Published in final edited form as:
Psychol Methods. 2012 June ; 17(2): 193–214. doi:10.1037/a0027539.

A Comparison of Methods for Estimating Quadratic Effects in Nonlinear Structural Equation Models

Jeffrey R. Harring, **Brandi A. Weiss**, and **Jui-Chen Hsu**
Department of Measurement, Statistics & Evaluation, University of Maryland.

Abstract

Two Monte Carlo simulations were performed to compare methods for estimating and testing hypotheses of quadratic effects in latent variable regression models. The methods considered in the current study were (a) a 2-stage moderated regression approach using latent variable scores, (b) an unconstrained product indicator approach, (c) a latent moderated structural equation method, (d) a fully Bayesian approach, and (e) marginal maximum likelihood estimation. Of the 5 estimation methods, it was found that overall the methods based on maximum likelihood estimation and the Bayesian approach performed best in terms of bias, root-mean-square error, standard error ratios, power, and Type I error control, although key differences were observed. Similarities as well as disparities among methods are highlight and general recommendations articulated. As a point of comparison, all 5 approaches were fit to a reparameterized version of the latent quadratic model to educational reading data.

Keywords

structural equation modeling; nonlinear models; quadratic; maximum likelihood; Bayesian

With the arrival of new estimation techniques, structural equation modeling (SEM) has been extended in recent years to accommodate nonlinearities between endogenous and exogenous latent variables. Positing a structural model with nonlinear terms is often necessary to adequately account for the complexities often underlying real-world phenomena found in the behavioral and social sciences (Moosbrugger, Schermelleh-Engel, Kelava, & Klein, 2009). Of the numerous nonlinear relations that are possible (see, e.g., Klein & Muthén, 2007; Wall, 2009), a simple quadratic function with a single exogenous variable and its squared manifestation remains an attractive alternative among practitioners due to its effectiveness in adequately summarizing many experimental processes within the observed range of the data (Cudeck & du Toit, 2002).

Estimation of regression models with quadratic effects composed of continuous latent variables has garnered a great deal of attention from methodologists since the seminal article by Kenny and Judd (1984). Many estimation alternatives exist within the more general polynomial specification, yet the vast majority of methodological studies that investigate the relative effectiveness of different methods have focused on structural models with multiplicative terms representing interactions among exogenous latent variables.

In contrast, estimation methods for models with only second-degree polynomial terms have been examined with less frequency (Marsh, Wen, & Hau, 2006). The methodological

Correspondence concerning this article should be addressed to Jeffrey R. Harring, Department of Measurement, Statistics & Evaluation, 1230 Benjamin Building, University of Maryland, College Park, MD 20742-1115. harring@umd.edu.

approaches that currently exist to test for the necessity of quadratic effects generally fall into distinct categories: (a) latent variable score approaches (e.g., Schumacker, 2002; a method of moments two-step estimation procedure; Wall & Amemiya, 2003); (b) product indicator approaches (e.g., Bollen, 1996; Hayduk, 1987; Jaccard & Wan, 1995; Ping, 1996), which may be further categorized as fully constrained (Algina & Moulder, 2001; Jöreskog & Yang, 1996), partially constrained (generalized appended product indicator; Wall & Amemiya, 2001), and unconstrained (Marsh, Wen, & Hau, 2004); (c) maximum likelihood estimation (Wall, 2009), which may be further categorized as marginal maximum likelihood (Cu-deck, Harring, & du Toit, 2009; Klein & Moosbrugger, 2000) and an approximate maximum likelihood estimation scheme (Klein & Muthén, 2007); and finally (d) Bayesian estimation (Lee, 2007; Lee, Song, & Poon, 2004; Wall, 2009).

Although some simulation studies have compared a small subset of these methods for estimating quadratic effects, the primary objective of this study was to do a more comprehensive investigation of a broader set of methods. Moreover, we illustrate the practical implications of using the different methods under investigation by applying them to data from a reading comprehension assessment.

We have divided this article into the following main sections. In the next section we review a nonlinear structural equation model that specifies a quadratic relation between endogenous and exogenous latent variables. We then review the five estimation methods used in the comparisons vis-à-vis Monte Carlo simulation. We then describe the factors thought to impact the performance of these estimation methods and summarize what other simulation studies have found. We outline the first of two simulations that examine the performance of the five estimation methods in terms of parameter bias, root-mean-square error (RMSE), and standard error ratio. In a subsequent section, we describe the second simulation designed to compare the estimation methods in terms of their power to detect the presence of the quadratic effect as well as their ability to control Type I error rate. We draw conclusions based on the results of the simulation and provide recommendations as to which methods are preferred across examined simulation conditions. Lastly, we fit reading data using all five methods from a sample of fourth-grade students.

Nonlinear Structural Equation Models

SEM, initially conceived as an analytic method for modeling linear relations among latent variables, can be extended, at least conceptually, to include nonlinear effects in a straightforward manner. Like its linear counterpart (e.g., Wall, 2009), a nonlinear structural equation model defines two regression models that play distinctive roles in formulating various substantive problems: (a) a measurement model that defines the relation of the latent independent (exogenous) and dependent (endogenous) variables to their observed variable indicators, and (b) the structural model, which delineates the effects of the exogenous latent variables on the endogenous variable.

Figure 1 displays a prototypical nonlinear structural equation model with six observed variables and two latent variables, one with a linear and a quadratic term and one with only a linear term; the structure of this model is the one that we use for the simulation study in this article.

Specifically, on the top and right-hand side, shown in squares, are six manifest variables that can be collected in a vector $\mathbf{Z}_i = (Z_i, \ldots, Z_{6i})'$ whose scores are observed for the ith individual. In the center, shown in circles, are the three unobserved (i.e., latent) variables, f_1, f_2, and f_2^2, that arise from the two core latent variables, f_1 and f_2, that can be collected in a vector $\mathbf{f}_i = (f_{1i}, f_{2i})'$. The full nonlinear structural equation model can then be specified as follows:

$$\mathbf{Z}_i=\tau+\Lambda\mathbf{f}_i+\mathbf{e}_i, \quad (1)$$

$$f_{1i}=\gamma_0+\gamma_1 f_{2i}+\gamma_2 f_{2i}^2+d_i, \quad (2)$$

where, in Equation 1, $\boldsymbol{\tau}$ is a ($p \times 1$) vector of intercepts, $\boldsymbol{\Lambda}$ is the ($p \times q$) matrix of factor loadings relating each latent variable in $\mathbf{f}_i$ to its measured variable indicators in $\mathbf{Z}_i$, and $\mathbf{e}_i$ represents the ($p \times 1$) vector of unique factors independent of $\mathbf{f}_i$ with $E(\mathbf{e}_i) = \mathbf{0}$ and $\mathrm{Var}(\mathbf{e}_i) = \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is diagonal. The regression coefficients $\boldsymbol{\gamma}_0$, $\boldsymbol{\gamma}_1$, and $\boldsymbol{\gamma}_2$ are the intercept and direct path estimates of the first-order latent variable f_{2i} and quadratic term f_{2i}^2, respectively. The sign of $\boldsymbol{\gamma}_2$ indicates whether the endogenous–exogenous curvilinear relation is concave (negative implies curving downward) or convex (positive implies curving upward). Figure 2 illustrates a scenario with only a quadratic component, in which change in the latent criterion f_1 is increasing across initial values of the latent predictor, f_2; attains some maximum level of performance or proficiency; and finally declines as the values of the latent predictor continue to increase. A convex relation would show initial decline of the criterion to a minimum value at which point the criterion values rebound and begin to increase at larger values of the latent predictor.

Lastly, a common assumption is that the exogenous factor f_{2i} comes from a normal distribution having mean $\boldsymbol{\kappa}_1$ and variance $\boldsymbol{\phi}_1$. For the structural model in Equation 2, it is assumed the residuals d_i have $E(d_i) = 0$ and $\mathrm{Var}(d_i) = \sigma_d^2$, and are independent of f_{2i} as well as $\mathbf{e}_i$.

For the purpose of our simulation studies, we use a model with two latent variables with three observed indicators each for both endogenous and exogenous latent variables. For identification purposes and to set the scale of the latent variables, we chose to set a single factor loading to 1 and corresponding intercept to 0 for both f_{1i} and f_{2i}. The measurement model for six observed variables, x_{mi} and y_{mi} for $m = 1, 2, 3$, is

$$\underbrace{\begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ x_{1i} \\ x_{2i} \\ x_{3i} \end{pmatrix}}_{\mathbf{Z}_i} = \underbrace{\begin{pmatrix} 0 \\ \tau_{y_2} \\ \tau_{y_3} \\ 0 \\ \tau_{x_2} \\ \tau_{x_3} \end{pmatrix}}_{\tau} + \underbrace{\begin{pmatrix} 1 & 0 \\ \lambda_{y_2} & 0 \\ \lambda_{y_3} & 0 \\ 0 & 1 \\ 0 & \lambda_{x_2} \\ 0 & \lambda_{x_3} \end{pmatrix}}_{\Lambda} \underbrace{\begin{pmatrix} f_{1i} \\ f_{2i} \end{pmatrix}}_{\mathbf{f}_i} + \underbrace{\begin{pmatrix} e_{y_{1i}} \\ e_{y_{2i}} \\ e_{y_{3i}} \\ e_{x_{1i}} \\ e_{x_{2i}} \\ e_{x_{3i}} \end{pmatrix}}_{\mathbf{e}_i}. \quad (3)$$

As discussed at the outset of this article, the parameters in such a nonlinear structural equation model can be estimated with a variety of methods, which we describe in the following section. Note that we are specifically interested in reliably testing whether quadratic latent variable effects, such as those represented by f_2^2 in Figure 1 above, are statistically necessary.

Estimation Methods

Latent Variable Scores

One traditional method that has been used to test for nonlinear relations between latent variables involves a two-step process based on latent variable (i.e., factor) scores. Latent variable scores (referred to in the simulations as LVS) represent estimates of individuals' scores on an underlying latent factor. In the first step, factor scores for the endogenous and exogenous latent variables f_1 and f_2 are computed. In the second step, for each individual's estimated latent variable score, $\hat{f}_{2i}$ is squared to form the latent variable score for the quadratic effect, $\widehat{f_{2i}^2}$, and finally all latent variable scores are submitted to a multiple regression analysis (Schumacker, 2002). Procedures for testing quadratic effects through multiple regression analyses can be found in Cohen, Cohen, West, and Aiken (2003).

Due to the straightforward manner in which latent variable scores are computed, coupled with the simplicity of least-squares estimation in subsequent regression analyses, use of factor scores for structural models with quadratic effects seems like a sensible approach. However, potential disadvantages exist that might impede their widespread use for these types of methods. As Bollen (1989) pointed out, because latent variable scores represent only estimates of individuals' true scores on the underlying latent factors, their scores may still contain measurement error. As a consequence, many methodological researchers contend that factor score regression will produce biased and/or inconsistent estimates of the structural model parameters (e.g., Bartholomew, 1987; Lastovicka & Thamodaran, 1991). These drawbacks notwithstanding, using latent variable scores does reduce measurement error, and consequently, this approach for estimating quadratic effects between latent variables is still advantageous over using observed scores within multiple regression analyses.

Secondly, because the number of latent variables and errors of measurement is greater than the number of observed variables, there exist an infinite number of potential solutions for the specific factor scores. This is known as the *factor score indeterminacy problem* (e.g., Mulaik, 2009). As a consequence, numerous estimation methods have been put forth to compute these factor scores. Several studies have been conducted to compare the various methods of estimating latent variable scores (e.g., Gorsuch, 1974; Lastovicka & Thamodaran, 1991; Mulaik, 1972). Lastovicka and Thamodaran (1991) conducted a parameter recovery simulation study comparing the least squares regression method, Bartlett's method, Anderson and Rubin's method, and another method proposed by Thurstone (as cited in Lastovicka & Thamodaran, 1991). In addition, Lastovicka and Thamodaran used an ad hoc procedure using a factor score extension proposed by Dwyer (as cited in Lastovicka & Thamodaran, 1991), as well as the commonly used method of simply summing an individual's responses on all variables (assuming they are coded in the same direction). Similar results were found among the six estimation methods. The Dwyer extension method resulted in the closest recovery of multiple R and had the lowest standard error of measurement associated with the regression beta weights. The Anderson and Rubin (1956) method, a revision of Bartlett's method, resulted in the most accurate recovery of the beta weights and had a comparable standard error of measurement associated with the regression beta weights to that of the Dwyer extension method. The other three methods were somewhat comparable.

In this study we use a modification of Anderson and Rubin's (1956) method as described in Jöreskog (2000). The procedure has the advantage of producing factor scores that have the same means and covariance matrix as the latent variables themselves. Jöreskog showed that when the mean vector fits perfectly—which is always the case when $\boldsymbol{\tau}$ in Equation 2 is unconstrained—the sample factor score estimates will be unbiased. Consequently, the estimated latent variable scores used as observed variables will get the same parameter estimates for coefficients in Equation 3. This is the default procedure in LISREL 8.80 (Jöreskog & Sörbom, 2006), the program we used to carry out the estimation of the latent variable score approach.

Unconstrained Product Indicator Approach

For the proposed quadratic model, the unconstrained method requires researchers to use all paired products of observed variables as indicators of the latent quadratic term, f_2^2 (e.g., $x_2 = x_1 \cdot x_1$). This is what Marsh et al. (2004) referred to as a matched-pairs strategy. That is, for the measurement model in Equation 3, the three products of the observed variables x_{mi} that would need to be constructed in the data set are

$$x_{1i}^2 = x_{1i} \cdot x_{1i} \quad x_{2i}^2 = x_{2i} \cdot x_{2i} \quad x_{3i}^2 = x_{3i} \cdot x_{3i}. \quad (4)$$

Note that the observed variables x_{mi} are not mean centered before constructing their products. Although multicollinearity will certainly exist between the first-order terms and their corresponding derived quadratic terms (Aiken & West, 1991), Wall (2009) demonstrated the equivalence of the regression coefficient of the highest order term (in this case the quadratic term) and its standard error when the observed variables were mean centered and when they were not. Because our focus is on reliably estimating the coefficient of the quadratic term, we proceed with not mean centering the indicators of the first-order predictor, f_2.

The modified measurement model incorporating the newly formed product indicators of f_2^2 is

$$\underset{\tilde{\mathbf{Z}}_i}{\begin{pmatrix} y_{1i} \\ y_{2i} \\ y_{3i} \\ x_{1i} \\ x_{2i} \\ x_{3i} \\ x_{1i}^2 \\ x_{2i}^2 \\ x_{3i}^2 \end{pmatrix}} = \underset{\tilde{\tau}}{\begin{pmatrix} 0 \\ \tau_{y_2} \\ \tau_{y_3} \\ 0 \\ \tau_{x_2} \\ \tau_{x_3} \\ 0 \\ \tau_{x_2^2} \\ \tau_{x_3^2} \end{pmatrix}} + \underset{\tilde{\Lambda}}{\begin{pmatrix} 1 & 0 & 0 \\ \lambda_{y_2} & 0 & 0 \\ \lambda_{y_3} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{y_2} & 0 \\ 0 & \lambda_{y_3} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{x_2}^2 \\ 0 & 0 & \lambda_{x_3}^2 \end{pmatrix}} \underset{\tilde{\mathbf{f}}_i}{\begin{pmatrix} f_{1i} \\ f_{2i} \\ f_{2i}^2 \end{pmatrix}} + \underset{\tilde{\mathbf{e}}_i}{\begin{pmatrix} e_{y_{1i}} \\ e_{y_{2i}} \\ e_{y_{3i}} \\ e_{4i} \\ e_{5i} \\ e_{6i} \\ e_{7i} \\ e_{8i} \\ e_{9i} \end{pmatrix}} \quad (5)$$

Note the difference in notation in Equation 5 compared with Equation 3. Under the assumption that the exogenous latent variable is normally distributed, product indicator approaches were proposed that demonstrated how parameters of the model corresponding to the nonlinear latent variable could be expressed as nonlinear functions of other model parameters (see, e.g., Algina & Moulder, 2001; Jöreskog & Yang, 1996, for constrained estimation approaches; Wall & Amemiya, 2003, for a partially constrained approach). Though admittedly ad hoc in nature, these methods were introduced primarily to make maximum likelihood estimation of nonlinear structural models feasible. Aside from the regression coefficients for the structural model, other model parameters to be estimated include observed variable intercepts, factor loadings, unique variable variances and corresponding covariances (see, e.g., Marsh et al., 2006, for more detail), and factor means, variance, and covariances.

