Moosbrugger, H., Schermelleh-Engel, K., Kelava; A. & Klein, A. G. (in press). Testing Multiple Nonlinear Effects in Structural Equation Modeling: A Comparison of Alternative Estimation Approaches. Invited Chapter in T. Teo & M. S. Khine (Eds.), *Structural Equation Modelling in Educational Research: Concepts and Applications*. Rotterdam, NL: Sense Publishers.

HELFRIED MOOSBRUGGER, KARIN SCHERMELLEH-ENGEL,
AUGUSTIN KELAVA, AND ANDREAS G. KLEIN

6. TESTING MULTIPLE NONLINEAR EFFECTS IN STRUCTURAL EQUATION MODELING: A COMPARISON OF ALTERNATIVE ESTIMATION APPROACHES

INTRODUCTION

The estimation of nonlinear relations between variables is an important concern in different areas of the social and behavioral sciences. Several theories do not only incorporate linear but also nonlinear relations between variables. The most-often investigated nonlinear effects are interaction and quadratic effects.

An *interaction effect* implies that a relationship between a predictor and a criterion is weakened or strengthened by a second predictor variable (also called "moderator variable"; Aiken & West, 1991). In social psychology, for example, an interaction effect is hypothesized in an extension of the theory of planned behavior (Ajzen, 1991). This theory suggests that a given behavior is dependent on the individual's intention to perform a specific behavior and an individual's perceived ease or difficulty of performing this behavior (perceived behavioral control). In an extension of this theory, Elliott, Armitage, and Baughan (2003) could demonstrate in a study on compliance with speed-limits that prior behavior of exceeding speed limits while driving in built-up areas moderated the relationship between perceived behavioral control and subsequent behavior: increasing frequency of prior non-compliance with speed limits was associated with a decrease in the relationship between perceived behavioral control and driver's subsequent reported non-compliance with speed limits.

A *quadratic effect* implies that predictor variables interact with themselves. In health psychology, for example, a quadratic effect is hypothesized in research dealing with adolescents' reputations of peer status and health behaviors. Wang, Houshyar, and Prinstein (2006) investigated adolescent boys' weight-related health behaviors and cognitions expecting a curvilinear association between perceived body size and reputation-based popularity. The results showed the expected inverted U-shaped curve: Lower levels of popularity were associated with self-reported body shapes at each extreme of the silhouette scale (thin and heavy silhouettes), whereas higher levels of popularity were associated with self-reported muscular silhouettes. These findings confirm boys' body ideals toward body shapes that are neither thin nor heavy but muscular.

Most studies investigate either interaction or quadratic effects. But it is also sometimes of interest to combine both types of nonlinear effects in a more complex "multiple" nonlinear model. We will cite two examples in order to explain such cases in empirical research:

In educational psychology, for example, theory suggests a negative interaction between parent's education on child's educational expectations (Ganzach, 1997): When the level of education of one parent is high, the educational expectations of the child will also be high, even if the level of education of the other parent is quite low. However, analyses also revealed two quadratic effects, a positively accelerated relationship between mother's education and child's educational expectations as well as a positively accelerated relationship between father's education and child's educational expectations.

In studying the relationship between teachers' expectations and students' perceived competence in physical education classes, Trouilloud, Sarrazin, Bressoux, and Bois (2006) hypothesized a quadratic effect, that is a negatively accelerated relationship between teachers' early expectations and students' later perceived competence. However, analyses also revealed an interaction effect: Teachers' early expectations had a stronger effect on students' later perceived competence when the classroom motivational climate was low in autonomy support.

In this chapter, we will investigate methods for the simultaneous analysis of multiple nonlinear relations, i.e., latent interaction and latent quadratic effects. We begin by briefly explaining nonlinear regression using observed (manifest) variables and some problems of introducing nonlinear terms in the regressions equation. We then move to nonlinear structural equation modeling and present four alternative methods that attempt to estimate the nonlinear effects. Finally, the performance of theses methods is compared by analyzing artificial data sets of a Monte Carlo study.

NONLINEAR REGRESSION WITH MANIFEST VARIABLES

In a multiple regression analysis, the variables are usually assumed to be linearly related, that is, the criterion variable Y is a linear function of two predictor variables, e.g., X and Z. However, as has been shown in the empirical examples cited above, it may be theoretically plausible that – in addition to the linear effects of X and Z on Y – there are also nonlinear effects if the relation between X and Y is moderated by Z or by the predictor variable X itself.

In order to analyze the nonlinear effects together with the linear effects in the regression equation, new terms must be created by forming products of the predictor variables. For the analysis of an interaction effect, the product of X and Z, the interaction term XZ, is formed, which is then included in the regression equation as a third variable. For the analysis of quadratic effects, the products of X with itself and of Z with itself, the quadratic terms X^2 and Z^2, are formed which are then entered as a fourth and a fifth variable into the equation. As it is often meaningful that researchers should investigate interaction and quadratic effects simultaneously, nonlinear regression models can contain several nonlinear terms, e.g., one or more interaction terms and one or more quadratic terms, depending on the number of independent or predictor variables in the equation. In the case of two predictors (see Equation 1), a model with multiple nonlinear effects includes a criterion variable *Y*, two predictor variables *X* and *Z*, an interaction term *XZ*, two quadratic terms X^2 and Z^2, and a disturbance term ε:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + \beta_4 X^2 + \beta_5 Z^2 + \varepsilon \qquad (1)$$

In Equation (1), β_0 is the intercept, β_1 and β_2 are the linear effects, β_3 is the interaction effect, and β_4 and β_5 are the quadratic effects.

Generally, hypotheses regarding interaction and quadratic effects between continuous variables have been analyzed by means of nonlinear multiple regression analysis, although it is well-known that nonlinear regression is plagued by several methodological problems (e.g., Aiken & West, 1991; Cohen, Cohen, West, & Aiken, 2006; Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007; Moosbrugger, Schermelleh-Engel, & Klein, 1997).

METHODOLOGICAL PROBLEMS OF NONLINEAR REGRESSION

There are two main methodological problems associated with nonlinear regression models, namely measurement error and multicollinearity.

Measurement Error

Multiple regression models implicitly assume that all observed variables are measured without error, although in most cases observed variable have a considerable amount of measurement error and are therefore not perfectly reliable. The consequence of this lack of reliability is that the true effects (parameter values) may be underestimated. Ignoring measurement error can therefore lead to biased estimates of the effects, a problem that will even be aggravated when nonlinear terms are included in the multiple regression equation. The reliability of a nonlinear term (interaction term, quadratic term) is affected by measurement error to an even greater extent than the reliability of a linear term. The estimated regression weight associated with this term greatly underestimates the population coefficient.

The reliability of the interaction term is usually lower than the reliabilities of the predictor variables used to form the interaction term. However, the reliability of the interaction term does not only depend on the reliability of the predictor variables, but also on the correlation between the predictors (“multicollinearity”). In case of two predictor variables *X* and *Z*, the reliability of the interaction term is as follows (cf. Busemeyer & Jones, 1983; Dimitruk et al., 2007):

$$Rel(XZ) = \frac{Rel(X)Rel(Z) + [Corr(X,Z)]^2}{1 + [Corr(X,Z)]^2} \qquad (2)$$

The consequences of Equation (2) are shown in Figure 1: When *X* and *Z* are uncorrelated ($Corr(X, Z) = 0$), the reliability of the interaction term is just the product of the reliabilities of the two predictors. If, for example, *X* and *Z* have reliabilities of *Rel*(*X*) = *Rel*(*Z*) = .60, then *Rel*(*XZ*) = .36. Even with rather reliable measures (e.g. *Rel*(*X*) = *Rel*(*Z*) = .80), the reliability of the interaction term is only .64. With increasing correlation, the reliability of *XZ* also increases. For less reliable predictor variables (*Rel* = .60), the reliability of *XZ* increases to .40 for *Corr*(*X*,*Z*) = .25 and to .49 for *Corr*(*X*,*Z*) = .50. For reliable predictor variables (*Rel* = .80), the reliability of *XZ* increases to .66 for *Corr*(*X*,*Z*) = .25 and to .71 for *Corr*(*X*,*Z*) = .50. It is obvious that increasing multicollinearity enhances the reliability of the interaction term, but the interaction term cannot reach the reliability of the predictor variables even with higher correlated predictor variables.

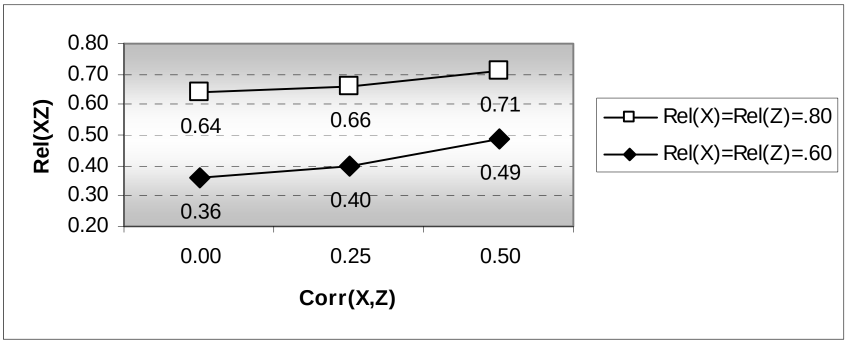

Figure 1. Relation between the correlation of the predictors, Corr(X,Z), and the reliability of the interaction term, Rel(XZ), for Rel(X) = Rel(Z) = .60, .80, and Corr(X,Z) = .00, .25, .50.

