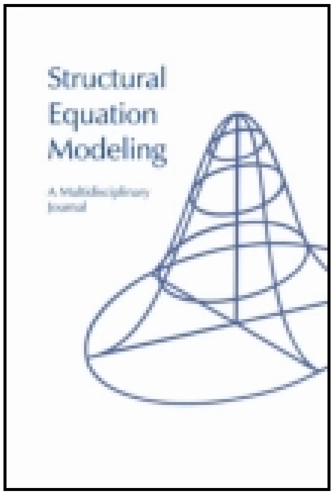

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

A Nonlinear Structural Equation Mixture Modeling Approach for Nonnormally Distributed Latent Predictor Variables

Augustin Kelava[a], Benjamin Nagengast[a] & Holger Brandt[a]
[a] Eberhard Karls Universität Tübingen, Tübingen, Germany
Published online: 05 Jun 2014.

To cite this article: Augustin Kelava, Benjamin Nagengast & Holger Brandt (2014) A Nonlinear Structural Equation Mixture Modeling Approach for Nonnormally Distributed Latent Predictor Variables, Structural Equation Modeling: A Multidisciplinary Journal, 21:3, 468-481, DOI: 10.1080/10705511.2014.915379

To link to this article: http://dx.doi.org/10.1080/10705511.2014.915379

Structural Equation Modeling: A Multidisciplinary Journal, 21: 468–481, 2014

ISSN: 1070-5511 print / 1532-8007 online
DOI: 10.1080/10705511.2014.915379

TEACHER'S CORNER

A Nonlinear Structural Equation Mixture Modeling Approach for Nonnormally Distributed Latent Predictor Variables

Augustin Kelava, Benjamin Nagengast, and Holger Brandt
Eberhard Karls Universität Tübingen, Tübingen, Germany

Structural equation models with interaction and quadratic effects have become a standard tool for testing nonlinear hypotheses in the social sciences. Most of the current approaches assume normally distributed latent predictor variables. In this article, we describe a nonlinear structural equation mixture approach that integrates the strength of parametric approaches (specification of the nonlinear functional relationship) and the flexibility of semiparametric structural equation mixture approaches for approximating the nonnormality of latent predictor variables. In a comparative simulation study, the advantages of the proposed mixture procedure over contemporary approaches [Latent Moderated Structural Equations approach (LMS) and the extended unconstrained approach] are shown for varying degrees of skewness of the latent predictor variables. Whereas the conventional approaches show either biased parameter estimates or standard errors of the nonlinear effects, the proposed mixture approach provides unbiased estimates and standard errors. We present an empirical example from educational research. Guidelines for applications of the approaches and limitations are discussed.

Keywords: estimators, interaction, latent variables, mixture distribution, moderator, nonlinear structural equation mixture models, quadratic, unconstrained approach

In the social and behavioral sciences linear regression has been a standard tool for modeling linear relationships between variables (e.g., Cohen, Cohen, West, & Aiken, 2003). If a relationship between two variables depends on the size of a third variable (the so-called moderator), the inclusion of products of variables is often used to model the nonlinear relationship (Aiken & West, 1991). Equation 1 shows a regression equation with one dependent variable (y), one predictor variable (x_1), and one moderator (x_2):

$$y = \alpha + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_1 x_2 + \zeta = \alpha + (\gamma_1 + \gamma_3 x_2)\, x_1 + \gamma_2 x_2 + \zeta \tag{1}$$

where α is an intercept, the γs are regression coefficients, and ζ is an error of prediction. As can be seen, Equation 1 can be rearranged such that the relationship of y and x_1 is expressed as a function of x_2. This means that the new coefficient $(\gamma_1 + \gamma_3 x_2)$ is not a constant and the relationship of y and x_1 is nonlinear. When a relationship between the dependent variable (y) and the predictor variable (x_1) is not constant, but changes by the amount of the predictor variable (x_1), the resulting model is a quadratic:

$$y = \alpha + \gamma_1 x_1 + \gamma_2 x_1^2 + \zeta = \alpha + (\gamma_1 + \gamma_2 x_1)\, x_1 + \zeta \tag{2}$$

Most variables in the social and behavioral sciences are measured with less than perfect reliability, which leads to biased estimates of regression coefficients (Carroll, Ruppert, Stefanski, & Crainiceanu, 2006; Lord & Novick, 1968). The effects of product variables including interaction and quadratic terms are particularly affected (e .g., x_1x_2, x_1^2, x_2^2; Bohrnstedt & Marwell, 1978; MacCallum & Mar, 1995).

Correspondence should be addressed to Augustin Kelava, Eberhard Karls Universität Tübingen, Center for Educational Science and Psychology, Europastr. 6, 72072 Tübingen, Germany. E-mail: augustin.kelava@uni-tuebingen.de

Structural equation models overcome this problem by controlling for measurement error and can be used to obtain unbiased estimates of the relationships of the latent entities (Bollen, 1989). Structural equation models with latent product variables have become part of the standard methodology over the last decade. The numerous *parametric approaches* for the estimation of nonlinear effects (for a review see Algina & Moulder, 2001; Marsh, Wen, & Hau, 2004, 2006; Schumacker & Marcoulides, 1998), including product indicator approaches (e.g., Algina & Moulder, 2001; Bollen, 1995; Jaccard & Wan, 1995; Jöreskog & Yang, 1996; Kelava & Brandt, 2009; Kenny & Judd, 1984; Little, Bovaird, & Widaman, 2006; Marsh et al., 2004, 2006; Ping, 1995), distribution analytic approaches (Klein & Moosbrugger, 2000; Klein & Muthen, 2007), Bayesian approaches (e.g., Arminger & Muthen, 1998; Lee, Song, & Tang, 2007), and method of moments based approaches (Mooijaart & Bentler, 2010; Wall & Amemiya, 2003) are a testament to this development. In parametric nonlinear structural equation models, the following latent structural model with one latent dependent variable (η) and two latent predictor variables (ξ_1, ξ_2) including one interaction and two quadratic effects is typically estimated:

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_1\xi_2 + \gamma_4\xi_1^2 + \gamma_5\xi_2^2 + \zeta \qquad (3)$$

One substantial drawback is that product indicator approaches (e.g., Jaccard & Wan, 1995; Jöreskog & Yang, 1996; Kelava & Brandt, 2009; Kenny & Judd, 1984; Little et al., 2006; Marsh, Little, Bovaird, & Widaman, 2007; Marsh et al., 2004, 2006; Ping, 1995), distribution analytic approaches (Klein & Moosbrugger, 2000; Klein & Muthén, 2007), and some Bayesian approaches (e.g., Lee et al., 2007; Song & Lu, 2010), usually assume that the latent linear predictor variables (e.g., ξ_1, ξ_2) are multivariate normally distributed. When this assumption is violated, parameter estimates in the nonlinear structural model will be biased because of additional shared variance between the latent first-order and the product variables that is not accounted for in the model (Kelava et al., 2011; Marsh et al., 2004). Some parametric approaches (e.g., the weighted least squares approach of Joreskog & Yang, 1996, and the two-step least squares approach of Bollen, 1995) that have used alternative asymptotically distribution-free estimators do not perform well in finite samples, leading to large standard errors and a low power for detecting the effects (Schermelleh-Engel, Klein, & Moosbrugger, 1998).

One conceptually different approach to modeling nonlinear relations among latent variables is the use of *semiparametric* structural equation mixture models (SEMM; Arminger & Stein, 1997; Arminger, Stein, & Wittenberg, 1999; Bauer, 2005; Bauer & Curran, 2004; Dolan & van der Maas, 1998; Jedidi, Jagpal, & DeSarbo, 1997a, 1997b; Muthen, 2001; Pek, Losardo, & Bauer, 2011; Pek, Sterba, Kok, & Bauer, 2009). Here, finite mixtures of linear structural equation models are used to approximate the unknown nonlinear relationship of the latent dependent and predictor variables. Whereas parametric approaches specify the functional form of their relationship a priori (see Equation 3), the SEMM approach does not require assumptions about this functional form. In addition, the SEMM approach does not require the assumption of normally distributed latent variables and disturbances inherent in conventional structural equation modeling (SEM), but allows for flexible approximations of nonnormal distributions. Hence, the SEMM approach is a flexible tool for predicting the latent dependent variable when there is nonnormality and when obtaining a strict parametric representation of the functional relation does not have the highest priority (for a discussion, see Bauer, 2005). Another way of approximating nonnormal latent predictor distributions has been developed in the context of semiparametric Bayesian SEM (e.g., Chow, Tang, Yuan, & Song, 2011; Yang & Dunson, 2010). In these approaches, mixtures of discrete or continuous distributions are used as prior distributions to flexibly approximate the multivariate latent predictor distribution. In contrast, the structural model is parametric. On the one hand, these approaches have highly desirable properties in terms of approximating an arbitrary multivariate distribution of the latent predictors, but on the other hand, the interpretation of the latent variables and identifiability require centering procedures for discrete and continuous mixtures (Yang & Dunson, 2010). Their transfer to structural models with nonlinear effects is not straightforward (for the nonmixture nonlinear SEM case, see Moosbrugger, Schermelleh-Engel, & Klein, 1997) and has not been discussed in non-Bayesian and Bayesian nonlinear latent variable mixture models.

