Moment testing for interaction terms in structutal equation modeling

Mooijaart, A.; Satorra, A.

PSYCHOMETRIKA—VOL. 77, NO. 1, 65–84
JANUARY 2012
DOI: 10.1007/S11336-011-9232-6

MOMENT TESTING FOR INTERACTION TERMS IN STRUCTURAL EQUATION MODELING

AB MOOIJAART

LEIDEN UNIVERSITY

ALBERT SATORRA

UNIVERSITAT POMPEU FABRA AND BARCELONA GSE

Starting with Kenny and Judd (Psychol. Bull. 96:201–210, 1984) several methods have been introduced for analyzing models with interaction terms. 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 are aiming to adding just a selection of the third-order moments. The key issue in this paper is to develop theoretical results that will allow practitioners to evaluate the strength of different third-order moments in assessing interaction terms of the model. To select the third-order moments, we propose to be guided by the power of the goodness-of-fit test of a model with no interactions, which varies with each selection of third-order moments. A theorem is presented that relates the power of the usual goodness-of-fit test of the model with the power of a moment test for the significance of third-order moments; the latter has the advantage that it can be computed without fitting a model. The main conclusion is that the selection of third-order moments can be based on the power of a moment test, thus assessing the relevance in the analysis of different sets of third-order moments can be computationally simple. The paper gives an illustration of the method and argues for the need of refraining from adding into the analysis an excess of higher-order moments.

Key words: structural equation modeling, goodness-of-fit testing, moment test, third-order moments, interaction terms, equivalent models, saturated model.

Introduction

In Mooijaart and Satorra (2009) it has been shown that, under some general conditions, the normal theory test statistics, which are based on means and covariances only, are not able to assess interactions among observable or latent variables of the model. One conclusion is that for analyzing models with interactions more information than just means and covariances need to be brought into the analysis. Several methods have been proposed for analyzing models with non-linear (interactions) relationships. Originally, the main approach was to bring into the model as new variables the product of indicators of exogenous factors; see, e.g., Kenny and Judd (1984) and Jöreskog and Yang (1996), among many others. For implementing that approach, a key issue is the choice of the product indicators; see, e.g., Marsh, Wen, and Hau (2004). In that approach it was assumed that the latent predictor variables are normally distributed. More recently, the maximum likelihood (ML) approach that assumes normality for all the independent stochastic constituents of the model has been promoted. In formulating the likelihood function, this ML approach has to deal with a multivariate integral issue which, in the way it is tackled, yields several ML alternatives: normal mixtures were used by Klein and Moosbrugger (2000) in what they call LMS (latent moderated structural) method; the method of Muthén and Muthén (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

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. E-mail: Mooijaart@Fsw.LeidenUniv.nl

method. A different approach, although in fact it also deals with finding maximum likelihood estimates, is the Bayesian approach combined with the MCMC method as discussed by Lee and Zhu (2002) and Lee (2007). Models with interaction terms have also been analyzed by methods that involve factor score estimates; see, e.g., Wall and Amemiya (2000, 2007) and Klein and Schermelleh-Engel (2010). Although interesting, such methods need to circumvent the classical issue of inconsistency of the maximum likelihood method under the presence of nuisance parameters (Neyman & Scott 1948); further, their regression-type perspective deviates from the classical structural equation model (SEM) approach where a goodness-of-fit test of the model naturally arises.

In this paper instead of the ML approach we use the moment estimation method based on fitting first-, second-, and a selection of third-order moments, as in Mooijaart and Bentler (2010). We expand Mooijaart–Bentler's work by developing theory for selecting the third-order moments to be included in the analysis. We conjecture that expanding the set of first- and second-order moments with just a selection of third-order moments yields a more accurate analysis, in terms of robustness against small samples and against deviation from distributional assumptions, than using methods that involve full distributional specification such as ML. Like in the traditional Kenny–Judd's approach when using product indicators, here we are also confronted with the issue of which third-order moments should be included in the analysis. In contrast with ML, the advantage of the moment structure approach is that a goodness-of-fit test of the model is obtained. We recall that the ML approach faces the problem of assessing the distribution of the likelihood ratio test under the null model (see Klein & Moosbrugger, 2000), and Klein & Schermelleh-Engel 2010, for a discussion of this feature). In fact, in the present paper, the model goodness-of-fit test guides the selection of the most informative third-order moments for specific interaction parameters; more specifically, the third-order moments that maximize the power of the model test will be the ones to be included in the analysis.

A key result of the paper is a theorem that shows the connection between the power of the goodness-of-fit test of a model and the power of a moment test based on multivariate moments. The theorem will allow us to circumvent parameter estimation and model fit when assessing the importance of a specific set of third-order moments.

The remaining of the paper is structured as follows. Section 1 presents the class of models considered, estimation issues, and the model and moment tests; Section 2 presents an illustration with simulated data that motivates the import of the paper; Section 3 develops the theorem of the paper; Section 4 classifies the third-order moments into various classes and types; a forward-selection procedure for higher-order moments is outlined in Section 5; Section 6 concludes. Proofs and technical results that are not essential for the flow of the paper are confined in appendices.

1. Formulation of the Model and Estimation and Testing

In LISREL formulation, a model with interaction terms is written as follows:

$$\eta = \alpha + B_0\eta + \Gamma_1\xi + \Gamma_2(\xi \otimes \xi) + \zeta, \tag{1}$$

$$y = \nu_y + \Lambda_y\eta + \epsilon, \tag{2}$$

$$x = \nu_x + \Lambda_x\xi + \delta, \tag{3}$$

where y and x are, respectively, the indicators of endogenous and exogenous variables, of dimensions p and q, respectively; η and ξ are, respectively, the vectors of endogenous and exogenous factors; ζ is the disturbance term of the structural model equation; and ϵ and δ are vectors of measurement error (or unique factors). In the developments of the present paper, the vector variables ξ, ζ, ϵ and δ will be assumed to be independent of each other, with ξ normally distributed. (Often these stochastic terms are assumed to be only uncorrelated, and in the ML analysis they are also assumed to be normally distributed.) The vector $\xi \otimes \xi$ collects the interaction factors, and the elements of the matrix Γ_2 are the magnitudes of the interactions. Whenever Γ_2 is zero, we say the model is linear, there are no interactions. Note that interaction is used as a general term encompassing product variable and quadratic terms. The coefficient matrices B_0 and Γ_1 contain the usual linear effects among endogenous and exogenous variables. Here α, ν_y, ν_x are intercept vectors.

For further use, we define $B = I - B_0$, a matrix that is assumed to be non-singular. The variances and covariances of independent variables of the model, namely $\Phi = \operatorname{cov}(\xi)$ and $\Psi = \operatorname{cov}(\zeta)$, can be structured as a function of more basic parameters. The vector of observable variables is $z = (y', x')'$. The model equations (1) to (3), with the added assumptions on the stochastic constituents of the model, imply that the means, variances and covariances, and third-order moments of z can be written as a function of the model parameters. Let σ_1 be the vector of first-order moments of z; σ_2 the vector of non-redundant second-order moments of z; and σ_3 a vector of a selection of third-order moments of z. Then σ_1, σ_2, and σ_3 can be expressed as a function of the model parameters (e.g., formula (2) of Mooijaart & Bentler, 2010).

