Structural Equation Models: The General Case[1]

STA2053 Fall 2022

[1] See last slide for copyright information.

Features of Structural Equation Models

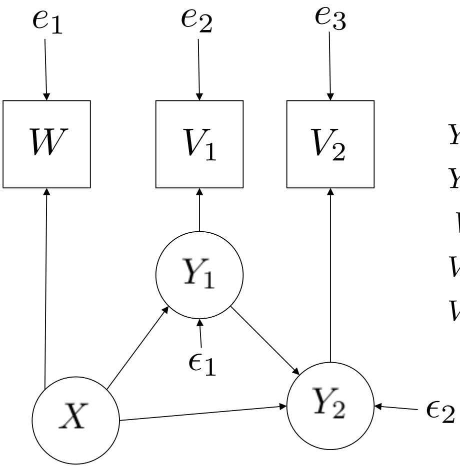
- Multiple equations.
- All the variables are random.
- An explanatory variable in one equation can be the response variable in another equation.
- Models are represented by path diagrams.
- Identifiability is always an issue.
- The statistical models are models of influence. They are often called *causal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

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

$$
\begin{aligned}
\mathbf{y}_i &= \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \left(\begin{array}{c} \mathbf{x}_i \\ \hline \mathbf{y}_i \end{array} \right) \\
\mathbf{d}_i &= \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i, \text{ where}
\end{aligned}
$$

- $\mathbf{y}_i$ is a $q \times 1$ 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 $p \times 1$ random vector with expected value $\boldsymbol{\mu}_x$ and positive definite covariance matrix $\boldsymbol{\Phi}_x$.
- $\boldsymbol{\epsilon}_i$ is a $q \times 1$ 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 $(p+q) \times 1$ 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 $k \times 1$ 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 $k \times 1$ 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.

Example: A Path Model with Measurement Error



$$
\begin{aligned}
Y_{i,1} &= \alpha_1 + \gamma_1 X_i + \epsilon_{i,1} \\
Y_{i,2} &= \alpha_2 + \beta Y_{i,1} + \gamma_2 X_i + \epsilon_{i,2} \\
W_i &= \nu_1 + \lambda_1 X_i + e_{i,1} \\
V_{i,1} &= \nu_2 + \lambda_2 Y i, 1 + e_{i,2} \\
V_{i,2} &= \nu_3 + \lambda_3 Y i, 2 + e_{i,3}
\end{aligned}
$$

Truth $\approx$ Original Model $\rightarrow$ Surrogate Model 1 $\rightarrow$ Surrogate Model 2 $\ldots$

Surrogate Mother

https://en.wikipedia.org/wiki/File:Harlow%27s_Monkey.png

Surrogate Models

Truth $\approx$ Original Model $\rightarrow$ Surrogate Model 1 $\rightarrow$ Surrogate Model 2 ...

- 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

$$
\begin{aligned}
\mathbf{y}_i &= \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \left(\frac{\mathbf{x}_i}{\mathbf{y}_i} \right) \\
\mathbf{d}_i &= \boldsymbol{\nu} + \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i,
\end{aligned}
$$

where ...

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

$$
\begin{aligned}
\stackrel{c}{\mathbf{y}}_i &= \boldsymbol{\beta} \stackrel{c}{\mathbf{y}}_i + \boldsymbol{\Gamma} \stackrel{c}{\mathbf{x}}_i + \boldsymbol{\epsilon}_i \\
\stackrel{c}{\mathbf{F}}_i &= \left(\begin{array}{c} \stackrel{c}{\mathbf{x}}_i \\ \hline \stackrel{c}{\mathbf{y}}_i \end{array} \right) \\
\stackrel{c}{\mathbf{d}}_i &= \boldsymbol{\Lambda} \stackrel{c}{\mathbf{F}}_i + \mathbf{e}_i
\end{aligned}
$$

- 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.

A General Two-Stage Centered Model

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

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

$$
\begin{aligned}
\mathbf{y}_i &= \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \left(\begin{array}{c} \mathbf{x}_i \\ \hline \mathbf{y}_i \end{array} \right) \\
\mathbf{d}_i &= \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i
\end{aligned}
$$

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

$\mathbf{y}_i = \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i \quad \mathbf{F}_i = \left(\begin{array}{c} \mathbf{x}_i \\ \hline \mathbf{y}_i \end{array} \right) \quad \mathbf{d}_i = \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i$

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

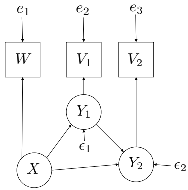
Covariance matrices
More notation

$$
\begin{aligned}
\mathbf{y}_i &= \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \left(\frac{\mathbf{x}_i}{\mathbf{y}_i} \right) \\
\mathbf{d}_i &= \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i
\end{aligned}
$$

$$
\begin{aligned}
cov(\mathbf{x}_i) &= \boldsymbol{\Phi}_x \\
cov(\boldsymbol{\epsilon}_i) &= \boldsymbol{\Psi} \\
cov(\mathbf{F}_i) &= \boldsymbol{\Phi} = \left(\begin{array}{cc} cov(\mathbf{x}_i) & cov(\mathbf{x}_i, \mathbf{y}_i) \\ cov(\mathbf{y}_i, \mathbf{x}_i) & cov(\mathbf{y}_i) \end{array} \right) = \left(\begin{array}{cc} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{12}^\top & \boldsymbol{\Phi}_{22} \end{array} \right) \\
cov(\mathbf{e}_i) &= \boldsymbol{\Omega} \\
cov(\mathbf{d}_i) &= \boldsymbol{\Sigma}
\end{aligned}
$$

- Collect the unique elements of $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}_x$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$ into a parameter vector $\boldsymbol{\theta}$.
- $\boldsymbol{\theta}$ is a *function* of the original model parameters.

