Global Model Fit Test for Nonlinear SEM

Rebecca Büchner, Andreas Klein, & Julien Irmer

Goethe-University Frankfurt am Main

Meeting of the SEM Working Group, 2018

Nonlinear SEM

Measurement Models:

$$\boldsymbol{x} = \Lambda_{\boldsymbol{x}}\xi + \delta$$
$$\boldsymbol{y} = \Lambda_{\boldsymbol{y}}\eta + \epsilon$$

Structural Model:

$$\eta = \alpha + \Gamma\xi + \xi'\Omega\xi + \zeta$$

$\rightarrow \eta$ and $\boldsymbol{y}$ are non-normally distributed.

$\boldsymbol{x}$ and $\boldsymbol{y}$: observed variables; Λ_x and Λ_y: factor loadings; ξ and η: latent variables, ξ multivariate normally distributed; δ, ϵ, and ζ: multivariate normally distributed error terms; Ω and Γ: coefficients

Model Fit Tests for Nonlinear SEM

- χ^2 difference tests (Gerhard et al., 2015)
- Information criteria (AIC, BIC, ...)
- Fit measures to detect omitted nonlinear terms (Klein & Schermelleh-Engel, 2010, Gerhard, Büchner, Klein & Schermelleh-Engel, 2017)
- Inferential tests:
  - The χ^2 test is inappropriate for nonlinear SEM (cf. Mooijaart & Satorra, 2009)
  - For nonlinear SEM no other inferential test has yet been developed

Aim

Development of a new inferential test for nonlinear SEM similar to the χ^2 test.

Procedure

1. Estimation Using Quasi-ML
2. Saturated Model
3. A Quasi-Likelihood Ratio Test
4. Simulation Study

Quasi-Maximum Likelihood

- Estimation method very similar to ML
- Difference: distributional assumptions are not fully met
- Correct standard errors and the distribution of likelihood ratio test statistics can be calculated

Klein and Muthén (2007) applied quasi-ML for the estimation of nonlinear SEM (QML).

Simplified QML (sQML)

$$
\begin{aligned}
\mathrm{f}(\boldsymbol{x}, \boldsymbol{y}) &= \mathrm{f}_1(\boldsymbol{x})\mathrm{f}_2(\boldsymbol{y}|\boldsymbol{x}) \\
&\approx f_1(\boldsymbol{x}) f_2^*(\boldsymbol{y}|\boldsymbol{x})
\end{aligned}
$$

Idea

$\mathrm{f}_2(\boldsymbol{y}|\boldsymbol{x})$ is approximated by a multivariate normal distribution $\mathrm{f}_2^*(\boldsymbol{y}|\boldsymbol{x})$

sQML - Log-likelihood Function

$f_2(\boldsymbol{y}|\boldsymbol{x})$:

- $\mu_T(\boldsymbol{x})$ is a polynomial of degree two in $\boldsymbol{x}$
- Model implied covariance matrix $\Sigma_{\boldsymbol{y}|\boldsymbol{x}}$
- Unconstrained covariance matrix $\Sigma^{\boldsymbol{T}}_{\boldsymbol{y}|\boldsymbol{x}}$

$$
\begin{aligned}
LL^T_\vartheta(\boldsymbol{x},\boldsymbol{y}) =& \frac{1}{N}\sum_{i=1}^{N}\left(\ln f_1(\boldsymbol{x_i}) + \ln f_2^*(\boldsymbol{y_i}|\boldsymbol{x_i}=\boldsymbol{x})\right) \\
=& c - \frac{1}{2}\left(\ln|\Sigma_{\boldsymbol{x}}| + \mathrm{tr}\left(\boldsymbol{S_x}\Sigma_{\boldsymbol{x}}^{-1}\right) + \ln|\Sigma_{\boldsymbol{y}|\boldsymbol{x}}|\right) + \\
& \frac{1}{2}\mathrm{tr}\left(\Sigma^{\boldsymbol{T}}_{\boldsymbol{y}|\boldsymbol{x}}\Sigma^{-1}_{\boldsymbol{y}|\boldsymbol{x}}\right)
\end{aligned}
$$