In contrast to the reliability of the interaction term *XZ*, the reliability of the quadratic term X^2 (or Z^2, respectively) is calculated differently (see Equation 3). Although it seems reasonable to assume that the reliability of the quadratic term is formed analogously to the reliability of the interaction term, this is not correct (Dimitruk et al., 2007). This is quite obvious when *Z* in Equation (2) is substituted by *X*. Including the correlation of *X* with itself would be incorrect due to the fact that not only the true scores but also the error scores are perfectly correlated. As the reliability is defined as the ratio of the true score variance to the total variance, only the correlation of the true scores should be included in the numerator of the equation. Therefore the reliability of the quadratic term X^2 is just the squared reliability of the variable *X* (see Equation 3).

$$Rel(X^2) = [Rel(X)]^2 \qquad (3)$$

Thus, as is shown in Figure 2, the reliability of the quadratic term X^2, the square of *Rel*(*X*), is usually smaller than the reliability of the predictor variable *X*. Both variables *X* and X^2 will only reach a reliability of 1.0 when *X* is measured without error.

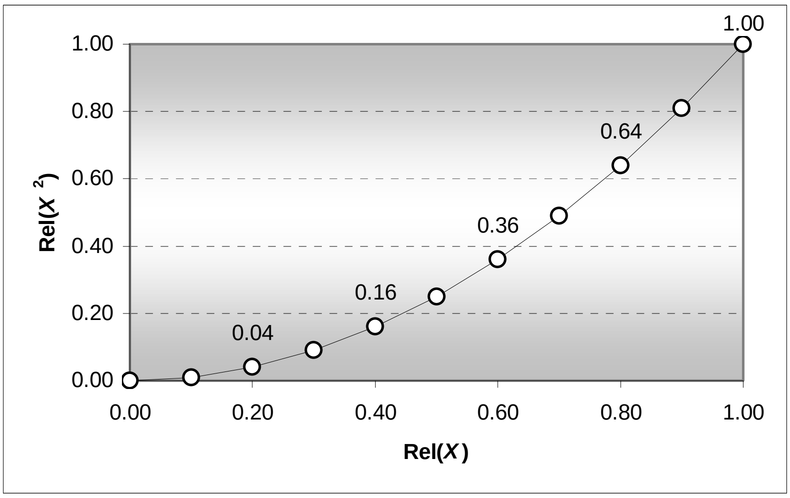

Figure 2. Relation between the reliability of the predictor variable X and the reliability of its quadratic term X^2.

Therefore it is recommended that nonlinear multiple regression analysis is only appropriate for manifest variables with very high reliabilities. In all other cases, nonlinear structural equation analysis (see below), which accounts for measurement error, should be used.

MULTICOLLINEARITY

The second problem that has gained attention in the literature of the past years is the relationship between interaction and quadratic terms when multicollinearity between the predictor variables is present (e.g., Busemeyer & Jones, 1983; Ganzach, 1997; Kelava, Moosbrugger, Dimitruk, & Schermelleh-Engel, 2008; Lubinski & Humphreys, 1990; MacCallum & Marr, 1995). When predictor variables are uncorrelated, the value of a regression weight remains unchanged regardless of all other predictor variables included in the regression equation. But when predictor variables are correlated, however, the value of a regression weight depends on which other variables are included in the model. Thus, when predictors are correlated, a regression coefficient does not simply reflect an inherent effect of the predictor variable on the criterion variable but rather a partial effect. On these grounds, estimated regression coefficients may vary widely from one data set to another (cf. Dimitruk et al., 2007).

When predictor variables are highly correlated, it is often difficult to distinguish between variance explained by *X* and variance explained by *Z* due to a high amount of shared variance. This problem is even aggravated when nonlinear terms are added in the regression equation because the nonlinear terms XZ, X^2, and Z^2 are then also correlated. While multicollinearity of predictors enhances the reliability of an interaction term (see above), it also causes severe estimation and interpretation problems because of the high correlations between all variables in the nonlinear regression model.

There are two sources of multicollinearity between predictor variables and their nonlinear terms (cf. Aiken & West, 1991; Marquardt, 1980): The first is nonessential multicollinearity that exists only due to scaling and that disappears when the predictors are centered before forming the product terms. The second source is essential multicollinearity, correlations that exist because of nonnormality or any nonsymmetry in the distribution of the predictor variables. Essential multicollinearity cannot be eliminated.

An example for nonessential multicollinearity can be seen in the correlation matrix in Table 1. In this matrix, the correlation between X and Z is .50. As a consequence, very high correlations exist between the predictor variables and their nonlinear terms with coefficients up to 1.0 between X and X^2 as well as Z and Z^2. In addition to this, the correlations between the nonlinear terms are also high, but only part of these correlations is due to nonessential multicollinearity.

Table 1. Nonessential multicollinearity in a correlation matrix of uncentered raw score variables X, Z (Corr(X,Z) = .50), and their nonlinear terms XZ, X^2, and Z^2.

	X	Z	XZ	X^2	Z^2
X	1.00				
Z	.50	1.00			
XZ	.87	.87	1.00		
X^2	1.00	.50	.87	1.00	
Z^2	.50	1.00	.87	.50	1.00

Centering the predictor variables X and Z by subtracting the mean from each predictor variable eliminates nonessential multicollinearity between the predictor variables and their nonlinear terms. Using the centered variables X_c and Z_c, the new interaction term X_cZ_c and quadratic terms X_c^2 and Z_c^2 are formed. In contrast to the correlations based on raw scores (cp. Table 1), the correlations between the centered variables X_c and Z_c and their nonlinear terms X_cZ_c, X_c^2, and Z_c^2 drop to zero (Table 2).

By centering the predictor variables, *nonessential multicollinearity* between the *nonlinear terms* could also be removed. The correlations between the interaction term X_cZ_c and the quadratic terms X_c^2 and Z_c^2 are smaller than the correlations based on raw scores, but they are still high. As is shown in Table 2, the correlations between the interaction term and the quadratic terms drop from .86 to .63, while the correlation between both quadratic terms is now .25 compared to the coefficient of .50 in Table 1.

Table 2. Remaining multicollinearity in a correlation matrix of centered variables X_c, Z_c and nonlinear terms X_cZ_c, X_c^2, and Z_c^2.

	X_c	Z_c	X_cZ_c	X_c^2	Z_c^2
X_c	1.00				
Z_c	.50	1.00			
X_cZ_c	.00	.00	1.00		
X_c^2	.00	.00	.63	1.00	
Z_c^2	.00	.00	.63	.25	1.00

With increasing correlation of the predictor variables the multicollinearity problem is exacerbated, insofar as the interaction term and the quadratic terms would share more and more variance. As a consequence it would be increasingly difficult to detect these effects separately (MacCallum & Mar, 1995; Ganzach, 1997). Another consequence of shared variance is that detected nonlinear effects might be an artifact if the analyzed model had been misspecified. For example, if the true model is a quadratic model and the analyzed model is an interaction model, this could result in a significant but spurious interaction effect (Klein, Schermelleh-Engel, Moosbrugger, & Kelava, 2009).

The problem of multicollinearity is not only present in nonlinear regression but also in nonlinear structural equation modeling. Multicollinearity leads in general to parameter estimates with higher standard errors, so that the power of detecting true effects is lowered. In order to eliminate *nonessential multicollinearity* it is recommended to always center the (manifest or latent) predictor variables.

NONLINEAR STRUCTURAL EQUATION MODELING

Over the last few years, nonlinear structural equation modeling (SEM) has received much attention and has become increasingly popular in the context of applied behavioral and social science research (for an overview, see Schumacker & Marcoulides, 1998). Nonlinear SEM provides many advantages compared to analyses based on manifest variables. It is, however, more complicated to conduct and is hindered by methodological problems that are different from those in multiple regression analysis (Dimitruk et al., 2007).

Several estimation methods have been developed during the last years. All methods aim at providing unbiased and efficient parameter estimates for the nonlinear effects. A nonlinear SEM with three nonlinear terms, one interaction term and two quadratic terms, is depicted in Figure 3.

