*

The following Lemma 1 gives a fundamental result for this paper (the lemma is proved in the Appendix).

Lemma 1: *Under Condition 1 assume that there is un-correlation among ξ, δ, ζ and ϵ, and ξ is normally distributed. Consider the WLS-fit of H_0 and H_1 to all the first- and second-order moments plus some or all third-order moments. Assume the weight matrix W is block diagonal with respect the vectors s_{12} of first and second order moments and the vector s_3 of third-order moments. Let $\sigma_a = (\sigma'_{12a}, \sigma'_{3a})'$ be the analyzed moment vector (here σ_{3a} is the vector of third order moments only) and let $\hat{\sigma}$ be the fitted vector when fitting H_0 to σ_a. Then*

$$(\hat{\sigma})_{12} = (\sigma_a)_{12},$$

where " $()_{12}$ " denotes first and second-order moments only. ■

Note that the conclusion of the Lemma can be written as

$$(\sigma_a - \hat{\sigma}_0) = \begin{pmatrix} 0 \\ \sigma_{a3} \end{pmatrix}$$

Note that Lemma 1 implies that when Condition 1 is verified there will be an exact WLS fit for the first- and second-order moments in any WLS analysis with $W =$ block-diagonal$(W_{12,12}, W_{3,3})$ where $W_{12,12}$ a partition conformably with $\sigma = (\sigma_{12}', \sigma_3')')$. [1] Lemma 1 in fact establishes that for the class of models and assumptions considered, when σ_a satisfies exactly H_1 and the fitted model iis H_0, then $\hat{\sigma}_{12}$ fits exactly the first- and second-order moments of σ_a.

Given the form (6), we have that

$$F' = \begin{pmatrix} G' & 0 \\ 0 & I \end{pmatrix}$$

with $G'\dot{\sigma}_{12,1} = 0$. Consequently it is easy to prove by some matrix manipulation and by using the inverse of partitioned matrices that

$$\lambda_{WLS}(\sigma_a \mid H_0) = n\sigma_3'(\hat{\Gamma}_{3,3}^{-1} - \hat{\Gamma}_{3,12}G(G'\hat{\Gamma}_{12,12}G)^{-1}G'\hat{\Gamma}_{12,3})^{-1}\sigma_3 \qquad (9)$$

[1] In that case, since the third-order fitted moments under H_0 are zero,

$$(s - \sigma(\theta))'W(s - \sigma(\theta)) = (s_{12} - \sigma_{12}(\theta))'W_{12,12}(s_{12} - \sigma_{12}(\theta)) + s_3'W_{33}s_3,$$

where we assumed that $\hat{\Gamma}_{3,3}$ is nonsingular. Here we considered the partitioning of $\Gamma = \text{avar}\,(s)$ as

$$\Gamma = \begin{pmatrix} \Gamma_{12,12} & \Gamma_{12,3} \\ \Gamma_{3,12} & \Gamma_{3,3} \end{pmatrix}$$

where $\Gamma_{3,12}$ is the asymptotic covariance matrix between the first-, second- and third-order moments, that is $\Gamma_{12,3} = \text{cov}\,(s_{12}, s_3)$ and $\Gamma_{12,12}$ and $\Gamma_{3,3}$ are the asymptotic variance matrices of s_{12} and s_3, respectively.

Clearly, the corresponding noncentrality parameter for the moment test defined in (5) is

$$\lambda_{MT}(\sigma_a) = n\sigma'_{a3}\Gamma^{-1}_{3,3}\sigma_{a3} \tag{10}$$

and, comparing (9) and (10) , we see that

$$\lambda_{MT}(\sigma_a) = \lambda_{WLS}(\sigma_a \mid H_0) \quad \text{iff} \quad G'\Gamma_{12,3} = 0 \tag{11}$$

We will now show that under standard regularity conditions on the model and on the distribution of the latent constituents, and the block diagonal structure of W, the condition of $G'\Gamma_{12,3} = 0$ do holds. For that we require to write the models H_1 and H_0 as a linear latent variable model (e.g., Satorra, 1992). Clearly H_1 can be written as

$$y = \nu_y + \Lambda_y B^{-1}\alpha + \Lambda_y B^{-1}\Gamma_1\xi + \Lambda_y B^{-1}\Gamma_2(\xi \otimes \xi) + \Lambda_y B^{-1}\zeta + \epsilon_y \tag{12}$$

$$x = \nu_x + \Lambda_x \xi + \epsilon_x \tag{13}$$

By setting $z = (y', x')'$ and $\epsilon = (\epsilon_y', \epsilon_x')'$,

$$z = \mu + [\Lambda_2 \zeta + \epsilon] + [\Lambda_1 \xi + \Lambda_3 (\xi \otimes \xi)]$$

where

$$\mu = \begin{pmatrix} \nu_y + \Lambda_y B^{-1} \alpha \\ \nu_x \end{pmatrix}, \Lambda_1 = \begin{pmatrix} \Lambda_y B^{-1} \Gamma_1 \\ \Lambda_x \end{pmatrix}, \Lambda_2 = \begin{pmatrix} \Lambda_y B^{-1} \\ 0 \end{pmatrix}$$

and

$$\Lambda_3 = \begin{pmatrix} \Lambda_y B^{-1} \Gamma_2 \\ 0 \end{pmatrix}$$

Thus, in compact expression, H_1 is

$$z = \mu + A\delta = \mu + A_1 \delta_1 + A_2 \delta_2 \tag{14}$$

where $A = (A_1, A_2)$, $\delta = (\delta_1', \delta_2')'$, $\delta_1 = (\zeta', \epsilon')'$, $\delta_2 = (\xi', (\xi \otimes \xi)')'$, $A_1 = (\Lambda_2, I)$ and $A_2 = (\Lambda_1, \Lambda_3)$. where $B = I - B_0$ and B is assumed to be invertible. The null hypothesis H_0 of no interaction terms can now be expressed as $\Lambda_3 = 0$.