Let σ be all first-, second- and a selection of third-order moments of z, and let s be the usual sample moment estimator of σ based on an i.i.d. sample of z of size n. Since $\sigma = \sigma(\theta)$, where $\sigma(\theta)$ is a continuously differentiable function of the model parameters θ, estimation will be undertaken by minimizing the weighted least squares (WLS) fitting function

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

where W is a weight matrix that converges in probability (when $n \to +\infty$) to W_0, a positive definite matrix. A natural choice of W is the inverse of an estimate of the covariance matrix of vector s. In covariance structure analysis, it has been shown, however, that the use of this general weight matrix leads to biased estimates when sample size is not too large (e.g., Boomsma & Hoogland, 2001). In our case, where in addition to the means and covariances we fit a selection of third-order moments, the bias of estimates can be expected to be even larger; so, often, a typical fitting function is the LS one, i.e. the one where W is the identity matrix.

Let Γ be the asymptotic covariance matrix of s (i.e., the asymptotic limit of $\operatorname{cov}(\sqrt{n}s)$). A well-known test statistic for testing the goodness-of-fit of the model is defined by

$$T_{\mathrm{WLS}} = \bigl(s - \sigma(\theta)\bigr)' \bigl(\hat{\Gamma}^{-1} - \hat{\Gamma}^{-1} \hat{\dot{\sigma}} \bigl(\hat{\dot{\sigma}}' \hat{\Gamma}^{-1} \hat{\dot{\sigma}}\bigr)^{-1} \hat{\dot{\sigma}}' \hat{\Gamma}^{-1}\bigr) \bigl(s - \sigma(\theta)\bigr), \tag{4}$$

where $\hat{\dot{\sigma}}$ is the Jacobian of $\sigma(\theta)$ evaluated at the WLS estimate $\hat{\theta}$, and $\hat{\Gamma}$ is a consistent estimate of Γ. Under standard conditions it can be shown (Browne, 1984; Satorra, 1989) that T_{WLS} is asymptotically (central) chi-square distributed when the model $\sigma = \sigma(\theta)$ holds, and it is non-central chi-square with non-centrality parameter λ_{WLS} when the analyzed model does not hold (but it is not too deviant from the null). The degrees of freedom of the test is equal to the dimension of s minus the number of independent parameters of the model. This implies that a saturated model is the one that leaves all the moments involved unrestricted. Note that the saturated model will change depending on the selection of the third-order moments included in the analysis.

The specific expression for λ_{WLS} is now developed. We need to introduce a bit of notation. Partition $\sigma = (\sigma_{12}', \sigma_3')'$ where σ_{12} contains the first- and second-order moments and σ_3 is the vector of the selected third-order moments included in the analysis. Further, let the vector θ of model parameters be partitioned as $\theta = (\theta_1', \theta_3')'$, where θ_3 contains all the parameters involved in the interactions, the free elements of Γ_2. Consider a null model H_0 with only linear terms, that is, Γ_2 equal to zero. We note that, in contrast with Mooijaart and Satorra (2009), σ_3 is now present in the analysis, and that θ_3 is present or not depending on whether the model fitted is H_0 or H_1. It holds

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

where $\dot{\sigma}_{12,1}$ and $\dot{\sigma}_{3,1}$ are, respectively, the Jacobian of σ_{12} and σ_3 with respect to θ_1; and $\dot{\sigma}_{12,3}$ and $\dot{\sigma}_{3,3}$ are, respectively, the Jacobian of σ_{12} and σ_3 with respect to the interaction term parameters θ_3. In this set-up, the Jacobian matrix associated to model H_0 is

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

Furthermore, when the model fitted is H_0, we have $\sigma_3(\theta) = 0$ independently of θ_1, so we get $\dot{\sigma}_{3,1} = 0$, and thus

$$\dot{\sigma}_{|H_0} = \begin{pmatrix} \dot{\sigma}_{12,1} \\ 0 \end{pmatrix}. \tag{5}$$

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

Now, let σ_a be a moment vector under the specification H_1 but deviant from H_0 (i.e., σ_a complies with the model equations (1) to (3) with at least one non-zero element in Γ_2). Consider the fit of H_0 to σ_a and let $\hat{\sigma}_0$ be the fitted moment vector. Then the non-centrality parameter (ncp) associated to T_{WLS} is (Satorra, 1989)

$$\lambda_{\mathrm{WLS}}(\sigma_a \mid H_0) = n(\sigma_a - \hat{\sigma}_0)'\bigl(\Gamma^{-1} - \Gamma^{-1}\dot{\sigma}\bigl(\dot{\sigma}'\Gamma^{-1}\dot{\sigma}\bigr)^{-1}\dot{\sigma}'\Gamma^{-1}\bigr)(\sigma_a - \hat{\sigma}_0). \tag{6}$$

For further use, let σ_{a3} be the sub-vector of σ_a involving only the third-order moments. The non-centrality parameter (6) and the degrees of freedom of the model test determine the power of the test against the deviation σ_a from H_0. The vector σ_a deviates from H_0 by having specific non-zero values for interaction parameters of Γ_2. We are interested in those third-order moments that, when included in the analysis (i.e., included in s_3), yield higher power for specific interaction parameters of Γ_2. In principle this would require computing the ncp $\lambda_{\mathrm{WLS}}(\sigma_a \mid H_0)$ of (6) for each set of third-order moments to be evaluated for inclusion in s_3. This would be a computationally cumbersome task, since it requires a different model fit for each selection of third-order moments.

Fortunately we will be able to circumvent this computational difficulty by using a moment test based just on multivariate raw data.

Consider the partition $s = (s_{12}', s_3')'$ of the sample moments and the associated partition of its variance matrix,

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

where $\Gamma_{3,3}$ is the asymptotic variance matrix of the vector s_3 of the selected third-order moments. A moment test (MT) for testing the null hypothesis $\sigma_3 = 0$ is simply

$$T_{\mathrm{MT}} = n s_3' \hat{\Gamma}_{3,3}^{-1} s_3, \tag{7}$$

where $\hat{\Gamma}_{3,3}$ is a consistent estimate of $\Gamma_{3,3}$. The number of degrees of freedom of the test is equal to the dimension of s_3. Since T_{MT} does not involve specifying a model nor a test for model fit, it is computationally easy to obtain using just multivariate raw data. The corresponding non-centrality parameter when $\sigma_3 = \sigma_{a3}$ is

$$\lambda_{\mathrm{MT}}(\sigma_a) = n\sigma_{a3}' \Gamma_{3,3}^{-1} \sigma_{a3}. \tag{8}$$

Computation of $\lambda_{\mathrm{MT}}(\sigma_a)$ does not involve fitting a model, thus it is rather easy to automatize. Given the difficulties of computing $\lambda_{\mathrm{WLS}}(\sigma_a \mid H_0)$, it would be useful to obtain it from $\lambda_{\mathrm{MT}}(\sigma_a)$. Section 3 develops conditions under which the two non-centrality parameters are in fact equal. The use of $\lambda_{\mathrm{MT}}(\sigma_a)$ to assess the power of T_{WLS} will be the basis of the procedure for selecting third-order moments proposed in the present paper. The next section motivates the need for researching this.

2. A Motivating Illustration

Using simulated data, we now illustrate a case where the choice of the third-order moment changes substantially the power of the goodness-of-fit test of the model, and where the ncp's of the model and moment test do in fact coincide.

