Including Measurement Error in the Regression Model: A First Try[1]

STA2053 Fall 2022

[1] See last slide for copyright information.

Overview

Moments and Moment Structure Equations

Model: $d \sim P_\theta$

- *Moments* of a distribution are quantities such $E(X)$, $E(Y^2)$, $Var(X)$, $E(X^2Y^2)$, $Cov(X,Y)$, and so on.
- *Moment structure equations* are a set of equations expressing moments of the distribution of the observable data in terms of the model parameters: $m = g(\theta)$
- If there are just variances and covariances, the moment structure equations are called *covariance structure equations*.

Important process

- Calculate the moments of the distribution: $m = g(\theta)$.
- Solve the moment structure equations for the parameters: $\theta = g^{-1}(m)$.
- Method of Moments: $\widehat{\theta} = g^{-1}(\widehat{m})$.
- By LLN and Continuous mapping, $\widehat{\theta} \stackrel{p}{\rightarrow} \theta$
- Showing that consistent estimation is possible.

Multivariate multiple regression

With just observed variables

$$\mathbf{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{x}_i + \boldsymbol{\epsilon}_i$$

where

$\mathbf{y}_i$ is an $q \times 1$ random vector of observable response variables, so the regression can be multivariate; there are q response variables.

$\boldsymbol{\beta}_0$ is a $q \times 1$ vector of unknown constants, the intercepts for the q regression equations. There is one for each response variable.

$\mathbf{x}_i$ is a $p \times 1$ observable random vector; there are p explanatory variables. $\mathbf{x}_i$ has expected value $\boldsymbol{\mu}_x$ and variance-covariance matrix $\boldsymbol{\Phi}$, a $p \times p$ symmetric and positive definite matrix of unknown constants.

$\boldsymbol{\beta}_1$ is a $q \times p$ matrix of unknown constants. These are the regression coefficients, with one row for each response variable and one column for each explanatory variable.

$\boldsymbol{\epsilon}_i$ is the error term of the latent regression. It is a $q \times 1$ random vector with expected value zero and variance-covariance matrix $\boldsymbol{\Psi}$, a $q \times q$ symmetric and positive definite matrix of unknown constants. $\boldsymbol{\epsilon}_i$ is independent of $\mathbf{x}_i$.

$$\boldsymbol{\theta} = (\boldsymbol{\beta}_0, \boldsymbol{\mu}_x, \boldsymbol{\Phi}, \boldsymbol{\beta}_1, \boldsymbol{\Psi})$$

$\mathbf{d}_i = \left(\frac{\mathbf{x}_i}{\mathbf{y}_i} \right)$: Write $E(\mathbf{d}_i)$ and $cov(\mathbf{d}_i)$ as partitioned matrices

$$\boldsymbol{\mu} = \left(\frac{E(\mathbf{x}_i)}{E(\mathbf{y}_i)} \right) = \left(\frac{\boldsymbol{\mu}_1}{\boldsymbol{\mu}_2} \right)$$

and

$$\boldsymbol{\Sigma} = cov \left(\frac{\mathbf{x}_i}{\mathbf{y}_i} \right) = \left(\begin{array}{c|c} cov(\mathbf{x}_i) & cov(\mathbf{x}_i, \mathbf{y}_i) \\ \hline cov(\mathbf{x}_i, \mathbf{y}_i)^\top & cov(\mathbf{y}_i) \end{array} \right) = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{12}^\top & \boldsymbol{\Sigma}_{22} \end{array} \right)$$

$$\mathbf{m} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22})$$

Moment structure equations

Based on $\mathbf{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{x}_i + \boldsymbol{\epsilon}_i$

$$\boldsymbol{\theta} = (\boldsymbol{\beta}_0, \boldsymbol{\mu}_x, \boldsymbol{\Phi}, \boldsymbol{\beta}_1, \boldsymbol{\Psi})$$
$$\mathbf{m} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22})$$

$$\begin{aligned}
\boldsymbol{\mu}_1 &= \boldsymbol{\mu}_x \\
\boldsymbol{\mu}_2 &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{\mu}_x \\
\boldsymbol{\Sigma}_{11} &= \boldsymbol{\Phi} \\
\boldsymbol{\Sigma}_{12} &= \boldsymbol{\Phi} \boldsymbol{\beta}_1^\top \\
\boldsymbol{\Sigma}_{22} &= \boldsymbol{\beta}_1 \boldsymbol{\Phi} \boldsymbol{\beta}_1^\top + \boldsymbol{\Psi}.
\end{aligned}$$

