Psychology 405: Latent Variable Modeling
How do you know if a model works?

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

March, 2015

Outline

A number of tests of fit taken from Marsh et al. (2005)

1. Marsh, Hau & Grayson (2005) lists 40 different proposed measures of goodness of fit
2. Measures of absolute fit
   - I_o = index of fit for original or baseline model
   - I_t = index of fit for target or "true" model
3. Measures of incremental fit Type I
   - $\frac{|I_t - I_o|}{Max(I_o, I_t)}$ which is either
     - $\frac{I_o - I_t}{I_o}$
     - or $\frac{I_t - I_o}{I_t}$
4. Measures of incremental fit Type II
   - $\frac{|I_t - I_o|}{E(I_t - I_o)}$ which is either
     - $\frac{I_o - I_t}{I_o - E(I_t)}$
     - or $\frac{I_t - I_o}{E(I_t) - I_o}$

Fit functions from Jöreskog

1. Ordinary least squares $F = \frac{1}{2}tr(S - \Sigma)^2$
   - The squared difference between the observed (S) and model (Σ) covariance matrices
   - tr means trace of the sum of the diagonal values of the matrix of squared deviations
2. Generalized least squares $F = \frac{1}{2}tr(I - S^{-1}\Sigma)^2$
   - I is the identity matrix
   - if the model = data, then $S^{-1}\Sigma$ should be I
   - weight the fit by the inverse of the observed covariances
3. Maximum Likelihood $F = log|\Sigma| + tr(S\Sigma^{-1}) - log|S| - p$
   - weight the fit by the inverse of the modeled covariance
   - p is the number of variables
   - tr (I) = p, and thus the ML should be 0 if the model fits the data

Fit-function based indices

1. Fit Function Minimum fit function (FF)
   - $FF = \frac{\chi^2}{(N-1)}$
2. Likelihood ratio $LHR = e^{-1/2FF}$
3. χ^2 (minimum fit function chi square)
   - $\chi^2 = tr(\Sigma^{-1}S - I) - log|\Sigma^{-1}S| = (N-1)FF$
4. $p(\chi^2)$ probability of observing a χ^2 this larger or larger given that the model fits
5. $\frac{\chi^2}{df}$ has expected value of 1

Non-centrality based indices

1. Dk: Rescaled noncentrality paramter (McDonald & Marsh, 1990)
   - $Dk = FF - df/(N-1) = \frac{\chi^2 - df}{N-1}$
2. PDF (population discrepancy function = DK normed to be non-negative)
   - $PDF = max(\frac{\chi^2 - df}{N-1}, 0)$
3. Mc: Measure of centrality (CENTRA, MacDonald Fit Index (MFI)
   - $Mc = e^{\frac{-(\chi^2 - df)}{2(N-1)}}$
4. Non-centrality parameter
   - $NCP = \chi^2 - df$

Error of approximation indices

How large are the residuals, estimated several different ways

1. RMSEA (root mean square error of approximation)
   - $RMSEA = \sqrt{PDF/df} = \sqrt{\frac{max(\frac{\chi^2 - df}{N-1}, 0)}{df}}$
   - based upon the non-central χ^2 distribution to find the error of fit
2. MSEA (mean square error of approximation – unnormed version of RMSEA)
   - $MSEA = \frac{Dk}{df} = \frac{\chi^2 - df}{(N-1)df}$
3. RMSEAP (root mean square error of approximation of close fit)
   - RMSEA $< .05$
4. RMR Root mean square residual (or, if S and Σ are standardized, the SRMR). Just
   - square root of the average squared residual
   - $\sqrt{\frac{2\sum(S-\Sigma)^2}{p*(p+1)}}$

Information indices

Compare the information of a model with the number of parameters used for the model. These allow for comparisons of different models with different degrees of freedom.

