MULTIVARIATE BEHAVIORAL RESEARCH, *42*(4), 647–673

Quasi-Maximum Likelihood Estimation of Structural Equation Models With Multiple Interaction and Quadratic Effects

Andreas G. Klein
The University of Western Ontario

Bengt O. Muthén
University of California, Los Angeles

In this article, a nonlinear structural equation model is introduced and a quasi-maximum likelihood method for simultaneous estimation and testing of multiple nonlinear effects is developed. The focus of the new methodology lies on efficiency, robustness, and computational practicability. Monte-Carlo studies indicate that the method is highly efficient and that the likelihood ratio test of nonlinear effects is robust and outperforms alternative testing procedures. The new method is applied to empirical data of middle-aged men, where a latent interaction between physical fitness and flexibility in goal adjustment on complaint level is hypothesized. A model with 5 simultaneous nonlinear effects is analyzed, and the hypothesized interaction is quantified and tested positively against an additive model with quadratic and linear effects.

Over the last 2 decades, structural equation modeling (SEM) has become a common statistical tool for modeling relationships between variables that cannot be observed directly, but only with measurement error. The relationships between these unobservable, latent variables are formulated in structural equations, and they are measured with errors by indicator variables in a measurement model. By the development of software packages for covariance structure analysis such

Correspondence concerning this article should be sent to Andreas G. Klein, The University of Western Ontario, Department of Psychology, SSC, London, ON, N6A 5C2, Canada. E-mail: aklein25@uwo.ca. Bengt O. Muthén, University of California, Los Angeles, Graduate School of Education and Information Studies, 2023 Moore Hall, 405 Hilgard Avenue, Los Angeles, CA 90095-1521.

as Amos (Arbuckle, 1997), EQS (Bentler, 1995; Bentler & Wu, 1993), LISREL (Jöreskog & Sörbom, 1993, 1996), or Mplus (Muthén & Muthén, 2004), SEM has become available to a large community of researchers.

Although ordinary SEM incorporates linear relationships among latent variables, applied researchers sometimes wish to estimate an SEM with quadratic forms in the latent variables, for example, models with latent interaction or quadratic effects. Several researchers have called for estimation methods for nonlinear latent variable models, and numerous substantive theories in education and psychology call for analysis of nonlinear models (Ajzen, 1987; Ajzen & Fishbein, 1980; Ajzen & Madden, 1986; Cronbach, 1975; Cronbach & Snow, 1977; Fishbein & Ajzen, 1975; Karasek, 1979; Lusch & Brown, 1996; Snyder & Tanke, 1976). Also, a need for nonlinear extensions of ordinary SEM has been expressed from a methodological perspective (Aiken & West, 1991; Busemeyer & Jones, 1983; Cohen & Cohen, 1975; Jaccard, Turrisi, & Wan, 1990), and different ad-hoc estimation approaches have been developed. Models with nonlinear latent variable structures are special because products of normal variates are no longer normally distributed; the model lies outside the framework of covariance structure analysis or the general linear model; and product indicators for latent variable products often lead to a low communality, a large sampling fluctuation, multicollinearity problems, and nonnormal error terms (Jöreskog & Yang, 1996; Klein, 2000; Moosbrugger, Schermelleh-Engel, & Klein, 1997).

Therefore, methodological difficulties arise in estimating these models because the known methods for ordinary linear SEM either have to be modified or extended when using product indicators, or other approaches that do not need product indicators had to be newly developed. There have been different estimation approaches. Hayduk (1987) established the estimation of an elementary interaction model with one latent product term proposed by Kenny and Judd (1984). Using LISREL 7 (Jöreskog & Sörbom, 1989), they formed products of indicators for measuring the latent product term. Two-step LISREL approaches for this elementary model were proposed by Moosbrugger, Frank, and Schermelleh-Engel (1991) and Ping (1995, 1996a, 1996b, 1998), who implemented a stepwise LISREL procedure by estimating the measurement model in a first step and the parameters of the structural equation in a second step.

Other approaches aimed at estimating the nonlinear model within the framework of covariance structure analysis. The technique of forming products of indicators was improved by Jaccard and Wan (1995), Jöreskog and Yang (1996, 1997), and Yang Jonsson (1997), who used nonlinear parameter constraints for estimation of the elementary interaction model under LISREL 8 (Jöreskog & Sörbom, 1996). Simulation studies showed that the LISREL-ML (maximum likelihood) estimation procedure could be used for parameter estimation of the elementary interaction model (Yang Jonsson, 1997), but applicability seemed to be limited to elementary quadratic or cross-product models because of unstable sampling characteristics of the covariance matrices, which include covariances of products of indicators. Also, the distributional assumptions of LISREL-ML are violated for a nonlinear structural equation model, and standard errors and χ^2-statistics can be erroneous and require an adjustment for bias (Yang-Wallentin & Jöreskog, 2001). Moreover, simulation studies for the elementary interaction model indicated that the LISREL parameter estimators did not have optimal efficiency (Klein & Moosbrugger, 2000; Schermelleh-Engel, Klein, & Moosbrugger, 1998).

As an alternative to covariance structure analysis, a two-stage least squares (2SLS) estimation technique was developed by Bollen (1995, 1996) and Bollen and Paxton (1998) using instrumental variables for estimating an elementary quadratic or interaction model. But although no distributional assumptions were violated for this method, simulation studies showed that 2SLS estimators were substantially less efficient when compared to alternative estimation techniques (Klein & Moosbrugger, 2000; Schermelleh-Engel et al., 1998). Using Bayesian estimation techniques, Arminger and Muthén (1998) proposed a computationally intensive method and demonstrated it for elementary models with one latent product term. Blom and Christoffersson (2001) developed an estimation method based on the empirical characteristic function of the distribution of the indicator variables. But both approaches seemed to be limited to elementary models because of their computational burden.

More recently, the LISREL approach of Jöreskog and Yang, which involves a set of specific, nonlinear constraints on the model parameters and carries out estimation in LISREL-ML, has been modified and improved. A constrained approach has been developed and tested by Algina and Moulder (2001) and Marsh, Wen, and Hau (2004), who modified the constraints for the mean structure. Wall and Amemiya (2000, 2001, 2003) proposed a generalized appended product indicator (GAPI) approach where some of the constraints related to the covariance matrix of the latent predictor variables were removed from the constraint list. It is a method of moments approach to estimate latent interactions. In particular under conditions where the predictors were nonnormal, their partially constrained approach yielded improved parameter estimates. Technical details about the difference between the Quasi-ML approach used here and their method of moments technique are given in the section "Quasi-ML Estimation Procedure." This technique was further modified and an unconstrained approach was derived where the complicated nonlinear parameter constraints present in the earlier approaches could be removed (Marsh et al.).

With the latent moderated structural equations (LMS) method, Klein and Moosbrugger (2000) first introduced a maximum likelihood estimation technique for latent interaction models with multiple latent product terms. In the LMS method, the latent independent variables and the error variables were assumed to be normally distributed, and in LMS the nonnormality caused by the latent product terms was now explicitly taken into account. The LMS method has been adopted in Mplus 3.0 (Muthén & Muthén, 2004). In contrast to the LISREL type approaches, no product indicators were needed. Going beyond the product indicator approaches, LMS also allowed for likelihood ratio tests, which can test for the significance of one or several nonlinear effects simultaneously. Simulation studies for the elementary interaction model indicated that LMS provides efficient parameter estimators and a fairly reliable likelihood ratio test, and standard errors were unbiased (Klein, 2000; Klein & Moosbrugger, 2000; Schermelleh-Engel et al., 1998). But, although models with several latent product terms could be analyzed with the LMS method, the method could become computationally extremely intensive for models with three or more product terms involved in the structural equation. Also, LMS turned out to be not yet robust enough when its distributional assumptions were violated.

In this article, a Quasi-ML estimation method for structural equation models with quadratic forms is proposed. Quasi-ML has been developed to aim for an efficient, computationally feasible, and more robust estimation technique for nonlinear structural equation models with quadratic forms of predictor variables, which cannot be analyzed optimally under LMS or the product indicator approaches mentioned earlier.

The new approach was applied to an empirical study taken from psychology of aging on the relationship between coping style, subjective and objective fitness, and complaints about one's physical or psychological situation. A data set of 304 middle-aged men was examined, and the variables, except for objective fitness, were measured using inventories and questionnaires previously developed in the research literature. Two hypothesized interaction effects with the two fitness variables interacting with coping style on the complaint level as dependent variable were tested against an additive model structure with possible quadratic and linear effects. A model with five nonlinear effects was analyzed with Quasi-ML for this purpose. Using the likelihood ratio test under Quasi-ML, the interaction effect between subjective fitness and coping style was found to be significant beyond the quadratic and linear effects of these predictors.

For the Quasi-ML method proposed in this article, an appropriate transformation of the indicator variables is carried out, which reduces the number of nonnormally distributed components of the original indicator vector to one nonnormally distributed component of the transformed indicator vector. After this transformation, the model is treated as a variance function model (Carroll, Ruppert, & Stefanski, 1995), and mean and variance functions for the nonlinear model are calculated. Then, the quasi-likelihood estimation principle is applied and the density function of the indicator vector is approximated by a product of an unconditionally normal and a conditionally normal density function. A Quasi-ML estimator is established by maximizing the log-likelihood function based on the approximating density function. The distributional assumptions on which the Quasi-ML estimator are based are less rigid than the normality assumption for the predictor variables under LMS. This is because Quasi-ML only assumes approximate conditional normality for the latent criterion variable where in LMS the criterion variable is assumed to exactly follow the distribution of a polynomial function of normal variates. Therefore, Quasi-ML can be expected to be more robust than LMS against the violation of the normality assumptions. Simulation studies also presented in this article were carried out to investigate the properties and robustness of Quasi-ML. Quasi-ML outperformed the other approaches under a variety of conditions. The only limitation of Quasi-ML found so far from the simulation studies is that when the predictor variables are clearly skewed and the size of a nonlinear effect is very large, some bias could occur in the estimation of the structural parameters.

