Title: Structural Model Fit and Construct Reliability: A Monte Carlo Examination of Available Methods
Authors:

1. Graham G. Rifenbark, PhD, University of Kansas: grahamr@ku.edu, **Presenter**

2. D. Betsy McCoach, PhD, University of Connecticut: betsy.mccoach@uconn.edu

3. H. Jane Rogers, PhD, University of Connecticut: jane.rogers@uconn.edu

Choice of conference section:

1. Research Methods

Background/Context

The evaluation of misfit in structural equation models (SEM) is an area of great importance, as the structural parameters are those we utilize to make population level inferences about some causal process. Recently, structural fit indices (SFIs) have been advanced due to the influence of the measurement model on the approximate fit indices (AFIs). First, AFIs are overly weighted by the measurement model (McDonald & Ho, 2002) and this increases when the number of indicators per factor increases (p:f). Second, AFI cut-offs were not determined in the context of varying measurement quality (MQ); as a rexsult, model fit appears to improve as measurement quality decreases, known as the *reliability paradox* (Hancock & Mueller, 2011). Gagne and Hancock (2006) declare that MQ the number of indicators per factor (p:f) are pivotal to construct reliability. In SEM, MQ corresponds to standardized factor loadings. Therefore, as MQ increases, so does construct reliability. On the other hand, as p:f increases the resulting latent construct becomes more reliable as its domain is more fully captured. The approach advanced by Hancock and Mueller (2011) requires two stages of estimation - hereafter referred to as SM-MV, whereas, the approach advanced by Lance, Beck, Fan, and Carter (2016) is accomplished by simultaneous estimation of all model parameters - hereafter referred to as SM-LV, this is also the case for the RMSEA-P of McDonald and Ho (2002). An inherent issue with two-stage estimation approaches is their inability to account for the uncertainty around the parameter estimates from the initial stage in the subsequent stage (Levy, 2017). The SM-MV approach makes no attempt to account for the uncertainty around the model-implied latent variance-covariance matrix. McNeish and Hancock (2018) recommends the SM-MV approach and asserts that it provides an avenue to evaluate structural model fit in isolation while being immune to MQ. McNeish and Hancock (2018) also asserts that the SM-MV approach can be used to bolster the performance of the SM-LV approach.

Research Questions

How does model size, MQ, and group sample size affect the performance of AFIs, SFIs (e.g., SM-MV, SM-LV, RMSEA-P), and traditional tests (e.g., χ^2 and $\triangle\chi^2$) in the context of multiple group models given a structural model that is: 1) correctly specified, 2) mis-specified mean structure, 3) mis-specified covariance structure, 4) or simultaneous misspecifications in the mean and covariance structures.

It was hypothesized that the SFIs estimated using the SM-MV approach would be negatively impacted by MQ, due to the two-stage nature of the approach. On the other hand, it was hypothesized that the SM-LV and RMSEA-P SFIs would be unimpacted by MQ. With respect to the RMSEA-P, it was hypothesized that it would be negatively impacted by model size, especially when there are few indicators. We suspected that the SM-LV and the RMSEA-P would outperform the AFIs when detecting a mis-specified model.

Simulation Conditions

Type I Error simulation. We systematically varied levels of MQ, p:f (model size), and group sample sizes. These design factors were fully crossed leading to 18 unique simulation conditions. See Table 1.

Power simulation. We investigated the type of misspecification and the severity of the structural misspecification. A total of 15 misspecification conditions were examined, when crossed with the 18 conditions above, this led to 270 unique simulations (e.g., 15*18).

Data Generation/Analysis

See Figure 1 for the data generation model. With respect to the measurement parameters, the manifest residuals (or θ) depends on MQ, therefore, these parameters were set to $1 - \lambda^2$. Manifest intercepts were all set to 0.0.

In the Type I Error simulation, data was generated according to full measurement and structural invariance. For the analysis, full measurement and structural invariance was modeled (i.e., no misspecifications are introduced). See Tables 2 through 4 for all population values for the structural model.