Matrix Form



$$
\begin{aligned}
Y_{i,1} &= \gamma_1 X_i + \epsilon_{i,1} \\
Y_{i,2} &= \beta Y_{i,1} + \gamma_2 X_i + \epsilon_{i,2} \\
W_i &= X_i + e_{i,1} \\
V_{i,1} &= Y_{i,1} + e_{i,2} \\
V_{i,2} &= Y_{i,2} + e_{i,3}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{y}_i &= \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \left(\frac{\mathbf{x}_i}{\mathbf{y}_i}\right) \\
\mathbf{d}_i &= \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i
\end{aligned}
$$

$$
\begin{array}{ccccccccc}
\mathbf{y}_i & = & \boldsymbol{\beta} & \mathbf{y}_i & + & \boldsymbol{\Gamma} & \mathbf{x}_i & + & \boldsymbol{\epsilon}_i \\
\begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} & = & \begin{pmatrix} 0 & 0 \\ \beta & 0 \end{pmatrix} & \begin{pmatrix} Y_{i,1} \\ Y_{i,2} \end{pmatrix} & + & \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} & X_i & + & \begin{pmatrix} \epsilon_{i,1} \\ \epsilon_{i,2} \end{pmatrix}
\end{array}
$$

$$
\begin{array}{ccccccc}
\mathbf{d}_i & = & \boldsymbol{\Lambda} & \mathbf{F}_i & + & \mathbf{e}_i \\
\begin{pmatrix} W_i \\ V_{i,1} \\ V_{i,2} \end{pmatrix} & = & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} & \begin{pmatrix} X_i \\ Y_{i,1} \\ Y_{i,2} \end{pmatrix} & + & \begin{pmatrix} e_{i,1} \\ e_{i,2} \\ e_{i,3} \end{pmatrix}
\end{array}
$$

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

$$\begin{aligned}
\mathbf{y}_i &= \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \left(\frac{\mathbf{x}_i}{\mathbf{y}_i}\right) \\
\mathbf{d}_i &= \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i
\end{aligned}$$

$$\begin{aligned}
cov(\mathbf{x}_i) &= \boldsymbol{\Phi}_x \\
cov(\boldsymbol{\epsilon}_i) &= \boldsymbol{\Psi} \\
cov(\mathbf{F}_i) &= \boldsymbol{\Phi} = \begin{pmatrix} cov(\mathbf{x}_i) & cov(\mathbf{x}_i, \mathbf{y}_i) \\ cov(\mathbf{y}_i, \mathbf{x}_i) & cov(\mathbf{y}_i) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{12}^\top & \boldsymbol{\Phi}_{22} \end{pmatrix} \\
cov(\mathbf{e}_i) &= \boldsymbol{\Omega} \\
cov(\mathbf{D}_i) &= \boldsymbol{\Sigma}
\end{aligned}$$

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

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

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

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

So,

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

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

$\mathbf{y}_i = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{y}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i$ yields $(\mathbf{I} - \boldsymbol{\beta})\mathbf{y}_i = \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\epsilon}_i$. Suppose $(\mathbf{I} - \boldsymbol{\beta})^{-1}$ does not exist. Then the rows of $\mathbf{I} - \boldsymbol{\beta}$ are linearly dependent, and there is a $q \times 1$ non-zero vector of constants $\mathbf{a}$ with $\mathbf{a}^\top(\mathbf{I} - \boldsymbol{\beta}) = 0$. So,

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

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

A hole in the parameter space

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

More calculations

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

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

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

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 $(\boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}_x, \boldsymbol{\Psi})$ can be recovered from $\boldsymbol{\Phi} = cov(\mathbf{F}_i)$.
- Solve $\left(\begin{array}{cc} cov(\mathbf{x}_i) & cov(\mathbf{x}_i, \mathbf{y}_i) \\ cov(\mathbf{y}_i, \mathbf{x}_i) & cov(\mathbf{y}_i) \end{array} \right) = \boldsymbol{\Phi} = \left(\begin{array}{cc} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{12}^\top & \boldsymbol{\Phi}_{22} \end{array} \right)$ for $(\boldsymbol{\beta}, \boldsymbol{\Gamma}, \boldsymbol{\Phi}_x, \boldsymbol{\Psi})$?
- Show the parameters of the measurement model $(\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Omega})$ can be recovered from $\boldsymbol{\Sigma} = cov(\mathbf{d}_i)$.
- This means all the parameters can be recovered from $\boldsymbol{\Sigma}$.
- Break a big problem into two smaller ones.
- Develop *rules* for checking identifiability at each stage.
- Just look at the path diagram.

Copyright Information