The contents of this article are as follows. In the section "Structural Equation Models With Quadratic Forms," we introduce a general structural equation model with a quadratic form of latent variables. In the section "Quasi-ML Estimation Procedure," the Quasi-ML estimation procedure is developed. In the section "Simulation Studies," the finite sample properties of the Quasi-ML estimators with respect to bias and efficiency are examined, and the Quasi-ML likelihood ratio test is evaluated. In the section "Empirical Example," the applicability of the proposed method to the analysis of empirical data sets is shown for the empirical example on psychology of aging.

STRUCTURAL EQUATION MODELS WITH QUADRATIC FORMS

In this section, a structural equation model with a general quadratic form of latent independent variables (predictor variables) is introduced. The elementary interaction model proposed by Kenny and Judd (1984) with one latent interaction effect and the model proposed by Klein and Moosbrugger (2000) with multiple interaction effects but no quadratic effects are special cases of the model specified here. The model we propose covers structural equations with a polynomial of degree two of predictor variables. We propose the following structural equation model with a quadratic form:

$$\eta_t = \alpha + \mathbf{\Gamma}\boldsymbol{\xi}_t + \boldsymbol{\xi}_t'\mathbf{\Omega}\boldsymbol{\xi}_t + \zeta_t, t = 1, \ldots, N, \tag{1}$$

where η_t is a latent dependent variable (criterion variable), α is an intercept term, $\boldsymbol{\xi}_t$ is a $(n \times 1)$ vector of latent predictor variables, $\mathbf{\Gamma}$ is a $(1 \times n)$ coefficient matrix, $\mathbf{\Omega}$ is a symmetric $(n \times n)$ coefficient matrix, and ζ_t is a disturbance variable. We call the parameters given by the matrices $\mathbf{\Gamma}$ and $\mathbf{\Omega}$ the structural parameters of the model. The quadratic form $\boldsymbol{\xi}_t'\mathbf{\Omega}\boldsymbol{\xi}_t$ of the structural equation (1) distinguishes the model from ordinary linear SEM. The vector $\boldsymbol{\xi}_t$ is assumed to be multivariate normally distributed with $E(\boldsymbol{\xi}_t) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\xi}_t, \boldsymbol{\xi}_t') = \boldsymbol{\Phi}$. The disturbance variable ζ_t is assumed to be normally distributed with $E(\zeta_t) = \mathbf{0}$, $\mathrm{Var}(\zeta_t) = \psi$, and $\mathrm{Cov}(\zeta_t, \boldsymbol{\xi}_t') = \mathbf{0}$. The latent variables of the structural equation are measured with error via measurement models:

$$\mathbf{x}_t = \boldsymbol{\Lambda}_x \xi_t + \delta_t, t = 1, \ldots, N, \tag{2}$$

$$\mathbf{y}_t = \boldsymbol{\Lambda}_y \eta_t + \varepsilon_t, t = 1, \ldots, N, \tag{3}$$

where $\mathbf{x}_t$ is a $(q \times 1)$ vector of observed indicators of $\boldsymbol{\xi}_t$, $\boldsymbol{\Lambda}_x$ is a $(q \times n)$ factor loading matrix, and $\boldsymbol{\delta}_t$ is a $(q \times 1)$ vector of measurement errors. Similarly, $\mathbf{y}_t$ is a $(p \times 1)$ vector of observed indicators of η_t, $\boldsymbol{\Lambda}_y$ is a $(p \times 1)$ factor loading matrix, and $\boldsymbol{\varepsilon}_t$ is a $(p \times 1)$ vector of measurement errors. The error vectors $\boldsymbol{\delta}_t$ and $\boldsymbol{\varepsilon}_t$ are assumed to be multivariate normally distributed with $E(\boldsymbol{\delta}_t) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, a diagonal covariance matrix $\mathrm{Cov}(\boldsymbol{\delta}_t, \boldsymbol{\delta}_t') = \Theta_\delta$, a diagonal covariance matrix $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_t') = \Theta_\varepsilon$, $\mathrm{Cov}(\boldsymbol{\delta}_t, \boldsymbol{\xi}_t') = \mathbf{0}$, $\mathrm{Cov}(\boldsymbol{\delta}_t, \boldsymbol{\varepsilon}_t') = \mathbf{0}$, and $\mathrm{Cov}(\boldsymbol{\varepsilon}_t, \boldsymbol{\xi}_t') = \mathbf{0}$, $\mathrm{Cov}(\zeta_t, \boldsymbol{\delta}_t') = \mathbf{0}$, and $\mathrm{Cov}(\zeta_t, \boldsymbol{\varepsilon}_t') = \mathbf{0}$. It is further assumed that the identifiability of the measurement model for the x-variables is guaranteed by certain restrictions on the parameters of the measurement model. For the identification of the measurement model for the y-variables, it is assumed that y_{1t} is a scaling indicator with loading $\lambda_{y11} = 1$. It can be assumed without loss of generality that the x-variables and the y-variables have zero mean, which implies that $\alpha = -\mathrm{tr}(\boldsymbol{\Omega}\boldsymbol{\Phi})$. This restriction does not limit the applicability of the model in practice because the indicator variable means can be regarded as nuisance parameters here. The nonlinear structural equation implies that the vector $\mathbf{y}_t$ is nonnormally distributed in general (Jöreskog & Yang, 1996; Klein & Moosbrugger, 2000).

The identification of the proposed nonlinear structural model is not different from an ordinary linear structural model: For each latent predictor variable, at least two indicator variables are needed for identification. Up to the present, an appropriate fit index, which is comparable to the chi-square fit measure for linear structural equation models, is not available. The reason for this lies in the fact that an unrestricted covariance structure is not a saturated model for a nonlinear SEM, and thus there is no straightforward computation of a fit measure. Future research is needed to address this important issue.

QUASI-ML ESTIMATION PROCEDURE

The major characteristic of a nonlinear structural equation model lies in the fact that the latent dependent variable η_t given by the structural equation (1) is nonnormally distributed in general, which also implies that the vector $\mathbf{y}_t$

of indicator variables is nonnormally distributed. The distribution type of η_t has been investigated for special cases in the literature (Gurland, 1955; Imhof, 1961; Press, 1966; Shah, 1963), but as Johnson and Kotz (1970) point out, these representations of the nonnormal density function are neither theoretically nor practically useful because of their complicated structure. The characteristic function of η_t or $\mathbf{y}_t$ can be expressed in closed form (Srivastava & Khatri, 1979), but it is not possible to derive the density function from the characteristic function by integrating out the transform variable because the integral cannot be solved in an analytically closed form.

The Quasi-ML estimation procedure developed in this article is based on an approximation of the nonnormal density function $f(x, y)$ of the indicator vector $(\mathbf{x}_t', \mathbf{y}_t')'$ by a nonnormal density $f^*(x, y)$, which is a product of a normal and a conditionally normal density. This approximation makes use of the concept of variance function models (Carroll et al., 1995, pp. 269–272), where the mean and variance function of a dependent variable conditional on the independent variables is specified. More precisely, for Quasi-ML the critical assumption about the structural part of the model is that the conditional distribution of the latent criterion variable given the x-variables can be approximated by a normal distribution. In Quasi-ML, the conditional mean and variance are derived under the assumption that the latent predictors and the error variables are normal, but this latter part of the model could be modified when a distribution different from the normal is assumed for the predictors. The distributional assumption for Quasi-ML is less rigid than it is for the LMS method, which assumes that the criterion variable is exactly distributed as a specific polynomial of normal variates. In the Quasi-ML method, the model parameters of the nonlinear structural equation model are simultaneously estimated. The maximization of the quasi-log-likelihood function derived from the approximating density $f^*(x, y)$ yields the Quasi-ML parameter estimates. The Quasi-ML principle has been applied to a specific heterogeneous latent growth curve model before (Klein & Muthén, 2006), but the development of the estimation procedure for the cross-sectional model given here is more complex and more general.

In the Quasi-ML method, a transformation of the nonnormally distributed indicator vector $(\mathbf{x}_t', \mathbf{y}_t')' \rightarrow (\mathbf{x}_t', y_{1t}, \mathbf{u}_t' = \mathbf{y}_t'\mathbf{R}')'$ is carried out such that only one component of the transformed indicator vector, namely, the scaling variable y_1 with loading one, is nonnormally distributed. Then, the conditional mean and variance of the nonnormally distributed component y_1 are derived. The technical details of these parts are described in the Appendix. Finally, the derived mean and variance function are used to develop the Quasi-ML estimation procedure. Also, standard errors for the parameter estimates and a quasi-likelihood ratio test statistic are computed.