In the Power simulation true differences between groups were generated. Specifically, the mean of X3 (ν_{13}) or the regression of Y2 on X2 (β_{52}), when simultaneous differences were generated, these parameters were again chosen. For the analysis, full measurement and structural invariance was modeled, thus introducing misspecifications. See Table 5 for the population values when the mean structure was misspecified and Table 6 when the covariance structure was misspecified.

Data were generated in R (R Core Team, 2017) using the simsem (Pornprasertmanit, Miller, & Schoemann, 2016) package. Latent variable models were executed in lavann (Rosseel, 2012). A total of 1000 replications were executed for each condition.

Simulation Outcomes

Type I Error simulation. We conducted an ANOVA for each of the estimated measures across all 18 simulation conditions. To assess the effect of MQ, p:f, and group size, we tested all main and interaction effects. We consulted the effect size, partial η^2 and considered 0.01, 0.06, and 0.14 as small, medium, and large effect sizes, respectively

(Cohen, 1988). Measures that were found to be unaffected were retained and empirical cut-offs were established at the 95^{th} percentile.

Power simulation. Using the above cut-offs, we generated hit rates for each of the measures and subjected these hit rates to ANOVA and utilized partial η^2 to guide us.

Findings/Results

Type I Error simulation. See Table 7 for the effects on global measures. The SM-LV SFIs (C9-C10), RMSEA-P, and the $\triangle\chi^2$ were unaffected by all design factors. MQ had a large effect on the SM-MV measures, see Table 8. To visualize the effect of MQ see Figure 2 (TLI) and Figure 3 (RMSEA). Due to these findings, we established empirically derived cut-offs for the TLI, Mc, RMSEA, SM-LV (C9 and C10), and the RMSEA-P. See Table 9 for these cut-offs values.

Power simulation. We found that measures of fit possessed more statistical power to correctly reject a model when the mean structure was mis-specified rather than the covariance structure – see Table 10 and 11 for the global and structural measures of fit, respectively. When the structures were simultaneously mis-specified, all measures (except TLI) were adequately powered to detect the structural model misfit – see Table 12. When the covariance structure was mis-specified, MQ had a larger influence on power rates, than when only the mean structure was mis-specified: see Figures 4 (RMSEA), 5 (RMSEA-P), and 6 (SM-LV, C9).

Conclusions

In the Type I Error simulation, we illustrated the importance of estimating the measurement and structural model simultaneously. We the authors, do not endorse the method put forth by Hancock and Mueller (2011) and that McNeish and Hancock (2018) promotes. When detecting structural misspecification(s) we suggest the use of $\triangle\chi^2$, SM-LV, and RMSEA-P.

References

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences.* Hillsdale, NJ: erlbaum.

Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*(1), 65–83.

Hancock, G. R., & Mueller, R. O. (2011). The reliability paradox in assessing structural relations within covariance structure models. *Educational and Psychological Measurement*, *71*(2), 306–324.

Lance, C. E., Beck, S. S., Fan, Y., & Carter, N. T. (2016). A taxonomy of path-related goodness-of-fit indices and recommended criterion values. *Psychological methods*, *21*(3), 388.

Levy, R. (2017). Distinguishing outcomes from indicators via bayesian modeling. *Psychological methods*, *22*(4), 632.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological methods*, *7*(1), 64.

McNeish, D., & Hancock, G. R. (2018). The effect of measurement quality on targeted structural model fit indices: A comment on lance, beck, fan, and carter (2016).

Pornprasertmanit, S., Miller, P., & Schoemann, A. (2016). simsem: Simulated structural equation modeling [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=simsem` (R package version 0.5-13)

R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from `http://www.jstatsoft.org/v48/i02/`

Table 1
Type I Error Simulation Conditions
Note. MQ = measurement quality; p:f = number of manifest variables per latent variable; G1.n = group one sample size; G2.n = group two sample size.