We need now the following distributional assumption of symmetry (SI) on stochastic components of the linear structure (14) implied by the model assumption H_1.

Assumption SI: *The model H_1 holds and the distribution of δ_1 of (14)* *is symmetric and independent of* δ_2.

Consider now the specification of a LISREL model H_0 like (1) to (3) with Γ_2 set to 0, i.e. H_0 does not contemplate interaction terms. Consider the vector of first- and second-order moments for z, $\sigma_{12} = (\sigma_1', \sigma_2')'$, where $\sigma_1 = E[z]$ and $\sigma_2 = \text{vech}\, E[(z-\mu)\otimes(z-\mu)]$. Clearly, under the specification H_0, the vector σ_{12} is structured as a function $\sigma_{12} = \sigma_{12}(\theta)$ of the vector of parameters θ. Let the parameter vector θ be partitioned as $\theta = (\theta_1', \theta_2')'$, where $\theta_1 = (\theta_\alpha', \theta_{\Gamma_1}', \theta_\Phi', \theta_{B_0}', \theta_\Psi')'$, θ_α, θ_{Γ_1}, θ_Φ, θ_{B_0} and θ_Ψ denoting the vectors of free parameters associated to the free components in α, Γ_1, Φ, B_0 and Ψ respectively. Assume that H_0 satisfies the following assumption

Functional parameter independence (FPI) assumption: *The parameter vectors* θ_α θ_{Γ_1}, θ_Φ, θ_{B_0} *and* θ_Ψ *are functionally independent (no cross constraints among them are allowed).*

Consider the partitioned Jacobian

$$\dot\sigma_{12} = \begin{pmatrix} \dot\sigma_1 \\ \dot\sigma_2 \end{pmatrix}$$

where $\dot\sigma_1 = \partial\sigma_1/\partial\theta_1'$ and $\dot\sigma_2 = \partial\sigma_2/\partial\theta_1'$. Clearly,

$$\dot\sigma_j = \frac{\partial\sigma_j}{\partial\alpha' \quad \partial(\text{vec}\,\Gamma_1)' \quad \partial(\text{vech}\,\Phi)' \quad \partial(\text{vec}\,B_0)' \quad \partial(\text{vech}\,\Psi)'} R\,, \quad j = 1, 2$$

where, in virtue of FPI,

$$R = \text{block-diagonal}\left[R_\alpha, R_{\Gamma_1}, R_\Phi, R_{B_0}, R_\Psi\right],$$

$R_\alpha = \partial\alpha/\partial\theta'_\alpha$, $R_{\Gamma_1} = \partial\text{vec}\,(\Gamma_1)/\partial\theta'_{\Gamma_1}$, $R_\Phi = \partial\text{vech}\,(\Phi)/\partial\theta'_\Phi$, $R_{B_0} = \partial\text{vec}\,(B_0)/\partial\theta'_{B_0}$ and $R_\Psi = \partial\text{vech}\,(\Psi)/\partial\theta'_\Psi$. Further, by differentiation it can easily be seen that

$$\dot{\sigma}_{12} = \begin{pmatrix} A_{11} & 0 & 0 & A_{14} & 0 \\ 0 & A_{22} & A_{23} & A_{24} & A_{25} \end{pmatrix}$$

where

$$A_{11} = \frac{\partial\sigma_1}{\partial\alpha'} = \begin{pmatrix} \Lambda_y B^{-1} \\ 0 \end{pmatrix} R_\alpha$$

$$A_{14} = \frac{\partial\sigma_1}{\partial(\text{vec}\,B_0)'} = \begin{pmatrix} (\alpha' \otimes \Lambda_y)(B^{-1} \otimes B^{-1}) \\ 0 \end{pmatrix} R_{B_0}$$

$$A_{22} = \frac{\partial\sigma_2}{\partial(\text{vec}\,\Gamma_1)'} = D^+_{p+q}(\Lambda_1\Phi \otimes \Lambda_2)R_{\Gamma_1}$$

$$A_{23} = \frac{\partial\sigma_2}{\partial(\text{vech}\,\Phi)'} = 2D^+_{p+q}(\Lambda_1 \otimes \Lambda_1)D_n R_\Phi$$

$$A_{24} = \frac{\partial\sigma_2}{\partial l(\text{vec}\,B_0)'} = D^+_{p+q}\left[(\Lambda_1 \otimes \Lambda_2)(\Phi\Gamma' B^{-1} \otimes I_m) + (\Lambda_2 \otimes \Lambda_2)(\Psi B^{-1} \otimes I_m)\right] R_{B_0}$$

$$A_{25} = \frac{\partial\sigma_2}{\partial(\text{vech}\,\Psi)'} = 2D^+_{p+q}(\Lambda_2 \otimes \Lambda_2)D_m R_\Psi$$

Below we will assume that the vector and matrices α, Γ_1 and Φ are unrestricted so then R_α, R_{Γ_1} and R_Φ are identity matrices.

Let G be a matrix orthogonal to $\dot{\sigma}_{12}$, that is, $G'\dot{\sigma}_{12} = 0$, and partition it as $G' = (G_1', G_2')$, so that we have $G_1'\dot{\sigma}_1 + G_2'\dot{\sigma}_2 = 0$. Because only the means of the y variables are functions of some model parameters it makes sense to define the following partitioning $G_1 = (G_{1y}, G_{1x})$. Then we have the following equations in which the means are involved:

$$G_{1y}'\Lambda_y B^{-1} R_\alpha = 0 \tag{15}$$

$$G_{1y}'(\alpha' \otimes \Lambda_y)(B^{-1} \otimes B^{-1})R_{B_0} + G_2' A_{24} = 0 \tag{16}$$

Under the assumption that α is unconstrained, R_α is the identity and thus $G_{1y}'\Lambda_y B^{-1} = 0$; so, it follows

$$G_{1y}'(\alpha' \otimes \Lambda_y)(B^{-1} \otimes B^{-1}) = G_{1y}'(\alpha' B^{-T} \otimes \Lambda_y B^{-1}) = 0$$

This equality to zero follows from noting that $\alpha' B^{-T}$ is a row vector, so $\alpha' B^{-T} \otimes \Lambda_y B^{-1}$ consists of scalars times $\Lambda_y B^{-1}$. Thus equation (16) results in $G_2' A_{24} = 0$, which is (19) below.

