Testing the RMSEA in multiple-group analysis

Mike Cheung

August 4, 2016

Generate some data

- The data are identical in these two groups. Therefore, the RMSEA should be identical in analyzing one group or two groups (Steiger, 1998).
- Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*(4), 411-419. http://doi.org/10.1080/10705519809540115

```
library(lavaan)
library(OpenMx)

## One-factor model
Lambda <- matrix(c(0.4, 0.5, 0.6, 0.4, 0.5, 0.6), ncol=1)
## Introduce correlated errors so that RMSEA will not be zero
Psi <- matrix(0.3, ncol=6, nrow=6)
diag(Psi) <- c(0.4, 0.5, 0.4, 0.5, 0.4, 0.5)
Sigma <- Lambda %*% t(Lambda) + Psi

set.seed(1031334)
## Single group data
my.data1 <- mvrnorm(n=500, mu=rep(0, 6), Sigma=Sigma)
my.data1 <- as.data.frame(my.data1)

## Multiple group data with identical data
my.data2 <- rbind(my.data1, my.data1)
my.data2$group <- rep(c("A", "B"), each=500)
```

OpenMx

Single-group analysis

- Reported RMSEA = 0.05596148 (correct)

```
observed <- paste0("V", 1:6)
latent <- "f"

model1 <- mxModel("Group1", type="RAM",
                manifestVars=observed, latentVars=latent,
                mxPath(from=latent, to=observed,
                       free=c(F,T,T,T,T,T),
                       values=c(1,0,0,0,0,0),
                       labels=c(NA,"l2_1","l3_1","l4_1","l5_1","l6_1")),
                mxPath(from=observed, arrows=2,
                       labels=c("e1_1","e2_1","e3_1","e4_1","e5_1","e6_1")),
                mxPath(from=latent, arrows=2,labels=c("phi_1")),
                mxPath(from="one", to=c(observed, latent), arrows=1,
                       free=c(T,T,T,T,T,T,F),
                       values=c(0,0,0,0,0,0,0),
                       labels=c("m1_1","m2_1","m3_1","m4_1","m5_1","m6_1",NA)),
                mxData(my.data1, type="raw"))

mx.fit1 <- mxRun(model1)
```

```
## Running Group1 with 18 parameters
```

```
summary(mx.fit1, refModels=mxRefModels(mx.fit1, run=TRUE))
```

```
## Running Saturated Group1 with 27 parameters
```

```
## Running Independence Group1 with 12 parameters
```

```
## Summary of Group1
##
## free parameters:
##     name matrix row col    Estimate   Std.Error A
## 1   l2_1      A  V2   f 1.054691054 0.039745888
## 2   l3_1      A  V3   f 1.198075653 0.035652934
## 3   l4_1      A  V4   f 1.023531549 0.038343353
## 4   l5_1      A  V5   f 1.123992573 0.032394776
## 5   l6_1      A  V6   f 1.213038645 0.039553244
## 6   e1_1      S  V1  V1 0.098207090 0.007720274
## 7   e2_1      S  V2  V2 0.219820269 0.015342497
## 8   e3_1      S  V3  V3 0.117016558 0.009706923
## 9   e4_1      S  V4  V4 0.205345021 0.014357146
## 10  e5_1      S  V5  V5 0.092581606 0.007902942
## 11  e6_1      S  V6  V6 0.176917434 0.013307448
## 12 phi_1      S   f   f 0.440054060 0.033782857
## 13  m1_1      M   1  V1 0.005578164 0.032806788
## 14  m2_1      M   1  V2 0.011840756 0.037661062
## 15  m3_1      M   1  V3 0.006965001 0.038691052
## 16  m4_1      M   1  V4 0.040506856 0.036503251
## 17  m5_1      M   1  V5 0.006655595 0.036010555
## 18  m6_1      M   1  V6 0.007131772 0.040602307
##
## observed statistics:  3000
## estimated parameters:  18
## degrees of freedom:  2982
## fit value ( -2lnL units ):  4285.451
## saturated fit value ( -2lnL units ):  4262.359
## number of observations:  500
## chi-square:  X2 ( df=9 ) = 23.09259,  p = 0.005991683
## Information Criteria:
##       |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
## AIC:      -1678.549               4321.451                       NA
## BIC:     -14246.510               4397.314                 4340.181
## CFI: 0.9954453
## TLI: 0.9924089   (also known as NNFI)
## RMSEA:  0.05596148  [95% CI (0.02140528, 0.08971861)]
## Prob(RMSEA <= 0.05): 0.3240696
## timestamp: 2016-08-04 11:01:19
## Wall clock time (HH:MM:SS.hh): 00:00:00.60
## optimizer:  SLSQP
## OpenMx version number: 2.6.9
## Need help?  See help(mxSummary)
```

