This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike License. Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.
Introduction to Structural Equations with Latent Variables

Statistics for Psychosocial Research II: Structural Models

Qian-Li Xue
Test of causal hypotheses?

- Yes
  - SEM (Origin: Path Models)
    - Yes
      - Continuous endogenous var. and Continuous LV?
        - Yes
          - Latent Class Reg.
        - No
          - Categorical indicators and Categorical LV?
            - Yes
              - Latent Trait
            - No
              - Latent Profile
    - No
      - Classic SEM

- No
  - Ordinary Regression

Longitudinal Data?

- Yes
  - Adv. SEM I: latent growth curves)
- No
  - Multi-level Data?
    - Yes
      - Adv. SEM II: Multi-level Models
    - No
      - Classic SEM
Adding Latent Variables to the Model

- So far...we've only included observed variables in our “path models”
- Extension to latent variables:
  - need to add in a “measurement piece”
  - how do we “define” the latent variable (think factor analysis)
  - more complicated to look at, but same general principles apply
Path Diagram
Notation for Latent Variable Model

- $\eta$ = latent *endogenous* variable (eta)
- $\xi$ = latent *exogenous* variable (ksi, pronounced “kah-see”)
- $\zeta$ = latent error (zeta)
- $\beta$ = coefficient on latent *endogenous* variable (beta)
- $\gamma$ = coefficient on latent *exogenous* variable (gamma)

\[
\begin{align*}
\eta_1 &= \gamma_{11} \xi_1 + \zeta_1 \\
\eta_2 &= \beta_{21} \eta_1 + \gamma_{21} \xi_1 + \zeta_2
\end{align*}
\]

E.g. $\xi_1$ is childhood home environment, $\eta_1$ is social support, and $\eta_2$ is cancer coping ability.

\[\eta = B \eta + \Gamma \xi + \zeta, \text{Cov}(\xi) = \Phi, \text{Cov}(\zeta) = \Psi\]
Notation for Measurement Model

\[ y = \text{observed indicator of } \eta \]
\[ x = \text{observed indicator of } \xi \]
\[ \varepsilon = \text{measurement error on } y \]
\[ \delta = \text{measurement error on } x \]
\[ \lambda_x = \text{coefficient relating } x \text{ to } \xi \]
\[ \lambda_y = \text{coefficient relating } y \text{ to } \eta \]

\[ x_1 = \lambda_1 \xi_1 + \delta_1 \]
\[ x_2 = \lambda_2 \xi_1 + \delta_2 \]
\[ x_3 = \lambda_3 \xi_1 + \delta_3 \]
\[ y_1 = \lambda_4 \eta_1 + \varepsilon_1 \]
\[ y_2 = \lambda_5 \eta_1 + \varepsilon_2 \]
\[ y_3 = \lambda_6 \eta_1 + \varepsilon_3 \]
\[ y_4 = \lambda_7 \eta_2 + \varepsilon_4 \]
\[ y_5 = \lambda_8 \eta_2 + \varepsilon_5 \]
\[ y_6 = \lambda_9 \eta_2 + \varepsilon_6 \]

\[ y = \Lambda_y \eta + \varepsilon, \quad \text{Cov}(\varepsilon) = \Theta_\varepsilon \]

\[ x = \Lambda_x \xi + \delta, \quad \text{Cov}(\delta) = \Theta_\delta \]
Example: Model Specification

- $\eta_2$ is democracy in 1965
- $\eta_1$ is democracy in 1960
- $\xi_1$ is industrialization in 1960

$$\begin{bmatrix} 
\eta_1 \\
\eta_2 
\end{bmatrix} = \begin{bmatrix} 0 & 0 \\
\beta_{21} & 0 
\end{bmatrix} \begin{bmatrix} \eta_1 \\
\eta_2 
\end{bmatrix} + \begin{bmatrix} \gamma_{11} \\
\gamma_{21} 
\end{bmatrix} \begin{bmatrix} \xi_1 
\end{bmatrix} + \begin{bmatrix} \zeta_1 \\
\zeta_2 
\end{bmatrix}$$

$$\eta = B\eta + \Gamma \xi + \zeta, \quad \text{Cov}(\xi) = \Phi, \text{Cov}(\zeta) = \Psi$$
Example: Model Specification

- $Y_1, Y_5$: freedom of press
- $Y_2, Y_6$: freedom of group opposition
- $Y_3, Y_7$: fairness of election
- $Y_4, Y_8$: effectiveness of legislative body
- $x_1$ is GNP per capita
- $x_2$ is energy consumption per capita
- $x_3$ is % labor force

Bollen pp322-323
Model Estimation

- In multiple regression, estimation is based on individual cases via LS, i.e. minimization of
  \[ \sum_n (\hat{Y} - Y)^2 \]