Clearly, the equations in which the covariances are involved are the following ones:

$$G_2'\, D_{p+q}^{+}(\Lambda_1 \Phi \otimes \Lambda_2) R_{\Gamma_1} = 0 \tag{17}$$

$$G_2'\, D_{p+q}^{+}(\Lambda_1 \otimes \Lambda_1) D_n R_\Phi = 0 \tag{18}$$

$$G_2' D_{p+q}^{+} \left[(\Lambda_1 \otimes \Lambda_2)(\Phi\Gamma' B^{-1} \otimes I_m) + (\Lambda_2 \otimes \Lambda_2)(\Psi B^{-1} \otimes I_m)\right] R_{B_0} = 0 \tag{19}$$

$$G_2'\, D_{p+q}^{+}(\Lambda_2 \otimes \Lambda_2) D_m R_{\Psi} = 0 \tag{20}$$

Lemma 2 *Consider the specification H_0 under the separability assumption FPI. Assume Condition 1 with Φ and Ψ of full rank. Then*

$$G_2' D_{p+q}^{+}[(\Lambda_1, \Lambda_2) \otimes (\Lambda_1, \Lambda_2)] = 0$$

where

$$\Lambda_1 = \begin{pmatrix} \Lambda_y B^{-1} \Gamma_1 \\ \Lambda_x \end{pmatrix} \quad \textit{and} \quad \Lambda_2 = \begin{pmatrix} \Lambda_y B^{-1} \\ 0 \end{pmatrix}.$$

$G' = (G_1', G_2')$, conformably with the matrix product above, and G any matrix that $G'\dot{\sigma}_{12} = 0$.

Proof: Using Lemma 6 of the Appendix, it holds

$$G_2' D_{p+q}^{+}[(\Lambda_1, \Lambda_2) \otimes (\Lambda_1, \Lambda_2)] = G_2' D_{p+q}^{+}[[(\Lambda_1 \otimes \Lambda_1), (\Lambda_2 \otimes \Lambda_1), (\Lambda_1 \otimes \Lambda_2), (\Lambda_2 \otimes \Lambda_2)] E$$

where E is a permutation matrix (square and of full rank). So for proving the Lemma it suffices to show that $G_2' D_{p+q}^{+}(\Lambda_i, \Lambda_j) = 0$ for $i, j = 1, 2$.

Since $G_2'\, D_{p+q}^{+}(\Lambda_1 \Phi \otimes \Lambda_2) = G_2'\, D_{p+q}^{+}(\Lambda_1 \otimes \Lambda_2)(\Phi \otimes I)$, using (17), the non-singularity of Φ, and Γ_1 unrestricted (so that $R_{\Gamma_1} = I$), yields

$$G_2'\, D_{p+q}^{+}(\Lambda_1 \otimes \Lambda_2) = 0 \tag{21}$$

Since $G_2'\, D_{p+q}^{+}(\Lambda_1 \otimes \Lambda_2) = G_2' D_{p+q}^{+} K_{p+q,p+q} (\Lambda_2 \otimes \Lambda_1) K_{n,m} = 0$ (the Ks are commutation matrices). Because the commutation matrix is square nonsingular and $D^+_{p+q}K_{p+q,p+q} = D^+_{p+q}$ it follows $G'_2\, D^+_{p+q}(\Lambda_1 \otimes \Lambda_2) = G'_2\, D^+_{p+q}(\Lambda_2 \otimes \Lambda_1)K_{n,m}$, so we prove

$$G'_2\, D^+_{p+q}(\Lambda_2 \otimes \Lambda_1) = 0 \tag{22}$$

Since Φ is symmetric and unrestricted, (18) implies $G'_2\, D^+_{p+q}(\Lambda_1 \otimes \Lambda_1)D_n = 0$, and so $G'_2\, D^+_{p+q}(\Lambda_1 \otimes \Lambda_1)D_n D^+_n = 0$. Further, since $D^+_{p+q}(\Lambda_1 \otimes \Lambda_1)D_n D^+_n = D^+_{p+q}(\Lambda_1 \otimes \Lambda_1)N_n$, where $N_n = \frac{1}{2}(I + K_n)$ and K_n is a commutation matrix (see Magnus and Neudecker, 2000). Because it holds $D^+_{p+q}(\Lambda_1 \otimes \Lambda_1)K_n = D^+_{p+q}K_{p+q}(\Lambda_1 \otimes \Lambda_1) = D^+_{p+q}(\Lambda_1 \otimes \Lambda_1)$, since $(\Lambda_1 \otimes \Lambda_1)K_n = K_n(\Lambda_1 \otimes \Lambda_1)$ and $D^+_{p+q}K_{p+q} = D^+_{p+q}$ (see, e.g., Theorem 7.37 of Schott, 1997), it follows $D^+_{p+q}(\Lambda_1 \otimes \Lambda_1)N_n = D^+_{p+q}(\Lambda_1 \otimes \Lambda_1)$ and thus

$$G'_2\, D^+_{p+q}(\Lambda_1 \otimes \Lambda_1) = 0 \tag{23}$$

From (21) it follows that the first term in (19) is 0. Combining this result and (20) it holds

$$G'_2\, D^+_{p+q}(\Lambda_2 \otimes \Lambda_2)[(\Psi B^{-1} \otimes I_m)R_{B_0}, D_m R_\Psi] = 0$$

