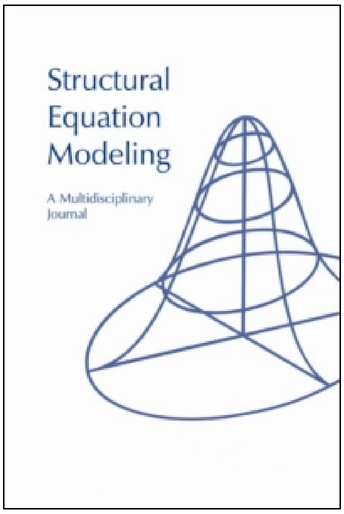

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653699

An Alternative Approach for Nonlinear Latent Variable Models

Ab Mooijaart[a]; Peter M. Bentler[b]
[a] Leiden University, [b] University of California, Los Angeles

Online publication date: 08 July 2010

To cite this Article Mooijaart, Ab and Bentler, Peter M.(2010) 'An Alternative Approach for Nonlinear Latent Variable Models', Structural Equation Modeling: A Multidisciplinary Journal, 17: 3, 357 — 373
To link to this Article: DOI: 10.1080/10705511.2010.488997
URL: http://dx.doi.org/10.1080/10705511.2010.488997

Structural Equation Modeling, 17:357–373, 2010

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705511.2010.488997

Psychology Press
Taylor & Francis Group

An Alternative Approach for Nonlinear Latent Variable Models

Ab Mooijaart
Leiden University

Peter M. Bentler
University of California, Los Angeles

In the last decades there has been an increasing interest in nonlinear latent variable models. Since the seminal paper of Kenny and Judd, several methods have been proposed for dealing with these kinds of models. This article introduces an alternative approach. The methodology involves fitting some third-order moments in addition to the means and covariances. This article discusses how the model equations can be formulated and how several standard tests, like the model fit and Lagrange multiplier tests, can be performed. The new method compares favorably with the maximum likelihood method in several studies and can provide evidence of interaction that earlier approaches might ignore.

In this article we deal with nonnormally distributed observed variables in latent variable structural equation models. As reviewed by Schumacker and Marcoulides (1998) and more recently by Dimitruk, Schermelleh-Engel, Kelava, and Moosbrugger (2007) and Coenders, Batista-Foguet, and Saris (2008), it can be hypothesized that this nonnormality arises from a nonlinear relationship between the latent variables. In particular, such a nonlinear relationship might reflect the existence of interaction or quadratic factors, as proposed in the seminal paper of Kenny and Judd (1984). Kenny and Judd pioneered the use of product indicators to identify the model and estimate its parameters. An extension of this approach was developed by Jöreskog and Yang (1996) and others. A key issue is the choice of the product indicators. Marsh, Wen, and Hau (2004) studied this issue via Monte Carlo and made some recommendations. These methods assume that the observed indicators, except the product indicators, are normally distributed. The currently more widely accepted maximum likelihood (ML) method involves

Correspondence should be addressed to Ab Mooijaart, Department of Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: Mooijaart@fsw.leidenuniv.nl

the same assumption. Although not all observed variables are normally distributed, a proper likelihood function can be defined when the nonnormality arises as a function of normal factor(s). The unattractive part of this approach is that the likelihood function involves a multivariate integral that does not have an explicit form. There are several ways to tackle this problem (see Cudeck, Harring, & du Toit, 2009; Klein, 2007; Klein & Moosbrugger, 2000; Lee & Zhu, 2002; as well as quasi-likelihood methods making fewer assumptions, Klein & Muthén, 2007). Further approaches that have been developed include Bayesian methods (Arminger & Muthén, 1998; Lee, 2007; Lee, Song, & Tang, 2007) and two-stage instrumental variables as borrowed from econometrics (Bollen & Paxton, 1998a, 1998b). Still other methods involve the use of sequential estimation (Ping, 1996), factor scores (Wall & Amemiya, 2003), and mixture models applied to obtain approximations to nonlinear models (Bauer, 2005).

Many of these methods require nonstandard and sometimes difficult ways of thinking about and implementing the model, for example, in requiring the use of nonlinear constraints on parameters. Although setups can be simplified to minimize such constraints (Coenders et al., 2008), the procedures are still highly specialized and technical. Such difficulties might explain the absence of discussions on nonlinear terms in introductory structural equation modeling (SEM) literature (e.g., Mulaik, 2009; Raykov & Marcoulides, 2006). In this article, we propose a method that is a minor extension of standard mean and covariance structure analysis.[1] Because the methodology involves standard SEM estimation procedures, little specialized knowledge is needed to use the method, and as usual, consistent estimates of the parameters are available and a goodness-of-fit test is an automatic by-product. As noted earlier, the crucial point of our method is that we fit, in addition to the usual means and covariances, a selection of third-order moments. See also Mooijaart (1985, 2008). The goodness-of-fit test is based on the residuals between the observed moments and the estimated moments. In addition to this overall fit test, we develop a test that gives an indication whether postulating one or more interactions between the latent variables might result in a significantly better fit of the model. As in standard SEM, this test is based on the Lagrange multiplier principle. This test is important because ML can yield incorrect conclusions on the necessity of an interaction term (Mooijaart & Satorra, 2009a, 2009b).

The first section of this article provides a formulation of the model equations. The second section discusses estimation of the model parameters and model tests. The subsequent sections discuss several examples for evaluating our method. This includes a study of the bias and standard errors of the estimates. We then show how the goodness-of-fit chi-square test statistic behaves. When a model is misspecified, we show that our new Lagrange multiplier test can indicate which parameter has to be added to the model to result in an improved model fit. We also evaluate the power of the test. Finally, we study our method when the indicators are nonnormally distributed. Because in the case of normally distributed indicators the ML method will yield the "best" estimates, we compare our results with those from the ML method as given by M*plus* (L. K. Muthén & Muthén, 1998–2007).

In this article we do not go into detail on the interpretation of interaction and quadratic terms in latent variable SEM models. There is a lot of literature on this subject. Although most of this literature deals with interaction models for observed variables, and in our models we

[1]The proposed method is implemented in an experimental version of EQS (Bentler, 2000–2008) and will be generally available in EQS 7.

deal with interactions of latent variables, the principles and problems involved (e.g., centering predictors, multicollinearity, and presentation and interpretation of results) are the same, and we refer the reader to the excellent book of Aiken and West (1991; see also Jaccard & Wan, 1996). Interpreting such effects causally, however, is a more difficult task (Hargens, 2009).