1. AIC (Akaike Information Criterion for a model penalizes for using up df)
   - $AIC = \chi^2 + p * (p+1) - 2df = \chi^2 + 2K$
   - where $K = \frac{p*(p+1)}{2} - df$
2. Baysian Information Criterion
   - $-2Log(L) + plog(N) = \chi^2 - Klog(N(.5(p(p+1))$

Goodness of fit indices

1. GFI from LISREL
   - $GFI = 1 - \frac{tr(\Sigma^{-1}S - I)^2}{tr(\Sigma^{-1}S)^2}$
2. Adjusted Goodness of Fit (Lisrel)
   - $AGFI = 1 - \frac{p(p+1)}{2df}(1 - GFI)$
3. Unbiased GFI (from Steiger)
   - $\text{GFI} = \frac{p}{2\frac{(\chi^2 - df)}{(N-1)} + p}$

Comparing solutions to solutions

1. Incremental fit indices without correction for model complexity
   - RNI (relative non-centrality) McDonald and Marsh
   - CFI Comparative fit index (normed version of RNI) Bentler
   - Normed Fit index (Bentler and Bonett)
2. Incremental fit indices correcting for model complexity
   - Tucker - Lewis Index
   - Normed Tucker Lewis
   - Incremental Fit index
   - Relative Fit Index
3. Parsimony indices

Incremental fit indices without correction for model complexity

1. RNI (relative non-centrality) McDonald and Marsh
   - $RNI = 1 - \frac{Dk_t}{Dk_n}$
   - where $DK = \frac{\chi^2 - df}{N-1}$ for either the null or the tested model
2. CFI Comparative fit index (normed version of RNI) Bentler
   - Just norm the RNI to be greater than 0.
   - $CFI = 1 - \frac{MAX(NCP_t, 0)}{MAX(NCP_n, 0)}$
3. Normed Fit index (Bentler and Bonett)

Fitting functions from Loehlin

1. Let S be the “strung out” data matrix
2. Let Σ be the “strung out” model matrix
3. $Fit = (S - \Sigma)' W^{-1} (S - \Sigma)$
4. Where W =
   - Ordinary Least Squares $W = I$
   - Generalized Least Squares $W = SS'$
   - Maximum likelihood $W = \Sigma\Sigma'$

Practical advice

1. Taken from Kenny
   - http://davidkenny.net/cf/fit.htm
2. Bentler and Bonnet Normed Fit Index
   - $\frac{\chi^2_{Null} - \chi^2_{Model}}{\chi^2_{Null}}$
   - Between .90 and .95 is acceptable
   - $>$ .95 is "good"
3. RMSEA
   - if $\chi^2 < df$, then RMSEA $= 0$
   - "good" models have $RMSEA < .05$
   - "poor" models have $RMSEA > .10$
4. p of close fit
   - Null hypothesis is that RMSEA is .05
   - test if $RMSEA > .05$
   - Claim good fit if $p(RMSEA > .05) > .05$

Fits and sample size

1. See associated simulation results
    -
2. -

Considering rules of thumb and fit

1. Fit functions have distributions and thus are susceptible to problems of type I and type II error.
   - Compare the fits for correct model as well as those for a simple incorrect
2. Should we just use chi square and reject models that don't fit, or should we reason about why they don't fit

What does it mean if the model does not fit

1. Model is wrong
2. Measurement is wrong
3. Structure is wrong
4. Assumptions are wrong
5. at least one of above, but which one?

Specification & Respecification

1. Is the measurement model consistent
    - revise it
        - evaluate loadings
        - evaluate error variances
        - more or fewer factors
        - correlated errors?
2. Structural model:
    - adjust paths
    - drop paths
    - add paths
3. Equivalent models
    - What models are equivalent
    - Do they make equally good sense

44 ways to fool yourself with SEM

Adapted from Rex Kline; Principals and Practice of Structural Equation Modeling, 2005

1. Specification
2. Data
3. Analysis and Respecicaton
4. Interpretation

Specification errors

1. Specifying the model after the data are collected.
   - Particularly a problem when using archival data.
2. Are key variables omitted?
3. Is the model identifiable?
4. Omitting causes that are correlated with other variables in the structural model.
5. Failure to have sufficient number of indicators of latent variables.
   - "Two might be fine, three is better, four is best, anything more is gravy" (Kenny, 1979)
6. Failure to give careful consideration to directionality.
   - Path techniques are good for estimating strengths if we know the underlying model, but are not good for determining the model (Meehl and Walker, 2002)

Specification errors (continued)

7. Specifying feedback loops (“non recursive models”) as a way to mask uncertainty
8. Overfit the model, ignoring parsimony
9. Add disturbances (“measurement error correlations” aka “correlated residuals”) with substantive reason
10. Specifying indicators that are multivocal without substantive reason