Define $Z = [(\Psi B^{-1} \otimes I_m)R_{B_0}, D_m R_\Psi] = [Z_1 R_{B_0}, Z_2 R_\Psi]$, then this can be written as $Z = (\Psi B^{-1} \otimes I_m, D_m)R_{B_0,\Psi}$ where $R_{B_0,\Psi} = \begin{pmatrix} R_{B_0} & 0 \\ 0 & R_\Psi \end{pmatrix}$. Let $H = B^{-1}\Psi B^{-T}$, then H has $m(m+1)/2$ different non-duplicated elements. Now it holds that H is completely unrestricted if the Jacobian of H w.r.t.

the parameters has $m(m+1)/2$ columns and is of full column rank. This Jacobian can be written as:

$$\frac{\partial \text{vec}\,(H)}{\partial[(\text{vec}\,(B_0))', (\text{vech}\,(\Psi))']} = (Z_1 R_{B_0}, Z_2 R_\Psi) = (Z_1, Z_2) \begin{pmatrix} R_{B_0} & 0 \\ 0 & R_\Psi \end{pmatrix} = Z R_{B_0,\Psi}$$

where Z is of full column rank. So the condition for un-restrictedness of H is that $R_{B_0,\Psi}$ is of full column rank equal to $m(m+1)/2$. That is, $H = B^{-1}\Psi B^{-T}$ free is equivalent to Z being of full column rank, this rank being equal to $m(m+1)/2$. Two typical conditions under which this holds, is when Ψ is a diagonal matrix with unconstrained elements $(B_0)_{ij}$, or when Ψ is an unconstrained free matrix and $(B_0)_{ij}$ constrained.

Now from $G_2'\, D_{p+q}^+(\Lambda_2 \otimes \Lambda_2) Z = 0$ it follows $G_2'\, D_{p+q}^+(\Lambda_2 \otimes \Lambda_2) D_m D_m^+ Z = G_2'\, D_{p+q}^+(\Lambda_2 \otimes \Lambda_2) D_m = 0$ and so

$$G_2'\, D_{p+q}^+(\Lambda_2 \otimes \Lambda_2) = 0, \tag{24}$$

completing the proof of the lemma. ∎

3.1 The covariance matrix $\Gamma_{12,3}$

We now elaborate on the expression for the covariance matrix $\Gamma_{12,3}$ among the vector of first and second-order moments and the vector of third-order moments. Under H_1, we have $z = \mu + A\delta$, where δ is partitioned as $\delta = (\delta_1', \delta_2')'$, δ_2 containing the main factors and the interaction/quadratic factors, and δ_1 the rest of the factors (errors and disturbances).

As argued above, under H_1,

$$z = \mu + A\delta = \mu + A_1\delta_1 + A_2\delta_2 = \mu + (\Lambda_2\zeta + \epsilon) + (\Lambda_1\xi + \Lambda_3(\xi \otimes \xi))$$

where matrix Λ_3 consists of regression weights of interaction and/or quadratic terms of the ξ - variables. Obviously, H_0 is defined by $\Lambda_3 = 0$. where the δ_1 and δ_2 are independent of each other, and matrix A is partitioned as $A = (A_1, A_2)$.

The following lemma is needed:

Lemma 3: *Under the notation of model H_1 of (14), assume the hypothesis SI holds, then*

$$\Gamma_{z,1,23} = D^+_{p+q}(A_2 \otimes A_2)D\Gamma_{\delta_2,1,23}T'(A_2 \otimes A_2 \otimes A_2)'T^{+\prime}_p$$

where $A_2 = (\Lambda_1, \Lambda_2\Gamma_2)$.

Proof: Given the form $z = \mu + A_1\delta_1 + A_2\delta_2$ for z implied by H_1, it holds

$$\begin{aligned}\Gamma_{z,1,23} &= D^+_{p+q}(A_1 \otimes A_1)D\Gamma_{\delta_1,1,23}T'(A_1 \otimes A_1 \otimes A_1)'T^{+\prime}_p \\ &+ D^+_{p+q}(A_2 \otimes A_2)D\Gamma_{\delta_2,1,23}T'(A_2 \otimes A_2 \otimes A_2)'T^{+\prime}_p \qquad (25)\end{aligned}$$

where $\Gamma_{\delta_i,1,23}$ for $i = 1, 2$ is the covariance matrix of the first-, second-order and third-order moments of δ_1 and δ_2 respectively. Because δ_1 has a symmetric distribution, $\Gamma_{\delta_1,1,23} = 0$, and so the first term on the right hand side of (25) is equal to 0. ■

Another crucial lemma, to reach the basic theorem for the paper, is the following.

Lemma 4: *Under the notation above, assume*

1. *Condition 1*

2. $G'\dot{\sigma}_{12} = 0$

then $G'\Gamma_{z,1,23} = 0$

PROOF:

$$D^+_{p+q}(A_2 \otimes A_2) = D^+_{p+q}\left[(\Lambda_1, \Lambda_3) \otimes (\Lambda_1, \Lambda_3)\right]$$

$$= D^+_{p+q}\left[(\Lambda_1, \Lambda_2\Gamma_2) \otimes (\Lambda_1, \Lambda_2\Gamma_2)\right]$$

$$= D^+_{p+q}\left[(\Lambda_1, \Lambda_2)\begin{pmatrix} I & 0 \\ 0 & \Gamma_2 \end{pmatrix} \otimes (\Lambda_1, \Lambda_2)\begin{pmatrix} I & 0 \\ 0 & \Gamma_2 \end{pmatrix}\right]$$

$$= D^+_{p+q}\left[((\Lambda_1, \Lambda_2) \otimes (\Lambda_1, \Lambda_2))\left(\begin{pmatrix} I & 0 \\ 0 & \Gamma_2 \end{pmatrix}\begin{pmatrix} I & 0 \\ 0 & \Gamma_2 \end{pmatrix}\right)\right]$$

Now if Φ, $B^{-1}\Psi B^{-T}$ are unrestricted matrices, then because Lemma 2, we have $G'D^+_p(A_2 \otimes A_2) = 0$. ■

Let $\tilde{\Lambda}_1$ and $\tilde{\Lambda}_2$ be matrices computed with the WLS estimates obtained when fitting the null model H_0, keeping the notation Λ_1 and Λ_2 for the parameters corresponding to true values.

