Implementing An SEM Analysis

I. Measurement and Structural Models

A. Confirmatory Factor Analysis

The model of classical, exploratory factor analysis is

$$\boldsymbol{x} = \boldsymbol{\Lambda}_{\boldsymbol{x}}\boldsymbol{\xi} + \boldsymbol{\delta} , \qquad \textbf{(1)}$$

where:

x = observed variable

Λ = loadings that the x have on the factors

ξ = latent factors (typically there are fewer ξ than x variables)

δ = residuals (errors); the δ are uncorrelated with the ξ

If Φ is the correlation matrix and Θ is the diagonal residual matrix, the factor model can be cast as follows. The covariance Σ of the x variables is

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Phi\Lambda}^{\boldsymbol{T}} + \boldsymbol{\Theta} \qquad \textbf{(2)}$$

Problems with classical factor analysis:

- number of factors unknown
- loading pattern unknown
- **indeterminacy problem**: when there is more than 1 factor, there is the following indeterminacy in (2): any nonsingular transformation of ξ changes Λ and in general also Φ, but not Σ (which we wish to explain). The usual solution, that is, setting $\Phi = I$ and $\Lambda^T\Lambda$ to be diagonal, leads to an arbitrary set of factors. Rotation may yield a more meaningful interpretation.

In contrast, in **confirmatory factor analysis** there is knowledge about (1) number of factors, (2) loading pattern, (3) factor intercorrelations etc.

Illustration: consider a problem where 5 variables are supposed to load on two factors. This problem can be expressed in terms of (1) as follows:

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & \lambda_{42} \\ 0 & \lambda_{52} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \end{bmatrix} . \tag{3}$$

This is equivalent to the following five structural equations:

$$x_1 = \lambda_{11}\xi_1 + \delta_1 ,$$

$$x_2 = \lambda_{21}\xi_1 + \delta_2 ,$$

$$x_3 = \lambda_{31}\xi_1 + \delta_3 ,$$

$$x_4 = \lambda_{42}\xi_2 + \delta_4 ,$$

$$x_5 = \lambda_{52}\xi_2 + \delta_5 .$$

Data Example (Confirmatory1A)

Data: nine biological variables, supposed to be grouped in the three factors, **Body Size, Reproductive Fitness,** and **Food**. Each factor has three **indicators**.

Model 1 a: Fits fine, yet, the χ^2 is somewhat high (suggests rejecting model); modification index suggests loosening factor structure (let one variable, patch size load on two factors)

Model 1 b: Fits very well, in all respects

B. Structural Model

1. Example: Structural 1A.

2. Nested Models (Structural 1bA)

II. Model Identification

Identification Problems

Example: a simple identification problem

Consider the factor, ξ, that has the two indicators, x_1 and x_2. The factor model for this situation is

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_x \boldsymbol{\Phi} \boldsymbol{\Lambda}_x^T + \boldsymbol{\Theta}_\delta \tag{9}$$

the correlation between the two indicator variables is **r**. (1) can equivalently be recast as follows:

$$\begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 \end{bmatrix} + \begin{bmatrix} \theta_{\delta_{11}} & 0 \\ 0 & \theta_{\delta_{22}} \end{bmatrix} \tag{10}$$

Performing the matrix multiplication and addition on the right hand side of (2) yields

$$\begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix} = \begin{bmatrix} \lambda_1^2 + \theta_{\delta_{11}} & \lambda_1\lambda_2 \\ \lambda_1\lambda_2 & \lambda_2^2 + \theta_{\delta_{22}} \end{bmatrix} \tag{11}$$

Equation (3) implies the following system of three equations:

$$\begin{aligned} 1 &= \lambda_1^2 + \theta_{\delta_{11}} \\ 1 &= \lambda_2^2 + \theta_{\delta_{22}} \\ r^2 &= \lambda_1\lambda_2 \end{aligned} \tag{12}$$

All parameters on the right hand sides of (4) are unknown. Thus, we have an equation system with three equations and four unknowns ⇨ **this system is not identified**

Options for Solutions

- fixing of parameters ⇨ still three equations but three or fewer unknowns
- setting parameters equal to each other

The t-rule

(necessary but not sufficient condition for identifiability)

$$t \leq \frac{1}{2}(p + q)(p + q + 1) \ , \tag{13}$$

where the product on the right hand side of the equation gives the number of **non-redundant elements in** Σ; p+q is the number of observed variables, and t is the number of true parameters (unknowns) (for more detail see, e.g., Bollen, 1989)

Models are said to be:

Under Identified

Just Identified

Over Identified

III Model Fit (Evaluate and Respecify Your Model)

