# Fitting Nonlinear Structural Equation Models in R with Package nlsem

**Nora Umbach**
University of Tübingen

**Katharina Naumann**
University of Tübingen

**Holger Brandt**
University of Tübingen

**Augustin Kelava**
University of Tübingen

## Abstract

This article is a (slightly) modified version of Umbach, Naumann, Hoppe, and Brandt (2017), published in the Journal of Statistical Software.

Structural equation mixture modeling (SEMM) has become a standard procedure in latent variable modeling over the last two decades (Jedidi, Jagpal, and DeSarbo 1997b; Muthén and Shedden 1999; Muthén 2001, 2004; Muthén and Asparouhov 2009). SEMM was proposed as a technique for the approximation of nonlinear latent variable relationships by finite mixtures of linear relationships (Bauer 2005, 2007; Bauer, Baldasaro, and Gottfredson 2012). In addition to this semiparametric approach to nonlinear latent variable modeling, there are numerous parametric nonlinear approaches for normally distributed variables (e.g., LMS in **Mplus**; Klein and Moosbrugger 2000). Recently, an additional semiparametric nonlinear structural equation mixture modeling (NSEMM) approach was proposed by Kelava, Nagengast, and Brandt (2014) that is capable of dealing with nonnormal predictors. In the **nlsem** package presented here, the SEMM, two distribution analytic (QML and LMS) and NSEMM approaches can be specified and estimated. We provide examples of how to use the package in the context of nonlinear latent variable modeling.

*Keywords*: interaction effect, quadratic effect, nonlinear effect, mixture model, nonnormality, semiparametric, latent variables, R.

## 1. Introduction

The analysis of nonlinear relationships between latent variables in the structural equation modeling (SEM) framework has been conducted primarily with two different classes of models. The first class of models consists of parametric SEMs that use latent product terms to analyze specific types of nonlinear relationships, such as interaction or quadratic effects (for an overview, see Kelava, Werner, Schermelleh-Engel, Moosbrugger, Zapf, Ma, Cham, Aiken, and West 2011; Marsh, Wen, and Hau 2004, 2006; Moulder and Algina 2002). For the purpose of assessing these parametric nonlinear effects, a variety of approaches have been developed such as the product-indicator approaches (e.g., Bollen 1995; Jaccard and Wan 1995; Jöreskog and Yang 1996; Kelava and Brandt 2009; Kenny and Judd 1984; Little, Bovaird, and Widaman 2006; Marsh *et al.* 2004, 2006; Ping 1995, 1996; Wall and Amemiya 2001),

distribution-analytic approaches (Klein and Moosbrugger 2000; Klein and Muthén 2007), moment-based approaches (Mooijaart and Bentler 2010; Wall and Amemiya 2000, 2003), and Bayesian approaches (Arminger and Muthén 1998; Lee 2007). For these kinds of models, the functional relationship (e.g., quadratic) needs to be specified a priori, and the size of the nonlinear effects can be calculated (Wen, Marsh, and Hau 2010).

The second class of models consists of semiparametric structural equation mixture models (SEMM; Arminger and Stein 1997; Arminger, Stein, and Wittenberg 1999; Bauer 2005; Bauer and Curran 2004; Dolan and van der Maas 1998; Jedidi, Jagpal, and DeSarbo 1997a; Jedidi *et al.* 1997b; Muthén 2001; Pek, Sterba, Kok, and Bauer 2009; Pek, Losardo, and Bauer 2011). In this approach, finite mixtures of linear SEMs are used to approximate unknown nonlinear relationships of the latent variables. Because the functional form of their relationship is not specified a priori in this case, the SEMM approach does not require assumptions about this functional form. The primary scope of these kinds of models is to identify curvilinear relationships among latent variables. Due to the semiparametric approximation of nonlinear relationships, no effect sizes can be calculated, nor is it possible to test specific hypotheses concerning the functional form of the relationship at present (i.e., it is not possible to infer a specific parametric model from the curvilinearity that is extracted, but see Bauer *et al.* 2012, for two approaches how to evaluate nonlinearity in the data visually).

Recently, a method was proposed that combines parametric and semiparametric models to estimate nonlinear effects (Kelava and Nagengast 2012; Kelava and Brandt 2014; Kelava *et al.* 2014). These models allow for an analysis of specific functional forms of relationships among latent variables in the mixture modeling framework while relaxing the distributional assumptions that are needed for the parametric models. This approach is an extension of the semiparametric approaches (e.g., Bauer 2005; Jedidi *et al.* 1997b) and can therefore not only be used to model nonlinear data in a semiparametric way but also has the additional capability to define the functional form of the nonlinear effect and estimate its effect size. In the following, parametric and semiparametric approaches to dealing with nonlinearity in a SEM framework will be introduced in more detail.

### 1.1. Parametric approach

A structural model that includes different parametric nonlinear effects is modeled by

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}_1\boldsymbol{\xi} + \boldsymbol{\Gamma}_2 h(\boldsymbol{\xi}) + \boldsymbol{\zeta}, \tag{1}$$

where $\boldsymbol{\alpha}$ is the latent intercept vector, $\boldsymbol{B}$ is a matrix of coefficients for the dependent variables, $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ are coefficient matrices that include the linear and nonlinear effects of the independent variables, respectively, $\boldsymbol{\xi}$ is a vector of latent predictor variables, $\boldsymbol{\eta}$ is a latent outcome variable vector, and $\boldsymbol{\zeta}$ is a latent residual vector. The function $h(\cdot)$ maps the vector $\boldsymbol{\xi}$ to a vector of product terms; it includes, for example, $\xi_1\xi_2$ for an interaction effect or $\xi_1^2$ for a quadratic effect. The measurement model is defined as

$$\mathbf{x} = \boldsymbol{\nu}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{2}$$

$$\mathbf{y} = \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{3}$$

with intercepts $\boldsymbol{\nu}_x$ and $\boldsymbol{\nu}_y$, factor loading matrices $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_y$, and measurement error variable vectors $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$.

An efficient maximum-likelihood (ML) estimator that fits the model in Equation 1 is the LMS approach (Klein and Moosbrugger 2000), and an alternative quasi-ML estimator is QML (Klein and Muthén 2007). The advantage of LMS is that it is theoretically more efficient, while QML is more robust to distributional assumptions (Kelava *et al.* 2011). Furthermore, QML is considerably less computationally intensive than LMS. However, both approaches assume that the latent variables $\boldsymbol{\xi}$ as well as the residual variables $\boldsymbol{\zeta}, \boldsymbol{\delta}, \boldsymbol{\varepsilon}$ are multivariate normally distributed. If this is not the case, different approaches may be needed.

## 1.2. Semiparametric approach

The SEMM approach is defined as a traditional linear SEM with latent classes (indicated by the subscript $c$)

$$\boldsymbol{\eta}_c = \boldsymbol{\alpha}_c + \boldsymbol{B}_c \boldsymbol{\eta}_c + \boldsymbol{\Gamma}_{1,c} \boldsymbol{\xi}_c + \boldsymbol{\zeta}_c. \tag{4}$$

The latent classes are modeled as normal mixtures with class-specific expected values ($\mu_c$) and covariance matrices ($\Sigma_c$) for the observed variable vector ($\boldsymbol{x}, \boldsymbol{y}|c$) within each class. The SEMM approach does not require the assumption of normally distributed latent variables, but allows for flexible approximations of nonnormal distributions. In a *direct* application of the model, the latent classes are interpreted as subgroups with class-specific linear relationships (Dolan and van der Maas 1998; Titterington, Smith, and Makov 1985). In an *indirect* application, the model can be used to approximate curvilinear relationships by means of a weighted average of the linear effects across classes (Bauer 2005; Bauer and Curran 2004; Dolan and van der Maas 1998; Pek *et al.* 2011, 2009). Thus, the SEMM approach can be used for predicting nonnormal latent dependent variables when obtaining an explicit parametric formulation of the functional relationship is not of primary interest.