Lemma 5: *Under the same conditions as in Lemma 2; if*

$$G_2' D_{p+q}^{+}[(\tilde{\Lambda}_1, \tilde{\Lambda}_2) \otimes (\tilde{\Lambda}_1, \tilde{\Lambda}_2)] = 0$$

then

$$G_2' D_{p+q}^{+}[(\Lambda_1, \Lambda_2) \otimes (\Lambda_1, \Lambda_2)] = 0$$

PROOF: From the equivalence of H_0 and H_1, it holds $B^{-1}\Gamma_1 = \tilde{B}^{-1}\tilde{\Gamma}_1$ and $\tilde{B}^{-1}\tilde{\Psi}\tilde{B}^{-T} = B^{-1}(\Psi + Q)B^{-T}$ where $Q = \Gamma_2 \mathrm{cov}\,(\xi \otimes \xi)\Gamma_2'$. So because $B^{-1}\Gamma_1 = \tilde{B}^{-1}\tilde{\Gamma}_1$ it follows immediately $\Lambda_1 = \tilde{\Lambda}_1$. Furthermore, it is easy to prove that $\tilde{\Lambda}_2 = \Lambda_2 V$, where $V = (\Psi + Q)B^{-1}\tilde{B}'\tilde{\Psi}^{-1}$, which is non-singular in general. So

$$(\tilde{\Lambda}_1, \tilde{\Lambda}_2) \otimes (\tilde{\Lambda}_1, \tilde{\Lambda}_2) = (\Lambda_1, \Lambda_2 V) \otimes \Lambda_1, \Lambda_2 V)$$

$$= [(\Lambda_1, \Lambda_2) \otimes (\Lambda_1, \Lambda_2)] \left[\begin{pmatrix} I & 0 \\ 0 & V \end{pmatrix} \otimes \begin{pmatrix} I & 0 \\ 0 & V \end{pmatrix} \right]$$

$$= [(\Lambda_1, \Lambda_2) \otimes (\Lambda_1, \Lambda_2)]\, W$$

where $W = \begin{pmatrix} I & 0 \\ 0 & V \end{pmatrix} \otimes \begin{pmatrix} I & 0 \\ 0 & V \end{pmatrix}$, which implies the conclusion of the theorem since W is non-singular. ■

From here we can move to prove the main Theorem.

Theorem 1: *Under the conditions of Lemma 4,*

$$\lambda_{WLS}(\sigma_0 \mid H_0) = \lambda_{MT}(\sigma_0)$$

PROOF: This follows from (11) and Lemma 4 . ■

4 Types of third-order moments on the bases of power function

We have seen that the form of the power function changes with the type of third-order moment added to the analysis. In fact, we have seen that for some cases, the power function can be constant, when the third order moment does not involve the misspecified parameter, monotonic increasing as in the case of V1V3V5, or non-monotonic, as in the case of V5V5V5. On the base of this behaviour of the power function we will classify the third-order moments in C, MI, NMI. Given a model, it will be interesting to disentangle the power status for the different higher order moment in relation to a given interaction term.

Here we discuss examples of each. We will express the noncentrality parameter (ncp) in terms of model parameters to see the nature of the different type of third order moments.

We concentrate here on the interaction model and fitting, besides means and covariances of the variables, third order moments. We further assume that the null hypothesis is that the interaction term, $\beta_{12} = 0$. In this section we concentrate on the moment test, and more particularly on the univariate moment test.

To better illustrate the issue we consider a simplified version of the model. The model: there are two observed independent variables, called x_1 and x_2, and and one dependent variable, called y. There are three different types of third order for the latent variables: $\mu_{yx_1x_2}$, $\mu_{y^2x_1}$, $\mu_{y^2x_2}$ and μ_{y^3}. Corresponding to these third order moments there are three different types of ncp's. These can be written as:

$$ncp(1) = \frac{(\mu_{yx_1x_2})^2}{\text{var}(m_{yx_1x_2})}, \quad ncp(2) = \frac{(\mu_{y^2x_1})^2}{\text{var}(m_{y^2x_1})}, \quad ncp(3) = \frac{(\mu_{y^3})^2}{\text{var}(m_{y^3})} \tag{26}$$

The regression model is written as :

$$y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + e$$

where it is assumed that x and e are centered variables. Because in the third order moments we discuss in this paper, the means of the observed dependent variables are zero, we write the regression equation as

$$y = y^* - E[y^*] = \beta_1 x_1 + \beta_2 x_2 + \beta_{12}(x_1 x_2 - \phi_{12}) + e \tag{27}$$

It can be shown that the different types of third order moments can be written as:

$$
\begin{aligned}
\text{type}(1) : \mu_{yx_1x_2} &= \beta_{12}(\phi_{11}\phi_{22} + \phi_{12}) \\
\text{type}(2) : \mu_{y^2x_k} &= 2\beta_{12}\left(2\beta_1\phi_{kk}\phi_{12} + \beta_2(\phi_{11}\phi_{22} + \phi_{12}^2)\right), k = 1, 2 \\
\text{type}(3) : \mu_{y^3} &= 6\left(\beta_1^2\phi_{11}\phi_{12} + \beta_2^2\phi_{22}\phi_{12} + \beta_1\beta_2(\phi_{11}\phi_{22} + \phi_{12}^2)\right)\beta_{12} + (6\phi_{11}\phi_{12}\phi_{22} + 2\phi_{12}^3)\beta_{12}^3
\end{aligned}
$$

where the ϕs are the covariances among the x's. These expressions have been deduced assuming bivariate normality for the variables x_1 and x_2.