- In SEM, estimation is based on covariances
  If Model is correct and \( \theta \) is known
  \[ \Sigma = \Sigma(\theta) \]

Where \( \Sigma \) is the population covariance matrix of observed variables and \( \Sigma(\theta) \) is the covariance matrix written as a function of \( \theta \)
Model Estimation

- Regression analysis, confirmatory factor analysis are special cases
- E.g. \( y = \gamma x + \zeta, \ \zeta \perp \gamma, \ E(\zeta) = 0 \)

\[
\begin{bmatrix}
\text{VAR}(y) \\
\text{COV}(x, y) & \text{VAR}(x)
\end{bmatrix}
= \begin{bmatrix}
\gamma^2\text{VAR}(x) + \text{VAR}(\zeta) \\
\gamma\text{VAR}(x) & \text{VAR}(x)
\end{bmatrix}
\]

\[\text{COV}(x, y) = \gamma\text{VAR}(x) \Rightarrow \gamma = \frac{\text{COV}(x, y)}{\text{VAR}(x)}\]

- Does \( \gamma \) look familiar?
  Remember in SLR, \( \beta = (x'x)^{-1}x'y \)
In reality, neither population covariances $\Sigma$ nor the parameters $\theta$ are known.

What we have is sample estimate of $\Sigma$: $S$.

Goal: estimate $\theta$ based on $S$ by choosing the estimates of $\theta$ such that $\hat{\Sigma}$ is as close to $S$ as possible.

But how close is “close”?

Define and minimize objective functions or called “fitting functions”: $F(S, \hat{\Sigma})$.

E.g. $F(S, \hat{\Sigma}) = S - \hat{\Sigma}$.
Common Fitting Functions

- **Maximum Likelihood (ML)**
  To minimize $F_{ML} = (1/2) \text{tr}([S-\Sigma(\theta)] \Sigma^{-1}(\theta))^2$ or $F_{ML} = \log|\Sigma(\theta)| + \text{tr}(S\Sigma^{-1}(\theta)) - \log|S| - (p+q)$
  - Explicit solutions of $\theta$ may not exist
  - Iterative numeric procedure is needed
  - Asymptotic properties of ML estimators:
    - Consistent, i.e. $\hat{\theta} \rightarrow \theta$ as $n \rightarrow \infty$
    - Efficient, i.e. smallest asymptotic variance
    - Asymptotic normality
Common Fitting Functions

- **Maximum Likelihood (ML)**
  
  **Advantages**
  
  - Scale invariant
    - \( F(S, \Sigma(\theta)) \) if scale invariance if \( F(S, \Sigma(\theta)) = F(DSD, D\Sigma(\theta)D) \), where \( D \) is a diagonal matrix with positive elements
    - E.g. \( D \) consists of inverses of standard deviation of observed variables, \( DSD \) becomes a correlation matrix
    - More general, the value of \( F \) is the same for any change of scale (e.g. dollars to cents)
  
  - Scale free
    - Knowing \( D \), we can calculate \( \hat{\theta}^* \) (based on transformed data) from \( \hat{\theta} \) (based on non-transformed data) without actually rerunning the model
    - Test of overall model fit for overidentified model based on the fact: \( (N-1)F_{ML} \) is a \( \chi^2 \) distribution with \( \frac{1}{2}(p+q)(p+q+1)-t \)
  
  **Disadvantage**
  
  - Assumption of multivariate normality
Common Fitting Functions

- **Unweighted Least Squares (ULS)**
  - To minimize
  \[
  F_{ULS} = \frac{1}{2} \text{tr}[(S-\Sigma(\theta))^2]
  \]
  - Analogous to OLS, minimize the sum of squares of each element in the residual matrix \(S-\Sigma(\theta)\)
  - Give greater weights to off covariance terms than variance terms
  - Explicit solutions of \(\theta\) may not exist
  - Iterative numeric procedure is needed
  - Advantages of ULS estimators:
    - Intuitive
    - Consistent, i.e. \(\hat{\theta} \to \theta\) as \(n \to \infty\)
    - No distributional assumptions
  - Disadvantages
    - Not most efficient
    - Not scale invariant, not scale free
Common Fitting Functions

- **Generalized Least Squares (GLS)**
  To minimize
  \[ F_{GLS} = \frac{1}{2} \text{tr} \left( \left\{ [S-\Sigma(\theta)]S^{-1} \right\}^2 \right) \]
  - Weights the elements of \((S- \Sigma(\theta))\) according to variances and covariances
  - \(F_{ULS}\) is a special case of \(F_{GLS}\) with \(S^{-1}=I\)
  - Advantages of ULS estimators:
    - Intuitive
    - Consistent, i.e. \(\hat{\theta} \to \theta\) as \(n \to \infty\)
    - Asymptotic normality (availability of significance test)
    - Asymptotically efficient
    - Scale invariant and scale free
    - Test of overall model fit for overidentified model based on the fact: \((N-1)F_{GLS}\) is a \(\chi^2\) distribution with \(\frac{1}{2}(p+q)(p+q+1)-t\)
  - Disadvantages
    - Sensitive to “fat” or “thin” tails
Example: Model Specification

