Extrinsic Convergent Validity to Prevent Jingle and Jangle Fallacies

2. Source Code

Below is annotated code to conduct tests for the difference of two dependent correlations in Mplus and R using lavaan. They all use the same example correlation matrix at the top of the code page. The tests compared if the correlation of V1 (r13=.4) and V2 (r23=.5) with V3 are the same in a sample size of N=103. This is the same example used in Steiger (1980).

$$\textbf{cordata} = \begin{bmatrix} 1.00 & .1 & .4 \\ .1 & 1.00 & .5 \\ .4 & .5 & 1.00 \end{bmatrix}$$

A.1 - Williams' Formula in R

The Williams' (1959) formula was programed in R as function *williams* (Lines 1-7) with the following inputs: the two correlations being compared, correlation between predictors, and sample size. The degrees of freedom then are computed (Line 8). Using the t-value from Williams' equation and the degrees of freedom, a p-value is computed (Line 9-10). The function then returns the correlation difference, the t-value, degrees of freedom, and p-value as a matrix (Line 12-15). With all of the inputs, the *williams* function can be used to test the difference of two dependent correlations (Line 17). Below is the R output for the example. The correlation difference was not statistically significant, which is consistent with the ECV hypothesis.

```
williams<-function(r13=r13,r23=r23,r12=r12,n=n){

cor_diff=r13-r23
williams_num=(n-1)*(1+r12)
Rdet=1-r13^2-r23^2-r12^2+(2*r12*r13*r23)
williams_den=2*((n-1)/(n-3))*Rdet+mean(c(r13,r23))^2*((1-r12)^3)
williams_rat=cor_diff*(sqrt(williams_num/williams_den))
df=n-3
pval_will=ifelse(williams_rat<0,pt(williams_rat,n-3)*2,
  pt(williams_rat,n-3,lower=FALSE)*2)

ret1=matrix(c(cor_diff,williams_rat,df,pval_will),ncol=4,nrow=1)
colnames(ret1)=c('Corr Diff','t-value','df','p_value')
return(ret1)
}

williams(r13=.40,r23=.50,r12=.10,n=103)
```

```
##      Corr Diff    t-value  df   p_value
## [1,]      -0.1 -0.8912799 100 0.3749185
```

A.2 - Delta Method in R

In order to use the delta method, the variance-covariance matrix of the correlations needs to be estimated using SEM. Therefore, one can call lavaan (Line 2) to specify and fit a model (Line 4-9) where the three variables are correlated. The variance-covariance matrix and the labels for the matrix are then extracted

(Line 10-11). We provide the function *mdelta* to test for the difference between two dependent correlations with the delta method using the following inputs: correlation 1 with output, variance of the correlation 1 estimate, correlation 2 with output, variance of correlation 2 estimate, and the covariance between correlation 1 and correlation 2 (Line 13-19). The function then returns the correlation difference, the z-value, and p-value as a matrix (Line 21-24). With all of the inputs, the *mdelta* function can be used to test the difference of two dependent correlations (Line 25). Below is the R output for the example. The correlation difference was not statistically significant, which is consistent with the ECV hypothesis.

```
require(lavaan)
```

```
## Loading required package: lavaan
```

```
## This is lavaan 0.6-4
```

```
## lavaan is BETA software! Please report any bugs.
```

```
N <- 103
cor_data<-rbind(c(1, .1, .4), c(.1, 1, .5), c(.4, .5, 1))
colnames(cor_data)<- c("V1", "V2", "V3")
rownames(cor_data)<-c("V1", "V2", "V3")
m1='
V1~~a*V2
V1~~b*V3
V2~~c*V3
'
model.fit=cfa(m1,sample.cov=cor_data,sample.nobs=103)
vfit=vcov(model.fit)
coef(model.fit)
```

```
##      a      b      c V1~~V1 V2~~V2 V3~~V3
##  0.099  0.396  0.495  0.990  0.990  0.990
```

```
mdelta<-function(r13=r13,var13=var13,r23=r23,
var23=var23,cov12=cov12){
cor_diff=r13-r23
mdelta_se=sqrt(var13+var23-2*cov12)
mdelta_rat=cor_diff/mdelta_se
pval_delta=ifelse(mdelta_rat<0,pnorm(mdelta_rat)*2,
pnorm(mdelta_rat, lower=FALSE)*2)

ret1=matrix(c(cor_diff,mdelta_rat,pval_delta),ncol=3,nrow=1)
colnames(ret1)=c('Corr Diff','z-value','p_value')
return(ret1)
}
mdelta(r13=.40,var13=.011,r23=.50,var23=.012,cov12=.003)
```

```
##      Corr Diff   z-value   p_value
## [1,]      -0.1 -0.766965 0.4431023
```