Obviously it holds for all moments that the third-order moment are 0 if the interaction parameter β_{12} is zero. An important remark to make here is that the third-order moments of type 1 and 2 are linear functions of the interaction parameter β_{12}, whereas for type 3 this function is a nonlinear function.

Example: In this example we take as model parameters the same model parameters as in the structural part of the Kenny and Judd model. This means that the measurement errors are not involved in our model. So the parameters are $\beta_0 = 1$, $\beta_1 = .2$, $\beta_2 = .4$ and $\text{var}(e) = .2$.

The interest of this example is the influence of the interaction parameter (β_{12}) on the size of the noncentrality parameter. Unfortunately, there is no analytical expression for the ncp's in terms of the model parameters, because the denumerator is hard to express in terms of the model parameters. For instance, it is easy to verify that for type 3 third-order moments the variance of the third-order moments depends on moments of order 12.

A small MC study is carried out. In this study $100,000$ samples with sample size 600 are drawn from a population which is specified by the model parameters as given above. Table 3 gives the results of this study for two different third-order moments (x_1x_2y and y^3) for different values of the interaction parameters.

Table 3: Monte Carlo values for the mean and variance of two types of third-order moments [1]

	moment x_1x_2y				moment y^3			
β_{12}	μ	m	γ	ncp	μ	m	γ	ncp
.0	.000	.000	.400	.000	.000	.000	.700	.000
.1	.037	.037	.417	1.939	.035	.035	.794	.939
.2 .	074	.074	.469	6.942	.074	.073	1.144	2.823
.3	.111	.110	.554	13.250	.117	.117	1.878	4.385
.4	.148	.147	.680	19.222	.170	.169	3.339	5.160
.5	.184	.184	.839	24.321	.233	.233	5.962	5.457
.6	.221	.221	1.032	28.518	.311	.311	10.507	5.524
.7	.258	.258	1.261	31.715	.405	.405	17.965	5.486
.8	.295	.295	1.504	34.729	.519	.519	29.570	5.457
.9	.332	.332	1.812	36.467	.656	.655	47.731	5.391
1.0	.369	.369	2.156	37.904	.818	.818	75.047	5.353

[1] Note that γ is defined as the sample size (600) times the variance of the third-order moment.

We now summarize the main findings in Table 3. i) The Monte Carlo mean of the third-order moments given in column m over all MC trials are about equal to the expectations given in column μ matching what should be expected from theory; ii) The ncp's for the type-1 third-order moments are always (substantially) larger than for the type-3 third-order moment;

iii) The variance of the moments increases (and so γ too) if the interaction parameter increases, although this variance increases sharper for the third-order moment y^3; and iv) When the interaction effect increases, the ncp associated to x_1x_2y increases also. This is not the case for y^3; we see that the ncp do in fact decreases when β_{12} is larger than .6 This non-linear relation between the size of the interaction term and the ncp was noted above for the type (3) third-order moment.

Point iv) above is the most important issue because it is counter-intuitive. The reason is the variance of the third-order moment (which is the denumerator of the ncp) increases sharply if the interaction parameter increases. The consequence of this is that the ncp may be decreasing while the interaction term is increasing. This explains why in Table 2 where a type (3) third-order moment was fitted, the power was decreasing for large interactions.

5 A forward selection procedure

In Table 4 the results for our forward-selection procedure of the third-order moments are given. The model as in Section 2 is used. The results are given for the true parameter $\beta_{12} = .4$. In the procedure first the third-order moment is looked for which has the highest estimate of the noncentrality parameter, or to put it differently, which has the highest power. Then subsequently, a second- and third-order moment is looked for which, together with the first one, results in the highest power. Table 4 gives in column 1 the third-order moment, then in column 2 and 3 the univariate noncentrality parameter (univ-ncp) and the multivariate noncentrality parameter (mult-ncp). This latter ncp is based on the third order moment in the same row and the third-order moments in all previous rows. In addition, in column 4 to 6, the bias, standard deviation and standard errors of the interaction parameters are given. The number of replications is 250.

Conclusion of Table 4 : The estimates based on the first three third-order moments compare well (in terms of bias and standard errors) to the estimates based on all third-order moments. An importation observation to make is that the gains in terms of efficiency of estimates are large when adding the three third-order moments; however, adding additional third-order moments do not yield improvement on the estimates of the interaction effect. It can be conjecture that each higher-order moment that is added to the analysis deteriorates the robustness against small sample sizes. The last column of the table shows the power values against deviations from zero of the interaction term associated to model test with as many interaction terms as accumulated in the respective row. We see that after adding three interactions terms, adding additional ones do in fact yield a decrease of power. This is in line with the above conclusion that in the model set-up considered, to take more than three third-order moments is not to be recommended.

Table 4: Monte Carlo results for the selection procedure

Moment	univ-ncp	mult-ncp	bias	se($\hat{\beta}$)	sd($\hat{\beta}$)	χ^2	df	Power
V1V3V5	14.675	14.675	0.007	.132	.129	6.81	7	.81
V2V5V5	7.844	16.369	0.000	.109	.102	7.77	8	.84
V4V5V5	8.201	17.068	-0.006	.098	.094	8.59	9	.84
V1V4V5	10.414	17.283	0.005	.100	.099	9.93	10	.83
V2V3V5	7.891	17.445	-0.005	.098	.095	10.80	11	.82
V1V5V5	12.875	17.632	0.003	.098	.102	12.00	12	.81
V1V1V5	3.893	17.870	0.005	.099	.095	12.54	13	.80
V5V5V5	8.190	18.095	-0.008	.091	.086	13.54	14	.79
V3V5V5	10.468	18.440	-0.008	.091	.089	14.93	15	.79
V3V3V5	5.282	18.616	-0.006	.091	.094	15.57	16	.78
V1V2V5	3.259	18.684	-0.005	.092	.092	16.10	17	.77
V3V4V5	5.146	18.731	0.003	.095	.092	17.11	18	.76
V2V4V5	5.479	18.760	-0.005	.092	.088	18.75	19	.75
V2V2V5	0.946	18.766	0.002	.095	.091	19.96	20	.74
V4V4V5	2.359	18.770	-0.002	.095	.095	20.69	21	.73

