General structural model – Part 2: RAM, special constraints and instrumental variables

Alternative representation

- LISREL notation does not allow:
  - Loadings from $\boldsymbol{\xi}$ to $\mathbf{y}$ or $\boldsymbol{\eta}$ to $\mathbf{x}$, or from an indicator to another
  - Covariances between different kinds of error terms (e.g., between $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$)
- These “unacceptable” parameters are a limitation of the LISREL notation, not genuine limitation of SEM
- Reticular Action Model (RAM) provides an alternative, apparently a more general SEM representation

LISREL model represented by RAM

$$\tilde{\boldsymbol{\eta}} = \tilde{\mathbf{B}}\tilde{\boldsymbol{\eta}} + \tilde{\boldsymbol{\zeta}}$$

$$\tilde{\mathbf{y}} = \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\eta}}$$

$$\tilde{\boldsymbol{\eta}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \\ \boldsymbol{\eta} \\ \boldsymbol{\xi} \end{bmatrix}, \quad \tilde{\boldsymbol{\zeta}} = \begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{\delta} \\ \boldsymbol{\zeta} \\ \boldsymbol{\xi} \end{bmatrix}, \quad \tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{x} \end{bmatrix}$$

$$\tilde{\mathbf{B}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_y & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_x \\ \mathbf{0} & \mathbf{0} & \mathbf{B} & \boldsymbol{\Gamma} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \tilde{\boldsymbol{\Lambda}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \tilde{\boldsymbol{\Psi}} = \begin{bmatrix} \boldsymbol{\Theta}_\varepsilon & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta}_\delta & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Phi} \end{bmatrix}$$

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \tilde{\boldsymbol{\Lambda}}(\mathbf{I} - \tilde{\mathbf{B}})^{-1}\tilde{\boldsymbol{\Psi}}(\mathbf{I} - \tilde{\mathbf{B}})^{-1\prime}\tilde{\boldsymbol{\Lambda}}'$$

LISREL & RAM really different?

- Either notation can represent almost all models --- compare Fig 9.1 (a) and (b), p. 398
  - All ξ are rewritten as η along with correlated ζ
  - All ε are rewritten as ζ with y turning into η
- RAM is more efficient in # of equations (2 vs. 3) and in kinds of parameters (6 vs. 15), while it requires much bigger parameter sets with most of them constrained to 0 or 1
- By RAM, the lesson is not what RAM can do more than LISREL, but what SEM can do as far as a model is identifiable
- Most programs (including LISREL) overcome the limitation of the LISREL notation

Equality & inequality constraints

- Constraints at constant values and of equality are easy to impose in terms of numerical optimization, as readily available in all SEM programs

- In contrast, parameter estimation is very difficult with constraints of inequality or a functional form --- such constraints are not widely available (EQS and COSAN seem to be exceptions)

  - Inequality --- $\lambda_{12} > c,\ \gamma_{12} > \beta_{23},\ \gamma_{12} + \beta_{23} - 2\beta_{34} > c$
  - linear equality --- $\gamma_{12} + \beta_{23} - 2\beta_{34} + c = 0$
  - nonlinear equality --- $\gamma_{12}^2 + \beta_{23} - 2\beta_{34}^3 + c = 0$

Alternative modeling for some complex constraints

- Any SEM program with the equality constraint feature can be tricked as follows (due to Rindskopf, 1983 & 1984):
  - Create a "phantom" latent variable <u>without error term</u> so that its preceding variable fully explains its variance
  - Set paths involving the phantom variable and equality constraints such that they indirectly satisfy the desired inequality or some functional relationship
- E.g., see Fig. 9.2, p. 402

Quadratic and interaction terms

- Key idea of the Kenny-Judd (1984) approach adapted to SEM is to model quadratic and product terms of latent variables through their proper indicators

- Like standard latent variable modeling, quadratic and interaction terms of latent variables must be identifiable --- typically attained by indicators of product and squared variables (see examples in pp. 403-406)

- Consider the example of quadratic term in p. 403:

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_1^2 + \zeta_1$$

$$x_1 = \xi_1 + \delta_1, \quad x_2 = \lambda_{21}\xi_1 + \delta_2, \quad y_1 = \eta_1$$

- New indicators to measure the quadratic LV

$$x_1^2 = \xi_1^2 + 2\xi_1\delta_1 + \delta_1^2$$

$$x_2^2 = \lambda_{21}^2\xi_1^2 + 2\lambda_{21}\xi_1\delta_1 + \delta_2^2$$

$$x_1 x_2 = \lambda_{21}\xi_1^2 + \lambda_{21}\xi_1\delta_1 + \xi_1\delta_1 + \delta_1\delta_2$$

- Note that, e.g., for new latent variables $\xi_1\delta_1,$ and $\xi_1\delta_2$

$$\operatorname{cov}\left(\xi_1,\delta_1\right) = E\left(\xi_1\delta_1\right) = 0$$

$$\operatorname{var}\left(\xi_1\delta_1\right) = E\left(\xi_1^2\delta_1^2\right) = \phi_{11}\operatorname{var}\left(\delta_1\right) \neq 0$$

$$\operatorname{cov}\left(\xi_1\delta_1,\xi_1\delta_2\right) = E\left(\xi_1^2\delta_1\delta_2\right)$$

- If the multinormality assumption holds for the original exogenous variables ($\xi_1, \ \delta_1, \ \delta_2$) all covariances between new LVs ($\xi_1^2, \ \xi_1\delta_1, \ \delta_1^2,$ etc.) become 0 (Kenny & Judd, 1984) and their variances are simple functions of variances of $\xi_1, \ \delta_1, \ \delta_2$ --- the only new parameter λ_{12} is identifiable (see p. 404) and so is the whole model

- However, we will need more flexible constraints such as $\lambda_i = \lambda_j^2$ (e.g., the loadings from ξ_1 and ξ_1^2, respectively, to x_1 and x_1^2), which is not readily available (e.g., unavailable in AMOS; available in Mplus and Mx); or the alternative approach may be used

- More problems:

  - Product or powered indicators can’t be normal (highly kurtotic and skewed) --- the ML and GLS estimators are still consistent but chi-square test and individual parameter test will be all invalid; alternatively, ADF might be used with large samples

  - What if the nonproduct LVs are not normal? --- Kenny-Judd approach hinges on this normality to eliminate all covariances between new LVs (e.g., ξ_1^2, $\xi_1\delta_1$, δ_1^2, etc.), and so any resulting estimates will be biased

Corrected covariance matrix (Busemeyer & Jones, 1983)

- Based on rather strong assumptions:
  - Only one indicator per nonproduct latent variable
  - All error terms are <u>independent</u> of predictors in equations and of each other
  - All nonproduct latent IVs are normal
  - Measurement error variances (or reliability of measures) known

For example, $\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \gamma_{13}\xi_1\xi_2 + \zeta_1$

$$x_1 = \xi_1 + \delta_1, \quad x_2 = \xi_2 + \delta_2, \quad y_1 = \eta_1 + \varepsilon_1$$

then estimator $\mathbf{\Gamma}' = \left(\mathbf{\Sigma}_{xx} - \mathbf{\Theta}_{\delta}\right)^{-1}\mathbf{\Sigma}_{xy}$ is consistent

Instrumental-variable estimator

$$y = \gamma x + \zeta$$

- Fundamental assumption in regression equations is that predictors be uncorrelated with residual term, $E(x\zeta) = 0$
- If unsatisfied, the OLS estimator for regression weights will be inconsistent
- Estimation by instrumental variables is to correct for such correlations of the residual term with predictors

Robust least-squares estimation by IV

- Necessary conditions of IV, z:

$$\mathrm{cov}(z,\zeta)=0, \quad \mathrm{cov}(z,x)\neq 0$$

- Then, $\mathrm{cov}(y,z)=\mathrm{cov}(\gamma x+\zeta,z)=\gamma\,\mathrm{cov}(x,z)$

$$\gamma=\frac{\mathrm{cov}(y,z)}{\mathrm{cov}(x,z)}$$

- Given multiple causal relations (equations) in measurement and path models, any variable that meets the IV conditions may be used for parameters that do not directly involve that variable --- nicely suited particularly to CFA where indicators has common causes, ξ