An integration of parametric and semiparametric approaches that allows for the explicit formulation of nonlinear relations and simultaneously accounts for nonnormality of the data with semiparametric mixture models is called NSEMM (nonlinear structural equation mixture models; Kelava *et al.* 2014). It extends Equation 1 by adding a latent class framework (indicated by the subscript $c$)

$$\boldsymbol{\eta}_c = \boldsymbol{\alpha}_c + \boldsymbol{B}_c \boldsymbol{\eta}_c + \boldsymbol{\Gamma}_{1,c} \boldsymbol{\xi}_c + \boldsymbol{\Gamma}_{2,c} h_c(\boldsymbol{\xi}_c) + \boldsymbol{\zeta}_c. \tag{5}$$

The model can be applied in two different ways. First, it can be used to identify latent classes within a heterogeneous population with class-specific nonlinear relationships. In this kind of direct application, the mixture components are interpreted as representing distinct subpopulations. Second, when the regression models are constrained across classes ($\boldsymbol{\alpha}_c = \boldsymbol{\alpha}, \boldsymbol{B}_c = \boldsymbol{B}, \boldsymbol{\Gamma}_{1,c} = \boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_{2,c} = \boldsymbol{\Gamma}_2$, and $h_c(\cdot) = h(\cdot)$), the model can be used to approximate nonnormal distributions per se, without assuming meaningful distinct subgroups in a population (McLachlan and Peel 2000). This indirect application has the advantage of providing a semiparametric alternative to parametric nonlinear models when distributional assumptions for traditional SEMs are violated and when a straightforward interpretation of the nonlinear effects is still possible.

The remainder of this article is organized as follows: In the next section, we will give a brief overview of current latent variable modeling software with a particular focus on its capability to estimate nonlinear structural equation models. Then, we will introduce the

| Software | PI approaches | QML | LMS | SEMM | NSEMM | 2SMM | MM |
|---|:---:|:---:|:---:|:---:|:---:|:---:|:---:|
| **AMOS** | ✓ | – | – | – | – | – | – |
| **EQS** | ✓ | – | – | – | – | – | ✓ |
| **LISREL** | ✓ | – | – | – | – | – | – |
| **Mplus** | ✓ | – | ✓ | ✓ | ✓ | – | – |
| R: **lavaan/OpenMx/sem** | ✓ | – | – | – | – | – | – |
| R: **nlsem** | ✓ | ✓ | ✓ | ✓ | ✓ | – | – |
| **SAS** | ✓ | – | – | – | – | ✓ | – |
| Experimental | – | – | – | – | – | ✓[1] | ✓[2] |

Table 1: Overview of latent variable software capable of estimating nonlinear latent structural relationships. Note: PI approaches = product-indicator approaches, '✓' = available, '–' = not available, 1 = experimental syntax in R, 2 = experimental syntax in MATLAB.

R (R Core Team 2017) package **nlsem** (Umbach *et al.* 2017) for fitting nonlinear structural equation (mixture) models. **nlsem** can model nonlinear structural equation (mixture) models as depicted in Equations 1, 4, and 5 with the (current) limitation that $\eta$ is a scalar for Equations 1 and 5. The functionality of the package will then be demonstrated using two empirical examples. Package **nlsem** is available from the Comprehensive R Archive Network (CRAN) at `https://CRAN.R-project.org/package=nlsem`.

# 2. Other software packages for fitting nonlinear SEMs

As can be seen in Table 1, there are several software packages that are capable of estimating nonlinear structural equation models. Within the class of *parametric* nonlinear models, traditional product-indicator approaches (Ping 1995; Marsh *et al.* 2004; Kelava and Brandt 2009) that rely on the inclusion of (manifest) product-indicator variables to identify the latent product terms can be feasibly implemented in standard SEM software packages such as **LISREL** (Jöreskog, Sörbom, Toit, and Toit 1999), **EQS** (Bentler 2000–2008), **Mplus** (Muthén and Muthén 1998–2012), **AMOS** (Arbuckle 2011), and in the R packages **lavaan** (Rosseel 2012), **sem** (Fox 2006), and **OpenMx** (Boker, Neale, Maes, Wilde, Spiegel, Brick, Spies, Estabrook, Kenny, Bates, Mehta, and Fox 2011). When more than a single nonlinear effect is estimated simultaneously or when variables are nonnormally distributed, a detailed specification of the residual covariance matrix is necessary (see Kelava and Brandt 2009).

Distribution analytic approaches have been implemented for LMS (Klein and Moosbrugger 2000) in **Mplus** (using the `XWITH` command) and for QML (Klein and Muthén 2007) in a standalone software package.

In the class of moment-based approaches, the method of moments approach (MM; Mooijaart and Bentler 2010) has been implemented in **EQS** (Mooijaart and Bentler 2010) and MATLAB (MATLAB 2010; Brandt, Kelava, and Klein 2014). For the two-stage method of moments (2SMM; Wall and Amemiya 2000, 2003), syntax is available for an interaction or a quadratic model in **SAS** (SAS Institute Inc. 2003; Wall and Amemiya 2003) and for a model that includes both interaction and quadratic effects in an experimental R syntax (Brandt *et al.* 2014).

Bayesian approaches (Arminger and Stein 1997; Feng, Wang, Wang, and Song 2015; Kelava and Nagengast 2012; Kelava *et al.* 2014; Lee 2007; Song and Lu 2010) can be feasibly imple-

mented in Bayesian software packages like **WinBUGS** (Lunn, Thomas, Best, and Spiegelhalter 2000) or **OpenBUGS** (Lunn, Spiegelhalter, Thomas, and Best 2009).

*Semiparametric* SEMMs can be estimated within a ML framework in **Mplus** or within a Bayesian framework in **WinBUGS** or **OpenBUGS**. The nonlinear extensions of the SEMM framework (Kelava and Nagengast 2012; Kelava and Brandt 2014; Kelava *et al.* 2014) can also be specified in these software packages.

As can be seen in the brief description above, there is a wide variety of other software packages that can be used to estimate nonlinear SEMs. Although, the one with the most capabilities is probably **Mplus**, R packages such as **lavaan** and **OpenMx** are growing rapidly. Not only can R be regarded as state-of-the-art statistical software but it also has the additional features of great data manipulation capabilities and graphical facilities. It is therefore desirable to make the full functionality for fitting linear as well as nonlinear SEMs available in R. The R package **nlsem** introduced in this paper provides a wide range of methods dealing with nonlinearity and nonnormality in the SEM framework. It has an easy-to-use user interface and aims to complement R packages with greater functionality such as **lavaan**.

# 3. Implementation in R

**nlsem** is a straightforward implementation of the equations provided in Klein and Moosbrugger (2000, LMS), Klein and Muthén (2007, QML), Jedidi *et al.* (1997b, SEMM), and Kelava *et al.* (2014, NSEMM). This section will describe these approaches for fitting nonlinear structural equation models in detail and demonstrate how LMS, QML, SEMM, and NSEMM are implemented. In each of these papers – with the exception of the QML approach – an expectation-maximization (EM) algorithm (Dempster, Laird, and Rubin 1977) is used for parameter estimation. The EM algorithm is a common tool to estimate parameters for finite mixture models (McLachlan and Peel 2000). As a first step, posterior probabilities for the $r$th iteration

$$\mathsf{P}^{(r)}(\mathsf{c} = c|x_i, y_i) = \frac{w_c f_c(x_i, y_i|\mu_c, \Sigma_c)}{f(x_i, y_i)} \tag{6}$$

are computed with a set of starting parameters. $f(x_i, y_i)$ denotes the joint density of a mixture distribution of multivariate normal distributions $f_c(x_i, y_i|\mu_c, \Sigma_c)$ of the indicator variables **x** (for the exogenous latent variables) and **y** (for the endogenous latent variables), with $i = 1, \ldots, N$, where $N$ is the number of observations and $c = 1, \ldots, C$ indexes the mixture components. For SEMMs this is equivalent to the latent class. The weights of the normal mixture components are denoted by $w_c$. These posterior probabilities are understood as a function of the latent class $c$, indicated by the notation $\mathsf{c} = c$. Hence, posterior probabilities for each observation and each mixing proportion are computed. As a second step, the log-likelihood for $N$ observations and $C$ mixture components is maximized for the parameters provided in $\mu_c$ and $\Sigma_c$

$$\theta^{(r)} = \arg\max_{\theta} \left\{ \sum_{i=1}^{N} \sum_{c=1}^{C} \mathsf{P}^{(r)}(\mathsf{c} = c|x_i, y_i) \log f_c(x_i, y_i|\mu_c, \Sigma_c) \right\}, \tag{7}$$

where $f_c$ are normal distributions with $(x_i, y_i|c) \sim N(\mu_c, \Sigma_c)$.

