

November 8, 2007
SC705: Advanced Statistics
Instructor: Natasha Sarkisian
Class notes: Measurement Model

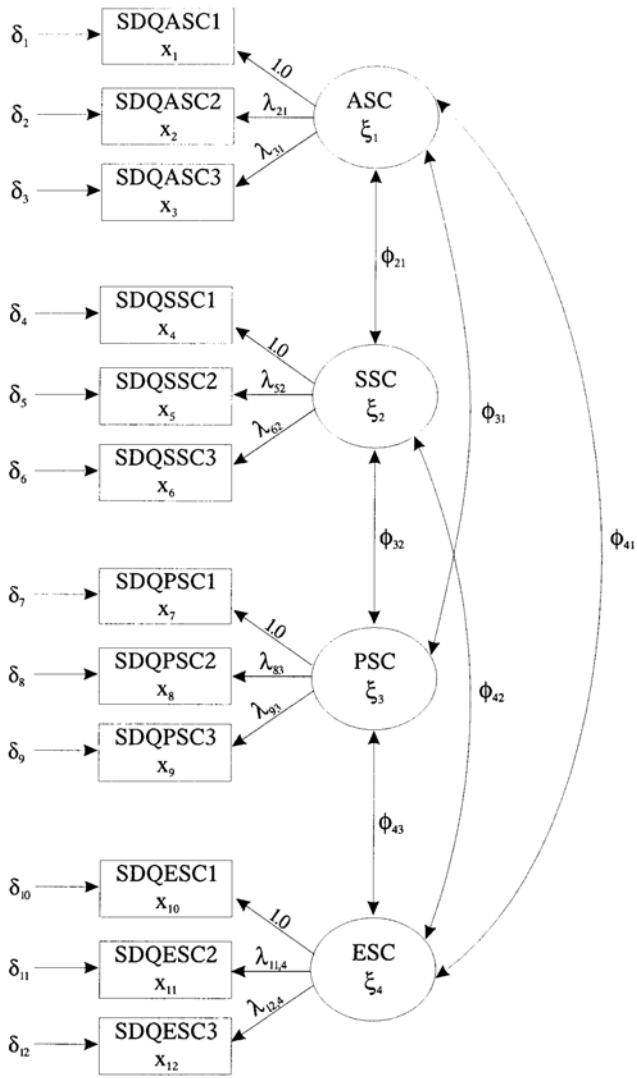
Simple path analysis ignores the possibility of measurement error – it assumes that each variable is measured perfectly. Measurement errors are less problematic for the endogenous variables – they become incorporated into the disturbance terms, so they don't affect the actual regression coefficients, although they do affect the proportion of variance explained (they would, however, affect the standardized regression coefficients because the total variance is affected).

The errors of measurement for exogenous variables will affect the regression coefficients, however, and therefore they are more problematic. Sometimes it is possible to incorporate measurement error based on known reliability for the measure, but it is also problematic if we are not very sure about that reliability estimate. One can do sensitivity analyses to see how various estimates of reliability affect the structural model results. But a better way to deal with measurement error is to have multiple indicators and to specify a measurement model, so for now, we'll focus on that.

While the structural model (path analysis portion of SEM) is based on regression, the measurement model is based on Confirmatory Factor Analysis (CFA). Note that there are some major differences between CFA and typical Exploratory Factor Analysis that many of you might be familiar with:

- EFA is atheoretical, CFA is based on theory
- In EFA, all indicators are related to all latent variables, only the strength of these correlations differs. In CFA, only some indicators are related to each of the latent variables; typically they do not overlap (i.e. each indicator is linked to only one latent variables, although there are exceptions).
- Related to the previous point, EFA models are always underidentified and therefore multiple solutions are possible; all are equally good, and the best solution is usually selected on the basis of producing a desirable structure of loadings (i.e. that each indicator has high loading for only one latent variable, and only weak loadings for the other variables). CFA models, in contrast, should be just-identified or overidentified.
- In EFA, the latent variables (factors) are usually assumed uncorrelated with each other (so called Principal Components Analysis). CFA, in contrast, is based on common factor analysis, and the factors are not considered orthogonal – they are, using the factor analysis terminology, oblique.
- Another difference between the PCA used in EFA and the common factor analysis used in CFA is in the utilization of variances and covariances – the PCA models redistribute all variance in the data across factor loadings, while the common factor analysis models partition the variance into common variance and residual variance.

