Nonlinear structural equation modeling

Holger Brandt

University of Kansas

KU Summer Stats Camp

Structure

1. Introduction
2. Product indicator approaches
   - Constrained approach
   - Means and (co)variances of product variables
   - Unconstrained pi approach
3. Alternative approaches for nonlinear SEM
   - LMS
   - QML
   - 2SMM
   - Application of LMS, QML, and 2SMM
4. Summary
5. Some simulation results for different approaches

Introduction

Linear models: Limitations

- Linear models such as regression models or SEM often assume linear relationships between the variables.
- This assumption is often either wrong or a very crude approximation of the actual functional relationship.
- From an applied perspective, interaction effects are important because they indicate if the relationship between two variables is moderated by a third variable.
- In this workshop, we will discuss a variety of options to analyze deviations from linearity in the latent variable framework.

Why latent variables?

- Latent variable models, or more specifically here: Structural equation models (SEM) allow to consider a variety of aspects in data analysis:
  - They take measurement error into account by using multiple indicators and by specifying relationships between latent variables that are theoretically measurement error free
  - More complex relationships can be analyzed
  - Model fit can be investigated
- Particularly for nonlinear effects, the measurement error aspect is important because a typical effect size in social sciences is around 2.5%.

What is nonlinear?

- Nonlinearity can involve very different aspects:
  - Interaction effects between latent variables
  - Curvilinear relationships between latent variables
  - Nonlinear measurement models, for example for dichotomous or count data
  - Nonlinear parameters (such as λ^2)
- In this workshop, we focus on the first two aspects.

Interaction effects

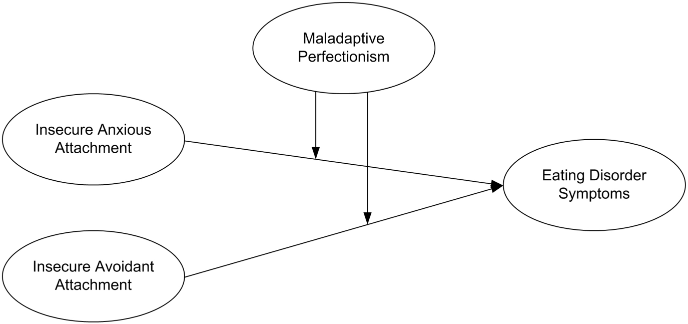


Figure: Taken from Dakanalis et al. (2014)

Quadratic and other polynomial effects

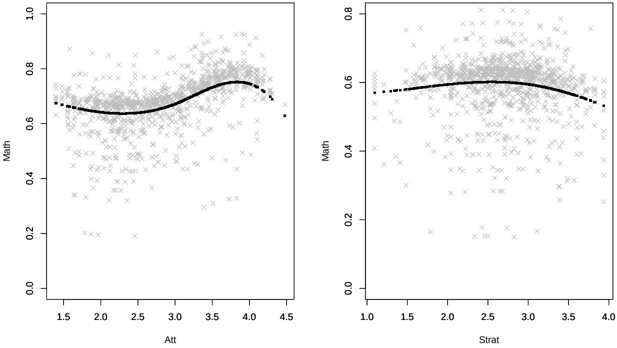


Figure: Taken from Kelava and Brandt (2015). Relationship between between pupils' math skills (Math) and their attitude toward reading (Att; left), and the reported teaching strategies (Strat; right).

Overview workshop

1. Interaction effects: Traditional product indicator approaches
2. Interaction and quadratic effects: Other approaches
3. Mixture models for curvilinear relationships
4. Mixture models for interaction and quadratic effects
5. Multilevel models with interaction effects

Overview sem-3

1. Interaction effects: Traditional product indicator (PI) approaches
   - Constrained approach
   - GAPI approach
   - Unconstrained approach
   - Main package: lavaan
2. Interaction and quadratic effects: Other approaches
   - Moment-based approaches
   - "Distribution-analytic" maximum likelihood approaches (LMS)
   - Standardization and illustration of results
   - Main package: nlsem and experimental syntax for 2SMM