A.3 - Asymptotic Formulas for Delta Method in R

As noted in the text, there are asymptotic formulas for the variance and covariance of correlation estimates (Olkin & Siotani, 1976). We provide the *delta_method* function (Line 1) with the following inputs: the two correlations being compared, correlation between predictors, and sample size. The function estimates the asymptotic variance of the correlation estimates being compared (Line 3-4) and the asymptotic covariance between the correlations estimates (Line 6). The difference between the two correlation coefficients and the standard error of the difference is then estimated (Line 7-8), and the ratio of the difference to the standard error is compared to the standard normal distribution (Line 9-11). The function returns the difference between correlation coefficients, the z-value, and the p-value associated with the test (Line 13-16). Below is the R output for the example. The correlation difference was not statistically significant, which is consistent with the ECV hypothesis.

```
delta_method<-function(r13=r13,r23=r23,r12=r12,n=n){

var13=(1-r13^2)^2/n
var23=(1-r23^2)^2/n
var12=(1-r12^2)^2/n
cov13_23=(.5*(2*r12-r13*r23)*(1-r12^2-r23^2-r13^2)+r12^3)/n
cordif=r13-r23
difse=sqrt(var13+var23-2*cov13_23)
mdelta_rat=cordif/difse
pval_delta=ifelse(mdelta_rat<0,pnorm(mdelta_rat)*2,
pnorm(mdelta_rat, lower=FALSE)*2)

ret1=matrix(c(cordif,mdelta_rat,pval_delta),ncol=3,nrow=1)
colnames(ret1)=c('Corr Diff','z-value','p_value')
return(ret1)
}

delta_method(r13=.40,r23=.50,r12=.10,n=103)
```

```
##      Corr Diff    z-value   p_value
## [1,]      -0.1 -0.9019545 0.3670811
```

A.4 - Bootstrap Confidence Intervals in R

The full dataset is needed to carry out the bootstrap, however we do not have a full dataset in this example. For illustrative purposes only, we simulated 103 cases from the example correlation structure (Line 3) so that simulated dataset is bootstrapped. We do not recommend bootstrapping simulated datasets from summary information because they do not accurately reflect the original correlations. You would put your data in this section. First, a blank R list is created to hold the bootstrap results (Line 4). Then, the bootstrap is programmed as a loop where rows of the dataset are resampled with replacement. In each of the 500 datasets the difference between the correlations of interest is estimated and saved in the blank list (Line 5-11). The list is then compiled and the 2.5th and 97.5th percentile from the correlation differences are obtained (Line 13). Below is the R output for the example. The confidence interval contains zero, so the results are consistent with the ECV hypothesis.

```
#For the example we simulated 103 cases with cor_data as the
#true correlation matrix
require(MASS)
p3=mvrnorm(103,Sigma=cor_data,mu=rep(0,3))
diff_boot=list(NULL)
boot1=for(k in 1:500){
  ss=sample(1:nrow(p3),nrow(p3),replace=T)
  p_boot=p3[ss,]
  p_boot2=scale(p_boot)
  cor_boot=cor(p_boot2)
  diff_boot[[k]]=cor_boot[1,3]-cor_boot[2,3] #cor's compared
}

boot_ci=quantile(do.call(rbind,diff_boot),c(.025,.975))
```

```
boot_ci
```

A.5 - Chi-square Test of Model Fit in R

A structural equation model estimating the correlations between the three variables is specified and estimated in *lavaan* (Line 2-10), however the correlations of V1 and V2 with the outcome V3 are constrained to be the same by giving the parameters the same label, *b*. Parameter estimates and fit information (Lines 12-14) are then extracted. Below is the output from the example. The correlation difference was not statistically significant, which is consistent with the ECV hypothesis.

```
require(lavaan)

m1='
V1~~a*V2
V1~~b*V3
V2~~b*V3
'

model.fit.cons=cfa(m1,sample.cov=cor_data,sample.nobs=103)
mcons=coef(model.fit.cons)
mfit=fitMeasures(model.fit.cons,fit.measures=c('chisq','df',
'pvalue','cfi', 'srmr','rmsea'))
mfit
```

```
##  chisq     df pvalue    cfi   srmr  rmsea
##  0.574  1.000  0.449  1.000  0.039  0.000
```