The application of the variance function modeling concept is based on the idea that the conditional distribution of $(y_1|\mathbf{x} = x, \mathbf{u} = u)$ is approximated by a normal distribution. For this approximating normal distribution, the mean function $E[y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u]$ and the variance function $\mathrm{Var}(y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u)$ are used. The application of the quasi-likelihood principle suggests the following approximation $f^*(x, y)$ of the density function $f(x, y)$ of the indicator vector $(\mathbf{x}_t', \mathbf{y}_t')'$:

$$\begin{aligned} f(x, y) &= f_2(x, \mathbf{R}y) f_3(y_1|\mathbf{x}_t = x, \mathbf{u}_t = \mathbf{R}y) \\ &\approx f_2(x, \mathbf{R}y) f_3^*(y_1|\mathbf{x}_t = x, \mathbf{u}_t = \mathbf{R}y) \\ &=: f^*(x, y) \end{aligned} \tag{4}$$

where $f_2(x, u)$ is the normal density function of $(\mathbf{x}_t', \mathbf{u}_t')'$ and $f_3^*(y_1|\mathbf{x}_t = x, \mathbf{u}_t = \mathbf{R}y)$ is a univariate normal density with mean $E[y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = \mathbf{R}y]$ and variance $\mathrm{Var}(y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = \mathbf{R}y)$. It should be noted that the approximating density function $f^*(x, y)$ is nonnormal in general. The Quasi-ML method maximizes the quasi-log-likelihood function, which is the log-likelihood function based on the approximating density $f^*(x, y)$, for the parameter vector $\boldsymbol{\theta}$ by application of standard numerical methods. Technically, this maximization is executed in two stages: In the first stage of the maximization process, the single-step iteration method (Isaacson & Keller, 1966; Schwarz, 1993) is used; in the second stage, the Newton-Raphson algorithm is applied. The maximization algorithm is programmed in Delphi Pascal program code and executed on an IBM compatible computer (Pentium, 1.4 GHz). The typical computing time for one data set lies between 2 and 10 seconds.

The Quasi-ML software is available for download under https://netfiles.uiuc.edu/agklein/QML/ or by contacting Andreas G. Klein for the latest version. The program comes with a manual that describes the model commands and sample data, and it is running under the Windows operating system. The program requires the data to be in ASCII format, and the model is specified using matrix commands and an input file. Fully standardized solutions, standard errors for the parameters, and likelihood ratio tests for testing simultaneous nonlinear effects can also be obtained.

For structural equation models with interaction and quadratic effects among latent variables, the two-step method of moments technique proposed by Wall and Amemiya (2000, 2003) is a procedure that uses only part of the statistical information used in the Quasi-ML method. The difference between Quasi-ML and their two-step approach can be best illustrated by regarding a model with one y-variable only, where $y = \mathbf{v}'\boldsymbol{\beta} + \zeta$ and $\mathbf{v}$ is a vector that carries the constant 1, the latent ξ-variables, and the product terms among ξ-variables that are involved in the structural equation. In the first step of Wall and Amemiya's approach, the parameters of the measurement model for $\boldsymbol{\xi}$ are estimated, for example, by using ML factor analysis. In the second step, the quadratic matrix $E(\mathbf{vv}') \approx N^{-1}\Sigma_i E(\mathbf{vv}'|x_i)$ of product moments is estimated, where the conditional product moments of the matrices $E(\mathbf{vv'}|x_i)$ are computed using the measurement model estimated in step one. Similarly, the vector $E(\mathbf{v}y)$ is estimated. In analogy to ordinary linear regression, estimates for the coefficient vector $\boldsymbol{\beta}$ are then obtained by solving the estimating equation $E(\mathbf{vv'})\boldsymbol{\beta} = E(\mathbf{v}y)$ for $\boldsymbol{\beta}$. It can be shown that the Quasi-ML method leads to an equivalent estimating equation for $\boldsymbol{\beta}$ if in Quasi-ML the variance function $\mathrm{Var}(y|x)$ is constrained to a constant variance term and the parameter estimates of the measurement model for $\boldsymbol{\xi}$ are fixed at the values that are obtained in step one of Wall and Amemiya's approach. Wall and Amemiya's (2000, 2003) approach combines a stepwise estimation of parameters with a method of moments estimation technique. It does not utilize the statistical information provided by the variance function that is deduced from the latent product terms for the Quasi-ML method proposed in this paper. In general, the Quasi-ML method could as well be algebraically extended to polynomial structural equation models with model-implied variance functions, although this is not pursued here because higher order polynomials might be of limited interest for practical applications.

The difference between the LMS method (Klein & Moosbrugger, 2000) and Quasi-ML is that LMS provides an ML estimation under the assumption that the predictor and error variables are all normally distributed. The LMS method relies on a numerical approximation of the nonnormal density function, which becomes computationally very intensive when the number of latent product terms increases. In contrast to this, Quasi-ML depends on the normality assumption about the predictor and error variables only insofar as the mean and variance functions are derived under this assumption. In practice, the mean and variance functions may still be good approximations even if the predictor and error variables are nonnormal. Moreover, the major difference between LMS and Quasi-ML lies in the fact that the computational burden does not critically increase for Quasi-ML when many latent product terms are included in the model, but it does critically increase for LMS.

For the computation of confidence intervals for the Quasi-ML parameter estimates, standard errors can be computed. The calculation of standard errors under Quasi-ML is straightforward and uses the "sandwich estimator" $\mathbf{J}^*$ (Carroll et al., 1995), which estimates the covariance matrix of the Quasi-ML estimator. The sandwich estimator is given by

$$\mathbf{J}^* = N^{-1}\mathbf{H}^{-1}\mathbf{J}\mathbf{H}^{-1} \tag{5}$$

where N is the sample size, and $\mathbf{H}$ and $\mathbf{J}$ are the matrices.

$$\begin{aligned} H &= \left(-E_{\hat{\theta}}\left[\frac{\partial^2 \ln f^*(\mathbf{x},\mathbf{y})}{\partial\theta_i\,\partial\theta_j}\right]\right), \\ J &= \left(E_{\hat{\theta}}\left[\left(\frac{\partial \ln f^*(\mathbf{x},\mathbf{y})}{\partial\theta_i}\right)\left(\frac{\partial \ln f^*(\mathbf{x},\mathbf{y})}{\partial\theta_j}\right)\right]\right). \end{aligned} \tag{6}$$

The adjustment of the information matrix $\mathbf{J}$ given by (5) is necessary in order to correct for a bias of estimation of standard errors. For the computation of the expectation values in the matrices $\mathbf{H}$ and $\mathbf{J}$, two alternative ways have been developed: They are either computed by stochastic integration, using a generated large sample of the indicator vector $(\mathbf{x}_t', \mathbf{y}_t')'$ (expected Fisher information), or they are computed by using the observed sample for stochastic integration (observed Fisher information). In the first case, typically a sample size between 20.000 and 30.000 is used for data generation, depending on the complexity of the model. The sample is generated according to the model equations specified by equations (1), (2), (3), where the Quasi-ML estimates are chosen for the parameter values. Simulation studies not reported here have indicated that standard errors based on the observed Fisher information are often preferable, which is in line with the suggestions by Efron and Hinkley (1978).

With the calculation of likelihood ratio test statistics based on the quasi-log-likelihood function, model difference tests can be carried out under the Quasi-ML method, and nested nonlinear structural equation models can be tested for significant differences in the model structure. For example, the statistical significance of multiple latent nonlinear effects can be tested simultaneously. Because the quasi-log-likelihood function is only an approximation of the correct log-likelihood function, the likelihood ratio test statistic $L = -2(l(\hat{\mathbf{\Theta}}_0) - l(\hat{\mathbf{\Theta}}))$ cannot be expected to asymptotically follow an exact chi-square distribution under the null hypothesis. However, simulation results not further reported here suggest that it is very close to the theoretical chi-square distribution for models with multiple latent interaction effects.

SIMULATION STUDIES

The Quasi-ML method has been specifically developed with the goal to provide a robust, efficient, and computationally inexpensive analysis of structural equation models with multiple nonlinear effects. In this section, the finite sample properties of the Quasi-ML estimation method are compared with four alternative estimation techniques: the latent moderated structural equations (LMS) approach (Klein & Moosbrugger, 2000), the unconstrained approach (Marsh et al., 2004), the constrained approach (Marsh et al., 2004), and the generalized appended product indicator (GAPI) approach (Wall & Amemiya, 2001). We conduct four major simulation studies. In Studies I and II, we compare Quasi-ML with LMS for an elementary interaction model under the condition that the distributional assumptions are met or violated, respectively. In Study III we compare the robustness of Quasi-ML with LMS for a complex model with three latent interaction effects. In Study IV we repeat a previously published simulation study for Quasi-ML (Vers. 2.61) and compare its performance with simulation results known for the unconstrained approach, the constrained approach, and the GAPI approach.

Quasi-ML Versus LMS (Studies I, II, and III)

The LMS method is a full information maximum likelihood approach which assumes that the ξ-variables are normally distributed. The normality assumption might be too rigid in practice when empirical data sets are analyzed. LMS is based on the expectation-maximization (EM) algorithm, and it becomes computationally very intensive for more complex models with multiple interaction terms because it heavily relies on multidimensional numerical integration, and the dimensionality of the integration depends on the number of predictor variables involved in the latent product terms. Quasi-ML only assumes that the conditional distribution of the latent criterion given the x-variables can be approximated as a normal distribution. Quasi-ML can be expected to be more robust than LMS, when the normality assumption for the ξ-variables is violated. For Quasi-ML the computational complexity does not increase significantly when multiple product terms are involved in the nonlinear part of the model. This expected difference in robustness between Quasi-ML and LMS may become even more evident when a complex model is analyzed. Apart from the potential robustness problem for LMS stays the question whether Quasi-ML is as efficient as the ML estimator of LMS when the distributional assumptions are met and only a simple model is analyzed. To investigate this, we conducted three simulation studies: First, we repeated the two studies described in Klein and Moosbrugger (2000, p. 468) for an elementary interaction model with exactly the same parameter values and simulation conditions and then analyzed the simulated data sets with Quasi-ML. For the first study (Study I), data were generated according to the normality assumption. For the second study (Study II), we investigated the robustness, and data were generated for the same elementary interaction model as in Study I, but with the distributional assumptions violated. In Study II, the nonnormal data were generated exactly as described in Klein and Moosbrugger (2000, p. 469). For the generation of the nonnormal data, the EQS program was used, which allows the specification of values for skewness and kurtosis for data generation of variables. In this way, the data for the exogenous variables were generated in EQS, and then the data for the observed variables were computed from these according to the model equations, using a Pascal program. An example of an EQS data generation file for skewed data can be downloaded under https://netfiles.uiuc.edu/agklein/QML/ or by contacting Andreas G. Klein. For the third study (Study III), we conducted a robustness study for a complex model with three latent interaction terms with distributional assumptions violated and then analyzed data sets both with Quasi-ML and LMS.