In this study, we follow the unconstrained approach (referred to in the simulations as UNC) introduced by Marsh et al. (2004) in which error variances, intercepts, factor loadings, and factor means associated with the nonlinear factor are allowed to be freely estimated, thereby eliminating the potential for incorrectly mis-specifying these nonlinear relations under real data-analytic modeling conditions (Marsh et al., 2006). Of the product indicator approaches, the unconstrained approach has been shown to be more robust than its competitors under nonnormal distributional conditions and can be implemented in current SEM software programs in a straightforward manner. For the purposes of this study, we used LISREL 8.80 (Jöreskog & Sörbom, 2006) to carry out the estimation of the unconstrained product indicator approach.

Approaches Utilizing the Likelihood

Some methods to estimating nonlinear structural models such as Equation 2 require working directly with the likelihood based on the joint distribution of observed data $\mathbf{Z}_i$ and latent variables $\mathbf{f}_i$. Maximum likelihood estimation of nonlinear structural equation models of this type has been considered most extensively in an important series of articles by Lee and his colleagues (Lee, Song, & Lee, 2003; Lee et al., 2004; Lee & Zhu, 2002).

Given the measurement model and nonlinear quadratic structural model in Equation 2 and Equation 3, respectively, for the ith individual, this likelihood function associated with a sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ can be written as in Equation 3. Following the notation in Wall (2009), let $\boldsymbol{\theta}_m$ represent the measurement model parameters (i.e., parameters in Λ, Θ) and $\boldsymbol{\theta}_s$ denote the nonlinear structural parameters (i.e., $\boldsymbol{\gamma}_0$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, σ_d^2). Note that $\theta=\left(\theta_m', \theta_s'\right)'$. The likelihood can then be written from the joint distribution of the response vector $\mathbf{Z}_1$ and the random latent variable $\mathbf{f}_i$ conditional on the parameter vector $\boldsymbol{\theta}$ as follows:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} p(\mathbf{Z}_i|\theta) \\ &= \prod_{i=1}^{n} \int p(\mathbf{Z}_i, \mathbf{f}_i|\theta)\, \partial \mathbf{f}_i \\ &= \prod_{i=1}^{n} \int p(\mathbf{Z}_i|\mathbf{f}_i, \theta)\, p(\mathbf{f}_i|\theta)\, \partial \mathbf{f}_i \\ &= \prod_{i=1}^{n} \int p(\mathbf{Z}_i|f_{1i}, f_{2i}, \theta_m)\, p(f_{1i}|f_{2i}, \theta_s)\, p(f_{2i})\, \partial f_{2i}. \end{aligned} \quad (6)$$

When either the measurement or structural model has nonlinear relations in the conditional means, or the underlying exogenous factor and/or error terms are not normally distributed, the integral in Equation 6 is analytically intractable.

Moreover, the likelihood in Equation 6 is difficult to estimate due to the computation of the potentially high-dimension integral over the latent quantities. This is especially true for nonlinear structural models that have multiple nonlinear effects (Moosbrugger et al., 2009). However, modern statistical estimation approaches and optimization schemes have made these computationally intensive problems more accessible to practitioners. Though several methods have been espoused (see, e.g., Skrondal & Rabe-Hesketh, 2004), in the context of estimating quadratic structural models, two general estimation schemes are (a) to approximate the integral in Equation 6 directly or (b) to circumvent the integration by implementing the expectation–maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977). The EM algorithm is central to the latent moderated structural equation approach described below.

Latent Moderated Structural Equations

Rather than directly approximate the integral in Equation 6, Klein and Moosbrugger (2000) proposed the latent moderated structural equation method (referred to in the simulations as LMS), which does not require the creation of indicators for the quadratic latent variable (a condition for the unconstrained approach) and recognizes the nonnormal distribution of the quadratic latent variable. It does, however, assume that the exogenous latent variable and structural equation errors are normally distributed. The EM algorithm is used to compute maximum likelihood estimates of the parameters.

Following Klein and Moosbrugger (2000), the form of a general interaction model with squared terms of two exogenous continuous latent variables as well as their cross product is

$$\eta=\alpha+\Gamma\xi+\xi^{'}\Omega\xi+\zeta,$$

where $\boldsymbol{\eta}$ is an endogenous latent variable, $\boldsymbol{\alpha}$ is an intercept term, $\boldsymbol{\Gamma}$ is a $(1 \times k)$ parameter vector associated with the linear latent exogenous variables, $\boldsymbol{\xi}$ is a $(k \times 1)$ vector of latent exogenous variables, $\boldsymbol{\Omega}$ is a $(k \times k)$ matrix of parameters that correspond to nonlinear factors of $\boldsymbol{\xi}'\boldsymbol{\xi}$, and $\boldsymbol{\zeta}$ is the error term. In the case of the quadratic structural model in Equation 2, $\boldsymbol{\xi}$ and $\boldsymbol{\Omega}$ are both scalars,

$$\xi=f_2 \qquad \Omega=\gamma_2.$$

The quadratic model specified in Equation 1 and Equation 2 comprises a vector of indicators with six elements $\mathbf{Z} = (\mathbf{x}', \mathbf{y}') = (x_1, x_2, x_3, y_1, y_2, y_3)'$ and can be represented as a finite mixture of multivariate normal distributions. The indicator vector $\mathbf{x}$ is assumed to be normally distributed, whereas indicator vector $\mathbf{y}$ is not assumed to be normally distributed because the quadratic term f_2^2 is in the structural model. Linear and nonlinear effects are separated and decomposed into independent random $\mathbf{z}$ variables via the Cholesky decomposition of the covariance matrix of the manifest indicators, $\boldsymbol{\Sigma}$. The vector $\mathbf{z}$ is made up of vectors $\mathbf{z}_1$ and $\mathbf{z}_2$, which represent the linear and nonlinear effects, respectively. From this, a continuous mixture of normal densities with $\mathbf{z}_1$ as the mixing vector can be derived. A subsequent partitioned mean vector and covariance matrix can then be obtained. If the quadratic effect exists, and thus $\boldsymbol{\gamma}_2$ differs from 0, then the integral of the mixture cannot be solved analytically. In this case it is approximated by Gaussian–Hermite quadrature formulas of numerical integration (Pinheiro & Bates, 1995) within an EM algorithmic scheme from which maximum likelihood estimates of the mixture probabilities and mixture components are obtained (Klein & Moosbrugger, 2000).

The LMS method, which is currently implemented in Mplus (Muthén & Muthén, 2007), can be used to fit the quadratic model to data. In this study, we used Mplus 6.1 to carry out estimation of the quadratic structural model using this method.

Marginal Maximum Likelihood

Several methods have been proposed that attempt to handle the multidimensional integration in Equation 6 directly. These methods—Laplace's approximation, Gaussian–Hermite quadrature, adaptive Gaussian quadrature, and importance sampling—have been coined *exact methods* in the literature (Pinheiro & Bates, 1995), but all are trying to approximate the area of the distributions of the latent exogenous variables. The family of so-called exact likelihood methods maximizes the likelihood function directly using deterministic or stochastic approximation to handle the integral. From this class, Gaussian–Hermite quadrature (nonadaptive) was used in the simulation to facilitate marginal maximum likelihood estimation (referred to in the simulations as MML). Gaussian–Hermite quadrature approximates each integral by a weighted average of the integrand evaluated at specific points over a grid centered at 0. The more grid points that are evaluated, the more precise the approximation of the integral will be (Skrondal & Rabe-Hesketh, 2004). With this rule, an integral over a function of the type $v(t) = g(t)\exp(-t^2)$ is approximated by the sum

$$\int_t g(t)\exp\left(-t^2\right)dt \approx \sum_{k=1}^{Q} w_k g(x_k),$$

where w_k and x_k are the weights and nodes of the Hermite polynomial of degree Q. It can be shown (Stroud & Secrest, 1966, Section 1) that the approximation is exact if $g(t)$ is of degree $2Q - 1$. It can also be shown (Krommer & Ueberhuber, 1994, Section 4.2.6) that the approximation converges to the true integral as the number of quadrature points, Q, increases.

The primary reason to examine just the nonadaptive Gaussian– Hermite quadrature method is based on the rate of convergence. Through preliminary investigations, we found that convergence rates for integration methods (i.e., adaptive Gaussian quadrature, importance sampling, Laplace's approximation) in which the nodes are rescaled or estimated, relied too heavily on good starting values of the parameters—especially the variance and covariance parameters of the exogenous factor and unique factors—to make the simulation feasible. We found that nonadaptive quadrature was less sensitive to starting values, but required more quadrature points to approximate the integral with satisfactory precision.

In the current study, we used a nonadaptive quadrature with $Q = 30$ quadrature points to approximate the integral followed by a quasi-Newton optimization scheme. The number of points was chosen after some initial investigation of the performance of the method to converge to a reasonable solution. Moreover, because the LMS method also requires an integration scheme using Gaussian–Hermite quadrature within the EM algorithm, setting the number of points to a universal value could make the results between these methods more comparable. For the simulation study, Gaussian–Hermite quadrature was implemented in SAS PROC NLMIXED (Wolfinger, 1999).

Bayesian Estimation

In the MML method described above, the elements of θ in Equation 6 were considered fixed parameters in the population, and $\mathbf{f}_i$ were random latent variables coming from a distribution, $p(\mathbf{f}_i|\theta)$. The random latent variables in $\mathbf{f}_i$ were not estimated explicitly in the maximum likelihood procedure but marginalized out by integration. In a secondary step, predicted factor scores could be obtained for each individual. In a Bayesian approach (referred to in the simulation as BAY) there is no distinction between random latent variables and model parameters, as all are considered random quantities (Gelman, Carlin, Stern, & Rubin, 2004). That is, in the Bayesian approach, vector $\mathbf{f}_i$ contains parameters, and the parameter vector θ is assigned a distribution called a prior distribution, $p(\theta)$ (Lee, 2007).

A fully Bayesian approach to estimating nonlinear structural equation models relies on an application of Bayes' theorem, which states

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta) \cdot p(\theta)}{p(\mathbf{x})},$$

where $p(\mathbf{x} \mid \theta)$ is the conditional distribution of the data $\mathbf{x}$ given model parameters θ, $p(\theta)$ is the prior probability distribution of the parameters, and $p(\mathbf{x}) = \int p(\mathbf{x}|\theta)\partial\theta$ represents a normalizing constant to make $p(\theta \mid \mathbf{x})$ a proper density function. Direct computation of the integral in Equation 6 is bypassed by noticing that the posterior distribution is proportional to the product of the likelihood function and prior distributions:

$$p(\theta, \mathbf{f}_i|\mathbf{Z}_i) \propto \prod_{i=1}^{n} p(\mathbf{Z}_i|f_{1i}, f_{2i};\theta_m)\, p(f_{1i}|f_{2i};\theta_s)\, p(f_{2i})\, p(\theta). \quad (7)$$

Draws from the posterior distribution can be obtained via Markov chain Monte Carlo (MCMC) simulation methods. MCMC methods are useful in sampling from the multivariate densities, which are not easy to sample from, usually by breaking these densities down into more tractable univariate or multivariate densities (Lynch, 2007). That is, MCMC methods sidestep sampling directly from the joint posterior distribution by sampling from the full conditional posteriors of each parameter conditioned on the data and the most recent value of all other parameters (see, e.g., Lee, Song, & Tang, 2007; Wall, 2009).

As is true with any type of model for multivariate data, the key to specifying a nonlinear structural equation model within any MCMC framework is properly setting up the prior distributions on the model parameters. This section describes conjugate priors (of the same family as the likelihood) for each of the parameters. Note, however, that priors need not be from the same distributional family as the likelihood. Lynch (2007) observed that a feature of the Bayesian framework that is particularly attractive is that it provides a natural mechanism for incorporating known constraints on values of model parameters and other subject matter knowledge through the specification of suitable proper prior distributions.

Recall that in Equation 6 that $\boldsymbol{\theta}$ comprises all model measurement and structural parameters —$\boldsymbol{\tau}$, Λ, Θ, $\boldsymbol{\gamma}$, $\boldsymbol{\kappa}_1$, $\boldsymbol{\phi}$, σ_d^2—from Equation 2 and Equation 3. The conjugate priors for the intercepts, factor loadings, and unique factor variances from Equation 1 are for $k = 1, \ldots, p$:

$$\tau_k \sim N(\mu_{0k}, \varphi_{0k}), \theta_{ek}^{-1} \sim \text{Gamma}(\alpha_{0ek}, \beta_{0ek}), \lambda_k \sim N\left(\lambda_{0k}, \sigma_{0k}^2\right),$$

where $\boldsymbol{\mu}_{0k}$, $\boldsymbol{\varphi}_{0k}$, $\boldsymbol{\alpha}_{0ek}$, $\boldsymbol{\beta}_{0ek}$, $\boldsymbol{\lambda}_{0k}$, and α_{0k}^2 are hyperparameters. The conjugate priors for the regression coefficients, regression residuals, and exogenous latent variable are

$$\gamma_{\text{j}} \sim N\left(\mu_{0\gamma\text{j}}, \varphi_{0\gamma j}\right), f_{2i} \sim N\left(\mu_{0f}, \varphi_{0f}\right), d \sim N(\mu_{0d}, \varphi_{0d}), \mu_{0f} \sim N\left(\mu_f, \varphi_f\right) \varphi_{0f}^{-1} \sim \text{Gamma}\left(\alpha_{0\varphi}, \beta_{0\varphi}\right).$$

Again, $\boldsymbol{\mu}_{0\gamma j}$, $\boldsymbol{\varphi}_{0\gamma j}$, $\boldsymbol{\mu}_{0f}$, $\boldsymbol{\varphi}_{0\varphi}$, $\boldsymbol{\mu}_{0d}$, $\boldsymbol{\varphi}_{0d}$, $\boldsymbol{\mu}_f$, $\boldsymbol{\varphi}_{0f}$, $\boldsymbol{\alpha}_{0\varphi}$, and $\boldsymbol{\beta}_{0\varphi}$ are hyperparameters. If hyperparameters in the conjugate prior distributions are not known, then they may be treated as unknown parameters and thus have their own hyperparameters. Song and Lee (2001) pointed out conjugate prior distributions with known values work well for many SEM applications, and assigning specific values for hyperparameters would indicate available prior knowledge. In general, informative priors are assigned small variance in the corresponding distributions; otherwise large variance should be selected. In the latter case, these prior distributions are known as diffuse or noninformative. Following Wall (2009), we used the following noninformative priors for the simulation: $\boldsymbol{\tau}_k \sim N(0, 0.0001)$; $\boldsymbol{\lambda}_k \sim N(0, 0.0001)$; $\theta_{ek}^{-1} \sim \text{Gamma}(0.1, 0.1)$; $\boldsymbol{\gamma}_j \sim N(0, 0.0001)$; $\boldsymbol{\mu}_{0f} \sim N(0, 0.0001)$; $\varphi_{0f}^{-1} \sim \text{Gamma}(0.1, 0.1)$; $d \sim N(0, \boldsymbol{\varphi}_{0d}$ and $\varphi_{0d}^{-1} \sim \text{Gamma}(0.1, 0.1)$.