In the following, we will use the LISREL notation (cf. Jöreskog & Yang, 1996). The nonlinear structural equation model (see Figure 3) includes two latent exogenous variables ξ_1 and ξ_2, a latent interaction term $\xi_1\xi_2$, two latent quadratic terms ξ_1^2 and ξ_2^2, a latent endogenous variable η and a disturbance term ζ. Linear parameters are denoted by γ_{11}, γ_{12}, and - using Klein and Moosbrugger's (2000) notation - ω_{12} is the interaction effect of $\xi_1\xi_2$, and ω_{11} and ω_{22} are the quadratic effects of ξ_1^2 and ξ_2^2, respectively. The structural equation of the nonlinear model with an intercept term α is given in Equation (4). Each exogenous variable is measured by three observed indicators (X_1, X_2, X_3, and X_4, X_5, X_6, respectively, see Equation 5), and the endogenous variable η is measured by three observed indicators (Y_1, Y_2, Y_3, see Equation 6). The measurement model of ξ_1 and ξ_2 includes the factor loadings λ_{21}, λ_{31}, and λ_{42}, λ_{52} (X_1 and X_4 are the scaling variables with factor loadings $\lambda_{11} = 1$ and $\lambda_{42} = 1$) and the error variables $\delta_1, \ldots, \delta_6$. The measurement model of η_1 includes the factor loadings λ_{21}^y, λ_{31}^y (Y_1 is the scaling variable with factor loading $\lambda_{11}^y = 1$) and the error variables ε_1, ε_2, and ε_3.

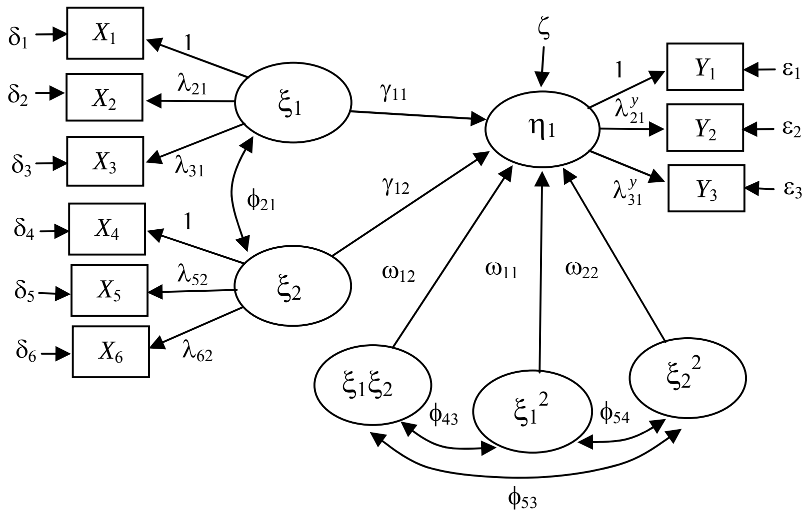

Figure 3. Nonlinear structural equation model with one latent criterion η, two latent predictors ξ_1 and ξ_2,a latent interaction term $\xi_1\xi_2$, and two latent quadratic terms ξ_1^2 and ξ_2^2 measured by nine indicator variables ($X_{1},\ldots,X_6$; $Y_1,\ldots,Y_3$).

In addition to the parameters that are estimated by all approaches for the analysis nonlinear SEM, several parameters resulting from the nonlinearity of the model have to be estimated that differ between the approaches.

While approaches especially developed for the analysis of nonlinear SEM only need the specification of the nonlinear terms $\xi_1\xi_2$, ξ_1^2, and ξ_2^2, and the estimation of the nonlinear effects ω_{12}, ω_{11}, and ω_{22}, product indicator approaches in the LISREL tradition have to additionally form interaction and quadratic terms from the manifest linear variables as indicators of the latent nonlinear terms, to specify the latent mean structure, and they need to estimate factor loadings and error variances of the nonlinear measurement model.

Structural equation:

$$\eta = \alpha + \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \omega_{12}\xi_1\xi_2 + \omega_{11}\xi_1{}^2 + \omega_{22}\xi_2{}^2 + \zeta \quad (4)$$

Measurement model of ξ-variables

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{pmatrix} \quad (5)$$

Measurement model of η-variable

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ \lambda_{21}^y \\ \lambda_{31}^y \end{pmatrix} \eta_1 + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix} \quad (6)$$

METHODOLOGICAL PROBLEMS OF NONLINEAR SEM

Nonnormality

The most severe problem of nonlinear SEM is the multivariate nonnormality of the nonlinear terms (cf. Dimitruk et al., 2007; Klein & Moosbrugger, 2000; Moosbrugger et al., 1997). Even if all indicators of the latent predictor variables ξ_1 and ξ_2 (and the latent predictors themselves) are normally distributed, the distributions of the latent nonlinear terms $\xi_1\xi_2$, $\xi_1{}^2$, and $\xi_2{}^2$ are *not* normal[1]. Furthermore, the latent criterion variable η will also be nonnormally distributed. The degree of nonnormality of the distribution of η depends on the nonnormality of $\xi_1\xi_2$, $\xi_1{}^2$ and $\xi_2{}^2$ as well as on the size of the nonlinear effects ω_{12}, ω_{11}, and ω_{22}.

For the analysis of nonlinear SEM, this fact has two possible consequences: First, adequate estimation methods should take the multivariate nonnormality explicitly into account (e.g., the heteroscedasticity utilizing approaches LMS and QML - see below). Second, if an estimation method is used under the assumption of normally distributed indicator variables that does not take the nonnormality into account, the robustness against violation of this assumption for the analysis of nonlinear models should be investigated thoroughly. For inferential statistics, the possible bias of the estimated standard errors can become critical in the method's performance. Simulation studies (see below) can decide whether this lowers the method's efficiency, which might become very low when sample size decreases.

Multicollinearity

While the reliability problem is solved by employing SEM, the multicollinearity problem remains and may even be exacerbated when using latent variable approaches (cf. Dimitruk et al. 2007; Kelava et al., 2008). In nonlinear SEM the correlation between latent predictor variables is generally higher than the correlation between manifest indicator variables due to attenuation by unreliability of the indicators. Thus, the multicollinearity increases. If, for example, the correlation between two manifest variables, each measured with a reliability of .90, is .50, the latent correlation is .56. But if the reliability of the indicator variables is as low as .63 and the manifest correlation is still .50, the latent correlation is .80.

A further exacerbation of the multicollinearity problem apparently occurs when all three nonlinear terms are included in the equation due to shared variance among them. With a correlation as high as .80 between the latent predictor variables, the correlations between the latent interaction term $\xi_1\xi_2$ and the quadratic terms ξ_1^2 and ξ_2^2 will be .88 (see Table 3).

Table 3. Correlation matrix of centered latent variables ξ_1 and ξ_2 and their nonlinear terms $\xi_1\xi_2$, ξ_1^2, and ξ_2^2, when the correlation of the manifest predictor variables is .50 (e.g., $Rel(X_1) = Rel(X_4) = .63$).

	ξ_1	ξ_2	$\xi_1\xi_2$	ξ_1^2	ξ_2^2
ξ_1	1.00				
ξ_2	.80	1.00			
$\xi_1\xi_2$	.00	.00	1.00		
ξ_1^2	.00	.00	.88	1.00	
ξ_2^2	.00	.00	.88	.64	1.00

All methods for the analysis of nonlinear SEM have to deal with the problem of multicollinearity. Consequences of multicollinearity may include estimation problems, biased estimates and standard errors that are not correctly estimated (cp. Kelava et al., 2008).

METHODS FOR THE ANALYSIS OF LATENT MULTIPLE NONLINEAR EFFECTS

In the last two decades there has been a huge amount of publications dealing with different estimation approaches for the analysis of interaction effects in structural equation models (for an overview see Marsh, Wen, & Hau, 2004; Moosbrugger et al., 1997; Schumacker & Marcoulides, 1998). In contrast to this kind of research, there are only few studies that investigate latent quadratic effects using nonlinear SEM (e.g., Klein & Muthén, 2007; Marsh, Wen, & Hau, 2006), and only few that investigate latent interaction and latent quadratic effects simultaneously in a nonlinear SEM (e.g., Dimitruk et al., 2007; Kelava et al, 2008; Ping, 1998; Wall & Amemiya, 2000).

In order to avoid misspecification, Ganzach (1997) has demanded that all nonlinear regression terms should be analyzed simultaneously, a demand that is also valid for nonlinear SEM (cf. Klein et al., 2009). Using simulation studies Ganzach (1997) demonstrated that it is often very useful to include quadratic terms in regression models when interactions are estimated, because without quadratic terms a nonsignificant effect may be observed in the presence of a true interaction effect or an estimated interaction effect may be found to have a positive sign although the true effect has a negative sign. When quadratic effects are being analyzed, it is also important to include the interaction effect in the nonlinear model because quadratic effects may either turn out to be nonsignificant in the presence of true quadratic effects or they may show a positive accelerated relation while a true negative accelerated relation exists (see also Kelava et al., 2008).

There are four approaches that we will consider for the analysis of multiple latent nonlinear effects: Latent Moderated Structural Equations (LMS; Klein & Moosbrugger, 2000), Quasi-Maximum Likelihood (QML; Klein & Muthén, 2007), the constrained approach (Jöreskog & Yang, 1996) extended by quadratic terms (Kelava et al., 2008), and the unconstrained approach (Marsh, Wen, & Hau, 2004) extended to multiple nonlinear terms. First, we will introduce these four approaches. Second, we will conduct simulation studies to investigate the performance of theses approaches for the simultaneous analysis of three latent nonlinear effects, i.e., an interaction effect and two quadratic effects.