Solve moment structure equations for the parameters

$\theta = g^{-1}(m)$

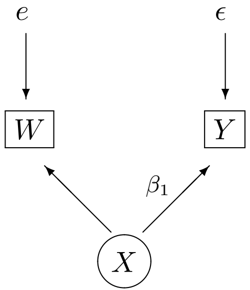
$$
\begin{aligned}
\boldsymbol{\beta}_0 &= \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{12}^\top \boldsymbol{\Sigma}_{11}^{-1} \, \boldsymbol{\mu}_1 \\
\boldsymbol{\mu}_x &= \boldsymbol{\mu}_1 \\
\boldsymbol{\Phi} &= \boldsymbol{\Sigma}_{11} \\
\boldsymbol{\beta}_1 &= \boldsymbol{\Sigma}_{12}^\top \boldsymbol{\Sigma}_{11}^{-1} \\
\boldsymbol{\Psi} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{12}^\top \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}
\end{aligned}
$$

- Just put hats on everything to get MOM estimates.
- Same as the MLEs in this case by invariance.

But let's admit it

In most applications, the explanatory variables are measured with error.

A first try at including measurement error in the explanatory variable



$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_i + \epsilon_i \\
W_i &= X_i + e_i,
\end{aligned}
$$

Observable data are the pairs (W_i, Y_i) for $i = 1, \ldots, n$.
Try to fit the true model.

Details

Make everything normal for simplicity

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

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \epsilon_i \\ W_i &= \nu + X_i + e_i, \end{aligned}$$

where

- X_i is normally distributed with mean μ_x and variance $\phi > 0$
- ϵ_i is normally distributed with mean zero and variance $\psi > 0$
- e_i is normally distributed with mean zero and variance $\omega > 0$
- X_i, e_i, ϵ_i are all independent.

Observable data are the pairs (W_i, Y_i) for $i = 1, \ldots, n$.

Model implies that the (W_i, Y_i) are independent bivariate normal

$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$
$W_i = \nu + X_i + e_i$

with

$$E \left(\begin{array}{c} W_i \\ Y_i \end{array} \right) = \boldsymbol{\mu} = \left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right) = \left(\begin{array}{c} \nu + \mu_x \\ \beta_0 + \beta_1 \mu_x \end{array} \right),$$

and variance-covariance matrix

$$cov \left(\begin{array}{c} W_i \\ Y_i \end{array} \right) = \boldsymbol{\Sigma} = [\sigma_{i,j}] = \left(\begin{array}{cc} \phi + \omega & \beta_1 \phi \\ \beta_1 \phi & \beta_1^2 \phi + \psi \end{array} \right).$$

Fit with maximum likelihood?

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-np/2} \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}) \right\}$$

Big problem revealed by the moment structure equations

$m = g(\theta)$. Solve to obtain $\theta = g^{-1}(m)$

$$\boldsymbol{\theta} = (\beta_0, \beta_1, \mu_x, \phi, \psi, \nu, \omega)$$

$$\begin{aligned}
\mu_1 &= \mu_x + \nu \\
\mu_2 &= \beta_0 + \beta_1 \mu_x \\
\sigma_{1,1} &= \phi + \omega \\
\sigma_{1,2} &= \beta_1 \phi \\
\sigma_{2,2} &= \beta_1^2 \phi + \psi
\end{aligned}$$

It is impossible to solve these five equations uniquely for the seven model parameters.