In the next iteration of the EM algorithm, the parameters estimated in the second step are used to compute new posterior probabilities $\mathsf{P}^{(r+1)}$ (see Equation 6). This is repeated until

some criterion is reached. Usually, this means that either the loglikelihood or the parameters do not change more than some defined $\epsilon$. In the **nlsem** package, the convergence criterion is based on the changes in the loglikelihood (and defaults to `convergence = .01`).

## 3.1. Model specification for LMS and QML

Klein and Moosbrugger's LMS approach is a distribution analytic approach that approximates the complex nonnormal multivariate density function $f(x_i, y_i)$ of the indicator variables **x** and **y** (see Equation 15 in Klein and Moosbrugger 2000) with the Gauss-Hermite quadrature

$$f(x_i, y_i) \approx \sum_{j=1}^{M} w_j \pi^{-k/2} \varphi\left(x_i, y_i | \mu(2^{1/2}\nu_j), \Sigma(2^{1/2}\nu_j)\right),$$  (8)

where $w_j \pi^{-k/2}$ are the weights and $\mu(2^{1/2}\nu_j)$ and $\Sigma(2^{1/2}\nu_j)$ are the expected value and covariance matrix for the mixture components (see Equation 29 in Klein and Moosbrugger 2000). $M$ is the number of node points used by the Gauss-Hermite quadrature.[1] In this approach, the mixture components are not interpreted as subgroups, and the mixture weights are not estimated but provided by the Gauss-Hermite quadrature. The parameters are then estimated via an EM algorithm as described above.

The QML approach by Klein and Muthén (2007) can also be classified as a distribution analytic approach. In contrast to the LMS approach, the nonnormal density function is approximated by a product of a multivariate normal density function $f_2$ and a univariate conditionally normal density function $f_3^*$ (see Equation 4 in Klein and Muthén 2007)

$$f(x_i, y_i) = f_2(x_i, \mathbf{R}y_i) f_3(y_{1i}|x_i, \mathbf{R}y_i) \approx f_2(x_i, \mathbf{R}y_i) f_3^*(y_{1i}|x_i, \mathbf{R}y_i) =: f^*(x_i, y_i),$$  (9)

with conditional mean $\mathsf{E}[y_{1i}|x_i, \mathbf{R}y_i]$ and variance $\mathsf{VAR}(y_{1i}|x_i, \mathbf{R}y_i)$. Here, the nonnormal function $f_3$ is substituted by the normal density $f_3^*$ which both have the same conditional expectation and variance. Parameters are now estimated by maximizing the quasi-loglikelihood function for the density $f^*(x_i, y_i)$ by means of standard numerical methods. For QML, the nonnormality of the observed indicator vector **y** is transformed using the matrix **R** such that only the first indicator ($y_1$) remains nonnormally distributed and all other indicator variables **Ry** are normally distributed (for details see the Appendix in Klein and Muthén 2007).

A model with interaction effects can be fitted in **nlsem** via the function `specify_sem()`.

```
R> library("nlsem")
R> model1 <- specify_sem(num.x = 6, num.y = 3, num.xi = 2, num.eta = 1,
+    xi = "x1-x3, x4-x6", eta = "y1-y3", num.classes = 1,
+    interaction = "xi1:xi2")
R> class(model1)

[1] "singleClass"
```

Assignment of x's to $\xi$'s is done with the argument `xi`. x's specified in the first slot before the comma measure $\xi_1$, while x's in the second slot measure $\xi_2$ and so on. The same applies

---

[1] The chosen value for $M$ should depend on the number of nonlinear effects in the model. Klein and Moosbrugger (2000) suggest $M = 16$ for one interaction effect, but $M = 24$ for two interaction effects.

to the assignment of y's to $\eta$'s in the argument `eta`. The function `specify_sem()` returns an object of class 'singleClass' when no latent classes are specified (meaning `num.classes = 1`). Models of class 'singleClass' can be fitted by using either the EM algorithm as introduced by Klein and Moosbrugger (2000, with function `em()`) or the QML method by Klein and Muthén (2007, with function `qml()`).

It should be pointed out that Klein and Moosbrugger (2000) and Klein and Muthén (2007) use a specific notation for the nonlinear effects with $\boldsymbol{\Gamma}_2 h(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top \boldsymbol{\Omega} \boldsymbol{\xi}$. Matrix $\boldsymbol{\Omega}$ includes the quadratic effects (e.g., $\omega_{11}$, $\omega_{22}$) on the diagonal and the interaction effects (e.g., $\omega_{12}$) above the diagonal. This notation is used by `specify_sem()` and limits the number of dependent latent variables to 1. This will be addressed in future releases. For more than one $\eta$, interactions need to be specified as follows: `interaction = "eta1~xi1:xi2, eta2~xi1:xi1"` with `specify_sem()` creating an array with a matrix $\Omega$ for each endogenous variable as a slice.

## 3.2. Model specification for SEMM

The semiparametric approach by Jedidi *et al.* (1997b) can also be fitted in **nlsem**. The class-conditional multivariate density function of **x** and **y** is given by (cf. Equation 7 in Jedidi *et al.* 1997b)

$$f_c(x_i, y_i) = (2\pi)^{-\left(\frac{p+q}{2}\right)} |\Sigma_c|^{-1/2} \exp\left\{ -\frac{1}{2} \left( (x_i, y_i)^\top - \mu_c \right)^\top \Sigma^{-1} \left( (x_i, y_i)^\top - \mu_c \right) \right\}, \quad (10)$$

with class-specific $\mu_c$ and $\Sigma_c$; $p$ and $q$ represent the number of indicators for the endogenous and exogenous latent variables, respectively. Then, the unconditional distribution of **x** and **y** is given by

$$f(x_i, y_i) = \sum_{c=1}^{C} w_c f_c(x_i, y_i), \quad (11)$$

with class weights $w_c$. As in Klein and Moosbrugger's LMS approach, the parameters in Jedidi *et al.*'s SEMM approach are estimated via an EM algorithm. Again, a model with two latent classes can be specified with `specify_sem()`; by specifying `interaction = "none"` and `num.classes = 2` (or more than 2) a model of class 'semm' is created, which will be fitted using the corresponding EM algorithm. Indirect and direct approaches as those mentioned above can be specified by setting the argument `constraints = c("indirect", "direct1", "direct2")`. For the indirect option, all parameters of Equations 2, 3, and 4 are constrained to be equal except for the mean vector and covariance matrix of the predictor variables. For `direct1` the parameters of all latent classes are freely estimated, while for `direct2` only the parameters for Equations 2 and 3 are constrained to be equal, thus allowing one to assume that the measurement model is identical for all classes (strict invariance; Meredith 1993).

```
R> model2 <- specify_sem(num.x = 3, num.y = 8, num.xi = 1, num.eta = 2,
+    xi = "x1-x3", eta = "y1-y4, y5-y8",
+    rel.lat = "eta1~xi1, eta2~xi1, eta2~eta1", num.classes = 2,
+    constraints = "direct1")
R> class(model2)


[1] "semm"
```