Condition No.	MQ	p:f	G1.n	G2.n
1	0.400	3	600	1400
2	0.600	3	600	1400
3	0.800	3	600	1400
4	0.400	5	600	1400
5	0.600	5	600	1400
6	0.800	5	600	1400
7	0.400	3	1000	1000
8	0.600	3	1000	1000
9	0.800	3	1000	1000
10	0.400	5	1000	1000
11	0.600	5	1000	1000
12	0.800	5	1000	1000
13	0.400	3	1400	600
14	0.600	3	1400	600
15	0.800	3	1400	600
16	0.400	5	1400	600
17	0.600	5	1400	600
18	0.800	5	1400	600

Table 2
***B**, Matrix of Latent Regressions*

	X1	X2	X3	Y1	Y2
X1	0	0	0	0	0
X2	0	0	0	0	0
X3	0	0	0	0	0
Y1	0.3	0.3	0.3	0	0
Y2	0	0.4	0	0.3	0

Table 3
$\boldsymbol{\Psi}$, Matrix of Latent Variances and Disturbances

	X1	X2	X3	Y1	Y2
X1	1	0.200	0.200	0	0
X2	0.200	1	0.200	0	0
X3	0.200	0.200	1	0	0
Y1	0	0	0	0.622	0
Y2	0	0	0	0	0.649

Table 4
ν, Vector of Latent Means and Intercepts

X1	X2	X3	Y1	Y2
0	0	0	1.5	0.75

Table 5
Population Values: Mean Structure X Group
Note. G1 = Group 1 and G2 = Group 2; ν_{13} = latent mean for X3; ν_{14} = latent intercept for Y1; ν_{15} = latent intercept for Y2

	ν_{11}	ν_{12}	ν_{13}	ν_{14}	ν_{15}
G1	0	0	0	1.500	0.750
G2 (small)	0	0	0.200	1.440	0.750
G2 (medium)	0	0	0.500	1.350	0.750
G2 (large)	0	0	0.800	1.260	0.750

Table 6
Population Values: Covariance Structure X Group
Note. G1 = Group 1 and G2 = Group 2; B_{52} = standardized regression weight $X2 \rightarrow Y2$; Ψ_{55} = Y2 disturbance

	Small (d = 0.2)	Medium (d = 0.4)	Large (d = 0.6)
G1:B_{52}	0.300	0.200	0
G1:Ψ_{55}	0.744	0.820	0.910
G2:B_{52}	0.500	0.600	0.600
G2:Ψ_{55}	0.534	0.399	0.399

Table 7
Effect of Design Factors: Global Measures
Partial $\eta^2 < 0.01$ *are left blank and empty columns are removed. MQ = measurement quality; pF = model size (p:f); bal.n = balanced groups; Mc = McDonald's measure of centrality; *.afi = global measure of fit.*

	pF(A)	MQ(B)	bal.n(C)	AxB
χ^2	0.980			
CFI		0.181		
TLI				
Mc				
rmsea.afi				
srmr.afi	0.096	0.342	0.063	0.029

Table 8
Effect of Design Factors: SM-MV Measures
Partial $\eta^2 < 0.01$ *are left blank and empty columns are removed. MQ = measurement quality; pF = model size (p:f); bal.n = balanced groups; Mc = McDonald's measure of centrality; *.sfi = structural measure of fit.*

	pF(A)	MQ(B)	bal.n(C)	AxB
χ^2.sfi	0.015	0.075		0.020
CFI.sfi	0.085	0.363		0.100
TLI.sfi	0.085	0.364		0.099
Mc.sfi	0.095	0.388		0.113
rmsea.sfi	0.175	0.646		0.104
srmr.sfi	0.219	0.704	0.027	0.161

Table 9
Empirically Derived Cut-Off values
Note. 95, $\alpha = 0.05$*;*
*Mc = McDonald's measure of centrality; *.afi = global measure of fit; ncp = non-centrality parameter; rmsea.path = RMSEA-P*

Measure	95
TLI	0.979
Mc	0.987
rmsea.afi	0.012
C9.ncp	0.994
C10.ncp	0.006
rmsea.path	0.017