Latent Moderated Structural Equations

Klein and Moosbrugger (2000) developed the Latent Moderated Structural Equations (LMS) method for the estimation of multiple latent interaction and quadratic effects that takes the nonnormality caused by the latent nonlinear terms explicitly into account. In LMS, no manifest nonlinear indicators are needed for the estimation of the nonlinear effects. Instead, since the latent criterion variable is nonnormally distributed when nonlinear effects are in the data, the distribution of the latent criterion is utilized to estimate the nonlinear effects. As the latent predictor and the latent moderator variable (the second predictor variable) are assumed to be bivariate normally distributed, it follows that for each value of the moderator variable the conditional distribution of the predictor variable and the conditional distribution of the latent criterion variable is normal. Therefore the nonnormal density function of the joint indicator vector (X, Y) is approximated by a finite mixture distribution of multivariate normally distributed components. In order to estimate the model parameters, model implied mean vectors and covariance matrices of the mixture components are utilized.

LMS only assumes that the latent predictor variables ξ_1 and ξ_2 and all error variables (δ_1 - δ_6, ε_1 - ε_3, ζ_1) are normally distributed. The ML estimates are computed with the Expectation-Maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977).

As simulation studies have shown, LMS provides efficient parameter estimators and unbiased standard errors (Klein, 2000; Klein & Moosbrugger, 2000; Schermelleh-Engel, Klein, & Moosbrugger, 1998).

The LMS method has been adopted in the M*plus* 5.0 program (Muthén & Muthén, 1998-2007). The syntax of the input file is quite easy to set up as only the equations for the measurement models (Equations 5 and 6) and the structural model (Equation 4) are needed and additionally, the nonlinear terms have to be specified (see Appendix A1). In M*plus*, latent nonlinear terms are demanded by using the XWITH command (Equation 7); the latent variables ξ_1 and ξ_2 are denoted by f1 and f2:

```
f1f2 | f1 XWITH f2;
f1f1 | f1 XWITH f1;        (7)
f2f2 | f2 XWITH f2;
```

Quasi-Maximum Likelihood

Klein and Muthén (2007) developed the Quasi-Maximum Likelihood (QML) method, a less demanding approach for the estimation of multiple interaction and quadratic effects. Just like LMS, QML takes the nonnormality caused by the latent nonlinear terms explicitly into account. Again, no manifest nonlinear indicators are needed. By applying the quasi-likelihood principle, the nonnormal density function of the joint indicator vector (X, Y) is approximated by a product of a normally distributed and a conditionally normal density function. In contrast to LMS, QML only assumes that the conditional distribution of the latent criterion variable given the X-variables can be approximated by normal distributions.

As QML is only an approximate ML estimator, a somewhat lower efficiency of the estimator should be expected; nevertheless, this approach should be more robust against the violation of distributional assumptions. Simulation studies have shown that QML provides parameter estimators that are almost as efficient as the LMS estimators when predictor variables are normally distributed (Dimitruk et al., 2007; Klein & Muthén, 2007). If the distributional assumptions are violated, QML provided a more efficient estimation of the interaction parameter ω_{12} than LMS does in simulation studies suggesting that QML is more robust against the violation of the distributional assumption (Klein & Muthén, 2007).

QML is currently not implemented in a commercial statistics software, but it is a stand-alone program. The syntax of the QML program (see Appendix A2) is matrix-based. One needs to specify all parameter matrices including factor loadings, linear effects, covariances of the latent predictor variables, and covariances of the error terms.

In addition to these matrices, a parameter matrix Ω for the nonlinear effects is needed (Klein, 2007; Klein & Moosbrugger, 2000). The matrix Ω is assumed to be an upper triangular matrix; the quadratic effects are listed on the diagonal, the interaction effects in the upper triangular, and zeros in the lower triangular. For the investigation of a nonlinear SEM with one interaction effect and two quadratic effects, Ω has to be specified as follows:

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} \\ 0 & \omega_{22} \end{pmatrix} \tag{8}$$

Extended Constrained Approach

The constrained approach (Jöreskog & Yang, 1996) and the unconstrained approach (Marsh et al., 2004) of the LISREL program were developed in the tradition of Kenny and Judd (1984). Just as in nonlinear regression, both approaches require that latent nonlinear terms are formed and that a measurement model for the latent nonlinear terms $\xi_1\xi_2$, ξ_1^2 and ξ_2^2 is specified. Using the maximum likelihood method for parameter estimation it is assumed that all latent variables including the nonlinear terms are normally distributed[2].

Kenny and Judd (1984) were the first to describe how a latent nonlinear model with an interaction or a quadratic effect can be analyzed. Their approach involves the formation of products of the indicators of linear predictors which serve as indicators of the latent interaction or quadratic terms.

Kenny and Judd (1984) suggested using all possible manifest product variables as indicators of the latent interaction term $\xi_1\xi_2$. Using the constrained approach for the example depicted in Figure 3 this would require the forming of 9 product indicators for the latent interaction term, X_1X_4, X_1X_5, X_1X_6, X_2X_4, X_2X_5, X_2X_6, X_3X_4, X_3X_5, and X_3X_6.

If X_2 is an indicator of ξ_1 and X_5 an indicator of ξ_2 (with $X_2 = \lambda_{21}\xi_1 + \delta_2$ and $X_5 = \lambda_{52}\xi_2 + \delta_5$), then the indicator X_2X_5 of the interaction term $\xi_1\xi_2$ would be specified as follows:

$$\begin{aligned} X_2X_5 &= (\lambda_{21}\xi_1+\delta_2)(\lambda_{52}\xi_2+\delta_5) \\ &= \lambda_{21}\lambda_{52}\xi_1\xi_2+(\lambda_{21}\xi_1\delta_5+\lambda_{52}\xi_2\delta_2+\delta_2\delta_5) \\ &= \lambda_{83}\xi_1\xi_2+\delta_8 \end{aligned} \tag{9}$$

As the parameters in this measurement equation cannot be estimated directly, several constraints are needed. The variance decomposition of the interaction indicator X_2X_5 required for model specification is given by

$$Var(X_2X_5)=\lambda_{83}^2 Var(\xi_1\xi_2)+Var(\delta_8) \tag{10}$$

and includes the following constraints:

$$\begin{aligned} \lambda_{83}^2 &= \lambda_{21}^2\lambda_{52}^2 \\ Var(\xi_1\xi_2) &= Var(\xi_1)Var(\xi_2)+Cov(\xi_1,\xi_2)^2 \\ Var(\delta_8) &= \lambda_{21}^2 Var(\xi_1)Var(\delta_5)+\lambda_{52}^2 Var(\xi_2)Var(\delta_2)+Var(\delta_2)Var(\delta_5) \end{aligned} \tag{11}$$

As a measurement model for all latent nonlinear terms that includes all possible manifest products would require many measurement equations (e.g., nine equations when two linear latent predictors are each measured by three indicators) with overlapping information, Marsh et al. (2004) suggested to use the matched-pair strategy: All indicators of the linear terms ξ_1 and ξ_2 should be used in the formation of the indicators of each latent nonlinear term, but each of the multiple indicators should be used only once for each nonlinear term. For our example depicted in Figure 3 this means that only three indicators for each nonlinear factor are needed: X_1X_4, X_2X_5, X_3X_6 for the latent interaction term $\xi_1\xi_2$, X_1^2, X_2^2, X_3^2 for the latent quadratic term ξ_1^2, and X_4^2, X_5^2, X_6^2 for the latent quadratic term ξ_2^2.

A further revision of the constrained approach was done by Algina and Moulder (2001). Their revision implied that the indicators of the linear terms have to be centered, a proposition that had also been suggested by Jaccard and Wan

(1995). The measurement model for the predictor variables of the nonlinear model is therefore as follows:

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_1X_4 \\ X_2X_5 \\ X_3X_6 \\ X_1^2 \\ X_2^2 \\ X_3^2 \\ X_4^2 \\ X_5^2 \\ X_6^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \lambda_{21} & 0 & 0 & 0 & 0 \\ \lambda_{31} & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & \lambda_{52} & 0 & 0 & 0 \\ 0 & \lambda_{62} & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \lambda_{21}\lambda_{52} & 0 & 0 \\ 0 & 0 & \lambda_{31}\lambda_{62} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \lambda_{21}^2 & 0 \\ 0 & 0 & 0 & \lambda_{31}^2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \lambda_{52}^2 \\ 0 & 0 & 0 & 0 & \lambda_{62}^2 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_1\xi_2 \\ \xi_1^2 \\ \xi_2^2 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \\ \delta_{13} \\ \delta_{14} \\ \delta_{15} \end{pmatrix} \tag{12}$$

Under the supposition that ξ_1, ξ_2, ζ, and all error terms δ and ε are multivariate normal, uncorrelated and have zero means (except that ξ_1 and ξ_2 are allowed to be correlated), Jöreskog and Yang (1996) proposed a model with a latent mean structure.