FORMULATION OF THE MODEL AND MODEL EQUATIONS

The structural equation models will be defined as in Bentler and Weeks (1980). In this approach, the model equations can be written as

$$\boldsymbol{\eta} = \boldsymbol{\beta}_0\boldsymbol{\eta} + \boldsymbol{\gamma}\boldsymbol{\xi},$$

where $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are the independent and dependent variables of the model and $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}$ are coefficient matrices. Both the independent and dependent variables can be either latent or observed variables. The measured variables are related to the dependent and independent variables via $\mathbf{y} = \mathbf{G}_y\boldsymbol{\eta}$ and $\mathbf{x} = \mathbf{G}_x\boldsymbol{\xi}$, where the $\mathbf{G}$ matrices are known selection matrices with 0 or 1 elements to select the observed from all the variables. Define now $\mathbf{z}' = (\mathbf{y}' \quad \mathbf{x}')$, $\mathbf{G}$ as a block-diagonal matrix with diagonal blocks $\mathbf{G}_y$ and $\mathbf{G}_x$, and $\boldsymbol{\upsilon}' = (\boldsymbol{\eta}' \quad \boldsymbol{\xi}')$. Then we can write

$$\boldsymbol{\upsilon} = \mathbf{B}_0\boldsymbol{\upsilon} + \boldsymbol{\Gamma}\boldsymbol{\xi}$$

$$\mathbf{z} = \mathbf{G}\boldsymbol{\upsilon},$$

where $\boldsymbol{\Gamma}' = (\boldsymbol{\gamma}' \quad \mathbf{I})$, and $\mathbf{B}_0$ has rows $(\boldsymbol{\beta}_0 \quad \mathbf{0})$ and $(\mathbf{0} \quad \mathbf{0})$. As a result, the measured variables are a linear combination of independent variables

$$\mathbf{z} = \mathbf{G}\mathbf{B}^{-1}\boldsymbol{\Gamma}\boldsymbol{\xi} = \mathbf{A}\boldsymbol{\xi},$$

where, assuming that $(\mathbf{I} - \mathbf{B}_0)$ is invertible, $\mathbf{B}^{-1} = (\mathbf{I} - \mathbf{B}_0)^{-1}$. Thus the observed variables are a linear combination of the independent variables. We deal only with latent variable models, so the number of variables and dimensionality of $\mathbf{z}$ is less than that of $\boldsymbol{\xi}$ (see Bollen, 2002).

Interaction and quadratic factors are specified by defining some elements of vector $\boldsymbol{\xi}$ as products of other elements in $\boldsymbol{\xi}$. For instance, a quadratic factor could be ξ_1^2, and an interaction factor $\xi_1\xi_2$. In general, with any number of such factors, the vector $\boldsymbol{\xi}$ can be written as

$$\boldsymbol{\xi} = \begin{pmatrix} 1 \\ \boldsymbol{\xi}_M \\ \boldsymbol{\xi}_{INT} \\ \boldsymbol{\xi}_{QUAD} \end{pmatrix}, \tag{1}$$

where $\boldsymbol{\xi}_M$ denotes the "main" factors, $\boldsymbol{\xi}_{INT}$ as the interaction factors, and $\boldsymbol{\xi}_{QUAD}$ as the quadratic factors. The element 1 in vector $\boldsymbol{\xi}$ is a constant, thus providing a convenient way to deal with the means of the variables and with intercepts in regression equations. An important assumption we make in our model formulation is that the vectors in $\boldsymbol{\xi}_M$ are normally distributed with mean zero; in the Discussion we note how this assumption can be relaxed. A consequence of this assumption is that the covariances and third-order moments of the interaction and quadratic factors can be written in terms of the covariances of the main factors.

Now, as in Bentler (1983) and Mooijaart (1985), we obtain the first-, second-, and third-order moments

$$\begin{aligned}
\boldsymbol{\sigma}_1 &\equiv \mathbf{D}_p^+ E[\mathbf{z}] = \mathbf{A}E[\boldsymbol{\xi}] = \mathbf{A}\boldsymbol{\varphi}_1 \\
\boldsymbol{\sigma}_2 &\equiv \mathbf{D}_p^+ E[(\mathbf{z}-\boldsymbol{\sigma}_1)\otimes(\mathbf{z}-\boldsymbol{\sigma}_1)] = \mathbf{D}_p^+(\mathbf{A}\otimes\mathbf{A})\mathbf{D}_r\boldsymbol{\varphi}_2 \qquad (2) \\
\boldsymbol{\sigma}_3 &\equiv \mathbf{T}_p^+ E[((\mathbf{z}-\boldsymbol{\sigma}_1)\otimes(\mathbf{z}-\boldsymbol{\sigma}_1)\otimes(\mathbf{z}-\boldsymbol{\sigma}_1))] = \mathbf{T}_p^+(\mathbf{A}\otimes\mathbf{A}\otimes\mathbf{A})\mathbf{T}_r\boldsymbol{\varphi}_3.
\end{aligned}$$

In Equation 2, $\boldsymbol{\sigma}_1$ defines the model equation for the means of the observed variables, and $\boldsymbol{\sigma}_2$ and $\boldsymbol{\sigma}_3$ define the model equations of the covariances and third-order moments of the observed variables. These population moments are all expressed as a weighted combination of the means $\boldsymbol{\varphi}_1$, covariances $\boldsymbol{\varphi}_2$, and third moments $\boldsymbol{\varphi}_3$ of the independent variables $\boldsymbol{\xi}$, for both the main factors as well as the quadratic and interaction factors. In the preceding definitions, the vectors $\boldsymbol{\sigma}_2$, $\boldsymbol{\sigma}_3$, $\boldsymbol{\varphi}_2$, $\boldsymbol{\varphi}_3$ are in reduced form; that is, vectors without the duplicated and triplicated elements. For properties of the duplication matrix $\mathbf{D}_p$ see Magnus and Neudecker (1988), and for the triplication matrix $\mathbf{T}_p$ see Meijer (2005).

In our model setup, any type of SEM can be accommodated, not only a specialized model form such as is required by Cudeck et al. (2009). Our only requirement is that some elements in $\boldsymbol{\varphi}_2$ and $\boldsymbol{\varphi}_3$ are taken as functions of the covariances of the main factors that are normally distributed. The Appendix gives some basic equations by which these functions are defined, and provides components needed to compute asymptotic distributions.