When estimating a measurement model, we first need to specify the model based on theory (i.e. specify which indicators measure which latent variables).



(Diagram from: Byrne 1998, p.27)

Note that the latent variables are all connected with double-ended arrows: CFA model typically allows all latent variables to covary.

We also need to decide on the reference indicator – i.e. one path per indicator should be selected as a reference indicator and set to 1 to identify the scale for the latent variable. Alternatively, we could allow all paths to be estimated freely but set the variance of each latent variable to 1 (i.e., we can either have a latent variable that is measured in units of one of the indicators, or we can standardize it).

The issue of identification for the measurement model is similar to that for the path model – we need to count the number of variances and covariances $p*(p+1)/2$, and we need to compare that to the number of estimated parameters.

Measurement model using LISREL

Let's estimate a measurement model with two latent variables, academic ability measured by two test scores, X1 and X2, and peer popularity, measured by choices of seating, choices during schoolwork, and playground choices (X3, X4, and X5). The number of cases N=100. Here's the correlation matrix:

	X1	X2	X3	X4	X5
X1	1.00				
X2	.28	1.00			
X3	.16	.10	1.00		
X4	.03	.04	.52	1.00	
X5	.15	.05	.59	.36	1.00

Formulas and equations:

$$X = \Lambda_x \xi + \delta$$

$$x_1 = 1*\xi_1 + 0*\xi_2 + \delta_1$$

$$x_2 = \lambda_{21}*\xi_1 + 0*\xi_2 + \delta_2$$

$$x_3 = 0*\xi_1 + 1*\xi_2 + \delta_3$$

$$x_4 = 0*\xi_1 + \lambda_{42}*\xi_2 + \delta_4$$

$$x_5 = 0*\xi_1 + \lambda_{52}*\xi_2 + \delta_5$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ 0 & 1 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix}$$

Other matrices:

Θ_δ (5x5 matrix of variances and covariances of measurement errors δ) – measurement errors vary but they do not covary. Therefore, we want to have this matrix to be diagonal and free – so LISREL default is what we need.

Φ (2x2 matrix of variances and covariances of exogenous variables ξ) – in a pure measurement model, we allow all latent variables to covary (all have to be connected by double-headed arrows). Therefore, all three elements of this matrix are estimated -- LISREL default (symmetric, free) is what we need.

DA NI=5 NO=100 MA=KM
 LA
 SCORE1 SCORE2 SEAT SCHOOL PLAY
 KM SY

1.00
 .28 1.00
 .16 .10 1.00
 .03 .04 .52 1.00
 .15 .05 .59 .36 1.00

MO NX=5 NK=2 LX=FU, FI

LK

ABILITY PEER

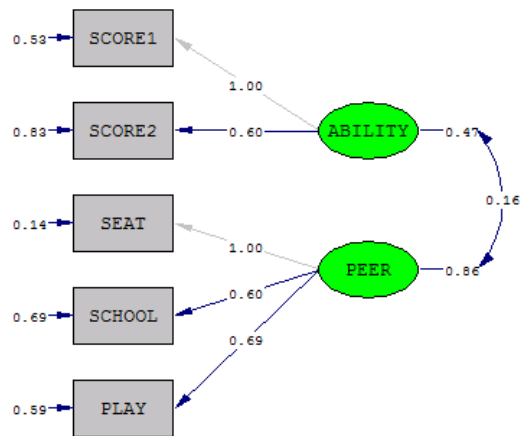
FR LX 2 1 LX 4 2 LX 5 2

VA 1.0 LX 1 1 LX 3 2

PD

OU

Output:



Chi-Square=1.08, df=4, P-value=0.89774, RMSEA=0.000

DA NI=5 NO=100 MA=KM

Number of Input Variables 5
 Number of Y - Variables 0
 Number of X - Variables 5
 Number of ETA - Variables 0
 Number of KSI - Variables 2
 Number of Observations 100