Multiple-group analysis with two identifical sets of data

- Reported RMSEA = 0.03957074 (incorrect)
- The correct RMSEA should be $\sqrt{K} * RMSEA_{reported} = \sqrt{2} * 0.03957074 = 0.05596148$, where K is the no. of groups.

```
model2 <- mxModel("Group2", type="RAM",
                manifestVars=observed, latentVars=latent,
                mxPath(from=latent, to=observed,
                       free=c(F,T,T,T,T,T),
                       values=c(1,0,0,0,0,0),
                       labels=c(NA,"l2_2","l3_2","l4_2","l5_2","l6_2")),
                mxPath(from=observed, arrows=2,
                       labels=c("e1_2","e2_2","e3_2","e4_2","e5_2","e6_2")),
                mxPath(from=latent, arrows=2,labels=c("phi_2")),
                mxPath(from="one", to=c(observed, latent), arrows=1,
                       free=c(T,T,T,T,T,T,F),
                       values=c(0,0,0,0,0,0,0),
                       labels=c("m1_2","m2_2","m3_2","m4_2","m5_2","m6_2",NA)),
                mxData(my.data1, type="raw"))

model.combined <- mxModel(model="Multiple group", model1, model2,
                          mxFitFunctionMultigroup(c("Group1", "Group2")))
mx.fit2 <- mxRun(model.combined)
```

```
## Running Multiple group with 36 parameters
```

```
summary(mx.fit2, refModels=mxRefModels(mx.fit2, run=TRUE))
```

```
## Running Saturated Multiple group with 54 parameters
```

```
## Running Independence Multiple group with 24 parameters
```

```
## Summary of Multiple group
##
## free parameters:
##     name   matrix row col    Estimate   Std.Error A
## 1   l2_1 Group1.A  V2   f 1.054690201 0.039746358
## 2   l3_1 Group1.A  V3   f 1.198074958 0.035652682
## 3   l4_1 Group1.A  V4   f 1.023530625 0.038343438
## 4   l5_1 Group1.A  V5   f 1.123991509 0.032394335
## 5   l6_1 Group1.A  V6   f 1.213038014 0.039552655
## 6   e1_1 Group1.S  V1  V1 0.098207044 0.007720272
## 7   e2_1 Group1.S  V2  V2 0.219820546 0.015342689
## 8   e3_1 Group1.S  V3  V3 0.117016681 0.009706975
## 9   e4_1 Group1.S  V4  V4 0.205344943 0.014357246
## 10  e5_1 Group1.S  V5  V5 0.092581492 0.007902946
## 11  e6_1 Group1.S  V6  V6 0.176917839 0.013307503
## 12 phi_1 Group1.S   f   f 0.440054869 0.033784349
## 13  m1_1 Group1.M   1  V1 0.005578453 0.032811579
## 14  m2_1 Group1.M   1  V2 0.011840814 0.037666048
## 15  m3_1 Group1.M   1  V3 0.006965169 0.038696885
## 16  m4_1 Group1.M   1  V4 0.040507144 0.036507281
## 17  m5_1 Group1.M   1  V5 0.006655972 0.036015737
## 18  m6_1 Group1.M   1  V6 0.007132195 0.040607507
## 19  l2_2 Group2.A  V2   f 1.054690685 0.039747412
## 20  l3_2 Group2.A  V3   f 1.198075784 0.035654888
## 21  l4_2 Group2.A  V4   f 1.023531463 0.038344662
## 22  l5_2 Group2.A  V5   f 1.123992490 0.032396456
## 23  l6_2 Group2.A  V6   f 1.213039101 0.039554757
## 24  e1_2 Group2.S  V1  V1 0.098207079 0.007720292
## 25  e2_2 Group2.S  V2  V2 0.219820589 0.015342605
## 26  e3_2 Group2.S  V3  V3 0.117016585 0.009706932
## 27  e4_2 Group2.S  V4  V4 0.205345024 0.014357249
## 28  e5_2 Group2.S  V5  V5 0.092581481 0.007902934
## 29  e6_2 Group2.S  V6  V6 0.176917457 0.013307412
## 30 phi_2 Group2.S   f   f 0.440053987 0.033784496
## 31  m1_2 Group2.M   1  V1 0.005577642 0.032811665
## 32  m2_2 Group2.M   1  V2 0.011840120 0.037666307
## 33  m3_2 Group2.M   1  V3 0.006964421 0.038697090
## 34  m4_2 Group2.M   1  V4 0.040506146 0.036507516
## 35  m5_2 Group2.M   1  V5 0.006655238 0.036016253
## 36  m6_2 Group2.M   1  V6 0.007131277 0.040607829
##
## observed statistics:  6000
## estimated parameters:  36
## degrees of freedom:  5964
## fit value ( -2lnL units ):  8570.903
## saturated fit value ( -2lnL units ):  8524.717
## number of observations:  1000
## chi-square:  X2 ( df=18 ) = 46.18519,  p = 0.0002787644
## Information Criteria:
##       |  df Penalty  |  Parameters Penalty  |  Sample-Size Adjusted
## AIC:      -3357.097               8642.903                       NA
## BIC:     -32626.950               8819.582                 8705.244
## CFI: 0.9954453
## TLI: 0.9924089   (also known as NNFI)
## RMSEA:  0.03957074  [95% CI (0.02263197, 0.05640769)]
## Prob(RMSEA <= 0.05): 0.8797571
## timestamp: 2016-08-04 11:01:24
## Wall clock time (HH:MM:SS.hh): 00:00:03.83
## optimizer:  SLSQP
## OpenMx version number: 2.6.9
## Need help?  See help(mxSummary)
```

```r
sessionInfo()
```

```
## R version 3.3.1 (2016-06-21)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 10586)
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] parallel  stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
## [1] OpenMx_2.6.9  Rcpp_0.12.6   Matrix_1.2-6  MASS_7.3-45   digest_0.6.10
## [6] lavaan_0.5-20
##
## loaded via a namespace (and not attached):
##  [1] lattice_0.20-33 quadprog_1.5-5  grid_3.3.1      stats4_3.3.1
##  [5] formatR_1.4     magrittr_1.5    evaluate_0.9    stringi_1.1.1
##  [9] pbivnorm_0.6.0  rmarkdown_1.0   tools_3.3.1     stringr_1.0.0
## [13] yaml_2.1.13     mnormt_1.5-4    htmltools_0.3.5 knitr_1.13
```