Overview sem-4

1. Structural equation mixture models (SEMM)
    - Direct and indirect applications
    - Growth curve mixture models (direct application)
    - Optional: Alternative models for heterogeneous growth
    - Example for curvilinear relationship (indirect application)
2. Background: EM algorithm
3. Nonlinear structural equation mixture models (NSEMM)
    - Standardization of effects in NSEMM
4. Main software: Mplus, nlsem and plotsemm

Overview sem-5

1. Multilevel models with interaction effects
    - Bayesian modeling (priors, logic of Bayesian modeling)
    - Multilevel models with and without interaction effects
    - Multilevel SEM with and without interaction effects
    - Main packages: rstan and stan

Regression model

- In regression models, the inclusion of an interaction effect is typically conducted by including a product term of two (or more) predictor variables:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \epsilon_i \quad (1)$$

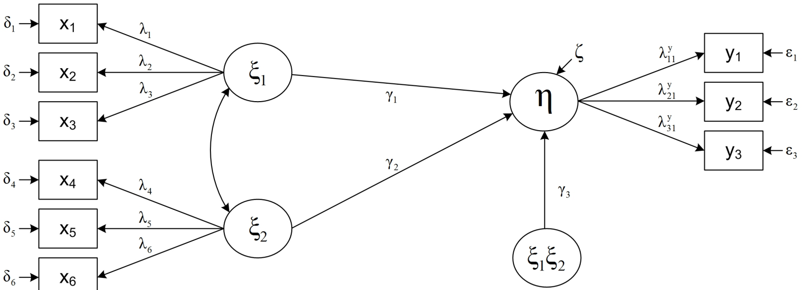
- A reformulation shows the moderation:

$$y_i = \beta_0 + \beta_1 x_{1i} + \underbrace{(\beta_2 + \beta_3 x_{1i})}_{\beta^*_{2i}} x_{2i} + \epsilon_i \quad (2)$$

  The relationship between x_2 and y depends on the value x_{1i} that may be different for each person i.

- The model is symmetric, i.e. the decision if x_1 or x_2 is the moderator is not a statistical but only applied question.

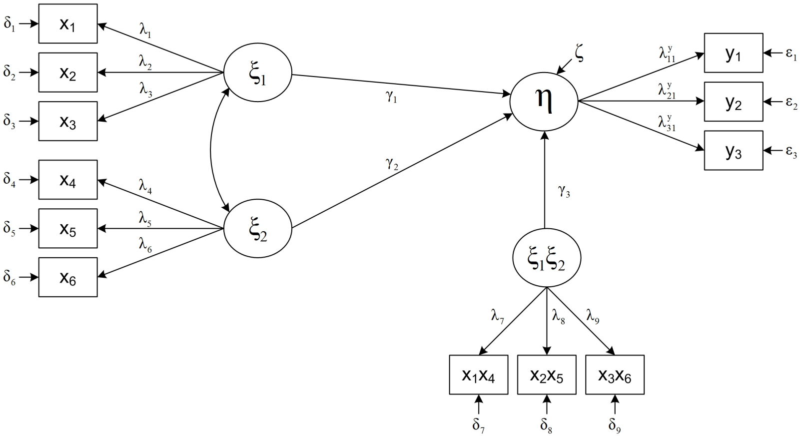
SEM with interaction effect



Problem of SEM: product terms

- The main question is: how can one form the product of ξ_1 and ξ_2?
- Intuitive solution (Kenny & Judd, 1984): build products of indicators for ξ_1 and ξ_2 and use them as "product indicators" for a latent interaction term $\xi_1 \cdot \xi_2$.
- Use standard SEM software to estimate the model (e.g., LISREL)

SEM with interaction effect and pi's



The constrained pi approach (Kenny & Judd, 1984)

- The first approach used the following logic: The parameters from the measurement model for the pi's are functions of the parameters of measurement model of the indicator. By constraining them accordingly, the model is still parsimonious.
- How does one derive these constraints?

$$x_1 \cdot x_4 = \lambda_7 \xi_1 \xi_2 + \delta_7 \tag{3}$$

- Example:

$$x_1 = \lambda_1 \xi_1 + \delta_1 \tag{4}$$
$$x_4 = \lambda_4 \xi_2 + \delta_4 \tag{5}$$

- And so

$$x_1 \cdot x_4 = (\lambda_1 \xi_1 + \delta_1) \cdot (\lambda_4 \xi_2 + \delta_4) \tag{6}$$