Table 10
Hit Rates: Global Measures
Note. μ = *hit rate;* σ = *standard deviation of hit rate; 95 =* α *of 0.05; tli = Tucker-Lewis index; mc = McDonald's measure of centrality; *.afi = global measure of fit*

Type of Misspecification:		Mean Structure		Covariance Structure	
Statistic	N	μ	σ	μ	σ
tli.afi.95	54,000	0.466	0.499	0.253	0.435
mc.afi.95	54,000	0.587	0.492	0.422	0.494
rmsea.afi.95	54,000	0.705	0.456	0.566	0.496
χ^2.95	54,000	0.677	0.468	0.535	0.499

Table 11
Hit Rates: Structural Measures
Note. μ = *hit rate;* σ = *standard deviation of hit rate; 95 =* α *of 0.05; ncp = non-centrality parameter; rmsea.p = RMSEA-P.*

Type of Misspecification:		Mean Structure		Covariance Structure	
Statistic	N	μ	σ	μ	σ
c9.ncp.95	54,000	0.769	0.421	0.743	0.437
c10.ncp.95	54,000	0.769	0.421	0.743	0.437
rmsea.p.95	54,000	0.817	0.387	0.760	0.427
$\triangle\chi^2$.95	54,000	0.827	0.378	0.771	0.420

Table 12
Hit Rates: Simultaneous Misspecification
Note. μ = *hit rate;* σ = *standard deviation of hit rate; 95 =* α *of 0.05; tli = Tucker-Lewis index; mc = McDonald's measure of centrality; *.afi = global measure of fit; N = total number of replications*

Measure	N	μ	σ
tli.afi.95	162,000	0.547	0.498
mc.afi.95	162,000	0.817	0.387
rmsea.afi.95	162,000	0.818	0.386
χ^2.95	162,000	0.850	0.357
c9.ncp.95	162,000	0.970	0.171
c10.ncp.95	162,000	0.970	0.171
rmsea.p.95	162,000	0.967	0.178
$\triangle\chi^2$.95	162,000	0.970	0.172

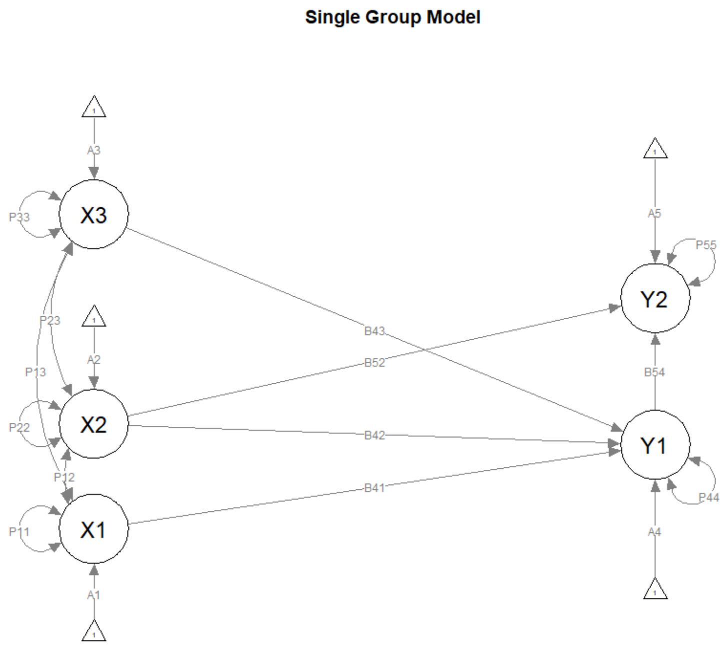

Figure 1. Path Diagram: Data Generating Model
Single group model with no misspecifications. Exogenous latent variables: X1, X2, X3. Endogenous latent variables: Y1 and Y2.

