Structural Equation Modeling 3

Psy 524

Andrew Ainsworth

Model Identification

- Only identified models can be estimated in SEM
- A model is said to be identified if there is a unique solution for every estimate
  - $Y = 10$
  - $Y = \alpha + \beta$
  - One of theme needs to fixed in order for there to be a unique solutions
  - Bottom line: some parts of a model need to be fixed in order for the model to be identified
  - This is especially true for complex models

Model Identification: Step 1

- Overidentification
  - More data points than parameters
  - This is a necessary but not sufficient condition for identification
- Just Identified
  - Data points equal number of parameters
  - Can not test model adequacy
- Underidentified
  - There are more parameters than data points
  - Can't do anything; no estimation
  - Parameters can be fixed to free DFs

Model Identification: Step 2a

- The factors in the measurement model need to be given a scale (latent factors don’t exist)
  - You can either standardize the factor by setting the variance to 1 (perfectly fine)
  - Or you can set the regression coefficient predicting one of the indicators to 1; this sets the scale to be equal to that of the indicator; best if it is a marker indicator
  - If the factor is exogenous either is fine
  - If the factor is endogenous most set the factor to 1

Model Identification: Step 2b

- Factors are identified:
  - If there is only one factor then:
    - at least 3 indicators with non-zero loadings
    - no correlated errors
  - If there is more than one factor and 3 indicators with non-zero loadings per factor then:
    - No correlated errors
    - No complex loadings
    - Factors covary

Model Identification: Step 2b

- Factors are identified:
  - If there is more than one factor and a factor with only 2 indicators with non-zero loadings per factor then:
    - No correlated errors
    - No complex loadings
    - None of the variances or covariances among factors are zero

Model Identification: Step 3

- Relationships among the factors should either be orthogonal or recursive to be identified
  - Recursive models have no feedback loops or correlated disturbances
  - Non-recursive models contain feedback loops or correlated disturbances
  - Non-recursive models can be identified but they are difficult

Model Estimation

- After model specification:
  - The population parameter are estimated with the goal of minimizing the difference between the estimated covariance matrix and the sample covariance matrix
  - This goal is accomplished by minimizing the Q function:

$$Q = (s - \sigma(\Theta))'W(s - \sigma(\Theta))$$

    - Where s is a vectorized sample covariance marix, σ is a vectorized estimated matrix and Θ indicates that σ is estimated from the parameters and W is a weight matrix

Model Estimation

- In factor analysis we compared the covariance matrix and the reproduced covariance matrix to assess fit
- In SEM this is extended into an actual test
- If the W matrix is selected correctly than (N – 1) * Q is Chi-square distributed
- The difficult part of estimation is choosing the correct W matrix

Model Estimation Procedures

- Model Estimation Procedures differ in the choice of the weight matrix
- Roughly 6 widely used procedures
  - ULS (unweighted least squares)
  - GLS (generalized least squares)
  - ML (maximum likelihood)
  - EDT (elliptical distribution theory)
  - ADF (asymptotically distribution free)
  - Satorra-Bentler Scaled Chi-Square (corrected ML estimate for non-normality of data)

Model Estimation Procedures

Estimation Method	Function Minimized	Interpretation of W, the Weight Matrix
Unweighted Least Squares[a] (ULS)	$F_{\text{ULS}} = \frac{1}{2}\operatorname{tr}[(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\Theta}))^2]$	$\mathbf{W} = \mathbf{I}$, the identity matrix
Generalized Least Squares (GLS)	$F_{\text{GLS}} = \frac{1}{2}\operatorname{tr}\{[(\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\Theta}))]\mathbf{W}^{-1}\}^2$	$\mathbf{W} = \mathbf{S}$. $\mathbf{W}$ is any consistent estimator of $\boldsymbol{\Sigma}$. Often the sample covariance matrix, $\mathbf{S}$, is used
Maximum Likelihood (ML)	$F_{\text{ML}} = \log\lvert\boldsymbol{\Sigma}\rvert - \log\lvert\mathbf{S}\rvert + \operatorname{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \rho$	$\mathbf{W} = \boldsymbol{\Sigma}^{-1}$, the inverse of the estimated population covariance matrix. The number of measured variables is ρ.
Elliptical Distribution Theory (EDT)	$F_{\text{EDT}} = \frac{1}{2}(\kappa + 1)^{-1}\operatorname{tr}\{[\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\Theta})]\mathbf{W}^{-1}\}^2 - \delta\{\operatorname{tr}[\mathbf{S} - \boldsymbol{\Sigma}(\boldsymbol{\Theta})]\mathbf{W}^{-1}\}^2$	$\mathbf{W}$ = any consistent estimator of $\boldsymbol{\Sigma}$. κ and δ are measures of kurtosis
Asymptotically Distribution Free (ADF)	$F_{\text{ADF}} = [\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\Theta})]'\mathbf{W}^{-1}[\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\Theta})]$	$\mathbf{W}$ has elements, $\mathbf{w}_{ijkl} = \boldsymbol{\sigma}_{ijkl} - \boldsymbol{\sigma}_{ij}\boldsymbol{\sigma}_{kl}$ ($\boldsymbol{\sigma}_{ijkl}$ is the kurtosis, $\boldsymbol{\sigma}_{ij}$ is the covariance)

Assessing Model Fit

- How well does the model fit the data?
- This can be answered by the Chi-square statistic but this test has many problems
  - It is sample size dependent, so with large sample sizes trivial differences will be significant
  - There are basic underlying assumptions are violated the probabilities are inaccurate

Assessing Model Fit

- Fit indices
  - Read through the book and you’ll find that there are tons of fit indices and for everyone in the book there are 5 – 10 not mentioned
  - Which do you choose?
  - Different researchers have different preferences and different cutoff criterion for each index
  - We will just focus on two fit indices
    - CFI
    - RMSEA

Assessing Model Fit

- Assessing Model FitFit Indices
  - Comparative Fit Index (CFI) – compares the proposed model to an independence model (where nothing is related)

$$CFI = 1 - \frac{t_{\text{est.model}}}{t_{\text{indep.model}}}$$

$$\text{where } t_{\text{indep.model}} = c^2_{\text{indep.model}} - df_{\text{indep.model}}$$

$$and \;\; t_{\text{est.model}} = c^2_{\text{est.model}} - df_{\text{est.model}}$$

Assessing Model Fit

- Root Mean Square Error of Approximation
  - Compares the estimated model to a saturated or perfect model

$$RMSEA = \sqrt{\frac{\widehat{F}_0}{df_{\text{model}}}}$$

$$where\ \widehat{F}_0 = \frac{c^2_{\text{model}} - df_{\text{model}}}{\text{N}}\ or\ \text{0 whichever is smaller and positive}$$

Model Modification

- Chi-square difference test
  - Nested models (models that are subsets of each other) can be tested for improvement by taking the difference between the two chi-square values and testing it at a DF that is equal to the difference between the DFs in the two models (more on this in lab)

Model Modification

- Langrange Multiplier test
  - This tests fixed paths (usually fixed to zero or left out) to see if including the path would improve the model
  - If path is included would it give you better fit
  - It does this both univariately and multivariately
- Wald Test
  - This tests free paths to see if removing them would hurt the model
  - Leads to a more parsimonious model