The constrained pi approach (Kenny & Judd, 1984)

- Based on this equation one can derive

$$x_1 \cdot x_4 = (\lambda_1\xi_1 + \delta_1) \cdot (\lambda_4\xi_2 + \delta_4) \quad (7)$$
$$= \underbrace{\lambda_1\lambda_4}_{=\lambda_7}\xi_1\xi_2 + \underbrace{\lambda_1\xi_1\delta_4 + \lambda_4\xi_2\delta_1 + \delta_1\delta_4}_{=\delta_7} \quad (8)$$

- Factor loading constraint:

$$\lambda_7 = \lambda_1\lambda_4 \quad (9)$$

The constrained pi approach (Kenny & Judd, 1984)

- Residual variance constraint

$$Var(\delta_7) = Var(\lambda_1\xi_1\delta_4 + \lambda_4\xi_2\delta_1 + \delta_1\delta_4) \quad (10)$$
$$= \lambda_1^2\phi_{11}\theta_{44} + \lambda_4^2\phi_{22}\theta_{11} + \theta_{11}\theta_{44} \quad (11)$$

- Constraints for latent variables

$$Var(\xi_1\xi_2) = \phi_{11}\phi_{22} + \phi_{12}^2 \quad (12)$$
$$E(\xi_1\xi_2) = \phi_{12} \quad (13)$$
$$Cov(\xi_1\xi_2, \xi_1) = Cov(\xi_1\xi_2, \xi_1) = 0 \quad (14)$$

- under the assumptions of
  - All latent variables have a mean of zero
  - All latent variables are normally distributed

The mean and variance of a product variable

- Mean of the product X and Y (Bohrnstedt & Goldberger, 1969)

$$E[XY] = E[X]E[Y] + C(X,Y) \tag{15}$$

which is only zero if X and Y are uncorrelated.

- Variance is

$$\begin{aligned} Var(XY) = {} & E^2[X]Var(Y) + E^2[Y]Var(X) + 2E[X]E[Y]C(X,Y) \\ & + Var(X)Var(Y) + C^2(X,Y) \end{aligned} \tag{16}$$

which simplifies for centered variables to

$$Var(XY) = Var(X)Var(Y) + C^2(X,Y) \tag{17}$$

The covariances of product variables

- The covariance between a variable Z and a product variable XZ depends on the multivariate skewness ν_{xyz} is

$$C(XY,Z) = E[X]C(Y,Z) + E[Y]C(X,Z) + \nu_{xyz} \tag{18}$$

which is only zero if X and Y have zero means *and* data is normally distributed.

- Similar results are obtained for the covariance between two product XY and UV which has a simplified formula for *centered* variables of

$$C(XY,UV) = -C(X,Y)C(U,V) + \nu_{xyuv} \tag{19}$$

with fourth central moment (kurtosis) ν_{xyuv} that is under normality

$$\nu_{xyuv} = C(X,Y)C(U,V) + C(X,U)C(Y,V) + C(X,V)C(Y,U) \tag{20}$$

What does that imply for the constrained approach?

- The formulated constraints are only valid for normally distributed and centered data.
- As soon as data is nonnormal, the model is *structurally* misspecified:
  - Variances of the latent variables
  - Covariances between the latent variables

  Both lead to biased estimates, also for the regression coefficients.
- But even under normality, the model leads to problems, because the normality assumption for the ML estimator is violated.

Approach under normality

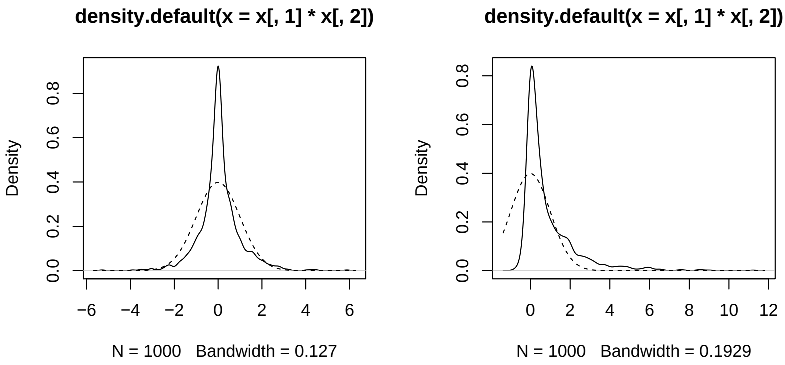


Figure: Distribution of product variables (left: uncorrelated, right: correlated). The normality assumption (dotted lines) for the ML estimator is clearly violated.