For Studies I and II, an elementary interaction model with the following structural equation was selected:

$$\eta_t = \alpha + \gamma_1 \xi_{1t} + \gamma_2 \xi_{2t} + \omega_{12} \xi_{1t} \xi_{2t} + \zeta_t. \tag{7}$$

The following parameter values were selected: $\alpha = 1.00$, $\gamma_1 = 0.20$, $\gamma_2 = 0.40$, $\omega_{12} = 0.70$, $\phi_{11} = 0.49$, $\phi_{21} = 0.235$, $\phi_{22} = 0.64$, $\lambda_{x21} = 0.60$, $\lambda_{x42} = 0.70$, $\psi = 0.20$. The predictor variables ξ_{1t}, ξ_{2t} were measured by two x-variables each; ξ_{1t} was measured with reliabilities .49, .22, and ξ_{2t} was measured with reliabilities .64, .38. The criterion variable η_t was measured by one y-variable without error: $y_t = \eta_t$. The selection of $\omega_{12} = 0.70$ gives a model where the percentage of variance of η explained by the interaction term is 33%, which is a very large interaction effect under substantive considerations. Thus, the study tested the performance of the method for an extremely nonlinear model. The reliabilities were selected to be not very high in order to evaluate the performance of the method under reasonably difficult conditions. The specified model has 14 parameters and five indicator variables. It has been investigated before by Jöreskog and Yang (1996), Schermelleh-Engel et al. (1998), and Klein and Moosbrugger (2000). For Study I, the data for the latent exogenous variables were generated according to the normal distribution. For Study II, the data for the latent exogenous variable ξ_1 were generated using EQS (Bentler, 1995) with a skewness of -2.0 and a kurtosis of 6.0; the data for the latent exogenous variable ξ_2 were generated with a skewness of $+1.5$ and a kurtosis of 5.0. The endogenous error variables were simulated as normally distributed variables in both Studies I and II. In Studies I and II, 500 data sets of sample size $N = 400$ for the five indicator variables were then generated, according to the model equations. The data sets were then analyzed with Quasi-ML. For computation of Quasi-ML standard errors, the observed Fisher information was used. Table 1 shows the simulation results for Quasi-ML in Studies I and II in comparison with the results previously published for LMS (Klein & Moosbrugger, 2000, pp. 468, 470; Schermelleh-Engel et al., 1998). We restrict the report of the simulation results to the three structural parameters of interest, γ_1, γ_2, and ω_{12}. Under the normal condition (Study I) both Quasi-ML and LMS showed virtually unbiased estimates and a relative bias within $\pm 3\%$ around the true values across all parameters. Under the nonnormal condition (Study II), the parameters not listed in Table 1 showed a relative bias within $\pm 12\%$ for Quasi-ML and within $\pm 22\%$ for LMS, suggesting that Quasi-ML is more robust against the violation of the distributional assumption.

In Table 1, under the nonnnormal condition (Study II), a bias for the parameter estimate of γ_1 is noticed for both methods, due to the skewness of the predictor variables. In the model simulated in Study II, the variance proportion of η explained by the interaction term $0.7\xi_1\xi_2$ is 31%, which is an extremely

TABLE 1
Estimation Results of Studies I and II for the Elementary Interaction Model With One Latent Interaction Effect (Equation 7)

	True Value	*Quasi-ML (Study I, Normal Condition) 500 Replications, N = 400*				*Quasi-ML (Study II, Nonnormal Condition) 500 Replications, N = 400*			
		M	*SD*	*SE*	*SE/SD*	*M*	*SD*	*SE*	*SE/SD*
γ_1	0.200	0.206	0.071	0.069	0.97	0.065	0.096	0.090	0.94
γ_2	0.400	0.395	0.063	0.062	0.98	0.409	0.077	0.072	0.94
ω_{12}	0.700	0.694	0.111	0.110	0.99	0.668	0.122	0.113	0.93
	True Value	*LMS (Normal Condition) (Study I, Normal Condition) 500 Replications, N = 400*				*LMS (Nonnormal Condition) (Study II, Nonnormal Condition) 500 Replications, N = 400*			
		M	*SD*	*SE*	*SE/SD*	*M*	*SD*	*SE*	*SE/SD*
γ_1	0.200	0.196	0.064	0.065	1.02	0.067	0.100	0.077	0.77
γ_2	0.400	0.411	0.061	0.061	1.00	0.394	0.069	0.061	0.88
ω_{12}	0.700	0.698	0.094	0.102	1.09	0.729	0.154	0.110	0.71

Note. In both studies, 500 data sets of sample size $N = 400$ were analyzed with Quasi-ML. The columns give for every model parameter the true value, the mean (M) of parameter estimates, the Monte-Carlo standard deviation (SD) of parameter estimates, the mean of estimated standard errors (SE) across all data sets, and the *SE/SD* ratios.

strong interaction effect compared with what is typically found in empirical studies (Jaccard & Wan, 1995). The reason for the bias lies in the fact that the variance function specified within Quasi-ML becomes incorrect when there is considerable skewness of the predictors involved the interaction term. Other studies not reported here revealed that the bias for a parameter appears only in the situation of strong skewness of the predictors in combination with a very strong interaction effect. In this case, the user of Quasi-ML should be aware of a possible bias. For a future improvement of Quasi-ML it is planned to modify the variance function appropriately to accommodate this special situation.

Under the normal condition (Study I), the Monte-Carlo standard deviations (SD) are somewhat smaller for LMS than they are for Quasi-ML. The largest difference occurs for ω_{12}, with $SD(\omega_{12})_{\text{LMS}}$ being 15% smaller than $SD(\omega_{12})_{\text{Quasi-ML}}$. This matches the expectation of the possibly somewhat lower efficiency of Quasi-ML under the normal condition because Quasi-ML is only an approximate ML estimator. But the simulation results under the nonnormal condition (Study II) show that in case of violated distributional assumptions, Quasi-ML clearly provides a more efficient estimation of the interaction parameter ω_{12} than LMS does.

Here, $SD(\omega_{12})_{\text{LMS}}$ is 26% larger than $SD(\omega_{12})_{\text{Quasi-ML}}$. The results further show that for Quasi-ML the standard errors are estimated without any critical bias: The *SE/SD* ratios for Quasi-ML are all close to one, whereas for LMS the ratio has an undesirably low value of 0.71 for ω_{12}.

The robustness of the Quasi-ML method in terms of bias, Type I error, and power was further examined by repeating the robustness study (Study II) with different values for the interaction parameter ($\omega_{12} = 0.0, 0.1, 0.2, 0.7$), while all other conditions of Study II were held constant. For the study with $\omega_{12} = 0.0$, the bias was 0.03 for ω_{12}, for the studies with $\omega_{12} = 0.1, 0.2$, the relative bias across all parameters lay within $\pm 7\%$, $\pm 8\%$ around the true value, respectively. This suggests that the bias under violation of the normality assumption does not become very large when the interaction effect size is not of an extreme size. Also, the robustness of the likelihood ratio test for testing the interaction parameter ω_{12} against zero was investigated in these studies. The observed Type I error/power for Quasi-ML under the four conditions was 6.5%, 46.5%, 88.0%, and 100.0% under the conditions $\omega_{12} = 0.0, 0.1, 0.2, 0.7$, respectively. These values give the rates of data sets for which the test was significant at the 5% Type I error level in the simulation studies. For LMS the observed Type I error/power was 10.4%, 48.0%, 93.5%, and 100.0%, respectively (Klein & Moosbrugger, 2000). Thus, with 6.5% the Type I error for Quasi-ML is not substantially inflated as compared with a more inflated Type I error of 10.4% for LMS. Under the conditions studied, the likelihood ratio test is more robust for Quasi-ML than it is for LMS. Altogether, Study II indicates that when the normality assumption is violated, Quasi-ML is more robust than LMS. For Study III, a complex model with multiple nonlinear effects was selected:

$$\begin{aligned}\eta_t = \alpha + \gamma_1\xi_{1t} + \gamma_2\xi_{2t} + \gamma_3\xi_{3t} + \omega_{12}\xi_{1t}\xi_{2t} \\ + \omega_{13}\xi_{1t}\xi_{3t} + \omega_{23}\xi_{2t}\xi_{3t} + \zeta_t.\end{aligned} \tag{8}$$

The following parameter values were selected: $\alpha = -0.05$, $\gamma_1 = 0.30$, $\gamma_2 = 0.40$, $\gamma_3 = 0.50$, $\omega_{12} = 0.10$, $\omega_{13} = -0.20$, $\omega_{23} = 0.20$, $\phi_{11} = \phi_{22} = \phi_{33} = 1.00$, $\phi_{21} = 0.30$, $\phi_{31} = 0.10$, $\phi_{32} = 0.20$, $\lambda_{x21} = \lambda_{x42} = \lambda_{x63} = 0.70$, $\psi = 0.40$. This model has three latent predictor variables, ξ_{1t}, ξ_{2t}, ξ_{3t}, and one latent criterion variable, η_t. The predictor variables ξ_{1t}, ξ_{2t} were measured by two x-variables each, which had reliabilities of .70 and .53. The criterion variable η_t was measured by one y-variable without error: $y_t = \eta_t$. The model includes three latent interaction effects among the predictor variables. The structural coefficients ω_{12}, ω_{13}, ω_{23} were selected to have opposite signs and represent relatively small effects to give a more realistic situation than in Studies I and II.