ESTIMATION AND TESTING

In this section, we first describe the statistical methodology, and then discuss the proposed methodology in relation to the ML method. For estimating the model parameters, we use the well-known SEM quadratic discrepancy function

$$f_H = f(\mathbf{s}, \boldsymbol{\sigma}|H) = (\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\theta}))'\mathbf{W}(\mathbf{s}-\boldsymbol{\sigma}(\boldsymbol{\theta})), \qquad (3)$$

where $\mathbf{s}$ is the vector of means, sample covariances, and selected third-order sample moments;[2] $\boldsymbol{\sigma}(\boldsymbol{\theta})$ is the corresponding vector of model moments as a function of all free parameters on the right side of Equation 2 placed in the vector $\boldsymbol{\theta}$, $\mathbf{W}$ is some weight matrix, and H is some special model, which might be the true model or not. We make the typical large sample distributional assumption

$$\sqrt{n}(\mathbf{s}-\boldsymbol{\sigma}) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Gamma}). \qquad (4)$$

[2]A sample third-order moment can be computed as averages of triple products of deviation scores $s_{ijk} = N^{-1}\Sigma_1^N(z_{it}-\bar{z}_i)(z_{jt}-\bar{z}_j)(z_{kt}-\bar{z}_k)$, where indexes i, j, and k can be the same or different. In EQS, such a moment is designated as (V_i, V_j, V_k).

It can be shown that the covariance matrix of the estimator is given by

$$\text{avar}(\hat{\boldsymbol{\theta}}) = n^{-1}(\dot{\boldsymbol{\sigma}}'\mathbf{W}\dot{\boldsymbol{\sigma}})^{-1}\dot{\boldsymbol{\sigma}}'\mathbf{W}\boldsymbol{\Gamma}\mathbf{W}\dot{\boldsymbol{\sigma}}(\dot{\boldsymbol{\sigma}}'\mathbf{W}\dot{\boldsymbol{\sigma}})^{-1}, \tag{5}$$

where $\dot{\boldsymbol{\sigma}}$ is the matrix of derivatives of $\boldsymbol{\sigma}$ with respect to the vector of parameters $\boldsymbol{\theta}$. Because we are dealing with third-order moments as well, matrix $\boldsymbol{\Gamma}$ will contain some elements of order 6. These elements can be quite unstable, in particular in small samples, and therefore we proceed in practice by specifying matrix $\mathbf{W}$ as the identity matrix. This means that we utilize the least squares discrepancy function. The corresponding test statistic can be written as

$$T_{LS} = n(\mathbf{s} - \hat{\boldsymbol{\sigma}})'(\hat{\boldsymbol{\Gamma}}^{-1} - \hat{\boldsymbol{\Gamma}}^{-1}\hat{\dot{\boldsymbol{\sigma}}}(\hat{\dot{\boldsymbol{\sigma}}}'\hat{\boldsymbol{\Gamma}}^{-1}\hat{\dot{\boldsymbol{\sigma}}})^{-1}\hat{\dot{\boldsymbol{\sigma}}}'\hat{\boldsymbol{\Gamma}}^{-1})(\mathbf{s} - \hat{\boldsymbol{\sigma}}), \tag{6}$$

where $\hat{\boldsymbol{\Gamma}}$ is a consistent estimator of $\boldsymbol{\Gamma}$ and $\hat{\dot{\boldsymbol{\sigma}}}$ is the Jacobian matrix evaluated at the estimator $\hat{\boldsymbol{\theta}}$. If the model is correctly specified, T_{LS} is central chi-square distributed, otherwise T_{LS} is noncentral chi-square distributed. In both cases the degrees of freedom is equal to the total number of means, second- and third-order moments minus the number of independent parameters. EQS also has implemented an appropriate Satorra–Bentler (1994) scaling correction to Equation 6 and the Yuan–Bentler (1998) residual-based F test. If the model is misspecified, and model H_0 is nested within model H, the noncentrality parameter of Equation 6 is equal to

$$\lambda(T_{LS}|H, H_0) = n(\boldsymbol{\sigma}_H - \hat{\boldsymbol{\sigma}}_{H_0})'(\boldsymbol{\Gamma}^{-1} - \boldsymbol{\Gamma}^{-1}\hat{\dot{\boldsymbol{\sigma}}}_{H_0}(\hat{\dot{\boldsymbol{\sigma}}}'_{H_{0'}}\boldsymbol{\Gamma}^{-1}\hat{\dot{\boldsymbol{\sigma}}}_{H_0})^{-1}\hat{\dot{\boldsymbol{\sigma}}}_{H_0}\boldsymbol{\Gamma}^{-1})(\boldsymbol{\sigma}_H - \hat{\boldsymbol{\sigma}}_{H_0}), \tag{7}$$

where $\boldsymbol{\sigma}_H$ is the vector of moments in the population, that is, under the correct model H, and $\hat{\boldsymbol{\sigma}}_{H_0}$ the estimate of $\boldsymbol{\sigma}_H$ under the misspecified model. For a more detailed development, regularity conditions, and proofs on the statistical theory underlying Equations 4 through 7, see Browne (1984), Bentler and Dijkstra (1985), and Satorra (1989).

Suppose a model H_0 has been fitted but it is unclear whether this model is a correct model or a misspecified model. Because H_0 is nested within model H, model H_0 must have restrictions on the parameter vector $\boldsymbol{\theta}$ that can be released. A test for checking whether a restriction has to be released is the so-called Lagrange multiplier test. In particular, we are interested in checking whether the model fit can be improved by adding an interaction or quadratic term to the model.