The model and simulations: We simulate data from the so-called Kenny and Judd (1984) model, the same model context used by Jöreskog and Yang (1996) and Klein and Moosbrugger (2000), among others. The Monte Carlo study consists on replicating (500 times) the generation of a sample of size $n = 600$ from Kenny and Judd's model with all the independent stochastic constituents of the model following a normal distribution. For each simulated sample, Kenny and Judd's model was fitted by LS. The analysis was carried out without centering the data, with a mean structure as part of the model. The power was computed as the percentage of rejections of the goodness-of-fit test T_{LS} across the 500 replications, when the model H_0 of no-interactions was analyzed. Essential to the illustration is that the theoretical value of the power of the test was also computed using the above formulas (6) and (8) of non-centrality parameters. Here power is the probability of rejecting the model H_0 that assumes zero interaction when in fact interactions are present in the model.

The model contains two latent factors plus an interaction term determining an observed dependent variable, V5. In addition, each factor has two indicators, $V1$ and $V2$ are indicators of the first factor, and $V3$ and $V4$ are indicators of the second factor. We are concerned with the interaction parameter β_{12} which in our Monte Carlo study is varied from 0.0 to 0.7. When the interaction equals zero the power is expected to be equal to the α-level (5%) of the test, and the power is expected to increase with the magnitude of the interaction term. Mooijaart and Bentler (2010) discuss a similar Monte Carlo study, however, they do not involve computation of theoretical power using the non-centrality parameter.

The present simulations aim to compare the theoretical power of the T_{LS} computed using (6) with the actual empirical power. Tables 1 and 2 show, for different sizes of the interaction parameter (coefficient β_{12}, first column of the table), the values of the ncp's for T_{LS} and T_{MT} (columns 2 and 3, respectively) computed using the formulas (6) and (8). Column 4 gives the theoretical power value for T_{LS} (using the ncp's of column 2 and the df $= 8$ of the model test). The last column of the table shows the empirical power deduced from the 500 replications. Tables 1 give the results for s_3 equal to the third-order moment V1V3V5, while Table 2 gives

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

β_{12}	$\lambda_{\mathrm{MT}}(1)$	$\lambda_{\mathrm{LS}}(8)$	powTh in %	powEmp in %
0.0	0	0	5.0	4.4
0.1	1.472	1.472	10.7	9.2
0.2	5.264	5.268	31.3	31.2
0.4	14.607	14.618	78.7	75.6
0.7	24.711	24.746	96.7	97.0

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

β_{12}	$\lambda_{\text{MT}}(1)$	$\lambda_{\text{LS}}(8)$	powTh in %	powEmp in %
0.0	0	0	5.0	4.6
0.1	2.215	2.217	14.2	14.0
0.2	5.752	5.777	34.3	32.4
0.4	8.204	8.359	49.7	42.6[a]
0.7	7.360	7.571	45.1	38.4[a]

[a]Difference from theoretical power is statistically significant at 5%-level.

the results when s_3 corresponds to V5V5V5. In the computations for the theoretical power we require the matrix Γ. This matrix is not exactly known in an application but can be estimated from the data. To avoid distorting the illustration with variation due to an estimate of Γ, this matrix was estimated by simulation with a sample of size 100,000 and it was kept fixed across all the simulations (this is similar to Satorra, 2003, where in covariance structure analysis power was computed for non-normal data).

From the tables we see, first, a substantial change on the power value of the model test depending on which third-order moment is incorporated as s_3, with V1V3V5 having more power than V5V5V5 (when β_{12} is greater than 0.2); second, we see that the two non-centrality parameters $\lambda_{\text{LS}}(\sigma_a \mid H_0)$ and $\lambda_{\text{MT}}(\sigma_a)$ are basically equal. Further, there is general agreement between the theoretical and empirical power values, with only two cells showing a significant difference between theoretical and empirical power. The significant differences correspond to cells related to the non-monotonicity of the power function to be commented on next.

One would expect that the power of the test increases monotonically with the magnitude of the misspecification inherent in the analyzed model H_0, i.e. when the absolute value of β_{12} increases. Clearly, this is the case for V1V3V5, but not for V5V5V5. It is remarkable that when s_3 is V5V5V5, the non-centrality parameter does not increase monotonically with the interaction parameter as one would expect. The empirical power shown in the last column of the table shows also such a decrease on power when misspecification increases. An explanation for this deviation from monotonicity will be given in Section 4.

3. Relation Between Power of the Model and Moment Test

This section develops a theorem setting up the conditions under which there is equality among the non-centrality parameters of the model and the moment tests.

The first condition we need to introduce is that the linear part of the structural model is saturated. This will guarantee that the models H_0 and H_1 are equivalent at the level of first- and second-order moments (see Mooijaart & Satorra, 2009).

Condition 1 (Saturation of structural equations). Under H_0 ($\Gamma_2 = 0$), parameterization of model equation (1) does NOT constrain α, Γ_1, Φ and the product matrix $B^{-1}\Psi B^{-T}$ (aside from symmetry).

As mentioned above, let σ_a satisfy the specification H_1, and let $\hat{\sigma}_0$ be the fitted vector when H_0 is fitted to σ_a. The next lemma shows that, basically under Condition 1, $\hat{\sigma}_0$ and σ_a coincide on the first- and second-order moments, i.e. $(\hat{\sigma}_0)_{12} = (\sigma_a)_{12}$ where the subscript "$_{12}$" denotes first- and second-order moments.

Lemma 1 (Model equivalence on first- and second-order moments). *Assume Condition 1; the variables ξ, δ, ζ and ϵ are uncorrelated; ξ is normally distributed; and the W of the WLS-analysis is block-diagonal on s_{12} and s_3. Let σ_a be a moment vector which will be fitted exactly by H_1, and $\hat{\sigma}_0$ be the fitted moment when H_0 is fitted to σ_a. Then*

$$(\hat{\sigma}_0)_{12} = (\sigma_a)_{12}.$$

Proof: See Appendix A. □

Note that the conclusion of the lemma can also be written as

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

Lemma 1 implies that under Condition 1 the WLS fit of H_0 to σ_a gives zero residuals for first- and second-order moments. This result needed $W = \text{block-diag}(W_{12,12}, W_{3,3})$, a partition conformable with $\sigma = (\sigma_{12}', \sigma_3')'$.[1]

The non-centrality parameter (6) can be written alternatively as (Satorra, 1989)

$$\lambda_{\text{WLS}}(\sigma_a \mid H_0) = n(\sigma_a - \hat{\sigma}_0)' F \big(F' \Gamma F\big)^{-1} F' (\sigma_a - \hat{\sigma}_0), \tag{9}$$

where F is an orthogonal complement of the matrix $\dot{\sigma}_{|H_0}$ defined above, i.e. $F'\dot{\sigma}_{|H_0} = 0$. Given the form (5) of the Jacobian $\dot{\sigma} \mid H_0$, we have

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

with $G'\dot{\sigma}_{12,1} = 0$. Thus, using the inverse of partitioned matrices and using Lemma 1, the non-centrality parameter of (9) can be written as

$$\lambda_{\text{WLS}}(\sigma_a \mid H_0) = n\sigma_{a3}' \big(\Gamma_{3,3}^{-1} - \Gamma_{3,12} G \big(G' \Gamma_{12,12} G\big)^{-1} G' \Gamma_{12,3}\big)^{-1} \sigma_{a3}, \tag{10}$$

where we assumed that $\Gamma_{3,3}$ is non-singular (recall the partition of Γ above).