6 Discussion

A central issue in the analysis of models with interaction is the selection of the higher-order moments to be involved in the analysis. Basically, the most relevant moments would be the ones that lead to higher power in detecting the failure of a null model H_0 with no interaction terms when in fact the interaction term is present in the data at hand. In this paper we have addressed the choice of the interactions terms to be used in the analysis. This paper spells out conditions (Condition 1) under which the power of the model test is exactly equal to the power of a moment test that can be easily implemented just from multivariate data analysis without reference to a SEM model. This fact is of importance since it simplifies considerably the practice of choosing higher order moments for the analysis of a model with interaction terms. The ncp $\lambda_{MT}(\sigma_a)$ can now be easily estimated by direct computation of the moment test $T_M = n\sigma'_{a3}\Gamma_{33}^{-1}\sigma_{a3}$ without involving any optimization process or model fit; so, parameter estimates are not needed to be computed as should when estimating the ncp in the form of $\lambda_{WLS}(\sigma_a)$.

For the results of this paper to hold distributional assumptions on the random constituents of the model are required. We use the assumptions that the distribution of the factors involved in the interaction term are in fact normally distributed (in which case the interaction factor itself will not be normally distributed). Other stochastic constituents of the model such as error of measurement and disturbances may however deviate from the normality assumption, though they are also restricted in distribution. In fact, assumption SI requires the disturbance terms and errors (the vector δ_1) to be symmetric and independent (not only uncorrelated) of δ_2 the vector that contains the factors and their interaction.

Note that both moment test and model test coincide on their non-centrality parameter but they will generally have different degrees of freedom, with generally the moment test having the smaller degrees of freedom. This implies that the moment test will have more power than the corresponding model test. This issue is of lower importance in our paper since we are mainly concerned in the variation of the power of the test when we change the choice of third-order moments.

The models encompassed by Theorem 1 are not so restricted as initially one can fair. The basic condition on the model (apart from the distributional assumptions of assumption SI) is that the structural part of the model is saturated, i.e. in the language of Mooijaart and Satorra (2009) the degrees of freedom of the structural part of the model is zero. The measurement part of the model can, however, contain restrictions that make the whole model with a high number regarding the degrees of freedom.

In this paper we have classified the third-order moments in three types depending on the number of times a dependent variable y appears in the third-order moment. In types (1) and (2) (when a variable y appears only once or two times in the third-order moment) the ncp is a monotonic increasing function of the interaction term for any of its value β_{12} (see the expressions (26)), while in the type (3) when the moment is just the product of three y variables, the ncp is just a non-linear function, increasing of course when β_{12} is close to zero. This non-linearity (more specifically, deviation from monotonicity) of the power function associated to type (3) moments, lead us to the recommendation of avoiding those moments when fitting models with interaction terms.

A final issue we want to discuss is how many third-order moments should be supplemented to first- and second-order moments to improve the analysis of a model with interaction terms. By looking at the last column of Table 4 we see that there is an slight improvement in the power of the model test when adding one additional moment to the first one, but the power do in fact decreases with the number of third-order moments added. In that example, if we want to optimize power, adding the two third-order moments $V1V3V5$ and $V2V5V5$ seems the optimal choice. From the computational perspective and robustness against small samples of our analysis it is clearly an advantage fitting a model with limited number of third-order moments. In practice where sample sizes tend to be not too large, each third-order moment that is added to the analysis is likely to induce more bias on parameter estimates and more inaccuracy of the asymptotic results. So even thought theoretically the asymptotic efficiency of estimates increase when the number of degrees of freedom increase, adding more moments may induce a deterioration on measures of accuracy as mean square error of estimates (square of bias plus variance). We therefore recommend researchers to refraining of adding higher-order moments much beyond the strictly necessary for identification purposes. The moment test can be easily integrated into an automatic procedure for assessing the number of third-order moments to be retained in the analysis, since the moment test can help to produce results as the ones in Table 4 without the need of fitting a model. All this aspects, however, even though we believe are of importance they go beyond the scope of the present paper.

7 Appendix : Lemma 6 and proof of Lemma 1

Lemma 6: *Given matrices A, B and C, it holds that*

$$(A, B) \otimes (C, D) = (A \otimes C, A \otimes D, B \otimes C, B \otimes D)E$$

where E is a permutation matrix .

Proof: We use basic properties of the right-kronecker product, namely $(A, B) \otimes C = (A \otimes C, B \otimes C)$ for conformable matrices A, B, C, so that $(A, B) \otimes (C, D) = (A \otimes (C, D), B \otimes (C, D))$. By definition of kronecker products it holds that the columns of $A \otimes (C, D)$ are either $a_i \otimes c_j$ or $a_i \otimes d_k$ where a_i,c_j and d_k are columns of the matrices A, C and D respectively. So $A \otimes (C, D) = (A \otimes C, A \otimes D)E$, where E is an elementary matrix which permutes the columns of $(A \otimes C, A \otimes D)$. Analogously, it can be written $B \otimes (C, D) = (B \otimes C, B \otimes D)E_2$ for a different permutation matrix E_2. So it holds $(A, B) \otimes (C, D) = (A \otimes (C, D), B \otimes (C, D)) = (A \otimes C, A \otimes D, B \otimes C, B \otimes D)E$, where E is a super 2×2 matrix with block diagonal matrices E_1 and E_2. ■

We will now prove Lemma 1 that basically shows that the mean and covariance structure implied by models H_1 and H_0 are equivalent; that is, if σ_{12} is a mean and covariance vector generated by one of the models will be fitted perfectly by the other model.