Demonstration 1

The constrained approach in lavaan. (for a simple model)

Summary constrained approach

- The approach uses product indicators that form a measurement model for the latent product terms that represent the interaction effects.
- The inclusion of pi's (e.g., $x_1 x_4$ lead to a violation of the normality assumption for the ML estimator which is typically used in SEM software.
- In situations with nonnormal data (e.g., x_1 is nonnormal) the approach is structurally misspecified.
- From a practical viewpoint: The formulation of constraints is extremely error prone and cumbersome.

GAPI and unconstrained pi approaches

- As a consequence to the potential problems of the constrained approach mainly two alternatives were suggested
- Wall and Amemiya (2001) proposed the Generalized Appended pi (gapi) approach that estimated the variances and covariances for the latent predictor variables freely (i.e. variances and covariances for ξ and their products)
- Marsh, Wen, and Hau (2004) proposed to estimate all parameters freely. This approach is widely known as unconstrained approach.

Unconstrained pi approach

- The approach shows quite robust results under normal and nonnormal data conditions for small models (e.g., one interaction effect).
- Its main advantages are
  - Easy to specify
  - Can be implemented in all standard software
- On the other hand problems remain:
  - The assumptions for the ML estimator are violated. Even a robust MLR estimator with sandwich type standard errors leads to increased type I error rates.
  - If data is nonnormal (e.g., x_1), covariances are misspecified because they are assumed to be zero, e.g. between $C(\xi_1, \xi_1\xi_2)$ or $C(\delta_1, \delta_7)$
  - If models are more complex, additional error covariances need to be estimated...
  - Selection of product terms

Unconstrained pi approach

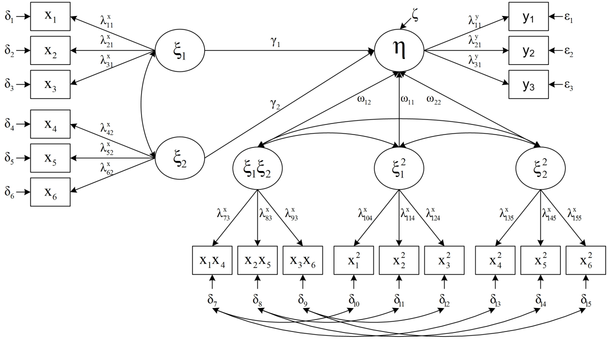


Figure: Unconstrained PI with two quadratic and one interaction effect. Additional error covariances need to be estimated (Kelava & Brandt, 2009).

Selection of indicators

- When forming product indicators, different possibilities exist:
  - Use non-redundant pairs indicators (only applicable for equal numbers of indicators per construct). This method is preferred.
  - Use all possible pairs. This leads to a very complex model and additional residual covariances need to be estimated freely (Kelava & Brandt, 2009)
- Even with non-redundant pairs, the model becomes fast very complex and instable when including more than two or three nonlinear effects.

Exercise II

1. Estimate the interaction model for the Kenny-Judd data using the upi approach.
2. Compare the findings of the cpi and the upi approach.
3. How do results change depending on the type of standard error estimation?

Exercise III

- The dataset was taken from PISA 2009 (OECD, 2010)
- Australian subsample of $N = 1,019$ students who took part in a reading test.
- Constructs:
  - *Predictor 1:* Students' attitude towards reading (Att)
  - *Predictor 2:* Reported online activities (Onl; i.e., read emails or chat online)
  - *Dependent variable:* Reading skills ($Read$)
- For each latent construct 3 item parcels were constructed that are saved in the file `pisa_online.dat`.

Exercise III

1. Investigate the data (distribution etc.)
2. Visualize the model in a path diagram. Include a model with an interaction and two quadratic effects.
3. Extend the `R` syntax used before and illustrate the results using the code available.

