

Structural Equation Models: The General Model¹

STA431 Spring 2023

¹See last slide for copyright information.

Features of Structural Equation Models

- Multiple equations.
- All the variables are random.
- An explanatory variable in one equation can be the response variable in another equation.
- Models are represented by path diagrams.
- Identifiability is always an issue.
- The statistical models are models of influence. They are *causal models*.

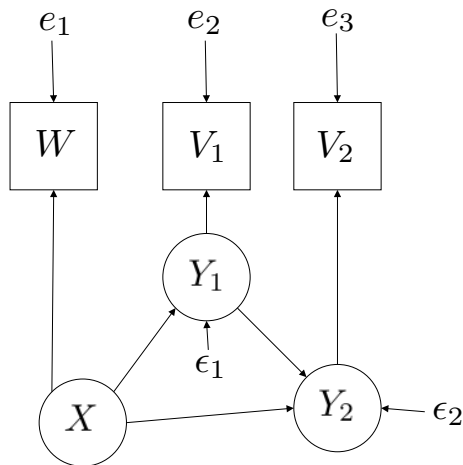
Modest changes in notation

$$Y_{i,1} = \alpha_1 + \gamma_1 X_{i,1} + \gamma_2 X_{i,2} + \epsilon_{i,1}$$

$$Y_{i,2} = \alpha_2 + \beta Y_{i,1} + \epsilon_{i,2}$$

- Regression coefficients (links between exogenous variables and endogenous variables) are now called gamma instead of beta.
- Betas are used for links between endogenous variables.
- Intercepts will soon disappear.

Example: A Path Model with Measurement Error



$$Y_{i,1} = \alpha_1 + \gamma_1 X_i + \epsilon_{i,1}$$

$$Y_{i,2} = \alpha_2 + \beta Y_{i,1} + \gamma_2 X_i + \epsilon_{i,2}$$

$$W_i = \nu_1 + \lambda_1 X_i + e_{i,1}$$

$$V_{i,1} = \nu_2 + \lambda_2 Y_{i,1} + e_{i,2}$$

$$V_{i,2} = \nu_3 + \lambda_3 Y_{i,2} + e_{i,3}$$

The General (original) Model: Independently for $i = 1, \dots, n$, let

$$\mathbf{y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i$$

$$\mathbf{F}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix}$$

$$\mathbf{d}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i, \text{ where}$$

- \mathbf{y}_i is a $q \times 1$ *latent* random vector.
- $\boldsymbol{\alpha}$ is a $q \times 1$ vector of constants.
- $\boldsymbol{\beta}$ is a $q \times q$ matrix of constants *with zeros on the main diagonal*.
- $\boldsymbol{\Gamma}$ is a $q \times p$ matrix of constants.
- \mathbf{x}_i is a $p \times 1$ *latent* random vector with expected value $\boldsymbol{\mu}_x$ and positive definite covariance matrix $\boldsymbol{\Phi}_x$.
- $\boldsymbol{\epsilon}_i$ is a $q \times 1$ latent random vector with expected value zero and positive definite covariance matrix $\boldsymbol{\Psi}$.
- \mathbf{F}_i (F for Factor) is a partitioned vector with \mathbf{x}_i stacked on top of \mathbf{y}_i . It is a $(p + q) \times 1$ latent random vector whose expected value is denoted by $\boldsymbol{\mu}_F$, and whose variance-covariance matrix is denoted by $\boldsymbol{\Phi}$.
- \mathbf{d}_i is a $k \times 1$ *observable* random vector. The expected value of \mathbf{d}_i will be denoted by $\boldsymbol{\mu}$, and the covariance matrix of \mathbf{d}_i will be denoted by $\boldsymbol{\Sigma}$.
- $\boldsymbol{\nu}$ is a $k \times 1$ vector of constants.
- $\boldsymbol{\Lambda}$ is a $k \times (p + q)$ matrix of constants.
- \mathbf{e}_i is a $k \times 1$ latent random vector with expected value zero and covariance matrix $\boldsymbol{\Omega}$, which need not be positive definite.
- \mathbf{x}_i , $\boldsymbol{\epsilon}_i$ and \mathbf{e}_i are independent.

