Double Measurement Regression

STA431: Spring 2013

Double Measurement Regression: A Two-Stage Model

$$\begin{aligned}
\mathbf{Y}_i &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \left(\begin{array}{c} \mathbf{X}_i \\ \mathbf{Y}_i \end{array} \right) \\
\mathbf{D}_{i,1} &= \boldsymbol{\nu}_1 + \mathbf{F}_i + \mathbf{e}_{i,1} \\
\mathbf{D}_{i,2} &= \boldsymbol{\nu}_2 + \mathbf{F}_i + \mathbf{e}_{i,2}
\end{aligned}$$

Observable variables are $D_{i,1}$ and $D_{i,2}$: both p+q by 1

$$
\begin{array}{rcl}
\mathbf{Y}_i & = & \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i & = & \left(\begin{array}{c} \mathbf{X}_i \\ \mathbf{Y}_i \end{array} \right) \\
\mathbf{D}_{i,1} & = & \boldsymbol{\nu}_1 + \mathbf{F}_i + \mathbf{e}_{i,1} \\
\mathbf{D}_{i,2} & = & \boldsymbol{\nu}_2 + \mathbf{F}_i + \mathbf{e}_{i,2}
\end{array}
\qquad E(\mathbf{X}_i) = \boldsymbol{\mu}_x,
$$

$$
\begin{array}{rcl}
E(\mathbf{D}_{i,1}) & = & \left(\begin{array}{c} \boldsymbol{\mu}_{1,1} \\ \boldsymbol{\mu}_{1,2} \end{array} \right) = \left(\begin{array}{c} \boldsymbol{\nu}_{1,1} + E(\mathbf{X}_i) \\ \boldsymbol{\nu}_{1,2} + E(\mathbf{Y}_i) \end{array} \right) = \left(\begin{array}{c} \boldsymbol{\nu}_{1,1} + \boldsymbol{\mu}_x \\ \boldsymbol{\nu}_{1,2} + \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{\mu}_x \end{array} \right) \\
E(\mathbf{D}_{i,2}) & = & \left(\begin{array}{c} \boldsymbol{\mu}_{2,1} \\ \boldsymbol{\mu}_{2,2} \end{array} \right) = \left(\begin{array}{c} \boldsymbol{\nu}_{2,1} + E(\mathbf{X}_i) \\ \boldsymbol{\nu}_{2,2} + E(\mathbf{Y}_i) \end{array} \right) = \left(\begin{array}{c} \boldsymbol{\nu}_{2,1} + \boldsymbol{\mu}_x \\ \boldsymbol{\nu}_{2,2} + \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{\mu}_x \end{array} \right)
\end{array}
$$

- $\boldsymbol{\nu}$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\mu}_x$ parameters appear only in expected value, not covariance matrix.
- Even with knowledge of $\boldsymbol{\beta}_1$, 2(p+q) equations in 3(p+q) unknown parameters.
- Identifying the expected values and intercepts is hopeless.
- Re-parameterize, swallowing them into $\boldsymbol{\mu}$.

$$
\begin{aligned}
\mathbf{Y}_i &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \left(\begin{array}{c} \mathbf{X}_i \\ \mathbf{Y}_i \end{array} \right) \\
\mathbf{D}_{i,1} &= \boldsymbol{\nu}_1 + \mathbf{F}_i + \mathbf{e}_{i,1} \\
\mathbf{D}_{i,2} &= \boldsymbol{\nu}_2 + \mathbf{F}_i + \mathbf{e}_{i,2}
\end{aligned}
$$

$$
V(\mathbf{X}_i) = \boldsymbol{\Phi}_{11},\ V(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi},\ V(\mathbf{e}_{i,1}) = \boldsymbol{\Omega}_1,\ V(\mathbf{e}_{i,2}) = \boldsymbol{\Omega}_2,
$$

$$
\mathbf{X}_i,\ \boldsymbol{\epsilon}_i,\ \mathbf{e}_{i,1} \text{ and } \mathbf{e}_{i,2} \text{ independent.}
$$

The main idea is that $\mathbf{D}_1$ and $\mathbf{D}_2$ are independent measurements of **F**, perhaps at different times using different methods. Measurement errors may be correlated within occasions (even Between explanatory and response variables), but not between occasions.