In Study III, it was of particular interest to see how precisely the Quasi-ML method estimates and separates small multiple nonlinear effects of opposite signs. For Study III, the data for the latent exogenous variables ξ_{1t}, ξ_{2t}, ξ_{3t}

were generated nonnormally using EQS with [skewness, kurtosis] = [−1.5, 4.0], [1.5, 5.0], [0.5, 5.0], respectively. The endogenous error variables were all simulated as normally distributed variables. The specified model has 23 parameters and seven indicator variables. Two hundred fifty replications of data sets with sample size $N = 400$ were generated using the model equations and then analyzed with the Quasi-ML method and the LMS method. The computational complexity of this model under LMS is already large and requires two-dimensional numerical integration; it can grow exponentially with the number of additional interacting predictor variables added. For Quasi-ML, the computational complexity mainly increases linearly with the number of parameters. For computation of Quasi-ML standard errors, the observed Fisher information was used. The estimation results for the structural parameters are given in Table 2.

The results show no substantial bias for Quasi-ML, but for LMS the estimates for γ_2 and ω_{12} are biased. For all six structural parameters, the Monte-Carlo standard deviations (*SD*) are all equal or less for Quasi-ML than they are for LMS, so Quasi-ML produces at least as efficient estimates as LMS. For the estimation of ω_{23}, Quasi-ML is clearly more efficient than LMS. The Monte-Carlo standard deviations are estimated with little bias for Quasi-ML, because the *SE/SD* ratios lie between 0.90 and 1.08. Opposed to this, the study reveals a clear underestimation of the Monte-Carlo standard deviations for LMS, for which the *SE/SD* ratios lie between 0.75 and 0.93. Altogether, Study III indicates that Quasi-ML is the more robust technique: it shows no substantial bias, is as efficient as LMS,

TABLE 2
Estimation Results of Study III for a Complex Interaction Model With Three Latent Interaction Effects (Equation 8)

		Quasi-ML (Study III, Nonnormal Condition) 250 Replications, N = 400				*LMS* (Study III, Nonnormal Condition) 250 Replications, N = 400			
	True Value	*M*	*SD*	*SE*	*SE/SD*	*M*	*SD*	*SE*	*SE/SD*
γ_1	0.300	0.296	0.056	0.057	1.02	0.264	0.057	0.053	0.93
γ_2	0.400	0.401	0.062	0.060	0.97	0.462	0.062	0.055	0.89
γ_3	0.500	0.493	0.062	0.057	0.92	0.504	0.066	0.055	0.83
ω_{12}	0.100	0.094	0.062	0.056	0.90	0.023	0.065	0.049	0.75
ω_{13}	−0.200	−0.210	0.059	0.064	1.08	−0.219	0.065	0.060	0.92
ω_{23}	0.200	0.225	0.055	0.057	1.04	0.230	0.071	0.055	0.77

Note. Two hundred fifty data sets of sample size $N = 400$ were analyzed with Quasi-ML and LMS. The columns give for every model parameter the true value, the mean (*M*) of parameter estimates, the Monte-Carlo standard deviation (*SD*) of parameter estimates, and the mean of estimated standard errors (*SE*) across all data sets, and the *SE/SD* ratios.

and provides more accurate confidence interval estimates for the structural parameters in this complex model, where the distributional assumptions for the predictor variables have been violated.

Quasi-ML Versus Other Approaches (Study IV)

Marsh et al. (2004) conducted a comprehensive simulation study where the performances of four methods developed for latent interaction models were compared: the unconstrained approach (Marsh et al.), the constrained approach (Marsh et al.), the generalized appended product indicator (GAPI) approach (Wall & Amemiya, 2001), and the Quasi-ML method (Vers. 1.31). In a complex simulation design for a model with two predictors, one latent interaction term, and nine indicator variables, they varied interaction effect size, sample size, correlation among predictor variables, and the distribution type of the endogenous variables in data generation. Interaction effect size was varied across two levels ($\omega_{12} = 0, 0.2$), sample size was varied across three levels ($N = 100, 200, 500$), correlation was varied across two levels ($\phi_{12} = .3, .7$), and distribution type was varied across three levels (normal, uniform, chi-square ($df = 6$)). Somewhat uncommon, for the nonnormal conditions not only the latent predictors but also the error variables were simulated according to the nonnormal distribution type, which results in severe violation of normality in particular under the condition "chi-square." This resulted in a simulation design with 36 conditions. Under each condition, they generated 250 replications of data sets. For the details of their study and design, we refer to Marsh et al.

For this article, we chose to repeat their simulation design using Quasi-ML (Vers. 2.61) and compare it to their results reported for the other methods. For the standard errors reported by Marsh et al. (2004), who used Quasi-ML (Vers. 1.31), the expected Fisher information was used because no other options were available at that time. Version 2.61 now uses improved standard errors based on the observed Fisher information. Also, we applied the likelihood ratio test in Quasi-ML for testing the interaction parameter against zero. We repeated their study for Quasi-ML (Vers. 2.61) for $N = 200$ under the remaining 12 conditions of their design. We generated the data according to their instructions given and then re-analyzed the data sets with Quasi-ML (Vers. 2.61). The results of our simulation study for Quasi-ML are given in Table 3 and compared with the results taken from Marsh et al. Only the results for the interaction parameter are given here, and Table 3 has a format similar to Tables 6, 7, and 8 given by Marsh et al. and can thus be directly compared. From the bias and the standard deviation (SD) of estimates, the root mean squared error (RMSE) was computed: RMSE $= \sqrt{Bias^2 + SD^2}$. The RMSE is the square root of the expected squared loss around the true parameter value. It is a suitable measure to compare the

TABLE 3
Quasi-ML Estimation Results of the Repeated Simulation Study for a Latent Interaction Model With One Interaction Parameter ω_{12}

	$\omega_{12} = 0$			$\omega_{12} = 0.2$		
Approach	*TI Error*	*RMSE*	*SE/SD*	*Power*	*RMSE*	*SE/SD*
	$N = 200$, Normal Condition, $\phi_{12} = .3$, see Marsh et al., (2004), Table 6					
Unc	.032	.125	.84	.49	.144	.80
Con	.056	.114	.88	.51	.123	.84
GAPI	.056	.128	.87	.46	.144	.81
Quasi-ML	.073 (.106)	.095 (.100)	.96 (.80)	.60 (.60)	.112 (.110)	.86 (.86)
	$N = 200$, Normal Condition, $\phi_{12} = .7$, see Marsh et al., (2004), Table 6					
Unc	.049	.084	.91	.69	.107	.82
Con	.048	.081	.93	.71	.098	.86
GAPI	.040	.091	.91	.66	.114	.83
Quasi-ML	.056 (.141)	.076 (.073)	.96 (.74)	.84 (.84)	.085 (.085)	.93 (.89)
	$N = 200$, Uniform Condition, $\phi_{12} = .3$, see Marsh et al., (2004), Table 7					
Unc	.044	.140	.86	.41	.175	.81
Con	.060	.118	.86	.54	.115	.93
GAPI	.060	.133	.88	.45	.146	.90
Quasi-ML	.045 (.096)	.093 (.101)	1.01 (.79)	.62 (.61)	.097 (.097)	1.03 (.98)
	$N = 200$, Uniform Condition, $\phi_{12} = .7$, see Marsh et al., (2004), Table 7					
Unc	.044	.131	.91	.38	.149	.96
Con	.072	.098	.86	.48	.100	.95
GAPI	.048	.130	.92	.36	.147	.97
Quasi-ML	.041 (.189)	.078 (.083)	1.09 (.64)	.56 (.67)	.084 (.084)	1.11 (.94)
	$N = 200$, Chi-Square Condition, $\phi_{12} = .3$, see Marsh et al., (2004), Table 8					
Unc	.032	.118	.92	.50	.159	.78
Con	.032	.104	.99	.65	.115	.94
GAPI	.036	.115	.96	.52	.140	.84
Quasi-ML	.057 (.114)	.091 (.106)	1.10 (.83)	.79 (.78)	.099 (.099)	1.12 (1.07)
	$N = 200$, Chi-Square Condition, $\phi_{12} = .7$, see Marsh et al., (2004), Table 8					
Unc	.040	.088	.89	.77	.100	.87
Con	.120	.091	.95	.95	.111	.85
GAPI	.048	.086	.92	.77	.099	.88
Quasi-ML	.085 (.201)	.084 (.092)	1.08 (.76)	.96 (.96)	.104 (.104)	1.09 (1.03)

Results are taken from 250 replications of sample size $N = 200$ under twelve conditions ($\omega_{12} = 0, 0.2$; $\phi_{12} = .3, .7$; distribution type = normal, uniform, chi-square ($df = 6$)). The simulation results for Quasi-ML are compared with the results for the unconstrained (Unc) approach, the constrained (Con) approach, and the generalized appended product indicator (GAPI) approach as given by Marsh et al. (2004). For Quasi-ML, the results reported earlier by Marsh et al. are given in parentheses. The columns give for the interaction parameter: the Type I error or power, the root mean squared error (RMSE), and the ratio between average estimated standard error and Monte-Carlo standard deviation of estimates (*SE/SD*).

precision of estimators across different methods because it combines bias and standard error.