Using the constrained approach extended for multiple nonlinear terms (Kelava et al., 2008) the mean vector and covariance matrix of ξ_1, ξ_2, $\xi_1\xi_2$, ξ_1^2, and ξ_2^2 are, respectively:

$$\boldsymbol{\kappa} = \begin{pmatrix} 0 \\ 0 \\ \phi_{21} \\ \phi_{11} \\ \phi_{22} \end{pmatrix}, \quad \Phi = \begin{pmatrix} \phi_{11} & & & & \\ \phi_{21} & \phi_{22} & & & \\ 0 & 0 & \phi_{11}\phi_{22} + \phi_{21}^2 & & \\ 0 & 0 & 2\phi_{11}\phi_{21} & 2\phi_{11}^2 & \\ 0 & 0 & 2\phi_{22}\phi_{21} & 2\phi_{21}^2 & 2\phi_{22}^2 \end{pmatrix} \tag{13}$$

Furthermore, even if ξ_1 and ξ_2 are centered so as to have zero means, $\kappa_3 = E(\xi_1\xi_2) = Cov(\xi_1, \xi_2) = \phi_{21}$ will typically not be zero. Extending Jöreskog and Yang's (1996) latent interaction model to a model with quadratic terms (Kelava et al., 2008), $\kappa_4 = E(\xi_1^2) = Var(\xi_1) = \phi_{11}$ and $\kappa_5 = E(\xi_2^2) = Var(\xi_2) = \phi_{22}$ will also not be zero. Hence, a mean structure is always necessary and should always be specified (cf. Marsh et al., 2004).

In contrast to QML, the nonlinear effects of the extended constrained approach are included in the parameter matrix of the linear effects so that $\omega_{12} = \gamma_{13}$, $\omega_{11} = \gamma_{14}$, and $\omega_{22} = \gamma_{15}$ in LISREL notation.

The syntax of the extended constrained approach (Kelava et al., 2008) using the LISREL program is quite complicated (see Appendix A3), extremely error-prone and therefore limited to few indicator variables. Nevertheless, simulation studies examining the performance of this approach have shown that parameter estimates of a model with only one nonlinear latent interaction effect are unbiased (cf. Kelava et al., 2008; Moulder & Algina, 2002; Schermelleh-Engel et al., 1998; Yang-Jonsson, 1998). However, simulation studies have also shown that using the maximum likelihood estimation method of the LISREL program with nonnormal product terms may lead to serious underestimation of standard errors and biased χ^2 values even for models that include only one latent interaction term (Marsh et al., 2004; Jöreskog & Yang, 1996; Kelava et al., 2008; Schermelleh-Engel et al., 1998).

Extended Unconstrained Approach

Marsh et al. (2004) revised the constrained approach in that they did not impose any complicated nonlinear constraints to define relations between product indicators and the latent nonlinear terms and denoted this new approach as "unconstrained approach".

In the unconstrained approach extended to quadratic nonlinear terms (Kelava et al., 2008) factor loadings as well as error variances and covariances (see Appendix A4) are estimated directly without using any constraints. Additionally, parameters based on assumptions of normality are also not constrained, so that this approach does not need any constraints on the latent variances, too. The covariances between the linear and nonlinear latent variables may be freely estimated if it is assumed that the variables are nonnormally distributed. The only constraints used in this approach are the constraints on the latent means (Equation 15).

$$\boldsymbol{\kappa} = \begin{pmatrix} 0 \\ 0 \\ \phi_{21} \\ \phi_{11} \\ \phi_{22} \end{pmatrix}, \quad \Phi = \begin{pmatrix} \phi_{11} & & & & \\ \phi_{21} & \phi_{22} & & & \\ 0 & 0 & \phi_{33} & & \\ 0 & 0 & \phi_{43} & \phi_{44} & \\ 0 & 0 & \phi_{53} & \phi_{54} & \phi_{55} \end{pmatrix} \tag{15}$$

An advantage of the extended unconstrained approach for applied researchers clearly is that the syntax is much easier to set up than for the extended constrained approach as no complicated, nonlinear constraints are required.

Simulation studies of latent interaction models showed that the unconstrained approach is comparable to the constrained approach in terms of unbiasedness of parameter estimates, but the parameter estimates may be biased for nonnormal data. Both approaches performed less well than QML with regard to standard error estimates and power (Klein & Muthén, 2007; Marsh et al., 2004). However, even when normality assumptions were met but sample size was small the parameter estimates of the unconstrained approach were more biased and standard errors were larger compared to the constrained approach and to QML.

MONTE CARLO STUDY

In nonlinear structural equation modeling, the analysis of latent quadratic effects is as important than the analysis of latent interaction effects (cf. Klein et al., 2009). However, while the latter has received considerable attention in the methodological literature, relatively few simulation studies have been conducted that compare the performance of different methods for the analysis of latent quadratic effects and almost none for the analysis of multiple nonlinear effects in a polynomial structural equation model.

In two simulation studies, we therefore investigated the performance of LMS, QML, the extended constrained approach and the extended unconstrained approach for the analysis of two nonlinear models that only differ in the size of the multicollinearity: the latent covariance was varied across the two levels $Cov(\xi_2, \xi_1) = .00$ and $Cov(\xi_2, \xi_1) = .50$. In the model with correlated predictor variables the latent nonlinear terms are of course also correlated as outlined above ($Cov(\xi_2\xi_1,\xi_1^2) = Cov(\xi_2\xi_1,\xi_2^2) = .63$, $Cov(\xi_1^2,\xi_2^2) = .25$). All other parameters were held constant in both models.

The following parameter values were selected for the population model: $\alpha = -.20$ for uncorrelated latent predictors ($\alpha = -.30$ for correlated latent predictors), $\gamma_{11} = \gamma_{12} = .316$, $\omega_{12} = .20$, $\omega_{11} = \omega_{22} = .10$, $\phi_{11} = \phi_{22} = 1.00$. According to Figure 3, the predictor variables ξ_1, ξ_2, and the criterion variable η_1 were each measured by three indicator variables with a reliability of .80. The given selection of nonlinear effects results in a model in which 4% (5%) of the variance of η_1 is explained by the interaction effect and 2% by each quadratic effect, while the linear effects each explain 10% of the variance of η_1. Hence, the study tested the performance of the four approaches with reasonably large nonlinear effects compared to those found in empirical studies. The data for the latent predictor variables were generated according to the normal distribution. Sample size was $N = 400$ with 500 data sets in each simulation study. We restrict the report of the simulation results to the three structural parameters of interest, ω_{12}, ω_{11}, and ω_{22}, and the parameters of the latent variances ϕ_{11}, ϕ_{22} and the latent covariance ϕ_{21}.
Data were generated using the PRELIS 2.7 program (Jöreskog & Sörbom, 1999). Analyses with the extended constrained approach and the extended unconstrained approach were conducted using LISREL 8.72 (Jöreskog & Sörbom, 1996). For the LMS analyses (Klein & Moosbrugger, 2000) M*plus* 5.0 (Muthén & Muthén, 1998-2007) was used, for the QML (Klein & Muthén, 2007) analyses QuasiML 3.10 (Klein, 2007).

Monte Carlo Statistics

In the two simulation studies, the means (*M*) of all 500 parameter estimates and their standard deviations (*SD*) as well as the means of the 500 estimated standard errors (*SE*) for each parameter were calculated. In addition to this, we also calculated the bias (underestimation or overestimation) of each parameter estimate and the bias of the estimated standard error. Furthermore, we examined the power of the approaches to detect the true effects.

The means of all parameter estimates (*M*) over all 500 data sets are estimates of the population values that we had defined before. The standard deviation of each parameter estimate is then an estimate for the true standard error (*SD*) of an approach's parameter estimator. The mean of the 500 estimated standard errors (*SE*) for each parameter was also calculated and compared to the estimated true *SD*. If the means of the *SE*s of an approach are generally smaller than the true standard errors (*SD*), the approach shows a progressive behavior so that effects are more often significant than they actually are in the population (increased Type I error rate).

The bias of a parameter estimator was calculated as follows:

$$Bias(\hat{\pi}) = \frac{\bar{\hat{\pi}} - \pi}{\pi} \quad (16)$$

where $\hat{\pi}$ denotes the parameter estimate, $\bar{\hat{\pi}}$ the mean of the 500 parameter estimates, and π the population parameter value.

The bias for each estimated standard error was calculated as follows:

$$Bias(SE) = \frac{\overline{SE} - SD}{SD} \quad (17)$$

where $\overline{SE}$ denotes the mean of the 500 estimated standard errors, and *SD* is the estimated true standard error of an approach's estimator calculated over 500 parameter estimates.

Power was examined based on the ratio of the parameter estimate to its standard error and the number of t-values were counted that exceeded the critical value of 1.96. As this measure may not be trustworthy when standard errors are underestimated by a method, we therefore calculated the corrected power by dividing the parameter estimates by their true standard error *SD* ("true power").

Results of the Simulation Studies

Study 1

The results of the first simulation study (see Table 4) with the condition of uncorrelated predictor variables (ϕ_{21} = 0.00) show that all four methods perform about equally well with regard to the estimated means (*M*) of the nonlinear effects ω_{12}, ω_{12}, and ω_{22}, their true standard errors (*SD*), the mean of the estimated standard errors (*SE*), and the true power to detect the nonlinear effects. The bias of the parameter estimates is smaller than 2% for all approaches and the bias of the standard errors is about the same for all approaches with the smallest bias for the extended unconstrained approach. The power to detect the interaction effect is high for all approaches (> 90%), while the power of the quadratic effects is somewhat lower (69% - 82%), especially the power of LMS and QML to detect ω_{22}. In case that the latent linear predictors are uncorrelated the approaches perform equally well in detecting nonlinear effects.