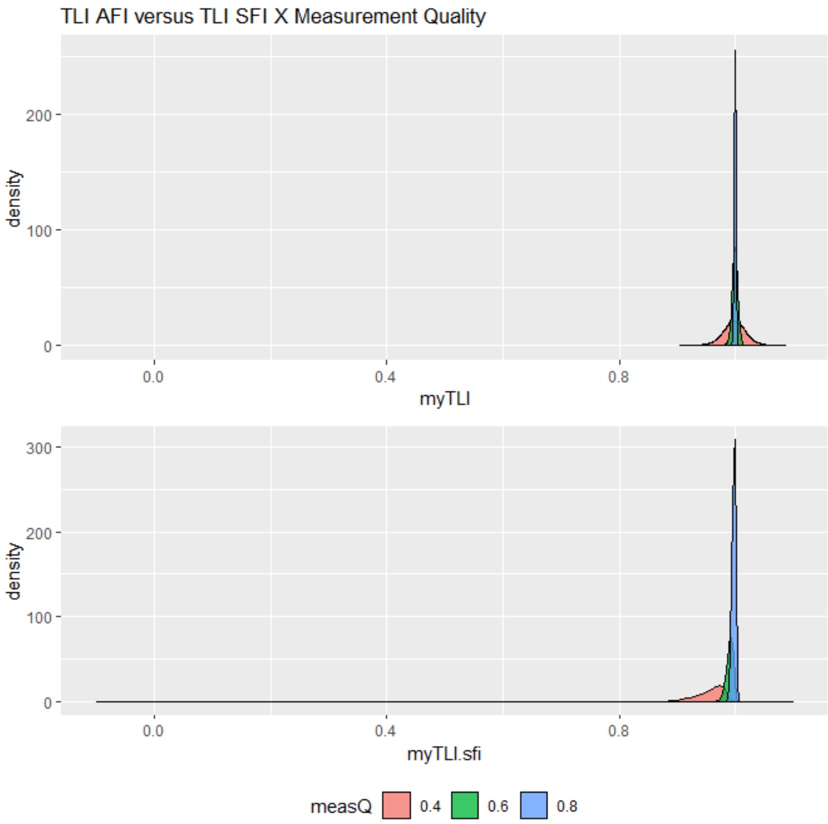

Figure 2. TLI Comparison
Global TLI distribution versus SM-MV TLI distribution. Note. myTLI utilizes the manually specified baseline model; *.sfi = structural measure of fit.