- $Y_1, Y_5$: freedom of press
- $Y_2, Y_6$: freedom of group opposition
- $Y_3, Y_7$: fairness of election
- $Y_4, Y_8$: effectiveness of legislative body

- $x_1$ is GNP per capita
- $x_2$ is energy consumption per capita
- $x_3$ is % labor force

Bollen pp322-323
Simple Case of SEM with Latent Variables:
Confirmatory Factor Analysis
Recap of Basic Characteristics of Exploratory Factor Analysis (EFA)

- Most EFA extract orthogonal factors, which is “boring” to SEM users
- Distinction between common and unique variances
- EFA is underidentified (i.e. no unique solution)
  - Remember rotation? Equally good fit with different rotations!
- All measures are related to each factor
Confirmatory Factor Analysis (CFA)

- Takes factor analysis a step further.
- We can “test” or “confirm” or “implement” a “highly constrained a priori structure that meets conditions of model identification”
- But be careful, a model can never be confirmed!!
- CFA model is constructed in advance
- number of latent variables (“factors”) is pre-set by analyst (not part of the modeling usually)
- Whether latent variable influences observed is specified
- Measurement errors may correlate
- Difference between CFA and the usual SEM:
  - SEM assumes causally interrelated latent variables
  - CFA assumes interrelated latent variables (i.e. exogenous)
Exploratory Factor Analysis

Two factor model:

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

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
  x_5 \\
  x_6 \\
\end{bmatrix}
= \begin{bmatrix}
  \lambda_{11} & \lambda_{12} \\
  \lambda_{21} & \lambda_{22} \\
  \lambda_{31} & \lambda_{32} \\
  \lambda_{41} & \lambda_{42} \\
  \lambda_{51} & \lambda_{52} \\
  \lambda_{61} & \lambda_{62} \\
\end{bmatrix}
\begin{bmatrix}
  \xi_1 \\
  \xi_2 \\
\end{bmatrix}
+ \begin{bmatrix}
  \delta_1 \\
  \delta_2 \\
  \delta_3 \\
  \delta_4 \\
  \delta_5 \\
  \delta_6 \\
\end{bmatrix}
\]
Two factor model: \[ x = \Lambda \xi + \delta \]
For the “matrix-challenged”

<table>
<thead>
<tr>
<th>CFA</th>
<th>EFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1 = \lambda_{11} \xi_1 + \delta_1$</td>
<td>$x_1 = \lambda_{11} \xi_1 + \lambda_{12} \xi_2 + \delta_1$</td>
</tr>
<tr>
<td>$x_2 = \lambda_{21} \xi_1 + \delta_2$</td>
<td>$x_2 = \lambda_{21} \xi_1 + \lambda_{22} \xi_2 + \delta_2$</td>
</tr>
<tr>
<td>$x_3 = \lambda_{31} \xi_1 + \delta_3$</td>
<td>$x_3 = \lambda_{31} \xi_1 + \lambda_{32} \xi_2 + \delta_3$</td>
</tr>
<tr>
<td>$x_4 = \lambda_{42} \xi_2 + \delta_4$</td>
<td>$x_4 = \lambda_{41} \xi_1 + \lambda_{42} \xi_2 + \delta_4$</td>
</tr>
<tr>
<td>$x_5 = \lambda_{52} \xi_2 + \delta_5$</td>
<td>$x_5 = \lambda_{51} \xi_1 + \lambda_{52} \xi_2 + \delta_5$</td>
</tr>
<tr>
<td>$x_6 = \lambda_{62} \xi_2 + \delta_6$</td>
<td>$x_6 = \lambda_{61} \xi_1 + \lambda_{62} \xi_2 + \delta_6$</td>
</tr>
</tbody>
</table>

$\text{cov}(\xi_1, \xi_2) = \varphi_{12}$

$\text{cov}(\xi_1, \xi_2) = 0$
More Important Notation

- \( \Phi \) (capital of \( \varphi \)): covariance matrix of factors

\[
\Phi = \text{var}(\xi) = \begin{bmatrix}
\varphi_{11} & \varphi_{12} \\
\varphi_{12} & \varphi_{22}
\end{bmatrix}
\]

\( \text{var}(\xi_1) = \varphi_{11} \)
\( \text{cov}(\xi_1, \xi_2) = \varphi_{12} \)