Latent moderated structures (LMS)

1. Klein and Moosbrugger (2000) developed a maximum likelihood estimator
2. This approach assumes normal data for the x variables (indicators of latent predictors) only.
3. No need for products of indicators.
4. The model takes the specific nonlinear structure of the means and covariances into account for estimation
5. Special software necessary
   - `Mplus` (Muthén & Muthén, 1998–2012)
   - `R` package `nlsem` (Umbach, Naumann, Brandt, & Kelava, 2017)

LMS: Background

- The main idea for LMS is to formulate a likelihood function for the observed data vector (x, y). For this, conditional means and covariance matrices need to be derived.
- In principle, this involves the following aspects
  - Cholesky decomposition of latent factors:

$$\xi = \mathbf{A}\mathbf{z}. \tag{21}$$

  - Decomposition of $\mathbf{z}$ in two components:

$$\mathbf{z} = \begin{bmatrix} \mathbf{z_1} \\ \mathbf{z_2} \end{bmatrix} = \begin{bmatrix} \mathbf{z_1} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{z_2} \end{bmatrix} \tag{22}$$

  only $\mathbf{z_1}$ is involved in nonlinear terms.

LMS: Background

- The conditional means and covariances are then derived (we skip the details here). The main point is that both are functions of z_1

$$\mu(\mathbf{z_1}) = \begin{bmatrix} \mu_x(\mathbf{z_1}) \\ \mu_y(\mathbf{z_1}) \end{bmatrix} \tag{23}$$

and

$$\Sigma(\mathbf{z_1}) = \begin{bmatrix} \Sigma_{\mathbf{xx}}(\mathbf{z_1}) & \Sigma_{\mathbf{xy}}(\mathbf{z_1}) \\ \Sigma'_{\mathbf{xy}}(\mathbf{z_1}) & \Sigma_{\mathbf{yy}}(\mathbf{z_1}) \end{bmatrix}. \tag{24}$$

- The observed variables are then conditional normal distributed:

$$[\mathbf{x}, \mathbf{y} | \mathbf{z_1} = z] \sim MVN(\mu(\mathbf{z_1}), \Sigma(\mathbf{z_1})) \tag{25}$$

LMS: Background

- These conditional means and covariances are then used for the density function

$$f(\mathbf{x} = x, \mathbf{y} = y) = \int_{\mathbf{R^k}} \varphi_{\mathbf{0},\mathbf{I}}(\mathbf{z_1}) \varphi_{\mu(\mathbf{z_1}),\Sigma(\mathbf{z_1})}(x, y)\, d\mathbf{z_1}. \qquad (26)$$

  where the $\varphi_{\mathbf{0},\mathbf{I}}(\mathbf{z_1})$ is a multivariate standard normal density and $\varphi_{\mu(\mathbf{z_1}),\Sigma(\mathbf{z_1})}(x, y)$ is a conditional multivariate distribution.
- This distribution can be viewed as continuous mixture distribution with mixing weights $\mathbf{z_1}$.

LMS: Background

- This mixture distribution cannot be solved analytically. However, it can be approximated with a finite mixture distribution using a Hermite-Gauss-quadrature

$$f(\mathbf{x} = x, \mathbf{y} = y) \approx \sum_{m=1}^{M} \pi_m \varphi_{\mu(2^{1/2}\nu_{\mathbf{m}}),\Sigma(2^{1/2}\nu_{\mathbf{m}})}(x, y). \quad (27)$$

  where $m = 1, \ldots, M$ are nodes and π_m are mixing weights.
- Intuitive idea: the standard normal $\mathbf{z_1}$ is replaced by a grouping variable where scores are grouped in intervals defined by the nodes.
- The estimation of finite mixtures will be part of tomorrow's workshop slides (EM algorithm).

