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

STA431 Winter/Spring 2015

[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 data in terms of the model parameters. $m = g(\theta)$
- If the moments involved are limited to variances and covariances, the moment structure equations are called *covariance structure equations*.

Important process

- Calculate the moments of the distribution (usually means, variances and covariances) in terms of the model parameters, obtaining a system of moment structure equations. $m = g(\theta)$
- Solve the moment structure equations for the parameters, expressing the parameters in terms of the moments. $\theta = g^{-1}(m)$
- Method of Moments: $\widehat{\theta} = g^{-1}(\widehat{m})$
- By SLLN and Continuous mapping, $\widehat{\theta} \stackrel{a.s.}{\rightarrow} \theta$
- So even if we're not going to use the Method of Moments, solving $\theta = g^{-1}(m)$ shows that consistent estimation is possible.

Recall multivariate multiple regression

$$\mathbf{Y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_i + \boldsymbol{\epsilon}_i$$

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 an $q \times 1$ multivariate normal 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 $\mathbf{X}_i$.

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

Data vectors are multivariate normal

$$\mathbf{D}_i = \left(\frac{\mathbf{X}_i}{\mathbf{Y}_i} \right)$$

- $\mathbf{D}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Write $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as partitioned matrices.

Write $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as partitioned matrices

$$\boldsymbol{\mu} = \left(\begin{array}{c} E(\mathbf{X}_i) \\ \hline E(\mathbf{Y}_i) \end{array} \right) = \left(\begin{array}{c} \boldsymbol{\mu}_1 \\ \hline \boldsymbol{\mu}_2 \end{array} \right)$$

and

$$\boldsymbol{\Sigma} = V\left(\begin{array}{c} \mathbf{X}_i \\ \hline \mathbf{Y}_i \end{array} \right) = \left(\begin{array}{c|c} V(\mathbf{X}_i) & C(\mathbf{X}_i, \mathbf{Y}_i) \\ \hline C(\mathbf{X}_i, \mathbf{Y}_i)^\top & V(\mathbf{Y}_i) \end{array} \right) = \left(\begin{array}{c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \hline \boldsymbol{\Sigma}_{12}^\top & \boldsymbol{\Sigma}_{22} \end{array} \right)$$

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

Moment structure equations

Based on $\mathbf{D}_i = (\mathbf{X}_i^\top | \mathbf{Y}_i^\top)^\top$ with $\mathbf{Y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_i + \boldsymbol{\epsilon}_i$

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

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

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

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 just 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

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

Big problem revealed by the moment structure equations

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

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

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

Impossible to solve the moment structure equations for the parameters

- Even with perfect knowledge of the probability distribution of the data (and for the multivariate normal, that means knowing $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, period), it would be impossible to know the model parameters.
- All data can ever tell you is the approximate distribution from which they come.
- So how could we expect to successfully *estimate* $\boldsymbol{\theta}$ based on sample data?

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*.
- The model parameters cannot be recovered from the distribution of the sample data.
- And all you can ever learn from sample data is the distribution from which it comes.
- So there will be problems using the sample data for estimation and inference.
- This is true even when *the model is completely correct.*

Definitions

Connected to parameter identifiability

- A *Statistical Model* is a set of assertions that partly specify the probability distribution of a set of observable data.
- 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 $P_{\boldsymbol{\theta}_1} \neq P_{\boldsymbol{\theta}_2}$.
- On the other hand, if there exist distinct $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in Θ with $P_{\boldsymbol{\theta}_1} = P_{\boldsymbol{\theta}_2}$, the parameter $\boldsymbol{\theta}$ is *not identifiable.*

An equivalent definition of identifiability

Full proof of equivalence deferred for now

- If the parameter vector is a function of the probability distribution of the observable data, it is identifiable.
- That is, if the parameter vector can somehow be recovered from the distribution of the data, it is identifiable.
- If two different parameter values gave the same distribution of the data, this would be impossible because functions yield only one value.

Regression models with no measurement error

- The mean and covariance matrix are functions of the probability distribution (calculate expected values).
- We solved for all the parameters from the mean and covariance matrix.
- Therefore the parameters are a function of the probability distribution.
- Thus they are identifiable.

Identifiability is a big concept

- It means *knowability* of the parameters from the distribution of the data.
- We will do mathematical proofs that show whether certain information can be known.
- Call it the **algebra of the knowable**.

Theorem

If the parameter vector is not identifiable, consistent estimation for all points in the parameter space is impossible.

- Let $\theta_1 \neq \theta_2$ but $P_{\theta_1} = P_{\theta_2}$
- Suppose $T_n = T_n(D_1, \ldots, D_n)$ is a consistent estimator of θ for all $\theta \in \Theta$, in particular for θ_1 and θ_2.
- So the distribution of T_n is identical for θ_1 and θ_2.

Identifiability of *functions* of the parameter vector

If $g(\boldsymbol{\theta}_1) \neq g(\boldsymbol{\theta}_2)$ implies $P_{\boldsymbol{\theta}_1} \neq P_{\boldsymbol{\theta}_2}$ for all $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ in Θ, the function $g(\boldsymbol{\theta})$ is said to be identifiable.

Some sample questions will be based on this model:

Let $W = X + e$, where

- $X \sim N(\mu, \phi)$
- $e \sim N(0, \omega)$
- X and e are independent.
- Only W is observable (X is a latent variable).

How does this fit the definition of a *model*?

Sample questions

Let $W = X + e$, where

- $X \sim N(\mu, \phi)$
- $e \sim N(0, \omega)$
- X and e are independent.
- Only W is observable (X 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.

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 $\boldsymbol{\theta} \in \Theta$.
- Let $g(\boldsymbol{\theta})$ be a function of the parameter vector. If $g(\boldsymbol{\theta}_0) \neq g(\boldsymbol{\theta})$ implies $P_{\boldsymbol{\theta}_0} \neq P_{\boldsymbol{\theta}}$ for all $\boldsymbol{\theta} \in \Theta$, then the function $g(\boldsymbol{\theta})$ is said to be identifiable at the point $\boldsymbol{\theta}_0$.

If the parameter (or function of the parameter) is identifiable at at every point in Θ, it is identifiable according to the earlier definitions.

Local identifiability

The parameter is said to be *locally identifiable* at a point $\boldsymbol{\theta}_0$ if there is a neighbourhood of points surrounding $\boldsymbol{\theta}_0$, none of which yields the same probability distribution as $\boldsymbol{\theta}_0$.

If the parameter is identifiable at a point, it is locally identifiable there, but the converse is not true.

The Parameter Count Rule

A necessary but not sufficient condition for 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.

So a necessary condition for parameter identifiability is that there be at least as many moment structure equations as parameters.

Example

Two latent explanatory 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

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

- Three equations in 4 unknowns.
- 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:

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

- $\psi_1 = \sigma_{1,1} - \sigma_{1,2}$
- $\psi_2 = \sigma_{2,2} - \sigma_{1,2}$
- 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

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

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 $L(\boldsymbol{\Sigma}(\boldsymbol{\theta})) = L_2(\boldsymbol{\theta})$.

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

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

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

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. What would it mean?

Remember,

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

Could the maximum of the likelihood function be outside the parameter space?

Testing hypotheses about $\boldsymbol{\theta}$

It is possible. Remember, the model implies

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

But likelihood ratio tests are out. All the theory depends on a unique maximum.

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.

Copyright Information