Model Identification

$$\begin{aligned} \theta &\in \Theta \\ \mathbf{D} &\sim P_\theta = g(\theta) \in \mathcal{P} \\ g &: \Theta \rightarrow \mathcal{P} \end{aligned}$$

If the function g is one to one, then the model is identified.

Consistent Estimation is Impossible for a non-identified Model

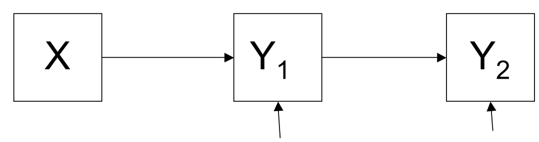
Suppose $\theta_1 \neq \theta_2$ with $P_{\theta_1} = P_{\theta_2}$

To prove model identification

- Show that the parameter can be recovered from the distribution of the observed data.
- In practice, recover it from the moments (usually, the covariance matrix).

- Sometimes, only a function of the parameter is identified.
- A model can be identified at some points but not others.

Remember the example



$$\boldsymbol{\theta} = (\gamma, \beta, \phi, \psi_1, \psi_2)$$

$$\boldsymbol{\Sigma} = \left[\begin{array}{ccc} \phi & \gamma\phi & \beta\gamma\phi \\ \gamma\phi & \gamma^2\phi + \psi_1 & \beta(\gamma^2\phi + \psi_1) \\ \beta\gamma\phi & \beta(\gamma^2\phi + \psi_1) & \beta^2(\gamma^2\phi + \psi_1) + \psi_2 \end{array} \right]$$

Solve the identifying equations

$$\boldsymbol{\Sigma} = \left[\begin{array}{ccc} \sigma_{1,1} & \sigma_{1,2} & \sigma_{1,3} \\ & \sigma_{2,2} & \sigma_{2,3} \\ & & \sigma_{3,3} \end{array}\right] = \left[\begin{array}{ccc} \phi & \gamma\phi & \beta\gamma\phi \\ & \gamma^2\phi + \psi_1 & \beta(\gamma^2\phi + \psi_1) \\ & & \beta^2(\gamma^2\phi + \psi_1) + \psi_2 \end{array}\right]$$

$$\text{For } \boldsymbol{\theta} = (\gamma, \beta, \phi, \psi_1, \psi_2)$$

An Identified model can be

- Just-Identified (saturated): Same number of parameters and identifying equations
- Over-identified: More equations than unknowns

Identification Rules

- A necessary condition for most Models
- Sufficient conditions for models with just observed variables
- Sufficient conditions for measurement models (factor analysis)
- Sufficient conditions for combined models

Parameter Count

- The number of parameters must be no more than the number of unique elements in the covariance matrix of the observed variables.
- Necessary for most models
- If this condition fails, seek to prove the model is *not* identified.

Observed variable models

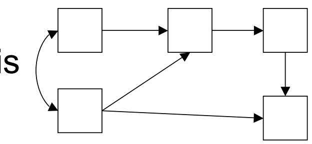
$$\mathbf{Y} = \boldsymbol{\beta}\mathbf{Y} + \boldsymbol{\Gamma}\mathbf{X} + \boldsymbol{\zeta}$$

Identified if $Cov(\mathbf{X}, \boldsymbol{\zeta}) = \mathbf{0}$ and

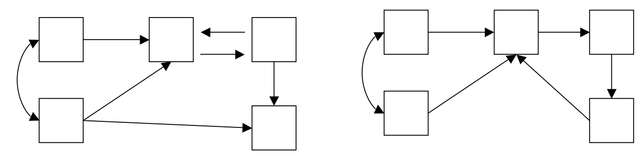
- $\boldsymbol{\beta} = \mathbf{0}$ (Regression model), or
- Model is Recursive, and $Var(\boldsymbol{\zeta})$ is diagonal

Recursive means no Feedback Loops

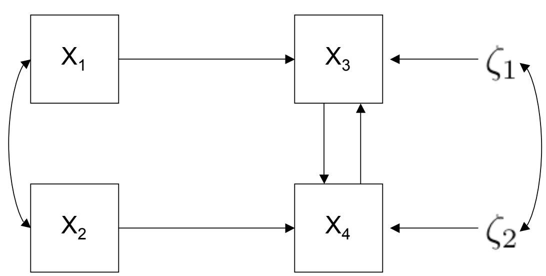
- Like this



- Not this

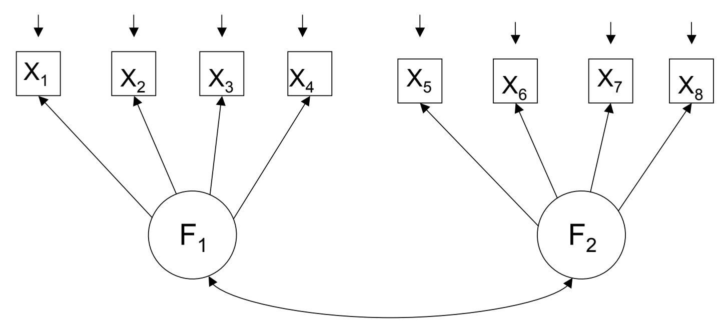


A just-identified, Non-recursive Model



Measurement Models (Confirmatory Factor Analysis)

$$\mathbf{X} = \boldsymbol{\Lambda}\mathbf{F} + \mathbf{e}$$



Rules for Confirmatory Factor Analysis

- Three-indicator rules
- Two-indicator rules

"Pick a Scale" for each factor

- F is obesity
- X_1 is percent body fat from immersion
- X_2 is triceps skin fold
- X_3 is Body Mass Index

$$\begin{aligned} X_1 &= \lambda_1 F + e_1 \\ X_2 &= \lambda_2 F + e_2 \\ X_3 &= \lambda_3 F + e_3 \end{aligned} \longrightarrow \begin{aligned} X_1 &= F + e_1 \\ X_2 &= \lambda_2 F + e_2 \\ X_3 &= \lambda_3 F + e_3 \end{aligned}$$

Three-indicator rules

- At least three variables per factor
- Each variable caused by only one factor
- Errors uncorrelated with factors and with each other
- Pick a scale for each factor
- No restrictions on Var(**F**)

Two-Indicator Rules

- At least two variables per factor
- Each variable caused by only one factor
- Errors uncorrelated with factors and with each other
- Pick a scale for each factor
- At least one non-zero off-diagonal element in each row of Var(**F**)

Combined Models

- Consider the latent part of the model as a model for observed variables. Verify identification.
- Consider the measurement part of the model as a confirmatory factor analysis, ignoring structure in Var(**F**). Verify identification.

Fixing up non-identified models

- Negotiation
- Deeper study may be rewarded
- “Model” is not identified.

- Consider identification before collecting data!

Further Issues

- Normality
- Numerical problems
- Sample size
- Categorical data

Software

- LISREL, EQS, RAM
- Mplus
- Stata
- R
- AMOS (Graphical interface)
- SAS proc calis