A.6 - Mplus to extract information for Multivariate Delta Method

Mplus can be used to estimate the variance-covariance matrix of the correlations so that the multivariate delta method test can be computed using the appendix section A.2. First, information about the dataset is provided, such as the file, type of data (in this case it is a correlation matrix), number of observations, and variable names (Line 1-5). The model is then specified allowing all variables to correlate with each other (Line 7-11). Finally, sample statistics, tech1 and tech3 are specified in the output line (Line 13). Tech3 is used to obtain the estimated variance-covariance matrix of the correlation parameter estimates and tech1 provides information to read tech3 a bit easier. The tech3 information is located toward the bottom of the output file and it provides the information needed to use the *mdelta* function described in section A.2. Information from tech3 is the following: Variance of (a) = 0.0110; Variance of (b) = 0.0119, Covariance of (a,b) = 0.0029

```
Title: ECV multivariate delta by hand;
Data: file is cor_data.csv;
type=fullcorr;
nobservations=103;
variable: names are v1-v3;

model:

v1 with v2(a);
v1 with v3(b);
v2 with v3(c);

output: tech1 tech3 sampstat;
```

A.7 - Mplus for the Bootstrap

Mplus also can be used to estimate confidence intervals for the difference of two dependent correlations. Again, for illustrative purposes only, we bootstrapped the simulated data described in the appendix section A.4. First, information about the dataset is provided, such as the file, type of data (in this case it is the full dataset), and variable names (Line 1-3). Then the bootstrap with 500 datasets is specified under the analysis command (Line 5-6). The model is then specified allowing all variables to correlate with each other and assigning parameter labels in parentheses (Line 7-11). Next, under the model constraint command, a new variable is defined (using the parameter labels in the model command) as the difference between the correlation estimates of interest. Finally, sample statistics, tech1, and confidence intervals are specified in the output line (Line 17). Below is the *Mplus* output for the example. The confidence interval contains zero, so the results are consistent with the ECV hypothesis.

parameter	Lower .5%	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%	Upper .5%
cor_diff	-0.512	-0.377	-0.343	-0.089	0.134	0.181	0.266

```
Title: ECV bootstrap;
Data: file is full_data.csv; !This is a simulated dataset
Variable: names are v1-v3;

Analysis:
bootstrap=500;
model:

v1 with v2(a);
v1 with v3(b);
v2 with v3(c);

Model constraint:
New(cor_diff);
cor_diff=b-c;

output: tech1 cinterval(bootstrap) sampstat;
```

A.8 - Chi-square test of model fit in Mplus

A structural equation model estimating correlation between the three variables is specified and estimated in Mplus (Line 1-12), however the correlations of V1 and V2 with the outcome V3 are constrained to be the same by giving the parameters the same label, b. Fit information is provided towards the top of the Mplus output file. Below is the Mplus output from the example. The chi-square test of model fit was not statistically significant, which is consistent with the ECV hypothesis. Chi-Square Test of Model Fit

```
Value                           0.574
Degrees of Freedom              1
P-Value                         0.4487
```

Mplus to extract information for Multivariate Delta Method

```
Title: ECV chi-square for model fit;
Data: file is cor_data.csv;
type=fullcorr;
```

```
nobservations=103;
variable: names are v1-v3;

model:

v1 with v2(a);
v1 with v3(b);
v2 with v3(b);
output: tech1 sampstat;
```

A.9 - Mplus Cor Difference through Model Constraint

The model constraint command can be used to obtain an estimate and a standard error for the difference between two dependent correlations. A structural equation model estimating correlation between the three variables is specified and estimated in Mplus (Line 1-10), assigning parameter labels in parentheses. Under the model constraint command, a new variable is defined using the parameter labels in the model command, as the difference between the correlation estimates of interest (Line 12-15). The new estimate and its standard error are reported along with the individual parameters. Below is the Mplus output from the example. The correlation difference was not statistically significant, which is consistent with the ECV hypothesis.

```
              Estimate    S.E.   Est./S.E.  Two-Tailed P-Value
   COR_DIFF    -0.099    0.131   -0.754       0.451
```

```
Title: ECV delta method in constraints;
Data: file is cor_data.csv;
type=fullcorr;
nobservations=103;
variable: names are v1-v3;
model:

v1 with v2(a);
v1 with v3(b);
v2 with v3(c);

Model constraint:
New(cor_diff);
cor_diff=b-c;
output: tech1 sampstat;
```