Double Measurement Regression Part One: A small example[1]

STA 2101 Fall 2019

[1] See last slide for copyright information.

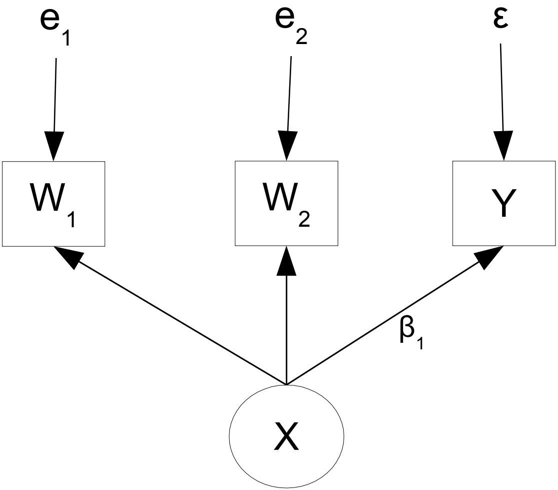
Seeking identifiability

We have seen that in simple regression, parameters of a model with measurement error are not identifiable.

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

- For example, X might be number of acres planted and Y might be crop yield.
- Plan the statistical analysis in advance.
- Take 2 independent measurements of the explanatory variable.
- Say, farmer's report and satellite photograph.

Double measurement
Of the explanatory variable



Model

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

$$\begin{array}{rcl} W_{i,1} & = & \nu_1 + X_i + e_{i,1} \\ W_{i,2} & = & \nu_2 + X_i + e_{i,2} \\ Y_i & = & \beta_0 + \beta_1 X_i + \epsilon_i, \end{array}$$

where

- X_i is normally distributed with mean μ_x and variance $\phi > 0$
- ϵ_i is normally distributed with mean zero and variance $\psi > 0$
- $e_{i,1}$ is normally distributed with mean zero and variance $\omega_1 > 0$
- $e_{i,2}$ is normally distributed with mean zero and variance $\omega_2 > 0$
- $X_i, e_{i,1}, e_{i,2}$ and ϵ_i are all independent.

Does this model pass the test of the Parameter Count Rule?

$$
\begin{aligned}
W_{i,1} &= \nu_1 + X_i + e_{i,1} \\
W_{i,2} &= \nu_2 + X_i + e_{i,2} \\
Y_i &= \beta_0 + \beta_1 X_i + \epsilon_i,
\end{aligned}
$$

$\boldsymbol{\theta} = (\nu_1, \nu_2, \beta_0, \mu_x, \beta_1, \phi, \psi, \omega_1, \omega_2)$: 9 parameters.

- Three expected values, three variances and three covariances: 9 moments.
- Yes. There are nine moment structure equations in nine unknown parameters. Identifiability is possible, but not guaranteed.

What is the distribution of the sample data?

Calculate the moments as a function of the model parameters

The model implies that the triples $\mathbf{D}_i = (W_{i,1}, W_{i,2}, Y_i)^\top$ are independent multivarate normal with

$$E(\mathbf{D}_i) = E\left(\begin{array}{c} W_{i,1} \\ W_{i,1} \\ Y_i \end{array}\right) = \left(\begin{array}{c} \mu_1 \\ \mu_2 \\ \mu_3 \end{array}\right) = \left(\begin{array}{c} \mu_x + \nu_1 \\ \mu_x + \nu_2 \\ \beta_0 + \beta_1 \mu_x \end{array}\right),$$

and variance covariance matrix $cov(\mathbf{D}_i) = \boldsymbol{\Sigma} =$

$$\left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array}\right) = \left(\begin{array}{ccc} \phi + \omega_1 & \phi & \beta_1 \phi \\ & \phi + \omega_2 & \beta_1 \phi \\ & & \beta_1^2 \phi + \psi \end{array}\right).$$

Are the parameters *in the covariance matrix* identifiable?

Six equations in five unknowns

$$\left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \color{blue}{\sigma_{23}} \\ & & \sigma_{33} \end{array}\right) = \left(\begin{array}{ccc} \phi+\omega_1 & \phi & \beta_1\phi \\ & \phi+\omega_2 & \color{blue}{\beta_1\phi} \\ & & \beta_1^2\phi+\psi \end{array}\right).$$

$$\begin{array}{rcl} \phi & = & \sigma_{12} \\ \omega_1 & = & \sigma_{11}-\sigma_{12} \\ \omega_2 & = & \sigma_{22}-\sigma_{12} \\ \beta_1 & = & \frac{\sigma_{13}}{\sigma_{12}} \\ \psi & = & \sigma_{33}-\beta_1^2\phi = \sigma_{33}-\frac{\sigma_{13}^2}{\sigma_{12}} \end{array}$$

Yes.

What about the expected values?

