R function `swain`

Correcting structural equation model fit statistics and indexes under small-sample and/or large-model conditions

Version 1.2 – August 20, 2013

Anne Boomsma & Walter Herzog

University of Groningen, The Netherlands
WHU, Germany

1. Purpose of the function

The R function `swain` can be used as a practical tool for the evaluation of structural equation models with small sample sizes and/or large model sizes. More specifically, it can be employed for correcting model fit statistics from the output of standard structural equation modeling (SEM) software programs like AMOS, EQS, `lavaan` (Rosseel's R package), LISREL, M*plus*, MX, and SEPATH, for example. The function provides small-sample and large-model robust estimates of the following model fit statistics and indexes, as well as confidence interval estimates for some corresponding population parameters:

- regular maximum likelihood chi-square model fit statistic (T_{ML})

- root mean square error of approximation (RMSEA)

- confidence intervals with arbitrary confidence levels for the RMSEA

- test of close fit: $H_0 : \text{RMSEA} < 0.05$

- Steiger's Gamma Index (Γ_1)

- confidence intervals with arbitrary confidence levels for Steiger's Γ_1

- Tucker-Lewis Index (TLI)

- Comparative Fit Index (CFI)

The `swain` function calculates corrections of these statistics based on a correction derived by Swain (1975); see Herzog, Boomsma, & Reinecke (2007). Under the assumption of multivariate normality of the observed variables, Monte Carlo studies indicate that the correction leads to bias reduction of the model fit estimators listed above, as long as the ratio of sample size N to the number of estimated parameters t, the so-called *N:t* ratio, is at least *2:1* (Herzog, Boomsma, & Reinecke, 2007). This performance is independent of the degree of model misspecification (Herzog & Boomsma, 2006, 2009). In case of clearly nonnormal data, it might be better to use the Satorra-Bentler corrected 'scaled' model fit statistic as an argument of this function (see, for example, EQS or M*plus* program output), instead of the regular maximum likelihood chi-square statistic.

2. Installation

The user may take the following steps to have the R program installed and the `swain` function loaded.

- Download the open source R software from the uniform resource locator `http://cran.r-project.org/`. The freely available R program is a powerful environment for statistical computation, statistical data analysis, and graphics.

- From URL `http://www.gmw.rug.nl/~boomsma`, copy the R script `swain.R` into a selected working directory of R.

- Start R, after which the R console window pops up.

- In the status bar at the top, click *File → Open script →* select `swain.R`, after which the R editor window pops up.

- After a right-mouse click in this editor window, select the complete script (*Select all*).

- After another right-mouse click, select *Run line or selection*. In principle, one can also use keyboard function `F5`.

- The `swain` function has been loaded now. It is thus available to apply. At the termination of an R session, the system will ask whether to save the workspace image. Pressing *Yes* will save the `swain` function with it. When starting a new R session, the previously saved workspace image can be reloaded by entering command `load(".RData")`.

3. How to use the function

Consider a vector of p random variables $\mathbf{z}$ $(p \times 1)$ and a corresponding empirical sample covariance matrix $\mathbf{S}$ $(p \times p)$ based on N independent observations. Given the sample covariance matrix $\mathbf{S}$, the structural equation model M_j, i.e., the model under study, was estimated using maximum likelihood estimation procedures. The primary objective of the `swain` function is to obtain proper estimates of a number of model fit statistics and indexes. Both regular and corrected estimates are calculated, given the values of a number of arguments of the function.

In principle, the `swain` function can be applied to any structural equation model of interest. The function is fairly easy to use, as the following example may serve to illustrate.

- Suppose the sample data and the model under study provide the following specifications:
    - sample size, `N = 50`.
    - number of observed variables, `p = 25`.
    - number of degrees of freedom, `d = 265`.

- Further suppose that the following estimated chi-square model fit statistics are printed in the output of some structural equation modeling software program:
    - the maximum likelihood chi-square statistic for the model under study, `TML = TMLj = 400`.
    - the maximum likelihood chi-square estimate for the independence model, `TMLi = 1100`. This estimate is used for the calculation of the corrected incremental fit indexes TLI and CFI. It should be noted that, unlike `TML`, `TMLi` itself is not corrected (for details, see Herzog & Boomsma, 2006, 2009).