Robust LS estimation of loadings by IV

- Consider a single factor model with 4 indicators:

$$x_1 = \xi + \delta_1, \quad x_2 = \lambda_2 \xi + \delta_2, \quad x_3 = \lambda_3 \xi + \delta_3, \quad x_4 = \lambda_4 \xi + \delta_4$$

- By combining equations for x_1 and x_2, we have

$$x_2 = \lambda_2 x_1 + \left(\delta_2 - \lambda_2 \delta_1\right) = \lambda_2 x_1 + u$$

- As before by taking covs with IVs, say x_3 and x_4, we have

$$\begin{bmatrix} \sigma_{23} \\ \sigma_{24} \end{bmatrix} = \lambda_2 \begin{bmatrix} \sigma_{13} \\ \sigma_{14} \end{bmatrix}$$

$$\hat{\lambda}_2 = \left(\begin{bmatrix} s_{13} & s_{14} \end{bmatrix} \begin{bmatrix} s_{13} \\ s_{14} \end{bmatrix} \right)^{-1} \begin{bmatrix} s_{13} & s_{14} \end{bmatrix} \begin{bmatrix} s_{23} \\ s_{24} \end{bmatrix} = \begin{bmatrix} s_{13} \\ s_{14} \end{bmatrix}^{+} \begin{bmatrix} s_{23} \\ s_{24} \end{bmatrix}$$

- The LS estimation by IV provides consistent estimates, but not efficient --- more efficient estimator uses inverse of cov matrix of all IVs as a weight matrix like weighted LS

- If measurement models are factorially complex (more than uni-factorial)

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_A \\ x_i \\ \mathbf{x}_C \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n \\ \boldsymbol{\lambda}'_i \\ \boldsymbol{\Lambda}_C \end{bmatrix} \boldsymbol{\xi} + \boldsymbol{\delta}, \qquad i = n+1, \ldots, q$$

$$\boldsymbol{\lambda}'_B = \left(\mathbf{S}_{AC} \mathbf{S}_{CC}^{-1} \mathbf{S}'_{AC} \right)^{-1} \mathbf{S}_{AC} \mathbf{S}_{CC}^{-1} \mathbf{s}_{Ci}$$

where $\mathbf{x}_A$ contains n indicators that determine the scale of $\boldsymbol{\xi}$ and $\mathbf{x}_C$ contains $q-n-1$ IVs to estimate loadings for x_i; and, e.g., $\mathbf{S}_{AC} = E(\mathbf{x}_A \mathbf{x}'_C)$

- Once all loadings are estimated by IVs, LS estimates of $\mathbf{\Theta}_{\delta}$ and $\mathbf{\Phi}$ are easily computed; also, ACOV of $\mathbf{\Lambda}_{x}$ is known (see Eqs. 9.60, 9.61, 9.66 and 9.67)

- 2-stage least squares (2SLS; available in AMOS), like identification of general models, first ignores all causal relationships between LVs and treat the model as CFA and estimate all parameters by the IV procedure; then, path coefficients between LVs ($\mathbf{B}$ and $\mathbf{\Gamma}$) are estimated using IVs and $\mathbf{\Psi}$ is accordingly computed

- A big advantage of IV estimation is that it’s not iterative, and so IV estimates may be used as starting values for other (more desirable) estimators, yielding faster and more optimal results

Limitations of IV estimation

- Correlated error terms --- use only uncorrelated ones as IVs; in practice, we often don't know which error terms are correlated with which
- IV estimation uses only part of information in given data (one equation considered at a time) --- less efficient which is a cost for robust estimation
- Asymptotic SE available only for path coefficients and loadings, not for $\mathbf{\Phi}$, $\mathbf{\Psi}$ and $\mathbf{\Theta}$