Model equations again:

$$
\begin{aligned}
W_{i,1} &= \nu_1 + X_i + e_{i,1} \\
W_{i,2} &= \nu_2 + X_i + e_{i,2} \\
Y_i &= \beta_0 + \beta_1 X_i + \epsilon_i,
\end{aligned}
$$

Expected values:

$$
\begin{aligned}
\mu_1 &= \nu_1 + \mu_x \\
\mu_2 &= \nu_2 + \mu_x \\
\mu_3 &= \beta_0 + \beta_1 \mu_x
\end{aligned}
$$

Four parameters appear only in the expected values: $\nu_1, \nu_2, \mu_x, \beta_0$.

- Three equations in four unknowns, even with β_1 identified from the covariance matrix.
- Parameter count rule applies.
- But we don't need it because these are linear equations.
- Re-parameterize.

Re-parameterize

$\mu_1 = \nu_1 + \mu_x \qquad \mu_2 = \nu_2 + \mu_x \qquad \mu_3 = \beta_0 + \beta_1\mu_x$

- Absorb $\nu_1, \nu_2, \mu_x, \beta_0$ into $\boldsymbol{\mu}$.
- Parameter was $\boldsymbol{\theta} = (\nu_1, \nu_2, \beta_0, \mu_x,\; \beta_1, \phi, \psi, \omega_1, \omega_2)$
- Now it's $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3,\; \beta_1, \phi, \psi, \omega_1, \omega_2)$.
- Dimension of the parameter space is now one less.
- We haven't lost much.
- Especially because the model was already re-parameterized.
- Of course there is measurement error in Y. Recall

$$
\begin{aligned}
Y &= \beta_0 + \beta_1 X + \epsilon \\
V &= \nu_0 + Y + e \\
&= \nu_0 + (\beta_0 + \beta_1 X + \epsilon) + e \\
&= (\nu_0 + \beta_0) + \beta_1 X + (\epsilon + e) \\
&= \beta_0' + \beta X + \epsilon'
\end{aligned}
$$

Re-parameterization

- Re-parameterization makes maximum likelihood possible.
- Otherwise the maximum is not unique and it's a mess.
- Estimate $\boldsymbol{\mu}$ with $\overline{\mathbf{D}}$ and it simply disappears from

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

- This step is so common it becomes silent.
- Model equations are often written without intercepts.
- It's more compact, and we don't have to look at parameters we can't estimate anyway.

Back to the covariance structure equations

$$\left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array} \right) = \left(\begin{array}{ccc} \phi+\omega_1 & \phi & \beta_1\phi \\ & \phi+\omega_2 & \beta_1\phi \\ & & \beta_1^2\phi+\psi \end{array} \right).$$

- Notice that the model dictates $\sigma_{1,3} = \sigma_{2,3}$.
- There are two ways to solve for β_1:
  $\beta_1 = \frac{\sigma_{13}}{\sigma_{12}}$ and $\beta_1 = \frac{\sigma_{23}}{\sigma_{12}}$.
- Does this mean the solution for β_1 is not "unique?"
- No; everything is okay. Because $\sigma_{1,3} = \sigma_{2,3}$, the two solutions are actually the same.
- If a parameter can be recovered from the moments in any way at all, it is identifiable.

Testing goodness of fit.

$$\left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array}\right) = \left(\begin{array}{ccc} \phi+\omega_1 & \phi & \beta_1\phi \\ & \phi+\omega_2 & \beta_1\phi \\ & & \beta_1^2\phi+\psi \end{array}\right).$$

- $\sigma_{1,3} = \sigma_{2,3}$ is a *model-induced constraint* upon $\boldsymbol{\Sigma}$.
- It's a testable null hypothesis.
- If rejected, the model is called into question.
- Likelihood ratio test comparing this model to a completely unrestricted multivariate normal model:

$$G^2 = -2\log\frac{L\left(\overline{\mathbf{D}}, \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)}{L(\overline{\mathbf{D}}, \widehat{\boldsymbol{\Sigma}})}$$

- Widely used even if the data are not normal.

The Reproduced Covariance Matrix

- $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})$ is called the *reproduced covariance matrix.*
- It is the covariance matrix of the observable data, written as a function of the model parameters and evaluated at the MLE.

$$\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}) = \left(\begin{array}{ccc} \widehat{\phi} + \widehat{\omega}_1 & \widehat{\phi} & \widehat{\beta}_1 \widehat{\phi} \\ & \widehat{\phi} + \widehat{\omega}_2 & \widehat{\beta}_1 \widehat{\phi} \\ & & \widehat{\beta}_1^2 \widehat{\phi} + \widehat{\psi} \end{array} \right)$$

- The reproduced covariance matrix obeys all model-induced constraints, while $\widehat{\boldsymbol{\Sigma}}$ does not.
- But if the model is right they should be close.
- This is a way to think about the likelihood ratio test for goodness of fit.