If Equation 3 is minimized under some restrictions that can be written as $c_i(\boldsymbol{\theta}) = 0$, for $i = 1, \ldots, r$, using Lagrange multipliers it is known that first derivatives of the constrained function imply

$$\mathbf{g} + \dot{\mathbf{c}}'\boldsymbol{\lambda} = \mathbf{0}, \tag{8}$$

where vector $\boldsymbol{\lambda}$ is a $(r \times 1)$ vector with Lagrange multipliers, g is the gradient of Equation 3 with respect to the parameters $\boldsymbol{\theta}$, and $\dot{\mathbf{c}}(\boldsymbol{\theta}) = \partial\mathbf{c}/\partial\boldsymbol{\theta}'$. For a more detailed discussion, Bentler and Dijkstra (1985) and Satorra (1989). As shown in Bentler (2000–08), it follows that

$$\hat{\boldsymbol{\lambda}} = \mathbf{D}'\dot{\boldsymbol{\sigma}}'(\mathbf{s} - \boldsymbol{\sigma}(\hat{\boldsymbol{\theta}})), \tag{9}$$

where $\mathbf{D}' = (\dot{\mathbf{c}}\boldsymbol{\Delta}\dot{\mathbf{c}}')^{-1}\dot{\mathbf{c}}\boldsymbol{\Delta}$ with $\boldsymbol{\Delta} = (\dot{\boldsymbol{\sigma}}'\dot{\boldsymbol{\sigma}})^{-1}$, and the Lagrange multiplier test can be written as

$$LM = n\hat{\boldsymbol{\lambda}}'(\mathbf{D}'\dot{\boldsymbol{\sigma}}'\boldsymbol{\Gamma}\dot{\boldsymbol{\sigma}}\mathbf{D})^{-1}\hat{\boldsymbol{\lambda}}. \tag{10}$$

Under the assumption that the restrictions $c_i(\boldsymbol{\theta}) = 0$ are correctly specified, the Lagrange multiplier statistic (Equation 10) is central chi-square distributed with r degrees of freedom. This test is used primarily to test whether a set of restrictions is correctly specified or not.

A few words are in order on our methodology as compared to potential alternatives that exist in the SEM literature. First, although we proposed using the Lagrange multiplier test (Equation 10), another standard approach in SEM, especially with ML, is to implement a model difference test when comparing two nested models. With our method, such a test is immediately available by standard moment structure theory (Satorra, 1989). However, when the variables are not normally distributed, it is unclear whether the ML difference test gives a statistic that is chi-square distributed. Under some regularity conditions for nonnormally distributed variables the likelihood ratio test gives a statistic that is asymptotically chi-square distributed. However, when two models are compared where one model contains interaction parameters and the other does not, the situation becomes more complex. The reason is that in this latter case one of the likelihood functions is related to the normal distribution and the other likelihood function to a nonnormal distribution. Theoretically in such a case it is unclear what the distribution of the likelihood ratio is, in particular for nonlarge samples. Evidently this is a topic for further research, as Cudeck et al. (2009) computed such a chi-square difference test based on their marginal ML method.

Next consider testing the goodness-of-fit of the model. With normally distributed variables, the sufficient statistics of the multivariate distribution are the means and the covariances. Assuming that there always is a saturated model that fits the data perfectly (in particular a model that has the same vector of means and covariance matrix as in the sample), a difference test between the hypothesized and this saturated model is in fact a test whether the model fits the data or not. However, this reasoning does not work in the case of non-normally distributed variables, because in general then there might be no model that fits the data perfectly and so testing against such a saturated model is impossible. As a consequence, the ML method that could be used for the models discussed in this article will not give a statistical test to evaluate whether a model fits the data. Cudeck et al. (2009) did not provide such a model fit test. This limitation does not exist in our proposed method. We fit a set of moments (of the first, second, third order) for which there always is a saturated model. Hence our model test is in fact a test of whether the model fits the chosen moments perfectly or not. Next we turn to some studies of the proposed methodology.

STUDY 1

In this example we reanalyze a model given by Sun and Willson (2009) and Muthén and Asparouhov (2003). The model is a four-time point latent growth model with interaction between “initial status” and a covariate. In this model it is assumed that the “slope” of the growth model is determined by the initial status, the covariate, and the interaction between the initial status and the covariate.

The model equations are

$$V_1 = 1F_1 + 0F_2 + E1$$

$$V_2 = 1F_1 + 1F_2 + E2$$

$$V_3 = 1F_1 + 2F_2 + E3$$

$$V_4 = 1F_1 + 3F_2 + E4$$

$$F_2 = \beta_0 + \beta_1 F_1 + \beta_3 F_3 + \beta_{13} F_1 F_3 + D_2$$

In this model variable F_3 is an observed covariate, however, we treat it as a latent variable that is equal to an observed variable (i.e., V5 = F3). As usual, the factor loadings of the V variables on the growth curve factors F are fixed values, either 0, 1, 2, or 3.

Data are generated according to the preceding model, which is exactly equal to the one given by Sun and Willson (2009), who give, in an appendix, the M*plus* code for generating the data. We do make a minor adjustment to the Sun and Willson setup. In their study, the mean of their covariate, our factor F_3, is not equal to zero. This means that in the interaction term one of the factors has mean unequal to zero, which is not acceptable in our approach. As is well known (Aiken & West, 1991, ch. 3), changing the mean of a variable involved in an interaction changes the coefficients for the other noninteracting variables. Here, when the mean of F_3 is shifted by a certain amount, say α, the equation for F_2 is modified as

$$\begin{aligned} F_2 &= \beta_0 + \beta_1 F_1 + \beta_3 F_3 + \beta_{13} F_1 F_3 + D_2 \\ &= \beta_0 + \beta_1 F_1 + \beta_3(\alpha + F_3) + \beta_{13} F_1(\alpha + F_3) + D_2 \\ &= (\beta_0 + \alpha\beta_3) + (\beta_1 + \alpha\beta_{13})F_1 + \beta_3 F_3 + \beta_{13} F_1 F_3 + D_2 \end{aligned}$$

Although the interaction coefficient is not affected, the intercept and the regression weight of F_1 are modified; that is, they are not free of the intercept of F_3. In our study we shifted the covariate such that the mean is equal to zero. The interaction effect for a covariance of .5 between the factors F_1 and F_3 is 4.97%.

We first analyze the data without an interaction factor by the ML method; that is, we estimate a misspecified model. Then, because some of the variables are nonnormally distributed, we analyze data with a third-order moment and apply the Lagrange multiplier test to get an indication of whether there is an interaction factor or not. Finally, we add an interaction factor to the model and estimate all the model parameters.