### 3.3. Model specification for NSEMM

The NSEMM approach introduced by Kelava *et al.* (2014) can be understood as a combination of the approaches introduced above. The joint density function for **x** and **y** is defined identically to Equation 11, but $f_c(\cdot)$ for each class corresponds to Equation 8 (or alternatively to Equation 9). In our implementation the parameters are first estimated by applying the LMS or the QML procedure within each latent class. These parameters are subsequently entered into the E-step of the EM algorithm for the SEMM approach in order to obtain class weights. Then, the M-step of the SEMM approach is applied and the likelihood is optimized. The resulting parameters are entered in the fitting algorithm for LMS or QML, respectively, for each latent class, and so on. A model with interaction effects and latent classes of class 'nsemm' can be specified as follows.

```
R> model3 <- specify_sem(num.x = 6, num.y = 3, num.xi = 2, num.eta = 1,
+    xi = "x1-x3, x4-x6", eta = "y1-y3", num.classes = 2,
+    interaction = "xi1:xi2", rel.lat = "eta1~xi1+xi2",
+    constraints = "indirect")
R> class(model3)

[1] "nsemm"
```

### 3.4. Model specification using lavaan syntax

As of **nlsem** version 0.7 and **lavaan** version 0.5-21, **lavaan** syntax for specifying structural equation models with latent interactions can be used to specify all models introduced in the previous sections. The following shows an example for a NSEMM model with three latent classes and different nonlinear effects in each class.

```
R> lav.model <- '
    class: 1
    eta =~ y1 + y2 + y3 + y4
    xi1 =~ x1 + x2 + x3 + x4 + x5
    xi2 =~ x6 + x7 + x8 + x9 + x10 + x11

    eta ~ xi1 + xi2 + xi1:xi1

    class: 2
    eta =~ y1 + y2 + y3 + y4
    xi1 =~ x1 + x2 + x3 + x4 + x5
    xi2 =~ x6 + x7 + x8 + x9 + x10 + x11

    eta ~ xi1 + xi2 + xi1:xi2 + xi1:xi1

    class: 3
    eta =~ y1 + y2 + y3 + y4
    xi1 =~ x1 + x2 + x3 + x4 + x5
    xi2 =~ x6 + x7 + x8 + x9 + x10 + x11
```

```
    eta ~ xi1 + xi2 + xi1:xi2'

R> model <- lav2nlsem(lav.model, constraints = "direct1", class.spec = "class")
```

Function `lav2nlsem()` can be used to generate model matrices from a string using **lavaan** notation. The resulting object has the same structure as an object created with `specify_sem()` and can then be handed to functions `em()` or `qml()` (see the following section for details). Please note, that it is only feasible to specify models in **lavaan** syntax that can be fitted with **nlsem**, meaning only models with continuous latent variables and latent interactions. Also, the full functionality of parameter restrictions that **lavaan** provides cannot be handled by **nlsem**. Equality restrictions, for example, are handled by the `constraints` argument in `em()` and will be ignored in a model specified in **lavaan syntax**. Simply fixing certain parameters to predefined values is, however, possible. The following code snippet shows an example. With `model$matrices` one can look at the parameter restrictions within the matrices.

```
R> lav.model <- '
    eta =~ 1*y1 + 1*y2 + y3 + y4
    xi1 =~ x1 + x2 + x3 + x4 + x5
    xi2 =~ x6 + x7 + x8 + x9 + x10 + x11

    eta ~ xi1 + 0*xi2 + xi1:xi1

R> model <- lav2nlsem(lav.model, constraints = "direct1", class.spec = "class")
R> model$matrices
```

### 3.5. Fitting the models

`specify_sem()` sets default parameter constraints for model identification. For most models, no additional constraints need to be considered. However, identifiability of a model is not automatically checked. For complex models, this needs to be taken into consideration by the user. The default constraints set by `specify_sem()` are:

- First factor loading for each $\xi$ and $\eta$ is fixed to 1.

- First intercept for each $\xi$ and $\eta$ is fixed to 0.

- $\Theta_\delta$ and $\Theta_\varepsilon$ are diagonal matrices.

Additionally, all constraints can be adjusted manually. The easiest way to do this is to use functions `as.data.frame()` and `create_sem()`. The following shows an example of how to manually constrain all factor loadings for the indicators of $\xi_1$ and $\xi_2$ to 1. First, the model is specified and then transformed to a data frame with `as.data.frame()`, where the first column holds the parameter labels, the second column holds the values for `class1` and so on. `NA` denotes that the parameter will be freely estimated. Second, the values for the parameters can be adjusted manually by changing the values in the respective column. Then, the function `create_sem()` can be used to create a new model with custom constraints.

```
R> model <- specify_sem(num.x = 6, num.y = 3, num.xi = 2, num.eta = 1,
+    xi = "x1-x3, x4-x6", eta = "y1-y3", interaction = "xi1:xi2")
R> specs <- as.data.frame(model)
R> head(specs)

      label class1
1 Lambda.x1      1
2 Lambda.x2     NA
3 Lambda.x3     NA
4 Lambda.x4      0
5 Lambda.x5      0
6 Lambda.x6      0

R> specs[specs$label %in% paste0("Lambda.x", c(2, 3, 11, 12)), "class1"] <- 1
R> model.custom <- create_sem(specs)
```

Each of the models specified above can be fitted with the function `em()`. Depending on the model class, the EM algorithms as introduced in Klein and Moosbrugger (2000) and Jedidi *et al.* (1997b) or a combination of both for the NSEMM model will be used. The single-class model can additionally be fitted using the function `qml()`. Below, we will illustrate how the `em()` function can be used with a standard example from the SEM literature (Bollen 1989). The code fits the SEMM with two latent classes introduced above (`model2`).

```
R> data("PoliticalDemocracy", package = "lavaan")
R> dat <- as.matrix(PoliticalDemocracy[, c(9:11, 1:8)])
R> set.seed(911)
R> pars.start <- runif(count_free_parameters(model2))
R> res <- em(model2, dat, pars.start, convergence = 0.1, max.iter = 200)
R> summary(res)
R> plot(res)
```

The object created with `em()` gives back the parameter estimates, the number of iterations, the final loglikelihood as well as the loglikelihoods obtained in each iteration step of the EM algorithm. It contains an approximation of the observed information and some information pertaining to the fitted model such as how many indicators and latent variables were specified. On this object the usual extractor functions like `summary()`, `coef()`, `AIC()`, `BIC()`, and `logLik()` can be used. For nested models of class 'singleClass' fitted with `em()`, the function `anova()` performs a likelihood ratio test. There is also a plot method for objects of class 'emEst' that plots the loglikelihoods for each iteration of the EM algorithm. This allows for a visual inspection of the convergence of the algorithm. In order to simulate data from a certain model, one can use the function `simulate()`, which takes a model specified with `specify_sem()` and an argument `parameters` with a vector of true parameter values.

## 4. Computational details

**nlsem** depends on the R packages **gaussquad** (Novomestky 2013) and **mvtnorm** (Genz and Bretz 2009; Genz, Bretz, Miwa, Mi, Leisch, Scheipl, and Hothorn 2017). Both are needed for

|  | Iterations | | Time (sec) | | |
|---|---|---|---|---|---|
|  | Mean | SD | Mean | SD | Converged |
| QML | 111.071 | 6.944 | 15.611 | 1.749 | 14 |
| LMS | 16.700 | 4.555 | 26.491 | 5.678 | 20 |
| SEMM | 33.214 | 11.450 | 22.170 | 6.092 | 14 |
| NSEMM (LMS) | 28.700 | 14.217 | 136.841 | 66.787 | 20 |
| NSEMM (QML) | 38.800 | 10.414 | 35.588 | 9.467 | 15 |

Table 2: Some information on computational resources needed to fit the models implemented in **nlsem**. For each model, models were fitted 20 times with varying starting parameters. Mean values are shown for number of iterations and running time.

parameter estimation within the EM algorithms. In order to obtain the observed information necessary to estimate the standard errors for parameter estimates, the function `fdHess()` is imported from **nlme** (Pinheiro, Bates, DebRoy, Sarkar, and R Core Team 2017).

Parameters in the M-step of the EM algorithm can be estimated with two different optimizers. Estimation of variances is bound so that negative variances cannot occur. Therefore, the user can choose between `nlminb()` and `optim()` (with `method = "L-BFGS-B"`), both of which allow for lower (and upper) bounds during parameter estimation. Choosing the optimizer can be done by setting the argument `optimizer = c("nlminb", "optim")`.