Table 4. Performance of LMS, QML, the constrained and the unconstrained approach for estimating the nonlinear effects ω_{12}, ω_{11}, ω_{22} when predictors are uncorrelated (ϕ_{21} = 0.00).

ϕ_{21} = 0.00							
Parameter	True Value	*M*	Bias	*SD*	*SE*	*SE*-Bias	True Power
LMS							
ω_{12}	0.20	0.199	-0.50%	0.052	0.050	-4.00%	96.60%
ω_{11}	0.10	0.102	1.96%	0.036	0.036	0.00%	82.00%
ω_{22}	0.10	0.100	0.00%	0.039	0.036	-8.33%	69.20%
QML							
ω_{12}	0.20	0.198	-1.01%	0.053	0.051	-3.92%	93.60%
ω_{11}	0.10	0.102	1.96%	0.037	0.036	-2.78%	78.60%
ω_{22}	0.10	0.100	0.00%	0.039	0.036	-8.33%	68.80%
Extended Constrained Approach							
ω_{12}	0.20	0.198	-1.00%	0.051	0.048	-5.88%	97.80%
ω_{11}	0.10	0.099	-1.00%	0.036	0.034	-5.56%	78.60%
ω_{22}	0.10	0.099	-1.00%	0.037	0.034	-8.11%	76.00%
Extended Unconstrained Approach							
ω_{12}	0.20	0.201	0.50%	0.053	0.052	-1.87%	97.00%
ω_{11}	0.10	0.101	1.00%	0.037	0.036	-2.70%	78.40%
ω_{22}	0.10	0.101	1.00%	0.039	0.037	-5.13%	74.60%

We were also interested in investigating whether the covariances between the latent predictors would be estimated correctly. For no correlation between the latent linear predictors (no multicollinearity) large differences between the approaches are detected (see Table 5).

While LMS and QML perform very well and estimate the latent variances and the latent covariance with a high precision, both LISREL approaches produce standard errors (*SE*) that underestimate the true standard errors *SD* (see column *SE*-Bias). The estimated standard errors (*SE*) underestimate the *SD* by about 30 % (extended unconstrained approach) or even by more then 60% (extended constrained approach). While the bias of the parameter estimates is quite small for LMS, QML, and the constrained approach, the unconstrained approach overestimates the latent variances. Nevertheless, the latent covariance of the latent predictor variables (ϕ_{21} = .00) is estimated on average correctly by all approaches.

Table 5. Performance of LMS, QML, the constrained and the unconstrained approach for estimating the latent variances and covariance ϕ_{11}, ϕ_{21}, ϕ_{22} when predictors are uncorrelated (ϕ_{21} = .00).

ϕ_{21} = .00						
Parameter	True Value	*M*	Bias	*SD*	*SE*	*SE*-Bias
LMS						
ϕ_{11}	1.00	0.996	-0.40%	0.089	0.089	0.00%
ϕ_{21}	0.00	0.002	n.d.	0.052	0.054	3.85%
ϕ_{22}	1.00	0.994	-0.60%	0.085	0.088	3.53%
QML						
ϕ_{11}	1.00	1.001	0.10%	0.087	0.089	2.30%
ϕ_{21}	0.00	-0.001	n.d.	0.056	0.054	-3.57%
ϕ_{22}	1.00	1.008	0.79%	0.086	0.089	3.49%
Extended Constrained Approach						
ϕ_{11}	1.00	1.039	3.75%	0.098	0.037	-62.24%
ϕ_{21}	0.00	0.000	0.00%	0.061	0.020	-67.21%
ϕ_{22}	1.00	1.032	3.10%	0.095	0.037	-61.05%
Extended Unconstrained Approach						
ϕ_{11}	1.00	1.152	15.20%	0.091	0.064	-29.67%

ϕ_{21}	0.00	0.000	0.00%	0.054	0.041	-31.71%
ϕ_{22}	1.00	1.146	14.60%	0.090	0.063	-30.00%

n.d. = not defined

Study 2

With multicollinearity in the data (ϕ_{21} = .50), there are now some important differences in the performance of the methods. As is obvious (see Table 6), the extended unconstrained approach now shows a somewhat lower performance in detecting nonlinear effects than the other approaches. A major finding is that the estimated standard errors (*SE*) of the extended constrained approach underestimate the true standard errors (*SD*) by about 10% when the predictor variables are correlated while the other three approaches do not show substantial underestimation. Power is now lower for all approaches indicating that larger sample sizes may be needed for nonlinear analyses with correlated predictors that include all three nonlinear effects.

The mean estimates of the nonlinear effects are unbiased for all approaches with the exception of the unconstrained approach which overestimates the latent quadratic effects by 4% - 5%. While the true standard errors (*SD*) of the parameter estimates of LMS, QML, and the extended constrained approach are almost identical, the parameter estimates of the unconstrained approach are less precise which implies higher true standard errors. The standard errors estimated by the extended constrained and the extended unconstrained approach substantially underestimate the true standard errors (*SD*) of the estimators of all nonlinear parameters, so that hypothesis testing using estimated standard errors for confidence intervals can be fallible and spurious effects may be produced.

Table 6. Performance of LMS, QML, the constrained and the unconstrained approach for estimating the nonlinear effects ω_{12}, ω_{11}, ω_{22} when predictors are correlated (ϕ_{21} = .50).

ϕ_{21} = .50							
Parameter	True Value	*M*	Bias	*SD*	*SE*	*SE*-Bias	True Power
LMS							
ω_{12}	0.20	0.197	-1.52%	0.070	0.070	0.00%	78.60%
ω_{11}	0.10	0.100	0.00%	0.046	0.044	-4.35%	60.60%
ω_{22}	0.10	0.102	1.96%	0.045	0.044	-2.22%	60.40%
QML							
ω_{12}	0.20	0.199	-0.50%	0.071	0.070	-1.41%	78.16%
ω_{11}	0.10	0.100	0.00%	0.047	0.044	-6.38%	58.98%
ω_{22}	0.10	0.100	0.00%	0.045	0.044	-2.22%	59.39%
Extended Constrained Approach							
ω_{12}	0.20	0.196	-2.00%	0.073	0.065	-10.96%	74.40%
ω_{11}	0.10	0.103	3.00%	0.045	0.041	-8.89%	63.60%
ω_{22}	0.10	0.102	2.00%	0.046	0.041	-10.87%	60.00%
Extended Unconstrained Approach							
ω_{12}	0.20	0.197	-1.50%	0.078	0.074	-5.13%	71.60%
ω_{11}	0.10	0.105	5.00%	0.048	0.047	-2.08%	59.00%
ω_{22}	0.10	0.104	4.00%	0.049	0.046	-6.12%	56.80%

The estimates of the latent variances and the covariance are comparable to the estimates of the analysis with uncorrelated predictors. Bias of the parameter estimates is again quite small for LMS, QML, and the constrained approach, while the unconstrained approach overestimates the latent variances - a result that seems to be independent of the amount of multicollinearity in the data. Nevertheless, the latent covariance of the latent predictor variables (ϕ_{21} = .50) is again estimated correctly by all approaches. Overestimation of the latent variances implies that the latent correlations are underestimated: The estimated latent correlation of the constrained approach is .48, the estimated latent correlation of the unconstrained approach is .43.

Table 7. Performance of LMS, QML, the constrained and the unconstrained approach for estimating the latent variances and covariance when predictors are correlated (ϕ_{21} = .50).

ϕ_{21} = .50						
Parameter	True Value	*M*	Bias	*SD*	*SE*	*SE*-Bias
LMS						
ϕ_{11}	1.00	0.997	-0.30%	0.088	0.089	1.14%
ϕ_{21}	0.50	0.496	-0.81%	0.058	0.062	6.90%
ϕ_{22}	1.00	0.996	-0.40%	0.092	0.088	-4.35%
QML						
ϕ_{11}	1.00	0.998	-0.20%	0.088	0.085	-3.41%
ϕ_{21}	0.50	0.496	-0.81%	0.059	0.056	-5.08%
ϕ_{22}	1.00	0.996	-0.40%	0.092	0.084	-8.70%
Extended Constrained Approach						
ϕ_{11}	1.00	1.029	2.90%	0.097	0.036	-62.89%
ϕ_{21}	0.50	0.494	-1.20%	0.068	0.022	-67.65%
ϕ_{22}	1.00	1.032	3.20%	0.093	0.036	-61.29%
Extended Unconstrained Approach						
ϕ_{11}	1.00	1.149	14.90%	0.093	0.063	-32.26%
ϕ_{21}	0.50	0.504	0.80%	0.063	0.045	-28.57%
ϕ_{22}	1.00	1.150	15.00%	0.091	0.063	-30.77%

While there is no bias of the parameter estimators of LMS, QML, and the constrained approach, the extended unconstrained approach overestimates the latent variances by more than 14%. Both LISREL approaches again produce estimates of the parameters less precisely than LMS and QML because the true standard errors (*SD*) are larger than those of LMS and QML, and the estimated standard errors (*SE*) are again severely underestimated by both LISREL approaches similar to the results for uncorrelated latent predictor variables. While the standard error estimates (*SD*) of the extended unconstrained approach were underestimated by the *SE* by about 30%, the extended constrained approach underestimated the standard errors by more than 60%.