Comparing (10) and (8), it holds

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

So, for the equality of the non-centrality parameters, we require the rather technical matrix equality $G'\Gamma_{12,3} = 0$. Appendix B shows that this matrix equality is also ensured by Condition 1, provided mild additional conditions apply: symmetry and independence of a vector of random constituents of the model (condition SI) and no constraints across different parameter matrices (condition FPI).

The theorem to be proven in this section makes use of the form of the covariance matrix among s_{12} and s_3, the matrix $\Gamma_{12,3}$ that is implied by the model equations (1) to (3). In the derivations of Appendix B, under the model H_1 the vector of observable variables z is written as

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

[1] In that case

$$\big(s - \sigma(\theta)\big)' W \big(s - \sigma(\theta)\big) = \big(s_{12} - \sigma_{12}(\theta)\big)' W_{12,12} \big(s_{12} - \sigma_{12}(\theta)\big) + s_3' W_{3,3} s_3,$$

since $\sigma_3(\theta) = 0$ when fitting H_0.

where matrix Λ_3 consists of regression weights of interaction and/or quadratic terms of the ξ variables. Note that under H_0, $\Lambda_3 = 0$; furthermore, δ_1 and δ_2 are independent of each other, and matrix A is partitioned as $A = (A_1, A_2)$. Note that $\delta = (\delta_1', \delta_2')'$, with δ_2 containing the main factors and the interaction/quadratic factors. The vector δ_1 collects the rest of the factors (errors and disturbances).

The following lemma is needed:

Lemma 2. *Under H_1 and the assumption SI (symmetry and independence) of Appendix B,*

$$\Gamma_{z,12,3} = D^+(A_2 \otimes A_2)D\Gamma_{\delta_2,12,3}T'(A_2 \otimes A_2 \otimes A_2)'T^{+\prime}$$

holds, where $A_2 = (\Lambda_1, \Lambda_2\Gamma_2)$, with D and T being duplication and triplication matrices respectively (Magnus & Neudecker, 1999; Meijer, 2005).

Proof: From $z = \mu + A_1\delta_1 + A_2\delta_2$, it follows that

$$\begin{aligned}\Gamma_{z,12,3} &= D^+(A_1 \otimes A_1)D\Gamma_{\delta_1,12,3}T'(A_1 \otimes A_1 \otimes A_1)'T^{+\prime} \\ &\quad + D^+(A_2 \otimes A_2)D\Gamma_{\delta_2,12,3}T'(A_2 \otimes A_2 \otimes A_2)'T^{+\prime},\end{aligned} \tag{12}$$

where $\Gamma_{\delta_i,12,3}$, $i = 1, 2$, is the covariance matrix of the first-, second- and third-order moments of δ_1 and δ_2, respectively. Because δ_1 has a symmetric distribution, $\Gamma_{\delta_1,12,3} = 0$, and so the first term on the right-hand side of (12) vanishes. □

For the main theorem of the paper, we need an additional lemma.

Lemma 3. *Assume Condition 1 and $G'\dot{\sigma}_{12} = 0$; then $G'\Gamma_{z,12,3} = 0$.*

Proof:

$$\begin{aligned}D^+(A_2 \otimes A_2) &= D^+\big[(\Lambda_1, \Lambda_3) \otimes (\Lambda_1, \Lambda_3)\big] \\ &= D^+\big[(\Lambda_1, \Lambda_2\Gamma_2) \otimes (\Lambda_1, \Lambda_2\Gamma_2)\big] \\ &= D^+\left[(\Lambda_1, \Lambda_2)\begin{pmatrix} I & 0 \\ 0 & \Gamma_2\end{pmatrix} \otimes (\Lambda_1, \Lambda_2)\begin{pmatrix} I & 0 \\ 0 & \Gamma_2\end{pmatrix}\right] \\ &= D^+\left[\big((\Lambda_1, \Lambda_2) \otimes (\Lambda_1, \Lambda_2)\big)\left(\begin{pmatrix} I & 0 \\ 0 & \Gamma_2\end{pmatrix} \otimes \begin{pmatrix} I & 0 \\ 0 & \Gamma_2\end{pmatrix}\right)\right].\end{aligned}$$

Since by Condition 1 the matrices Φ and $B^{-1}\Psi B^{-T}$ are unrestricted, using Lemma B2, we obtain $G'D^+(A_2 \otimes A_2) = 0$, from which we obtain the result of the lemma. □

So far, all the matrices were evaluated at the true population values, the same values as when fitting H_1 to σ_a. The theorem to be proven involves matrices evaluated at the fitted values under the restricted model H_0. Appendix C presents Lemma C1 that relates expressions involving both sets of matrices. Now we are ready to state and prove the main theorem of the paper.

Theorem 1. *Under the conditions of Lemma 3,*

$$\lambda_{\mathrm{WLS}}(\sigma_a \mid H_0) = \lambda_{\mathrm{MT}}(\sigma_a).$$

Proof: Simply, combine (11) with Lemma C1. □

This theorem will be exploited in the next section to yield a classification of third-order moments attending to their power functions.

4. Classes of Third-Order Moments

In Section 2 we presented an example where the choice of third-order moments determines the shape of the power function of the model test. In this section we investigate analytically such variation of the power function for different third-order moments. In principle, to study this variation of the power function we would need to compute the expression of the ncp arising from (6). That expression involves fitting a model for each set of third-order moments considered. The theorem of the previous section equates the ncp of the model test with the ncp of the moment test, and thus allows investigation of the power of the model test without requiring fitting a model.

For a given interaction term, we distinguish three classes of third-order moments: those for which the power does not vary with the size of interaction, to be called the CP (constant power) class; those for which the power increases monotonically with the size of the interaction term, to be called the MP (monotonic power) class; and, finally, those for which the power does not increase monotonically with the size of misspecification, to be called the NMP (non-monotonic power) class. The three classes of third-order moments will be illustrated using a model example similar to the one in Section 2.

We consider a simple model set-up of two observed independent variables x_1 and x_2, with a single dependent variable y. Note that this example is closely related to the example discussed in the illustration of Section 2 (now, however, we do not include the measurement part of the model). This section aims to address the non-monotonicity between power and size of interaction noted in Section 2 for some third-order terms. The model considered is

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

where the x's and e are centered variables. This model equation can be re-written as

$$y = y^* - E\big[y^*\big] = \beta_1 x_1 + \beta_2 x_2 + \beta_{12}(x_1 x_2 - \phi_{12}) + e, \tag{13}$$

where ϕ_{12} is $E(x_1 x_2)$. In this example, the following types of third-order moment can be distinguished: $\mu_{yx_1x_2}$, $\mu_{y^2x_1}$, $\mu_{y^2x_2}$ and μ_{y^3}. From the section above, we know that the power of the goodness-of-fit test T_{WLS} is determined by its ncp which has the same value as the ncp associated to the moment test T_{MT}. That is, we have the following three types of expression for the non-centrality parameters (up to a sample size scaling) (we used Theorem 1):

$$\mathrm{ncp}(1) = \frac{(\mu_{yx_1x_2})^2}{\gamma_{yx_1x_2}}, \qquad \mathrm{ncp}(2) = \frac{(\mu_{y^2x_1})^2}{\gamma_{y^2x_1}}, \qquad \mathrm{ncp}(3) = \frac{(\mu_{y^3})^2}{\gamma_{y^3}}, \tag{14}$$