As mentioned above, **nlsem** is a straightforward implementation of the equations provided in Klein and Moosbrugger (2000), Klein and Muthén (2007), Jedidi *et al.* (1997b), and Kelava *et al.* (2014) directly in R. These equations are given in matrix notation, which greatly facilitates implementation. However, SEMs generally come with a great number of parameters that need to be estimated. When we apply an approach with two or three latent classes, the number of parameters doubles and triples, which suggests that the estimation of parameters in **nlsem** can be slow, especially for complex models.

Table 2 shows details on running times for what is sometimes referred to as the "elementary interaction model" (Klein and Moosbrugger 2000)

$$\eta = \alpha + \gamma_1\xi_1 + \gamma_2\xi_2 + \omega_{12}\xi_1\xi_2 + \zeta \tag{12}$$

fitted with LMS and QML. The NSEMM (with LMS or QML within the latent classes and indirect approach) was fitted to the following model

$$\eta_c = \alpha + \gamma_1\xi_{1,c} + \gamma_2\xi_{2,c} + \omega_{12}\xi_{1,c}\xi_{2,c} + \zeta, \tag{13}$$

with $c = 1, 2$; for the SEMM the interaction term was omitted and parameters were fitted with a direct approach

$$\eta_c = \alpha_c + \gamma_{1,c}\xi_{1,c} + \gamma_{2,c}\xi_{2,c} + \zeta_c. \tag{14}$$

The models were fitted 20 times to data simulated from the model presented in Equation 12 for 400 observations. The convergence criterion value was set to 1. The results show that choosing the starting parameters influences not only how long the models need to converge but also whether they converge. This is a common problem for multidimensional optimization (Jedidi *et al.* 1997b). The LMS approach is most robust to different starting parameters and almost always converges. In our example, all NSEMM models with LMS converged as well. The EM algorithm is known to converge slowly, so setting a laxer convergence criterion

greatly speeds up the model fitting. This particular approach could be appropriate when several models are fitted and compared to each other. For a final parameter estimation for the chosen model, convergence should be set to a smaller value (e.g., its default `convergence = .01`).

For more complex models, running time increases considerably. The results in Table 2 show that QML is indeed much faster than LMS. Since NSEMM uses finite mixtures of the distribution analytic approaches to calculate latent interaction effects, QML can also speed up the estimation for these models considerably. The user can decide which approach to use by setting `qml = TRUE` when using `em()` for fitting a model of class 'nsemm'. The NSEMM model with LMS takes the longest to converge. Setting `qml = TRUE` for the NSEMM yields faster results but is more sensitive to the choice of starting parameters.

The next section demonstrates the functionality of **nlsem** using two data examples from education research. Data analyses for the four possible approaches implemented in **nlsem** (QML, LMS, SEMM, and NSEMM) were conducted.

# 5. Empirical examples with nlsem

## 5.1. Predictors of career aspirations in science

The data stem from the large-scale assessment study Program for International Student Assessment 2006 (PISA; Organisation for Economic Co-Operation and Development 2009) in which 15-year-old students' reading, mathematics and science literacy is assessed using nationally representative samples in 3-year cycles. In this example, data from the student background questionnaire for PISA 2006 were used. In line with Nagengast, Marsh, Scalas, Xu, Hau, and Trautwein (2011) and Kelava *et al.* (2014), we tested the relation between the following variables: career aspirations in science ($\eta$, 4 items), enjoyment of science ($\xi_1$, 5 items), and academic self-concept in science ($\xi_2$, 6 items)[2]. The data stem from the Jordan sample of PISA 2006. Responses to the questionnaire items used as indicators for academic self-concept and enjoyment were nonnormally distributed. Only data of students with complete responses to all 15 items ($N = 6,038$) were considered. The analyses are not representative but instead serve as an illustration of nonnormal latent predictor distributions on parameter estimates in nonlinear SEMs.

First, a model with one interaction effect was fitted to the data applying the LMS approach

$$CAREER = \alpha + \gamma_1 ENJ + \gamma_2 SC + \omega_{11} ENJ^2 + \omega_{12} ENJ \cdot SC + \omega_{22} SC^2 + \zeta, \qquad (15)$$

with *CAREER* being career aspirations in science, *ENJ* enjoyment of science, and *SC* academic self-concept in science. As can be seen in Table 3 (first panel), enjoyment ($\gamma_1 = .523$, $p < .001$) and self-concept ($\gamma_2 = .456, p < .001$) showed effects on career as well as small but significant interaction and quadratic effects ($\omega_{12} = -.062$, $p = .007$, $\omega_{11} = .033$, $p = .011$, $\omega_{22} = .014$, $p = .368$).

Since the indicators for the predictor variables were nonnormal, assumptions to apply LMS

---

[2]Career aspirations: items ST29Q01–ST29Q04; enjoyment of science: items ST16Q01–ST16Q05; self-concept in science: items ST37Q01–ST37Q06.

|       |              | Estimate | Std. Error | z value  | $P(> \lvert z \rvert)$ |
|-------|--------------|----------|------------|----------|------------|
| LMS   | $\alpha$     | $-.009$  | .011       | -.844    | .398       |
|       | $\gamma_1$   | .523     | .017       | 31.144   | $< .001$   |
|       | $\gamma_2$   | .456     | .019       | 24.203   | $< .001$   |
|       | $\omega_{11}$ | .033    | .013       | 2.542    | .011       |
|       | $\omega_{12}$ | $-.062$ | .023       | $-2.683$ | .007       |
|       | $\omega_{22}$ | .014    | .015       | .900     | .368       |
| SEMM  | $P(c = 1)$   | .396     |            |          |            |
|       | $\alpha_1$   | $-.153$  | .025       | $-6.049$ | $< .001$   |
|       | $\gamma_{1,1}$ | .399   | .029       | 13.889   | $< .001$   |
|       | $\gamma_{2,1}$ | .335   | .029       | 11.607   | $< .001$   |
|       | $P(c = 2)$   | .604     |            |          |            |
|       | $\alpha_2$   | .093     | .012       | 7.652    | $< .001$   |
|       | $\gamma_{1,2}$ | .595   | .023       | 26.421   | $< .001$   |
|       | $\gamma_{2,2}$ | .241   | .021       | 11.701   | $< .001$   |
| NSEMM | $\alpha$     | .000     | .003       | .133     | .894       |
|       | $\gamma_1$   | .521     | .003       | 195.710  | $< .001$   |
|       | $\gamma_2$   | .442     | .003       | 143.159  | $< .001$   |
|       | $\omega_{11}$ | .014    | .003       | 4.793    | $< .001$   |
|       | $\omega_{12}$ | $-.008$ | .003       | $-2.774$ | .006       |
|       | $\omega_{22}$ | $-.019$ | .005       | $-3.702$ | $< .001$   |

Table 3: Parameter estimates for example analyses. Standard errors for the NSEMM model were bootstrapped.

were violated. Therefore, we fitted a SEMM with two latent classes ($c = 1, 2$) to the data

$$CAREER_c = \alpha_c + \gamma_{1,c}ENJ_c + \gamma_{2,c}SC_c + \zeta_c. \tag{16}$$

Results are shown in Table 3 (second panel). Again, we found that enjoyment and self-concept both had an effect on career aspiration (see Figure 1). Additionally, the SEMM approach revealed two groups with different sets of significant parameters, thus addressing the nonnormality and nonlinearity in the data.