DISCUSSION

Our simulation studies revealed the important results that all four estimation approaches are able to estimate the nonlinear effects without bias, but that there exist substantial differences between the methods with regard to the standard errors and the latent variances and covariances. As could be expected for the extended constrained approach, standard errors of the nonlinear effects are underestimated when the predictor variables are correlated (cf. Marsh et al., 2004; Jöreskog & Yang, 1996; Kelava et al., 2008; Schermelleh-Engel et al., 1998). In contrast to LMS and QML, the standard errors of the latent variances and covariances of both LISREL approaches are underestimated to a large extent, namely by more than 30% - 60%, so that the efficiency for these estimators is quite low.

The efficiency of a parameter estimator corresponds to its true standard error (*SD*): a low true standard error indicates high efficiency. Thus, the standard error (*SD*) of an estimator is a measure of its precision, and the lower the standard error (the higher the efficiency), the higher is the power or detection capability for true effects (i.e., the lower is Type II error) of the estimator. The LMS and QML estimators of all parameters are clearly most efficient under the two investigated conditions. The estimators of the LISREL approaches are about as efficient as the LMS and QML estimators when the predictors are uncorrelated, but the true standard deviations of the latent variances and covariance of the linear predictors are always underestimated to a large extent.

Multicollinearity plays an important role in the analyses of nonlinear SEM. As Klein et al. (2009) demonstrated using simulation studies, latent interaction effects may be overestimated when the quadratic effects are not estimated simultaneously together with the interaction effect. If multicollinearity is present in the data, i.e., when linear predictors as well as nonlinear terms are correlated, estimates of the nonlinear effects are still unbiased for all approaches, but the power to detect these effects is definitely lower for all approaches than in the uncorrelated condition so that Type II errors are larger. The problem of low true power is more apparent for both LISREL approaches. Concerning the extended constrained approach, true standard errors are grossly underestimated: While power may be too low to detect nonlinear effects reliably, the confidence interval around the parameter estimates looks very small caused by underestimated standard errors so that effects would be erroneously judged too often to be significant.

There are of course some limitations to this study that should be noted:

– First, an important limitation is that only normally distributed variables were generated in the simulation study, while empirical data often have the problem that they are nonnormally distributed. Nonnormality would lead to essential multicollinearity between the nonlinear terms that cannot be reduced by centering predictor variables. Therefore it can be expected that differences between the approaches will be larger when nonnormal variables are used.

- Second, all generated variables were interval-scaled while in empirical applications often ordinal-scaled variables are used. It would be of interest to investigate whether LMS and QML would behave differently because QML as a more robust method may be able to deal with this problem better than LMS.
- Third, in our simulation study we used a modified and extended version of the Jöreskog and Yang (1996) approach. While Jöreskog and Yang proposed to use all possible products of indicators, we followed Marsh et al.'s (2004) suggestion and used only three indicators for each latent construct. Therefore we cannot tell whether the inclusion of all possible product indicators would have resulted in a different performance of this approach.

Future research needs to be carried out to further investigate the practical applicability of LMS and QML for more complex models and to further investigate possible differences between both approaches. For example, in empirical research it may be of interest to investigate models with more than three nonlinear effects, e.g., three interaction effects and three quadratic effects. Until now, we do not know anything about the performance of the approaches to deal with such complex models.

In summary, the results of the Monte Carlo study indicate the following:

- Applied researchers should use nonlinear SEM when nonlinear relations are hypothesized.
- LMS implemented in M*plus* as well as QML both estimate the nonlinear effects reliable and most efficient. The approaches are equally well suited for the analysis of latent interaction and quadratic effects.
- Both LISREL approaches performed quite well when the linear predictors were uncorrelated. In order to reduce nonessential multicollinearity, the means of the latent predictors - but not the means of the latent nonlinear terms – should always be fixed to zero. While the syntax of the extended constrained approach using the LISREL program is quite complicated and error-prone, the syntax of the extended unconstrained approach is much easier to set up. But the advantage of the unconstrained approach that the complicated nonlinear constraints are eliminated is offset by the disadvantage of overestimated latent variances and covariances so that standardized effects may be underestimated.
- When there is multicollinearity in the data, LMS and QML should be preferred to the LISREL approaches.

NOTES

1 Please note that the nonlinear terms should never be normalized, standardized, or centered (cf. Aiken & West, 1991).

2 Interaction and quadratic terms are always nonnormally distributed: Even if linear predictor variables are normally distributed, the distributions of nonlinear terms are always highly kurtotic, and additionally the distribution of a quadratic term is censored below at zero (cf. Dimitruk et al., 2007).

REFERENCES

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

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211.

Busemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the casual variables are measured with error. *Psychological Bulletin*, *93*, 549–562.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

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

Elliott, M. A., Armitage, C. J., & Baughan, C. J. (2003). Drivers' compliance with speed limits: An application of the theory of planned behavior. *Journal of Applied Psychology*, *88*, 964–972.

Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods*, *3*, 235–247.

Jaccard, J., & Wan, C. K. (1995). Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin*, *117*, 348–357.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Lincolnwood, IL: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1999). *PRELIS 2: User's reference guide*. Lincolnwood, IL: Scientific Software International.

Jöreskog, K. G., & Yang, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In G. Marcoulides & R. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 57–87). Mahwah, NJ: Lawrence Erlbaum Associates.

Kelava, A., Moosbrugger, H., Dimitruk, P., & Schermelleh-Engel, K. (2008). Multicollinearity and missing constraints: A comparison of three approaches for the analysis of latent nonlinear effects. *Methodology*, *4*, 51–66.

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 3.10 - Quick reference manual*. Unpublished Paper, The University of Western Ontario.

Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., & Kelava, A. (2009). Spurious interaction effects. In T. Teo & M. S. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications*. Amsterdam, NL: Sense Publishers.

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.

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious "moderator effects": Illustrated substantively with the hypothesized ("synergistic") relation between spatial and mathematical ability. *Psychological Bulletin*, *107*, 385–393.

MacCallum, R. C., & Marr, C. M. (1995). Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin*, *118*, 405–421.

Marquardt, D. W. (1980). you should standardize the predictor variables in your regression models. *Journal of the American Statistical Association*, *75*, 87–91.

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.

Marsh, H. W., Wen, Z. L., & Hau, K. T. (2006). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R. O. Mueller (Eds.), *A second course in structural equation modeling*. Greenwich, CT: Information Age Publishing.

Moosbrugger, H., Schermelleh-Engel, K., & Klein, A. (1997). Methodological problems of estimating latent interaction effects. *Methods of Psychological Research Online*, *2*, 95–111.

Moulder, B. C., & Algina, J. (2002). Comparison of methods for estimating and testing latent variable interactions. *Structural Equation Modeling*, *9*, 1–19.

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

Ping, R. A., Jr. (1998). EQS and LISREL examples using survey data. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (pp. 63–100). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Schermelleh-Engel, K., Klein, A., & Moosbrugger, H. (1998). Estimating nonlinear effects using a latent moderated structural equations approach. In R. E. Schumacker & G. A. Marcoulides (Eds.), *Interaction and nonlinear effects in structural equation modeling* (pp. 203–238). Mahwah, NJ: Lawrence Erlbaum Associates.

Trouilloud, D., Sarrazin, P., Bressoux, P., & Bois, J. (2006). Relation between teachers' early expectations and students' later perceived competence in physical education classes: Autonomy-supportive climate as a moderator. *Journal of Educational Psychology*, *98*, 75–86.

Wall, M. M., & Amemiya, Y. (2000). Estimation for polynomial structural equation models. *Journal of the American Statistical Association*, *95*, 929–940.

Wang, S. S., Houshyar, S., & Prinstein, M. J. (2006). Adolescent girls' and boys' weight-related health behaviors and cognitions: Associations with reputation- and preference-based peer status. *Health Psychology*, *25*, 658–663.

Yang-Jonsson, F. (1997). *Nonlinear structural equation models. Simulation studies of the Kenny-Judd model*. Stockholm: Gotab.