Surrogate Models

Truth \approx Original Model \rightarrow Surrogate Model 1 \rightarrow Surrogate Model 2 \dots

- We more or less accept the original model, but we can't identify the parameters.
- So we re-parameterize, obtaining a surrogate model. Repeat.
- We will carefully keep track of the *meaning* of the new parameters in terms of the parameters of the original model.

The Original Model

$$\begin{aligned}\mathbf{y}_i &= \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i \\ \mathbf{F}_i &= \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix} \\ \mathbf{d}_i &= \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i\end{aligned}$$

where ...

- Carefully count the parameters that appear *only* in $E(\mathbf{d}_i) = \boldsymbol{\mu}$ and not in $cov(\mathbf{d}_i)$.
- There are more of these parameters than elements of $E(\mathbf{d}_i)$.
- Parameter count rule.

Center the model

- There are too many expected values and intercepts to identify.
- Center all the random variables in the model by adding and subtracting expected values.
- Obtain a *centered surrogate model*

$$\overset{c}{\mathbf{y}}_i = \boldsymbol{\beta} \overset{c}{\mathbf{y}}_i + \boldsymbol{\Gamma} \overset{c}{\mathbf{x}}_i + \boldsymbol{\epsilon}_i$$

$$\overset{c}{\mathbf{F}}_i = \begin{pmatrix} \overset{c}{\mathbf{x}}_i \\ \overset{c}{\mathbf{y}}_i \end{pmatrix}$$

$$\overset{c}{\mathbf{d}}_i = \boldsymbol{\Lambda} \overset{c}{\mathbf{F}}_i + \mathbf{e}_i$$

- Same $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Lambda}$, same variances and covariances.

Change of variables

- Centering is a change of variables.
- Expected values and intercepts are gone, and the dimension of the parameter space is reduced.
- Drop the little c over the random vectors.

The General Centered Model

Independently for $i = 1, \dots, n$,

$$\mathbf{y}_i = \beta \mathbf{y}_i + \Gamma \mathbf{x}_i + \boldsymbol{\epsilon}_i$$

$$\mathbf{F}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix}$$

$$\mathbf{d}_i = \Lambda \mathbf{F}_i + \mathbf{e}_i$$

- \mathbf{d}_i (the data) are observable. All other variables are latent.
- $\mathbf{y}_i = \beta \mathbf{y}_i + \Gamma \mathbf{x}_i + \boldsymbol{\epsilon}_i$ is called the *Latent Variable Model*.
- The latent vectors \mathbf{x}_i and \mathbf{y}_i are collected into a *factor* \mathbf{F}_i .
- $\mathbf{d}_i = \Lambda \mathbf{F}_i + \mathbf{e}_i$ is called the *Measurement Model*.

$$\mathbf{y}_i = \boldsymbol{\beta} \mathbf{y}_i + \boldsymbol{\Gamma} \mathbf{x}_i + \boldsymbol{\epsilon}_i \quad \mathbf{F}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix} \quad \mathbf{d}_i = \boldsymbol{\Lambda} \mathbf{F}_i + \mathbf{e}_i$$

- \mathbf{y}_i is a $q \times 1$ latent random vector.
- $\boldsymbol{\beta}$ is a $q \times q$ matrix of constants with zeros on the main diagonal.
- \mathbf{x}_i is a $p \times 1$ latent random vector.
- $\boldsymbol{\Gamma}$ is a $q \times p$ matrix of constants.
- $\boldsymbol{\epsilon}_i$ is a $q \times 1$ vector of error terms.
- \mathbf{F}_i (F for Factor) is just \mathbf{x}_i stacked on top of \mathbf{y}_i . It is a $(p + q) \times 1$ latent random vector.
- \mathbf{d}_i is a $k \times 1$ observable random vector. Sometimes, $\mathbf{d}_i = \begin{pmatrix} \mathbf{w}_i \\ \mathbf{v}_i \end{pmatrix}$.
- $\boldsymbol{\Lambda}$ is a $k \times (p + q)$ matrix of constants: “factor loadings.”
- \mathbf{e}_i is a $k \times 1$ vector of error terms.
- \mathbf{x}_i , $\boldsymbol{\epsilon}_i$ and \mathbf{e}_i are independent.