- For the estimation of confidence intervals for RMSEA and Γ_1, the researcher wants the confidence level, $1 - \alpha$, to be 0.90, which is the `default` level:
    - `conf.level = 0.90`.

- The preceding information provides the necessary values of the arguments of the `swain` function: `N, p, d, TML, TMLi, conf.level`. In the R console window, the user can now type the following command and press the *return* button: `swain(N=50, p=25, d=265, TML=400, TMLi=1100)`. One could also use the short command: `swain(50, 25, 265, 400, 1100)`.

- As a result, the following output is generated:

```
Sample size, N = 50
Number of observed variables, p = 25
Number of degrees of freedom, d = 265
Chi-square test statistic, TML = 400, p-value: 1.5592e-07
Swain's correction factor: 0.80433
Swain-corrected chi-square test statistic: 321.73, p-value: 0.0097438
Chi-square test statistic of the independence model, TMLi = 1100

Estimate of the Root Mean Square Error of Approximation (RMSEA): 0.10196
90% confidence interval for RMSEA: (0.080977, 0.12186)
Test of close fit: H_0: RMSEA < .05, p-value: 0.00013680
Swain-corrected RMSEA: 0.066099
Swain-corrected 90% confidence interval for RMSEA: (0.034600, 0.090403)
Swain-corrected test of close fit: H_0: RMSEA < .05, p-value: 0.17136

Estimate of Gamma_1: 0.8194
90% confidence interval for Gamma_1: (0.76056, 0.87795)
Swain-corrected Gamma_1: 0.91523
Swain-corrected 90% confidence interval for Gamma_1: (0.85232, 0.97525)

Tucker-Lewis Index (TLI): 0.80896
Swain-corrected Tucker-Lewis Index: 0.91972
Comparative Fit Index (CFI): 0.83125
Swain-corrected Comparative Fit Index: 0.92908
```

- A comparison of the uncorrected fit estimates and the Swain-corrected estimates demonstrates the effect of Swain's corrective procedure. In the preceding example, the corrected estimates indicate a much better model fit than the uncorrected ones.

- The confidence level for the confidence intervals of RMSEA and Γ_1 can be changed from its default 90% level to another value by using function argument `conf.level`. The next command changes the significance level to 80%: `swain(N=50,p=25,d=265,TML=400,TMLi=1100,conf.level=0.80)`.

4. In case of problems

In case of any problems one is requested to write an e-mail to `a.boomsma@rug.nl` and/or `walter.herzog@whu.edu`. Users are also invited to contact the authors if possible extensions or improvements of the function should be considered.

5. License

The R **function** `swain` is free software: one can redistribute it and/or modify it under the terms of the GNU General Public License as published by the *Free Software Foundation*, either version 3 of the License, or (at one's option) any later version; see `http://www.gnu.org/licenses/`.

References

Swain, A.J. (1975). *Analysis of parametric structures for variance matrices*. Unpublished doctoral dissertation, Department of Statistics, University of Adelaide, Australia.

Herzog, W., & Boomsma, A. (2006, July). *Finite sample corrections for RMSEA estimation*. Paper presented at the annual meeting of the Psychometric Society, Montreal, Canada.

Herzog, W., & Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling*, **16**, 1–27.

Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model-size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling*, **14**, 361–390.

Summary R function swain

Correcting structural equation model fit statistics and indexes under small-sample and/or large-model conditions

Anne Boomsma & Walter Herzog

Description

Calculates Swain-corrected model fit statistics and indexes for the evaluation of structural equation models.

Usage

```
swain(N, p, d, TML, TMLi, conf.level=0.90)
```

Arguments

`N`	sample size
`p`	number of observed variables
`d`	number of degrees of freedom for model M_j
`TML`	chi-square likelihood-ratio test statistic for model M_j, i.e., T_{ML_j}
`TMLi`	chi-square test statistic for the independence model M_i, i.e., T_{ML_i}
`conf.level`	confidence level for interval estimation (default 0.90)

Details

In principle the corrections can be applied for any structural equation model under maximum likelihood estimation.

Examples

```
swain(N=50, p=25, d=265, TML=400, TMLi=1100, conf.level=0.90)
swain(50, 25, 265, 400, 1100)
swain(50, 25, 265, 400, 1100, 0.80)
swain(7929, 40, 730, 23262.1, 76214.9, 0.90)
```