Although the flexibility of Bayesian estimation is in part due to the ability to specify prior distributions, this is also one of the main criticisms of the estimation framework (Paap, 2002). Estimates can sometimes be quite sensitive to the choice of priors and hyperparameters, even when conjugate priors are used. Furthermore, although large sample sizes can sometimes negate the impact of priors, there are many instances where the specification of prior distributions can have a large impact on estimates. Thus, an initial exploration as to the sensitivity of the expected results based on different priors for the current simulation was completed. Each of the priors was set for each parameter and tested to make sure the distributions did not overwhelm the data.

Bayesian estimation for the current study was executed with WinBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2002). To begin running the full-scale simulation with MCMC estimation, we investigated an adequate number of burn-in and estimation cycles with a reasonable number of simulated draws from the posterior distribution. The number of burn-in iterations can be determined by plots of the simulated sequences generated from different starting values of the parameter under investigation. At convergence, parallel sequences generated with different starting values should "mix" together well. Examples of sequences for which convergence looks reasonable and sequences that have not reached convergence are presented in Figure 3.

Based on different starting values, the Gelman–Rubin convergence statistic (as modified by Brooks and Gelman, 1998) can also be computed to help assess convergence. The Gelman–Rubin convergence statistic as implemented in WinBUGS finds the ratio of between-chain variability to within-chain variability. Due to how it is scaled, values of the statistic close to 1 represent convergence. In preliminary analyses, we examined convergence plots and descriptive statistics, statistics of posterior parameters, history of cycles, and densities of the posterior data under both severe and nonsevere conditions to select the number of burn-in cycles and adequate estimate cycles per replication in the final simulation analyses. Consequently, we took a burn-in phase of 3,000 iterations and collected 2,000 observations after convergence for obtaining the Bayesian point estimates of the parameters and their corresponding standard error estimates.

Simulation Study 1

The design of the first simulation study was a 5 (sample size levels) × 3 (observed variable reliability levels) × 3 (latent variable distribution levels) × 2 (observed variable distribution levels) × 5 (methods of estimation) completely crossed factorial design resulting in 450 possible combinations. Of primary concern were those factors that have been shown to impact methods of estimation in past methodological studies. Levels of the manipulated factors were chosen to (a) correspond to realistic analytic environments found in practice and (b) push the various estimation methods to the breaking point. As a reviewer pointed out, this latter motive might produce more divergent performance of the various methods and thus provide clearer guidance for practitioners, who likely will use these methods under diverse data conditions. The independent manipulated factors and rationale of choices for the levels of those factors are described in detail in the following subsections.

Indicator Reliability

One advantage of structural equation models over traditional regression models for observed variables is that, by definition, latent variables are measured without random error. Thus, for fully latent estimation methods in which estimation of the measurement model and structural model is performed simultaneously, unreliability of the indicators may pose fewer problems. However, for nonfully latent estimation methods, such as multiple regression analyses with predicted factor scores, reliability of the indicators has been known to impact estimates of structural coefficients (Dimitruk, Schermelleh-Engel, Kelava, & Moosbrugger, 2007). In other simulation studies (see, e.g., Weiss, 2010), indicator reliability has been shown to affect power to detect the interaction effect in a latent variable interaction model. For these reasons it will be a factor manipulated in the design. Commensurate with past simulation studies (see, e.g., Algina & Moulder, 2001; Jaccard & Wan, 1995), reliabilities of observed variables were simulated to be .65 and .85, representing fair and good values, respectively. A more extreme reliability level (indicating poor reliability) of .45 was also examined. We chose the indicator reliabilities to be equal across the indicator variables. This decision was made in part by the practical realities of executing the simulation, although we do recognize that in real-world applications the reliabilities for each observed variable indicator could very well be distinct.

Sample Size

We used five sample sizes ($n = 50$, $n = 100$, $n = 250$, $n = 500$, and $n = 1{,}000$) reflecting various degrees of estimation precision for measurement and structural parameters, respectively. Three of the sample sizes—$n = 100$, $n = 250$, $n = 500$—reflect those used in past simulation designs that investigated nonlinear effects (see, e.g., Klein & Muthén, 2007; Marsh et al., 2004; Moulder & Algina, 2002), whereas the other two represent more extreme conditions. A sample size of 50 was chosen to reflect a very small, perhaps even unrealistic, sample size; $n = 1{,}000$ was chosen on the upper end to reflect a very large sample size. From preliminary investigations across both high-risk and low-risk condition combinations, it was determined that for even more extreme sample sizes (e.g., $n = 2{,}000$), the increase in estimation precision relative to that observed when $n = 1{,}000$ was negligible.

Nonnormal Distributions

Several estimation methods, including four under investigation in this study, rely on the assumption that either the observed indicator variables follow a multivariate normal distribution or the exogenous latent predictor is normally distributed, or both. Maximum likelihood estimation tends to be fairly robust to violations of normality in terms of bias of parameter estimates (e.g., Boomsma, 1983). However, results from past simulation studies do suggest that nonnormality leads to serious underestimation of standard errors and biased chi-square values (Curran, West, & Finch, 1996; Hu, Bentler, & Kano, 1992; Jöreskog & Yang, 1996). Of importance to the current study, underestimation of standard errors is an indication that Type I error rates may be much greater than the nominal level. For testing the quadratic effect, false rejection of the null hypothesis could occur with greater frequency than what the nominal level would suggest.

Even in the best circumstances within a maximum likelihood estimation framework, when the observed variables follow a multivariate normal distribution and the exogenous latent predictor is normally distributed, the quadratic term f_2^2 will necessarily not be normally distributed. Consequently, the endogenous dependent variable, f_1, will not be normally distributed (Jöreskog & Yang, 1996; Klein & Moosbrugger, 2000). Marsh et al. (2004) suggested examining the effects of nonnormality in both observed and latent variables, although more recent studies (e.g., Klein & Muthén, 2007) have noted that if the exogenous latent variable is nonnormal, then indicators formed from them would be nonnormal as well. However, for completeness, we examined nonnormality for both observed and latent variables.

It should be pointed out that previous studies have only investigated the impact of mild nonnormality on estimating nonlinear effects (Klein & Moosbrugger, 2000; Klein & Muthén, 2007; Marsh et al., 2004; Wall & Amemiya, 2001). Specifically, for the nonnormal conditions within these studies, data were generated for exogenous latent variables from distributions with skew ranging from –2.0 to 1.5 and kurtosis ranging from –1.5 to 6.0. Kline (2011) suggested that extreme skew is defined by skew values greater than an absolute value of 3.0 and extreme kurtosis is defined by absolute kurtosis values ranging from 8.0 to over 20.0. He further suggested that kurtosis values greater than the absolute value of 20 may indicate serious problems with nonnormality. On the basis of these heuristic values for skew and kurtosis, the skew and kurtosis values for previous studies are categorized to be fairly mild.

Fleishman's (1978) polynomial transformation procedure was used to generate combined levels of skewness and kurtosis for the nonnormal distributions of the exogenous latent variable. Three levels were assessed: (a) skewness = 0, kurtosis = 0 (normal condition); (b) skewness = 2, kurtosis = 7 (moderate nonnormality); and (c) skewness = 3 and kurtosis = 20 (severe nonnormality). The same procedure was used to generate two levels of observed variable nonnormality: (a) skewness = 0, kurtosis = 0 (normal condition), and (b) skewness = 2, kurtosis = 6 (moderate nonnormality). Descriptive statistics were computed for a number of data sets to ensure that the population levels of skew and kurtosis were obtained. In all there are 3 × 2 = 6 fully crossed distributional conditions. A summary of all manipulated conditions can be found in Table 1.

Nonmanipulated Factors

In past methodological studies (see e.g., Jaccard & Wan, 1995; Kelava, Moosbrugger, Dimitruk, & Schermelleh-Engel, 2008; Marsh et al., 2004), the effects of multicollinearity have been investigated primarily for latent interaction models where the correlation between exogenous latent first-order predictors could be manipulated. For the quadratic model, multicollinearity does exist between the first-order latent predictor and its squared manifestation because the latter is derived from the former. In practice, researchers might center the first-order exogenous latent predictor before squaring that term to eliminate what Cohen et al. (2003) have coined *nonessential multicollinearity*. This is the correlation due to the scaling of the variables. Because the regression coefficient and its standard error of the highest order term in the model (in this case the quadratic term) are mathematically equivalent (Wall, 2009), differences between centered and uncentered models that address issues inherent with multicollinearity will not be examined further.

Data Generation

The first simulation examines parameter recovery and accuracy of the regression coefficients and their corresponding standard errors under different (a) sample sizes, (b) reliability of the indicators, (c) various distributions of the indicators as well as the latent variables, and (d) estimation methods.

All simulated variables were derived based on the population model depicted in Figure 1 following structural and measurement models in Equation 2 and Equation 3, respectively. The following population values were selected for the regression parameters: $\boldsymbol{\gamma}_0 = 3$, $\boldsymbol{\gamma}_1 = -1$, and $\boldsymbol{\gamma}_2 = 0.25$. The errors, $e_{1i}, \ldots, e_{6i}$, were generated under the distributional conditions previously described, with the variances of the errors chosen so that the reliability of each indicator adhered to the study levels detailed above. Error term d_i was generated from a normal distribution with mean 0 and with Var(d) taken so that $R^2 = .5$. The latent factor, f_{2i}, was generated under the distributional study conditions with mean 3 and variance 1.

For each of the 450 possible condition combinations, 500 data sets were generated with SAS.[1] As for the number of replications, Bandalos (2006) suggested that 500 replications were large for SEM Monte Carlo simulation studies. She argued that this number of replications would provide stable standard error estimates even when data were generated to come from a nonnormal distribution. Once the data were generated, they were analyzed with the aforementioned statistical software.

[1]SAS code for data generation can be downloaded from the first author's website (http://education.umd.edu/EDMS/fac/Harring/webpage.html). The LISREL code, Mplus code, WinBUGS code, and SAS NLMIXED code to run each of the five methods can be found in the Appendix.

Evaluation Criteria

The empirical performance of each estimation method was evaluated on the basis of $\hat{\theta}_b, b=1, \ldots, 500$. Bias for each parameter was calculated as the difference between the mean of the estimates obtained from the 500 replications and the true value. Hoogland and Boomsma (1998) suggested as a criterion that the absolute value of bias be less than .05 for parameter estimates in a particular condition, or across conditions in a study, to be considered unbiased. Other criteria have been considered for other measures of bias, such as relative bias (see, e.g., Gagné, 2004; Hoogland & Boomsma, 1998). In the current study, we adopted Hoogland and Boomsma's suggested value of .05 to demarcate bias and unbiased parameter estimates. RMSE is the square root of the expected squared loss around the true parameter value and was computed for the mth element of parameter vector θ as

$$\hat{\theta}(m)_{\text{RMSE}} = \left\{ \sqrt{500^{-1} \sum_{b=1}^{500} \left[\hat{\theta}_b(m) - \theta_0(m) \right]^2} \right\},$$

where $\theta_0(m)$ is the true parameter for the mth element of θ and $\hat{\theta}_b(m)$ is the estimate obtained by the approach under consideration. The accuracy of the estimated standard errors was evaluated by calculating the ratio $SE\left[\hat{\theta}(m)\right]/SD\left[\hat{\theta}(m)\right]$ (see, e.g., Lee, Poon, & Bentler, 1995), where $SE\left[\hat{\theta}(m)\right]$ is the square root of the mean of the variance of the estimates of $\hat{\theta}(m)$ obtained from the 500 replications and $SD\left[\hat{\theta}(m)\right]$ is the sample standard deviation obtained from $\hat{\theta}_b(m)$. If the standard errors are accurate, $SE\left[\hat{\theta}(m)\right]$ should be close to $SD\left[\hat{\theta}(m)\right]$, and the ratio $SE\left[\hat{\theta}(m)\right]/SD\left[\hat{\theta}(m)\right]$ should be approximately 1.0. Values less than 1 would indicate that the standard errors are being systematically underestimated, which would lead to inflated Type I error when the parameter null hypotheses are true. On the other hand, standard error ratios greater than 1 would lead to inflated Type II error rates for detecting nonzero parameters.

Convergence to a proper solution was also tracked, as some methods demonstrated a lack of convergence to a proper solution at the more extreme levels of the conditions. In situations in which replicates did not converge, replacement replications were not generated. Thus, the computations of the outcome criteria are based on only those converged replications.

Research Questions

Specific research questions we address in this simulation study are

1. Do differences exist among the five estimation methods in terms of bias? If so, which manipulated study conditions influence the accuracy of parameter recovery?
2. Do differences exist among the five estimation methods in terms of variability of estimates as measured by RMSE? If so, which manipulated study conditions influence variability of the parameter estimates?
3. Do differences exist among the five estimation methods in terms of the accuracy of computing standard errors of the parameter estimates? If so, which manipulated study conditions influence standard error estimation accuracy?

Results of Simulation Study 1

Convergence

Convergence was monitored, as some methods were thought to be negatively impacted by the extreme sample size and distributional conditions. For the LVS method, which used ordinary least squares estimation, convergence was not an issue. Similarly, there were few, if any, convergence issues with the LMS, MML, and BAY approaches, although the MML method had a small number of replicates that did not converge, particularly when the sample size was small ($n = 50$ and $n = 100$) and the exogenous predictor distribution was severely nonnormal (skew = 3, kurtosis = 20). However, the number of nonconvergent solutions was less than 1% of all replicates in any particular cell of the design. For the UNC method, sample size, nonnormality, and indicator reliability did have a modest negative impact on the rate of convergence.

When the exogenous predictor was severely nonnormally distributed, the indicator reliability was poor, and the sample size was 50 or 100, the UNC approach was unable to converge for approximately 1%–8% of the replicate data sets. When sample size increased to 250 or higher, the indicator reliability increased, or the exogenous predictor deviated modestly from normality, the UNC approach reached convergence for all replicates.

Bias, RMSE, and Standard Error Ratio

Bias, RMSE, and standard error ratio from Simulation Study 1 for the intercept, $\boldsymbol{\gamma}_0$, and the first-order linear term, $\boldsymbol{\gamma}_1$, paralleled the results for the quadratic coefficient, and as such are not shown to conserve space. Tables 2, 3, and 4 display the effects of sample size, nonnormality, and estimation method on parameter bias, RMSE, and standard error ratio of estimates of the quadratic coefficient $\boldsymbol{\gamma}_2$ across the three levels of indicator reliability.

Bias—Bias was positive for all methods except the LVS method, which produced estimates that were consistently negative. Marginally across all other conditions, the UNC method showed the least amount of bias, followed closely by the MML and LMS estimation methods. This result is commensurate with results from past simulation studies that have examined nonlinear effects in structural equation models (e.g., Marsh et al., 2004). In reference to the criterion of .05, the UNC method produced bias estimates, particularly when the reliability of the indicators was poor coupled with small sample size. The LVS method demonstrated the greatest bias (negative), with average values of bias exceeding .05 in absolute value. Again, this occurred when the reliability was moderate (.65) or poor (.45) and the sample size was small to moderate. The MML, BAY, and LMS methods produced estimates that were less biased compared with the UNC and LVS methods and were insensitive to the effects of unreliable indicators and sample size. As anticipated, bias across conditions decreased as sample size increased, but there was a pattern indicative of diminishing returns for sample sizes larger than 500. Bias in parameter estimates produced by the MML, BAY, and LMS procedures were comparable to the UNC method when reliability of the indicators was good (e.g., reliability = .85). Perhaps not too surprisingly, bias was greatest for all methods (relative to themselves) at the extreme small sample size ($n = 50$) coupled with poor indicator reliability (reliability = .45) and exogenous factor nonnormality (NN-4, NN-5, and NN-6). Essentially, all methods produced more biased estimates in these nonnormal conditions compared with case when the indicators were nonnormal and the factors were normally distributed.