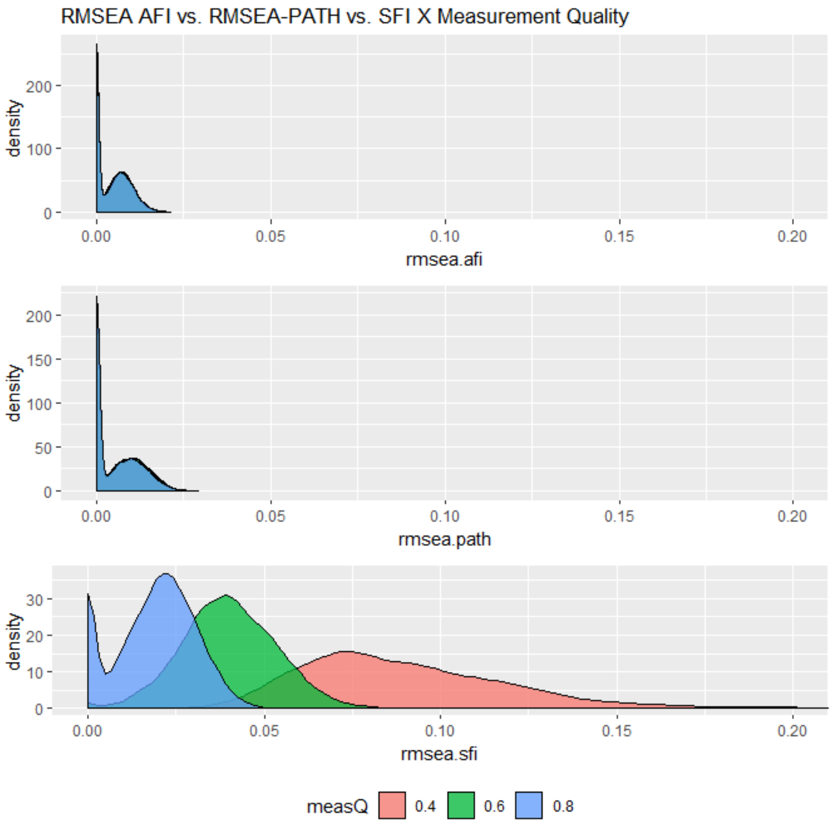

Figure 3. RMSEA Comparison
Global RMSEA versus RMSEA-P versus SM-MV RMSEA Distributions. *.sfi = structural measure of fit.

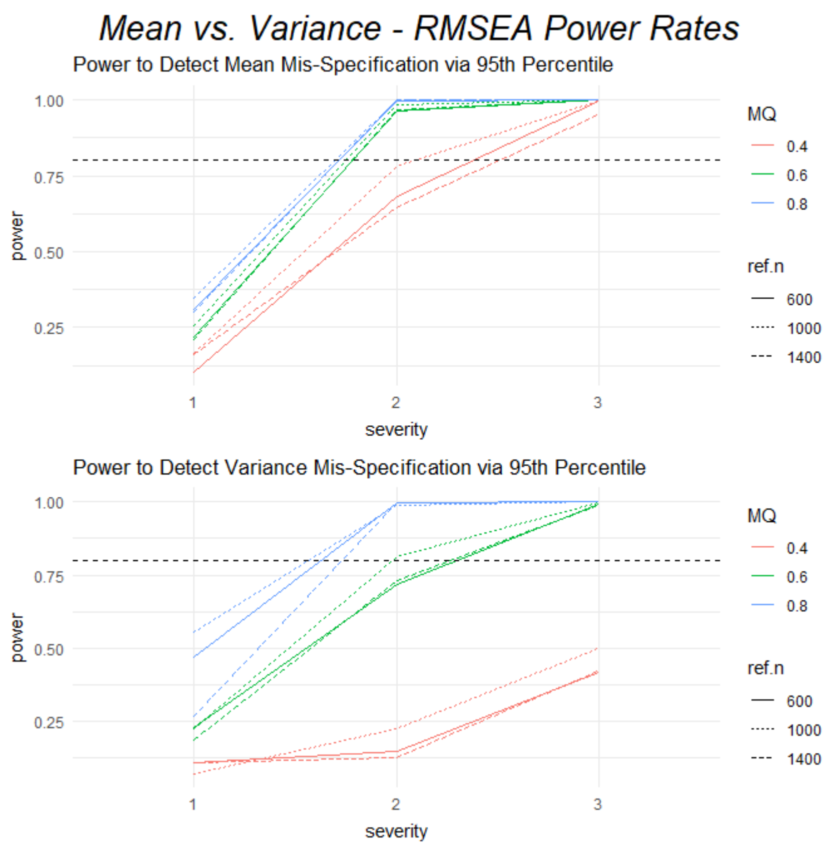

Figure 4. Mean versus covariance structure misspecification
severity 1 = small, 2 = medium, 3 = large