Little is known about the importance of the normality assumption for latent variables with respect to SEM. Researchers in psychology, education, and other social sciences typically do not explicate their assumptions about the distributional form of the constructs they are studying, conveniently assuming that normality will be appropriate. However, this assumption is unlikely to hold for each and every variable. For example, well-known tendencies for self-enhancement (e.g., reviewed by Sedikides & Gregg, 2008) are likely to skew the distribution of self-evaluations such as academic self-concepts (Marsh, 2007) to the right. People are more likely to perceive themselves in a positive way than would be expected from a strictly symmetric normal distribution. Although nonnormal latent variable modeling has received increasing attention over the last few years (e.g., Bauer, 2005; Bauer & Curran, 2004; Curran, West, & Finch, 1996; Dolan & van der Maas, 1998), systematic evaluations of its consequences and model alternatives, particularly for parametric models with nonlinear relations between latent variables, are rare.

With respect to the parametric and semiparametric approaches already presented, non-Bayesian approaches are needed that unify the strengths of semiparametric approaches to approximate nonnormal distributions and the strengths of parametric approaches to specify a simple functional nonlinear relationship between latent variables.

AIMS OF THE ARTICLE

In this article, we present a model that approximates the multivariate nonnormal distribution of the latent predictor variables using normal mixtures. We begin with a didactic introduction into manifest and latent variable mixture modeling. Next, we extend structural equation mixture models with a nonlinear structural model and discuss how this comprehensive nonlinear structural equation mixture model (NSEMM) compares with semiparametric approaches to modeling nonlinear relations between latent variables and to modeling latent variable distributions. In a comparative simulation study, we show the properties of NSEMM for varying degrees of nonnormality demonstrating the power and flexibility of the normal mixture approach to modeling nonnormality. An empirical example from educational psychology is used as an illustration. The article concludes with a discussion of the approaches, giving specific recommendations for applications and pointing out directions for future research.

MIXTURES IN OBSERVED AND LATENT VARIABLE MODELS

In this section, we first introduce mixture variables as a useful tool for the approximation of nonnormal variables or variables coming from heterogeneous subpopulations. We then show that the concept of mixture variables can be transferred to latent variable models either to detect unobserved groups with varying relationships of latent variables or to approximate nonlinear relationships.

Mixture Variables

The basic idea in mixture modeling is to represent a given nonnormal distribution as a combination of two or more normal distributions having different means and variances[1]. The density of the nonnormal variable is then a weighted sum of two or more normal densities. An obvious example is the distribution of the height of adults in the United States, which is (slightly) nonnormal in the total population (skewness = 0.0277, kurtosis = 2.8431). This nonnormality can be explained by gender: Both males' ($M = 69.41$, $SD = 4.48$ in.) and females' heights ($M = 63.86$, $SD = 4.39$ in., McDowell, Fryar, Ogden, & Flegal, 2008) follow approximately normal distributions. The sex ratio of males (49.24%) to females (50.76%) is 0.97, hence combining these two conditional distributions into one distribution yields the nonnormal distribution in the entire population. The mixture density is given in Equation 4 and illustrated in Figure 1:

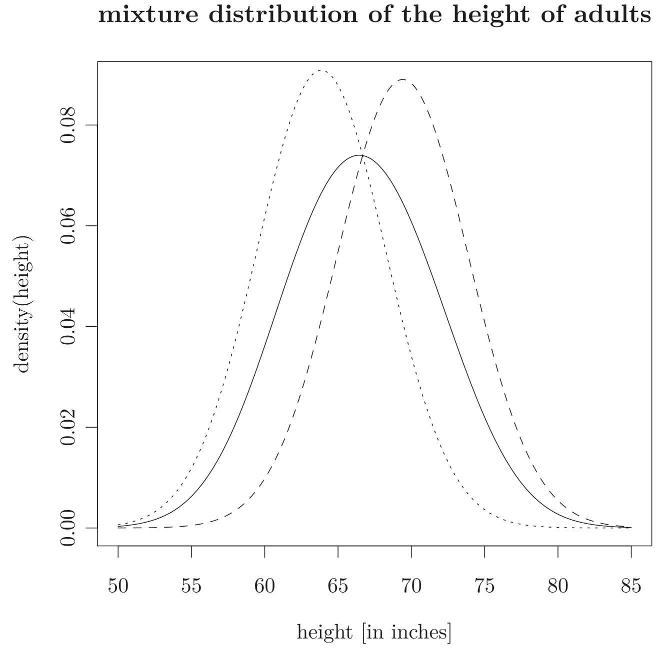

FIGURE 1 Mixture density of the height of adults in the United States (solid line). The dotted line represents the (unweighted) distribution of the height of females (with $M = 63.86$, $SD = 4.39$ in.). The dashed line represents the (unweighted) distribution of the height of males (with $M = 69.41$, $SD = 4.48$ in.). Equation 4 describes the mixture distribution in the population.

$$\begin{aligned} \text{density (height)} = 0.4924 \cdot N(69.41, 4.48) \\ + 0.5076 \cdot N(63.86, 4.39) \end{aligned} \tag{4}$$

Mixtures of normal distributions can be used to approximate a variety of nonnormal distributions including highly skewed or multimodal distributions (for numerous examples, see McLachlan & Peel, 2000). Figure 2 shows two distributions that are left skewed. The distribution in the left panel of Figure 2 is unimodal and the distribution in the right panel of Figure 2 is bimodal. Both are based on weighted sums of normal distributions illustrating the flexibility of the mixture approach.

The concept of structural equation mixture models. In nonmixture structural equation model multivariate relationships of observed and latent variables are examined (Bollen, 1989). Typically, it is assumed that the multivariate normal distribution of the indicator vectors $\mathbf{y} = (y_1, y_2, \ldots, y_p)'$ and $\mathbf{x} = (x_1, x_2, \ldots, x_q)'$ can be explained

[1]In principle, other distributions than the normal distribution could be used as mixture distributions (e.g., gamma). For reasons of simplicity and its flexibility we use the normal distribution here.

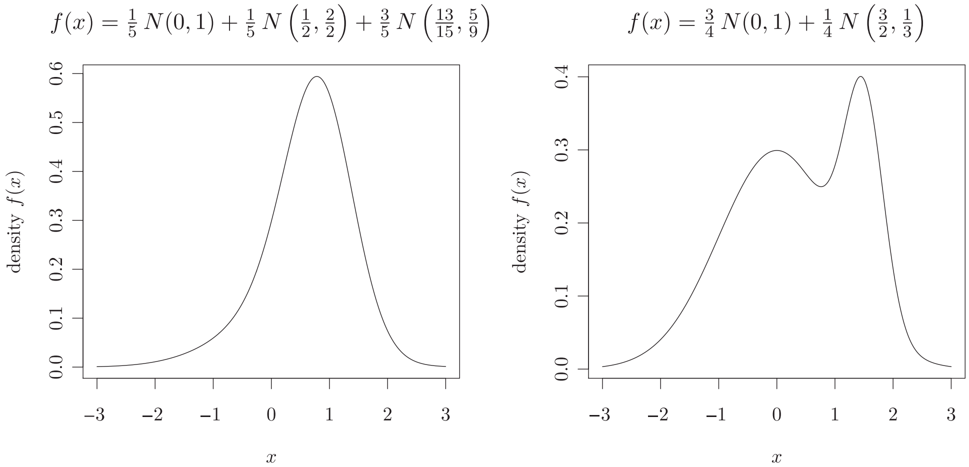

FIGURE 2 Two left skewed mixture distributions with given densities. The first distribution is unimodal and the second distribution is bimodal.

by underlying $(m \times 1)$ latent dependent variables $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_m)'$ and $(n \times 1)$ latent independent variables $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)'$ and the relations among them. Here, the observed variables $\boldsymbol{y}$ are indicators of $\boldsymbol{\eta}$ and the observed variables $\boldsymbol{x}$ are indicators of $\boldsymbol{\xi}$. According to the distributional assumptions and specified measurement, $\boldsymbol{y}$ and $\boldsymbol{x}$ jointly follow a multivariate normal distribution with the density:

$$f\left((\boldsymbol{y}', \boldsymbol{x}')'\right) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{5}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the model implied mean structure and covariance matrix, respectively. As can be seen from Equation 5, in (nonmixture) linear structural equation models, the joint distribution of $\boldsymbol{y}$ and $\boldsymbol{x}$ is described by a single multivariate normal component representing one population, as opposed to Equation 4, where the distribution of height was a weighted sum of two normal components representing the male and female populations.