$$
\begin{aligned}
\mathbf{Y}_i &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_i + \boldsymbol{\epsilon}_i \\
\mathbf{F}_i &= \begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} \\
\mathbf{D}_{i,1} &= \boldsymbol{\nu}_1 + \mathbf{F}_i + \mathbf{e}_{i,1} \\
\mathbf{D}_{i,2} &= \boldsymbol{\nu}_2 + \mathbf{F}_i + \mathbf{e}_{i,2}
\end{aligned}
$$

$$
V(\mathbf{X}_i) = \boldsymbol{\Phi}_{11},\ V(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi},\ V(\mathbf{e}_{i,1}) = \boldsymbol{\Omega}_1,\ V(\mathbf{e}_{i,2}) = \boldsymbol{\Omega}_2,
$$

$\mathbf{X}_i$, $\boldsymbol{\epsilon}_i$, $\mathbf{e}_{i,1}$ and $\mathbf{e}_{i,2}$ independent.

Stage One

$$
V(\mathbf{F}_i) = \boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{12}' & \boldsymbol{\Phi}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{11}\boldsymbol{\beta}_1' \\ \boldsymbol{\beta}_1 \boldsymbol{\Phi}_{11} & \boldsymbol{\beta}_1 \boldsymbol{\Phi}_{11}\boldsymbol{\beta}_1' + \boldsymbol{\Psi} \end{pmatrix}
$$

$\boldsymbol{\Phi}_{11}$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\Psi}$ can be recovered from $\boldsymbol{\Phi}$

The Measurement Model (Stage 2)

$$\begin{aligned} \mathbf{Y}_i &= \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{X}_i + \boldsymbol{\epsilon}_i \\ \mathbf{F}_i &= \begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} \\ \mathbf{D}_{i,1} &= \boldsymbol{\nu}_1 + \mathbf{F}_i + \mathbf{e}_{i,1} \\ \mathbf{D}_{i,2} &= \boldsymbol{\nu}_2 + \mathbf{F}_i + \mathbf{e}_{i,2} \end{aligned}$$

$$\boldsymbol{\Sigma} = V \begin{pmatrix} \mathbf{D}_{i,1} \\ \mathbf{D}_{i,2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Phi} + \boldsymbol{\Omega}_1 & \boldsymbol{\Phi} \\ \boldsymbol{\Phi} & \boldsymbol{\Phi} + \boldsymbol{\Omega}_2 \end{pmatrix}$$

$\boldsymbol{\Phi}$, $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ can easily be recovered from $\boldsymbol{\Sigma}$

All the parameters in the covariance matrix are identifiable

- $\mathbf{\Phi}$, $\mathbf{\Omega}_1$ and $\mathbf{\Omega}_2$ may be recovered from $\mathbf{\Sigma}$
- $\mathbf{\Phi}_{11}$, $\boldsymbol{\beta}_1$ and $\mathbf{\Psi}$ may be recovered from $\mathbf{\Phi}$

- Correlated measurement error within sets is allowed (a big plus), because it's reality
- Correlated measurement error between sets must be ruled out by careful data collection
- No need to do the calculations ever again

The BMI Health Study

- Body Mass Index: Weight in Kilograms divided by Height in Meters Squared
- Under 18 means underweight, Over 25 means overweight, Over 30 means obese
- High BMI is associated with poor health, like high blood pressure and high cholesterol
- People with high BMI tend to be older and fatter
- **BUT**, what if you have a high BMI but are in good physical shape – low percent body fat?

The Question

- If you control for age and percent body fat, is BMI still associated with indicators for poor health?
- But percent body fat (and to a lesser extent, age) are measured with error. Standard ways of controlling for them with regression are highly suspect.
- Use the double measurement design.

True variables (all latent)

- X_1 = Age
- X_2 = BMI
- X_3 = Percent body fat
- Y_1 = Cholesterol
- Y_2 = Diastolic blood pressure

Measure twice with different personnel at different locations and by different methods

	Measurement Set One	Measurement Set Two
Age	Self report	Passport or Birth Certificate
BMI	Dr. Office Measurement	Lab technician, no shoes, gown
% Body Fat	Tape and calipers	Submerge in water tank
Cholesterol	Lab 1	Lab 2
Diastolic BP	Blood pressure cuff, Dr. Office	Digital readout, mostly automatic

Set two is of generally higher quality

Correlation of measurement errors is less likely between sets

Copyright Information