where $\gamma_{yx_1x_2} = \mathrm{var}(m_{yx_1x_2})$, $\gamma_{y^2x_1} = \mathrm{var}(m_{y^2x_1})$ and $\gamma_{y^3} = \mathrm{var}(m_{y^3})$ involve six-order moments (they are elements of the matrix Γ). Equations (14) express the link between the ncp's and third-order moments. Model equation (13) implies the following expression of the third-order moments as a function of model parameters:

$$\text{Type}(1): \quad \mu_{yx_1x_2} = \beta_{12}(\phi_{11}\phi_{22} + \phi_{12}),$$

$$\text{Type}(2): \quad \mu_{y^2x_k} = 2\beta_{12}\big(2\beta_1\phi_{kk}\phi_{12} + \beta_2\big(\phi_{11}\phi_{22} + \phi_{12}^2\big)\big), \quad k = 1, 2,$$

$$\text{Type}(3): \quad \mu_{y^3} = 6\big(\beta_1^2\phi_{11}\phi_{12} + \beta_2^2\phi_{22}\phi_{12} + \beta_1\beta_2\big(\phi_{11}\phi_{22} + \phi_{12}^2\big)\big)\beta_{12} + \big(6\phi_{11}\phi_{12}\phi_{22} + 2\phi_{12}^3\big)\beta_{12}^3.$$

Here the ϕs denote the covariances among the x's. (To derive those expressions we used bivariate normality for the variables x_1 and x_2.) We see that, as should be expected, the third-order moments are zero when the interaction parameter β_{12} is zero. Importantly, note that the third-order moments of Type 1 and 2 are linear functions of the interaction parameter β_{12} while the Type 3 is non-linear on the interaction, thus inducing the non-monotonicity of the power function. Hence we see that moments of Type 1 and 2 are of the MP class, while moments of Type 3 are of the NMP class. We could have also considered third-order moments involving only X variables; these are, obviously, of CP (constant power) class. Theorem 1 has thus allowed us to relate the power of the T_{WLS} test with the form of the third-order moments as a function of the interactions. The power function is further investigated in the following simulation study.

Simulation 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 $\beta_0 = 1$, $\beta_1 = 0.2$, $\beta_2 = 0.4$ and $\text{var}(e) = 0.2$.

In this example we aim to assess the influence of the interaction parameter (β_{12}) on the size of the ncp. Unfortunately, there is no analytical expression for the ncp's in terms of the model parameters, because the denominator 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 up to order twelve. A small Monte Carlo 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 and parameter values described 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.

The results shown in Table 3 are summarized as follows: (i) As expected, the means across replications of the third-order moments, column m, are close to the population values shown in column μ. (ii) The ncp's for the MP (monotonic power) third-order moments are always

TABLE 3.
Monte Carlo results for the mean and variance of two types of third-order moment[a].

	Moment x_1x_2y				Moment y^3			
β_{12}	μ	m	γ	ncp	μ	m	γ	ncp
0.0	0.000	0.000	0.400	0.000	0.000	0.000	0.700	0.000
0.1	0.037	0.037	0.417	1.939	0.035	0.035	0.794	0.939
0.2	0.074	0.074	0.469	6.942	0.074	0.073	1.144	2.823
0.3	0.111	0.110	0.554	13.250	0.117	0.117	1.878	4.385
0.4	0.148	0.147	0.680	19.222	0.170	0.169	3.339	5.160
0.5	0.184	0.184	0.839	24.321	0.233	0.233	5.962	5.457
0.6	0.221	0.221	1.032	28.518	0.311	0.311	10.507	5.524
0.7	0.258	0.258	1.261	31.715	0.405	0.405	17.965	5.486
0.8	0.295	0.295	1.504	34.729	0.519	0.519	29.570	5.457
0.9	0.332	0.332	1.812	36.467	0.656	0.655	47.731	5.391
10.0	0.369	0.369	2.156	37.904	0.818	0.818	75.047	5.353

[a]Note that γ is defined as the sample size (600) times the variance of the third-order moment. Columns μ and m indicate the population and the mean (over the 100,000 replications) of the corresponding third-order moment. The "ncp" column correspond to the value of the non-centrality parameter computed using the moment test associated to the specific third-order moment.

(substantially) larger than for the NMP (non-monotonic power) third-order moments. (iii) The variance of the moments increases (and so does γ) when the interaction parameter increases, although this variance increases more sharply for the third-order moment y^3. (iv) When the interaction effect increases, the ncp associated with x_1x_2y increases also, but not the ncp for y^3, where we see that the ncp does in fact decrease when β_{12} is larger than 0.6. This empirical non-linear relation between the size of the interaction term and the ncp was noted above analytically for the NMP (non-monotonic power) class of third-order moments.

Point (iv) is a counter-intuitive result that needs to be commented on. Our explanation of this result is that the variance of the third-order moment (the denominator of the ncp) increases sharply with the increase of the interaction parameter, so the ncp may in fact be decreasing while the interaction term (the numerator of the ncp) is increasing. This explains why in Table 2, involving a NMP class of third-order moment, the power does not vary monotonically with the size of the interaction.

A forward-selection procedure for third-order moments is discussed in the next section.

5. A Forward-Selection Procedure

In the context of the same model as in Section 2, and for the interaction parameter β_{12}, Table 4 presents non-centrality parameters, bias and standard errors (s.e.'s) for estimates of interaction, mean of (chi-square) goodness-of-fit values, and theoretical power, for a sequence of forward nested sets of third-order moments. The sequence starts with the third-order product term V1V3V5 and adds one additional third-order moment in each stage of the sequence. The first

TABLE 4.
Monte Carlo results for the selection procedure.[a]

Moment	univ-ncp	mult-ncp	bias	se($\hat{\beta}$)	sd($\hat{\beta}$)	χ^2	df	Power
V1V3V5	14.675	14.675	0.007	0.132	0.129	6.81	7	0.81
V2V5V5	7.844	16.369	0.000	0.109	0.102	7.77	8	0.84
V4V5V5	8.201	17.068	−0.006	0.098	0.094	8.59	9	0.84
V1V4V5	10.414	17.283	0.005	0.100	0.099	9.93	10	0.83
V2V3V5	7.891	17.445	−0.005	0.098	0.095	10.80	11	0.82
V1V5V5	12.875	17.632	0.003	0.098	0.102	12.00	12	0.81
V1V1V5	3.893	17.870	0.005	0.099	0.095	12.54	13	0.80
V5V5V5	8.190	18.095	−0.008	0.091	0.086	13.54	14	0.79
V3V5V5	10.468	18.440	−0.008	0.091	0.089	14.93	15	0.79
V3V3V5	5.282	18.616	−0.006	0.091	0.094	15.57	16	0.78
V1V2V5	3.259	18.684	−0.005	0.092	0.092	16.10	17	0.77
V3V4V5	5.146	18.731	0.003	0.095	0.092	17.11	18	0.76
V2V4V5	5.479	18.760	−0.005	0.092	0.088	18.75	19	0.75
V2V2V5	0.946	18.766	0.002	0.095	0.091	19.96	20	0.74
V4V4V5	2.359	18.770	−0.002	0.095	0.095	20.69	21	0.73