- \( \Psi \) (capital of \( \psi \)): covariance matrix of errors

\[
\Psi = \text{var}(\Delta) = \begin{bmatrix}
\psi_{11} & \psi_{12} & \psi_{13} & \psi_{14} \\
\psi_{21} & \psi_{22} & \psi_{23} & \psi_{24} \\
\psi_{31} & \psi_{32} & \psi_{33} & \psi_{34} \\
\psi_{41} & \psi_{42} & \psi_{43} & \psi_{44} \\
\psi_{51} & \psi_{52} & \psi_{53} & \psi_{54} \\
\psi_{61} & \psi_{62} & \psi_{63} & \psi_{64} \\
\psi_{65} & \psi_{66}
\end{bmatrix}
\]

\( \text{var}(\delta_1) = \psi_{11} \)
\( \text{cov}(\delta_1, \delta_2) = \psi_{12} \)

**NOTE:** usually, \( \psi_{ij} = 0 \) if \( i \neq j \)
Model Estimation

- In our previous CFA example, we have six equations, and many more unknowns.
- In this form, not enough information to uniquely solve for $\lambda$ and the factor correlation.
- What if we multiple both sides of $x = \Lambda \xi + \delta$ by $X'$

$$xx' = (\lambda \xi + \delta)(\lambda \xi + \delta)'$$
$$= (\lambda \xi)(\lambda \xi)' + (\lambda \xi)\delta' + \delta(\lambda \xi)' + \delta \delta'$$

because $\xi$ and $\delta$ are independent.

$$xx' = (\lambda \xi)(\lambda \xi)' + \delta \delta'$$
$$= \lambda \xi \xi' \lambda' + \delta \delta'$$

$$\Sigma = \lambda \Phi \lambda + \Psi = \Sigma(\Theta)$$

where $\Phi$ is the covariance matrix of factors $\xi$, and $\Psi$ is error covariance matrix.
Model Constraints

- Hallmark of CFA
- Purposes for setting constraints:
  - Test a priori theory
  - Ensure identifiability
  - Test reliability of measures
Model Constraints: Identifiability

- Latent variables (LVs) need some constraints
- Because factors are unmeasured, their variances can take different values
- Recall EFA where we constrained factors:
  \[ F \sim N(0,1) \]
- Otherwise, model is not identifiable.
- Here we have two options:
  - Fix variance of latent variables (LV) to be 1 (or another constant)
  - Fix one path between LV and indicator
Necessary Constraints

Fix variances:

Fix path:
Model Parametrization

Fix variances:
\[ x_1 = \lambda_{11} \xi_1 + \delta_1 \]
\[ x_2 = \lambda_{21} \xi_1 + \delta_2 \]
\[ x_3 = \lambda_{31} \xi_1 + \delta_3 \]
\[ x_4 = \lambda_{42} \xi_2 + \delta_4 \]
\[ x_5 = \lambda_{52} \xi_2 + \delta_5 \]
\[ x_6 = \lambda_{62} \xi_2 + \delta_6 \]
\[ \text{cov}(\xi_1, \xi_2) = \varphi_{12} \]
\[ \text{var}(\xi_1) = 1 \]
\[ \text{var}(\xi_2) = 1 \]

Fix path:
\[ x_1 = \xi_1 + \delta_1 \]
\[ x_2 = \lambda_{21} \xi_1 + \delta_2 \]
\[ x_3 = \lambda_{31} \xi_1 + \delta_3 \]
\[ x_4 = \xi_2 + \delta_4 \]
\[ x_5 = \lambda_{52} \xi_2 + \delta_5 \]
\[ x_6 = \lambda_{62} \xi_2 + \delta_6 \]
\[ \text{cov}(\xi_1, \xi_2) = \varphi_{12} \]
\[ \text{var}(\xi_1) = \varphi_{11} \]
\[ \text{var}(\xi_2) = \varphi_{22} \]
Model Constraints: Reliability Assessment

- Parallel measures
  - $T_{x1} = T_{x2} [= E(x)]$ (True scores are equal)
  - $T$ affects $x1$ and $x2$ equally
  - $\text{Cov}(\delta_1, \delta_2) = 0$ (Errors not correlated)
  - $\text{Var}(\delta_1) = \text{Var}(\delta_2)$ (Equal error variances)
Model Constraints: Reliability Assessment

- **Tau equivalent measures**
  - \( T_{x1} = T_{x2} \)
  - \( T \) affects \( x1 \) and \( x2 \) equally
  - \( \text{Var} (\delta_1) \neq \text{Var} (\delta_2) \)
  - Note: for standardized measures, it makes no sense to constrain the loadings without also constraining the residuals, since \( \text{Var}(x) = 1.0 \)