The relative bias of Quasi-ML for the parameters of the linear effects of the two predictors was within ±5% under all 12 design conditions. The bias for the interaction parameter ω_{12} under the conditions "$\omega_{12} = 0$, normal" and "$\omega_{12} = 0$, uniform" was smaller than 0.01. The bias for ω_{12} under the condition "$\omega_{12} = 0$, chi-square, $\phi_{12} = .3$" was 0.03, and under the condition "$\omega_{12} = 0$, chi-square, $\phi_{12} = .7$" it was 0.05. The bias for ω_{12} under the conditions "$\omega_{12} = 0.2$, normal" and "$\omega_{12} = 0.2$, uniform" was smaller than 0.02. Under the conditions "$\omega_{12} = 0.2$, chi-square, $\phi_{12} = .3$" the bias for ω_{12} was with 0.040 (= +20%) somewhat larger, and under "$\omega_{12} = 0.2$, chi-square, $\phi_{12} = .7$" it was again larger with 0.065 (= +32%). Apart from the structural parameters, the relative bias of all the other parameters was within ±5% under all 12 design conditions. The bias for ω_{12} under the condition "$\omega_{12} = 0.2$, chi-square," for which the violation of the normality assumption is most severe, is similar to the bias found for Marsh's constrained approach (Unc), where additional constraints are imposed that are based on the normality assumption. To further validate the results for Quasi-ML, other indicators such as root mean squared error (RMSE) had to be investigated.

The results show that when the Quasi-ML simulation results are compared with the results reported for Quasi-ML by Marsh et al. (2004), their results under condition $\omega_{12} = 0.2$ could be replicated for power, RMSE, and *SE/SD*. But under the condition $\omega_{12} = 0$ the new results obtained under Version 2.61 deviate substantially from their results on Quasi-ML (Vers. 1.31). The new results do not show any critical inflation of Type I error as reported by them, and the Type I error for Quasi-ML does not substantially exceed the nominal level of 5% for the test of the hypothesis H0: $\omega_{12} = 0$. Table 3 also shows that under the condition $\omega_{12} = 0.2$, Quasi-ML clearly outperforms the other three methods with respect to power across all distribution types. Also, when compared with the other three methods, the RMSE is the lowest for Quasi-ML in 11 out of the 12 conditions; only under the condition with $\omega_{12} = 0.2$, chi-square distribution, $\phi_{12} = .7$, the RMSE for Quasi-ML is minimally larger than it is for the unconstrained and the GAPI approach. Thus, the results of this particular validation study indicate that Quasi-ML tends to give the most precise estimates for the interaction parameter.

The results also show that the RMSE for Quasi-ML increases only very moderately when the distributional assumptions are violated, with the largest increase from 0.85 to 0.104 under the condition $\omega_{12} = 0.2$, $\phi_{12} = .7$. For the *SE/SD* ratios, the table shows that the values for Quasi-ML lie between 0.86 and 1.12. Under the nonnormal distribution conditions, the Quasi-ML standard errors based on the observed Fisher information tend to overestimate the Monte-Carlo standard error up to 12%, whereas the other three methods tend to underestimate the Monte-Carlo standard error, with *SE/SD* ratios as low as 0.78 for Unc, as low as 0.85 for Con, and as low as 0.84 for GAPI. Although not displayed in Table 3, under the normal distribution condition, the Quasi-ML estimates showed no indication of bias, in line with what was previously stated by Marsh et al. (2004). Furthermore, under the normal distribution condition, the *SE/SD* ratios for Quasi-ML are closer to one than they are for the other three methods, indicating that Quasi-ML gives more precise interval estimates for the interaction parameter than the other three techniques, when the normality assumption holds. Future simulation research with different models, varied sample size, and varied degree of nonnormality in the variables will decide if these primary results about the robustness of Quasi-ML can be generalized.

EMPIRICAL EXAMPLE

This section covers an empirical example of an aging study in psychology for a nonlinear latent variable model with two interactions, three quadratic effects, and seven indicator variables. The empirical data set was collected by Thiele (1998), who investigated age-related effects of coping strategies and the maintenance of well-being for middle-aged males. The sample size was 302, and males in the age range from 35 to 64 years were examined. Starting from the theory of primary and secondary control in life span development (Heckhausen & Schulz, 1993) and the theory of assimilative versus accommodative coping (Brandstädter & Renner, 1990), Thiele examined the impact of different coping strategies. He investigated the effect of subjectively perceived fitness (ξ_{1t}), objective fitness (ξ_{2t}), and flexibility in goal adjustment (ξ_{3t}) on the level of complaining about one's mental or physical situation (η_t). He formulated the interaction hypotheses that the effect of flexibility in goal adjustment on the complaint level is high when individuals have low values on the fitness scales but that it is neutralized or weak for individuals with high values on the fitness scales. For participants with a high level of subjective fitness, the flexibility of goal adjustment is supposed to have only a small or negligible effect on complaint level, whereas for persons with a low perceived availability of bodily resources, the flexibility of goal adjustment is expected to be an important factor for the level of complaining.

The interaction hypotheses can be modeled by including appropriate product terms ($\xi_{1t}\xi_{3t}$ and $\xi_{2t}\xi_{3t}$) in a structural equation model. The interaction model itself is nonadditive, and the alternative model is a model where the predictor variables have only additive effects on the criterion, although these additive effects are not necessary linear. Ganzach (1997) provides arguments to include the quadratic terms of the interacting predictors in regression models with multiple interactions in addition to the interaction terms, in order to avoid the estimation of artifactual interactions due to overlooked quadratic effects and multicollinearity between interaction and quadratic terms. Therefore, we selected the following nonlinear structural equation for analysis:

$$\begin{aligned} \eta_t &= \alpha + \gamma_1\xi_{1t} + \gamma_2\xi_{2t} + \gamma_3\xi_{3t} + \omega_{11}\xi_{1t}^2 + \omega_{22}\xi_{2t}^2 \\ &\quad + \omega_{33}\xi_{3t}^2 + \omega_{13}\xi_{1t}\xi_{3t} + \omega_{23}\xi_{2t}\xi_{3t} + \zeta_t. \end{aligned} \tag{9}$$

This initial model has five nonlinear effects. The subjectively perceived fitness (ξ_{1t}) refers to the self-evaluation of the effectiveness with which one's body is functioning. It was measured by a split scale (x_1, x_2) of the Frankfurt self-concept scales of bodily efficiency (Deusinger, 1998), and the construction of the item split is described in detail by Thiele (1998). The objective fitness (ξ_{2t}) refers to the objective level of fitness. It was measured by lung volume (x_3). Flexibility in goal adjustment (ξ_{3t}) addresses the fact that individuals are more or less willing to adapt their goals to the limits given by their individual physical situation or health condition, which refers to the coping style of a person. The latent predictor variable ξ_{3t} was measured by two subscales (x_4, x_5) taken from a flexibility scale developed by Brandstädter and Renner (1990). Sample items were asking about recovery from disappointments or if a person gives up easily The complaint level (η_t) was measured by two indicators (y_1: psychological complaints, sample items were asking about feelings of mental exhaustion or having depressive symptoms; y_2: psychovegetative complaints, sample items were asking about dizziness or backache) given by the complaint inventory of Degenhardt and Schmidt (1994). The data were z-standardized and analyzed with the Quasi-ML method. The univariate skewness of the five x-variables was between -0.33 and 0.01, their univariate kurtosis was between -0.07 and 0.30.

Initially, the full model (8) with all five nonlinear effects was estimated, and the estimates for the structural coefficients are listed in Table 4. For the measurement model, the estimated communalities (percentage of explained variance) for the seven observed variables were .91 for x_1, .60 for x_2, 1.00 for x_3, .99 for x_4, .36 for x_5, .67 for y_1, and .75 for y_2. The estimated correlations between the three ξ-variables lay between -0.08 and 0.24. The analysis of the data set revealed that the subjectively perceived fitness has larger linear and quadratic effects than the objective fitness. A simultaneous interaction hypothesis (H0: $\omega_{13} = \omega_{23} = 0$) was tested using a likelihood ratio test statistic based on the quasi-likelihood, which gave a significant result ($\chi^2 = 7.25$, $p < .05$, $df = 2$), suggesting that there is significant interaction beyond the additive quadratic and linear effects. However, the standardized estimates for ω_{22} and ω_{23} were relatively small (0.029 and 0.08, respectively), and thus it was decided to modify and re-estimate the model with now ω_{22} and ω_{23} being fixed to zero (see Table 4).

TABLE 4
Parameter Estimates, Estimated Standard Errors, and Standardized Parameter Estimates for the Structural Equation Parameters and the Variance of the Disturbance Term Provided by the Quasi-ML Method

	Initial Model			*Re-estimated Model*		
Parameter	*Estimate*	*SE*	*Standardized Estimate*	*Estimate*	*SE*	*Standardized Estimate*
γ_1	−0.389	0.068	−0.452	−0.393	0.067	−0.455
γ_2	−0.104	0.048	−0.126	−0.099	0.048	−0.120
γ_3	−0.225	0.058	−0.274	−0.223	0.052	−0.269
ω_{11}	−0.051	0.030	−0.056	−0.047	0.029	−0.052
ω_{22}	0.024	0.030	0.029	—	—	—
ω_{33}	0.083	0.039	0.101	0.074	0.036	0.089
ω_{13}	0.106	0.055	0.123	0.125	0.056	0.145
ω_{23}	0.066	0.061	0.080	—	—	—
ψ	0.384	0.081	0.572	0.396	0.075	0.582

In the re-estimated model, again the likelihood ratio test (H0: $\omega_{13} = 0$) gave a significant result for the interaction effect ($\chi^2 = 5.53$, $p < .05$, $df = 1$) beyond the quadratic and linear effects.