RMSE—Variability of the quadratic parameter estimate as measured by RMSE showed predictable, systematic change across study conditions of sample size and indicator reliability, yet less anticipated results across nonnormality conditions. As for estimation

NIH-PA Author Manuscript

NIH-PA Author Manuscript

methods, the UNC, LMS, BAY, and MML methods showed comparable RMSE values across study conditions. The LVS estimates were more variable overall than those generated by the other four methods. However, this result may be explained, in part, by the fact that the LVS method showed the most consistent and greatest overall parameter bias among the estimation approaches. Because the RMSE comprises both parameter estimate variability and bias, it is difficult to disentangle the extent that these values for the LVS method were influenced by parameter bias when using this measure.

As predicted, the parameter was more precisely estimated as the sample size and indicator reliability increased. Unexpectedly, in the nonnormal conditions, RMSE values tended to decrease slightly in the more extreme exogenous factor nonnormality conditions.

Standard error ratio—Standard error ratio was perhaps the most revealing of the outcome variables and the one that seems to have received the smallest amount of attention in the literature. In its current usage, the ratio communicates how the standard error of the estimated parameter compares with an empirically derived standard deviation that represents the true variability of the sampling distribution of the regression coefficient in the population. A value of 0.80, for example, would indicate that the standard error that is computed could be up to 20% smaller than what would be realistic in the population. The implication is that for values straying too far below 1.0, a proliferation of Type I error is likely, given that standard error estimates will be on average too small. On the other hand, ratio values above 1.0 would indicate that Type II error is probable given that the standard error estimates will be on average too large.

The results of the simulation study in terms of the relative ratio are displayed graphically in Figure 4. As Figure 4 shows, the ratio of standard errors is not the same across estimation methods. The MML and LMS methods produced standard errors that are closest to the standard deviation computed from the empirical sampling distribution of $\boldsymbol{\gamma}_2$, as the median ratios from these two methods are near 1.0, 0.946, and 0.949, respectively. The UNC and LVS methods are distinctive in that they demonstrate up to a 19% decrease in the magnitude of the standard error (UNC = 0.858 and LVS = 0.810). Standard error ratios were more variable for the BAY approach compared with the MML and LMS approaches. The BAY approach demonstrated better ratio performance in smaller sample sizes (ratio = 0.975 for $n = 50$ and ratio = 0.967 for $n = 100$) than the likelihood-based competitors, but was consistently less than 1.0 overall and demonstrated poorer performance as the sample size increased. This result was counter to that found in a small simulation by Lee (2007).

Simulation Study 2

It is frequently the case that inferential procedures for nonlinear structural models involve testing whether the regression coefficient for the nonlinear term is statistically significant. To address the performance of such test procedures under the five methods of estimating latent variable quadratic effects, we conducted a second Monte Carlo simulation study. All variables were simulated to come from a population in which

$$f_{1i}=3+\gamma_1 f_{2i}+\gamma_2 f_{2i}^2+d_i, \quad (8)$$

where the mean and variance of f_{2i} were set to 3 and 1, respectively. The effect size to be manipulated represents the additional variance that the quadratic effect explains in f_1 above and beyond that which can be explained by the first-order effect, and is equal to the value expressed by (Marsh et al., 2004)[2]

$$R^2_{\gamma_2}=\frac{\gamma_2^2\left(2\phi_{11}^2\right)}{\sigma^2_{f_1}}. \quad (9)$$

Effect sizes for the quadratic effect were chosen to correspond to 0% (to investigate Type I error rates), 5%, and 10%, and are similar to values found in other studies (see, e.g., Dimitruk et al., 2007; Jaccard & Wan, 1995). To generate the data, the linear and quadratic coefficients in Equation 9 were computed to correspond to effect sizes above with constant variance—Var(d) = 10. In the situation with no effect, $R^2_{\gamma_2}=0$, $\boldsymbol{\gamma}_0$ = 3, $\boldsymbol{\gamma}_1$ = 3.162, and $\boldsymbol{\gamma}_2$ = 0. For $R^2_{\gamma_2}=.05$, $\boldsymbol{\gamma}_0$ = 3, $\boldsymbol{\gamma}_1$ = 0.082, and $\boldsymbol{\gamma}_2$ = 0.5. And lastly, for $R^2_{\gamma_2}=.10$, $\boldsymbol{\gamma}_0$ = 3, $\boldsymbol{\gamma}_1$ = −1.243, and $\boldsymbol{\gamma}_2$ = 0.707. Five sample sizes were examined (n = 50, n = 100, n = 250, n = 500, n = 1,000); reliability of the indicators followed those values chosen in Simulation Study 1, namely .45, .65, and .85. Because the impact of nonnormality of the observed variables was negligible, only latent variable nonnormal conditions were examined, and follow the conditions of the first simulation study. The three conditions were (a) skewness = 0, kurtosis = 0; (b) skewness = 2, kurtosis = 7; and (c) skewness = 3, kurtosis = 20.

Therefore, the design of the second simulation study was a 3 (effect size) × 5 (sample size levels) × 3 (observed variable reliability levels) × 3 (latent variable distribution levels) completely crossed factorial design resulting in 135 possible combinations. In each condition, 500 data sets were generated. The five estimation schemes were then employed to analyze the data sets in each cell.

Research Questions

4 Do differences exist among the five methods in terms of Type I error rate control? If so, which manipulated study conditions influence Type I error rates?

5 Do differences exist among the five estimation methods in terms of power to detect the quadratic effect? If so, which manipulated conditions influence power?

Type I Error Rate and Power

The empirical Type I error rates of the nominal size $\boldsymbol{\alpha}$ = .05 two-sided tests (under the null hypothesis, $H_0 : \boldsymbol{\gamma}_2 = 0$) when using the five estimation procedures are given in Table 5. The Type I error rate was computed as the proportion of converged solutions that had a statistically significant quadratic effect (at the .05 level) in the simulated data when H_0 was true. In addition, empirical power (probability of rejecting a false null hypothesis, $H_0 : \boldsymbol{\gamma}_2 = 0$) was computed and tabulated under the 5% and 10% effect size conditions. Bradley (1978) presented both conservative and liberal criteria for identifying conditions in which hypothesis testing procedures work adequately. His conservative criterion was .045 $\leq \boldsymbol{\alpha} \leq$.055, and his liberal criterion was .025 $\leq \boldsymbol{\alpha} \leq$.075. For this study, the liberal criterion was

[2]Marsh et al. (2004) computed an effect size for an interaction effect. Because the quadratic can be thought of as a latent variable interacting with itself, the quadratic effect above is a simplified version of that presented in Marsh et al.:

$$R^2_{\gamma_2}=\frac{\gamma_2^2\left(\phi_{11}\phi_{22}+\phi_{21}^2\right)}{\sigma^2_{\eta}}.$$

used to identify conditions for which the Type I error rate was unacceptable. These are indicated in italics in Table 5.

Type I error rate—It is evident from the reported values that when the sample size was large (i.e., n = 500 or n = 1,000) and data were normally distributed, all methods did a good job at controlling the Type I error rate under all conditions except the most unreliable (i.e., reliability = .45). Under this severe condition, all the methods rejected the null hypothesis more frequently than the nominal level would predict, except when coupled with large sample size.

When data were nonnormally distributed, the UNC and LVS approaches had better Type I error rates than the other methods, and furthermore, the Type I error rate was close to the desired alpha level as indicator reliability and sample size increased. Overall, The LMS, MML, and BAY methods had high Type I error rates when the data were nonnormal, although these rates were mitigated by the severity of the nonnormality (e.g., better Type I error control in the mild nonnormal condition). Also, as nonnormality became more severe, Type I error rate soared as high as 15% and 13% for the LMS and MML approaches, respectively, albeit the degree of inflation diminished as the reliability of the indicators increased from .45 to .85. The Type I error rate under the BAY approach also showed elevated levels across conditions of nonnormality and reliability but diminished as the sample size increased.

Empirical power—In response to Research Question 5, the values in Table 6 and Table 7 were examined. Empirical power is represented by the proportion of converged solutions that have a statistically significant quadratic effect in the simulated data when the population quadratic effect was not equal to 0. Empirical power rates for effect sizes $R^2_{\gamma_2}=.05$ and $R^2_{\gamma_2}=.10$ were computed with an alpha level of .05, and are displayed in Table 6 and Table 7, respectively.

As anticipated, empirical power increased as the size of the effect increased from 5% to 10% across methods and conditions. That is, when medium to large quadratic effects exist in the population, the methods were able to detect them with a great deal of certainty for moderate sample sizes under suboptimal reliability and normality conditions. This was the case even when the sample size was extremely small (n = 50) and the indicators were unreliable (reliability = .45).

In general, all the estimation approaches had lower power when the observed variables were measured unreliably, the sample size was small, and the data were nonnormal. When the reliability was low and data were normally distributed, the MML, BAY, and LVS approaches had the highest power to detect true quadratic effect, although the UNC and LMS methods did not lag far behind. Predictably, power increased across all estimation methods as reliability and sample size increased.

Conclusions and Recommendations

For the quadratic model, both the Bayesian approach and the methods based on maximum likelihood (MML, LMS, and UNC approaches) appeared to outperform the moderated multiple factor score regression method (LVS) in terms of bias in the coefficient for the quadratic term as well as in the precision of estimation. Of additional concern, the standard error ratios for the LVS method were well below 1.0, which may indicate a tendency for increased Type I error. Although power to detect effects with the LVS method was adequate across many conditions, when paired with the other results, we recommend that the LVS method not be used in lieu of other approaches.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The UNC, MML, LMS, and BAY approaches were comparable in terms of parameter bias and RMSE across the study conditions. Bias was affected by sample size and indicator reliability, but in predictable ways, commensurate with results from past simulation studies that have examined nonlinear effects in structural equation models (Marsh et al., 2004). The different methods were particularly similar for situations in which sample sizes were adequate ($n > 100$), indicator variables were somewhat reliable (reliability > .65), and exogenous factor distributions were normal or mildly nonnormal. It should be noted that for combinations of small sample size, poor indicator reliability, and nonnormality of the exogenous predictors, the methods performed much worse than when sample size was larger, the observed variables were moderately reliable, and the exogenous predictor distributions did not deviate greatly from normality; although the more extreme conditions appeared to impact the UNC method more than the MML, LMS, and BAY approaches. This simulation study did not incorporate the method recently introduced by Mooijaart and Bentler (2010) that effectively counteracts the effects of nonnormality. Because nonnormality is expected in nonlinear structural models, this particular approach may have some advantage over the methods investigated here, although the extent of its potential benefit would require additional investigation.

Greater differences among methods were observed for the standard error ratio. The MML and LMS methods were comparable and superior in terms of accuracy of the standard errors of the regression coefficients. As was true in the small simulation study conducted by Lee (2007), the BAY approach had ratios that worked well for small sample sizes. In contrast, however, for larger sample sizes the approach produced more erratic behavior in ratio values. This was an interesting finding, which we could not readily explain. The UNC method consistently underestimated the standard errors, resulting in the chance for inflated Type I error.

The UNC method held Type I error rates at the nominal level across different reliability and nonnormality conditions, and it provided sufficient power to detect medium to large quadratic effects due to low standard errors. Both MML and LMS methods demonstrated adequate Type I error control under normality, but under nonnormal conditions, rates were inflated. Both of these methods displayed more than enough power to detect real quadratic effects. The LVS method consistently demonstrated poor properties across many of the examined conditions. However, these results may be a consequence of measurement error at the regression modeling level and not the result of the manipulated conditions.

These results point us to the following recommendations:

1. The LVS approach is not recommended at this time. It consistently demonstrated poor properties across many of the examined conditions. These results may be a consequence of measurement error at the regression modeling level and not the result of the manipulated conditions, however. Before advocating its use, further investigation into what role measurement error plays is warranted.
2. Under ideal data-analytic conditions—high indicator reliability normality, and larger sample sizes—we recommend the MML, LMS, or BAY method, as they appear to be comparable in terms of parameter accuracy and estimation precision. The BAY approach is better for small sample sizes, whereas the LMS and MML methods are recommended for larger sample sizes ($n > 250$).
3. Under less optimal study conditions, especially for mild to severe nonnormality, the UNC method is recommended. The UNC method held Type I error rates at the nominal level across moderate reliability and nonnormality conditions, and it provided sufficient power to detect medium to large quadratic effects.

Empirical Example

For illustrative purposes, we examined the quadratic effect of latent decoding on latent fluency using all five estimation methods. Two of the principal skills implicated in reading comprehension are decoding, which is the ability to connect letters and sounds to read words, and fluency, which is accurate and automatic decoding at an appropriate pace. According to the verbal efficiency theory (Perfetti, 1985), decoding must be fluent in order to free up cognitive resources to focus on comprehension, the end goal of reading. A study of 230 fourth-grade students obtained from a large urban school district found that latent fluency mediates the relation between latent decoding and latent reading comprehension (Silverman, Speece, Harring, Ritchey, & Cutting, 2010). In addition, theory suggested that there might be a curvilinear relationship between decoding and fluency such that as decoding ability increases, fluency initially increases at a faster rate, reaches a peak, and then decreases or levels off once decoding skills reach an advanced level. In the current example, latent decoding (DEC) was measured by tasks assessing spelling, word discrimination, phonological elision, pseudoword repetition, word attack, and word identification, and is on the scale of the Woodcock–Johnson III Word Identification standard scores ($M = 100$, $SD = 15$; Woodcock, McGrew, & Mather, 2001). Latent fluency (FLU) was measured by tasks assessing oral passage reading fluency, word identification fluency, spelling fluency, rapid automatized letter naming, silent reading fluency, pseudoword decoding efficiency, and sight word efficiency, and is on the scale of oral passage reading fluency words correct per minute (norm referenced scores between 80 and 120; Hasbrouck & Tindal, 1992). Estimates of reliability from the sample for decoding ranged from .42 (phonological elision) to .87 (word identification). Estimates of reliability of measures of fluency ranged from .40 (pseudoword decoding efficiency) to .83 (word identification fluency). Note that these values were commensurate with the reliability conditions used in the simulation studies. A suggested measurement model for the six observed measures, DEC_j, $j = 1, \ldots, 6$, and seven observed measures for fluency, FLU_k, $k = 1, \ldots, 7$, is

$$\begin{pmatrix} \mathrm{FLU}_1 \\ \mathrm{FLU}_2 \\ \mathrm{FLU}_3 \\ \mathrm{FLU}_4 \\ \mathrm{FLU}_5 \\ \mathrm{FLU}_6 \\ \mathrm{FLU}_7 \\ \mathrm{DEC}_1 \\ \mathrm{DEC}_2 \\ \mathrm{DEC}_3 \\ \mathrm{DEC}_4 \\ \mathrm{DEC}_5 \\ \mathrm{DEC}_6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & 0 \\ \lambda_{51} & 0 \\ \lambda_{61} & 0 \\ \lambda_{71} & 0 \\ 0 & 1 \\ 0 & \lambda_{92} \\ 0 & \lambda_{102} \\ 0 & \lambda_{112} \\ 0 & \lambda_{122} \\ 0 & \lambda_{132} \end{pmatrix} \begin{pmatrix} \mathrm{FLU} \\ \mathrm{DEC} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \\ e_{13} \end{pmatrix}.$$

The structural model is the conventional quadratic model with decoding mean centered to make the interpretation of the intercept more sensible:

$$\mathrm{FLU} = \beta_0 + \beta_1 \left(\mathrm{DEC} - \overline{\mathrm{DEC}}\right) + \beta_2 \left(\mathrm{DEC} - \overline{\mathrm{DEC}}\right)^2 + d. \quad (10)$$

The coefficients in Equation 11 are interpreted as

β_0: Value of FLU at the mean of DEC

β_1: Instaneous rate of change at the mean of DEC

β_2: Curvature (changing rate of change) in FLU over values of DEC.

A comparison of the regression coefficients of the structural model and corresponding standard errors across the five estimation methods can be found in Table 8.