General pattern for testing goodness of fit

- Suppose there are k moment structure equations in p parameters, and all the parameters are identifiable.
- If $p < k$, call the parameter vector *over-identifiable*.
- Only needed p equations to solve for $\boldsymbol{\theta}$.
- Substituting the solutions (in terms of σ_{ij}) back into the unused equations would yield $k-p$ equality constraints on $\boldsymbol{\Sigma}$.
- Test those constraints with $G^2 = -2\log\frac{L\left(\overline{\mathbf{D}},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)}{L(\overline{\mathbf{D}},\widehat{\boldsymbol{\Sigma}})}$.
- $df = k-p$
- Don't need to actually derive the constraints unless asked – just count them.

With the same number of equations and parameters

- If the parameter is identifiable, call it *just identifiable*.
- Parameters are 1-1 with those of an unrestricted multivariate normal.
- Call the model "saturated."
- There are no equality constraints on $\boldsymbol{\Sigma}$.
- No likelihood ratio test because $G^2 = -2\log\frac{L\left(\overline{\mathbf{D}},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)}{L(\overline{\mathbf{D}},\widehat{\boldsymbol{\Sigma}})} = 0$.
- This is what happens in regression with all observed variables.

How to proceed

- Verify identifiability.
- If the model is over-identified, test goodness of fit.
- If it passes (non-significant), proceed.
- Now think of your model as the "full," or unrestricted model.
- Compared to some (even more) reduced model that is restricted by a null hypothesis like $\beta_1 = 0$.
- Fit the reduced model.
- Subtract G^2 goodness of fit statistics to test H_0.

Subtract goodness of fit statistics

G^2 tests the full model against the saturated model, and G_0^2 tests the reduced model against the saturated model.

$$
\begin{aligned}
G_0^2 - G^2 &= -2\log\frac{L\left(\overline{\mathbf{D}},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_0)\right)}{L(\overline{\mathbf{D}},\widehat{\boldsymbol{\Sigma}})} - -2\log\frac{L\left(\overline{\mathbf{D}},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)}{L(\overline{\mathbf{D}},\widehat{\boldsymbol{\Sigma}})} \\
&= -2\left(\log L\left(\overline{\mathbf{D}},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_0)\right) - \log L(\overline{\mathbf{D}},\widehat{\boldsymbol{\Sigma}}) - \log L\left(\overline{\mathbf{D}},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)\right. \\
&\qquad \left. + \log L(\overline{\mathbf{D}},\widehat{\boldsymbol{\Sigma}})\right) \\
&= -2\log\frac{L\left(\overline{\mathbf{D}},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_0)\right)}{L\left(\overline{\mathbf{D}},\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}})\right)}
\end{aligned}
$$

Further comments

- Models with non-identifiable parameters can imply testable equality constraints, but testing them is not automatic.
- Models can imply *inequality* constraints on $\boldsymbol{\Sigma}$, too.
- Using the solutions

$$\begin{aligned}
\phi &= \sigma_{12} \\
\omega_1 &= \sigma_{11} - \sigma_{12} \\
\omega_2 &= \sigma_{22} - \sigma_{12} \\
\beta_1 &= \frac{\sigma_{13}}{\sigma_{12}} \\
\psi &= \sigma_{33} - \beta_1^2\phi = \sigma_{33} - \frac{\sigma_{13}^2}{\sigma_{12}}
\end{aligned}$$

We get four inequality constraints.

Four inequality constraints on $\boldsymbol{\Sigma}$

$$\left(\begin{array}{ccc} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{array} \right) = \left(\begin{array}{ccc} \phi+\omega_1 & \phi & \beta_1\phi \\ & \phi+\omega_2 & \beta_1\phi \\ & & \beta_1^2\phi+\psi \end{array} \right).$$

$$\begin{array}{rcl} \phi & = & \sigma_{12} > 0 \\ \omega_1 & = & \sigma_{11} - \sigma_{12} > 0 \\ \omega_2 & = & \sigma_{22} - \sigma_{12} > 0 \\ \psi & = & \sigma_{33} - \frac{\sigma_{13}^2}{\sigma_{12}} > 0 \end{array}$$

Inequality constraints

- Inequality constraints arise because variances are positive.
- Or more generally, covariance matrices are positive definite.
- Could inequality constraints be violated in numerical maximum likelihood?
- Definitely.
- But only a little by sampling error if the model is correct.
- So maybe it's not so dumb to test hypotheses like $H_0 : \omega_1 = 0$.
- Since the model says $\omega_1 = \sigma_{11} - \sigma_{12}$ and $\sigma_{11} - \sigma_{12} > 0$ might not be true.

Copyright Information