DA NI=5 NO=100 MA=KM

Correlation Matrix

	SCORE1	SCORE2	SEAT	SCHOOL	PLAY
SCORE1	1.00				
SCORE2	0.28	1.00			
SEAT	0.16	0.10	1.00		
SCHOOL	0.03	0.04	0.52	1.00	
PLAY	0.15	0.05	0.59	0.36	1.00

DA NI=5 NO=100 MA=KM

Parameter Specifications

LAMBDA-X

	ABILITY	PEER
	-----	-----
SCORE1	0	0
SCORE2	1	0
SEAT	0	0
SCHOOL	0	2
PLAY	0	3

PHI

	ABILITY	PEER			
	-----	-----			
ABILITY	4				
PEER	5	6			
THETA-DELTA					
	SCORE1	SCORE2	SEAT	SCHOOL	PLAY
	-----	-----	-----	-----	-----
	7	8	9	10	11

DA NI=5 NO=100 MA=KM

Number of Iterations = 10

LISREL Estimates (Maximum Likelihood)

LAMBDA-X

	ABILITY	PEER
	-----	-----
SCORE1	1.00	- -
SCORE2	0.60 (0.63) 0.95	- -
SEAT	- -	1.00
SCHOOL	- -	0.60 (0.14) 4.24
PLAY	- -	0.69 (0.15) 4.52

PHI

	ABILITY	PEER
	-----	-----
ABILITY	0.47 (0.50) 0.92	
PEER	0.16 (0.10) 1.57	0.86 (0.21) 4.03

THETA-DELTA

SCORE1	SCORE2	SEAT	SCHOOL	PLAY
0.53	0.83	0.14	0.69	0.59
(0.50)	(0.21)	(0.16)	(0.11)	(0.11)
1.08	3.90	0.85	6.02	5.23

Squared Multiple Correlations for X - Variables

SCORE1	SCORE2	SEAT	SCHOOL	PLAY
0.47	0.17	0.86	0.31	0.41

Goodness of Fit Statistics

Degrees of Freedom = 4
 Minimum Fit Function Chi-Square = 1.09 (P = 0.90)
 Normal Theory Weighted Least Squares Chi-Square = 1.08 (P = 0.90)
 Estimated Non-centrality Parameter (NCP) = 0.0
 90 Percent Confidence Interval for NCP = (0.0 ; 1.72)

Minimum Fit Function Value = 0.011
 Population Discrepancy Function Value (F0) = 0.0
 90 Percent Confidence Interval for F0 = (0.0 ; 0.017)
 Root Mean Square Error of Approximation (RMSEA) = 0.0
 90 Percent Confidence Interval for RMSEA = (0.0 ; 0.066)
 P-Value for Test of Close Fit (RMSEA < 0.05) = 0.93
 Expected Cross-Validation Index (ECVI) = 0.26
 90 Percent Confidence Interval for ECVI = (0.26 ; 0.28)
 ECVI for Saturated Model = 0.30
 ECVI for Independence Model = 0.99

Chi-Square for Independence Model with 10 Degrees of Freedom = 88.07

Independence AIC = 98.07
 Model AIC = 23.08
 Saturated AIC = 30.00
 Independence CAIC = 116.10
 Model CAIC = 62.73
 Saturated CAIC = 84.08

Normed Fit Index (NFI) = 0.99
 Non-Normed Fit Index (NNFI) = 1.09
 Parsimony Normed Fit Index (PNFI) = 0.40
 Comparative Fit Index (CFI) = 1.00
 Incremental Fit Index (IFI) = 1.03
 Relative Fit Index (RFI) = 0.97

Critical N (CN) = 1208.77

Root Mean Square Residual (RMR) = 0.021
 Standardized RMR = 0.021
 Goodness of Fit Index (GFI) = 1.00
 Adjusted Goodness of Fit Index (AGFI) = 0.98
 Parsimony Goodness of Fit Index (PGFI) = 0.27