Parameters

More notation

$$\mathbf{y}_i = \beta \mathbf{y}_i + \Gamma \mathbf{x}_i + \epsilon_i$$

$$\mathbf{F}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix}$$

$$\mathbf{d}_i = \Lambda \mathbf{F}_i + \mathbf{e}_i$$

$$\text{cov}(\mathbf{x}_i) = \Phi_x$$

$$\text{cov}(\epsilon_i) = \Psi$$

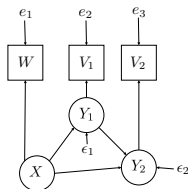
$$\text{cov}(\mathbf{F}_i) = \Phi = \begin{pmatrix} \text{cov}(\mathbf{x}_i) & \text{cov}(\mathbf{x}_i, \mathbf{y}_i) \\ \text{cov}(\mathbf{y}_i, \mathbf{x}_i) & \text{cov}(\mathbf{y}_i) \end{pmatrix} = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{pmatrix}$$

$$\text{cov}(\mathbf{e}_i) = \Omega$$

$$\text{cov}(\mathbf{d}_i) = \Sigma$$

- Collect the unique elements of β , Γ , Λ , Φ_x , Ψ and Ω into a parameter vector θ .
- θ is a *function* of the original model parameters.

Matrix Form



$$Y_{i,1} = \gamma_1 X_i + \epsilon_{i,1}$$

$$Y_{i,2} = \beta Y_{i,1} + \gamma_2 X_i + \epsilon_{i,2}$$

$$W_i = \lambda_1 X_i + e_{i,1}$$

$$V_{i,1} = \lambda_2 Y_{i,1} + e_{i,2}$$

$$V_{i,2} = \lambda_3 Y_{i,2} + e_{i,3}$$

$$\mathbf{y}_i = \beta \mathbf{y}_i + \Gamma \mathbf{x}_i + \boldsymbol{\epsilon}_i$$

$$\mathbf{F}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix}$$

$$\mathbf{d}_i = \Lambda \mathbf{F}_i + \mathbf{e}_i$$

$$\mathbf{y}_i = \beta \mathbf{y}_i + \Gamma \mathbf{x}_i + \boldsymbol{\epsilon}_i$$

$$\begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \beta & 0 \end{pmatrix} \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} X_i + \begin{pmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \end{pmatrix}$$

$$\mathbf{d}_i = \Lambda \mathbf{F}_i + \mathbf{e}_i$$

$$\begin{pmatrix} W_i \\ V_{i,1} \\ V_{i,2} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix} \begin{pmatrix} X_i \\ Y_{i,1} \\ Y_{i,2} \end{pmatrix} + \begin{pmatrix} e_{i,1} \\ e_{i,2} \\ e_{i,3} \end{pmatrix}$$

Observable variables in the “latent” variable model

$$\mathbf{y}_i = \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i$$

Fairly common

- These present no problem.
- Let $P(e_j = 0) = 1$, so $Var(e_j) = 0$.
- And $Cov(e_i, e_j) = 0$
- Because if $P(e_j = 0) = 1$,

$$\begin{aligned}Cov(e_i, e_j) &= E(e_i e_j) - E(e_i)E(e_j) \\&= E(e_i \cdot 0) - E(e_i) \cdot 0 \\&= 0 - 0 = 0\end{aligned}$$

- In $\boldsymbol{\Omega} = cov(\mathbf{e}_i)$, column j (and row j) are all zeros.
- $\boldsymbol{\Omega}$ singular, no problem.