Finally, we fitted a model with two latent classes, one interaction effect, and two quadratic effects to the data thus applying the NSEMM approach. We used the indirect approach, meaning that parameteres are constrained to be equal across classes

$$CAREER_c = \alpha + \gamma_1 ENJ_c + \gamma_2 SC_c + \omega_{11}ENJ_c^2 + \omega_{12}ENJ_c \cdot SC_c + \omega_{22}SC_c^2 + \zeta. \tag{17}$$

Results in Table 3 (third panel) show significant main effects for the indirect NSEMM approach ($\gamma_1 = .521$, $p < .001$, $\gamma_2 = .442$, $p < .001$). Also the interaction and quadratic effects are significant ($\omega_{12} = -.008$, $p = .006$, $\omega_{11} = .014$, $p < .001$, $\omega_{22} = -.019$, $p < .001$). Due to the manner in which the EM algorithm for NSEMM models is implemented, it is not possible to approximate the observed information based on the likelihood function. Therefore, standard errors cannot be obtained by taking the inverse of the observed information. However, standard errors can easily be bootstrapped with the functions provided in **nlsem** (Efron 1981); see Table 3 for the standard errors obtained with nonparametric bootstrapping. Please note that this, though simple, can be time-consuming.
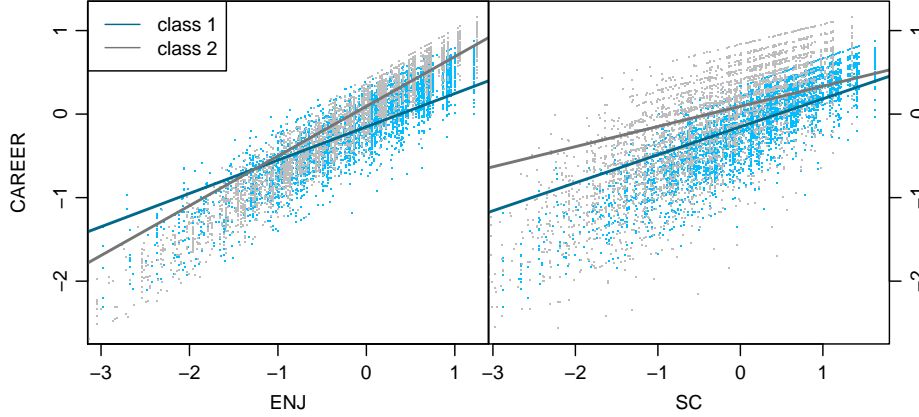
Figure 1: Scatterplots for latent predictors. Both self-concept ($SC$) and enjoyment of science ($ENJ$) influence career aspirations in science ($CAREER$). Result patterns look different for latent groups extracted by SEMM.

|            | Estimate | Std. Error | z value | $P(>|z|)$ |
|------------|----------|------------|---------|-----------|
| $\alpha$   | .305     | .145       | 2.094   | .036      |
| $\gamma_1$ | $-.078$  | .076       | $-1.020$| .308      |
| $\gamma_2$ | .238     | .076       | 3.147   | .002      |
| $\omega_{11}$ | .023  | .015       | 1.580   | .114      |
| $\omega_{12}$ | .041  | .023       | 1.788   | .074      |
| $\omega_{22}$ | $-.048$ | .016     | $-3.090$| .002      |

Table 4: Parameter estimates for the second example analysis. Parameters were estimated with QML.

## 5.2. Attitude towards reading, online activities, and reading skill

The second example is based on data from the Program for International Student Assessment 2009 (PISA; Organisation for Economic Co-Operation and Development 2010), which is publicly available under http://pisa2009.acer.edu.au/downloads.php. The sample is a subsample of $N = 1,069$ Australian students who took part in a reading test. Missing values were excluded from the analysis. Students' attitude towards reading ($Att$) and their reported online activities ($Online$; i.e., read emails or chat online) were selected as predictors of reading skill ($Read$).

Between 7 and 11 items were grouped together to make up three indicators for each latent variable (i.e., item parcels), which resulted in three indicators for each latent variable.[3] Again, this example merely serves to illustrate the use of the package and was chosen for didactic purposes only. The structural model was specified as

$$Read = \alpha + \gamma_1 Att + \gamma_2 Online + \omega_{11} Att^2 + \omega_{12} Att \cdot Online + \omega_{22} Online^2 + \zeta, \qquad (18)$$

with a normally distributed residual $\zeta \sim N(0, \psi)$ and predictors $(Att, Online)^\top \sim N(\boldsymbol{\tau}, \boldsymbol{\Phi})$. Parameters were estimated using QML. The results depicted in Table 4 show that read-

---

[3]Reading skill: average of items R06, R102, R219, R220, R414, R447, R452, and R458. Attitude towards reading: average of items ST24Q01–ST24Q11. Online activities: average of items ST26Q01–ST26Q07.
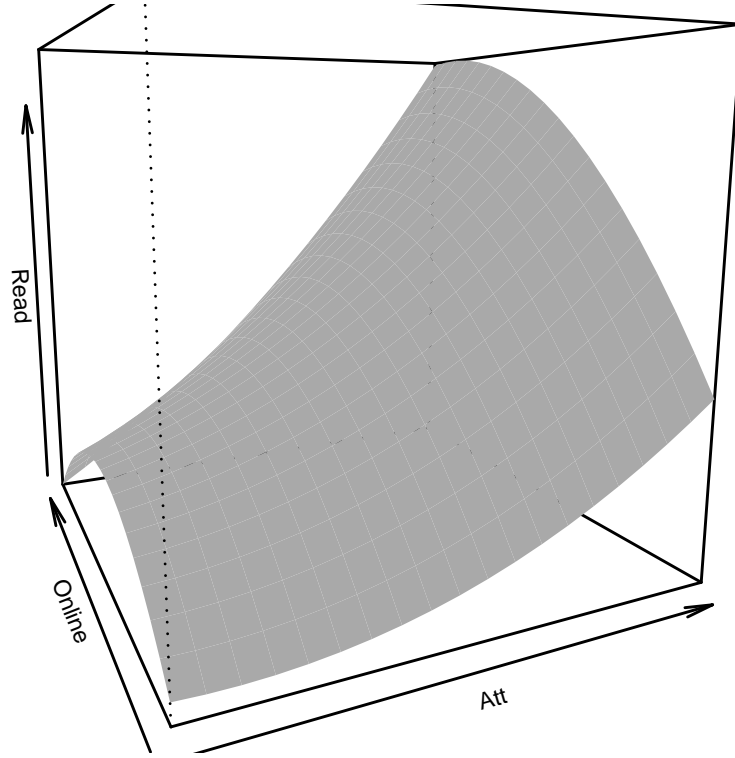
Figure 2: Predicted plane for estimated parameters shown in Table 4 showing how attitude towards reading (*Att*) and online activities (*Online*) predict reading skill (*Read*).

ing attitude and online activities influence reading skill nonlinearly ($\omega_{12} = .041$, $p = .074$, $\omega_{22} = -.048$, $p = .002$). Figure 2 shows that reading skill is highest when online activities are at a medium level. For high and low amounts of online activities reading skill is low when attitude towards reading is also low. Additionally, how much attitude towards reading increases reading skill seems to depend on the level of online activities: for higher online activities reading attitude has a higher influence.

# 6. Conclusion and outlook

This paper introduces the R package **nlsem** which provides an open source implementation to fit nonlinear SEMs and structural equation mixture models with an easy-to-use interface. The availability of several kinds of nonlinear approaches, such as LMS, QML, SEMM and NSEMM, to fit nonlinear SEMs in one package greatly facilitates data analysis for nonlinear data within the SEM framework and allows for an easy comparison of results for different approaches.

The user can choose between direct and indirect approaches for structural equation mixture models, using these approaches to model nonlinearity, nonnormality, or both. Two direct approaches are implemented where separate sets of coefficients are estimated for the prespecified latent classes. For the indirect approach, parameters are constrained to be equal for all latent classes, thereby modeling nonnormality in the data without strong assumptions about the interpretability of latent classes.

So far, LMS and QML have only been theoretically developed for one $\eta$ (Klein and Moosbrugger 2000; Klein and Muthén 2007). The equations and algorithms will be expanded in future relaeses so that more than one endogenous variable can be included in the model. The model specification can already handle the specification of interaction effects between latent predictors that influence different endogenous variables, meaning that the changes will expand the functionality of the package without changing the overall structure of the package.

Additionally, since standardized coefficients are often calculated in SEMs, a function (or argument) for parameter standardization will be added to **nlsem** in the future (Brandt, Umbach, and Kelava 2015). This standardization will allow for the comparison of linear and nonlinear effects across studies and thus facilitate meta studies.