The estimated structural part of the misspecified model is:

$$F_2 = \underset{(.013)}{.500} - \underset{(.026)}{.004}\, F_1 + \underset{(.020)}{.049}\, F_3 + D_2\ ,$$

where the estimated standard errors are given in parentheses. The model chi-square statistic is 9.384 with 8 *df*. This indicates (as do all the other fit indexes) that the model without the interaction term fits the data very well. This result seems to be remarkable, because the ML method suggests that the model fits a linear model, where it is known that the model is nonlinear. The explanation is that the ML method and the corresponding likelihood ratio test is, under some conditions that are satisfied here, not sensitive to detect nonlinear terms (see Mooijaart & Satorra, 2009a).

Variables V2, V3, and V4 are not normally distributed, and the skewness of these variables around −.30. Hence we now analyze the data with the same misspecified model, but add factor $F_4 = F_1 F_3$ with coefficient zero; we also add one third-order moment to the data to be modeled. This gives Lagrange multiplier = 12.232 for the coefficient of F_4, which is significant when referred to χ^2_1. Thus the Lagrange multiplier test indicates that there is an interaction factor that explains the nonnormality significantly. Estimating the model with the coefficient of this interaction factor as a free parameter yields

$$\begin{array}{lllll} F_2 = & .538 & + .006\ F_1 & + .054\ F_3 & - .088\ F_4 + D_2 \\ & (.012) & (.026) & (.021) & (.022) \end{array}$$

The goodness-of-fit test yields a value of 9.396, evaluated with χ^2_8. This also indicates that the model fits the data, but the difference here is that not only the means and the covariances are fitted, but also the third-order moment V4, V4, V5. Furthermore, as expected, the interaction effect is significant, because the estimate of the regression weight of F4 has a standard error of .022.

The corresponding model with ML, as obtained from M*plus,* gives

$$\begin{array}{lllll} F_2 = & .533 & + .003\ F_1 & + .047\ F_3 & - .072\ F_4 + D_2 \\ & (.014) & (.024) & (.021) & (.016) \end{array}$$

It is interesting to note that both ML and our method find that the coefficient of F_2 on F_1 is very small and not statistically significant.

Conclusion

The Lagrange multiplier test indicates that there is a significant interaction effect that was unnoticed by the standard ML method. Fitting the model with one higher order moment verifies that a significant interaction exists. The final model parameters are comparable to those of the ML method with interaction effects.

STUDY 2

In this study we reanalyze the same model as in Study 1, however, now we set up a small simulation study with 200 replications. Some key results are given in Tables 1 and 2.

Table 1 shows that the estimation algorithm converged without any problem in all 200 replications. The two methods hardly differ in terms of the bias of the estimates; it is clear that they are unbiased. Except for the regression weight of the interaction factor, the standard errors of the parameters are in both methods about equal. The main difference between the methods is the size of the interaction standard error. In general, the ML standard errors are smaller, as would be expected from large sample theory.

TABLE 1
Results of Simulation: Bias, Standard Deviation, and Standard Error

		Generalized Least Squares			*M*plus		
Parameters	*True*	*Bias*	*SD*	*SE*	*Bias*	*SD*	*SE*
Covariate mean							
α	2.300	−.003	.033	.033	−.003	.033	.034
Factor variances							
F1	.850	.000	.067	.063	.009	.066	.064
F3	.820	.005	.051	.051	−.001	.049	.052
Factor covariances							
F1, F3	.500	.003	.049	.045	.007	.048	.046
Error variances							
E1	.200	−.002	.026	.025	.002	.024	.026
E2	.200	.000	.018	.017	−.001	.018	.017
E3	.200	−.001	.018	.018	.001	.017	.018
E4	.200	−.001	.032	.028	.002	.032	.028
Disturbance variance							
D2	.040	−.001	.006	.006	−.001	.006	.006
Intercept							
β_0	.524	.000	.016	.012	.001	.014	.014
Regression weights, main							
β_1	−.023	−.002	.021	.022	.000	.019	.022
β_3	.045	.001	.019	.019	.000	.018	.019
Regression weight, interaction							
β_{13}	−.048	.001	.022	.021	−.002	.013	.013

Note. The execution time for 200 replications was 19 sec for EQS and 30 sec for M*plus*.

An important difference between ML and our method is that our method gives a goodness-of-fit test for the model. Table 2 shows that the mean of the model chi-square test and the corrected chi-square statistic are in the 95% confidence interval, and shows that the proportion of rejections is also in the 95% confidence interval. These are promising results and provide useful information beyond that given by M*plus,* which is still missing a test statistic. Of course, M*plus* gives the likelihood function values and some information criteria, like the Akiake's Information Criterion (AIC), Bayesian Information Criterion (BIC) and adjusted BIC criteria that can be used to choose among a set of potential models.

TABLE 2
Some Information on Goodness-of-Fit

Degrees of freedom	8
Chi-square statistic	8.47
Number of rejections, $\alpha = 5\%$	13 (6.5%)
Corrected chi-square statistic	8.29
Number of rejections, $\alpha = 5\%$	10 (5.0%)
F statistic	1.04
Number of rejections, $\alpha = 5\%$	10 (5.0%)

Note. 95% confidence interval: around mean chi-square distribution $df = 8$: 7.45–8.55; around $\alpha = 5\%$: 1.98–8.02%.

Conclusion

The estimates of the parameters seem to be unbiased for our method. Compared to the ML estimates, the estimated standard errors are larger, in particular for the interaction parameter. With respect to the mean of the chi-square test and the proportion of rejections, the model chi-square test behaves as theoretically expected.

STUDY 3

Next we study the power of the proposed goodness-of-fit model test, by evaluating the behavior of empirical power. Power is important for model selection and therefore it is important to investigate whether our proposed test discriminates between alternative models. In our study we reanalyze data that follow the so-called Kenny and Judd (1984) model, which was also analyzed by others (e.g., Klein & Moosbrugger, 2000). In our simulation study, we use the same setup as they did for their study of the Latent Moderated Structural Equations (LMS) method.

The model contains two latent variables that predict an observed variable where, besides the main effects of the predictors, there is also an interaction effect. The two latent predictors each have two observed indicators. The parameter that plays a key role is the "interaction" parameter. In our study this parameter will vary from .0 to .7. The sample size is 400 and there are 200 replications. In one study the third-order moment V5, V5, V5 is analyzed; in another study two third-order moments V5, V5, V5 and V1, V3, V5 are analyzed.