For the standardized solution of the re-estimated model, the negative sign of γ_2 and the additive effects $-0.455\xi_{1t} - 0.052\xi_{1t}^2$ and $-0.269\xi_{3t} + 0.089\xi_{3t}^2$ indicate the expected, generally negative relationship between the latent variables ξ_{1t}, ξ_{2t}, ξ_{3t}, and η_t: High values for subjective or objective fitness and high values for flexibility in goal adjustment predict a low complaint level. However, the positive sign of ω_{13} indicates the expected direction of the interaction between subjective fitness and flexibility in goal adjustment: Conditional on a high level of subjective fitness (ξ_{1t}), the effect of flexibility in goal adjustment (ξ_{3t}) on complaint level (η_t), which is given by the entire term $[-0.269 + 0.145\xi_{1t}]\xi_{3t} + 0.089\xi_{3t}^2$, becomes smaller because the moderator function $[-0.269 + 0.145\xi_{1t}]$ gets closer to zero and the variance of the term decreases. Analogously, at low levels of subjective fitness the flexibility level of goal adjustment has a more substantial impact on complaint level. Taking the high efficiency and the modeling capabilities of the Quasi-ML method into account, the new estimation method can be used to separate between multiple nonlinear effects among latent or observed variables and to test a model with a block of interaction effects against an additive quadratic model. If the model in (9) had been analyzed using one of the product indicator approaches, a large number of products of indicators would have been specified and made the estimation process very likely to be unstable because of sampling variation of and possible multicollinearity between the product indicators.

CONCLUSION

The Quasi-ML method proposed in this article has been developed for an efficient and computationally feasible estimation of multiple nonlinear effects in structural equation models with quadratic forms. In the Quasi-ML approach, the nonnormal density function of the joint indicator vector is approximated by a product of a normal density and a conditionally normal density. For the conditionally normal density, the concept of variance function models is applied.

The simulation studies carried out indicate that the Quasi-ML method is more robust than the LMS method against the violation of distributional assumptions (Studies II, III). When the distributional assumptions are met, Quasi-ML is almost as efficient as LMS (Study I). Quasi-ML provides robust standard errors for the structural parameters (Studies III, IV), and the likelihood ratio test for Quasi-ML showed no substantial inflation even under nonnormal conditions (Study II). Also, Quasi-ML can handle complex models with multiple nonlinear effects (Study III and empirical example with five nonlinear effects) without any computational problems. When compared with three alternative estimation techniques, Quasi-ML estimates were overall more precise and standard error estimates were less biased under the wide majority of the conditions studied (Study IV). Also, Quasi-ML outperformed the alternative methods clearly with respect to statistical power. Limitations of the Quasi-ML method arise when a very large interaction effect size is combined with considerable skewness of the latent predictors (Study II) or if both latent predictors and error variables are considerably skewed (Study IV, “chi-square” condition): Under these special conditions, parameter estimates for the structural parameters can be biased in Quasi-ML.

The applicability of Quasi-ML was further demonstrated by an empirical example, where a model with five latent nonlinear effects was estimated, and a latent interaction was positively tested against an additive model with both linear and quadratic effects. Based on the current state of research, Quasi-ML seems to be an efficient, fairly robust, computationally inexpensive method that can handle a wide range of nonlinear structural equation models. Future research needs to be carried out to further investigate the practical applicability of the new method and to further improve the robustness of Quasi-ML under the condition of skewed variables.

ACKNOWLEDGMENTS

This research was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) under Grant K02 AA 00230-01; by the National Institute of Mental Health (NIMH) and the National Institute on Drug Abuse (NIDA) under Grant MH40859; and by the Research Board of the University of Illinois, Urbana-Champaign, under Grant 03269. The Quasi-ML software is available from Andreas G. Klein upon request.

REFERENCES

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

Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 1–63). New York: Academic.

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice-Hall.

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology, 22,* 453–474.

Algina, J., & Moulder, B. C. (2001). A note on estimating the Jöreskog-Yang model for latent variable interaction LISREL 8.3. *Structural Equation Modeling, 8*, 40–52.

Arbuckle, J. L. (1997). *AMOS users' guide version 3.6.* Chicago: Small Waters Corporation.

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

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate Software.

Bentler, P. M., & Wu, E. J. C. (1993). *EQS/Windows user's guide.* Los Angeles: BMDP Statistical Software.

Blom, P., & Christoffersson, A. (2001). Estimation of nonlinear structural equation models using empirical characteristic functions. In R. Cudeck, S. Du Toit, & D. Soerbom (Eds.), *Structural equation modeling: Present and future* (pp. 443–460). Lincolnwood, IL: Scientific Software.

Bollen, K. A. (1995). Structural equation models that are non-linear in latent variables: A least squares estimator. In P. V. Marsden (Ed.), *Sociological methodology* (Vol. 25, pp. 223–251). Washington, DC: American Sociological Association.

Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, *61,* 109–121.

Bollen, K. A., & Paxton, P. (1998). 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). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Brandstädter, J., & Renner, G. (1990). Tenacious goal pursuit and flexible goal adjustment: Explication and age-related analysis of assimilative and accommodative strategies of coping. *Psychology and Aging, 5*, 58–67.

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

Carroll, R. J., Ruppert, D., & Stefanski, L. A. (1995). *Measurement error in nonlinear models* (1st ed.). London: Chapman & Hall.

Cohen, J., & Cohen, P. (1975). Applied multiple regression/correlational analysis for the behavioral sciences (1st ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116–127.

Cronbach, L. J., & Snow, R. E. (1977). Aptitudes and instructional methods. A handbook for research on interactions. New York: Irvington.

Degenhardt, A., & Schmidt, H. (1994). Physische Leistungsvariablen als Indikatoren für die Diagnose "Klimakterium Virile" [Physical efficiency variables as indicators for the diagnosis of "climacterium virile"]. *Sexuologie, 3,* 131–141.
Deusinger, I. (1998). *Frankfurter Körperkonzept-Skalen* [Frankfurt bodily self-concept scales]. Göttingen: Hogrefe.
Efron, B., & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika*, *65*, 475–487.
Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research.* Reading, MA: Addison-Wesley.
Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods, 2*(3), 235–247.
Gurland, J. (1955). Distribution of definite and of indefinite quadratic forms. *Annals of Mathematical Statistics, 26,* 122–127.
Hayduk, L. A. (1987). *Structural equation modeling with LISREL.* Baltimore: Johns Hopkins University Press.
Heckhausen, J., & Schulz, R. (1993). Optimisation by selection and compensation: Balancing primary and secondary control in life span development. *International Journal of Behavioral Development, 16*(2), 287–303.
Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika, 48,* 419–426.
Isaacson, E., & Keller, H. B. (1966). *Analysis of numerical methods*. New York: Wiley.
Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression.* Newbury Park, CA: Sage.
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.
Johnson, N. L., & Kotz, S. (1970). *Continuous univariate distributions-2.* New York: Wiley.
Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS.
Jöreskog, K. G., & Sörbom, D. (1993). *New features in LISREL 8*. Chicago: Scientific Software.
Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide.* Chicago: Scientific Software.
Jöreskog, K. G., & Yang, F. (1996). Non-linear structural equation models: The Kenny-Judd model with interaction effects. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 57–87). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
Jöreskog, K. G., & Yang, F. (1997). Estimation of interaction models using the augmented moment matrix: Comparison of asymptotic standard errors. In W. Bandilla & F. Faulbaum (Eds.), *SoftStat '97* (Advances in Statistical Software 6, pp. 467–478). Stuttgart: Lucius & Lucius.
Karasek, R. A. (1979). Job demands, job decision latitude, and mental strain: Implications for job redesign. *Administrative Quarterly, 24*, 285–307.
Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin, 96,* 201–210.
Klein, A. (2000). *Moderatormodelle. Verfahren zur Analyse von Moderatoreffekten in Strukturgleichungsmodellen* [Moderator models. Methods for the analysis of moderator effects in structural equation models]. Hamburg: Dr. Kovac.
Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika, 65*(4), 457–474.
Klein, A., & Muthén, B. O. (2006). Modeling heterogeneity of latent growth depending on initial status. *Journal of Educational and Behavioral Statistics, 31*(4), 357–375.

Krickeberg, K., & Ziezold, H. (1988). *Stochastische Methoden* [Stochastic methods]. Berlin: Springer.

Longford, N. T. (1995). Random coefficient models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 519–577). New York: Plenum.

Lusch, R. F., & Brown, J. R. (1996). Interdependency, contracting, and relational behavior in marketing channels. *Journal of Marketing, 60*, 19–38.

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*(3), 275–300.

Moosbrugger, H., Frank, D., & Schermelleh-Engel, K. (1991). Zur Überprüfung von latenten Moderatoreffekten mit linearen Strukturgleichungsmodellen [Estimating latent interaction effects in structural equation models]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 12,* 245–255.

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

Muthén, B. O., & Muthén, L. (2004). *Mplus user's guide, Version 3*. Los Angeles: Author.

Ping, R. A. (1995). A parsimonious estimating technique for interaction and quadratic latent variables. *Journal of Marketing Research, 32,* 336–347.

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

Ping, R. A. (1996b). Latent variable regression: A technique for estimating interaction and quadratic coefficients. *Multivariate Behavioral Research, 31,* 95–120.

Ping, R. A. (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, Inc.

Press, S. J. (1966). Linear combinations of non-central chi-square variates. *Annals of Mathematical Statistics, 37,* 480–487.