QML: Background

- The main difference between QML (Klein & Muthén, 2007) and LMS is the approximation of the nonnormal density function. In QML no mixture is used but instead a product of a multivariate normal density function f_2 and a univariate conditionally normal density f_3^*

$$\begin{aligned} f(x_i, y_i) &= f_2(x_i, \mathbf{R}y_i) f_3(y_{1i}|x_i, \mathbf{R}y_i) \\ &\approx f_2(x_i, \mathbf{R}y_i) f_3^*(y_{1i}|x_i, \mathbf{R}y_i) =: f^*(x_i, y_i), \end{aligned} \tag{28}$$

  with conditional mean $E[y_{1i}|x_i, \mathbf{R}y_i]$ and variance $Var(y_{1i}|x_i, \mathbf{R}y_i)$.
- Parameters are estimated by maximizing the quasi-loglikelihood function for the density $f^*(x_i, y_i)$ using standard numerical methods (e.g., Newton-Raphson).
- The main advantages are: QML is faster, and theoretically more robust to nonnormality

2SMM: Background

- The 2 stage method of moments approach uses a completely different approach:
  1. Calculate factor scores using a cfa model
  2. Calculate the structural model using a corrected minimization criterion (very similar to OLS regression)

$$\hat{\gamma} = \hat{\boldsymbol{M}}^{-1}\hat{\boldsymbol{m}}, \tag{29}$$

  where γ include the regression coefficients.
    - $\hat{\boldsymbol{M}}$: corrected sum of squares and cross-products of the predictor variables (equal to $X'X$ in a regression framework)
    - $\hat{\boldsymbol{m}}$: corrected cross-products between the predictor and dependent variables (equal to $X'y$ in a regression framework).

It is a so called error-in-variables regression that takes the unreliability of the factor score estimation from the first stage into account for the second stage.

Empirical example 1

- The dataset was taken from PISA 2009 (OECD, 2010)
- Australian subsample of $N = 1,019$ students who took part in a reading test.
- Constructs:
  - *Predictor 1:* Students' attitude towards reading (Att)
  - *Predictor 2:* Reported online activities (Onl; i.e., read emails or chat online)
  - *Dependent variable:* Reading skills ($Read$)
- For each latent construct 3 item parcels were constructed that are saved in the file `pisa_online.dat`.

KU

Summary

- Original approach used product indicators
- More recent approaches (LMS, QML, 2SMM) use different operationalization of latent product terms and provide more efficient and/or more robust estimators
- There are more approaches available, however they are typically not available in standard software or have been shown to be inefficient
- All approaches are somewhat sensitive to nonnormal distributions and assume specific functional forms of nonnormality
- Tomorrow, we will address these issues using
    - Mixture models
    - Discuss options for model fit

Results from simulation studies (Brandt, Kelava, & Klein, 2014; Kelava, Nagengast, & Brandt, 2014)

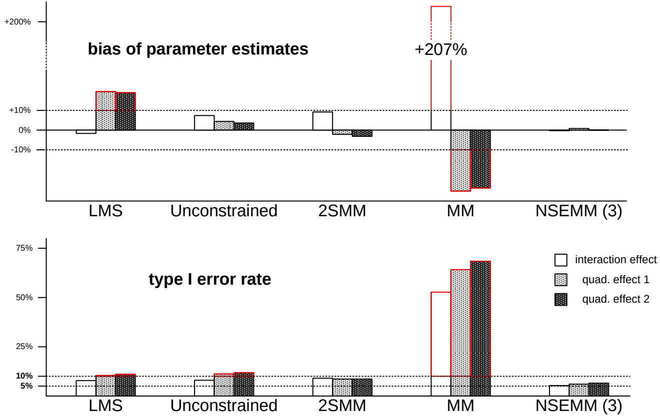


Figure: Bias of parameter estimation (top panel) and type I error rate (bottom panel) under the condition of nonnormally distributed data

Thank you for your attention.

Bohrnstedt, G. W., & Goldberger, A. S. (1969). On the exact covariance of products of random variables. *American Statistical Association Journal*, *64*, 1439–1442.

Brandt, H., Kelava, A., & Klein, A. G. (2014). A simulation study comparing recent approaches for the estimation of nonlinear effects in SEM under the condition of non-normality. *Structural Equation Modeling*, *21*, 181–195.

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., Nagengast, B., & Brandt, H. (2014). A nonlinear structural equation mixture modeling approach for non-normally distributed latent predictor variables. *Structural Equation Modeling*, *21*, 468–481.

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.

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.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7. ed.). Los Angeles, CA: Muthén & Muthén.

Umbach, N., Naumann, K., Brandt, H., & Kelava, A. (2017). Fitting nonlinear structural equation models in r with package nlsem. *Journal of Statistical Software*, *77*, 1–20. doi: doi:10.18637/jss.v077.i07

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