Two-stage Proofs of Identifiability[1]

STA431 Winter/Spring 2013

[1]See last slide for copyright information.

Overview

The two-stage model: $V(\mathbf{D}_i) = \boldsymbol{\Sigma}$

$$
\begin{aligned}
\mathbf{Y}_i &= \boldsymbol{\beta}\mathbf{Y}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \left(\begin{array}{c} \mathbf{X}_i \\ \mathbf{Y}_i \end{array} \right) \\
\mathbf{D}_i &= \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i
\end{aligned}
$$

- $V(\mathbf{X}_i) = \boldsymbol{\Phi}_{11}$, $V(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi}$
- $V(\mathbf{F}_i) = V\left(\begin{array}{c} \mathbf{X}_i \\ \mathbf{Y}_i \end{array} \right) = \boldsymbol{\Phi} = \left(\begin{array}{cc} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{12}^\prime & \boldsymbol{\Phi}_{22} \end{array} \right)$
- $V(\mathbf{e}_i) = \boldsymbol{\Omega}$

Identify parameter matrices in two steps

It does not matter which one you do first.

1. *Measurement model*: Show $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$ can be recovered from $\boldsymbol{\Sigma} = V(\mathbf{D}_i)$.
2. *Latent model*: Show $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}_{11}$ and $\boldsymbol{\Psi}$ can be recovered from $\boldsymbol{\Phi}$.

This means the parameters of the latent model are recovered from $\boldsymbol{\Sigma}$ as well.

- $\mathbf{D}_i = \boldsymbol{\Lambda}\mathbf{F}_i + \mathbf{e}_i$

  $V(\mathbf{F}_i) = \boldsymbol{\Phi}$, $V(\boldsymbol{e}_i) = \boldsymbol{\Omega}$
- $\mathbf{Y}_i = \boldsymbol{\beta}\mathbf{Y}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\epsilon}_i$

  $V(\mathbf{X}_i) = \boldsymbol{\Phi}_{11}$, $V(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi}$

Parameter count rule

A necessary condition

If a model has more parameters than covariance structure equations, the parameter vector can be identifiable on at most a set of volume zero in the parameter space. This applies to all models.

All the following rules

- Are sufficient conditions for identifiability, not necessary.
- Assume that errors are independent of exogenous variables that are not errors.
- Assume all variables have expected value zero.

Measurement Model Rules

Factor Analysis

In these rules, latent variables that are not error terms are described as "factors."

Double Measurement Rule

- Each factor is measured twice.
- All factor loadings equal one.
- There are two sets of measurements, set one and set two.
- Correlated measurement errors are allowed within sets of measurements, but not between sets.

Three-Variable Rule for Standardized Factors

- Errors are independent of one another.
- Each observed variable is caused by only one factor.
- The variance of each factor equals one.
- There are at least 3 variables with non-zero loadings per factor.
- The sign of one non-zero loading is known for each factor.

Factors may be correlated.

Three-Variable Rule for Unstandardized Factors

The parameters of a measurement model will be identifiable if

- Errors are independent of one another.
- Each observed variable is caused by only one factor.
- For each factor, at least one factor loading equals one.
- There are at least 2 additional variables with non-zero loadings per factor.

Factors may be correlated.

Two-Variable Rule for Standardized Factors

A factor with just two variables may be added to a measurement model whose parameters are identifiable, and the parameters of the combined model will be identifiable provided

- The errors for the two additional variables are independent of one another and of those already in the model.
- The two new variables are caused only by the new factor.
- The variance of the new factor equals one.
- Both new factor loadings are non-zero.
- The sign of one new loading is known.
- The new factor has a non-zero covariance with at least one factor already in the model.

Two-Variable Rule for Unstandardized Factors

A factor with just two variables may be added to a measurement model whose parameters are identifiable, and the parameters of the combined model will be identifiable provided

- The errors for the two additional variables are independent of one another and of those already in the model.
- The two new variables are caused only by the new factor.
- At least one new factor loading equals one.
- The other new factor loading is non-zero.
- The sign of one new loading is known.
- The new factor has a non-zero covariance with at least one factor already in the model.

Four-variable Two-factor Rule

The parameters of a measurement model with two factors and four observed variables will be identifiable provided

- All errors are independent of one another.
- All factor loadings are non-zero.
- For each factor, either the variance of the factor equals and the sign of one new loading is known, or at least one factor loading equals one.
- The covariance of the two factors does not equal zero.

Proof of the Four-variable Two-factor Rule
With standardized factors

The model equations are

$$\begin{aligned} D_1 &= \lambda_1 F_1 + e_1 \\ D_2 &= \lambda_2 F_1 + e_2 \\ D_3 &= \lambda_3 F_2 + e_4 \\ D_4 &= \lambda_4 F_2 + e_5, \end{aligned}$$

where all expected values are zero, $V(e_j) = \omega_j$ for $j = 1, \ldots, 4$, and

$$V \left[\begin{array}{c} F_1 \\ F_2 \end{array} \right] = \left[\begin{array}{cc} 1 & \phi_{12} \\ \phi_{12} & 1 \end{array} \right].$$

Also suppose no loadings $= 0$ and $\lambda_1 > 0$, $\lambda_3 > 0$.