A numerical example

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} \mu_x + \nu \\ \beta_0 + \beta_1 \mu_x \end{pmatrix}$$

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \phi + \omega & \beta_1 \phi \\ & \beta_1^2 \phi + \psi \end{pmatrix}$$

	μ_x	β_0	ν	β_1	ϕ	ω	ψ
$\boldsymbol{\theta}_1$	0	0	0	1	2	2	3
$\boldsymbol{\theta}_2$	0	0	0	2	1	3	1

Both $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ imply a bivariate normal distribution with mean zero and covariance matrix

$$\boldsymbol{\Sigma} = \left[\begin{array}{cc} 4 & 2 \\ 2 & 5 \end{array} \right],$$

and thus the same distribution of the sample data.

Parameter Identifiability

- No matter how large the sample size, it will be impossible to decide between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, because they imply exactly the same probability distribution of the observable data.
- The problem here is that the parameters of the regression are not identifiable.

Definition of Identifiability

One of many

If the probability distribution of the observable data is a one-to-one function of the parameter (vector), the parameter (vector) is said to be identifiable.

- The probability distribution of the data is always a function of the parameter.
- If the parameter is also a function of the probability distribution, the function is one-to-one and the parameter is identifiable.
- If the parameter can somehow be recovered from the distribution of the data, it is identifiable.
- If two different parameter values yield the same distribution of the data, the parameter is not identifiable.
- If the parameter is not knowable from the distribution of the data, there will be trouble with estimation.

Theorem

If the parameter vector is not identifiable, consistent estimation is impossible.

- Let $\theta_1 \neq \theta_2$ but $P_{\theta_1}(d_n) = P_{\theta_2}(d_n)$ for all n.
- So the distribution of $T_n = T_n(D_1, \ldots, D_n)$ is identical for θ_1 and θ_2.
- Suppose T_n is a consistent estimator of θ.
- Then $T_n \stackrel{p}{\rightarrow} \theta_1$ and $T_n \stackrel{p}{\rightarrow} \theta_2$.

- Impossible.

The zipper example

Identification of parameters from the moments

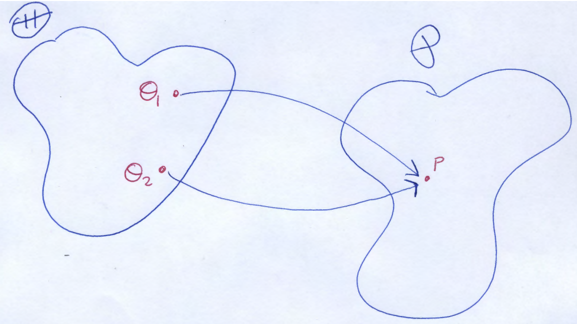
$$
\begin{tikzcd}
 & \theta \arrow[dr, "P_\theta = h(\theta)"] & \\
m \arrow[ur, "\theta = g^{-1}(m)"] & & P_\theta \arrow[ll, "m = g(\theta)"]
\end{tikzcd}
$$

- $m = g(\theta)$ are the moment structure equations.
- $\theta = g^{-1}(m)$ is the solution of the moment structure equations.
- In this course, parameters will be identified from $\mathbf{m} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (usually just $\boldsymbol{\Sigma}$), or not at all.

Non-identifiability

Parameter is identifiable if the probability distribution of the observable data is a one-to-one function of the parameter.

If two different parameter values yield the same distribution of the data, the parameter is not identifiable.



Identifiability of *functions* of the parameter vector

- If a function $g(\boldsymbol{\theta})$ can be recovered from the distribution of the observable data, that function of the parameter vector is said to be identifiable.
- This applies to individual parameters and subsets of the parameters.
- Frequently, not everything can be known, but informative *functions* of the parameter are knowable.

Some sample questions will be based on this model:

Let $W_i = X_i + e_i$, where

- $X_i \sim N(\mu_x, \phi)$
- $e_i \sim N(0, \omega)$
- X_i and e_i are independent.
- Only W_i is observable (X_i is a latent variable).