[a]Here "univ-ncp" is the non-centrality parameter of the moment test for an analysis that adds only the specific third-order moment. Corresponding to an analysis that uses the cumulative set of third-order moments: "mult-ncp" is the non-centrality parameter of the moment test; "bias" is the difference between the mean (across Monte Carlo replicates) of the estimate of interaction minus the true value; "se($\hat{\beta}$)" is the mean (across Monte Carlo replicates) of the standard errors; "sd($\hat{\beta}$)" is the standard deviation (across Monte Carlo replicates) of the estimates of interaction; χ^2 is the mean (across Monte Carlo replicates) of the goodness-of-fit test; "df" is the number of degrees of freedom of the goodness-of-fit test; "Power" corresponds to the asymptotic (theoretical) power associated with the moment test.

term of the sequence (in our case V1V3V5) is chosen as the one giving maximum (theoretical) non-centrality parameter when evaluated by the moment test approach discussed in Section 3. The third-order moment that is added in each step of the sequence is the one that yields the highest increase of the multivariate (overall) ncp. Column 3 showing the multivariate ncp will therefore increase when moving down by rows. Columns 4 to 7 of the table give, respectively, the mean of estimates minus the true value, the mean of estimates of s.e., the standard deviation of the estimates, and the mean of the goodness-of-fit test (means and standard deviations computed across replications). The last column of the table gives the theoretical power based on the multivariate ncp of column 3 and degrees of freedom reported in column 8. The reported sequence of increasing nested sets of third-order moments is like a forward-selection sequence encountered in variable selection procedures such as regression analysis. Note the key information for ordering the third-order moments arises from the ncp of the moment test. The Monte Carlo set-up specification used $\beta_{12} = 0.4$ and 250 replications.

Table 4 shows that 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. Remarkably, the gains in terms of efficiency of estimates are substantial when accumulating up to three third-order moments, but there is no substantial gain after that. It could be conjectured that adding third-order moments beyond the ones that improve on efficiency may deteriorate robustness against small samples. The last column of the table shows power values of the H_0 model test for the different sets of third-order terms. We see that, generally, after the inclusion of three third-order moments in the analysis, the power of the model test does in fact decrease if more third-order moments are added. This is another argument for refraining from including all the third-order moments.

6. Discussion

A central issue in the analysis of models with interactions is the selection of the higher-order moments to be included in the analysis. In this paper we argue that the most relevant moments for assessing specific interactions are those that lead to higher power in detecting the failure of H_0 (a model with no interaction terms) when the interaction terms are in fact present. In this paper we have addressed the choice of third-order moments to be used in the analysis. We have spelled out conditions (specially Condition 1) under which the power of the model test can be assessed using the power of a multivariate moment test available from raw data, without involving model fit. This fact is of importance since it simplifies considerably the practice of selecting higher-order moments to include in the analysis of a model with interaction terms. We have seen that the ncp $\lambda_{\mathrm{WLS}}(\sigma_a)$ can be computed as $\lambda_{\mathrm{MT}}(\sigma_a) = n\sigma_{a3}'\Gamma_{33}^{-1}\sigma_{a3}$, without need of a model fit.

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 normally distributed (note that then the interaction factor itself will not be normally distributed). Other stochastic constituents of the model, such as errors of measurement and disturbances, may, however, deviate from the normality assumption, though they are also subject to a mild distributional assumption: the assumption SI in Appendix B requires disturbances and error terms (the vector δ_1) to be symmetric and independent (not only uncorrelated) of the stochastic term δ_2, the vector of factors and interactions.

Note that, even though the model and moment tests have the same non-centrality parameter, they will generally have different degrees of freedom, the moment test having generally the smaller number of degrees of freedom. This implies that the moment test will have more power than the corresponding model test; this issue, however, is of minor relevance in our paper, which is mainly concerned with variation of power when changing the third-order moments included in s_3.

For Theorem 1 to hold, a basic condition is that the structural part of the model is saturated; that is, using the language of Mooijaart and Satorra (2009), *the degrees of freedom of the structural part of the model* need to be zero. The measurement part of the model can, however, have restrictions so that the degrees of freedom for the whole model can in fact be large.

In this paper we have classified the third-order moments in three classes according to the form of their power function. We have presented a simple model example, involving just one Y dependent variable and two X variables, where different types of third-order moment arise depending on the degree by which Y appears in the product term: when Y appears in degree one or two (e.g., YXX and YYX), the third-order moment is of the MP class, the power is monotonically increasing with the size of the interaction parameter (see expressions in (14)); when Y appears in degree three (YYY), the third-order moment is of the NMP class, the power is not monotonically increasing with the size of the interaction parameter. Deviation from monotonicity lead us to recommend avoiding the NMP types of third-order moment when fitting models with interaction terms. Note that there is also the CP (constant power) class of third-order moments, for example, the moments involving only the Xs. Since third-order moments of the CP class are not informative for specific interaction terms, they also should not be included in the analysis.

A final issue we want to discuss is how many third-order moments should be included, supplementing 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 a slight improvement in the power of the model test when adding two additional moments to the first one, but the power does in fact decrease when the number of third-order moments added is beyond two. Inclusion of higher-order moments deteriorates, generally, the robustness against small samples. In the case where sample size is not extremely large, each third-order moment that is added to the analysis is likely to induce more bias on parameter estimates and, generally, more inaccuracy in the asymptotic results. Even though theoretically asymptotic efficiency of estimates increases with the number of degrees of freedom, adding more moments may deteriorate accuracy measures such as mean square errors. We therefore recommend researchers to refrain from adding higher-order moments much beyond those strictly necessary for identification purposes.

Another argument for refraining from adding an excess of third-order moments relates to the concept of saturated model implicit in the model test. As noted above, a saturated model is the one that has enough parameters for the first-, second- and the selected higher-order moments to be unconstrained. Saturation of first- and second-order moments is a well-known topic. Adding regression effects, loadings, variances and covariances as free parameters of the model, can cause first- and second-order moments to be unrestricted. Saturation of higher-order moments is a much less explored land. Adding an interaction term in the model could possibly saturate third-order moments of the MP or NMP class, but not a third-order moment of the CP class. Saturating third-order moments of the CP class (for example, third-order moments that involve only Xs) may require the introduction of distributional parameters (e.g. skewness, kurtosis, etc.) for independent variables of the model. Not having those parameters in the model may amount to imposing distributional constraints on independent variables. So, to avoid distributional parameters, or restricting the distribution of independent variables, we recommend not to include moments of the CP class into the analysis.

In contrast to other approaches that require fitting a model for each set of third-order moments proposed, we now can assess the relevance of a specific set of third-order moments by direct computation of a moment test that does not involve fitting a model. Obviously, there remains issues to be investigated, such as the relative small sample size needed, effect-size issues, optimal step-wise method for selection of third-order moments, etc.. These are issues for further research that fall beyond the scope of the present paper.

Acknowledgements

Research of the second author is supported by grant SEJ2006-13537 from the Spanish Ministry of Science and Technology. The authors thank the AE and three reviewers for very constructive comments. The second author's work was mainly carried out while visiting UCLA.