Data Errors

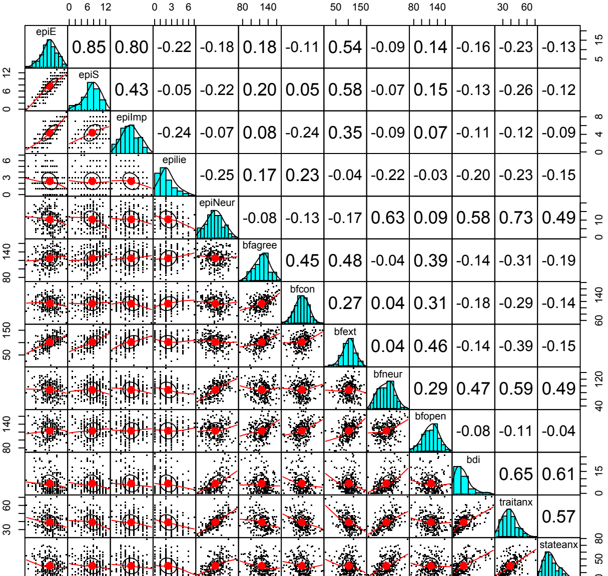
1. Failure to check the accuracy of data input or coding
   - Missing data codes (use a clear missing value)
   - Misytyped, mis-scanned data matrices
   - Improperly reversed items
     - Let the computer do it for you
     - Why reverse an item when a negative sign will do it for you?
2. Ignoring the pattern of missing data, is it random or systematic.
3. Failure to examine distributional characteristics
   - Weird data -> weird results
4. Failure to screen for outliers
   - Outliers due to mistakes
   - Outliers due to systematic differences

Describe the data

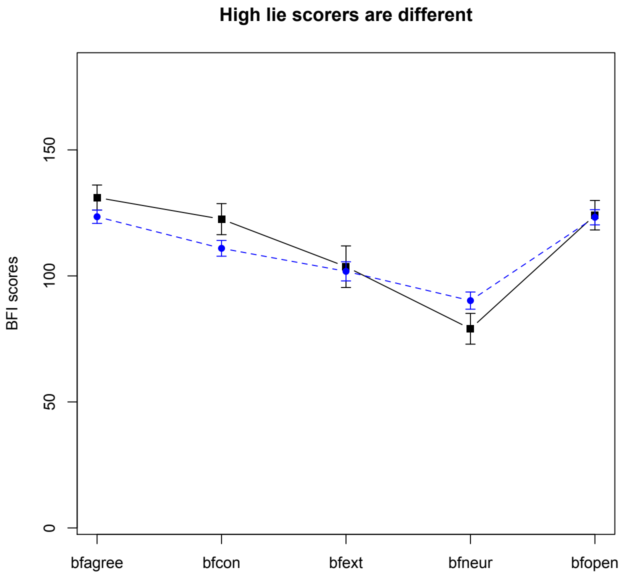
```
> describe(epi.bfi)

 pairs.panels(epi.bfi,pch=".",gap=0)   #mind the gap
```

	var	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
epiE	1	231	13.33	4.14	14	13.49	4.45	1	22	21	-0.33	-0.06	0.27
epiS	2	231	7.58	2.69	8	7.77	2.97	0	13	13	-0.57	-0.02	0.18
epiImp	3	231	4.37	1.88	4	4.36	1.48	0	9	9	0.06	-0.62	0.12
epilie	4	231	2.38	1.50	2	2.27	1.48	0	7	7	0.66	0.24	0.10
epiNeur	5	231	10.41	4.90	10	10.39	4.45	0	23	23	0.06	-0.50	0.32
bfagree	6	231	125.00	18.14	126	125.26	17.79	74	167	93	-0.21	-0.27	1.19
bfcon	7	231	113.25	21.88	114	113.42	22.24	53	178	125	-0.02	0.23	1.44
bfext	8	231	102.18	26.45	104	102.99	22.24	8	168	160	-0.41	0.51	1.74
bfneur	9	231	87.97	23.34	90	87.70	23.72	34	152	118	0.07	-0.55	1.54
bfopen	10	231	123.43	20.51	125	123.78	20.76	73	173	100	-0.16	-0.16	1.35
bdi	11	231	6.78	5.78	6	5.97	4.45	0	27	27	1.29	1.50	0.38
traitanx	12	231	39.01	9.52	38	38.36	8.90	22	71	49	0.67	0.47	0.63
stateanx	13	231	39.85	11.48	38	38.92	10.38	21	79	58	0.72	-0.01	0.76

Graphic descriptions using SPLOMs



High lie score subjects seem different



Data errors (continued)

5. Assuming all relationships are linear without checking
   - graphical techniques are helpful for non-linearities
   - Simple graphical techniques do not help for interactions
6. Ignoring lack of independence among observations
   - Nesting of subjects within pairs, within classrooms, with managers
   - Can we model the nesting?

