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:

\[ \mathbf{x} = \Lambda_x \xi + \delta \]
\[ \mathbf{y} = \Lambda_y \eta + \epsilon \]

Structural Model:

\[ \eta = \alpha + \Gamma \xi + \xi' \Omega \xi + \zeta \]

→ \( \eta \) and \( \mathbf{y} \) are non-normally distributed.

\( \mathbf{x} \) and \( \mathbf{y} \): observed variables; \( \Lambda_x \) and \( \Lambda_y \): factor loadings; \( \xi \) and \( \eta \): latent variables, \( \xi \) multivariate normally distributed; \( \delta, \epsilon, \) and \( \zeta \): multivariate normally distributed error terms; \( \Omega \) and \( \Gamma \): coefficients
Model Fit Tests for Nonlinear SEM

- $\chi^2$ 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 $\chi^2$ 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 $\chi^2$ 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

Simplified QML (sQML)

\[ f(x, y) = f_1(x)f_2(y|x) \]
\[ \approx f_1(x)f_2^*(y|x) \]

Idea

\[ f_2(y|x) \] is approximated by a multivariate normal distribution
\[ f_2^*(y|x) \]
sQML - Log-likelihood Function

\( f_2(y|x) \):
- \( \mu_T(x) \) is a polynomial of degree two in \( x \)
- Model implied covariance matrix \( \Sigma_{y|x} \)
- Unconstrained covariance matrix \( \Sigma_T^{y|x} \)

\[
LL^T_{\vartheta}(x, y) = \frac{1}{N} \sum_{i=1}^{N} \left( \ln f_1(x_i) + \ln f_2^*(y_i|x_i = x) \right)
\]

\[
= c - \frac{1}{2} \left( \ln |\Sigma_x| + \text{tr} \left( S_x \Sigma_x^{-1} \right) + \ln |\Sigma_{y|x}| \right) + \\
\frac{1}{2} \text{tr} \left( \Sigma_T^{y|x} \Sigma_{y|x}^{-1} \right)
\]

\( T \): target model; \( \vartheta \): vector of parameters in the target model; \( c \): a constant; \( S_x \): observed covariance matrix; \( f_1(x) \) is the density function of a multivariate normal distribution
A Saturated Model

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

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

S: saturated model; $p$ and $q$: number of parameters in the saturated and in the target model, respectively;
$\theta$: vector of parameters in the saturated model; $c$: a constant
A Quasi-Likelihood Ratio Test (Q-LRT)

Test statistic

\[ \Lambda(x, y) := -2N \left( LL^T_\theta (x, y) - LL^S_\theta (x, y) \right) \]

Distribution

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

- 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 $\chi^2$-Test.
- Robustness of Q-LRT and sQML: simulation study.
Many thanks for your attention!

References