It is clear that although the coefficients are not the same across methods, they are strikingly similar. Across methods the standard errors of the coefficients were small compared with their corresponding estimates, and each would be statistically significant at most practical nominal levels. The average fluency score for a subject with average decoding is approximately 130 (see β_0 in Table 8). The quadratic coefficient (see β_2 in Table 8) estimated by each method was small and negative, indicating downward curvature in the fitted function. Differences in the estimated structural regression parameters can best be seen by examining a plot of the fitted functions computed for each method. Under Equation 9 and with the results from the MML method, the variance that the quadratic effect of decoding explains in fluency above and beyond that which can be explained by the first-order effect is 0.027. In terms of the simulation study, this effect would be considered small to moderate.

With the output from all five methods, the fitted functions in Figure 5 give the estimated relation over the range of decoding: $60 \le \text{DEC} \le 150$. The fitted relations are consistent with the researcher's hypothesis that at advanced levels of decoding (about 2 standard deviations above the mean ≈ 140), there is a point at which fluency reaches its maximum. Further fluency at this level of decoding likely fails to add to improved reading comprehension, and may in fact hinder comprehension if students are reading too fast and not paying attention to what they are reading.

Why were the outcomes so similar? One explanation for these findings might be found by connecting the results of the simulation studies to the sample data used in the empirical example. From the simulation studies, combinations of poor reliability (e.g., .45), smaller sample sizes (e.g., $n = 50$ or 100), and moderate to extreme nonnormality (e.g., skew = 2, 3; kurtosis = 7, 20) appeared to have the greatest marginal impact on the methods in terms of bias and standard error ratio. Aside from a single indicator of both decoding and fluency, the reliability estimates would be categorized as fair and good, .65 to .85. The sample univariate skew and kurtosis measures for decoding ranged from 0.084 to –0.75, and kurtosis ranged in absolute value from 0.47 to 1.14. For fluency, sample skew and kurtosis indices ranged from –0.12 to –0.99 and 0.08 to 0.92, respectively. In terms of the study conditions, these values would indicate modest departures from normality. Lastly, the sample size, $n = 230$, would be considered moderate in terms of the simulation study conditions. Thus, it may not be too surprising, given that the UNC, MML, LMS, and BAY methods were similar in terms of parameter bias under more favorable data-analytic conditions, that the methods produced estimates that were quite close to one another. It should be noted that the estimate for the quadratic coefficient under the LVS approach was smaller than those produced by the other methods. This also coincides with the findings of the simulation study, as the LVS method consistently underestimated the quadratic coefficient across most study conditions. All the methods were able to find this small quadratic effect. This, too, corresponds to the results of Simulation Study 2, as the power was ample to detect moderate effects across methods and study conditions.

Discussion

Quadratic effects between continuous variables are often hypothesized in the social sciences because, when coupled with a linear component, they adequately approximate many curvilinear behavioral processes. Although many simulation studies have been conducted to compare estimation approaches of nonlinear effects in SEM, an overwhelming majority of these studies has focused on latent interactions (with the notable exception of Marsh et al., 2004). An examination of quadratic effects was conducted in the current study via Monte

Carlo simulation to compare five methods for estimating the quadratic relation between latent variables: the two-stage moderated regression approach using latent variable scores, the unconstrained approach, the latent moderated structural equation approach, marginal maximum likelihood, and a fully Bayesian approach. These methods span the continuum of approaches currently available to practitioners. These methods were not scrutinized in isolation but rather in realistic data-analytic conditions (nonnormality, reliability, sample size, and effect size) in which practitioners using latent variable methods often work and under more extreme conditions as well.

Our findings in the simulation studies, coupled with our real data analysis, are consistent with the literature in the following ways. As with past simulation studies, exogenous predictor non-normality and indicator variable reliability tended to have the greatest impact on the ability of the methods to accurately and precisely estimate the quadratic effect with sample size exerting less influence. All methods appeared to control Type I error fairly well except at the most extreme data-analytic conditions, within which all methods performed poorly.

There are also some important new insights that can prove useful for practitioners who are interested in modeling quadratic effects within an SEM framework. For example, unlike with other simulation studies, we studied not only the effects of real-world data-analytic conditions on the ability to estimate quadratic effects, but also those that pushed the boundaries of what might be deemed minimally acceptable in practice. Severe nonnormality, poor indicator reliability, and overly small sample sizes appear to have the greatest negative impact on the estimation accuracy, precision, and deflation of standard errors of the quadratic parameter. Because all the methods investigated here performed poorly under these circumstances, if data exhibit these characteristics in practice, statistical conclusions should be made cautiously. From the knowledge gained from carrying out this study, we are more convinced that the efficacy of the Bayesian approach should be investigated on its own merits and not necessarily compared with likelihood-based methods for this type of nonlinear structural equation model. Our uneasiness stems from the fact that we, like some others who have ventured down this path (e.g., Lee, 2007), have set up the simulation “game” somewhat unfairly. Aside from the philosophical differences that exist between frequentist and Bayesian approaches, the obvious advantages that the Bayesian framework offers were not exploited. For example, as was previously mentioned, in a Bayesian approach prior knowledge about model parameters including their distributional assumptions can be incorporated into the model formulation. In the simulation studies executed in this article, we used noninformative conjugate priors, which put the preponderance of weight in estimating the posterior distribution on the data (or the likelihood). We would expect the Bayesian method under this scenario to behave very similarly to the marginal maximum likelihood method. For this set of simulation conditions, it did. What has been learned? We encourage further exploration into the Bayesian approach that was not investigated here. Specifically, there is a need to better understand the impact of prior distributions (not only the values of hyperparameters characterizing the distributions but also the choices of distributions themselves) on parameter recovery and accuracy of standard errors of the estimates.

The five estimation approaches reviewed here constitute both popular estimators and approaches that are derived in a sensible but admittedly ad hoc manner. All these methods can be used to estimate the quadratic structural model developed here as well as the full quadratic model that includes the interaction between exogenous predictors. Unfortunately, not all these methods can be extended easily to the general class of nonlinear structural equation models. This is certainly the case for the LMS method, which allows only quadratic terms to be formulated. It is not obvious how indicator variables would be constructed for transcendental functions of exogenous latent variables (i.e., e^{f2} or $\ln(f_2)$) with the ad hoc UNC method. Even marginal maximum likelihood, with its ability to handle the most general cases, suffers the deficiency of lengthy estimation times for multidimensional integration of order greater than three or four. The Bayesian approach also allows for the most general model specification and has the ability to incorporate prior information about model parameters—which can be seen as both an asset and a liability depending on one's familiarity with specifying a model within this framework. Lastly, the LVS method, with all its statistical drawbacks, can also handle very broad nonlinear models. In the second stage of the estimation process, nonlinear least squares estimation would need to be employed for intrinsically nonlinear functions, otherwise ordinary least squares can be used. Despite the findings from this study and others, in the end, the complexity of the structural model may ultimately have the greatest impact on which estimation method or approach will be endorsed.

All five estimation methods were employed and fit to reading data from a sample of fourth-grade students. The estimated pa rameters and their corresponding standard errors were similar but not identical, which leaves some question as to the comparability of the methods. On the basis of the results of the simulation studies in this article, further exploration into the distribution of the variables and the reliability of the predictors paired with the moderate sample size may lead to choosing a particular method because it is optimal under these analytic circumstances.

Acknowledgments

We would like to thank André Rupp and Greg Hancock for their helpful feedback on earlier drafts of this article.

Appendix

Computer Code for Five Estimation Methods

Data for the simulation were coded in SAS Interactive Matrix Language. The following input statements were used to run each of the methods in the various statistical software programs.

Multiple Regression Using Latent Variable Scores (Implemented in LISREL)

LISREL code to create latent variable scores

Estimating the measurement model and latent variable scores

```
DA NI=6 NO=&SAMP

RA FI=C: \quadsim\Cell&COUNT2\quad_r_&COUNT1..dat

LA

Y1 Y2 Y3 ×1 ×2 ×3

MO NY=3 NE=1 NX=3 NK=1 TX=FU,FI TY=FU,FI

LY=FU,FI LX=

FU,FI C TE=DI,FR TD=DI,FR PH=SY,FR KA=FR AL=FR

LE
```

```
eta
LK
ksi1
FR LX(2,1) LX(3,1) LY(2,1) LY(3,1)
FR TX(2,1) TX(3,1) TY(2,1) TY(3,1)
VA 1.00 LX(1,1) LY(1,1)
VA 0.00 TX(1,1) TY(1,1)
PS=QUADRATIC.psf
OU
```

SAS code to read in new data file, create the quadratic term, and run the regression analysis using the latent variable scores

```
Data LVS;
Infile QUADRATICnew.raw dlm=′′ firstobs=1;
Input Y1-Y3 ×1-X3 eta ksi1;
ksi1ksi1=ksi1**2;
run;
Proc Reg Data=LVS tableout edf simple outest=params;
Model eta=ksi1 ksi1ksi1 / scorr2;
run;
```

Unconstrained Model (Implemented in LISREL)

LISREL code to run the unconstrained model

```
DA NI=9 NO=500
RA FI=C:\Quadratic\QUADRATIC.dat
LA
y1 y2 y3 x1 x2 x3 x1x1 x2x2 x3x3
MO NY=3 NE=1 NX=6 NK=2 TX=FU,FI TY=FU,
FI LY=FU,FR PS=FR C
PH=FU,FR TE=DI,FR TD=SY,FI AL=FR KA=FR
LE
```

```
eta
LK
ksi ksi*ksi
FI LY 1 1
FR LX 2 1 LX 3 1 LX 5 2 LX 6 2
FR TX 2 1 TX 3 1 TX 5 1 TX 6 1
FR TY 2 1 TY 3 1
VA 1 LY 1 1 LX 1 1 LX 4 2
VA 0 TX 1 1 TY 1 1
FR GA 1 1 GA 1 2
FR TD(1,1) TD(2,2) TD(3,3) TD(4,4) TD(5,5) TD(6,6)
FR TD(4,1)
FR TD(5,2)
FR TD(6,3)
OU AD = OFF IT = 5000 XM
```

Latent Moderated Structural (LMS) Model (Implemented in Mplus)

```
TITLE: Quadratic SEM Using LMS in Mplus
DATA: FILE IS C:\Quad\SASMplus\Cell1\Data\quad&rep.dat;
VARIABLE: NAMES = y1-y6;
USEVARIABLES ARE y1-y6;
ANALYSIS: TYPE = RANDOM;
ALGORITHM = INTEGRATION;
INTEGRATION = GAUSS(30);
ADAPTIVE = OFF;
STITERATIONS = 50;
ITERATIONS = 5000;
SDITERATIONS = 250;
MITERATIONS = 1000;
MODEL:
```

f2 BY y4 y5*.7 y6*.7; [y4@0];

f1 BY y1 y2*.7 y3*.7; [y1@0];

f1SQ | f1 XWITH f1;

f2 ON f1*-1 f1SQ*.25;

[f1*3 f2*3];

f1*1;

f2*.5;

y1*.4 y2*.4 y3*.4 y4*.4 y5*.4 y6*.4;

OUTPUT: Tech1 Tech8;

savedata: results are C:\Quad\SASMplus\Cell1\Output\out&rep.dat;

Bayesian Approach (Implemented in WinBUGS)

The Bayesian approach used the R2WinBUGS library and bugs function in R. R was utilized as the platform to call WinBUGS and collate results upon convergence of the program. There is a debugging option in the bugs function that allows monitoring of the iteration history and mixing. We used this extensively in the beginning to identify problematic code. The bugs function requires three files to call the WinBUGS program:

quad.sim <- bugs(d, init, parameters, "quadwin.txt", n.chains=3, n.iter=5000, n.burnin=floor(3000), debug=FALSE)

File 1: Initial Values (init)

init=function(){list(gam=c(3,-1,.25), lam=c(.7,.7,.7,.7), psiinv=c(.001,.001, .001,.001,.001,.001),muksi=3, phi1inv=1)}

File 2: Parameters to Monitor (parameters) parameters = c("gam", "lam", "psi", "muksi", "phi1")

File 3: Model Statement (quadwin.txt) model{for(i in 1:100){

#Specify the measurement model

z[i,1]~dnorm(mu[i,1],psiinv[1])

z[i,2]~dnorm(mu[i,2],psiinv[2])

z[i,3]~dnorm(mu[i,3],psiinv[3])

z[i,4]~dnorm(mu[i,4],psiinv[4])

z[i,5]~dnorm(mu[i,5],psiinv[5])

z[i,6]~dnorm(mu[i,6],psiinv[6])

mu[i,1] <- ksi[i]

```
mu[i,2] <- int[1] + lam[1]*ksi[i]

mu[i,3] <- int[2] + lam[2]*ksi[i]

mu[i,4] <- eta[i]

mu[i,5] <- int[3] + lam[3]*eta[i]

mu[i,6] <- int[4] + lam[4]*eta[i]

#Specify the nonlinear structural model

eta[i] <- gam[1] + gam[2]*ksi[i] + gam[3]*ksi[i]*ksi[i] + delta[i]

#Specify the random parts of the latent distribution

ksi[i]~dnorm(muksi,phi1inv)

delta[i]~dnorm(0.0,deltainv) }

#Priors for Psi

for (t in 1:6){ psiinv[ta]~dgamma(0.1,0.1) psi[ta] <- 1/psiinv[ta] }

#Priors for Lam for (k in 1:4){ lam[k]~dnorm(0.0,0.0001) }

#Priors for Int for (k in 1:4){ int[k]~dnorm(0.0,0.0001) }

#Priors for Gamma for (j in 1:3){ gam[j]~dnorm(0.0,0.0001) }

#Priors for muksi and phi1 and delta

muksi~dnorm(0.0,0.0001)

phi1inv~dgamma(0.1,0.1)

phi1 <- 1/phi1inv

deltainv~dgamma(0.1,0.1)

ddelta <- 1/deltainv }
```

Marginal Maximum Likelihood (Implemented in SAS PROC NLMIXED)

Note that the "general" function in PROC NLMIXED could have also been used to carry out maximum likelihood estimation. However, the data would need to be in a slightly different format.

```
proc nlmixed data=longquad noad qpoints=30 tech=quanew lis=2 lsp=.1

maxiter=5000 maxfu=10000;

***starting values;

parms mu21 = 0 mu31 = 0 mu52 = 0 mu62 = 0

lam21 = 0.7 lam31 = 0.7 lam52 = 0.7 lam62 = 0.7
```

```
alpha = 3 phi = 1 var_d = 10
psi1 = 0.4 psi2 = 0.4 psi3 = 0.4
psi4 = 10
psi5 = 4
psi6 = 6
gam0 = 3
gam1 = 0.082
gam2 = 0.5;
if (item eq 1) then do;
mu = 1*f1;
sig2 = psi1;
end;
if (item eq 2) then do;
mu = mu21 + lam21*f1;
sig2 = psi2;
end;
if (item eq 3) then do;
mu = mu31 + lam31*f1;
sig2 = psi3;
end;
f2 = gam0 + gam1*f1 + gam2*f1**2 + d;
if (item eq 4) then do;
mu = 1*f2;
sig2 = psi4;
end;
if (item eq 5) then do;
mu = mu52 + lam52*f2;
sig2 = psi5;
end;
```

```
if (item eq 6) then do;

mu = mu62 + lam62*f2;

sig2 = psi6;

end;

model y~normal(mu,sig2);

random f1 d~normal([alpha,0], [phi,0,var_d]) subject = id;

bounds var_d > 0;

run;
```

References

Aiken, LS.; West, SG. Multiple regression: Testing and interpreting interactions. Sage; Newbury Park, CA: 1991.

Algina J, Moulder BC. A note on estimating the Jöreskog–Yang model for latent variable interaction using LISREL 8.3. Structural Equation Modeling. 2001; 8:40–52. doi:10.1207/S15328007SEM0801_3.

Anderson, TW.; Rubin, H. Statistical inference in factor analysis.. In: Neyman, J., editor. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability. Vol. 5. University of California Press; Berkeley: 1956. p. 111-150.