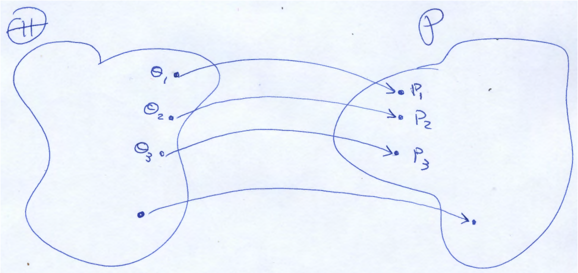
In the following questions, you may use the fact that the normal distribution corresponds uniquely to the pair (μ, σ^2).

1. What is the parameter vector $\boldsymbol{\theta}$?
2. What is the parameter space Θ?
3. What is the probability distribution of the observable data?
4. Give the moment structure equations.
5. Either prove that the parameter is identifiable, or show by an example that it is not. A simple numerical example is best.
6. Give two *functions* of the parameter vector that are identifiable.

A Useful Equivalent Definition of Identifiability

Equivalent to P_θ is a one-to-one function of θ

- Suppose a statistical model implies $\mathbf{D} \sim P_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta$. If no two points in Θ yield the same probability distribution, then the parameter $\boldsymbol{\theta}$ is said to be identifiable.
- That is, identifiability means that $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ implies $P_{\boldsymbol{\theta}_1} \neq P_{\boldsymbol{\theta}_2}$.



Pointwise identifiability

As opposed to global identifiability

- The parameter is said to be identifiable at a point $\boldsymbol{\theta}_0$ if no other point in Θ yields the same probability distribution as $\boldsymbol{\theta}_0$.
- That is, $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ implies $P_{\boldsymbol{\theta}} \neq P_{\boldsymbol{\theta}_0}$ for all $\boldsymbol{\theta} \in \Theta$.

If the parameter is identifiable at at every point in Θ, it is identifiable according to the earlier definitions.

Determining identifiability in practice

- In practice, identifiability means that the moment structure equations can be solved uniquely for the parameters.

$$
\begin{tikzcd}
 & \theta \arrow[dr, "P_\theta = h(\theta)"] & \\
m \arrow[ur, "\theta = g^{-1}(m)"] & & P_\theta \arrow[ll, "m = g(\theta)"]
\end{tikzcd}
$$

Proving identifiability

- You can explicitly solve the moment structure equations.
- You can use theorems.
- We will develop a collection of identifiability rules.
- These are really simple theorems about the existence of unique real solutions to equations.
- They are not well-known to mathematicians because they are too specific to be interesting.
- We will be able to look at a path diagram and verify that the parameters are identifiable. Usually.

Proving that a parameter is *not* identifiable

- You can carefully describe the set of points in the parameter space that yield the same mean and covariance matrix. It's a lot of work, even for small models.
- You can produce a numerical example of two different points that yield the same mean and covariance matrix. That settles it, but is can still be a lot of work for big models.
- You can use a big theorem.

Theorem

For us, the x variables are parameters and the y variables are moments.

Let

$$
\begin{array}{rcl}
y_1 & = & f_1(x_1, \ldots, x_p) \\
y_2 & = & f_2(x_1, \ldots, x_p) \\
\vdots & & \vdots \\
y_q & = & f_q(x_1, \ldots, x_p),
\end{array}
$$

If the functions $f_1, \ldots, f_q$ are analytic (posessing a Taylor expansion) and $p > q$, the set of points $(x_1, \ldots, x_p)$ where the system of equations has a unique solution occupies at most a set of volume zero in $\mathbb{R}^p$.

The Parameter Count Rule

For establishing non-identifiability

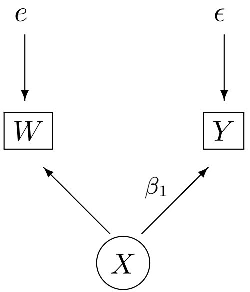
Suppose identifiability is to be decided based on a set of moment structure equations. If there are more parameters than equations, the set of points where the parameter vector is identifiable occupies a set of volume zero in the parameter space.