What should you be able to do?

- Given a path diagram, write the model equations and say which exogenous variables are correlated with each other.
- Given the model equations and information about which exogenous variables are correlated with each other, draw the path diagram.
- Given either piece of information, write the model in matrix form and say what all the matrices are.
- Calculate model covariance matrices.
- Check identifiability.

Recall the notation

$$\mathbf{y}_i = \beta \mathbf{y}_i + \mathbf{\Gamma} \mathbf{x}_i + \boldsymbol{\epsilon}_i$$

$$\mathbf{F}_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y}_i \end{pmatrix}$$

$$\mathbf{d}_i = \mathbf{\Lambda} \mathbf{F}_i + \mathbf{e}_i$$

$$\text{cov}(\mathbf{x}_i) = \mathbf{\Phi}_x$$

$$\text{cov}(\boldsymbol{\epsilon}_i) = \mathbf{\Psi}$$

$$\text{cov}(\mathbf{F}_i) = \mathbf{\Phi} = \begin{pmatrix} \text{cov}(\mathbf{x}_i) & \text{cov}(\mathbf{x}_i, \mathbf{y}_i) \\ \text{cov}(\mathbf{y}_i, \mathbf{x}_i) & \text{cov}(\mathbf{y}_i) \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi}_{11} & \mathbf{\Phi}_{12} \\ \mathbf{\Phi}_{12}^\top & \mathbf{\Phi}_{22} \end{pmatrix}$$

$$\text{cov}(\mathbf{e}_i) = \mathbf{\Omega}$$

$$\text{cov}(\mathbf{D}_i) = \mathbf{\Sigma}$$

Calculate a general expression for $\mathbf{\Sigma}(\boldsymbol{\theta})$.

For the latent variable model, calculate $\Phi = cov(\mathbf{F}_i)$

Have $cov(\mathbf{x}_i) = \Phi_x$, need $cov(\mathbf{y}_i)$ and $cov(\mathbf{x}_i, \mathbf{y}_i)$

$$\begin{aligned}\mathbf{y}_i &= \beta \mathbf{y}_i + \Gamma \mathbf{x}_i + \epsilon_i \\ \Rightarrow \mathbf{y}_i - \beta \mathbf{y}_i &= \Gamma \mathbf{x}_i + \epsilon_i \\ \Rightarrow \mathbf{I} \mathbf{y}_i - \beta \mathbf{y}_i &= \Gamma \mathbf{x}_i + \epsilon_i \\ \Rightarrow (\mathbf{I} - \beta) \mathbf{y}_i &= \Gamma \mathbf{x}_i + \epsilon_i \\ \Rightarrow (\mathbf{I} - \beta)^{-1} (\mathbf{I} - \beta) \mathbf{y}_i &= (\mathbf{I} - \beta)^{-1} (\Gamma \mathbf{x}_i + \epsilon_i) \\ \Rightarrow \mathbf{y}_i &= (\mathbf{I} - \beta)^{-1} (\Gamma \mathbf{x}_i + \epsilon_i)\end{aligned}$$

So,

$$\begin{aligned}cov(\mathbf{y}_i) &= (\mathbf{I} - \beta)^{-1} cov(\Gamma \mathbf{x}_i + \epsilon_i) (\mathbf{I} - \beta)^{-1\top} \\ &= (\mathbf{I} - \beta)^{-1} (cov(\Gamma \mathbf{x}_i) + cov(\epsilon_i)) (\mathbf{I} - \beta^\top)^{-1} \\ &= (\mathbf{I} - \beta)^{-1} \left(\Gamma \Phi_x \Gamma^\top + \Psi \right) (\mathbf{I} - \beta^\top)^{-1}\end{aligned}$$

Theorem: If the original model holds, $(\mathbf{I} - \beta)^{-1}$ exists.