Appendix A. Proof of Lemma 1

We now prove Lemma 1 that shows the equivalence of H_1 and H_0 for first- and second-order moments. Writing the class of models (1) to (3) in the linear-latent variable form (e.g., Satorra 1992)

$$y = \nu_y + \Lambda_y B^{-1}\alpha + \Lambda_y B^{-1}\Gamma_1\xi + \Lambda_y B^{-1}\Gamma_2(\xi \otimes \xi) + \Lambda_y B^{-1}\zeta + \epsilon_y, \tag{15}$$

$$x = \nu_x + \Lambda_x\xi + \epsilon_x, \tag{16}$$

we obtain expressions of the first- and second-order moments of observable variables as a function of the vector and matrix parameters; for the means:

$$\begin{aligned} E(y) &= \nu_y + \Lambda_y B^{-1}\bigl(\alpha + \Gamma_2 E(\xi \otimes \xi)\bigr), \\ E(x) &= \nu_x; \end{aligned} \tag{17}$$

for the variances and covariances:

$$\begin{aligned} \operatorname{cov}(y) &= \Lambda_y B^{-1}\Gamma_1\Phi\Gamma_1' B^{-T}\Lambda_y' \\ &\quad + \Lambda_y B^{-1}\bigl[\Gamma_2 \operatorname{cov}(\xi \otimes \xi)\Gamma_2' + \Psi\bigr]B^{-T}\Lambda_y' + \Theta_\epsilon \\ &= \Lambda_y B^{-1}\bigl[\Gamma_1\Phi\Gamma_1' + Q + \Psi\bigr]B^{-T}\Lambda_y' + \Theta_\epsilon, \\ \operatorname{cov}(x) &= \Lambda_x\Phi\Lambda_x' + \Theta_\delta, \\ \operatorname{cov}(x, y) &= \Lambda_x\Phi\Gamma_1' B^{-T}\Lambda_y' \end{aligned} \tag{18}$$

where $Q = \Gamma_2 \operatorname{cov}(\xi \otimes \xi)\Gamma_2'$, and Θ_ϵ and Θ_δ are the covariance matrices of ϵ and δ, respectively. The first- and second-order components of the moment vector σ_a are derived from the moment equations (17) and (18). The key issue now is whether such first- and second-order moments can be equated exactly under model H_0. The fitted matrices under model H_0, which are not necessarily equal to the ones obtained when fitting H_1 (i.e., the ones in the right-hand side of (17) and (18)), will be denoted with a tilde.

Let $\tilde{B}$ and $\tilde{\Phi}$ be the solutions under the specification H_0 of the matrix equality:

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

Such a solution exists by Condition 1, since saturation of the model at the level of the structural equations leaves the product matrix $\tilde{B}^{-1}\tilde{\Psi}\tilde{B}^{-T}$ unrestricted. Define

$$\begin{aligned} \tilde{\alpha} &= \tilde{B}B^{-1}\bigl(\alpha + \Gamma_2 D^+ E(\xi \otimes \xi)\bigr), \\ \tilde{\Gamma}_1 &= \tilde{B}B^{-1}\Gamma_1, \\ \tilde{\Phi} &= \Phi \end{aligned} \tag{19}$$

and the other vectors and parameter matrices (such as Λ_x, Λ_y, etc.) the same as under H_1 (these are parameters that can be constrained under both H_0 and H_1). Recall that α, Γ_1 and Φ are unconstrained by Condition 1, so $\widetilde{\alpha}$, $\tilde{\Gamma}_1$ and $\tilde{\Phi}$ could be feasible solutions under H_0. Simply by substitution it can be seen that

$$\begin{aligned} E(y) &= \nu_y + \Lambda_y \tilde{B}^{-1}\tilde{\alpha}, \\ E(x) &= \nu_x \end{aligned} \tag{20}$$

and

$$\begin{aligned} \operatorname{cov}(y) &= \Lambda_y \tilde{B}^{-1}\big[\tilde{\Gamma}_1 \tilde{\Phi} \tilde{\Gamma}_1' \tilde{B}^{-T} + \tilde{\Psi}\big]\tilde{B}^{-T}\Lambda_y' + \Theta_\epsilon, \\ \operatorname{cov}(x) &= \Lambda_x \tilde{\Phi} \Lambda_x' + \Theta_\delta, \\ \operatorname{cov}(x, y) &= \Lambda_x \tilde{\Phi} \tilde{\Gamma}_1' \tilde{B}^{-T} \Lambda_y'. \end{aligned}$$

Note that as expressed in our notation, the vectors ν_x and ν_y and the matrices Λ_y, Λ_x, Θ_ϵ, Θ_δ are the same under H_0 and H_1. Note that such matrices are allowed to be constrained under both models. When viewed as functions of the parameters, the matrices B and Ψ can be restricted provided $B^{-1}\Psi B^{-T}$ is unrestricted. These are the same conditions as the saturated model of Mooijaart and Satorra (2009).

Appendix B. Proof of $G'\Gamma_{12,3} = 0$

In this appendix we will make use of the following.

Lemma B1. *Given matrices A, B and C, we have*

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

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 the Kronecker products we see 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 $A \otimes (C, D) = (A \otimes C, A \otimes D)E_1$, where E_1 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 we have $(A, B) \otimes (C, D) = (A \otimes (C, D), B \otimes (C, D)) = (A \otimes C, A \otimes D, B \otimes C, B \otimes D)E$, where E is a super 2×2 matrix with block-diagonal matrices E_1 and E_2. □

We can re-write (15) and (16) as

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

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

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

and

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

Thus, in compact expression, H_1 can be written as

$$z = \mu + A\delta = \mu + A_1\delta_1 + A_2\delta_2, \tag{21}$$

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 to introduce two additional assumptions that will be needed for the theorem.

Assumption SI (Symmetry and independence). *The model H_1 holds and the distribution of δ_1 of* (21) *is symmetric and independent of δ_2.*

Furthermore,

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

Consider now the specification H_0, i.e. (1) to (3) with Γ_2 set to 0. Consider the vector of first- and second-order moments for z, $\sigma_{12} = (\sigma_1', \sigma_2')'$, where $\sigma_1 = E[z]$ and $\sigma_2 = \operatorname{vech} E[(z-\mu) \otimes (z-\mu)]$. 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.

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,

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

where, by virtue of FPI,

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

$R_\alpha = \partial\alpha/\partial\theta_\alpha'$, $R_{\Gamma_1} = \partial \operatorname{vec}(\Gamma_1)/\partial\theta_{\Gamma_1}'$, $R_\Phi = \partial \operatorname{vech}(\Phi)/\partial\theta_\Phi'$, $R_{B_0} = \partial \operatorname{vec}(B_0)/\partial\theta_{B_0}'$ and $R_\Psi = \partial \operatorname{vech}(\Psi)/\partial\theta_\Psi'$. 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

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

$$A_{14} = \frac{\partial \sigma_1}{\partial (\text{vec}\, B_0)'} = \begin{pmatrix} (\alpha' \otimes \Lambda_y)(B^{-T} \otimes B^{-1}) \\ 0 \end{pmatrix} R_{B_0},$$

$$A_{22} = \frac{\partial \sigma_2}{\partial (\text{vec}\, \Gamma_1)'} = 2D^+_{p+q}(\Lambda_1 \Phi \otimes \Lambda_2) R_{\Gamma_1},$$

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

$$A_{24} = \frac{\partial \sigma_2}{\partial (\text{vec}\, B_0)'} = 2D^+_{p+q}\big[(\Lambda_1 \otimes \Lambda_2)\big(\Phi \Gamma' B^{-T} \otimes I_m\big) + (\Lambda_2 \otimes \Lambda_2)\big(\Psi B^{-T} \otimes I_m\big)\big] R_{B_0},$$

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

Below we will assume that the vector and matrices α, Γ_1 and Φ are unrestricted, so 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 partitioning $G_1' = (G_{1y}', G_{1x}')$. Then we have the following equations in which the *means* are involved:

$$G_{1y}' \Lambda_y B^{-1} R_\alpha = 0, \tag{22}$$

$$G_{1y}'\big(\alpha' \otimes \Lambda_y\big)\big(B^{-T} \otimes B^{-1}\big) R_{B_0} + G_2' A_{24} = 0. \tag{23}$$

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

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

This expression being 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 (23) results in $G_2' A_{24} = 0$, which is (26) below.

