Confirmatory Factor Analysis: Identification and estimation

Psychology 588: Covariance structure and factor models

Identification

- Covariance structure of measurement model:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Lambda}_x \boldsymbol{\Phi} \boldsymbol{\Lambda}_x' + \boldsymbol{\Theta}_\delta$$

where we can impose various kinds of constraints (zero, equality, etc.) on selective entries of $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Phi}$; and free selective off-diagonal elements in $\boldsymbol{\Theta}_\delta$, provided that the resulting model is identifiable

- Though CFA has different parameter sets than the path model only with observed variables, the basic ideas hold:
  - A model is identified <u>if and only if</u> every single free parameter has a unique solution
  - A parameter is identified if it can be written as a function (or functions) of the data (variances/covariances) that is unique given an optimization function (e.g., ML)

Single-factor model

- As a simplistic case, consider only one factor measured by 2 or 3 indicators, with no correlated errors (see detailed algebraic derivation in pp. 240-242)
  - 2 indicators --- one more constraint (in addition to the scaling constraint) needed for just identification; e.g., known reliability, tau-equivalent measures
  - 3 indicators --- just identifiable with the scaling constraint
  - Multiple orthogonal factors with a uni-factorial loading pattern can also be considered as separate single-factor models for identification, except that any non-zero covariance between indicators of different factors will also contribute to misfit

t-rule

- With no knowledge whatsoever about parameters to constrain, the factor model (for both EFA and CFA) has free parameters as many as:
  - $q \times n$ factor loadings in $\boldsymbol{\Lambda}_x$
  - $n(n+1)/2$ nonredundant elements in $\boldsymbol{\Phi}$
  - $q(q+1)/2$ nonredundant elements in $\boldsymbol{\Theta}_\delta$
- Thus, constraints (including the scaling constraints) needed so as to satisfy

$$t \leq q\left(q+1\right)/2$$

- As before, t-rule is necessary, not sufficient

Three indicators per factor

- Three indicators or more per factor ($n \geq 2$) are sufficient if
  - All indicators are uni-factorial
  - Error terms are not correlated (diagonal $\boldsymbol{\Theta}_{\delta}$)
  - $\boldsymbol{\Phi}$ needs no constraints (such as orthogonality)
- Identification established since
  - λ_{ij}, ϕ_{jj}, δ_{ii} are identifiable given 3 or more indicators for each ξ_j, according to the 3-indicator condition for the single-factor model

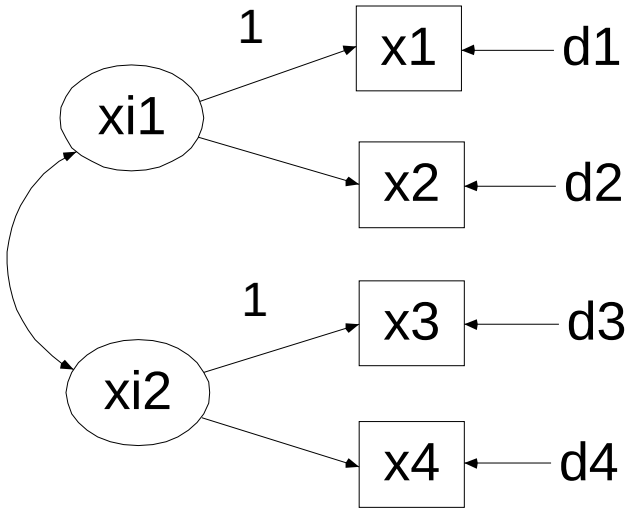
- Remaining parameters $\phi_{jk},\ j \neq k$ are identifiable by covariance between the indicators that determine the scale of ξ_j and ξ_k

$$x_i = \xi_j + \delta_i, \quad x_h = \xi_k + \delta_h \quad \rightarrow \quad \phi_{jk} = \sigma_{ih}$$

- Additional covariances between indicators for different factors will make the model overidentified, requiring an optimization for estimation --- general for any free parameters

Two indicators per factor

- Two indicators per factor ($n \geq 2$) are also <u>sufficient</u> if
  - all indicators are uni-factorial
  - diagonal $\boldsymbol{\Theta}_\delta$
  - $\phi_{jk} \neq 0,\ j \neq k$



- Consider a simple case of $n = 2$:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \phi_{11} + \mathrm{var}(\delta_1) & & & \\ \lambda_{21}\phi_{11} & \lambda_{21}^2\phi_{11} + \mathrm{var}(\delta_2) & & \\ \phi_{12} & \lambda_{21}\phi_{12} & \phi_{22} + \mathrm{var}(\delta_3) & \\ \lambda_{42}\phi_{12} & \lambda_{21}\lambda_{42}\phi_{12} & \lambda_{42}\phi_{22} & \lambda_{42}^2\phi_{22} + \mathrm{var}(\delta_4) \end{bmatrix}$$