# Acknowledgments

# References

Arbuckle JL (2011). *IBM SPSS AMOS 20 User's Guide.* Armonk.

Arminger G, Muthén BO (1998). "A Bayesian Approach to Nonlinear Latent Variable Models Using the Gibbs Sampler and the Metropolis-Hastings Algorithm." *Psychometrika*, **63**(3), 271–300. `doi:10.1007/bf02294856`.

Arminger G, Stein P (1997). "Finite Mixtures of Covariance Structure Models with Regressors." *Sociological Methods and Research*, **26**(2), 148–182. `doi:10.1177/0049124197026002002`.

Arminger G, Stein P, Wittenberg J (1999). "Mixtures of Conditional Mean- And Covariance-Structure Models." *Psychometrika*, **64**(4), 475–494. `doi:10.1007/bf02294568`.

Bauer DJ (2005). "A Semiparametric Approach to Modeling Nonlinear Relations among Latent Variables." *Structural Equation Modeling: A Multidisciplinary Journal*, **12**(4), 513–535. `doi:10.1207/s15328007sem1204_1`.

Bauer DJ (2007). "Observations on the Use of Growth Mixture Models in Psychological Research." *Multivariate Behavioral Research*, **42**(4), 757–786. `doi:10.1080/00273170701710338`.

Bauer DJ, Baldasaro R, Gottfredson NC (2012). "Diagnostic Procedures for Detecting Nonlinear Relationships between Latent Variables." *Structural Equation Modeling: A Multidisciplinary Journal*, **19**(2), 157–177. `doi:10.1080/10705511.2012.659612`.

Bauer DJ, Curran PJ (2004). "The Integration of Continous and Discrete Latent Variable Models: Potential Problems and Promising Opportunities." *Psychological Methods*, **9**(1), 3–29. `doi:10.1037/1082-989x.9.1.3`.

Bentler PM (2000–2008). ***EQS** 6 Structural Equations Program Manual.* Multivariate Software, Inc., Encino, CA.

Boker S, Neale M, Maes H, Wilde M, Spiegel M, Brick T, Spies J, Estabrook R, Kenny S, Bates T, Mehta P, Fox J (2011). "**OpenMx**: An Open Source Extended Structural Equation Modeling Framework." *Psychometrika*, **76**(2), 306–317. `doi:10.1007/s11336-010-9200-6`.

Bollen KA (1989). *Structural Equations with Latent Variables.* John Wiley & Sons, New York. `doi:10.1002/9781118619179`.

Bollen KA (1995). "Structural Equation Models That Are Nonlinear in Latent Variables: A Least Squares Estimator." In PV Marsden (ed.), *Sociological Methodology 1995*, pp. 223–251. Blackwell, Oxford.

Brandt H, Kelava A, Klein AG (2014). "A Simulation Study Comparing Recent Approaches for the Estimation of Nonlinear Effects in SEM under the Condition of Non-Normality." *Structural Equation Modeling: A Multidisciplinary Journal*, **21**(2), 181–195. `doi:10.1080/10705511.2014.882660`.

Brandt H, Umbach N, Kelava A (2015). "The Standardization of Linear and Nonlinear Effects in Direct and Indirect Applications of Structural Equation Mixture Models for Normal and Nonnormal Data." *Frontiers in Psychology*, **6**(1813), 1–13. `doi:10.3389/fpsyg.2015.01813`.

Dempster AP, Laird NM, Rubin DB (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society B*, **39**(1), 1–38.

Dolan CV, van der Maas HLJ (1998). "Fitting Multivariate Normal Finite Mixtures Subject to Structural Equation Modeling." *Psychometrika*, **63**(3), 227–253. `doi:10.1007/bf02294853`.

Efron B (1981). "Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods." *Biometrika*, **68**(3), 589–599. `doi:10.1093/biomet/68.3.589`.

Feng XN, Wang GC, Wang YF, Song XY (2015). "Structure Detection of Semiparametric Structural Equation Models with Bayesian Adaptive Group Lasso." *Statistics in Medicine*, **34**(9), 1527–1547. `doi:10.1002/sim.6410`.

Fox J (2006). "Structural Equation Modeling with the **sem** Package in R." *Structural Equation Modeling: A Multidisciplinary Journal*, **13**(3), 465–486. `doi:10.1207/s15328007sem1303_7`.

Genz A, Bretz F (2009). *Computation of Multivariate Normal and t Probabilities.* Lecture Notes in Statistics. Springer-Verlag, Heidelberg. `doi:10.1007/978-3-642-01689-9`.

Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T (2017). ***mvtnorm**: Multivariate Normal and t Distributions.* R package version 1.0-6, URL `https://CRAN.R-project.org/package=mvtnorm`.

Jaccard J, Wan CK (1995). "Measurement Error in the Analysis of Interaction Effects between Continuous Predictors Using Multiple Regression: Multiple Indicator and Structural

Equation Approaches." *Psychological Bulletin*, **117**(2), 348–357. `doi:10.1037/0033-2909.117.2.348`.

Jedidi K, Jagpal HS, DeSarbo WS (1997a). "Finite Mixture Structural Equation Models for Response Based Segmentation and Unobserved Heterogeneity." *Marketing Science*, **16**(1), 39–59. `doi:10.1287/mksc.16.1.39`.

Jedidi K, Jagpal HS, DeSarbo WS (1997b). "STEMM: A General Finite Mixture Structural Equation Model." *Journal of Classification*, **14**(1), 23–50. `doi:10.1007/s003579900002`.

Jöreskog KG, Sörbom D, Toit SD, Toit MD (1999). ***LISREL*** *8: New Statistical Features*. Chicago: Scientific Software International, Lincolnwood, IL.

Jöreskog KG, Yang F (1996). "Nonlinear Structural Equation Models: The Kenny-Judd Model With Interaction Effects." In GA Marcoulides, RE Schumacker (eds.), *Advanced Structural Equation Modeling: Issues and Techniques*, pp. 57–87. Erlbaum, Mahwah, NJ.

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, Brandt H (2014). "A General Nonlinear Multilevel Structural Equation Mixture Model." *Frontiers in Psychology*, **5**(748), 1–16. `doi:10.3389/fpsyg.2014.00748`.

Kelava A, Nagengast B (2012). "A Bayesian Model for the Estimation of Latent Interaction and Quadratic Effects When Latent Variables Are Non-Normally Distributed." *Multivariate Behavioral Research*, **47**(5), 717–742. `doi:10.1080/00273171.2012.715560`.

Kelava A, Nagengast B, Brandt H (2014). "A Nonlinear Structural Equation Mixture Modeling Approach for Non-Normally Distributed Latent Predictor Variables." *Structural Equation Modeling: A Multidisciplinary Journal*, **21**(3), 468–481. `doi:10.1080/10705511.2014.915379`.

Kelava A, Werner C, Schermelleh-Engel K, Moosbrugger H, Zapf D, Ma Y, Cham H, Aiken LS, West SG (2011). "Advanced Nonlinear Structural Equation Modeling: Distribution Analytic LMS and QML Estimators of Interaction and Quadratic Effects." *Structural Equation Modeling: A Multidisciplinary Journal*, **18**(3), 465–491. `doi:10.1080/10705511.2011.582408`.

Kenny D, Judd CM (1984). "Estimating the Nonlinear and Interactive Effects of Latent Variables." *Psychological Bulletin*, **96**(1), 201–210. `doi:10.1037/0033-2909.96.1.201`.

Klein AG, Moosbrugger H (2000). "Maximum Likelihood Estimation of Latent Interaction Effects with the LMS Method." *Psychometrika*, **65**(4), 457–474. `doi:10.1007/bf02296338`.

Klein AG, Muthén BO (2007). "Quasi Maximum Likelihood Estimation of Structural Equation Models with Multiple Interaction and Quadratic Effects." *Multivariate Behavioral Research*, **42**(4), 647–674. `doi:10.1080/00273170701710205`.

Lee SY (2007). *Structural Equation Modeling: A Bayesian Approach*. John Wiley & Sons, Chichester, England.