It is straightforward to extend the concept of mixture variables to latent variable modeling where mixtures can be used to represent separate expected value vectors and variance–covariance matrices in multiple subpopulations. By analogy to the example of male and female heights (see Equation 4), Equation 5 can be generalized to a density function of a multivariate normal mixture model, the SEMM (Arminger & Stein, 1997; Arminger et al., 1999; Bauer, 2005; Bauer & Curran, 2004; Dolan & van der Maas, 1998; Jedidi et al., 1997a, 1997b; Muthén, 2001; Pek et al., 2011, Pek et al., 2009):

$$f\left((\boldsymbol{y}', \boldsymbol{x}')'\right) = \sum\nolimits_{g=1}^{G} w_g N\left(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g\right) \tag{6}$$

where G is the number of mixture components, w_g are the mixture probabilities (with $w_g > 0$ and $\sum_{g=1}^{G} w_g = 1$). In each mixture g, $\boldsymbol{y}$, and $\boldsymbol{x}$ follow a mixture-specific multivariate normal distribution with an expectation μ_g and covariance matrix Σ_g.

In the measurement model $\boldsymbol{y}|g$ and $\boldsymbol{x}|g$ denote the observed variables represented by the g-th mixture component:

$$y\,|g = \tau_g^y + \lambda_g^y \eta_g + \epsilon_g \qquad x\,|g = \tau_g^x + \lambda_g^x \xi_g + \delta_g \tag{7}$$

where $\tau_g^y\,(p \times 1)$ and $\tau_g^x\,(q \times 1)$ are intercepts, $\lambda_g^y\,(p \times m)$ and $\lambda_g^x\,(q \times n)$ are factor loading matrices, and $\varepsilon_g\,(p \times 1)$ and $\delta_g\,(q \times 1)$ are centered normally distributed measurement error variables with covariance matrices Θ_g^{ε} and Θ_g^{δ}, respectively. The measurement model is identical to the measurement model for nonmixture SEM, except that all parameters and variables hold for the mixture component g. Within each mixture g, the latent predictor variables ξ_g are normally distributed with expectation μ^{ξ_g} and covariance matrix Φ^{ξ_g}.

Accordingly, for each mixture g, a separate structural model can be defined that describes the relationship of the latent dependent and independent variables:

$$\eta_g = \mathrm{B}_g \eta_g + \alpha_g + \Gamma_g \xi_g + \zeta_g \tag{8}$$

where $\mathbf{B}_g\,(m \times m)$ is a coefficient matrix denoting the effects among the latent dependent variables $\eta_g, \alpha_g\,(m \times 1)$ is a vector of latent intercepts, $\boldsymbol{\Gamma}_g\,(m \times n)$s is a coefficient matrix denoting the effect(s) of ξ_g on η_g, and $\zeta_g\,(m \times 1)$ is a vector of normally distributed latent disturbances. It is obvious from the preceding definitions that SEMMs are a very flexible tool for accounting for heterogeneous populations. In principle, both the structural and the measurement model could differ entirely between the different mixture components g; for example, allowing for different dimensionalities of the latent

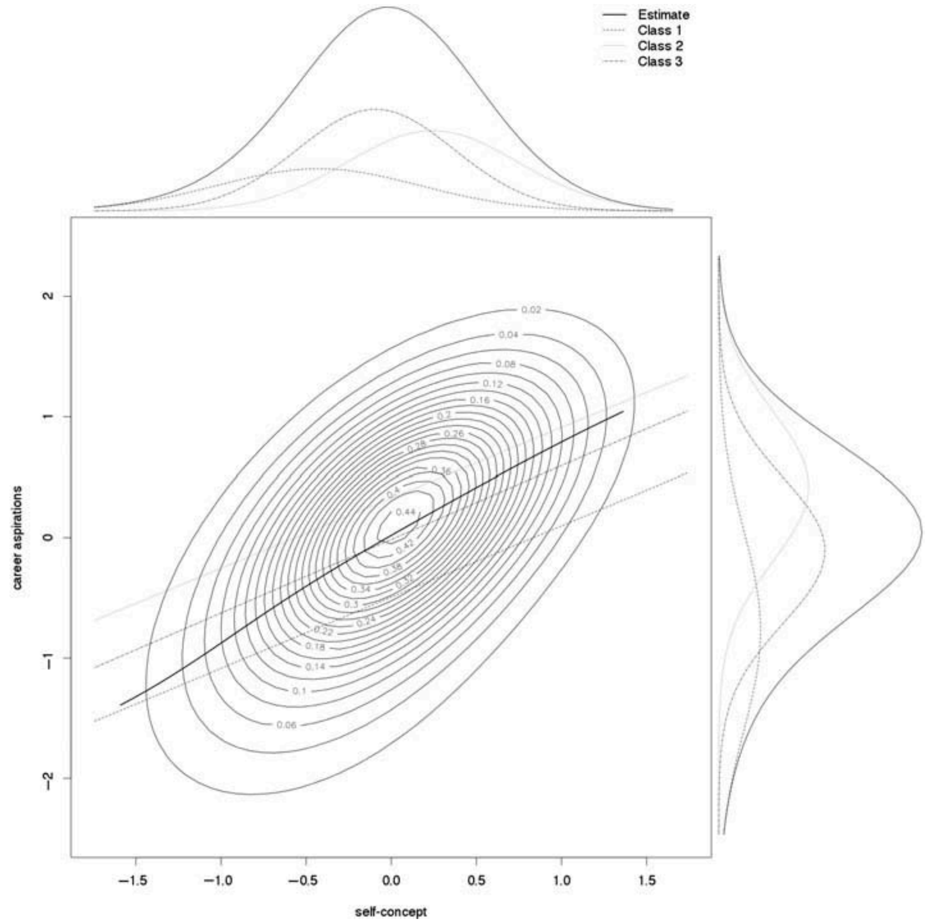

FIGURE 3 Semiparametric nonlinear relationship between students' career aspirations in science and academic self-concept in science obtained from a structural equation mixture model. Three mixture components (classes) are specified. The thick line represents the estimated joint relationship. The figure was created using the SEMMplot r-package (Pek, Sterba, Kok, & Bauer, 2009).

variable vectors, for different measurement models, and for different relations in the structural models. However, in practice, it is often useful (and necessary for identification) to restrain at least some parameters across the mixture components and to assume that parts of the measurement model and the structural model as well as the factor loadings are the same across the different mixture components (Bauer, 2005; Bauer & Curran, 2004; Pek et al., 2011; Pek et al., 2009).

Applications of SEMM

So far, our presentation of the SEMM has implicitly assumed that the mixture components represent existing distinct subpopulations and that they can be interpreted realistically (Borsboom, Mellenbergh, & van Heerden, 2003). Such interpretations are referred to as *direct applications* of SEMM (Dolan & van der Maas, 1998; Titterington, Smith, & Makov, 1985). In contrast, different mixture components might, in fact, not represent latent subpopulations, but could be brought about by nonnormal distributions of the latent variables and their indicators (McLachlan & Peel, 2000).

In *indirect applications* of SEMM, the mixtures are not given a substantive meaning and mixtures are only used to approximate nonlinearity and nonnormality (Bauer, 2005; Bauer & Curran, 2004; Dolan & van der Maas, 1998; Pek et al., 2011; Pek et al., 2009). Consider the following example in which we study the relations between academic self-concept and career aspirations of 15-year-old students in the Jordan subsample of the large-scale Program for International Student Assessment 2006 (PISA; Organisation for Economic Co-Operation and Development, 2009).[2] We specified a SEMM with three mixture components that varied with respect to their structural models, but had invariant measurement models. Figure 3 shows the three separate linear relationships in each component. As can be seen, the three components differ with respect to the slopes and the intercepts. As the measurement models were the same for each component, the differential linear relations in the

[2]In a later section we go into details about the data set.

structural models are indicative of nonlinearities in the relation of academic self-concept and career aspirations in the total population and nonnormal distributions of these variables. Figure 3 (thick line) also shows the result of combining the distributions in the mixture components to represent these nonlinear relations and nonnormal distributions (Pek et al., 2009). As can be seen, the relationship is slightly s-shaped, indicating that students' career aspirations in science rise more strongly with lower academic self-concept and that the growth decelerates for higher values of academic self-concept. As the approximation of the nonlinear relation without specifying an explicit functional form, indirect applications of the SEMM can be characterized as semiparametric models (Escobar & West, 1995; Everitt & Hand, 1981; Nagin, 1999; Verbeke & Lesaffre, 1996).