For each different value of the interaction parameter, the behavior of the X^2 statistic is presented by the mean of X^2 over the 200 replications and the proportion of rejections based on a nominal α-level of 5%. Obviously, if there is no misspecification of the model, the proportion of rejections of the model is the empirical α-level, otherwise it is the empirical power.

Because 200 replications are used, the means and the proportions will vary around some unknown value. Therefore, the 95% confidence intervals of the estimates are given. The 95% confidence intervals are: for $\alpha = 5\%$: 1.98–8.02%. The 95% confidence interval around the mean of some chi-square distributions are: for $df = 7$: 6.48–7.52, for $df = 8$: 7.45–8.55, and for $df = 9$: 8.41–9.59.

Tables 3, 4, and 5 show five main results. First, with standard ML, the empirical α is not significant from the nominal α, see Table 3, column "%rej." So even if the model is

TABLE 3
Results of Simulation: Maximum Likelihood Estimation

β_{12}	X_7^2	*% rej.*
.0	6.90	4.6 (2)
.1	6.61	3.1 (2)
.2	6.59	2.5 (3)
.4	6.54	2.6 (4)
.7	6.50	2.6 (4)

Note. Numbers in parentheses are the number of data sets that resulted in nonconvergence of the algorithm.

TABLE 4
Results of Simulation with Third-Order Moment V5, V5, V5

		β_{12} *Free*				$\beta_{12} = 0$	
β_{12}	X_7^2	*% rej.*	*Bias*	*SD*	*SE*	X_8^2	*% rej.*
.0	6.74	3.6 (3)	.00	.10	.10	7.80	6.1 (2)
.1	6.72	3.1 (3)	.00	.10	.10	8.28	6.6 (3)
.2	6.70	3.1 (3)	−.01	.10	.11	9.70	11.7 (4)
.4	6.66	4.1 (3)	−.01	.12	.12	12.12	22.0 (9)
.7	6.54	4.1 (6)	−.03	.16	.15	13.47	28.5 (14)

Note. Numbers in parentheses are the number of data sets that resulted in nonconvergence of the algorithm.

a nonlinear model (i.e., the interaction parameter is unequal to 0) and some variable is not normally distributed, the standard ML method indicates that a linear model fits the data. This is known from Mooijaart and Satorra (2009a) but is verified in this study. Although all means of the test statistics are smaller than the theoretical mean of their distributions, none of them is outside the range of the 95% confidence interval. Second, Tables 4 and 5 compare the results of our two studies with different third-order moments. Both tables show that some (in total 5) of the parameters are outside the 95% confidence interval. All other estimates are within the proper 95% interval. This falling outside the confidence interval is caused by the small sample size. For $N = 500$, all estimates were inside the 95% interval.

The left side of Tables 4 and 5 (columns 2–6) refers to correctly specified models, whereas the right side (columns 7–8) refers to the model in which the interaction parameter was set to zero. Thus column 3 refers to the empirical α, and column 8 to the empirical power. All empirical αs are fine, and the bias of the estimates is also small. However, Tables 4 and 5 differ with respect to empirical power. For instance, in Table 4 for $\beta_{12} = .7$ the empirical power is 28.5%, whereas in Table 5 the corresponding power is 87.4%. It might also be the case that another difference between the two tables is that standard errors of the estimate are smaller in Table 5.

TABLE 5
Results Simulation with Third-Order Moments V1, V3, V5 and V5, V5, V5

		β_{12} *Free*				$\beta_{12} = 0$	
β_{12}	X_8^2	*% rej.*	*Bias*	*SD*	*SE*	X_9^2	*% rej.*
.0	7.43[a]	3.1 (4)	.00	.08	.08	9.22	6.6 (2)
.1	7.41[a]	3.1 (4)	−.01	.09	.08	10.03	8.6 (3)
.2	7.39[a]	3.6 (3)	−.01	.09	.09	12.56	20.4 (4)
.4	7.45[a]	4.0 (2)	−.02	.12	.10	18.60	53.1 (8)
.7	7.41[a]	4.1 (3)	−.04	.16	.14	26.26	87.4 (9)

Note. Numbers in parentheses are the number of data sets that resulted in nonconvergence of the algorithm.
[a]The corresponding estimate is outside the 95% confidence interval.

Conclusion

With respect to the Type I level, the goodness-of-fit model test behaves as theoretically expected. The choice of the third-order moments might have an influence on the power of the test. However, this is not completely clear, so more study on the effect of choice and number of third-order moments is needed in the future.

STUDY 4

In this study, we evaluate the effect of nonnormality of the indicators. We study a factor analysis model with three factors and three indicators for each factor. Some of the indicators might not be normally distributed. These nonnormally distributed indicators load on one of the main factors and on one of the interaction factors. Again we make a comparison of the results for our method and those of the ML method.

For identification, Variables 1, 4, and 7 have factor loadings fixed equal to 1. We take the sample size at 500, and the number of replications at 100. In our method, we use three third-order moments for fitting the model, namely those of the following triples of variables: (V1, V3, V4), (V4, V6, V7), and (V4, V7, V9). Tables 6 and 7 present the results.

The results show that the estimation algorithm converged without any problem in all 100 replications with both methods, although the execution time (over 100 replications) is much longer with M*plus*. The two methods hardly differ in terms of the parameter estimate bias; it is clear that both are unbiased. The mean of the standard deviations and the standard errors are also almost equal in both methods, indicating that the estimated standard errors can be trusted in both methods.

The main difference between the methods is in the size of the standard errors of the factor loadings on the interaction factors. In general, the estimated standard errors by ML are smaller than those obtained by our method. As might be expected, the model chi-square test seems to reject the model too often. This is a commonly known effect with this test, in particular with small samples and large models. However, the corrected chi-square and the F test seem to behave nicely.

Conclusion

Nonnormality of the indicators can be specified in different ways. In this example, the nonnormality of the indicators is determined by the interaction factors. It is clear that this type of model can be analyzed easily with our method and that the results do not differ from the results in the previous simulation studies.