Appendix A: Syntax Files

A1 - LMS

```
Title: Full Nonlinear Model

DATA: FILE = LMS.dat;

VARIABLE: NAMES ARE Y1 - Y3 X1-X6;

ANALYSIS: TYPE = RANDOM;
 ALGORITH = INTEGRATION;
 ITERATIONS = 300;

MODEL: f1 BY X1 X2 X3;
 f2 BY X4 X5 X6;
 f3 BY Y1 Y2 Y3;

 f1f2 | f1 XWITH f2;
 f1f1 | f1 XWITH f1;
 f2f2 | f2 XWITH f2;
 f3 on f1 f2 f1f2 f1f1 f2f2;

OUTPUT: tech1; tech8;
```

A2 - QML (Klein & Muthén, 2007)

```
% Full Nonlinear Model

%Raw Data
$RD
QML.dat

%Parameter Output
$PO
QML.par

%Standard Error Output
$SO
QML.std

%Sample Size
$SA
400

%Number of Data Sets
$DS
1

%Number of x-variables
$q
6

%Number of y-variables
$P
3

%Number of Ksis
$N
2

%Gamma
$GA
* *

%Omega
$Om
* *
0 *
```

```
%Restriction Omega (for Model Difference Test)
$RO
0 0
0 0

%Restriction Gamma (for Model Difference Test)
$RG
* *

%Theta Delta
$TD
* 0 0 0 0 0
0 * 0 0 0 0
0 0 * 0 0 0
0 0 0 * 0 0
0 0 0 0 * 0
0 0 0 0 0 *

%Psi
$PS
*

%Theta Epsilon
$TE
* 0 0
0 * 0
0 0 *

%Lambda x
$LX
1 0
* 0
* 0
0 1
0 *
0 *


%Lambda y
$LY
1
*
*

END
```

A3 – Constrained Approach Extended for Multiple Nonlinear Effects

```
!Full Nonlinear Model - Extended Constrained Approach
DA NI=18 NO=400
RA FI=Constrained.dat
LA
Y1 Y2 Y3 X1 X2 X3 X4 X5 X6 X1X4 X2X5 X3X6 X1X1 X2X2 X3X3 X4X4
X5X5 X6X6

MO NY=3 NX=15 NE=1 NK=5 LY=FU,FR LX=FU,FR GA=FU,FR PH=SY
TE=DI,FR TD=SY PS=DI,FR AL=FR KA=FI

LK
KSI1 KSI2 KSI1KSI2 KSI1KSI1 KSI2KSI2

LE
ETA

PA LY
0
1
1
VA 1 LY(1,1)

PA LX
0 0 0 0 0
1 0 0 0 0
1 0 0 0 0
0 0 0 0 0
0 1 0 0 0
0 1 0 0 0
0 0 0 0 0
0 0 1 0 0
0 0 1 0 0
0 0 0 0 0
0 0 0 1 0
0 0 0 1 0
0 0 0 0 0
0 0 0 0 1
0 0 0 0 1
VA 1 LX(1,1) LX(4,2) LX(7,3) LX(10,4) LX(13,5)

!constrained nonlinear factor loadings
CO LX(8,3)=LX(2,1)*LX(5,2)     ! 2. interaction indicator X2X5
CO LX(9,3)=LX(3,1)*LX(6,2)     ! 3. interaction indicator X3X6
```

```
CO LX(11,4)=LX(2,1)*LX(2,1)     ! 2. Quad1 indicator X2X2
CO LX(12,4)=LX(3,1)*LX(3,1)     ! 3. Quad1 indicator X3X3
CO LX(14,5)=LX(5,2)*LX(5,2)     ! 2. Quad2 indicator X5X5
CO LX(15,5)=LX(6,2)*LX(6,2)     ! 3. Quad2 indicator X6X6

PA PH
1
1 1
0 0 1
0 0 1 1
0 0 1 1 1

!constrained nonlinear variances and covariances
CO PH(3,3)=PH(1,1)*PH(2,2)+PH(2,1)**2 ! variance KSI1*KSI2
CO PH(4,3)=2*PH(1,1)*PH(2,1)        ! covariance KSI1*KSI2,KSI1*KSI1
CO PH(5,3)=2*PH(2,2)*PH(2,1)        ! covariance KSI1*KSI2,KSI2*KSI2
CO PH(4,4)=2*PH(1,1)**2             ! variance KSI1*KSI1
CO PH(5,4)=2*PH(2,1)**2             ! covariance KSI1*KSI1,KSI2*KSI2
CO PH(5,5)=2*PH(2,2)**2             ! variance KSI2*KSI2

PA KA
0 0 1 1 1

!constrained latent means
CO KA(3)=PH(2,1)            !mean of Interaction term
CO KA(4)=PH(1,1)            !mean of Quad1 term
CO KA(5)=PH(2,2)            !mean of Quad2 term

! constrained error variances
CO   TD(7,7)=PH(1,1)*TD(4,4)+PH(2,2)*TD(1,1)+TD(1,1)*TD(4,4)   !error   var.
X1X4
CO TD(8,8)=LX(2,1)**2*PH(1,1)*TD(5,5)+LX(5,2)**2*PH(2,2)*TD(2,2)
+TD(2,2)*TD(5,5)                                          !error var. X2X5
CO TD(9,9)=LX(3,1)**2*PH(1,1)*TD(6,6)+LX(6,2)**2*PH(2,2)*TD(3,3)
+ TD(3,3) *TD(6,6)                                        !error var. X3X6
CO TD(10,10)=4*PH(1,1)*TD(1,1)+2*TD(1,1)**2               !error var. X1X1
CO TD(11,11)=4*LX(2,1)**2*PH(1,1)*TD(2,2)+2*TD(2,2)**2 !error var. X2X2
CO TD(12,12)=4*LX(3,1)**2*PH(1,1)*TD(3,3)+2*TD(3,3)**2 !error var. X3X3
CO TD(13,13)=4*PH(2,2)*TD(4,4)+2*TD(4,4)**2               !error var. X4X4
CO TD(14,14)=4*LX(5,2)**2*PH(2,2)*TD(5,5)+2*TD(5,5)**2 !error var. X5X5
CO TD(15,15)=4*LX(6,2)**2*PH(2,2)*TD(6,6)+2*TD(6,6)**2 !error var. X6X6

! constrained error covariances
CO TD(10,7)= 2*PH(2,1)*TD(1,1)                       !error cov. X1X4,X1X1
CO TD(11,8)= 2*LX(2,1)*LX(5,2)*PH(2,1)*TD(2,2)   !error cov. X2X5,X2X2
```

```
CO TD(12,9)= 2*LX(3,1)*LX(6,2)*PH(2,1)*TD(3,3)   !error cov. X3X6,X3X3
CO TD(13,7)= 2*PH(2,1)*TD(4,4)                   !error cov. X1X4,X4X4
CO TD(14,8)= 2*LX(2,1)*LX(5,2)*PH(2,1)*TD(5,5)   !error cov. X2X5,X5X5
CO TD(15,9)= 2*LX(3,1)*LX(6,2)*PH(2,1)*TD(6,6)   !error cov. X3X6,X6X6

OU ME=ML IT=500 AD=OFF ND=3 RS
```

A4 – Unconstrained Approach Extended for Multiple Nonlinear Effects

```
!Full Nonlinear Model - Extended Unconstrained Approach
DA NI=18 NO=400
RA FI=Unconstrained.dat
LA
Y1 Y2 Y3 X1 X2 X3 X4 X5 X6 X1X4 X2X5 X3X6 X1X1 X2X2 X3X3 X4X4
X5X5 X6X6
MO NY=3 NE=1 NX=15 NK=5 LY=FU,FI LX=FU,FI PH=SY,FR TE=DI,FR
TD=SY,FI PS=FU,FR AL=FR KA=FI
LE
ETA
LK
KSI1 KSI2 KSI1KSI2 KSI1KSI1 KSI2KSI2
PA PH
1
1 1
0 0 1
0 0 1 1
0 0 1 1 1

VA 1 LY(1,1)
FR LY(2,1) LY(3,1)
VA 1 LX(1,1) LX(4,2) LX(7,3) LX(10,4) LX(13,5)
FR LX(2,1) LX(3,1) LX(5,2) LX(6,2) LX(8,3) LX(9,3)
FR LX(11,4) LX(12,4) LX(14,5) LX(15,5)
FR GA(1,1) GA(1,2) GA(1,3) GA(1,4) GA(1,5)

PA KA
0 0 1 1 1

!constrained latent means
CO KA(3)=PH(2,1) !mean of latent interaction term
CO KA(4)=PH(1,1) !mean of first latent quadratic term
CO KA(5)=PH(2,2) !mean of second latent quadratic term

! error variances
FR TD(1,1) TD(2,2) TD(3,3) TD(4,4) TD(5,5) TD(6,6)
FR TD(7,7)    !measurement error variance X1X4
FR TD(8,8)    !measurement error variance X2X5
FR TD(9,9)    !measurement error variance X3X6
FR TD(10,10) !measurement error variance X1X1
FR TD(11,11) !measurement error variance X2X2
FR TD(12,12) !measurement error variance X3X3
FR TD(13,13) !measurement error variance X4X4
```

```
FR TD(14,14) !measurement error variance X5X5
FR TD(15,15) !measurement error variance X6X6

! error covariances
FR TD(10,7)   ! measurement error covariance X1X4_X1X1
FR TD(11,8)   ! measurement error covariance X2X5_X2X2
FR TD(12,9)   ! measurement error covariance X3X6_X3X3
FR TD(13,7)   ! measurement error covariance X1X4_X4X4
FR TD(14,8)   ! measurement error covariance X2X5_X5X5
FR TD(15,9)   ! measurement error covariance X3X6_X6X6

PD
OU ME=ML IT=500 ND=3 RS
```