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

$$G_2'\, D^+_{p+q}(\Lambda_1 \Phi \otimes \Lambda_2) R_{\Gamma_1} = 0, \tag{24}$$

$$G_2'\, D^+_{p+q}(\Lambda_1 \otimes \Lambda_1) D_n R_\Phi = 0, \tag{25}$$

$$G_2' D^+_{p+q}\big[(\Lambda_1 \otimes \Lambda_2)\big(\Phi \Gamma' B^{-T} \otimes I_m\big) + (\Lambda_2 \otimes \Lambda_2)\big(\Psi B^{-T} \otimes I_m\big)\big] R_{B_0} = 0, \tag{26}$$

$$G_2'\, D^+_{p+q}(\Lambda_2 \otimes \Lambda_2) D_m R_\Psi = 0. \tag{27}$$

Lemma B2. *Consider the specification H_0 under the separability assumption FPI. Assume Condition* 1 *with Φ and Ψ of full rank. Then*

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

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, such that $G'\dot\sigma_{12} = 0$.

Proof: Using Lemma B1, it holds

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

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 (24), the non-singularity of Φ and Γ_1 being unrestricted (so that $R_{\Gamma_1} = I$) yield

$$G_2' D_{p+q}^+(\Lambda_1 \otimes \Lambda_2) = 0. \tag{28}$$

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 non-singular 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

$$G_2' D_{p+q}^+(\Lambda_2 \otimes \Lambda_1) = 0. \tag{29}$$

Since Φ is symmetric and unrestricted, (25) 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 1999). Because we have $D_{p+q}^+(\Lambda_1 \otimes \Lambda_1)K_n = D_{p+q}^+ K_{p+q}(\Lambda_1 \otimes \Lambda_1) = D_{p+q}^+(\Lambda_1 \otimes \Lambda_1)$, and 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 that $D_{p+q}^+(\Lambda_1 \otimes \Lambda_1)N_n = D_{p+q}^+(\Lambda_1 \otimes \Lambda_1)$ and thus

$$G_2' D_{p+q}^+(\Lambda_1 \otimes \Lambda_1) = 0. \tag{30}$$

From (28) it follows that the first term in (26) is 0. Combining this result and (27) we have

$$G_2' D_{p+q}^+(\Lambda_2 \otimes \Lambda_2)\big[\big(\Psi B^{-1} \otimes I_m\big)R_{B_0}, D_m R_\Psi\big] = 0.$$

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 $R_{B_0,\Psi} = \left(\begin{smallmatrix} R_{B_0} & 0 \\ 0 & R_\Psi \end{smallmatrix}\right)$. Let $H = B^{-1}\Psi B^{-T}$; then H has $m(m+1)/2$ different non-duplicated elements. Now it holds true 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

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

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, $H = B^{-1}\Psi B^{-T}$ being free is equivalent to Z being of full column rank, this rank being equal to $m(m+1)/2$. There are two typical conditions under which this holds true: when Ψ is a diagonal matrix with unconstrained elements $(B_0)_{ij}$, or when Ψ is an unconstrained free matrix and $(B_0)_{ij}$ is constrained.

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

$$G_2' D_{p+q}^+(\Lambda_2 \otimes \Lambda_2) = 0. \tag{31}$$

Combination of (29) to (31) completes the proof of the lemma. □

Appendix C. Orthogonality Conditions for Matrices Fitted Under H_0 and H_1

Here we show that an orthogonality condition satisfied by the matrices under the fit of H_0 implies orthogonality when the matrices involved correspond to the fit of H_1. We use a tilde to denote the matrices fitted under H_0 that may have different values as when fitted under H_1.

Lemma C1. *Under the same conditions as in Lemma B2; if*

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

then

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

Proof: From the equivalence of H_0 and H_1, we have $B^{-1}\Gamma_1 = \tilde{B}^{-1}\tilde{\Gamma}_1$ and $\tilde{B}^{-1}\tilde{\Psi}\tilde{B}^{-T} = B^{-1}(\Psi + Q)B^{-T}$ where $Q = \Gamma_2 \operatorname{cov}(\xi \otimes \xi)\Gamma_2'$. So, because $B^{-1}\Gamma_1 = \tilde{B}^{-1}\tilde{\Gamma}_1$, it follows immediately that $\Lambda_1 = \tilde{\Lambda}_1$. 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,

$$\begin{aligned}
\big(\tilde{\Lambda}_1, \tilde{\Lambda}_2\big) \otimes \big(\tilde{\Lambda}_1, \tilde{\Lambda}_2\big) &= (\Lambda_1, \Lambda_2 V) \otimes (\Lambda_1, \Lambda_2 V) \\
&= \big[(\Lambda_1, \Lambda_2) \otimes (\Lambda_1, \Lambda_2)\big]\left[\begin{pmatrix} I & 0 \\ 0 & V \end{pmatrix} \otimes \begin{pmatrix} I & 0 \\ 0 & V \end{pmatrix}\right] \\
&= \big[(\Lambda_1, \Lambda_2) \otimes (\Lambda_1, \Lambda_2)\big] W
\end{aligned}$$

where $W = \left(\begin{smallmatrix} I & 0 \\ 0 & V \end{smallmatrix}\right) \otimes \left(\begin{smallmatrix} I & 0 \\ 0 & V \end{smallmatrix}\right)$, which implies the conclusion of the theorem since W is non-singular. □

References

Boomsma, A., & Hoogland, J.J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation models: Present and future. A Festschrift in honor of Karl Jöreskog* (pp. 139–168). Chicago: Scientific Software International.

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). Mahwah: 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.

Klein, A.G., & Schermelleh-Engel, K. (2010). Introduction of a new measure for detecting poor fit due to omitted nonlinear terms in SEM. *Advances in Statistical Analysis*, *94*, 157–166.

Lee, S.-Y. (2007). *Structural equation modeling: A Bayesian approach*. West Sussex: Wiley.

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

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

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., & Satorra, A. (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. (2007). *Mplus user's guide* (5th ed.). Los Angeles: Muthén & Muthén.

Neyman, J., & Scott, E.L. (1948). Consistent estimation from partially consistent observations. *Econometrica*, *16*, 1–32.
Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, *54*(1), 131–151.
Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology*, *22*, 239–278.
Satorra, A. (2003). Power of chi-square goodness-of-fit test in structural equation models: The case of non-normal data. In H. Yanai, A. Okada, K. Shigemasu, Y. Kano, & J.J. Meulman (Eds.), *New developments of psychometrics* (pp. 57–68). Tokyo: Springer.
Schott, J.R. (1997). *Matrix analysis for statistics*. New York: Wiley.
Wall, M.M., & Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association*, *26*, 1–29.
Wall, M.M., & Amemiya, Y. (2007). Nonlinear structural equation modeling as a statistical method. In *Handbook of computing and statistics with applications* (Vol. 1). Amsterdam: Elsevier.

Manuscript Received: 24 JUN 2010
Final Version Received: 13 APR 2011
Published Online Date: 15 OCT 2011