- All parameters in the $n = 2$ case are identified, i.e., can be written as functions of data variances and covariances (see p. 245)

- Further relaxed <u>sufficient</u> condition:
  - All indicators are uni-factorial
  - Diagonal $\boldsymbol{\Theta}_{\sigma}$
  - Each row of $\boldsymbol{\Phi}$ has at least one nonzero off-diagonal element --- meaning that for every ξ_j, there is at least one other ξ_k such that $\phi_{jk} \neq 0$, and so the identifiability of the $n = 2$ case holds per the j-k pair, and any additional $\phi_{jk'} \neq 0$ will make the model overidentified (and so will more indicators per factor)

Identification in practice

- None of the considered rules is unconditionally necessary and sufficient; and we often want to test a model with factorially-complex indicators, some correlated errors, and sometimes all orthogonal factors (e.g., Big 5 personality factors)

- Difficult to derive a general algebraic rule for a wide range of factor models (also true for the path model part since the necessary and sufficient rank condition only holds with free off-diagonal entries in $\boldsymbol{\Psi}$)

- A rule of thumb is 3 indicators per factor --- not necessarily sufficient and so an empirical test is useful for each specific model

Empirical test for identification (of general models)

- Local vs. global identification
  - $\boldsymbol{\theta}$ is globally identified if no $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ exist such that $\boldsymbol{\Sigma}(\boldsymbol{\theta}_1) = \boldsymbol{\Sigma}(\boldsymbol{\theta}_2)$ unless $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, while local identification means that $\boldsymbol{\theta}$ is unique only near a specific $\boldsymbol{\theta}_1$ --- global identifiability implies local identifiability, but not the reverse
  - Local identification can be numerically tested, while there is no feasible global test --- local test is useful to rule out unidentifiable models; and more importantly, once we have an optimal solution (e.g., a converged ML solution) with its local test satisfied, we can say “the model is uniquely identified within the nearby search space around this optimal solution”

Two local identification tests

- Wald's rank rule:

$$\text{rank}\left(\partial\boldsymbol{\sigma}/\partial\boldsymbol{\theta}\right) \text{ evaluated at } \boldsymbol{\theta}_1 = t$$

where $\boldsymbol{\sigma}$ is a vector of all distinctive elements of $\boldsymbol{\Sigma}$ --- implying that change of $\boldsymbol{\Sigma}$ due to a slight change of an element of $\boldsymbol{\theta}$ is not the same as what would result from a slight change of another element

- If the information matrix $\mathbf{G}$ for $\boldsymbol{\theta}$ evaluated at $\boldsymbol{\theta}_1$ is not singular (i.e., its ACOV matrix exists), $\boldsymbol{\theta}_1$ is locally unique

$$\mathbf{G} = \frac{N-1}{2} E\left(\frac{\partial^2 f\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) = \left(\text{acov}\left(\hat{\boldsymbol{\theta}}\right)\right)^{-1}$$

- Either test could be misleading since the local test is numerical, depending on fallible estimation (due to, e.g., sampling, measurement errors) --- thus, a locally unidentified model at a particular $\hat{\boldsymbol{\theta}}$ may be identifiable at the population value, and vice versa

- Consider the single-factor model with 3 indicators (just identified with $\lambda_{11} = 1$) for cases of (1) when $\lambda_{21} = 0$ in population and both s_{12} and s_{32} are somewhat different than 0; (2) when the population value of $\boldsymbol{\theta}$ is identified while the data are "measured" to numerically suggest an unidentified model

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \end{bmatrix} \xi_1 + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}, \qquad \lambda_{21} = \frac{\mathrm{cov}(x_2, x_3)}{\mathrm{cov}(x_1, x_3)}$$

Recommendation for identification

- Rule out quickly any model that doesn't pass t-rule
- See if any of the sufficient conditions applies
- Whenever feasible, see if all parameters can be written as functions of the data (see political demo example, pp. 251-254)
- Given an optimal solution (i.e., locally identified), see if there is any excessive sampling error for a parameter, which would suggest something wrong with the solution; or given a locally unidentified model, see if any reason to doubt about the numerical result
  - Apply multiple random starts to see if they yield the same result
  - Use random sub-samples (e.g., split halves or bootstrap samples) to empirically see sampling fluctuation

Estimation

- All of the ML, ULS and GLS fitting functions have the same form as we learned for observed-variables-only model, except that parameter sets are different:

  $\boldsymbol{\Lambda}_x, \boldsymbol{\Phi}, \boldsymbol{\Theta}_\delta$ instead of $\mathbf{B}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}, \boldsymbol{\Psi}$

- Non-convergence --- may occur due to too stringent stopping criterion for the optimal fit, too early stop of iteration, bad start (cf. “rational” start), degeneracy (due to poor specification), oscillation, etc.

  Local minima/maxima --- multiple random starts needed (perhaps also with different “seeds”)

- Improper solution (e.g., negative variance)