Some semiparametric Bayesian approaches replace finite normal mixtures with even more flexible latent variable distributions: Chow et al. (2011), Lee, Lu, and Song (2008), and Song, Xia, and Lee (2009) proposed to model nonnormal latent variables with an approximate truncation Dirichlet process with stick-breaking priors and the blocked Gibbs sampler (Ishwaran & James, 2001). Their approach was further generalized by Yang and Dunson (2010), who extended the distribution of the latent variables to mixtures of a centered Dirichlet process (Yang, Dunson, & Baird, 2010) that solves identifiability problems that hampered earlier approaches. In contrast to finite normal mixture models, these kinds of semiparametric approaches are guaranteed to flexibly approximate arbitrary distributions of the latent variables by adapting the number of mixture components. Despite these highly desirable properties, current semiparametric approaches do not offer the possibility of capturing the nonnormality and simultaneously specifying a parametric functional form of the nonlinear relationship of the latent variables. The centered mixture Dirichlet process has not been adapted and discussed in the context of nonlinear latent structural models. The adaptation is not trivial in terms of interpretability of the nonlinearity (Moosbrugger et al., 1997), the ontological status of the latent variables (Borsboom et al., 2003), and from a statistical perspective, because restrictions have to be relaxed then (e.g., related to the covariance structure of the latent variables). In summary, current semiparametric Bayesian approaches should be theoretically more flexible and superior in terms of approximating distributions of latent variables or nonlinear relationships (e.g., Song & Lu, 2010). However, they are either linear and not extended to nonlinear effects (Yang & Dunson, 2010) or not intended to answer the question concerning the parametric size of a nonlinear effect (e.g., How large is the interaction?; Chow et al., 2011), and their specification is difficult for substantive researchers.

None of the mixture modeling approaches discussed has been extended to non-Bayesian SEM hypothesizing parametric nonlinear, (i.e., quadratic and interactive) relations between outcome and predictor variables. In addition, little is known about the consequences of not accounting for nonnormal latent variable distributions in parametric SEM. In this article, we seek to fill this gap by introducing a comprehensive NSEMM that accounts for nonnormal distributions of the predictor variables with finite mixtures of normal distributions and that allows for a parametric specification of nonlinear effects. Hence, we combine the best of both worlds: the explicit model for the nonlinear relations from parametric modeling and the flexibility of accounting for nonnormal latent variables distributions from semiparametric approaches. The proposed NSEMM approach is implementable in readily available software for latent variable modeling (e.g., M*plus*; Muthén & Muthén, 1998–2010).

NONLINEAR STRUCTURAL EQUATION MIXTURE MODEL APPROACH

In this section we extend the SEMM approach and present a comprehensive NSEMM that is a finite mixture approach for nonnormally distributed latent predictor variables in nonlinear SEM.

Structural Model

The major difference between SEMM and the newly proposed NSEMM approach lies in the structural model. In the NSEMM approach, Equation 8 is extended by introducing an additional term that represents nonlinear effects to the structural model:

$$\eta_g = \mathbf{B}_g \eta_g + \alpha_g + \Gamma_{1g}\xi_g + \Gamma_{2g} h\left(\xi_g\right) + \zeta_g \qquad (9)$$

where $h(\cdot)$ is a function that maps the vector $\boldsymbol{\xi}_g$ to a $(k \times 1)$ vector of product terms of $\boldsymbol{\xi}_g$ (e.g., $\left(\xi_1\xi_2, \xi_1^2, \xi_2^2\right)'$). Γ_{2g} $(m \times k)$ is a coefficient matrix denoting the nonlinear effects of $h\left(\boldsymbol{\xi}_g\right)$ on $\boldsymbol{\eta}_g$. $\boldsymbol{\Gamma}_{1g}$ is still a coefficient matrix for the linear effects of ξ_g (see Equation 8).

Estimation

The specification of the proposed measurement and structural model for each mixture g leads to a specific mean structure $\boldsymbol{\mu}_g$ and covariance matrix $\boldsymbol{\Sigma}_g$ of the observed variables that is heteroscedastic (for details see Klein & Moosbrugger, 2000). Because the structural model is nonlinear in $h(\cdot)$, both the mean structure $\boldsymbol{\mu}_g$ and the covariance matrix $\boldsymbol{\Sigma}_g$ are complicated nonlinear functions (in contrast to linear models; for details see Kelava et al., 2011; Klein & Moosbrugger, 2000). This results in a nonnormal density f_g for the observed variables $\left(\boldsymbol{y}', \boldsymbol{x}'\right)'$ for each mixture g. The resulting finite mixture density f of the indicator vector $\left(\boldsymbol{y}', \boldsymbol{x}'\right)'$ can be expressed as follows:

$$f\left((\boldsymbol{y}',\boldsymbol{x}')'\right)=\sum_{g=1}^{G}w_g f_g\left((\boldsymbol{y}',\boldsymbol{x}')'\right) \tag{10}$$

The likelihood of the observed variables for a sample of N randomly drawn observations $(\boldsymbol{y}',\boldsymbol{x}')_i'$ with $i\in\{1,\ldots,N\}$ from the finite mixture is given as:

$$L=\Pi_{i=1}^{N}\left[\sum_{g=1}^{G}w_g f_g\left((\boldsymbol{y}',\boldsymbol{x}')_i'\right)\right] \tag{11}$$

L is a function of the unknown parameters $w_g, \tau_g^y, \tau_g^x, \lambda_g^y, \lambda_g^x, \Theta_g^\varepsilon, \Theta_g^\delta, \mu^{\xi_g}, \Phi^{\xi_g}, \mathrm{B}_g, \alpha_g, \Gamma_{1g}, \Gamma_{2g}$, and Ψ_g for $g\in\{1,\ldots,G\}$.

As with the SEMM approach, the unknown parameters in the likelihood L of the NSEMM approach can be estimated easily by applying an expectation-maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977), if standard rules for the identification of SEM are fulfilled. For linear SEMM, the estimation with the EM algorithm is straightforward (Jedidi et al., 1997b). However, because our structural model has nonlinear (and therefore nonnormal) mixture components f_g, we follow Klein and Moosbrugger's (2000) suggestion for their Latent Moderated Structural Equations approach (LMS) approach and approximate each mixture component f_g itself by a mixture distribution. Because f in the NSEMM approach is a weighted sum of mixture components f_g (see Equation 10), f is a higher dimensional mixture distribution. The parameters of the likelihood L can be estimated readily using M*plus* (Muthén & Muthén, 1998–2010).

Applications and Properties of the NSEMM Approach

As with conventional SEMM models, the NSEMM approach can be used in indirect applications to approximate the nonnormal distributions of latent variables. In this investigation, this is the major use of the model that we are studying and evaluating its performance on. We describe this use of the model next and discuss the restrictions to the measurement model that are necessary to obtain the appropriate interpretation of the latent variables. However, it is equally possible to use the model in a direct application and interpret the components as existing subpopulations or even average the nonlinear parametric within-component structural models to model complex relations between latent variables semiparametrically. We also briefly discuss these applications.

First and foremost, however, we use the NSEMM approach to combine a semiparametric approximation of the nonnormally distributed latent predictor variables and a separate specification of a parametric nonlinear model. The approximation of the latent predictor distribution is accomplished by using the mixture-specific $\boldsymbol{\xi}_g$s and creating a new composite latent variable $\boldsymbol{\xi}$:

$$\boldsymbol{\xi}\sim\sum_{g=1}^{G}w_g N(\mu^{\xi_g},\Phi^{\xi_g}) \tag{12}$$

For a parametric specification of an overall nonlinear structural model, additional restrictions on the parameters need to be imposed. All parameters except for the latent expectation μ^{ξ_g} and covariance matrix Φ^{ξ_g} of the predictor variables are constrained to be equal across the mixture components. This means that the measurement model parameters $\boldsymbol{\tau}_g^y, \boldsymbol{\tau}_g^x, \boldsymbol{\lambda}_g^x, \boldsymbol{\lambda}_g^x, \boldsymbol{\Theta}_g^\epsilon, \Theta_g^\delta$ and the structural model parameters $\mathbf{B}_g, \boldsymbol{\alpha}_g, \boldsymbol{\Gamma}_{1g}, \boldsymbol{\Gamma}_{2g}$, and Ψ_g for $g\in\{1,\ldots,G\}$ are invariant. The resulting model could be, for example, a simple interaction model $\eta_1=\gamma_1\xi_1+\gamma_2\xi_2+\gamma_3\xi_1\xi_2+\zeta_1$ that allows for nonnormally distributed ξ_1 and ξ_2 variables. The NSEMM approach is then complementary to parametric nonlinear SEM approaches and the semiparametric SEMM approach.