Consider the class of models defined by (1) to (3) and further elaborated in (12) and (13). For the population means and covariances of the observable variables we can write (here we assume the usual uncorrelation assumption among the basic random constituents of the model)

$$\begin{aligned} E(y) &= \nu_y + \Lambda_y B^{-1}(\alpha + \Gamma_2 E(\xi \otimes \xi)) \\ E(x) &= \nu_x \end{aligned} \tag{28}$$

and

$$\begin{aligned} \text{cov}(y) &= \Lambda_y B^{-1}\Gamma_1 \Phi \Gamma_1' B^{-T} \Lambda_y' + \Lambda_y B^{-1}[\Gamma_2 \text{cov}(\xi \otimes \xi)\Gamma_2' B^{-T}\Lambda_y' \\ &\quad + \Psi] B^{-T}\Lambda_y' + \text{cov}(\epsilon) \\ &= \Lambda_y B^{-1}[\Gamma_1 \Phi \Gamma_1' + Q + \Psi] B^{-T}\Lambda_y' + \text{cov}\,(\epsilon) \\ \text{cov}(x) &= \Lambda_x \Phi \Lambda_y' + \text{cov}(\delta) \\ \text{cov}(x, y) &= \Lambda_x \Phi \Gamma_1' B^{-T} \Lambda_y' \end{aligned} \tag{29}$$

where $Q = \Gamma_2 \text{cov}(\xi \otimes \xi)\Gamma_2'$.

Under the specification H_0, the means and variances are expressed as (we use the conditions of Lemma 1 of the normality assumption on ξ. In fact, we only need the assumption that $E\xi \otimes \xi \otimes \xi = 0$)

$$\begin{aligned} E(y) &= \nu_y + \Lambda_y \tilde{B}^{-1} \tilde{\alpha} \\ E(x) &= \nu_x \end{aligned} \tag{30}$$

and

$$\begin{aligned} \text{cov}(y) &= \Lambda_y \tilde{B}^{-1} \tilde{\Gamma}_1 \tilde{\Phi} \tilde{\Gamma}_1' \tilde{B}^{-T} \Lambda_y' \\ &\quad + \Lambda_y \tilde{B}^{-1} \tilde{\Psi} \tilde{B}^{-T} \Lambda_y' + \text{cov}(\epsilon) \\ \text{cov}(x) &= \tilde{\Lambda_x} \tilde{\Phi} \Lambda_y' + \text{cov}(\delta) \\ \text{cov}(x, y) &= \tilde{\Lambda_x} \tilde{\Phi} \tilde{\Gamma}_1' \tilde{B}^{-T} \Lambda_y' \end{aligned} \tag{31}$$

We have introduced the tildes to indicate different values for the parameters when fitted under model H_0 than when fitted under model H_1.

In the notation above, the tilde makes explicit that models H_1 and H_0 can differ on the matrices B, α, Φ and $\Psi = \text{cov}(\zeta)$, in addition to the difference with respect to matrix Γ_1. Clearly, equivalence of equations (28) and (30) implies

$$\tilde{\alpha} = \tilde{B} B^{-1} (\alpha + \Gamma_2 D^+ E(\xi \otimes \xi))) \tag{32}$$

It follows that if α is unconstrained this equation will not yield necessarily the same constraints for $\tilde{\alpha}$, so a general condition is that $\tilde{\alpha}$ is unconstrained

.

As expressed in the notation used, the vectors ν_x and ν_y and the matrices Λ_y, $\text{cov}(\epsilon)$, $\text{cov}(\delta)$ are the same in both pair of equations, (28) and (29), and (30) and (31).

The equivalence of (29) and (31) implies

$$\tilde{B}^{-1}(\tilde{\Gamma}_1\tilde{\Phi}\tilde{\Gamma}_1' + \tilde{\Psi})\tilde{B}^{-T} = B^{-1}(\Gamma_1\Phi\Gamma_1' + Q + \Psi)B^{-T} \tag{33}$$

The equivalence thus can be ensured by

$$\tilde{\Gamma}_1 = \tilde{B}B^{-1}\Gamma_1$$

$$\tilde{\Lambda_x}\tilde{\Phi} = \Lambda_x\Phi$$

and

$$\tilde{B}^{-1}\tilde{\Psi}\tilde{B}^{-T} = B^{-1}(Q + \Psi)B^{-T}$$

to be unconstrained under model H_0. So, the first- and second-order moments equations for models H_1 and H_0 are equivalent, when the model H_0 unconstraints the vector α and the matrices Γ_1, Φ and $B^{-1}\Psi B^{-T}$, precisely conditions that are ensured by Condition 1 of Lemma 1. ∎

Note that the equality claimed in Lemma 1 is confined to alternative moment vectors σ_a that satisfy exactly the specification H_1 of the moment equations (28) and (29).

References

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). Mahwak, 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. (2007). *Structural equation modeling: a Bayesian approach* . West Sussex: Wiley.

Lee, S.Y., & Zhu, H.T. (2002). Maximum likelihood mestimation of nonlinear structural equation models. *Psychometrika, 67,* 189-210.

Luijben, T.C.W. (1991). Equivalent models in covariance structure analysis. *Psychometrika, 56,* 653-665.

Magnus, J.R., & Neudecker, H. (1999). Matrix differential calculus with applications in statistics and econometrics(2nd edit.). Chichester: John Wiley & Sons.

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.

Meijer, E. (2005). *Matrix algebra for higher order moments. Linear Algebra and its Applications*, 410, 112-134.

Mooijaart, A., & Bentler, P.M. (2010). An alternative approach for non-linear latent variable models. *Structural Equation Modeling, 17*, 357-373.

Mooijaart, A. and A. Satorra (2009), "On insensitivity of the chi-square model test to non-linear misspecification in structural equation models", *Psychometrika, 74* , 443-455

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

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: a unified approach. Psychometrika, 54(1), 131-151.