- Note that the empty set has volume zero.
- The parameter count rule is really a theorem about the existence of unique real solutions to systems of equations.
- The moment structure equations need to have derivatives and mixed partial derivatives of all orders, but they usually do.

Back to the example

Trying to include measurement error in the model

- Recall the first attempt to include measurement error in the model.



- There were five moment structure equations in seven unknown parameters.
- The model failed the parameter count rule.
- Game over.

Again: The Parameter Count Rule

Suppose identifiability is to be decided based on a set of moment structure equations. If there are more parameters than equations, the set of points where the parameter vector is identifiable occupies a set of volume zero in the parameter space.

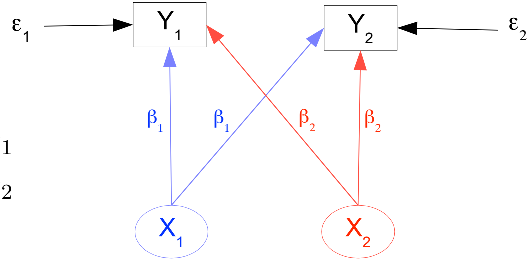
- So a necessary condition for parameter identifiability is that there be at least as many moment structure equations as parameters.
- There can be more equations than unknown parameters, and still no unique solution.
- There may be points in the parameter space where the parameter is identifiable, but if so, that set of points has volume zero.
- Failure of the parameter count rule means that it's impossible to identify the whole parameter vector.
- Useful functions of the parameters may be identifiable, maybe including what you really want to know.
- Maximum likelihood estimation depends on identifiability of the entire parameter vector (usually).

Example

To illustrate the parameter count rule.

There are two latent explanatory variables and two observable response variables.

$$
\begin{aligned}
Y_1 &= \beta_1 X_1 + \beta_2 X_2 + \epsilon_1 \\
Y_2 &= \beta_1 X_1 + \beta_2 X_2 + \epsilon_2
\end{aligned}
$$



where

- X_1, X_2, ϵ_1 and ϵ_2 are independent normal random variables with expected value zero, and
- $Var(X_1) = Var(X_2) = 1$, $Var(\epsilon_1) = \psi_1$ and $Var(\epsilon_2) = \psi_2$.
- Only Y_1 and Y_2 are observable.

The parameter vector is $\boldsymbol{\theta} = (\beta_1, \beta_2, \psi_1, \psi_2)$.

Calculate the covariance matrix of $(Y_1, Y_2)^\top$

Expected value is (zero, zero)

$$
\begin{aligned}
Y_1 &= \beta_1 X_1 + \beta_2 X_2 + \epsilon_1 \\
Y_2 &= \beta_1 X_1 + \beta_2 X_2 + \epsilon_2,
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \left(\begin{array}{cc} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_{2,2} \end{array} \right) \\
&= \left(\begin{array}{ll} \beta_1^2 + \beta_2^2 + \psi_1 & \beta_1^2 + \beta_2^2 \\ \beta_1^2 + \beta_2^2 & \beta_1^2 + \beta_2^2 + \psi_2 \end{array} \right)
\end{aligned}
$$

Covariance structure equations

$\boldsymbol{\theta} = (\beta_1, \beta_2, \psi_1, \psi_2)$

$$
\begin{aligned}
\sigma_{1,1} &= \beta_1^2 + \beta_2^2 + \psi_1 \\
\sigma_{1,2} &= \beta_1^2 + \beta_2^2 \\
\sigma_{2,2} &= \beta_1^2 + \beta_2^2 + \psi_2
\end{aligned}
$$

- Three equations in 4 unknowns, so the model fails.
- Parameter count rule does *not* say that a solution is impossible.
- It says that *the set of points in the parameter space where there is a unique solution (so the parameters are all identifiable) occupies a set of volume zero.*
- Are there any such points at all?