DISCUSSION

For practitioners, an important feature of the proposed methodology is its continuity with existing SEM methods. This continuity includes development of an explicit structural model for moments of the observed variables, measuring model fit as a weighted residual between

TABLE 6
Results of Simulation

		Generalized Least Squares			*M*plus		
Parameters	*True*	*Estimate*	*SD*	*SE*	*Estimate*	*SD*	*SE*
Intercepts							
V1	1.0	1.01	.05	.04	1.01	.05	.05
V2	1.0	1.01	.05	.04	1.01	.04	.04
V3	1.0	1.01	.05	.04	1.01	.04	.04
V4	1.0	1.01	.06	.04	1.00	.05	.05
V5	1.0	1.01	.05	.04	1.01	.04	.04
V6	1.0	1.01	.05	.04	1.00	.04	.04
V7	1.0	1.01	.06	.04	1.00	.05	.05
V8	1.0	1.00	.05	.04	1.00	.04	.04
V9	1.0	1.01	.06	.04	1.00	.05	.04
Factor variances							
F1	1.0	.98	.09	.08	1.01	.09	.09
F2	1.0	.99	.08	.08	1.01	.07	.08
F3	1.0	.99	.08	.08	1.01	.08	.08
Factor covariances							
F1, F2	.5	.50	.05	.05	.51	.05	.06
F1, F3	.5	.49	.06	.05	.50	.06	.06
F2, F3	.5	.50	.06	.05	.51	.06	.06
Error variances							
E1	.3	.30	.04	.04	.30	.04	.03
E2	.3	.29	.03	.03	.30	.03	.03
E3	.3	.29	.04	.04	.30	.02	.03
E4	.3	.29	.03	.03	.29	.03	.03
E5	.3	.29	.03	.03	.30	.02	.03
E6	.3	.30	.03	.03	.30	.03	.03
E7	.3	.29	.04	.03	.30	.04	.03
E8	.3	.30	.03	.03	.30	.03	.03
E9	.3	.29	.04	.03	.30	.03	.03
Factor loadings, main							
V2, F1	.8	.80	.04	.04	.80	.03	.04
V3, F1	.8	.80	.05	.04	.80	.05	.04
V5, F2	.8	.80	.03	.04	.80	.03	.04
V6, F2	.8	.80	.05	.04	.80	.04	.04
V8, F3	.8	.80	.04	.04	.80	.03	.04
V9, F3	.8	.80	.05	.04	.80	.04	.04
Factor loadings, interaction							
V3, F1×F2	.3	.30	.06	.06	.30	.03	.03
V6, F2×F3	.3	.29	.05	.04	.30	.03	.03
V9, F2×F3	.3	.30	.05	.04	.30	.03	.03

Note. The execution time for 100 replications was 59 sec for EQS and 10,049 sec for M*plus.*

sample moments and a model-implied moment structure, and the availability of standard statistical methods for model evaluation and modification. We illustrated this by developing a fairly standard SEM model goodness-of-fit test and a Lagrange multiplier test to evaluate the necessity of nonlinear terms in an equation. These tests were shown to have sufficient power to pick up nonlinear effects. In fact, because our methodology is just an application of moment structure analysis, an ordinary SEM user does not need much specialized knowledge to use the methodology. Further, our approach makes it possible to extend most useful existing

TABLE 7
Some Information on Goodness-of-Fit

Degrees of freedom	24
Chi-square statistic	25.35[a]
Number of rejections, $\alpha = 5\%$	11[b]
Corrected chi-square statistic	24.02
Number of rejections, $\alpha = 5\%$	5
F statistic	1.007
Number of rejections, $\alpha = 5\%$	6

[a]The corresponding estimate is outside the 90% confidence interval, inside the 95% confidence interval.
[b]The corresponding estimate is outside the 95% confidence interval.

SEM methodology for estimation, testing, sample-size free-fit evaluation, and so on, to models with quadratic and interaction latent variables. To illustrate, once the theory for the residual-based test (see Equation 6) was developed and implemented in EQS, the program more or less automatically also produced the Satorra–Bentler correction and Yuan–Bentler residual-based F statistic that behaved quite well under violation of distributional assumptions, as well as a variety of typical fit indexes (not reported) that are a standard part of SEM.

The tests proposed and studied in this article seem to perform well in practice, although, of course, further research with a wider range of conditions is needed to determine their boundary conditions. One of these conditions will no doubt be sample size, because the asymptotic distribution of third-order moments is unstable at smaller sample sizes. Of course, if sample size is not large enough to trust a third-order moment to reflect nonlinearity, it might not make sense to consider modeling such nonlinearity with latent variables in the first place—the model structure we developed in Equation 2 will hold in any case, even if the estimation method is not a standard moment structure method. Another boundary condition no doubt will involve the optimal selection of third-order moments to model, as well as the number of such moments. Because the chosen moments must reflect the nonnormality of the indicators of the quadratic and interaction factors, moments should be chosen to reflect the degree of skewness of those variables. Although our simulations did not yield definitive conclusions, it seems that using more than one third-order moment yields greater power. We suspect that modeling two or three marginal or joint moments will usually be sufficient, with more being appropriate when several variables are quite skewed. Additional recommendations for choice of moments are to involve those variables that have the most reliable indicators (Saris, Batista-Foguet, & Coenders, 2007) and to choose those moments that yield maximum power (Mooijaart & Satorra, 2009b). Further research will be needed to give definitive recommendations, but in any case, the selection of such moments can and should be done based on current knowledge and implemented automatically by the computer program used to fit the model. Of course, there should be an option for the user to override any default, as is done in EQS.

We do not claim that our methodology always will give the “best” results—we mainly want to emphasize that we have provided a new and promising approach to nonlinear models. In small samples, methods that incorporate additional information, especially Bayesian methods (e.g., Lee et al., 2007), are liable to have more stable estimates and power to detect nonlinearities.

These advantages, of course, have to be traded off against increased difficulty in implementation and in current limitations that only allow a single dependent nonlinear factor. The advantages of the Bayesian methods also will disappear in large samples, because the influence of prior distributions vanishes asymptotically. ML (Klein & Moosbrugger, 2000) certainly is valuable for its asymptotic optimality. In our simulations, ML standard errors were consistently smaller by a small amount as compared to our method. Although ML computations can be difficult when there are several nonlinear factors, the recently developed marginal ML methodology of Cudeck et al. (2009) promises to be computationally more feasible in certain situations. Another useful general alternative is quasi-ML (Klein & Muthén, 2007) because it, like the ML methods, provides difference tests. However, standard SEM statistics such as overall model goodness-of-fit chi-square statistics and Lagrange multiplier tests are not currently available in these methods.

