General structural model – Part 1: Covariance structure and identification
Latent variables

• Interchangeably used:
  Ø constructs --- substantively defined concepts
  Ø common factors or components --- something uncorrelated with measurement errors (cf. case of cause indicators)
  Ø true scores in measurement theory

• Generality vs. specificity --- latent variables are operationally defined to imply a set of indicators (but the reverse is not true), i.e., more generalizable than what’s intended by each indicator

• Good indicators
  Ø represent distinctive aspects of a defined concept
  Ø reasonably inter-correlated (i.e., internally reliable) since they possess some common characteristics
  Ø factorially simple; ideally uni-factorial
General structural models

• General structural equation models combine:
  - measurement models that operationally define a set of theoretical concepts, allowing for fallible measurement, with
  - a path model that explains complex relationships among a set of (mostly) latent variables

• All forms of models considered so far are special cases of GM (e.g., path models only with observed variables, CFA, MIMIC, etc.) --- rules and principles considered so far apply only to a part of GM, and so we will consider ways of combining them as well
• The general SE model subsumes most multivariate (causal) models, including MANOVA, discriminant function analysis, multivariate regression, canonical correlation analysis (CCA)

• The equivalence is about the model form, not about what optimization function is used
  
  ➢ The above listed models can be considered as special cases of CCA, which uses the OLS like loss function (i.e., maximum accounted variance of manifest variables)

  ➢ This type of variance maximization tends to fit more of variances than covariances --- cf. principal component analysis vs. common factor analysis

  ➢ Component model type of approach to SEM --- Partial Least Squares (Wold, 1974, European Economic Review, 67-86) and Generalized Structured Component Analysis (Hwang & Takane, 2004, Psychometrika, 81-99)
The model form

• By combining measurement models and a path model, we mean:

\[ \eta = B\eta + \Gamma \xi + \zeta = (I - B)^{-1} (\Gamma \xi + \zeta) \]

\[ y = \Lambda_y \eta + \epsilon = \Lambda_y (I - B)^{-1} (\Gamma \xi + \zeta) + \epsilon \]

\[ x = \Lambda_x \xi + \delta \]

where we have 4 parameter matrices of “regression weights” \((B, \Gamma, \Lambda_y, \Lambda_x)\) and 4 covariance matrices for exogenous latent variables \((\Phi, \Psi, \Theta_\epsilon, \Theta_\delta)\) --- any cause indicators included in \(\xi\)

• Based on this structural representation, we build covariance structure of observed variables \((y \text{ and } x)\)
Implied covariance matrix

\[
\Sigma(\theta) = \begin{bmatrix}
\Sigma_{yy}(\theta) & \Sigma_{yx}(\theta) \\
\Sigma_{xy}(\theta) & \Sigma_{xx}(\theta)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\Lambda_y (I - B)^{-1} (\Gamma \Phi \Gamma' + \Psi)(I - B)^{-1}' \Lambda_y' + \Theta \varepsilon & \Lambda_y (I - B)^{-1} \Gamma \Phi \Lambda_x' \\
\Lambda_x \Phi \Gamma' (I - B)^{-1}' \Lambda_y' & \Lambda_x \Phi \Lambda_x' + \Theta \delta
\end{bmatrix}
\]

• Compare to the case only with observed variables:

\[
\begin{bmatrix}
(I - B)^{-1} (\Gamma \Phi \Gamma' + \Psi)(I - B)^{-1}' & (I - B)^{-1} \Gamma \Phi \\
\Phi \Gamma' (I - B)^{-1}' & \Phi
\end{bmatrix}
\]

• What if some \( \varepsilon \) are correlated with \( \delta \)?
• \( \theta \) is globally identified if no vector \( \theta_1 \) and \( \theta_2 \) exist such that:

\[
\Sigma(\theta_1) = \Sigma(\theta_2), \quad \theta_1 \neq \theta_2
\]

• If any element of \( \theta \) is unidentifiable, estimates of all others, if any, shouldn’t be interpreted.

• Identifiability, as before, means parameters “can be” identified by algebraic form but doesn’t mean “will be” numerically.

  E.g., \( \gamma_{11} \) is algebraically solved to be \( \text{cov}(x_2, y_1)/\text{cov}(x_2, x_1) \) and sample estimate of \( \text{cov}(x_2, x_1) \) is very close to 0.

• What we’ve learned so far about identification equally applies here but only for parts of GM, not as a whole.
**t-rule**

- Number of free parameters in $\theta$ does not exceed distinctive elements in $S$:

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

- Necessary, not sufficient

- How many distinctive elements in \( S = \begin{bmatrix} 1.0 \\ 0.5 & 1.0 \\ 0.5 & 0.5 & 1.0 \end{bmatrix} \)
Two-step rule

• Step 1  --- ignore all paths between latent variables, treat them as simply mutually correlated factors, and use any feasible identification rule for CFA (with fully unconstrained $\Phi$)

• Step 2  --- if passes step 1, ignore all measurement relationships and use any feasible identification rule for SEM with observed variables

Any cause indicators can be treated as latent variables with loading of 1 and no measurement error  --- equivalent to simply including them in $\xi$

• Sufficient, not necessary; so, there will be identifiable models that do not pass two-step rule  --- local identification will be useful in such cases, though it’s empirical and fallible

• See Fig. 8.3 (p. 329), 8.4 (p. 330) and 8.5 (p. 333)
MIMIC rule

- Multiple Indicators and Multiple Causes (MIMIC) model --- a latent variable is measured with multiple (reflective) indicators and caused by multiple observed variables (formative or cause indicators, or covariates)

- A sufficient condition per \( \eta \) for identification of a MIMIC model:

\[
p \geq 2, \quad q \geq 1
\]

\[
S = \begin{bmatrix}
\phi \\
\gamma \phi \\
\lambda \gamma \phi \\
\gamma^2 \phi + \psi + \theta_{11} \\
\gamma^2 \phi + \lambda \psi \\
\lambda \gamma^2 \phi + \lambda^2 \psi + \theta_{22}
\end{bmatrix}
\]
• For cases when \( m > 1 \), if all cause indicators affect all \( \eta \)'s and each \( \eta \) is measured with at least 2 indicators, the 2-step rule applies; how?

• In general, too limited model condition in that no path modeling allowed between latent variables