# Online Supplement 1: scaling example

We demonstrate the latent variable scaling problem discusesd in the article by estimating a set of three CFA models on ten different samples. In the first sample, we drew 1,000 observations from a population with two perfectly correlated factors with three items each. The factor variance and errors were both 1, producing items with population variance of 2 (sd=1.41). Then we generated four rescaled versions of the data where the item reliabilities were the same, but their variances were different. Then we used these data to estimate three different CFA models, an unconstrained model where the factor covariance was freely estimated and two constrained models where either the covariance or the correlation between the factors was constrained to 1. In this model, there was a complete lack of discriminant validity, which was always detected by the model where the constraint was specified as a correlation, but was detected in only one sample when a model with the covariance constraint was estimated as shown in Table S1.

Table S1 Demonstration of the test with a two-factor CFA

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Indicator SD | Model 1: Unconstrained | Model 2: Covariance constrained at 1 | | | Model 3: Correlation constrained at 1 | | |
|  |  |  | p |  |  | P |
| Population correlation: 1 | | | | | | | |
| 0.71 | 13.24 | 415.7 | 402.46 | 0.00 | 13.4 | 0.17 | 0.68 |
| 1.06 | 13.24 | 74.64 | 61.4 | 0.00 | 13.4 | 0.17 | 0.68 |
| 1.41 | 13.24 | 14.17 | 0.93 | 0.33 | 13.4 | 0.17 | 0.68 |
| 1.77 | 13.24 | 65.98 | 52.75 | 0.00 | 13.4 | 0.17 | 0.68 |
| 2.12 | 13.24 | 151.48 | 138.25 | 0.00 | 13.4 | 0.17 | 0.68 |
| Population correlation: 0.5 | | | | | | | |
| 0.71 | 2.13 | 581.68 | 579.56 | 0.00 | 392.39 | 390.27 | 0.00 |
| 1.06 | 2.13 | 185.33 | 183.2 | 0.00 | 392.39 | 390.27 | 0.00 |
| 1.41 | 2.13 | 41.86 | 39.73 | 0.00 | 392.39 | 390.27 | 0.00 |
| 1.77 | 2.13 | 3.61 | 1.48 | 0.22 | 392.39 | 390.27 | 0.00 |
| 2.12 | 2.13 | 8.05 | 5.92 | 0.01 | 392.39 | 390.27 | 0.00 |

The second set of samples were generated similarly, except this time the two factors were clearly discriminant valid correlating only at 0.5. Table S1 shows that the correlation constrained model always rejected the null hypothesis of complete lack of discriminant validity, but using a covariance constrained model in one sample a researcher would incorrectly conclude that there was a severe discriminant validity problem.

## R code for the example

library(lavaan)

set.seed(1)  
  
N <- 1000  
F1 <- rnorm(N)  
F2 <- F1 \* .5 + rnorm(N, sd = sqrt(.75))  
  
b1 <- matrix(F1 + rnorm(3\*1000), ncol = 3, dimnames = list(1:1000,c("x1","x2","x3")))  
b2 <- matrix(F1 + rnorm(3\*1000), ncol = 3, dimnames = list(1:1000,c("x4","x5","x6")))  
b3 <- matrix(F2 + rnorm(3\*1000), ncol = 3, dimnames = list(1:1000,c("x4","x5","x6")))  
  
d1 <- cbind(b1,b2)  
d2 <- cbind(b1,b3)  
  
model1 <- "F1 =~ x1 + x2 + x3; F2 =~ x4 + x5 + x6"  
model2 <- "F1 =~ x1 + x2 + x3; F2 =~ x4 + x5 + x6; F1~~1\*F2"  
  
table <- matrix(NA,10,8)  
i <- 1  
for(data in list(d1,d2)){  
 for(sd in sqrt(2) \* (2:6\*.25)){  
 test1 <- cfa(model1,data \* sd/sqrt(2))@test[[1]]$stat  
 test2 <- cfa(model2,data \* sd/sqrt(2))@test[[1]]$stat  
 test3 <- cfa(model2,data \* sd/sqrt(2), std.lv = TRUE)@test[[1]]$stat  
 table[i,] <- c(sd,test1,   
 test2, test2-test1, 1-pchisq(test2-test1, df=1),  
 test3, test3-test1, 1-pchisq(test3-test1, df=1))  
 i <- i+1  
 }  
}

write.table(round(table, digits=2), file ="Chi2(1) demo results.csv",  
 row.names = FALSE,  
 col.names = FALSE)