In addition, it is also possible to use the flexibility of the proposed NSEMM approach in direct applications where the different mixture components are given substantive interpretations as latent subpopulations. For example, mixture-specific nonlinear models could be specified indicating that each unobserved latent mixture (class) has its own parametric nonlinear or linear relationship. This would be the case if for one subpopulation a relationship can be described by an interaction model (e.g., omitting the mixture index: $\eta_1=\gamma_1\xi_1+\gamma_2\xi_2+\gamma_3\xi_1\xi_2+\zeta_1$), whereas for another subpopulation a relationship can be described by a linear model ($\eta_1=\gamma_1\xi_1+\gamma_2\xi_2+\zeta_1$). When each mixture has its own parametric linear or nonlinear submodel, parameters vary freely across the mixtures. This is an extension of the direct applications of the SEMM approach to nonlinear modeling.

Thus, the NSEMM approach fills a gap between the parametric nonlinear structural equation model on the one hand, which leads to biased results if variables are nonnormally distributed (Kelava & Nagengast, 2012; Marsh et al., 2004), and on the other hand the semiparametric SEMM approach, which cannot answer the question about the size of a nonlinear effect. The NSEMM approach inherits the advantages of both approaches and can estimate the size of a nonlinear effect in the presence of nonnormal data.

A SIMULATION STUDY WITH NONNORMAL DATA

In this section we present a simulation study in which the LMS approach (Klein & Moosbrugger, 2000), the extended unconstrained approach (ExUC; Kelava, 2009; Kelava & Brandt, 2009; Marsh et al., 2004), and the proposed NSEMM approach are compared. The major aim of the simulation study was to examine the influence of varying degrees of nonnormality on the estimates of the nonlinear effects. Special emphasis was placed on the inspection of bias of the parameters and their standard error estimates, because even slightly biased nonlinear effects and confidence intervals are severe challenges for model inferences and interpretation with respect to spurious nonlinear effects (Kelava, Moosbrugger, Dimitruk, & Schermelleh-Engel, 2008; Klein, Schermelleh-Engel, Moosbrugger, & Kelava, 2009; MacCallum & Mar, 1995).

The NSEMM approach was implemented in M*plus* (version 6.12; Muthén & Muthén, 1998–2010) specifying two to four normal mixture components. The LMS approach and the ExUC approach were also implemented in M*plus*. For the ExUC approach, robust sandwich estimates of the standard errors were used (with the MLR option in M*plus*), because standard error estimates of the effects based on the expected (information) covariance matrix underestimate the standard errors in product indicator approaches (see Jöreskog & Yang, 1996; Kenny & Judd, 1984).

Design of the Simulation Study

In the simulation study data were generated according to the following structural model with latent interaction and quadratic effects:

$$\eta = .3\xi_1 + .3\xi_2 + .2\xi_1\xi_2 + .2\xi_1^2 + .2\xi_2^2 + \zeta \tag{13}$$

where ζ was normally distributed with $\zeta \sim N(0, .700)$. Nonnormality of the latent predictors was induced using the Vale and Maurelli (1983) method implemented in an R (R Development Core Team, 2011) script by Zopluoglu (2012). Three conditions of univariate skewness and kurtosis were selected for each latent predictor variable (in line with the values used by Curran et al., 1996): (a) normality with skewness 0 and kurtosis 0, (b) moderate nonnormality with skewness 2 and kurtosis 7, and (c) severe nonnormality with skewness 3 and kurtosis 21. The latent predictor variables were centered ($E(\xi_1) = E(\xi_2) = 0$) and correlated at .625 with variances of 1 ($Var(\xi_1) = Var(\xi_2) = 1$). For the linear latent predictor variables (ξ_1, ξ_2) and the latent outcome variable (η) three indicator variables were generated. The reliability of the three indicators was set to be .80. The indicators were unidimensional measures of the latent variables with loadings equal to 1 and with residual variables following a normal distribution. Hence, all observed deviations of the indicator distribution from normality are due to the nonnormal latent predictor variables and the nonlinear effects. The sample size was set to $N = 800$. A total of $rep = 1{,}000$ replications were generated from each of the resulting conditions. A solution was considered proper and selected when there were no negative variances or standard error estimates. Outliers were identified by analyzing plots and z scores (cf. Paxton, Curran, Bollen, Kirby, & Chen, 2001).

Results of the Simulation Study

First, we present results when the latent predictor variables are normally distributed. As can be seen from Table 1, each of the three approaches provided unbiased parameter estimates of the linear and nonlinear effects. The standard error estimates were unbiased for all approaches. As expected from its theoretical maximum likelihood properties, the LMS approach had an advantage in its power of detecting the nonlinear effects (interaction effect with 69.20% and quadratic effects with 98.20% and 97.90%). The ExUC approach showed the lowest power (interaction effect with 62.60% and quadratic effects with 96.80% and 97.00%). The power of the NSEMM approach was higher than the power of the ExUC approach (but below the LMS approach).

With moderate nonnormality (see Table 2), the NSEMM approach and the ExUC approach provided unbiased estimates of the linear and nonlinear effects. With the LMS approach, biased estimates of the nonlinear effects were obtained. The interaction effect was underestimated by 14.63% and the quadratic effects were overestimated by 18.92% and 17.77%, respectively. Their standard error estimates were acceptable. The ExUC approach produced underestimated standard error estimates for the interaction effect and for the first quadratic effect. Their power was therefore overestimated.

With severe nonnormality (see Table 3), the NSEMM approach and the ExUC approach provided unbiased estimates of the linear and nonlinear effects. With increasing nonnormality the bias of the nonlinear effects increased for the LMS approach. The interaction effect was underestimated by 17.6% and the first and second quadratic effects were overestimated by 23.25% and 21.32%, respectively. The standard errors of the two quadratic effects were underestimated and their power overestimated. Again, the ExUC approach showed underestimated standard errors of all three nonlinear effects, leading to a biased estimate of the power to detect the nonlinear effects. The NSEMM approach showed unbiased standard error estimates. With an increasing number of mixture components, a small loss of power for detecting the effects was observed.

EXAMPLE

We demonstrate the proposed NSEMM approach and how its results can differ from conventional approaches using data from the large-scale assessment study PISA 2006 (Organisation for Economic Co-Operation and Development, 2009). In 3-year cycles, the competencies of 15-year-old students in reading, mathematics, and Science are assessed using nationally representative samples. In addition to the achievement data, background information on students, parents, and schools is available for secondary data analyses. In our example, we use data from the student background questionnaire of PISA 2006 focusing on achievement and motivation in science. Following Nagengast et al. (2011), we test the relation between academic self-concept in science (5 items, representing expectancy of success), enjoyment of science (6 items, representing intrinsic value), and career aspirations in science (4 items, used as proxy variable for

TABLE 1
Interaction and Quadratic Effects Model: Skewness/Kurtosis: 0/0 (Normally Distributed Variables)

Parameter	*True Value*	*Mean Par. Est.*	*Bias %*	*SD*	*SE*	*SE/SD*	*Power*
NSEMM approach: 2 components							
γ_1	0.300	0.302	0.55%	0.051	0.053	1.050	100.00%
γ_2	0.300	0.299	−0.22%	0.050	0.054	1.087	100.00%
γ_3	0.200	0.199	−0.61%	0.082	0.085	1.036	66.50%
γ_4	0.200	0.200	0.10%	0.050	0.051	1.028	97.40%
γ_5	0.200	0.200	−0.20%	0.050	0.051	1.023	97.80%
NSEMM approach: 3 components							
γ_1	0.300	0.299	−0.37%	0.051	0.054	1.051	100.00%
γ_2	0.300	0.299	−0.35%	0.051	0.055	1.077	100.00%
γ_3	0.200	0.196	−1.85%	0.083	0.086	1.038	66.00%
γ_4	0.200	0.197	−1.36%	0.050	0.052	1.041	97.10%
γ_5	0.200	0.199	−0.28%	0.051	0.053	1.046	97.40%
NSEMM approach: 4 components							
γ_1	0.300	0.299	−0.22%	0.052	0.056	1.063	100.00%
γ_2	0.300	0.300	−0.14%	0.053	0.057	1.073	100.00%
γ_3	0.200	0.201	0.54%	0.084	0.088	1.048	65.30%
γ_4	0.200	0.197	−1.37%	0.051	0.054	1.050	96.90%
γ_5	0.200	0.198	−0.75%	0.052	0.055	1.052	97.00%
LMS approach							
γ_1	0.300	0.301	0.48%	0.050	0.053	1.063	100.00%
γ_2	0.300	0.301	0.21%	0.049	0.053	1.072	100.00%
γ_3	0.200	0.198	−0.82%	0.082	0.082	0.998	69.20%
γ_1	0.200	0.201	0.71%	0.049	0.049	0.991	98.20%
γ_5	0.200	0.201	0.41%	0.049	0.049	0.994	97.90%
ExUC approach							
γ_1	0.300	0.301	0.50%	0.050	0.050	0.994	100.00%
γ_2	0.300	0.301	0.24%	0.050	0.050	0.995	100.00%
γ_3	0.200	0.198	−0.92%	0.089	0.088	0.985	62.60%
γ_4	0.200	0.201	0.61%	0.053	0.052	0.997	96.80%
γ_5	0.200	0.202	0.75%	0.054	0.052	0.971	97.00%