Little TD, Bovaird JA, Widaman KF (2006). "On the Merits of Orthogonalizing Powered and Interaction Terms: Implications for Modeling Interactions among Latent Variables." *Structural Equation Modeling: A Multidisciplinary Journal*, **13**(4), 497–519. doi:10.1207/s15328007sem1304_1.

Lunn DJ, Spiegelhalter D, Thomas A, Best N (2009). "The BUGS Project: Evolution, Critique, and Future Directions." *Statistics in Medicine*, **28**(25), 3049–3067. doi:10.1002/sim.3680.

Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000). "**WinBUGS** – A Bayesian Modelling Framework: Concepts, Structure, and Extensibility." *Statistics and Computing*, **10**(4), 325–337. doi:10.1023/a:1008929526011.

Marsh HW, Wen Z, Hau KT (2004). "Structural Equation Models of Latent Interactions: Evaluation of Alternative Estimation Strategies and Indicator Construction." *Psychological Methods*, **9**(3), 275–300. doi:10.1037/1082-989x.9.3.275.

Marsh HW, Wen Z, Hau KT (2006). "Structural Equation Models of Latent Interaction and Quadratic Effects." In GR Hancock, RO Müller (eds.), *Structural Equation Modeling: A Second Course*, pp. 225–265. Information Age Publishing, Greenwich, CT.

McLachlan GJ, Peel D (2000). *Finite Mixture Models*. John Wiley & Sons, New York. doi:10.1002/0471721182.

Meredith W (1993). "Measurement Invariance, Factor Analysis and Factorial Invariance." *Psychometrika*, **58**(4), 525–543. doi:10.1007/bf02294825.

Mooijaart A, Bentler PM (2010). "An Alternative Approach for Nonlinear Latent Variable Models." *Structural Equation Modeling: A Multidisciplinary Journal*, **17**(3), 357–573. doi:10.1080/10705511.2010.488997.

Moulder BC, Algina J (2002). "Comparison of Methods for Estimating and Testing Latent Variable Interactions." *Structural Equation Modeling: A Multidisciplinary Journal*, **9**(1), 1–19. doi:10.1207/s15328007sem0901_1.

Muthén BO (2001). "Second-Generation Structural Equation Modeling with a Combination of Categorical and Continuous Latent Variables: New Opportunities for Latent-Class Growth Modeling." In LM Collins, A Sayer (eds.), *New Methods for the Analysis of Change*, pp. 291–322. American Psychological Association, Washington, DC.

Muthén BO (2004). "Latent Variable Analysis: Growth Mixture Modeling and Related Techniques for Longitudinal Data." In D Kaplan (ed.), *Handbook of Quantitative Methodology for the Social Sciences*, pp. 345–368. Sage, Newbury Park, CA.

Muthén BO, Asparouhov T (2009). "Growth Mixture Modeling: Analysis with Non-Gaussian Random Effects." In G Fitzmaurice, M Davidian, G Verbeke, G Molenberghs (eds.), *Longitudinal Data Analysis*, pp. 143–165. Chapman & Hall/CRC, Boca Raton, FL.

Muthén BO, Shedden K (1999). "Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm." *Biometrics*, **55**(2), 463–469. doi:10.1111/j.0006-341x.1999.00463.x.

Muthén LK, Muthén BO (1998–2012). **Mplus** *User's Guide*. 7th edition. Muthén & Muthén, Los Angeles, CA.

Nagengast B, Marsh HW, Scalas LF, Xu MK, Hau KT, 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**(8), 1058–1066. doi:10.1177/0956797611415540.

Novomestky F (2013). **gaussquad**: *Collection of Functions for Gaussian Quadrature*. R package version 1.0-2, URL https://CRAN.R-project.org/package=gaussquad.

Organisation for Economic Co-Operation and Development (2009). "PISA 2006: Science Competencies for Tomorrow's World." *Technical report*, OECD, Paris, France.

Organisation for Economic Co-Operation and Development (2010). "PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science." *Technical report*, OECD, Paris, France.

Pek J, Losardo D, Bauer DJ (2011). "Confidence Intervals for a Semiparametric Approach to Modeling Nonlinear Relations among Latent Variables." *Structural Equation Modeling: A Multidisciplinary Journal*, **18**(4), 537–553. doi:10.1080/10705511.2011.607072.

Pek J, Sterba SK, Kok BE, Bauer DJ (2009). "Estimating and Visualizing Nonlinear Relations among Latent Variables: A Semiparametric Approach." *Multivariate Behavioral Research*, **44**(4), 407–436. doi:10.1080/00273170903103290.

Ping RA (1995). "A Parsimonious Estimating Technique for Interaction and Quadratic Latent Variables." *Journal of Marketing Research*, **32**(3), 336–347. doi:10.2307/3151985.

Ping RA (1996). "Latent Variable Interaction and Quadratic Effect Estimation: A Two-Step Technique Using Structural Equation Analysis." *Psychological Bulletin*, **119**(1), 166–175. doi:10.1037/0033-2909.119.1.166.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2017). **nlme**: *Linear and Nonlinear Mixed Effects Models*. R package version 3.1-129, URL https://CRAN.R-project.org/package=nlme.

MATLAB (2010). *MATLAB, Version 7.10.0 (R2010a)*. Natick, MA. URL https://www.mathworks.com/.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org.

Rosseel Y (2012). "**lavaan**: An R Package for Structural Equation Modeling." *Journal of Statistical Software*, **48**(2), 1–36. doi:10.18637/jss.v048.i02.

SAS Institute Inc (2003). *SAS/STAT Software, Version 9.1*. Cary, NC. URL http://www.sas.com/.

Song XY, Lu ZH (2010). "Semiparametric Latent Variable Lodels with Bayesian P-Splines." *Journal of Computational and Graphical Statistics*, **19**(3), 590–608. doi:10.1198/jcgs.2010.09094.

Titterington DM, Smith AFM, Makov UE (1985). *Statistical Analysis of Finite Mixture Distributions.* John Wiley & Sons, Chichester, England.

Umbach N, Naumann K, Hoppe D, Brandt H (2017). **nlsem**: *Fitting Structural Equation Mixture Models.* R package version 0.7, URL https://CRAN.R-project.org/package=nlsem.

Wall MM, Amemiya Y (2000). "Estimation for Polynomial Structural Equation Models." *Journal of the Statistical American Association*, **95**(451), 929–940. doi:10.2307/2669475.

Wall MM, Amemiya Y (2001). "Generalized Appended Product Indicator Procedure for Nonlinear Structural Equation Analysis." *Journal of Educational and Behavioral Statistics*, **26**(1), 1–29. doi:10.3102/10769986026001001.

Wall MM, Amemiya Y (2003). "A Method of Moments Technique for Fitting Interaction Effects in Structural Equation Models." *British Journal of Mathematical and Statistical Psychology*, **56**(1), 47–63. doi:10.1348/000711003321645331.

Wen Z, Marsh HW, Hau KT (2010). "Structural Equation Models of Latent Interactions: An Appropriate Standardized Solution and Its Scale-Free Properties." *Structural Equation Modeling: A Multidisciplinary Journal*, **17**(1), 1–22. doi:10.1080/10705510903438872.

**Affiliation:**

Nora Umbach
Hector Research Institute of Education Sciences and Psychology
University of Tübingen
Europastr. 6
72072 Tübingen, Germany
E-mail: nora.umbach@uni-tuebingen.de
URL: http://homepages.uni-tuebingen.de/nora.umbach

Katharina Naumann
Department of Psychology
University of Tübingen
Schleichstr. 4
72074 Tübingen, Germany
E-mail: katharina.naumann@psycho.uni-tuebingen.de

Holger Brandt
Hector Research Institute of Education Sciences and Psychology
University of Tübingen
Europastr. 6
72072 Tübingen, Germany
E-mail: holger.brandt@uni-tuebingen.de

Augustin Kelava
Hector Research Institute of Education Sciences and Psychology
University of Tübingen
Europastr. 6
72072 Tübingen, Germany
E-mail: augustin.kelava@uni-tuebingen.de