$\mathbf{y}_i = \alpha + \beta \mathbf{y}_i + \Gamma \mathbf{x}_i + \epsilon_i$ yields $(\mathbf{I} - \beta) \mathbf{y}_i = \alpha + \Gamma \mathbf{x}_i + \epsilon_i$.
Suppose $(\mathbf{I} - \beta)^{-1}$ does not exist.

Then the rows of $\mathbf{I} - \beta$ are linearly dependent, and there is a $q \times 1$ non-zero vector of constants \mathbf{a} with $\mathbf{a}^\top (\mathbf{I} - \beta) = 0$. So,

$$\begin{aligned} \mathbf{a}^\top (\mathbf{I} - \beta) \mathbf{y}_i &= 0 = \mathbf{a}^\top \alpha + \mathbf{a}^\top \Gamma \mathbf{x}_i + \mathbf{a}^\top \epsilon_i \\ \Rightarrow \text{Var}(0) &= \text{Var}(\mathbf{a}^\top \Gamma \mathbf{x}_i) + \text{Var}(\mathbf{a}^\top \epsilon_i) \\ \Rightarrow 0 &= \mathbf{a}^\top \Gamma \Phi_x \Gamma^\top \mathbf{a} + \mathbf{a}^\top \Psi \mathbf{a} > 0. \end{aligned}$$

Contradicts $\mathbf{I} - \beta$ singular.

A hole in the parameter space

$|\mathbf{I} - \beta| \neq 0$ can create a hole in the parameter space.

More calculations

- Have $cov(\mathbf{y}_i) = (\mathbf{I} - \beta)^{-1} (\mathbf{\Gamma} \mathbf{\Phi}_x \mathbf{\Gamma}^\top + \mathbf{\Psi}) (\mathbf{I} - \beta^\top)^{-1}$.
- Know $cov(\mathbf{x}_i) = \mathbf{\Phi}_x$
- Easy to get $cov(\mathbf{x}_i, \mathbf{y}_i)$.

For the measurement model, calculate $\Sigma = cov(\mathbf{d}_i)$

$$\begin{aligned}\mathbf{d}_i &= \mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i \\ \Rightarrow cov(\mathbf{d}_i) &= cov(\mathbf{\Lambda}\mathbf{F}_i + \mathbf{e}_i) \\ &= cov(\mathbf{\Lambda}\mathbf{F}_i) + cov(\mathbf{e}_i) \\ &= \mathbf{\Lambda}cov(\mathbf{F}_i)\mathbf{\Lambda}^\top + cov(\mathbf{e}_i) \\ &= \mathbf{\Lambda}\Phi\mathbf{\Lambda}^\top + \mathbf{\Omega} \\ &= \mathbf{\Sigma}\end{aligned}$$

Two-stage Proofs of Identifiability

Stage 1 is the latent variable model and Stage 2 is the measurement model.

- Show the parameters of the latent variable model $(\beta, \Gamma, \Phi_x, \Psi)$ can be recovered from $\Phi = cov(\mathbf{F}_i)$.
- Solve $\begin{pmatrix} cov(\mathbf{x}_i) & cov(\mathbf{x}_i, \mathbf{y}_i) \\ cov(\mathbf{y}_i, \mathbf{x}_i) & cov(\mathbf{y}_i) \end{pmatrix} = \Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{12}^\top & \Phi_{22} \end{pmatrix}$ for $(\beta, \Gamma, \Phi_x, \Psi)$?
- Show the parameters of the measurement model (Λ, Φ, Ω) can be recovered from $\Sigma = cov(\mathbf{d}_i)$.
- This means all the parameters can be recovered from Σ .
- Break a big problem into two smaller ones.
- Develop *rules* for checking identifiability at each stage.
- Just look at the path diagram.

Copyright Information

This slide show was prepared by **Jerry Brunner**, Department of Statistical Sciences, University of Toronto. It is licensed under a **Creative Commons Attribution - ShareAlike 3.0 Unported License**. Use any part of it as you like and share the result freely. The \LaTeX source code is available from the course website:

<http://www.utstat.toronto.edu/brunner/oldclass/431s23>