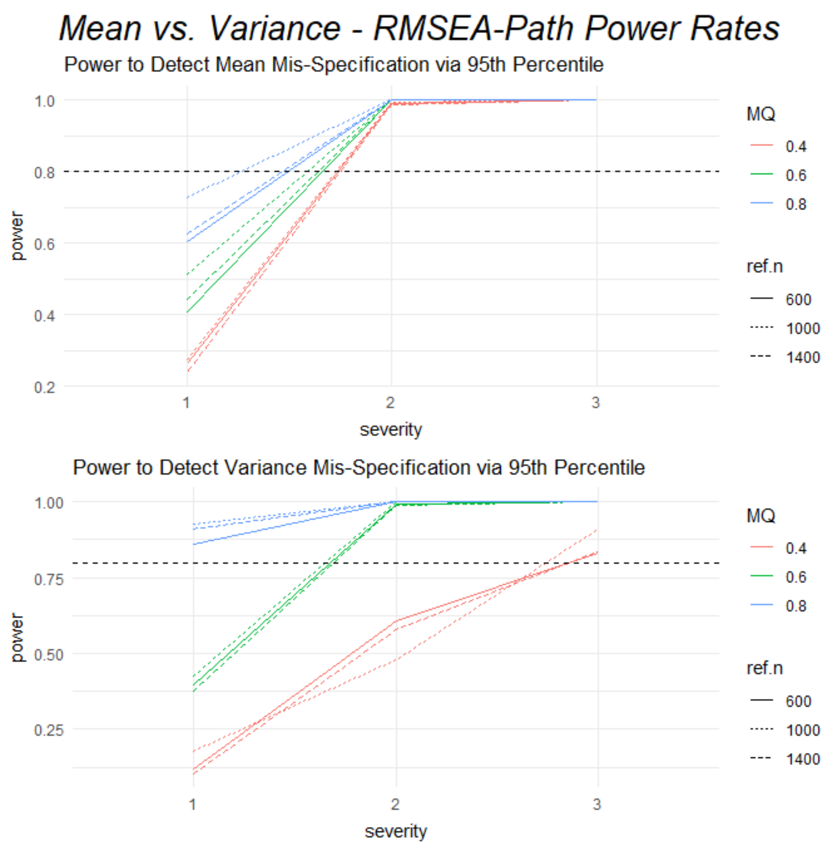

Figure 5. Mean versus covariance structure misspecification
severity 1 = small, 2 = medium, 3 = large

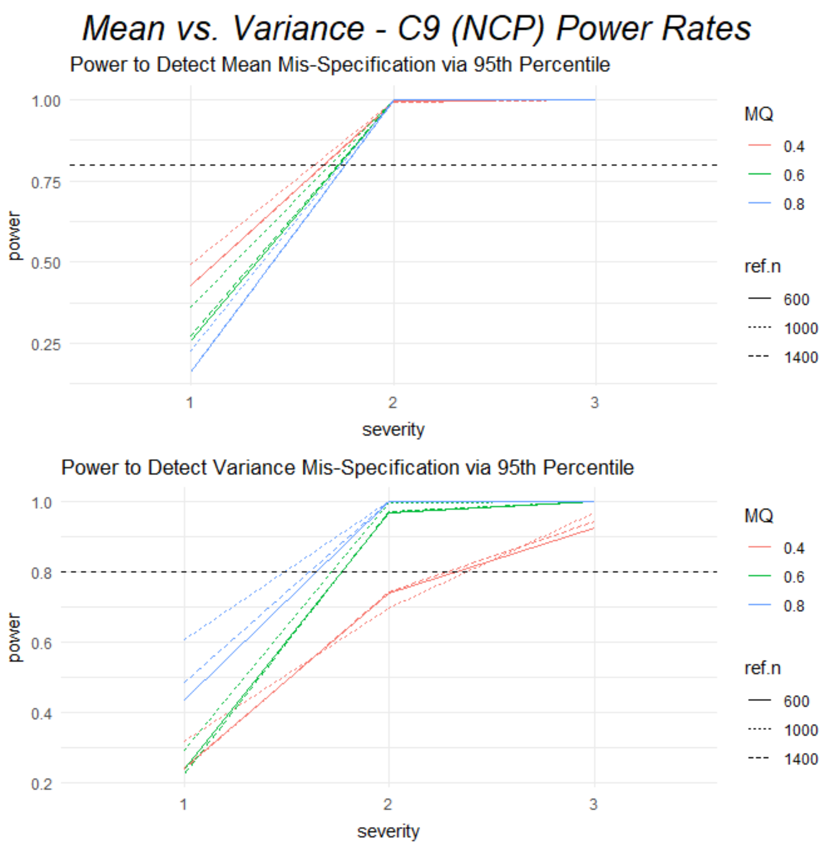

Figure 6. Mean versus covariance structure misspecification
severity 1 = small, 2 = medium, 3 = large