Covariance matrix

For the 4-variable 2-factor problem

$$
\begin{aligned}
D_1 &= \lambda_1 F_1 + e_1 \\
D_2 &= \lambda_2 F_1 + e_2 \\
D_3 &= \lambda_3 F_2 + e_4 \\
D_4 &= \lambda_4 F_2 + e_5,
\end{aligned}
$$

$$
\boldsymbol{\Sigma} = \begin{array}{c|cccc}
 & D_1 & D_2 & D_3 & D_4 \\ \hline
D_1 & \lambda_1^2 + \omega_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3\phi_{12} & \lambda_1\lambda_4\phi_{12} \\
D_2 & & \lambda_2^2 + \omega_2 & \lambda_2\lambda_3\phi_{12} & \lambda_2\lambda_4\phi_{12} \\
D_3 & & & \lambda_3^2 + \omega_3 & \lambda_3\lambda_4 \\
D_4 & & & & \lambda_4^2 + \omega_4
\end{array}
$$

Using the assumption that $\lambda_1 > 0$ and $\lambda_3 > 0$

$$\boldsymbol{\Sigma} = \begin{array}{c|cccc} & D_1 & D_2 & D_3 & D_4 \\ \hline D_1 & \lambda_1^2+\omega_1 & \lambda_1\lambda_2 & \lambda_1\lambda_3\phi_{12} & \lambda_1\lambda_4\phi_{12} \\ D_2 & & \lambda_2^2+\omega_2 & \lambda_2\lambda_3\phi_{12} & \lambda_2\lambda_4\phi_{12} \\ D_3 & & & \lambda_3^2+\omega_3 & \lambda_3\lambda_4 \\ D_4 & & & & \lambda_4^2+\omega_4 \end{array}$$

$$\frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} = \frac{\lambda_1^2\lambda_2\lambda_3\phi_{12}}{\lambda_2\lambda_3\phi_{12}} = \lambda_1^2$$

$$\Rightarrow \quad \lambda_1 = \sqrt{\frac{\sigma_{12}\sigma_{13}}{\sigma_{23}}}$$

Similarly, $\lambda_3 = \sqrt{\frac{\sigma_{34}\sigma_{23}}{\sigma_{24}}}$, and the rest is easy.

Please don't do both!

Don't set the variance *and* a factor loading to one!

- Setting the variance of factors to one looks arbitrary, but it's really a smart re-parameterization.
- Setting one loading per factor to one also is a smart re-parameterization.
- It's smart because the resulting models impose the *same restrictions on the covariance that the original model does.*
- And, the *meanings* of the parameters have a clear connection to the meanings of the parameters of the original model.
- But if you do both, it's a mess. Most or all of the meaning is lost.
- And you put an *extra* restriction on $\boldsymbol{\Sigma}$ that is not implied by the original model.

Combination Rule

Suppose that the parameters of two measurement models are identifiable by any of the rules above. The two models may be combined into a single model provided that the error terms of the first model are independent of the error terms in the second model. The additional parameters of the combined model are the covariances between the two sets of factors, and these are all identifiable.

Cross-over Rule

Suppose that

- The parameters of a measurement models are identifiable, and
- For each factor there is at least one observable variable that is caused only by that factor (with a non-zero factor loading).

Then any number of new observable variables may be added to the model and the result is a model whose parameters are all identifiable, provided that

- The error terms associated with the new variables are independent of the error terms in the existing model.

Each new variable may be caused by any or all of the factors, potentially resulting in a cross-over pattern in the path diagram. The error terms associated with the new set of variables may be correlated with one another.

Error-Free Rule

A vector of observable variables may be added to the factors of a measurement model whose parameters are identifiable. Suppose that

- The new observable variables are independent of the errors in the measurement model, and
- For each factor in the measurement model there is at least one observable variable that is caused only by that factor (with a non-zero factor loading).

Then the parameters of a new measurement model, where some of the variables are assumed to be measured without error, are identifiable. The practical consequence is that variables assumed to be measured without error may be included in the latent component of a structural equation model, provided that the measurement model for the other variables has identifiable parameters.

Latent Model Rules

- $\mathbf{Y}_i = \boldsymbol{\beta}\mathbf{Y}_i + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\epsilon}_i$
- Here, identifiability means that the parameters involved are functions of $V(\mathbf{F}) = \boldsymbol{\Phi}$.

Regression Rule

Someimes called the Null Beta Rule

Suppose

- No endogenous variables cause other endogenous variables.

- $\mathbf{Y}_i = \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\epsilon}_i$
- Of course $C(\mathbf{X}_i, \boldsymbol{\epsilon}_i) = \mathbf{0}$, always.
- $\boldsymbol{\Psi} = V(\boldsymbol{\epsilon}_i)$ need not be diagonal.

Then $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ are identifiable.

Acyclic Rule

Parameters of the Latent Variable Model are identifiable if the model is acyclic (no feedback loops through straight arrows) and $V(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$ has the following block diagonal structure.

- Organize the variables into sets. Set 0 consists of the exogenous variables.
- For $j = 1, \ldots, k$, each variable in set j is caused by at least one variable in set $j-1$, and also possibly by variables in earlier sets.
- Error terms for the variables in a set may (or may not) have non-zero covariances.
- All other covariances between error terms are zero.

Copyright Information