Note. *SE* = mean standard error; NSEMM = nonlinear equation mixture model; LMS = Latent Moderated Structural Equations approach; ExUC = extended unconstrained approach.

choice), testing predictions of expectancy-value theory of achievement motivation (Eccles (Parsons), 1983; Wigfield & Eccles, 2000). We analyze data from the Jordan sample of PISA 2006. Responses to the questionnaire items used as indicators for academic self-concept and enjoyment were nonnormally distributed. To simplify the analysis and because of the didactical nature of the example, we only consider students with complete responses to all 15 items ($N = 6{,}038$) and do not correct for the PISA sampling design by weighting the observations. Hence, the analyses are not representative of 15-year-old students in Jordan, but merely serve as an illustration of nonnormal latent predictor distributions on parameter estimates in nonlinear SEM.

Specifically we tested the following structural model with the proposed NSEMM approach, the LMS approach, and the ExUC approach:

$$CAREER = \gamma_1\, SC + \gamma_2\, ENJ + \gamma_3\, SC \cdot ENJ + \gamma_4\, SC^2 + \gamma_5\, ENJ^2 + \zeta, \tag{14}$$

where *CAREER* is the latent outcome variable career aspirations in science, *SC* is academic self-concept in science, and *ENJ* is enjoyment of science. The notation for the model parameters is similar to Equation 3, but without the latent intercept α.[3] All analyses were implemented in M*plus* (version 6.11; Muthen & Muthen, 1998–2010).

The summarized results for the three approaches are given in Table 4. The main effects were positive and significant in all approaches. However, there were important differences with respect to the nonlinear effects: For LMS and the ExUC approach, all nonlinear effects were small and statistically not significant. The NSEMM approach yielded nonsignificant interaction effects and a nonsignificant quadratic effect of academic self-concept (γ_4). In contrast to the conventional approaches, the NSEMM approach revealed a quadratic effect (γ_5) of enjoyment of science on career aspirations. This finding indicated that the relation of academic self-concept and enjoyment of science in predicting career

[3]Instead of estimating the latent intercept α, we specified latent measurement intercepts for the indicator variables, which is the standard setting in software for nonlinear SEM. See the earlier discussion of identification problems.

TABLE 2
Interaction and Quadratic Effects Model: Skewness/Kurtosis: 2/7 (Unimodal Nonnormally Distributed Variables)

Parameter	*True Value*	*Mean Par. Est.*	*Bias %*	*SD*	*SE*	*SE/SD*	*Power*
NSEMM approach: 2 components							
γ_1	0.300	0.285	−4.95%	0.138	0.140	1.014	59.70%
γ_2	0.300	0.296	−1.41%	0.134	0.137	1.021	63.10%
γ_3	0.200	0.204	1.79%	0.120	0.118	0.987	48.00%
γ_4	0.200	0.209	4.68%	0.069	0.069	1.004	84.50%
γ_5	0.200	0.208	3.84%	0.069	0.070	1.024	87.00%
NSEMM approach: 3 components							
γ_1	0.300	0.292	−2.77%	0.140	0.143	1.020	58.30%
γ_2	0.300	0.300	0.09%	0.136	0.139	1.024	62.30%
γ_3	0.200	0.203	1.41%	0.123	0.125	1.018	47.00%
γ_4	0.200	0.203	1.74%	0.070	0.072	1.023	82.30%
γ_5	0.200	0.203	1.38%	0.071	0.074	1.035	85.10%
NSEMM approach: 4 components							
γ_1	0.300	0.292	−2.51%	0.142	0.148	1.041	57.20%
γ_2	0.300	0.303	1.06%	0.137	0.141	1.028	60.80%
γ_3	0.200	0.198	−1.04%	0.125	0.130	1.035	46.50%
γ_4	0.200	0.202	0.98%	0.072	0.075	1.049	81.10%
γ_5	0.200	0.202	1.04%	0.072	0.077	1.060	84.00%
LMS approach							
γ_1	0.300	0.310	3.27%	0.122	0.120	0.981	75.20%
γ_2	0.300	0.320	6.60%	0.115	0.120	1.040	78.40%
γ_3	0.200	0.171	−14.63%	0.139	0.126	0.907	32.40%[a]
γ_4	0.200	0.238	18.92%	0.075	0.069	0.913	88.50%[a]
γ_5	0.200	0.236	17.77%	0.074	0.069	0.928	88.80%[a]
ExUC approach							
γ_1	0.300	0.295	−1.53%	0.147	0.141	0.960	57.40%
γ_2	0.300	0.306	2.07%	0.138	0.141	1.017	61.00%
γ_3	0.200	0.210	5.04%	0.125	0.113	0.900	48.40%[a]
γ_4	0.200	0.200	−0.22%	0.071	0.063	0.894	83.90%[a]
γ_5	0.200	0.198	−1.14%	0.069	0.064	0.926	85.20%

Note. *SE* = mean standard error; NSEMM = nonlinear structural equation mixture model; LMS = Latent Moderated Structural Equations approach; ExUC = extended unconstrained approach.

[a]Results are based on a biased mean parameter or *SE* estimate ($|bias| \geq 10\%$).

aspirations was not simply linear, but included a nonlinear component as well, the quadratic effect of enjoyment. This quadratic effect is important, as it contradicts most contemporary tests of expectancy-value models in educational psychology that only assume a linear relation between expectancy and value in predicting engagement and achievement. If the nonnormal predictor distribution had not been taken into account, this nonlinear relationship would have not have been detected.

DISCUSSION

In this article, we proposed an NSEMM approach that integrates the strength of parametric approaches of easy-to-interpret functional relationships and the flexibility of the semiparametric mixture models for approximating the nonnormality of latent variables. We began by introducing mixture modeling as a statistical tool for the approximation of nonnormal distributions and nonlinear relationships in manifest and latent variable models. The key idea was to use mixture distributions to approximate nonnormal distributions of the latent predictors (McLachlan & Peel, 2000). We presented the specification of the mixture measurement and structural models of the NSEMM approach. In a comparative simulation study we showed its properties for varying degrees of nonnormality. Finally, we used an empirical example from educational psychology as an illustration and showed that the results from several current (parametric) approaches led to different inferences.

Interpretation of the Results of the Simulation Studies

Regarding the results of the simulation study, the following conclusions concerning the three approaches can be drawn. First, in the situation of normally distributed indicators, LMS produced unbiased and efficient parameter estimates. If indicators were nonnormally distributed, parameter estimates were biased. Second, the ExUC approach led to biased standard error estimates that were more severely biased in the presence of nonnormally distributed indicators, showing that the finite sample properties of the sandwich estimator were not sufficient to produce unbiased standard errors. Third, the NSEMM approach provided unbiased parameter

TABLE 3
Interaction and Quadratic Effects Model: Skewness/Kurtosis: 3/21 (Unimodal Nonnormally Distributed Variables)