T: target model; ϑ: vector of parameters in the target model; c: a constant; $\boldsymbol{S_x}$: observed covariance matrix; $f_1(\boldsymbol{x})$ is the density function of a multivariate normal distribution

A Saturated Model

- $f_1(\boldsymbol{x})$: Observed covariance matrix $\boldsymbol{S_x}$ of $\boldsymbol{x}$
- $\mu_S(\boldsymbol{x})$
- Unconstrained covariance matrix $\Sigma^{\boldsymbol{S}}_{\boldsymbol{y}|\boldsymbol{x}}$

$$LL^S_\theta(\boldsymbol{x}, \boldsymbol{y}) = c - \frac{1}{2}\left(\ln|\boldsymbol{S_x}| + \ln|\Sigma^{\boldsymbol{S}}_{\boldsymbol{y}|\boldsymbol{x}}| + p + q\right)$$

S: saturated model; p and q: number of parameters in the saturated and in the target model, respectively;
θ: vector of parameters in the saturated model; c: a constant

A Quasi-Likelihood Ratio Test (Q-LRT)

Test statistic

$$\Lambda(\boldsymbol{x}, \boldsymbol{y}) := -2N\left(LL_{\vartheta}^{T}(\boldsymbol{x}, \boldsymbol{y}) - LL_{\theta}^{S}(\boldsymbol{x}, \boldsymbol{y})\right)$$

Distribution

It is possible to determine the distribution and critical values of $\Lambda(\boldsymbol{x}, \boldsymbol{y})$

Simulation Study - Example

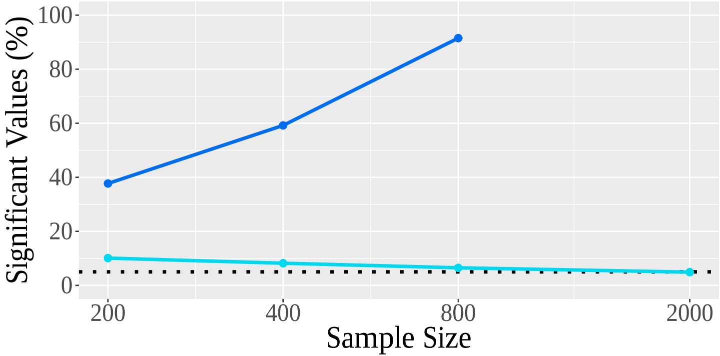
Population model:

$$\eta = -0.08 + 0.5\,\xi_1 + 0.4\,\xi_2 + 0.2\,\xi_1\xi_2 + \zeta$$

Analysis model:

Power: $\eta = \alpha + \gamma_1\,\xi_1 + \gamma_2\,\xi_2 + \zeta$

Type I error: $\eta = \alpha + \gamma_1\,\xi_1 + \gamma_2\,\xi_2 + \omega_{12}\,\xi_1\xi_2 + \zeta$



Simulation Study - General

- High power rates for various conditions, when sample size is sufficiently large
- Even for $N = 800$ Type I error rates are slightly elevated (between 5% and 7.7%)

Conclusion and Outlook

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

- Q-LRT (quasi-likelihood ratio test) is a suitable inferential test for nonlinear models, when sample size is sufficiently large
- Q-LRT is only appropriate for nonlinear SEM estimated with sQML
- Advantages and disadvantages of the χ^2-Test
- Robustness of Q-LRT and sQML: simulation study

Many thanks for your attention!

References

- Gerhard, C., Büchner, R. D., Klein, A. G., & Schermelleh-Engel, K. (2017). A fit index to assess model fit and detect omitted terms in nonlinear SEM. *Structural Equation Modeling, 24*, 414-427.
- Gerhard, C., Klein, A. G., Schermelleh-Engel, K., Moosbrugger, H., Gäde, J., & Brandt, H. (2015). On the performance of likelihood-based difference tests in nonlinear structural equation models. *Structural Equation Modeling, 22,* 276-287.
- Mooijaart, A. & Satorra, A. (2009). On insensitivity of the chi-square model test to nonlinear misspecification in structural equation models. *Psychometrika, 74,* 443-455.
- 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,* 647-673.