I. Diagnostics

A. Chi-Square and fit indices

1. Overall model fit

a. Sample size dependent

b. Rule of thumb for reasonable fit

$$\mathbf{X^2 \leq 2(\mathit{df})}$$

2. Nested models

a. More reliable

$$\mathbf{X^2(\mathit{model}1) - X^2(\mathit{model}2)}$$

df (model 1) - df (model 2)

3. GFI (Goodness of Fit Index)

B. Residuals

1. Fitted residuals

$$\boldsymbol{S - \Sigma}$$

a. Expressed in true metric

b. Value represents the sum of model effects

2. Standardized residuals
    a. Standardized metric
    b. Estimates the number of standard deviations the observed residuals are away from the zero residuals that would be provided by a perfectly fitting model.
    c. Random error expected within 2 standard deviations of 0.0
    d. Q-Q plots
        1. Expect residuals to be normally distributed.
        2. X-axis provides value of standardized residual
        3. Y-axis provides corresponding value from standard normal curve.
        4. Non-linear plot indicates lack of normality
        5. Line if normally distributed
            >45 normal dist., less variability
            < 45 normal dist. but more variability

C. T-values
    1. Coefficient estimate / Standard Error
    2. Provides the number of sampling distribution standard deviations the estimate is away from zero.
    3. Tests null hypothesis that the true parameter is zero.
        +- 2 standard deviations (95%)

D. Correlations of estimates - Examines correlations among estimates

E. Partial derivatives of the fit function
    1. Large positive or negative partial derivatives indicate that increasing or decreasing a particular coefficient's value is likely to substantially improve the model fit and chi-square.

2. Should be zero or model has not achieved maximization of the likelihood function.
3. Modification indices
    a. take into account the metric of the parameter

F. The Standardized Model

1. LISREL reports a rescaling of the maximum likelihood estimates such that all of the concepts are given a variance of 1.0 but the indicators remain in their original scales.
2. Scales Betta and Gamma between 0.0 and 1.0
3. Since the total variance of each Eta is 1.0 the Psi variance corresponding to each Eta gives the proportion of error variance in the prediction of Eta.
4. 1-Psi gives the variance explained in Eta.
5. Standardized solution can indicate estimation problems.
    a. Elements of Betta, Gamma or Psi > 1.0
    b. Observed variances other than 1.0 of Phi or Eta indicate problems

IV. Philosophy of Model Respecification

A. Exploratory and Confirmatory analysis

B. Model Over Fitting

V. Writing A LISREL Program

Application of LISREL

(Adapted from LISREL8 Manual, p. 99)

Step	Relevant Line in Command File
1. Title (optional)	
2. Data Description: number of variables, number of observations, number of groups, names for variables, type of information for input (raw data, covariance matrix, correlation matrix ...)	DA, LA
3. Data Input: from file, diskette etc.	LM, RA, KM, ...
4. Selecting and Reordering of variables	SE
5. Specification of General Model Characteristics: ▸ number of exogenous and endogenous latent variables ▸ number of exogenous and endogenous observed variables ▸ characteristics of matrices (symmetrical, diagonal, free, fixed ...)	MO
6. Specify Changes for Matrix Elements: ▸ free elements in fixed matrices ▸ fix elements in free matrices ▸ set parameters equal	LY, FR, FI, EQ ...
7. Specify Values ▸ for fixed parameters ▸ starting values for iterations (if needed)	VA, ST
8. Specify what Output Information is Needed	OU

Table 1.2
Parameter Matrices in LISREL: Their Possible Forms and Default Values

Name	Math Symbol	LISREL Name	Order	Possible Forms	Default Form	Default Mode
LAMBDA-Y	$\mathbf{\Lambda}_y$	LY	NY × NE	ID,IZ,ZI,DI,FU	FU	FI
LAMBDA-X	$\mathbf{\Lambda}_x$	LX	NX × NK	ID,IZ,ZI,DI,FU	FU	FI
BETA	$\mathbf{B}$	BE	NE × NE	ZE,SD,FU	ZE	FI
GAMMA	$\mathbf{\Gamma}$	GA	NE × NK	ID,IZ,ZI,DI,FU	FU	FR
PHI	$\mathbf{\Phi}$	PH	NK × NK	ID,DI,SY,ST	SY	FR
PSI	$\mathbf{\Psi}$	PS	NE × NE	ZE,DI,SY	SY	FR
THETA-EPSILON	$\mathbf{\Theta}_\epsilon$	TE	NY × NY	ZE,DI,SY	DI	FR
THETA-DELTA	$\mathbf{\Theta}_\delta$	TD	NX × NX	ZE,DI,SY	DI	FR

The meaning of the possible form values is as follows:

Th