Errors of analysis and respecification

1. Failure to check the accuracy of computer syntax
   - Direction of effects
   - Error specifications
   - Omitted paths
2. Respecifying the model based entirely on statistical criteria
   - Just because it does not fit does not mean it should be fixed
3. Failure to check for admissible solutions
   - Are some of the paths impossible?
   - Do some of the variables have negative variances?
4. Reporting only standardized estimates
   - These are sample based estimates and reflect variances (errorful) and covariances (supposedly error free)
5. Analyzing a correlation matrix when the covariance matrix is more appropriate
   - Anything that has growth or change component must be done with covariances

Errors of Analysis and respecification (continued)

6. Analyzing a data set with extremely high correlations
   - solution will either be unstable or will not work if variables are too "colinear"
7. Not enough subjects for complexity of the data
   - This is ambiguous – what is enough?
   - Remember, the standard error of a correlation reflects sample size $se_r = \frac{1}{\sqrt{(1-r^2)(n-2)}}$
   - And thus, the t value associated with any correlation is $\frac{r}{\sqrt{(1-r^2)(n-2)}}$

Errors of Analysis and respecification (continued)

8. Setting scales of latent variables inappropriately.
   - particularly a problem with multiple group comparisons
9. Ignoring the start values or giving bad ones.
   - Supplying reasonable start values helps a great deal
10. Do different start values lead to different solutions?
11. Failure to recognize empirical underidentification
    - for some data sets, the model is underidentified even though there are enough parameters
    - Failure to separate measurement from structural portion of model
      - Use the two or four step procedure

Errors of Analysis and respecification (continued)

12. Estimating means and intercepts without showing measurement invariance
13. Analyzing parcels without checking if parcels are in fact factorially homogeneous.
    - Factorial Homogeneous Item Domains (FHID)
    - Homogenous Item Composites (HIC)
    - (but consider contradictory advice on parcels)

Errors of Interpretation

1. Looking only at indexes of overall fit
   - need to examine the residuals to see where there is misfit, even though overall model is fine
2. Interpreting good fit as meaning model is "proved".
   - consider alternative models
   - better able to reject alternatives
3. Interpreting good fit as meaning that the endogenous variables are strongly predicted.
   - What is predicted is the covariance of the variables, not the variables
   - Are the residual covariances small, not whether the error variance is small
4. Relying solely on statistical criterion in model evaluation
   - What can the model not explain
   - What are alternative models
   - What constraints does the model imply

Errors of interpretation (continued)

5. Relying too much on statistical tests
   - significance of particular path coefficients does not imply effect size or importance
   - Overall statistical fit (χ^2) is sensitive to model misfit as f(N)
6. Misinterpreting the standardized solution in multiple group problems
7. Failure to consider equivalent models
   - Why is this model better than equivalent models?
8. Failure to consider non-equivalent models
   - Why is this model better than other, non-equivalent models?
9. Reifying the latent variables
   - Latent variables are just models of observed data
   - "Factors are fictions"
10. Believing that naming a factor means it is understood

Errors of interpretation (continued)

11. Believing that a strong analytical method like SEM can overcome poor theory or poor design.
12. Failure to report enough so that you can be replicated
13. Interpreting estimates of large effects as evidence for "causality"

Final Comments

1. Theory First
   - What are the alternative theories?
   - Are there specific differences in the theories that are testable?
2. Measurement Model
   - Comparison of measurement models?
   - How many latent variables? How do you know?
   - Measurement Invariance?
3. Structural Model
   - Comparison of multiple models?
   - What happens if the arrows are reversed?
4. Theory Last
   - What do we know now that we did not know before?
   - What do we have shown is not correct?

Conclusion

1. Latent variable models are a powerful theoretical aid but do not replace theory
2. Nor do latent modeling algorithms replace the need for good scale development
3. Latent variable models are a supplement to the conventional regression models of observed scores.
4. Other latent models (not considered) include
    - item Response Theory
    - Latent Class Analysis
    - Latent Growth Curve analysis

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of Fit in Structural Equation Models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary Psychometrics* chapter 10, (pp. 275–340). New York: Routledge.

McDonald, R. & Marsh, H. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*(2), 247–255.