PSY 9555A – SEM: Oct 2 – Steps in SEM Continued, Focus on CFA

Identification (from Kline, Ch. 6)

- Degrees of freedom >= 0
- Latent variables must be assigned a metric

- Understanding identification:

$a + b = 6$ (unidentified – more than one solution)

$a + b = 6$
$3a + 3b = 18$ (unidentified, linear dependence)

$a + b = 6$
$2a + b = 10$ ($a = 4$; $b = 2$; one unique solution, just identified)

$a + b = 6$
$2a + b = 10$
$3a + b = 12$ (3 observations; 2 parameters; no unique solution)

(continued next slide)

Identification (from Kline, Ch. 6)

a + b = 6
2a + b = 10
3a + b = 12 (3 observations; 2 parameters; no unique solution)

- Let's impose statistical criterion that leads to unique estimates

- Find values for a and b that will give us answers closest to 6, 10, 12
  - i.e., minimum total squared deviations

- Note that these unique estimates may not reproduce the observations 6, 10, 12 exactly

- Try the values a = 3.0 and b = 3.3
  - 3.0 + 3.3 = 6.3 (6.0 - 6.3)^2 = .09
  - 2(3.0) + 3.3 = 9.3 (10 – 9.3)^2 = .49
  - 3(3.0) + 3.3 = 12.3 (12 – 12.3)^2 = .09
  - .09 + .49 + .09 = .67 (this is the best we can get)

Additional Rules for Identification (from Kline, Ch. 6)

- Recursive structural models identified
- Nonrecursive models more complicated (details p. 132-137
- CFA models:
    - Three-indicator rule (one factor)
    - Two-indicator rule (two or more factors)
    - A bit more complicated with correlated residuals and/or cross loadings
    - See p. 140-142

Estimation

- We start with the data we have collected/observed:
  - a **sample variance-covariance matrix (S)**

- We can convert the information from a variance-covariance matrix into model parameters using mathematical rules
  - Sewall Wright – path analysis
  - Simple examples with basic regression models
  - We will see specific tracing rules later
  - Gets too complicated with complex models
    - No unique solution
    - Use estimation criterion (e.g., Maximum likelihood)

- Objective in SEM:
  - to obtain model parameter estimates that produce a **predicted var-cov matrix Σ** that resembles the **sample var-cov matrix (S)** as closely as possible

Estimation

- Recall that an over-identified model does not fit the data perfectly
  - You have decided that some parameter values are negligible
  - These have been fixed at zero (e.g., negligible cross loadings)
  - As a result your model is less complex with fewer parameters and more dfs.

- Fitting function to minimize the difference between the two matrices.

$$F_{ML} = ln|S| - ln|\Sigma| + trace[(S)(\Sigma^{-1})] - p$$

- p = the number of observed variables (order of the input matrix
- ML estimation: parameters that maximize the probability of observing the same var-cov matrix S if the data were collected from the same population again

Model Evaluation: Overview

How should we evaluate the model?

- **Model fit**: how close is the predicted/reproduced var-cov matrix Σ to sample matrix S?
  - Model test statistics (e.g., chi-square)
    - Ask whether deviations of Σ from S are due to sampling error
    - Backwardness of test – proving the null?
  - Approximate fit indexes
    - Values that express model-data fit

- Are hypothesized model parameters statistically significant and meaningful?

- Consider also predictive power of the model

- Consider also the reliability of the measurement model

- Residuals can help pinpoint specific areas of misfit in model

- Have you compared the fit of your model to other models?