Try to solve for the parameters
$\boldsymbol{\theta} = (\beta_1, \beta_2, \psi_1, \psi_2)$

Covariance structure equations:

$$\begin{array}{rcl}\sigma_{1,1} & = & \beta_1^2 + \beta_2^2 + \psi_1 \\ \sigma_{1,2} & = & \beta_1^2 + \beta_2^2 \\ \sigma_{2,2} & = & \beta_1^2 + \beta_2^2 + \psi_2\end{array}$$

- $\psi_1 = \sigma_{1,1} - \sigma_{1,2}$
- $\psi_2 = \sigma_{2,2} - \sigma_{1,2}$
- So those *functions* of the parameter vector are identifiable.
- What about β_1 and β_2?

Can we solve for β_1 and β_2?

$\boldsymbol{\theta} = (\beta_1, \beta_2, \psi_1, \psi_2)$

$$
\begin{aligned}
\sigma_{1,1} &= \beta_1^2 + \beta_2^2 + \psi_1 \\
\sigma_{1,2} &= \beta_1^2 + \beta_2^2 \\
\sigma_{2,2} &= \beta_1^2 + \beta_2^2 + \psi_2
\end{aligned}
$$

- $\sigma_{1,2} = 0$ if and only if Both $\beta_1 = 0$ and $\beta_2 = 0$.
- The set of points where all four parameters can be recovered from the covariance matrix is *exactly* the set of points where the parameter vector is identifiable.
- It is

$$\{(\beta_1, \beta_2, \psi_1, \psi_2) : \beta_1 = 0, \beta_2 = 0, \psi_1 > 0, \psi_2 > 0\}$$

- A set of infinitely many points in $\mathbb{R}^4$
- A set of volume zero, as the theorem says.

Suppose $\beta_1^2+\beta_2^2 \neq 0$

This is the case "almost everywhere" in the parameter space.

The set of infinitely many points $\{(\beta_1, \beta_2, \psi_1, \psi_2)\}$ such that

- $\psi_1 = \sigma_{1,1} - \sigma_{1,2}$
- $\psi_2 = \sigma_{2,2} - \sigma_{1,2}$
- $\beta_1^2 + \beta_2^2 = \sigma_{1,2}$

Substitute back into

$$cov \left(\begin{array}{c} Y_1 \\ Y_2 \end{array} \right) = \left(\begin{array}{cc} \beta_1^2 + \beta_2^2 + \psi_1 & \beta_1^2 + \beta_2^2 \\ \beta_1^2 + \beta_2^2 & \beta_1^2 + \beta_2^2 + \psi_2 \end{array} \right)$$

And see they all produce the covariance matrix

$$\boldsymbol{\Sigma} = \left(\begin{array}{cc} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_{2,2} \end{array} \right)$$

And hence the same bivariate normal distribution of $(Y_1, Y_2)^\top$.

Why are there infinitely many points in this set?

$\{(\beta_1, \beta_2, \psi_1, \psi_2)\}$ such that

- $\psi_1 = \sigma_{1,1} - \sigma_{1,2}$
- $\psi_2 = \sigma_{2,2} - \sigma_{1,2}$
- $\beta_1^2 + \beta_2^2 = \sigma_{1,2} \neq 0$

Because $\beta_1^2 + \beta_2^2 = \sigma_{1,2}$ is the equation of a circle with radius $\sqrt{\sigma_{1,2}}$.

Maximum likelihood estimation

$\boldsymbol{\theta} = (\beta_1, \beta_2, \psi_1, \psi_2)$

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-np/2} \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{x}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{x}} - \boldsymbol{\mu}) \right\} \\ L(\boldsymbol{\Sigma}) &= |\boldsymbol{\Sigma}|^{-n/2} (2\pi)^{-n} \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + \overline{\mathbf{x}}^\top \boldsymbol{\Sigma}^{-1} \overline{\mathbf{x}} \right\} \end{aligned}$$

Can write likelihood as either $L(\boldsymbol{\Sigma})$ or $L(\boldsymbol{\Sigma}(\boldsymbol{\theta})) = L_2(\boldsymbol{\theta})$.