Bandalos, DL. The use of Monte Carlo studies in structural equation modeling research.. In: Hancock, GR.; Mueller, RO., editors. Structural equation modeling: A second course. Information Age; Greenwich, CT: 2006. p. 385-426.

Bartholomew, D. Latent variable models and factor analysis. Griffin; London, England: 1987.

Bollen, KA. Structural equations with latent variables. Wiley; New York, NY: 1989.

Bollen KA. An alternative two-stage least squares (2SLS) estimate for latent variable equation. Psychometrika. 1996; 61:109–121. doi: 10.1007/BF02296961.

Boomsma, A. On the robustness of LISREL against small sample size and non-normality. University of Groningen; Groningen, the Netherlands: 1983. (Unpublished doctoral dissertation)

Bradley JV. Robustness? British Journal of Mathematical and Statistical Psychology. 1978; 31:144–152. doi:10.1111/j.2044-8317.1978.tb00581.x.

Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics. 1998; 7:434–455. doi:10.2307/1390675.

Cohen, J.; Cohen, P.; West, SG.; Aiken, LS. Applied multiple regression/correlational analysis for the behavioral sciences. 3rd ed.. Erlbaum; Hillsdale, NJ: 2003.

Cudeck R, du Toit SHC. A version of quadratic regression with interpretable parameters. Multivariate Behavioral Research. 2002; 37:501–519. doi:10.1207/S15327906MBR3704_04.

Cudeck R, Harring JR, du Toit SHC. Marginal maximum likelihood estimation of a latent variable model with interaction. Journal of Educational and Behavioral Statistics. 2009; 34:131–144. doi: 10.3102/1076998607313593.

Curran PJ, West SG, Finch JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychological Methods. 1996; 1:16–29. doi:10.1037/1082-989X.1.1.16.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B. Methodological. 1977; 39:1–38. doi: 10.2307/2984875.

Dimitruk P, Schermelleh-Engel K, Kelava A, Moosbrugger H. Challenges in nonlinear structural equation modeling. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences. 2007; 3:100–114. doi:10.1027/1614-2241.3.3.100.

Fleishman AI. A method for simulating non-normal distributions. Psychometrika. 1978; 43:521–532. doi:10.1007/BF02293811.

Gagné, P. Generalized confirmatory factor mixture models: A tool for assessing factorial invariance across unspecified populations. University of Maryland; College Park: 2004. (Unpublished doctoral dissertation)

Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. Bayesian data analysis. 2nd ed.. Chapman & Hall; London, England: 2004.

Gorsuch, RL. Factor analysis. Saunders; Philadelphia, PA: 1974.

Hasbrouck JE, Tindal G. Curriculum-based oral reading fluency norms for students in Grades 2 through 5. Teaching Exceptional Children. 1992; 24:41–44.

Hayduk, LA. Structural equation modeling with LISREL: Essentials and advances. Johns Hopkins University; Baltimore, MD: 1987.

Hoogland JJ, Boomsma A. Robustness studies in covariance structure modeling: An overview and meta-analysis. Sociological Methods & Research. 1998; 26:329–367. doi: 10.1177/0049124198026003003.

Hu L, Bentler PM, Kano Y. Can test statistics in covariance structure analysis be trusted? Psychological Bulletin. 1992; 112:351–362. doi:10.1037/0033-2909.112.2.351. [PubMed: 1454899]

Jaccard J, Wan CK. Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. Psychological Bulletin. 1995; 117:348–357. doi:10.1037/0033-2909.117.2.348.

Jöreskog, KG. Latent variable scores and their uses. 2000. Retrieved from http://www.ssicentral.com/lisrel/advancedtopics.html

Jöreskog, KG.; Sörbom, D. LISREL for Windows [Computer software]. Scientific Software International; Lincolnwood, IL: 2006.

Jöreskog, KG.; Yang, F. Nonlinear structural equation models: The Kenny–Judd model with interaction effects.. In: Marcoulides, GA.; Schumacker, RE., editors. Advanced structural equation modeling: Issues and techniques. Erlbaum; Mahwah, NJ: 1996. p. 57-88.

Kelava A, Moosbrugger H, Dimitruk P, Schermelleh-Engel K. Multicollinearity and missing constraints: A comparison of three approaches for the analysis of latent nonlinear effects. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences. 2008; 4:51–66. doi:10.1027/1614-2241.4.2.51.

Kenny DA, Judd CM. Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin. 1984; 96:201–210. doi:10.1037/0033-2909.96.1.201.

Klein A, Moosbrugger H. Maximum likelihood estimation of latent interaction effects with the LMS method. Psychometrika. 2000; 65:457–474. doi:10.1007/BF02296338.

Klein A, Muthén BO. Quasi-maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. Multivariate Behavioral Research. 2007; 42:647–673. doi:10.1080/00273170701710205.

Kline, RB. Principles and practice of structural equation modeling. 3rd ed.. Guilford Press; New York, NY: 2011.

Krommer, AR.; Ueberhuber, CW. Numerical integration on advanced computer systems. Springer-Verlag; New York, NY: 1994. doi: 10.1007/BFb0025796

Lastovicka JL, Thamodaran K. Common factor score estimates in multiple regression problems. Journal of Marketing Research. 1991; 28:105–112. doi:10.2307/3172730.

Lee, SY. Structural equation modeling: A Bayesian approach. Wiley; London, England: 2007. doi: 10.1002/9780470024737

Lee S-Y, Poon W-Y, Bentler PM. A two-stage estimation of structural equation models with continuous and polytomous variables. British Journal of Mathematical and Statistical Psychology. 1995; 48:339–358. doi:10.1111/j.2044-8317.1995.tb01067.x. [PubMed: 8527346]

Lee S-Y, Song X-Y, Lee JCK. Maximum likelihood estimation of nonlinear structural equation models with ignorable missing data. Journal of Educational and Behavioral Statistics. 2003; 28:111–134. doi:10.3102/10769986028002111.

Lee SY, Song XY, Poon WY. Comparison of approaches in estimating interaction and quadratic effects of latent variables. Multivariate Behavioral Research. 2004; 39:37–67. doi:10.1207/s15327906mbr3901_2.

Lee SY, Song XY, Tang NS. Bayesian methods for analyzing structural equation models with covariates, interaction, and quadratic latent variables. Structural Equation Modeling. 2007; 14:404–434. doi:10.1080/10705510701301511.

Lee S-Y, Zhu H-T. Maximum likelihood estimation of nonlinear structural equation models. Psychometrika. 2002; 67:189–210. doi: 10.1007/BF02294842.

Lynch, SM. Introduction to applied Bayesian statistics and estimation for social scientists. Springer; New York, NY: 2007. doi:10.1007/978-0-387-71265-9

Marsh HW, Wen Z, Hau KT. Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. Psychological Methods. 2004; 9:275–300. doi: 10.1037/1082-989X.9.3.275. [PubMed: 15355150]

Marsh, HW.; Wen, Z.; Hau, KT. Structural equation models of latent interaction and quadratic effects.. In: Hancock, GR.; Mueller, RO., editors. Structural equation modeling: A second course. Information Age; Greenwich, CT: 2006. p. 225-265.

Mooijaart A, Bentler PM. An alternative approach for nonlinear latent variable models. Structural Equation Modeling. 2010; 17:357–373. doi:10.1080/10705511.2010.488997.

Moosbrugger, H.; Schermelleh-Engel, K.; Kelava, A.; Klein, AG. Testing multiple nonlinear effects in structural equation modeling: A comparison of alternative estimation approaches.. In: Teo, T.; Khine, MS., editors. Structural equation modeling in educational research: Concepts and applications. Sense; Amsterdam, the Netherlands: 2009. p. 103-136.

Moulder BC, Algina J. Comparison of methods for estimating and testing latent variable interactions. Structural Equation Modeling. 2002; 9:1–19. doi:10.1207/S15328007SEM0901_1.

Mulaik, SA. The foundations of factor analysis. McGraw-Hill; New York, NY: 1972.

Mulaik, SA. Foundations of factor analysis. 2nd ed.. Chapman & Hall/CRC Press; Boca Raton, FL: 2009.

Muthén, LK.; Muthén, BO. Mplus user's guide. 5th ed.. Muthén & Muthén; Los Angeles, CA: 2007.

Paap R. What are the advantages of MCMC based inference in latent variable models? Statistica Neerlandica. 2002; 56:2–22. doi:10.1111/1467-9574.00060.

Perfetti, CA. Reading ability. Oxford University Press; New York, NY: 1985.

Ping RA Jr. Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. Psychological Bulletin. 1996; 119:166–175. doi:10.1037/0033-2909.119.1.166.

Pinheiro JC, Bates DM. Approximations to the log-likelihood function in the nonlinear mixed-effects model. Journal of Computational and Graphical Statistics. 1995; 4:12–35. doi:10.2307/1390625.

Schumacker RE. Latent variable interaction modeling. Structural Equation Modeling. 2002; 9:40–54. doi:10.1207/S15328007SEM0901_3.

Silverman, R.; Speece, D.; Harring, JR.; Ritchey, M.; Cutting, RJ. What is the role of fluency in reading comprehension? Poster presented at the Pacific Coast Research Conference; Coronado, CA. 2010, February;

Skrondal, A.; Rabe-Hesketh, S. Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. Chapman & Hall/CRC Press; Boca Raton, FL: 2004.

Song X-Y, Lee S-Y. Bayesian estimation and test for factor analysis model with continuous and polytomous data in several populations. British Journal of Mathematical and Statistical Psychology. 2001; 54:237–263. doi:10.1348/000711001159546. [PubMed: 11817092]

Spiegelhalter, DJ.; Thomas, A.; Best, NG.; Lunn, D. WinBUGS user manual (Version 1.4). MRC Biostatistics Unit 26; Cambridge, England: 2002.

Stroud, AH.; Secrest, D. Gaussian quadrature formulas. Prentice Hall; Englewood Cliffs, NJ: 1966.

Wall, MM. Maximum likelihood and Bayesian estimation for nonlinear structural equation models.. In: Millsap, RE.; Maydeu-Olivares, A., editors. The Sage handbook of quantitative methods in psychology. Sage; London, England: 2009. p. 540-567.

Wall MM, Amemiya Y. Generalized appended product indicator procedure for nonlinear structural equation analysis. Journal of Educational and Behavioral Statistics. 2001; 26:1–29. doi: 10.3102/10769986026001001.

Wall MM, Amemiya Y. A method of moments technique for fitting interaction effects in structural equation models. British Journal of Mathematical and Statistical Psychology. 2003; 56:47–63. doi: 10.1348/000711003321645331. [PubMed: 12803821]

Weiss, BA. A comparison of methods for testing interaction effects in structural equation modeling. University of Maryland; College Park: 2010. (Unpublished doctoral dissertation)

Wolfinger, RD. Fitting nonlinear mixed models with the new NLMIXED procedure (Paper 287). SAS Institute; Cary, NC: 1999.

Woodcock, R.; McGrew, K.; Mather, N. Woodcock–Johnson Test of Cognitive Abilities. 3rd ed.. Riverside; Itasca, IL: 2001.

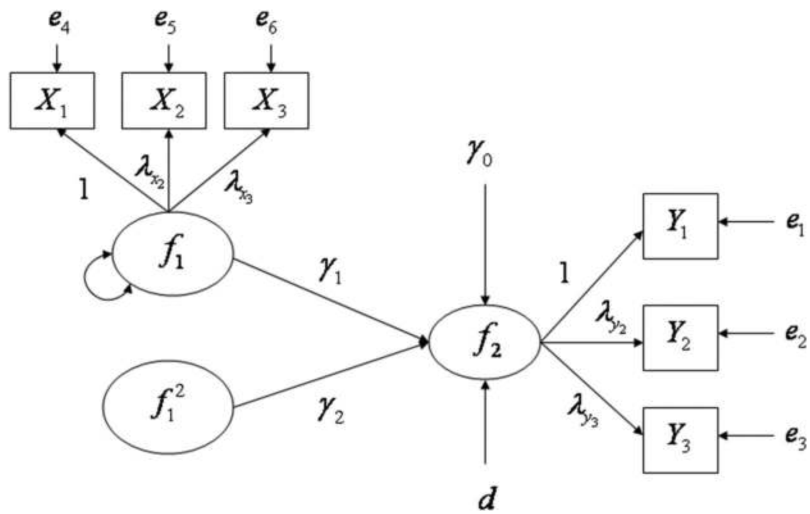

Figure 1.
Nonlinear structural equation model with one latent criterion, f_1; one latent exogenous predictor, f_2; and its corresponding latent quadratic term, f_2^2. The exogenous and endogenous latent variables are each measured by three indicators (X_1, X_2, X_3, and Y_1, Y_2, Y_3, respectively).

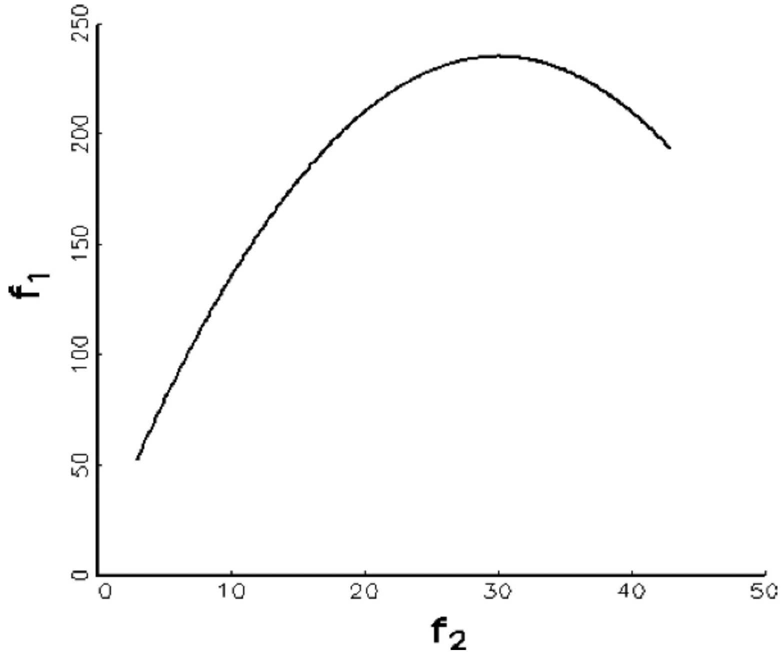

Figure 2.

Generic graph of a quadratic relationship between latent criterion, f_1, and exogenous predictor, f_2.

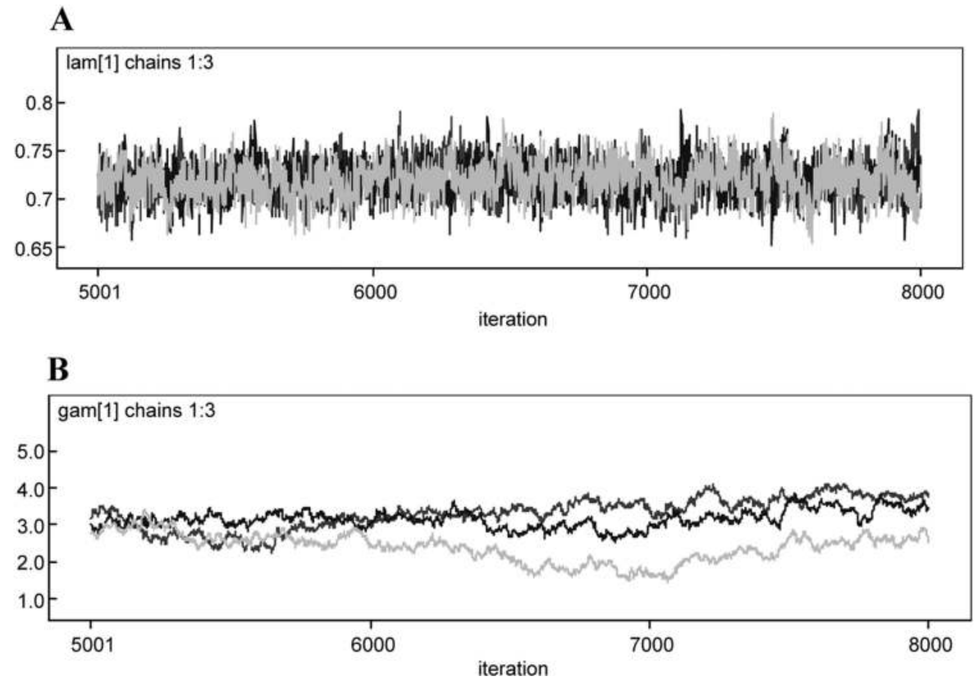

Figure 3.