All issues in nonlinear SEM models certainly have not been settled. A standard assumption of quadratic and interaction models, including ours, is that nonnormality in all variables stems from the underlying latent nonlinear factors. Although we found that our method possesses some robustness to violation of this assumption, and appropriate robust model tests performed well, all existing methods need extension to situations where some main factors, errors of measurement, and variables in the model might be nonnormal over and above any nonnormality that is due to quadratic or interaction factors. One source of such excess nonnormality might be a nonnormally distributed covariate. Although we have preliminary results showing that such an approach can give acceptable results by our method, detailed study of this and other approaches remains a challenge for the future.

ACKNOWLEDGMENTS

This research was supported in part by National Institute on Drug Abuse grants DA00017 and DA01070. Peter M. Bentler acknowledges a financial interest in EQS and its distributor, Multivariate Software.

REFERENCES

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Arminger, G., & Muthén, B. (1998). A Bayesian approach to nonlinear variable models using the Gibbs sampler and the Metropolis–Hastings algorithm. *Psychometrika, 63,* 271–310.

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

Bentler, P. M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. *Psychometrika, 48,* 493–517.

Bentler, P. M. (2000–2008). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Kirshnaiah (Ed.), *Multivariate analysis VI* (pp. 9–42). Amsterdam: North Holland.

Bentler, P. M., & Weeks, D. G. (1980). Linear structural equations with latent variable. *Psychometrika, 45,* 289–308.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53,* 605–634.

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

Bollen, K. A., & Paxton, P. (1998b). Two-stage least squares estimation of interaction effects. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (pp. 125–151). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Coenders, G., Batista-Foguet, J. M., & Saris, W. E. (2008). Simple, efficient and distribution-free approach to interaction effects in complex structural equation models. *Quality & Quantity, 42,* 369–396.

Cudeck, R., Harring, J. R., & du Toit, S. H. C. (2009). Marginal maximum likelihood estimation of a latent variable model with interaction. *Journal of Educational and Behavioral Statistics, 34,* 131–144.

Dimitruk, P., Schermelleh-Engel, K., Kelava, A., & Moosbrugger, H. (2007). Challenges in nonlinear structural equation modeling. *Methodology, 3,* 100–114.

Hargens, L. L. (2009). Product-variable models of interaction effects and causal mechanisms. *Social Science Research, 38,* 19–28.

Jaccard, J. J., & Wan, C. K. (1996). *LISREL approaches to interaction effects in multiple regression.* Thousand Oaks, CA: Sage.

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny–Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 57–88). Mahwah, NJ: Lawrence Erlbaum Associates.

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., & Muthén, B. O. (2007). Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research, 42,* 647–673.

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

Lee, S.-Y., Song, X.-Y., & Tang, N.-S. (2007). Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. *Structural Equation Modeling, 14,* 404–434.

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

Magnus, J. R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics.* Chichester, UK: 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. (1985). Factor analysis for non-normal variables. *Psychometrika, 50,* 323–342.

Mooijaart, A. (2008). Random polynomial and interaction latent variable models. In K. Shigemasu, A. Okada, T. Imaizumi, & T. Hoshino (Eds.), *New Trends in Psychometrics* (pp. 303–310). Tokyo: Universal Academy Press.

Mooijaart, A., & Satorra, A. (2009a). On insensitivity of the chi-square model test to nonlinear misspecification in structural equation models. *Psychometrika, 74,* 443–455.

Mooijaart, A., & Satorra, A. (2009b, July). *Testing for non-linear relationship in structural equation modeling.* Paper presented at the International Meeting of the Psychometric Society, Cambridge, UK.

Mulaik, S. A. (2009). *Linear causal modeling with structural equations.* Boca Raton, FL: Chapman & Hall/CRC.

Muthén, B. O., & Asparouhov, T. (2003). Modeling interactions between latent and observed continuous variables using maximum-likelihood estimation in M*plus*. M*plus* Web Notes, 6 (version 1).

Muthén, L. K., & Muthén, B. O. (1998–2007). *M*plus *user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Ping, R. A., Jr. (1996). Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin, 119,* 166–175.

Raykov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling* (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Saris, W. E., Batista-Foguet, J. M., & Coenders, G. (2007). Selection of indicators for the interaction term in structural equation models with interaction. *Quality & Quantity, 41,* 55–72.

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

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Schumacker, R. E., & Marcoulides, G. A. (Eds.). (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sun, R., & Willson, V. L. (2009). Evaluating intercept-slope interactions in latent growth modeling. *Structural Equation Modeling, 16,* 226–244.

Wall, M. M., & Amemiya, Y. (2003). A method of moment technique for fitting interaction effects in structural equation models. *The British Journal of Mathematical and Statistical Psychology, 56,* 47–63.

Yuan, K.-H., & Bentler, P. M. (1998). Normal theory based test statistics in structural equation modeling. *British Journal of Mathematical and Statistical Psychology, 51,* 289–309.

APPENDIX

Let the ξ's be normally distributed with means zero. Then the second-order mixed moment (covariance) is $\phi_{ij} = E[\xi_i \xi_j]$, and the fourth- and sixth-order mixed moments are defined as $\phi_{ijkl} = E[\xi_i \xi_j \xi_k \xi_l]$ and $\phi_{ijklmn} = E[\xi_i \xi_j \xi_k \xi_l \xi_m \xi_n]$. In view of normality, the third- and fifth-order mixed moments are zero. Then it holds

$$\phi_{ijkl} = \phi_{ij}\phi_{kl} + \phi_{ik}\phi_{jl} + \phi_{il}\phi_{jk}$$

$$\phi_{ijklmn} = \phi_{ij}\phi_{klmn} + \phi_{ik}\phi_{jlmn} + \phi_{il}\phi_{jkmn} + \phi_{im}\phi_{jkln} + \phi_{in}\phi_{jklm}$$

These equations are used to compute the elements of the vectors with second- and third-order moments, $\boldsymbol{\varphi}_2$ and $\boldsymbol{\varphi}_3$. For instance, $\mathrm{cov}(\xi_1\xi_1, \xi_2\xi_2) = \phi_{1122} - \phi_{11}\phi_{22} = \phi_{11}\phi_{22} + 2\phi_{12}^2 - \phi_{11}\phi_{22} = 2\phi_{12}^2$. Similarly, $\mathrm{cov}(\xi_1\xi_1\xi_1, \xi_2\xi_2\xi_2) = \phi_{111222} - \phi_{111}\phi_{222}$.