Parameter	*True Value*	*Mean Par. Est.*	*Bias %*	*SD*	*SE*	*SE/SD*	*Power*
NSEMM approach: 2 components							
γ_1	0.300	0.284	−5.25%	0.181	0.185	1.020	43.00%
γ_2	0.300	0.290	−3.18%	0.179	0.182	1.015	43.90%
γ_3	0.200	0.189	−5.71%	0.118	0.119	1.005	40.30%
γ_4	0.200	0.206	2.82%	0.068	0.070	1.025	81.20%
γ_5	0.200	0.205	2.55%	0.070	0.070	1.013	80.70%
NSEMM approach: 3 components							
γ_1	0.300	0.302	0.67%	0.188	0.194	1.032	40.80%
γ_2	0.300	0.287	−4.31%	0.184	0.189	1.028	41.50%
γ_3	0.200	0.199	−0.67%	0.125	0.127	1.016	39.80%
γ_4	0.200	0.199	−0.47%	0.069	0.069	1.001	81.40%
γ_5	0.200	0.199	−0.67%	0.071	0.074	1.045	78.50%
NSEMM approach: 4 components							
γ_1	0.300	0.293	−2.49%	0.192	0.203	1.055	39.50%
γ_2	0.300	0.302	0.68%	0.188	0.197	1.048	40.40%
γ_3	0.200	0.198	−1.20%	0.130	0.138	1.064	38.90%
γ_4	0.200	0.196	−1.87%	0.071	0.074	1.048	78.30%
γ_5	0.200	0.199	−0.30%	0.072	0.076	1.050	77.20%
LMS approach							
γ_1	0.300	0.306	2.14%	0.167	0.162	0.974	51.70%
γ_2	0.300	0.307	2.23%	0.162	0.163	1.003	52.00%
γ_3	0.200	0.165	−17.61%	0.152	0.139	0.910	27.40%[a]
γ_4	0.200	0.247	23.25%	0.078	0.069	0.886	88.30%[a]
γ_5	0.200	0.243	21.32%	0.080	0.070	0.875	88.70%[a]
ExUC approach							
γ_1	0.300	0.301	0.35%	0.205	0.191	0.935	41.10%
γ_2	0.300	0.304	1.37%	0.195	0.192	0.983	40.80%
γ_3	0.200	0.208	4.05%	0.132	0.115	0.870	48.70%[a]
γ_4	0.200	0.200	0.16%	0.071	0.058	0.811	87.10%[a]
γ_5	0.200	0.195	−2.28%	0.070	0.058	0.831	86.70%[a]

Note. *SE* = mean standard error; NSEMM = nonlinear structural equation mixture modeling; LMS = Latent Moderated Structural Equations approach; ExUC = extended unconstrained approach.

[a]Results are based on a biased mean parameter or *SE* estimate ($|bias| \geq 10\%$).

and standard error estimates under all conditions. With an increasing number of mixture components, the power to detect the effects slightly decreased.

Guidelines for the Application of the Approaches

As long as variables are normally distributed, we recommend using the LMS or the ExUC approach. With normally distributed latent predictors, the LMS approach provides optimal (maximum likelihood) estimates of the structural model. With a single mixture component for the latent predictors, the LMS approach is a special case of the NSEMM approach. When manifest predictor indicators are normally distributed, there is no need to overspecify the model with additional latent mixture components. The ExUC approach could be applied as well, but its standard errors are also asymptotically biased, even for normally distributed variables (Jöreskog & Yang, 1996).[4]

When latent predictors are skewed, we recommend the use of the NSEMM approach. The mixture components in the NSEMM approach allow for the approximation of an arbitrary distribution of the latent predictors (Bauer & Curran, 2004; McLachlan & Peel, 2000) and avoid overestimated and spurious nonlinear effects (in terms of biased parameter and standard error estimates). As nonlinear latent variables violate the distributional assumptions of the LMS approach, its estimates of the effects will be positively or negatively biased (depending on the skewness on shape of the latent predictors).

The ExUC approach showed unbiased effects in general, but underestimated standard errors, leading to an increased frequency of Type-I-errors. In addition, for more complex models (e.g., with multiple nonlinear effects: $\xi_1\xi_2$, $\xi_1\xi_3$, $\xi_2\xi_3$), the number of additional parameters for a correct specification (Kelava & Brandt, 2009; Kelava et al., 2011) increases substantially (measurement error covariances, latent intercepts, and covariances of the latent predictors). As the ratio of subjects to free parameters decreases, parameter estimates can become unstable. Furthermore,

[4]The reason is that the product of normally distributed variables is nonnormal (Aroian, 1944).

TABLE 4
Results of the Analysis of the Jordan Sample of PISA 2006 With Three Approaches

Parameter	*Par. Est.*	*SE*	*t Value*	*p*
Mixture approach: 2 components				
γ_1	.459	.026	17.772	<.001
γ_2	.555	.024	22.828	<.001
γ_3	−.059	.055	−1.069	.285
γ_4	.019	.042	.454	.649
γ_5	.066	.024	2.711	.007
LMS approach				
γ_1	.466	.024	19.248	<.001
γ_2	.524	.021	25.022	<.001
γ_3	−.047	.068	−0.689	.491
γ_4	.002	.050	0.035	.972
γ_5	.026	.029	0.904	.368
ExUC approach				
γ_1	.484	.027	18.003	<.001
γ_2	.541	.024	22.623	<.001
γ_3	.021	.039	0.526	.599
γ_4	.004	.022	0.173	.863
γ_5	.033	.020	1.669	.095

Note. LMS = Latent Moderated Structural Equations approach; ExUC = extended unconstrained approach.

from a practical perspective, the specification of product indicator approaches is error-prone. For example, the omission of the specification of some measurement error covariances in product indicator approaches can lead to wrong inferences because of underestimated and overestimated effects (Kelava et al., 2008).

An important practical question concerns the number of components that should be specified in the NSEMM approach. The answer to this question strongly depends on the shape of the multivariate distribution of the measured variables, on their reliability, and on the source of nonnormality. In principle, it is (technically) not possible to answer whether the nonnormality of the observed variables is the result of nonnormally distributed latent predictors or measurement error variables (Song et al., 2010). If there are hints that the measures are sufficiently reliable (we used a reliability of our measures of .80), as is usually the case for approved scales, then it could be assumed that the nonnormality of the measured variables is mostly due to nonnormal latent predictors. The distribution of the observed variables will be very similar to the distribution of the latent predictors because most of the indicators' variance will be explained by the latent predictors (and not by the measurement error variables; Bollen, 1989). The decision on the number of latent mixture components should be based on a combination of visual inspection of multivariate plots of the observed variables (e.g., using the scaling indicator of each latent predictor) and the minimum size of information criteria (e.g., Bayesian information criterion; Schwarz, 1978). McLachlan and Peel (2000) gave numerous examples of mixture distributions and numbers of components that can approximate arbitrary multimodal and severely skewed distributions. In our simulation study, we showed that the overspecification of the number of mixture components did not lead to biased estimates, but only to a small loss of power. Based on our experience and the results of simulation studies with two latent predictors, two to four mixtures should work quite well, which is in line with the experiences of Bauer and Curran (2004) and Bauer (2005). We are very cautious with recommending a specific number of components for models with more than two latent predictors (e.g., ξ_1, ξ_2, ξ_3).

Limitations and Directions for Further Research

There are also two limitations of the proposed procedure that are worthwhile to explore in further research. First, we proposed a frequentist approach for the approximation of the nonnormal latent distribution. Other semiparametric Bayesian mixture approaches should be superior under some conditions as they are even more flexible in approximating nonnormal distributions (Chow et al., 2011; Lee et al., 2008; Song et al., 2009; Yang & Dunson, 2010). However, these semiparametric mixture approaches require specialized software when applied to latent variable models and are not yet readily available to applied researchers. Their relative advantage compared to finite normal mixtures has not been investigated systematically for finite samples. Most important, these approaches are not intended to give an easy-to-interpret answer on the size of a nonlinear effect as parametric frequentist approaches do. To allow for a specification of parametric nonlinear effects in the structural models and to keep an ontological interpretability (Borsboom et al., 2003), the applied mixture Dirichlet processes have to be adapted and assumptions relaxed.

Second, we restricted our model to nonnormally distributed latent predictors and we assumed normality for the residual variables for the NSEMM approach. In situations where the reliability of indicators is low and indicator variables are nonnormally distributed, the assumption of the NSEMM approach thus might be violated. Recently, a method of moments based approach was presented by Mooijaart and Bentler (2010) that can be used to relax the assumption of normally distributed measurement errors. Further conceptual and empirical work is necessary to systematically compare these approaches and ascertain when one of them is more appropriate than the other (for recent work in the linear model context, see Molenaar, Dolan, & Verhelst, 2010).

Final Considerations and Future Perspectives

Finally, we conclude that the proposed new NSEMM approach offers a different perspective on modeling nonlinear relationships in SEM. One possibility of an extension is to specify nonlinear effects (e.g., interactions) of multivariate ordinal data, which should be useful for substantial researchers, too (see also Lee, Song, & Cai, 2010). Future work and simulation studies are needed to examine the strengths and limitations of mixture modeling in latent nonlinear models.