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, Inc.

Schwarz, H. R. (1993). *Numerische Mathematik* [Numerical mathematics]. Stuttgart: Teubner.

Shah, B. K. (1963). Distribution of definite and of indefinite quadratic forms from a non-central normal distribution. *Annals of Mathematical Statistics, 34*, 186–190.

Snyder, M., & Tanke, E. D. (1976). Behavior and attitude: Some people are more consistent than others. *Journal of Personality, 44,* 501–517.

Srivastava, M. S., & Khatri, C. G. (1979). *An introduction to multivariate statistics*. New York: Elsevier North Holland.

Thiele, A. (1998). *Verlust körperlicher Leistungsfähigkeit: Bewältigung des Alterns bei Männern im mittleren Lebensalter* [Loss of bodily efficacy: The coping of aging for men of medium age]. Idstein: Schulz-Kirchner-Verlag.

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

Wall, M. M., & Amemiya, Y. (2001). Generalized appended product indicator procedure for nonlinear structural equation analysis. *Journal of Educational and Behavioral Statistics*, *26*, 1–29.

Wall, M. M., & Amemiya, Y. (2003). A method of moments technique for fitting interaction effects in structural equation models. *British Journal of Mathematical and Statistical Psychology*, *56*, 47–64.

Yang Jonsson, F. (1997). Non-linear structural equation models: Simulation studies of the Kenny-Judd model. Uppsala, Sweden: University of Uppsala.

Yang-Wallentin, F., & Jöreskog, K. G. (2001). Robust standard errors and chi-squares for interaction models. In G. A. Marcoulides & R. E. Schumacker & (Eds.), *New developments and techniques in structural equation modeling* (pp. 159–171). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

APPENDIX

Transformation of Indicator Vector

Because the latent criterion variable η_t includes a quadratic form of normally distributed variates, the indicator vector $(\mathbf{x}_t', \mathbf{y}_t')'$ is in general not normally distributed. But by an appropriate transformation, the indicator vector can be transformed such that $q+p-1$ of the $q+p$ components of the transformed vector are normally distributed and only one component remains nonnormally distributed. Let $\boldsymbol{\beta} = (\lambda_{y21}, \ldots, \lambda_{yp1})'$ be the $((p-1)\times 1)$ subvector of the free parameters of $\boldsymbol{\Lambda}_y$, and let $\mathbf{R} = (-\boldsymbol{\beta}, \mathbf{I}_{(p-1)})$, where $\mathbf{I}_{(p-1)}$ is the $((p-1)\times(p-1))$ identity matrix. Let $\mathbf{u}_t = \mathbf{R}\mathbf{y}_t$, so that $\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t$. Then, according to the transformation theorem (Krickeberg & Ziezold, 1988), the density function $f(x, y)$ of the indicator vector $(\mathbf{x}_t', \mathbf{y}_t')'$ is

$$f(x, y) = f_1(x, y_1, \mathbf{R}y) = f_2(x, \mathbf{R}y) f_3(y_1 \mid \mathbf{x}_t = x, \mathbf{u}_t = \mathbf{R}y), \tag{10}$$

where $f_1(x, y_1, u)$ is the density function of the transformed indicator vector $(\mathbf{x}_t', y_{1t}, \mathbf{u}_t')'$, $f_2(x, u)$ is the density function of $(\mathbf{x}_t', \mathbf{u}_t')'$, and $f_3(y_1|\mathbf{x}_t = x, \mathbf{u}_t = u)$ is the conditional density function of y_{1t} under the condition $\mathbf{x}_t = x$, $\mathbf{u}_t = u$. The vector $(\mathbf{x}_t', \mathbf{u}_t')'$ is normally distributed, but y_{1t} is nonnormally distributed in general. The function f_2 is a normal density, but the density f_1 and the conditional density f_3 are nonnormal in general.

The indicator variable y_{1t} is a function of $\boldsymbol{\xi}_t$, ζ_t, and ε_{1t} (see Equations 1, 2), and the vector $(\mathbf{x}_t', \mathbf{u}_t', \boldsymbol{\xi}_t', \zeta_t, \varepsilon_{1t})'$ is multivariate normally distributed. The variable $(y_1|\mathbf{x}_t = x, \mathbf{u}_t = u)$ can be formally written as a sum of its mean function and a residual variable $e_t(x, u)$:

$$(y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u) \propto E[y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u] + e_t(x, u). \tag{11}$$

In Equation (11), $E[y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u]$ is the conditional expectation of y_{1t} given $\mathbf{x}_t = x$, $\mathbf{u}_t = u$.

Calculation of Mean and Variance Function

In this subsection, Equation (11) is further analyzed, and the conditional mean $E[y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u]$ and conditional variance $\mathrm{Var}(y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u)$ of y_{1t}

are derived. A straightforward application of a common result on the conditional normal distribution (Longford, 1995) to the conditional distribution of y_{1t} gives the mean function and the following expression for the residual:

$$E[y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u] = \mathrm{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_1) + \alpha + \boldsymbol{\Gamma}\mathbf{L}_1 x + x'\mathbf{L}_1'\boldsymbol{\Omega}\mathbf{L}_1 x + \mathbf{L}_2 u \quad (12)$$

$$e_t(x, u) \propto -\mathrm{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_1) + (\boldsymbol{\Gamma} + 2x'\mathbf{L}_1'\boldsymbol{\Omega})\mathbf{z}_{1t} + \mathbf{z}_{1t}'\boldsymbol{\Omega}\mathbf{z}_{1t} + e_{0t}, \quad (13)$$

where the distribution of $(\mathbf{x}_t', \mathbf{u}_t', \mathbf{z}_{1t}', e_{0t})'$ and the parameter matrices $\mathbf{L}_1$, $\mathbf{L}_2$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ are given by

$$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{u}_t \\ \mathbf{z}_{1t} \\ e_{0t} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Lambda}_X\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}\boldsymbol{\Theta}_\varepsilon\mathbf{R}' & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix}\right), \quad (14)$$

$$\mathbf{L}_1 = \boldsymbol{\Phi}\boldsymbol{\Lambda}_x'[\boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta]^{-1}, \mathbf{L}_2 = -\theta_{\varepsilon 11}\boldsymbol{\beta}'[\mathbf{R}\boldsymbol{\Theta}_\varepsilon\mathbf{R}']^{-1}, \quad (15)$$

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= \boldsymbol{\Phi} - \boldsymbol{\Phi}\boldsymbol{\Lambda}_x'[\boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta]^{-1}\boldsymbol{\Lambda}_x\boldsymbol{\Phi}, \\ \boldsymbol{\Sigma}_2 &= \psi + \theta_{\varepsilon 11} - \theta_{\varepsilon 11}^2\boldsymbol{\beta}'[\mathbf{R}\boldsymbol{\Theta}_\varepsilon\mathbf{R}']^{-1}\boldsymbol{\beta}. \end{aligned} \quad (16)$$

By using Equations (13), (14), the conditional variance $\mathrm{Var}(y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u)$ can now be calculated. Equation (13) shows that the residual $e_t(x, u)$ does not depend on u. It is composed of the term $(\boldsymbol{\Gamma} + 2x'\mathbf{L}_1'\boldsymbol{\Omega})\mathbf{z}_{1t} + e_{0t}$, which is linear in $\mathbf{z}_{1t}$ and e_{0t}, and the quadratic term $\mathbf{z}_{1t}'\boldsymbol{\Omega}\mathbf{z}_{1t}$. The linear and the quadratic term are uncorrelated. This gives for the variance function of y_{1t},

$$\begin{aligned} \mathrm{Var}(y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u) &= \mathrm{Var}(e_t(x, u)) \\ &= \mathrm{Var}((\boldsymbol{\Gamma} + 2x'\mathbf{L}_1'\boldsymbol{\Omega})\mathbf{z}_{1t} + e_{0t}) + \mathrm{Var}(\mathbf{z}_{1t}'\boldsymbol{\Omega}\mathbf{z}_{1t}) \\ &= (\boldsymbol{\Gamma} + 2x'\mathbf{L}_1'\boldsymbol{\Omega})\boldsymbol{\Sigma}_1(\boldsymbol{\Gamma} + 2x'\mathbf{L}_1'\boldsymbol{\Omega})' \\ &\quad + \boldsymbol{\Sigma}_2 + \mathrm{Var}(\mathbf{z}_{1t}'\boldsymbol{\Omega}\mathbf{z}_{1t}). \end{aligned} \quad (17)$$

Both the conditional mean $E(y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u)$ and the conditional variance $\mathrm{Var}(y_{1t}|\mathbf{x}_t = x, \mathbf{u}_t = u)$ are polynomials of degree two of the components of x. The variance $\mathrm{Var}(\mathbf{z}_{1t}'\boldsymbol{\Omega}\mathbf{z}_{1t})$ is a constant that does not depend on x, and the exact expression is computed by using a known result on expectation values of products of normal variates (Busemeyer & Jones, 1983):

$$\mathrm{Var}(\mathbf{z}_{1t}'\boldsymbol{\Omega}\mathbf{z}_{1t}) = \sum_{i,j,k,s} \omega_{ij}\omega_{ks}(\sigma_{ij}\sigma_{ks} + \sigma_{ik}\sigma_{js} + \sigma_{is}\sigma_{jk}) - [\mathrm{tr}(\boldsymbol{\Omega}\boldsymbol{\Sigma}_1)]^2, \quad (18)$$

where the ω_{ij} are the entries of $\boldsymbol{\Omega}$ and the σ_{ij} are the entries of the covariance matrix $\boldsymbol{\Sigma}_1$.