Sample traces of chains: traces of three chains for a factor loading for which convergence looks reasonable (A) and traces of three chains for a regression coefficient that have not yet reached convergence (B).

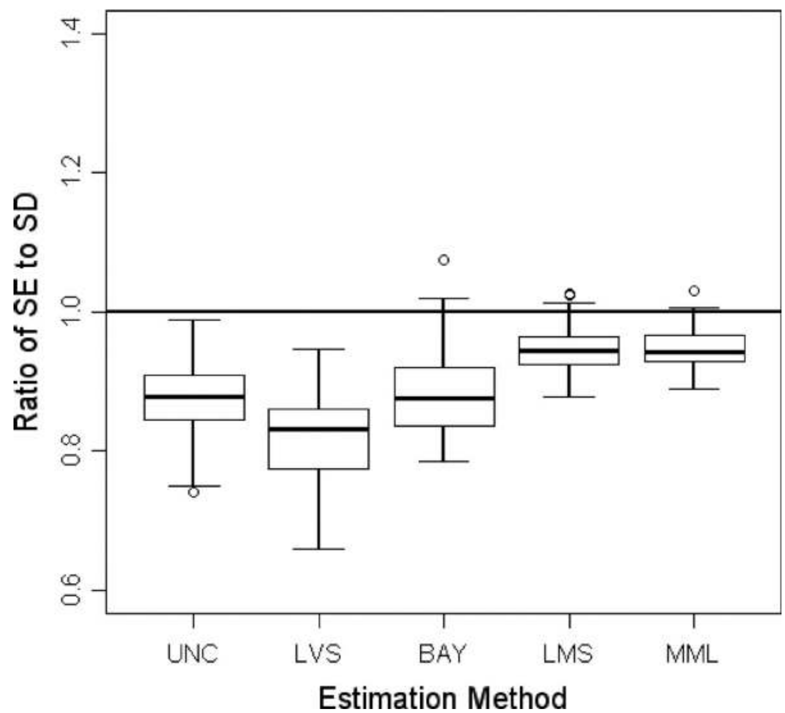

Figure 4.

Box plots for the estimated coefficients of f_2^2 in structural model (2) averaged across sample size, distributional, and reliability conditions. The five methods are unconstrained (UNC), latent variable score (LVS), Bayesian (BAY), latent moderated structural equations (LMS), and marginal maximum likelihood (MML). Error bars indicate standard errors.

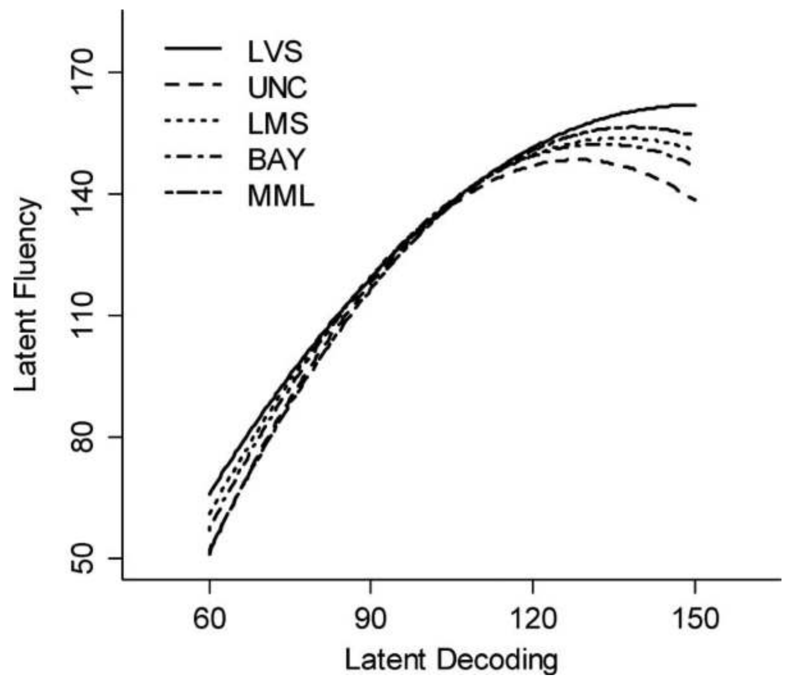

Figure 5.

Fitted quadratic structural model for latent fluency versus latent decoding across the five estimation methods. UNC = unconstrained method; LVS = latent variable score method; BAY = Bayesian method; LMS = latent moderated structural equation method; MML = marginal maximum likelihood method.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Summary of Manipulated Factors for Simulation Study 1 by Level

Factor	**1**	**2**	**3**	**4**	**5**	**6**
Sample size	50	100	250	500	1,000	
Indicator reliability	.45	.65	.85			
Observed variable distribution : Exogenous latent variable distribution	Normal : normal	Normal : skew = 2, kurtosis = 7	Normal : skew = 3, kurtosis = 20	Skew = 2, kurtosis = 7 : normal	Skew = 2, kurtosis = 7 : skew = 2, kurtosis = 7	Skew = 2, kurtosis = 7 : skew = 3, kurtosis = 20
Estimation method	LVS	UNC	LMS	MML	BAY	

Note. LVS = latent variable score; UNC = unconstrained; LMS = latent moderated structural equations; MML = marginal maximum likelihood; BAY = Bayesian.

Table 2

Simulation Study 1: Performance of Five Estimation Methods for Estimating γ_2 in Terms of Parameter Bias, Root Mean Square Error, and Standard Error Ratio Across Conditions of Sample Size and Indicator and Latent Variable Nonnormality

	NN-1			NN-2			NN-3			NN-4			NN-5			NN-6		
Method	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**
									n = 50									
UNC	0.145	0.167	0.917	0.151	0.171	0.917	0.156	0.179	0.920	0.143	0.140	0.919	0.144	0.168	0.820	0.169	0.170	0.906
LVS	–0.106	0.111	0.858	–0.105	0.103	0.856	–0.100	0.114	0.867	–0.103	0.099	0.860	– 0.094	0.100	0.870	–0.100	0.112	0.750
BAY	0.021	0.106	0.851	0.031	0.109	0.858	0.038	0.094	0.877	0.017	0.100	0.856	0.039	0.108	0.886	0.044	0.115	0.879
LMS	0.019	0.102	0.938	0.024	0.100	0.939	0.025	0.097	0.940	0.020	0.096	0.940	0.028	0.109	0.946	0.029	0.107	0.925
MML	0.021	0.109	0.945	0.028	0.099	0.943	0.026	0.099	0.956	0.022	0.098	0.940	0.028	0.110	0.953	0.031	0.107	0.905
									n = 100									
UNC	0.052	0.100	0.928	0.057	0.129	0.888	0.062	0.135	0.901	0.049	0.120	0.925	0.059	0.130	0.901	0.059	0.127	0.910
LVS	–0.097	0.091	0.869	–0.071	0.079	0.845	–0.093	0.109	0.858	–0.101	0.081	0.740	–0.074	0.097	0.857	–0.086	0.103	0.786
BAY	0.017	0.086	0.869	0.012	0.076	0.819	0.015	0.085	0.832	0.020	0.093	0.879	0.009	0.091	0.830	0.008	0.100	0.890
LMS	0.007	0.081	0.932	0.008	0.086	0.927	–0.004	0.091	0.917	0.005	0.090	0.936	0.010	0.098	0.939	–0.008	0.100	0.945
MML	0.006	0.079	0.941	0.010	0.088	0.940	–0.004	0.090	0.935	0.007	0.096	0.941	0.009	0.095	0.933	0.018	0.105	0.939
									n = 250									
UNC	0.018	0.071	0.889	0.020	0.073	0.860	0.032	0.081	0.849	0.014	0.076	0.890	0.016	0.069	0.913	0.018	0.062	0.925
LVS	–0.075	0.091	0.850	–0.069	0.077	0.851	–0.095	0.088	0.842	–0.068	0.073	0.859	–0.069	0.074	0.879	–0.080	0.083	0.802
BAY	0.006	0.078	0.858	0.004	0.069	0.901	0.002	0.078	0.901	0.003	0.064	0.851	0.003	0.061	0.908	0.003	0.073	0.918
LMS	0.003	0.054	0.922	–0.005	0.051	0.954	–0.004	0.055	0.955	0.003	0.082	0.919	–0.003	0.084	0.944	–0.002	0.070	0.991
MML	0.004	0.055	0.958	0.007	0.052	0.987	0.005	0.051	0.940	0.004	0.083	0.943	0.008	0.081	0.967	0.003	0.076	1.005
									n = 500									
UNC	0.004	0.049	0.885	0.011	0.046	0.857	0.019	0.049	0.854	0.003	0.049	0.881	0.007	0.040	0.890	0.008	0.044	0.885
LVS	–0.076	0.078	0.843	–0.068	0.081	0.815	–0.085	0.078	0.804	–0.071	0.066	0.836	–0.050	0.068	0.849	– 0.054	0.070	0.837
BAY	0.004	0.056	0.851	0.001	0.048	0.824	–0.002	0.052	0.814	0.002	0.048	0.826	0.002	0.054	0.850	0.001	0.056	0.868
LMS	0.002	0.041	0.919	<0.001	0.047	0.909	–0.002	0.041	0.936	0.002	0.050	0.956	0.001	0.054	0.932	–0.001	0.049	0.987
MML	0.002	0.045	0.932	<0.001	0.043	0.923	0.001	0.040	0.915	0.002	0.046	0.943	0.001	0.055	0.941	0.001	0.051	0.991
									n = 1,000									
UNC	0.001	0.037	0.804	0.004	0.034	0.822	0.003	0.037	0.844	<0.001	0.039	0.809	0.002	0.035	0.856	0.002	0.039	0.803

Method	NN-1 Bias	NN-1 RMSE	NN-1 Ratio	NN-2 Bias	NN-2 RMSE	NN-2 Ratio	NN-3 Bias	NN-3 RMSE	NN-3 Ratio	NN-4 Bias	NN-4 RMSE	NN-4 Ratio	NN-5 Bias	NN-5 RMSE	NN-5 Ratio	NN-6 Bias	NN-6 RMSE	NN-6 Ratio
LVS	–0.019	0.073	0.765	–0.028	0.076	0.812	–0.041	0.067	0.809	–0.021	0.060	0.779	–0.031	0.062	0.883	–0.041	0.064	0.870
BAY	0.001	0.049	0.821	<0.001	0.046	0.830	<0.001	0.061	0.850	–0.001	0.050	0.819	–0.001	0.053	0.870	<0.001	0.054	0.901
LMS	<0.001	0.039	0.901	0.001	0.035	0.915	<0.001	0.032	0.930	<0.001	0.045	0.900	0.001	0.051	0.948	<0.001	0.054	0.995
MML	<0.001	0.034	0.899	–0.001	0.031	0.930	<0.001	0.036	0.945	<0.001	0.043	0.890	0.001	0.050	0.952	<0.001	0.049	0.994

Note. Indicator reliability is .45. NN = nonnormality; RMSE = root-mean-square error; UNC = unconstrained; LVS = latent variable score; BAY = Bayesian; LMS = latent moderated structural equations; MML = marginal maximum likelihood.

Table 3

Simulation Study 1: Performance of Five Estimation Methods for Estimating γ_2 in Terms of Parameter Bias, Root Mean Square Error, and Standard Error Ratio Across Conditions of Sample Size and Indicator and Latent Variable Nonnormality

	NN-1			NN-2			NN-3			NN-4			NN-5			NN-6		
Method	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**
n = 50																		
UNC	0.120	0.131	0.925	0.134	0.130	0.913	0.128	0.133	0.920	0.101	0.127	0.922	0.111	0.129	0.897	0.151	0.134	0.896
LVS	–0.091	0.107	0.861	–0.092	0.103	0.852	–0.100	0.107	0.758	–0.055	0.087	0.722	–0.061	0.100	0.752	–0.072	0.109	0.730
BAY	0.010	0.091	0.852	0.080	0.089	0.849	0.005	0.090	0.821	0.072	0.083	0.861	0.062	0.101	0.901	0.081	0.101	0.852
LMS	0.005	0.090	0.947	0.008	0.091	0.937	0.001	0.086	0.918	0.053	0.081	0.917	0.061	0.104	0.964	0.062	0.100	0.923
MML	0.006	0.089	0.950	0.008	0.092	0.940	0.005	0.088	0.920	0.049	0.084	0.922	0.058	0.105	0.971	0.063	0.099	0.905
n = 100																		
UNC	0.028	0.126	0.941	0.030	0.132	0.878	0.020	0.128	0.989	0.014	0.114	0.922	0.014	0.128	0.873	0.019	0.129	0.903
LVS	–0.075	0.097	0.874	–0.071	0.094	0.838	–0.086	0.108	0.778	–0.062	0.089	0.735	–0.063	0.090	0.763	–0.067	0.094	0.728
BAY	0.008	0.080	0.872	0.013	0.082	0.819	0.001	0.086	0.787	0.057	0.099	0.845	0.056	0.097	0.892	0.057	0.100	0.860
LMS	0.004	0.085	0.951	0.008	0.086	0.931	–0.004	0.089	0.923	0.046	0.098	0.916	0.047	0.099	0.994	0.048	0.104	0.928
MML	0.005	0.087	0.974	0.010	0.087	0.945	–0.003	0.089	0.903	0.043	0.097	0.930	0.044	0.097	0.981	0.046	0.104	0.910
n = 250																		
UNC	0.007	0.065	0.864	0.005	0.061	0.859	0.011	0.076	0.862	0.003	0.058	0.846	0.002	0.059	0.764	0.012	0.066	0.831
LVS	–0.076	0.085	0.830	–0.077	0.085	0.840	–0.084	0.093	0.801	–0.061	0.070	0.791	–0.063	0.073	0.713	–0.063	0.073	0.741
BAY	0.007	0.077	0.889	0.005	0.076	0.897	0.001	0.085	0.809	0.055	0.096	0.891	0.050	0.095	0.859	0.057	0.099	0.878
LMS	<–0.001	0.050	0.952	–0.001	0.048	0.962	–0.002	0.052	0.972	0.048	0.067	1.024	0.043	0.065	0.950	0.051	0.071	1.026
MML	<0.001	0.051	0.970	<0.001	0.049	0.977	<0.001	0.052	0.948	0.046	0.066	1.030	0.040	0.063	0.967	0.048	0.069	0.993
n = 500																		
UNC	0.004	0.046	0.802	0.003	0.043	0.816	0.004	0.049	0.858	0.002	0.041	0.741	0.001	0.039	0.751	0.005	0.043	0.750
LVS	–0.076	0.081	0.769	–0.077	0.081	0.807	–0.088	0.092	0.771	–0.060	0.066	0.670	–0.060	0.066	0.673	–0.029	0.036	0.803
BAY	0.004	0.080	0.829	0.003	0.082	0.825	–0.007	0.086	0.786	0.057	0.102	0.826	0.055	0.100	0.809	0.061	0.104	0.832
LMS	<–0.001	0.037	0.897	<–0.001	0.036	0.916	–0.007	0.038	0.933	0.048	0.060	0.927	0.050	0.060	1.002	0.050	0.061	0.967
MML	<0.001	0.037	0.909	<0.001	0.036	0.932	–0.007	0.038	0.909	0.046	0.058	0.892	0.046	0.057	0.957	0.047	0.059	0.919
n = 1,000																		
UNC	0.001	0.038	0.800	0.002	0.031	0.824	0.003	0.034	0.872	0.001	0.037	0.762	0.001	0.039	0.786	0.005	0.041	0.799

	NN-1			NN-2			NN-3			NN-4			NN-5			NN-6		
Method	Bias	RMSE	Ratio	Bias	RMSE	Ratio	Bias	RMSE	Ratio	Bias	RMSE	Ratio	Bias	RMSE	Ratio	Bias	RMSE	Ratio
LVS	–0.025	0.073	0.754	–0.028	0.075	0.816	–0.041	0.093	0.786	–0.031	0.066	0.694	–0.038	0.063	0.719	–0.019	0.036	0.794
BAY	0.002	0.074	0.841	0.002	0.080	0.835	–0.004	0.090	0.830	0.032	0.099	0.837	0.048	0.094	0.832	0.052	0.100	0.884
LMS	<0.001	0.026	0.901	<0.001	0.035	0.924	–0.003	0.038	0.932	0.049	0.051	0.925	0.049	0.059	1.005	0.045	0.054	0.993
MML	<0.001	0.027	0.908	<0.001	0.035	0.929	–0.005	0.036	0.911	0.041	0.057	0.913	0.046	0.059	0.987	0.042	0.056	1.001

Note. Indicator reliability is .65. NN = nonnormality; RMSE = root-mean-square error; UNC = unconstrained; LVS = latent variable score; BAY = Bayesian; LMS = latent moderated structural equations; MML = marginal maximum likelihood.