ACKNOWLEDGMENTS

The authors would like to thank Editor George A. Marcoulides and three anonymous reviewers for their valuable comments on an earlier version of the article.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (DFG; Grants No. KE 1664/1–1).

REFERENCES

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

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

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

Arminger, G., & Stein, P. (1997). Finite mixtures of covariance structure models with regressors. *Sociological Methods and Research*, *26*, 148–182.

Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika*, *64*, 475–494.

Aroian, L. A. (1944). The probability function of the product of two normally distributed variables. *The Annals of Mathematical Statistics*, *18*, 265–271.

Bauer, D. J. (2005). A semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*, *12*, 513–535.

Bauer, D. J., & Curran, P. J. (2004). The integration of continous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, *9*, 3–29.

Bohrnstedt, G., & Marwell, G. (1978). The reliability of the products of two random variables. In K. Schuessler (Ed.), *Sociological methodology* (pp. 254–273). San Francisco, CA: Jossey-Bass.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: A least squares estimator. *Sociological Methodology*, *1995*, 223–251.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–218.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. (2006). *Measurement error in nonlinear models: A modern perspective*. Boca Raton, FL: CRC Press.

Chow, S.-M., Tang, N., Yuan, Y., & Song, X. (2011). Bayesian estimation of semiparametric nonlinear dynamic factor analysis models using Dirichlet process prior. *British Journal of Mathematical and Statistical Psychology*, *64*, 69–106.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analyses for the behavioral sciences* (3rd ed.). Hillsdale, NJ: Erlbaum.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29.

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

Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, *63*, 227–253.

Eccles (Parsons), J. S. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives. Psychological and sociological approaches* (pp. 75–146). San Francisco, CA: Freeman.

Escobar, M. D., & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, *90*, 577–588.

Everitt, B., & Hand, D. J. (1981). *Finite mixture distributions*. London, UK: Chapman & Hall.

Ishwaran, H., & James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, *96*, 161–173.

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

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997a). Finite-mixture structural equation models for response based segmentation and unobserved heterogeneity. *Marketing Science*, *16*, 39–59.

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997b). STEMM: A general finite mixture structural equation model. *Journal of Classification*, *14*, 23–50.

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

Kelava, A. (2009). *Multikollinearität in nicht-linearen latenten Strukturgleichungsmodellen [Multicollinearity in nonlinear structural equation models]*. (Unpublished doctoral dissertation). Goethe Universität Frankfurt, Frankfurt, Germany. Retrieved from http://publikationen.ub.uni-frankfurt.de/volltexte/2009/6336/

Kelava, A., & Brandt, H. (2009). Estimation of nonlinear latent structural equation models using the extended unconstrained approach. *Review of Psychology*, *16*, 123–131.

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

Kelava, A., & Nagengast, B. (2012). A Bayesian approach for the estimation of latent interaction and quadratic effects when latent variables are non-normally distributed. *Multivariate Behavioral Research*, *47*, 717–742.

Kelava, A., Werner, C., Schermelleh-Engel, K., Moosbrugger, H., Zapf, D., Ma, Y., Cham, H., Aiken, L. S., & West, S. G. (2011). Advanced nonlinear structural equation modeling: Theoretical properties and empirical application of the distribution-analytic LMS and QML estimators. *Structural Equation Modeling*, *18*, 465–491.

Kenny, D., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.

Klein, A. G., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.

Klein, A. G., & Muthén, B. O. (2007). Quasi maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects. *Multivariate Behavioral Research*, *42*, 647–674.

Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., & Kelava, A. (2009). Assessing spurious interaction effects. In T. Teo & M. Khine (Eds.), *Structural equation modeling in educational research: Concepts and applications* (pp. 13–28). Rotterdam, Netherlands: Sense.

Lee, S.-Y., Lu, B., & Song, X.-Y. (2008). Semiparametric Bayesian analysis of structural equation models with fixed covariates. *Statistics in Medicine*, *27*, 2341–2360.

Lee, S.-Y., Song, X.-Y., & Cai, J.-H. (2010). A Bayesian approach for nonlinear structural equation models with dichotomous variables using logit and probit links. *Structural Equation Modeling*, *17*, 280–302.

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

Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and interaction terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling*, *13*, 497–519.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

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

Marsh, H. W. (2007). *Self-concept theory*, *measurement and research into practice: The role of self-concept in educational psychology*. Leicester, UK: British Psychological Society.

Marsh, H. W., Little, T. D., Bovaird, J. A., & Widaman, K. F. (2007). Unconstrained structural equation models of latent interactions: Contrasting residual- and mean-centered approaches. *Structural Equation Modeling*, *14*, 570–580.

Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: Evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, *9*, 275–300.

Marsh, H. W., Wen, Z., & Hau, K.-T. (2006). Structural equation models of latent interaction and quadratic effects. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 225–265). Greenwich, CT: Information Age.

McDowell, M. A., Fryar, C. D., Ogden, C. L., & Flegal, K. M. (2008). Anthropometric reference data for children and adults: United States, 2003–2006. *National Health Statistics Report*, *10*, 1–45. Retrieved from http://www.cdc.gov/nchs/data/nhsr/nhsr010.pdf

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

Molenaar, D., Dolan, C. V., & Verhelst, N. D. (2010). Testing and modelling non-normality within the one-factor model. *British Journal of Mathematical and Statistical Psychology*, *63*, 293–317.

Mooijaart, A., & Bentler, P. M. (2010). An alternative approach for nonlinear latent variable models. *Structural Equation Modeling*, *17*, 357–373.

Moosbrugger, H., Schermelleh-Engel, K., & Klein, A. G. (1997). Methodological problems of estimating latent interaction effects. *Methods of Psychological Research Online*, *2*, 95–111. Retrieved from http://www.mpr-online.de

Muthén, B. O. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In A. Sayer & L. Collins (Eds.), *New methods for the analysis of change* (p. 291–322). Washington, DC: American Psychological Association.

Muthén, L. K., & Muthén, B. O. (1998–2010). *M*plus *user's guide*. (6th ed.). Los Angeles, CA: Muthén & Muthén.

Nagengast, B., Marsh, H. W., Scalas, L. F., Xu, M. K., Hau, K.-T., & Trautwein, U. (2011). Who took the "x" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science*, *22*, 1058–1066.

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods*, *4*, 139–157.

Organisation for Economic Co-Operation and Development. (2009). *PISA 2006: Science competencies for tomorrow's world*. Paris, France: OECD.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, *8*, 287–312.

Pek, J., Losardo, D., & Bauer, D. J. (2011). Confidence intervals for a semiparametric approach to modeling nonlinear relations among latent variables. *Structural Equation Modeling*, *18*, 537–553.

Pek, J., Sterba, S. K., Kok, B. E., & Bauer, D. J. (2009). Estimating and visualizing nonlinear relations among latent variables: A semiparametric approach. *Multivariate Behavioral Research*, *44*, 407–436.

Pek, J., Sterba, S., Kok, B. E., & Bauer, D. J. (2009). Visualizing nonlinear relations among latent variables: An online plotting utility and R package. Multivariate Behavioral *Research*, *44*, 407–436.

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

R Development Core Team. (2011). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria: Author. Retrieved from http://www.R-project.org

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

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

Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement food for thought. *Perspectives on Psychological Science*, *3*, 102–116.

Song, X.-Y., & Lu, Z.-H. (2010). Semiparametric latent variable models with Bayesian p-splines. *Journal of Computational and Graphical Statistics*, *19*, 590–608.

Song, X.-Y., Pan, J.-H., Kwok, T., Vandenput, L., Ohlsson, C., & Leung, P.-C. (2010). A semiparametric Bayesian approach for structural equation models. *Biomatrical Journal*, *52*, 314–332.

Song, X.-Y., Xia, Y.-M., & Lee, S.-Y. (2009). Bayesian semiparametric analysis of structural equation models with mixed continuous and unordered categorical variables. *Statistics in Medicine*, *28*, 2253–2276.

Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Chichester, UK: Wiley.

Vale, C., & Maurelli, V. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, *48*, 465–471.

Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogenity in the random-effects population. *Journal of the American Statistical Association*, *91*, 217–221.

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

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81.

Yang, M., & Dunson, D. B. (2010). Bayesian semiparametric structural equation models with latent variables. *Psychometrika*, *75*, 675–693.

Yang, M., Dunson, D. B., & Baird, D. (2010). Semiparametric Bayes hierarchical models with mean and variance constriants. *Computational Statistics & Data Analysis*, *54*, 2172–2186.

Zopluoglu, C. (2012). *Gen.nonnormal: Generating multivariate nonnormal data*. Retrieved from http://www.tc.umn.edu/~zoplu001/Rscript.html