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \left(\begin{array}{ll} \beta_1^2 + \beta_2^2 + \psi_1 & \beta_1^2 + \beta_2^2 \\ \beta_1^2 + \beta_2^2 & \beta_1^2 + \beta_2^2 + \psi_2 \end{array} \right)$$

Likelihood $L_2(\boldsymbol{\theta})$ has non-unique maximum

- $L(\boldsymbol{\Sigma})$ has a unique maximum at $\boldsymbol{\Sigma} = \widehat{\boldsymbol{\Sigma}}$.
- For every positive definite $\boldsymbol{\Sigma}$ with $\sigma_{1,2} \neq 0$, there are infinitely many $\boldsymbol{\theta} \in \Theta$ which produce that $\boldsymbol{\Sigma}$, and have the same height of the likelihood.
- This includes $\widehat{\boldsymbol{\Sigma}}$.
- So there are infinitely many points $\boldsymbol{\theta}$ in Θ with $L_2(\boldsymbol{\theta}) = L(\widehat{\boldsymbol{\Sigma}})$.
- A circle in $\mathbb{R}^4$.

A circle in $\mathbb{R}^4$ where the likelihood is maximal

$\{(\beta_1, \beta_2, \psi_1, \psi_2)\} \subset \mathbb{R}^4$ such that

- $\psi_1 = \widehat{\sigma}_{1,1} - \widehat{\sigma}_{1,2}$
- $\psi_2 = \widehat{\sigma}_{2,2} - \widehat{\sigma}_{1,2}$
- $\beta_1^2 + \beta_2^2 = \widehat{\sigma}_{1,2}$

Some Questions

Remembering that if the model is true,

- $\psi_1 = \sigma_{1,1} - \sigma_{1,2}$
- $\psi_2 = \sigma_{2,2} - \sigma_{1,2}$
- $\beta_1^2 + \beta_2^2 = \sigma_{1,2}$

What would happen in the numerical search for $\widehat{\boldsymbol{\theta}}$ if ...

- $\widehat{\sigma}_{1,2} > \widehat{\sigma}_{1,1}$?
- $\widehat{\sigma}_{1,2} > \widehat{\sigma}_{2,2}$?
- $\widehat{\sigma}_{1,2} < 0$?

These could not *all* happen, but one of them could. When numerical maximum likelihood search leaves the parameter space, it may indicate that the model is incorrect. Or it might be just a bad starting value.

Testing hypotheses about $\boldsymbol{\theta}$

Some hypotheses are testable if the model is true, but direct likelihood ratio tests are out. All the theory depends on a unique maximum.

Remember,

$$cov \left(\begin{array}{c} Y_1 \\ Y_2 \end{array} \right) = \left(\begin{array}{cc} \beta_1^2 + \beta_2^2 + \psi_1 & \beta_1^2 + \beta_2^2 \\ \beta_1^2 + \beta_2^2 & \beta_1^2 + \beta_2^2 + \psi_2 \end{array} \right)$$

- How would you test $H_0 : \beta_1 = \beta_2 = 0$?
- If you did a large-sample likelihood ratio test, what would the degrees of freedom be?

Lessons from this example

- A parameter may be identifiable at some points but not others.
- Identifiability at infinitely many points is possible even if there are more unknowns than equations. But this can only happen on a set of volume zero.
- Some parameters and functions of the parameters may be identifiable even when the whole parameter vector is not.
- Lack of identifiability can produce multiple maxima of the likelihood function – even infinitely many.
- A model whose parameter vector is not identifiable may still be falsified by empirical data.
- Numerical maximum likelihood search may leave the parameter space. This may be a sign that the model is false. It can happen when the parameter is identifiable, too.
- Some hypotheses may be testable when the parameter is not identifiable, but these will be hypotheses about functions of the parameter that are identifiable in the part of the parameter space where the null hypothesis is true. $H_0 : \beta_1 = \beta_2 = 0$

Copyright Information