Table 4

Simulation Study 1: Performance of Five Estimation Methods for Estimating γ_2 in Terms of Parameter Bias, Root Mean Square Error, and Standard Error Ratio Across Conditions of Sample Size and Indicator and Latent Variable Nonnormality

	NN-1			NN-2			NN-3			NN-4			NN-5			NN-6		
Method	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**
n = 50																		
UNC	0.006	0.082	0.918	0.007	0.083	0.916	0.006	0.078	0.908	0.003	0.079	0.918	0.003	0.086	0.910	0.004	0.088	0.900
LVS	–0.040	0.091	0.890	–0.039	0.094	0.853	–0.051	0.103	0.843	–0.041	0.099	0.852	–0.039	0.100	0.848	–0.051	0.104	0.813
BAY	0.009	0.076	0.921	0.010	0.079	0.950	0.010	0.082	0.951	0.008	0.076	0.950	0.009	0.082	0.975	0.062	0.096	0.976
LMS	0.005	0.078	0.941	0.006	0.081	0.952	0.007	0.081	0.953	0.026	0.070	0.879	0.072	0.085	0.949	0.082	0.099	0.909
MML	0.003	0.071	0.950	0.005	0.072	0.955	0.005	0.076	0.932	0.022	0.070	0.899	0.069	0.082	0.947	0.080	0.107	0.904
n = 100																		
UNC	0.004	0.069	0.906	0.002	0.066	0.920	0.007	0.070	0.917	0.003	0.072	0.845	0.002	0.068	0.872	–0.003	0.070	0.851
LVS	–0.038	0.066	0.926	–0.038	0.064	0.946	–0.038	0.066	0.919	–0.030	0.063	0.838	–0.031	0.061	0.881	–0.036	0.067	0.818
BAY	0.011	0.061	0.930	0.011	0.057	0.976	0.013	0.063	0.898	0.037	0.070	0.954	0.038	0.069	1.011	0.033	0.068	0.986
LMS	–0.002	0.061	0.949	–0.002	0.057	1.006	<–0.001	0.063	0.943	0.020	0.068	0.889	0.021	0.065	0.954	0.015	0.067	0.893
MML	–0.001	0.062	0.954	<0.001	0.059	0.983	0.001	0.063	0.940	0.019	0.068	0.909	0.021	0.066	0.949	0.016	0.066	0.924
n = 250																		
UNC	0.003	0.041	0.871	–0.001	0.041	0.877	0.005	0.040	0.935	<0.001	0.040	0.818	<0.001	0.037	0.848	0.003	0.040	0.814
LVS	–0.034	0.048	0.867	–0.038	0.051	0.873	–0.036	0.049	0.908	–0.030	0.044	0.796	–0.029	0.043	0.809	–0.030	0.045	0.774
BAY	0.017	0.061	0.924	0.010	0.064	0.883	0.013	0.064	0.892	0.035	0.070	0.976	0.039	0.068	1.003	0.041	0.069	1.020
LMS	0.002	0.038	0.948	–0.003	0.038	0.944	0.001	0.036	1.001	0.021	0.043	0.933	0.020	0.041	0.956	0.021	0.042	0.934
MML	0.001	0.038	0.940	–0.003	0.038	0.948	0.001	0.037	0.986	0.022	0.043	0.936	0.020	0.041	0.943	0.022	0.043	0.931
n = 500																		
UNC	0.001	0.028	0.885	0.001	0.028	0.885	0.001	0.028	0.910	–0.001	0.027	0.801	0.001	0.025	0.849	0.003	0.026	0.848
LVS	–0.035	0.042	0.900	–0.036	0.043	0.885	–0.039	0.046	0.866	–0.030	0.038	0.769	–0.029	0.036	0.812	–0.063	0.069	0.660
BAY	0.017	0.062	0.904	0.014	0.063	0.909	0.013	0.064	0.875	0.033	0.068	0.977	0.032	0.063	1.074	0.038	0.070	0.995
LMS	<0.001	0.026	0.970	<–0.001	0.026	0.967	–0.002	0.026	0.978	0.020	0.032	0.951	0.022	0.033	0.999	0.023	0.033	1.009
MML	<0.001	0.026	0.946	–0.001	0.028	0.909	–0.002	0.027	0.960	0.019	0.032	0.933	0.022	0.032	1.007	0.022	0.033	0.987
n = 1,000																		
UNC	<0.001	0.019	0.892	0.001	0.020	0.891	0.003	0.023	0.872	0.001	0.021	0.872	0.002	0.018	0.860	0.004	0.028	0.829

	NN-1			NN-2			NN-3			NN-4			NN-5			NN-6		
Method	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**	**Bias**	**RMSE**	**Ratio**
LVS	0.010	0.035	0.910	–0.019	0.039	0.916	–0.024	0.040	0.902	–0.014	0.042	0.821	–0.023	0.038	0.834	–0.020	0.051	0.770
BAY	0.001	0.031	0.925	0.002	0.043	0.921	–0.004	0.050	0.870	0.026	0.061	0.966	0.027	0.071	1.014	0.049	0.065	0.991
LMS	<0.001	0.021	1.001	<0.001	0.024	0.973	–0.001	0.029	0.963	0.042	0.031	0.961	0.046	0.025	1.012	0.048	0.028	1.001
MML	<0.001	0.019	0.995	<0.001	0.024	0.979	0.002	0.028	0.970	0.045	0.033	0.954	0.040	0.031	0.992	0.050	0.035	1.005

Note. Indicator reliability is .85. NN = nonnormality; RMSE = root-mean-square error; UNC = unconstrained; LVS = latent variable score; BAY = Bayesian; LMS = latent moderated structural equations; MML = marginal maximum likelihood.

Table 5

Type I Error Rates ($R^2_{\gamma_2}=0$) Across Study Conditions

	n = 50			*n* = 100			*n* = 250			*n* = 500			*n* = 1,000		
Method	**N**	**NN1**	**NN2**	**N**	**NN1**	**NN2**	**N**	**NN1**	**NN2**	**N**	**NN1**	**NN2**	**N**	**NN1**	**NN2**
								Reliability = .45							
LVS	.06	.07	*.08*	.06	.07	.07	.04	.05	.05	.03	.04	.06	.04	.05	.06
UNC	*.12*	*.11*	*.12*	*.10*	*.09*	*.08*	*.09*	*.09*	.07	.06	*.08*	.07	.05	.05	.06
LMS	*.08*	*.10*	*.15*	*.08*	*.10*	*.11*	*.07*	*.09*	*.10*	.05	.07	*.08*	.04	.07	.07
MML	*.09*	*.10*	*.12*	.07	*.11*	*.12*	.06	*.10*	*.11*	.05	*.10*	*.12*	.05	*.09*	*.10*
BAY	*.10*	*.11*	*.12*	*.10*	*.10*	*.12*	*.09*	*.10*	*.11*	*.07*	*.08*	*.10*	.06	*.08*	.07
								Reliability = .65							
LVS	.05	.05	.07	.04	.06	.07	.03	.05	.06	.03	.05	.06	*.02*	.04	.05
UNC	*.10*	*.10*	*.12*	*.08*	*.08*	*.09*	.07	.06	*.08*	.05	.07	*.08*	.04	.05	.06
LMS	.07	*.09*	*.13*	.07	*.08*	*.11*	.06	*.08*	*.09*	.04	.05	.07	.03	.05	.06
MML	*.08*	*.11*	*.13*	.07	*.10*	*.12*	.05	*.08*	*.09*	.03	.07	*.08*	.03	.05	.07
BAY	*.09*	*.10*	*.11*	*.09*	*.12*	*.09*	.07	.07	*.09*	.07	*.08*	*.11*	.07	*.09*	.07
								Reliability = .85							
LVS	.04	.05	.07	.03	.05	.06	.03	.05	.05	.04	.04	.05	.03	.04	.05
UNC	.07	*.08*	*.10*	.06	.07	*.09*	.05	.06	.06	.04	.05	.06	.04	.04	.05
LMS	.07	*.09*	*.10*	.06	*.09*	*.09*	.05	.06	*.09*	.04	.05	.06	.03	.04	.06
MML	*.08*	*.10*	*.10*	.06	*.08*	*.09*	.05	.06	.07	.05	.07	.07	.04	.06	.06
BAY	*.08*	*.09*	*.10*	.07	*.09*	*.08*	.07	.07	*.08*	.06	*.08*	*.09*	.05	.07	.06

Note. Values in italics indicate conditions for which the Type I error rate was unacceptable. Normality conditions: N = normal; NN = nonnormality; NN1 = skew = 2, kurtosis = 7; NN2 = skew = 3, kurtosis = 20. Estimation methods: LVS = latent variable score; UNC = unconstrained; LMS = latent moderated structural equations; MML = marginal maximum likelihood; BAY = Bayesian.

Table 6

Empirical Power ($R^2_{\gamma_2}$=.05) Across Study Conditions

	n = 50			*n* = 100			*n* = 250			*n* = 500			*n* = 1,000		
Method	**N**	**NN1**	**NN2**	**N**	**NN1**	**NN2**	**N**	**NN1**	**NN2**	**N**	**NN1**	**NN2**	**N**	**NN1**	**NN2**
Reliability = .45															
LVS	.78	.80	.81	.79	.77	.80	.81	.81	.82	.83	.82	.80	.81	.81	.82
UNC	.71	.72	.73	.70	.71	.74	.75	.78	.80	.80	.83	.85	.82	.80	.85
LMS	.80	.81	.80	.79	.84	.84	.81	.82	.87	.85	.83	.84	.86	.86	.84
MML	.83	.86	.82	.82	.87	.86	.84	.84	.87	.87	.87	.88	.87	.86	.86
BAY	.96	.97	.98	.98	.99	.99	.98	.98	.99	.99	.97	.99	1.00	1.00	1.00
Reliability = .65															
LVS	.84	.85	.85	.83	.87	.88	.86	.86	.87	.89	.89	.89	.90	.90	.88
UNC	.87	.85	.87	.89	.88	.88	.90	.88	.91	.90	.90	.92	.89	.90	.91
LMS	.87	.87	.85	.89	.90	.90	.92	.91	.92	.90	.93	.97	.99	.99	1.00
MML	.88	.86	.87	.88	.92	.90	.94	.91	.93	.94	.95	.98	.99	1.00	.99
BAY	.98	.99	1.00	1.00	.96	.98	.99	1.00	.99	1.00	1.00	1.00	1.00	1.00	1.00
Reliability = .85															
LVS	.89	.90	.91	.91	.92	.90	.90	.92	.92	.93	.95	.94	.97	.99	.99
UNC	.90	.92	.93	.92	.92	.93	.95	.97	.97	.95	.99	.98	.97	1.00	1.00
LMS	.90	.90	.91	.94	.95	.94	.96	.98	.98	.97	.96	.98	1.00	1.00	1.00
MML	.90	.91	.90	.93	.95	.95	.95	.96	.96	.97	.98	.98	.99	1.00	1.00
BAY	.98	.99	.98	.98	.99	.99	.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note. Normality conditions: N = normal; NN = nonnormality; NN1 = skew = 2, kurtosis = 7; NN2 = skew = 3, kurtosis = 20. Estimation methods: LVS = latent variable score; UNC = unconstrained; LMS = latent moderated structural equations; MML = marginal maximum likelihood; BAY = Bayesian.

Table 7

Empirical Power ($R^2_{\gamma_2}=.10$) Across Study Conditions

Method	n = 50 N	n = 50 NN1	n = 50 NN2	n = 100 N	n = 100 NN1	n = 100 NN2	n = 250 N	n = 250 NN1	n = 250 NN2	n = 500 N	n = 500 NN1	n = 500 NN2	n = 1,000 N	n = 1,000 NN1	n = 1,000 NN2
Reliability = .45															
LVS	.89	.89	.90	.90	.90	.92	.95	.96	.95	.95	.96	.98	.94	.98	.98
UNC	.87	.88	.90	.88	.93	.93	.91	.92	.89	.90	.94	.96	.94	.99	.99
LMS	.84	.85	.86	.85	.85	.84	.85	.88	.89	.90	.88	.91	.89	.93	.93
MML	.84	.86	.87	.86	.85	.84	.89	.88	.90	.91	.90	.90	.91	.92	.91
BAY	.99	.99	.98	.98	1.00	.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Reliability = .65															
LVS	.90	.89	.90	.92	.92	.91	.93	.92	.92	.99	.99	1.00	1.00	1.00	1.00
UNC	.91	.93	.92	.92	.93	.95	.95	.93	.96	.98	.98	.99	1.00	1.00	1.00
LMS	.92	.91	.90	.91	.92	.91	.93	.97	.97	.98	1.00	.99	1.00	1.00	1.00
MML	.90	.90	.94	.93	.91	.94	.95	.97	.99	.99	1.00	1.00	1.00	1.00	1.00
BAY	1.00	1.00	1.00	1.00	1.00	1.00	.99	1.00	.99	1.00	1.00	1.00	1.00	1.00	1.00
Reliability = .85															
LVS	.90	.90	.94	.94	.97	.96	.95	.95	.98	.99	.99	1.00	.99	1.00	1.00
UNC	.93	.93	.95	.95	.96	.98	.97	.99	.99	.99	1.00	1.00	1.00	1.00	1.00
LMS	.95	.94	.95	.95	.95	.97	.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MML	.94	.97	.97	.96	.98	.98	.99	.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
BAY	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Note. Normality conditions: N = normal; NN = nonnormality; NN1 = skew = 2, kurtosis = 7; NN2 = skew = 3, kurtosis = 20. Estimation methods: LVS = latent variable score; UNC = unconstrained; LMS = latent moderated structural equations; MML = marginal maximum likelihood; BAY = Bayesian.

Table 8

Estimates and Standard Errors From the Five Estimation Methods for the Quadratic Model

	LVS		UNC		LMS		BAY		MML	
Parameter	**Estimate**	***SE***	**Estimate**	***SE***	**Estimate**	***SE***	**Estimate**	***SE***	**Estimate**	***SE***
β_0	132.43	0.57	131.86	2.07	132.80	1.50	132.96	1.64	131.30	1.47
β_1	1.19	0.04	1.18	0.07	1.16	0.10	1.18	0.08	1.22	0.08
β_2	–0.012	0.003	–0.021	0.006	–0.016	0.005	–0.018	0.004	–0.017	0.005

Note. LVS = latent variable score; UNC = unconstrained; LMS = latent moderated structural equations; BAY = Bayesian; MML = marginal maximum likelihood.