Omitted Variables[1]

STA431 Spring 2023

[1]See last slide for copyright information.

Overview

1. Omitted Variables

2. Instrumental Variables

A Practical Data Analysis Problem

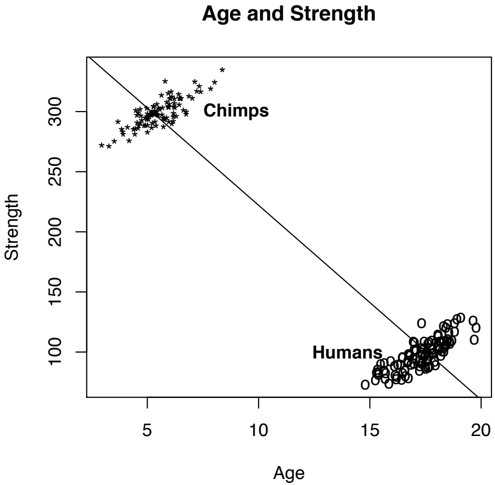
When more explanatory variables are added to a regression model and these additional explanatory variables are correlated with explanatory variables already in the model (as they usually are in an observational study),

- Statistical significance can appear when it was not present originally.
- Statistical significance that was originally present can disappear.
- Even the signs of the $\widehat{\beta}$s can change, reversing the interpretation of how their variables are related to the response variable.

An extreme, artificial example

To make a point

Suppose that in a certain population, the correlation between age and strength is $r = -0.93$.



The fixed x regression model

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,p-1} + \epsilon_i \\
&= \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i,
\end{aligned}
$$

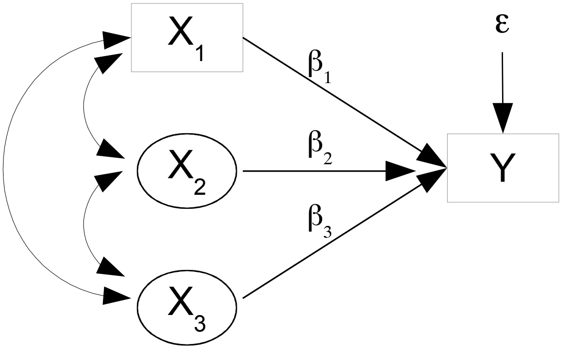
with $\epsilon_1 \ldots, \epsilon_n \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.

- If viewed as conditional on $\mathcal{X}_i = \mathbf{x}_i$, this model implies independence of ϵ_i and $\mathcal{X}_i$, because the conditional distribution of ϵ_i given $\mathcal{X}_i = \mathbf{x}_i$ does not depend on $\mathbf{x}_i$.
- What is ϵ_i? *Everything else* that affects Y_i.
- So the usual model says that if the explanatory varables are random, they have *zero covariance* with all other variables that are related to Y_i, but are not included in the model.
- For observational data (no random assignment), this assumption is almost always violated.
- Does it matter?

Example: $Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \epsilon_i$

As usual, the explanatory variables are random.

Suppose that the variables X_2 and X_3 affect Y and are correlated with X_1, but they are not part of the data set.



Statement of the model

The explanatory variables X_2 and X_3 influence Y and are correlated with X_1, but they are not part of the data set.

The values of the response variable are generated as follows:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \epsilon_i,$$

independently for $i = 1, \ldots, n$, where $\epsilon_i \sim N(0, \sigma^2)$. The explanatory variables are random, with expected value and variance-covariance matrix

$$E \left(\begin{array}{c} X_{i,1} \\ X_{i,2} \\ X_{i,3} \end{array} \right) = \left(\begin{array}{c} \mu_1 \\ \mu_2 \\ \mu_3 \end{array} \right) \text{ and } cov \left(\begin{array}{c} X_{i,1} \\ X_{i,2} \\ X_{i,3} \end{array} \right) = \left(\begin{array}{ccc} \phi_{11} & \phi_{12} & \phi_{13} \\ & \phi_{22} & \phi_{23} \\ & & \phi_{33} \end{array} \right),$$

where ϵ_i is independent of $X_{i,1}$, $X_{i,2}$ and $X_{i,3}$. Values of the variables $X_{i,2}$ and $X_{i,3}$ are latent, and are not included in the data set.

Absorb X_2 and X_3

Since X_2 and X_3 are not observed, they are absorbed by the intercept and error term.

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \beta_3 X_{i,3} + \epsilon_i \\
&= (\beta_0 + \beta_2 \mu_2 + \beta_3 \mu_3) + \beta_1 X_{i,1} + (\beta_2 X_{i,2} + \beta_3 X_{i,3} - \beta_2 \mu_2 - \beta_3 \mu_3 + \epsilon_i) \\
&= \beta_0' + \beta_1 X_{i,1} + \epsilon_i'.
\end{aligned}
$$

And,

$$
\begin{aligned}
Cov(X_{i,1}, \epsilon_i') &= Cov(X_{i,1}, \beta_2 X_{i,2} + \beta_3 X_{i,3} - \beta_2 \mu_2 - \beta_3 \mu_3 + \epsilon_i) \\
&= \beta_2 Cov(X_{i,1}, X_{i,2}) + \beta_3 Cov(X_{i,1}, X_{i,3}) + Cov(X_{i,1}, \epsilon_i) \\
&= \beta_2 \phi_{12} + \beta_3 \phi_{13} \neq 0.
\end{aligned}
$$

The “True” Regression Model

Almost always closer to the truth than the usual model, for observational data

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where $E(X_i) = \mu_x$, $Var(X_i) = \sigma_x^2$, $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma_\epsilon^2$, and $Cov(X_i, \epsilon_i) = c$.

Under this model,

$$\sigma_{xy} = Cov(X_i, Y_i) = Cov(X_i, \beta_0 + \beta_1 X_i + \epsilon_i) = \beta_1 \sigma_x^2 + c$$

Estimate β_1 as usual with least squares

Recall $Cov(X_i, Y_i) = \sigma_{xy} = \beta_1 \sigma_x^2 + c$

$$
\begin{aligned}
\widehat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2} \\
&= \frac{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2} \\
&= \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x^2} \xrightarrow{p} \frac{\sigma_{xy}}{\sigma_x^2} \\
&= \frac{\beta_1 \sigma_x^2 + c}{\sigma_x^2} \\
&= \beta_1 + \frac{c}{\sigma_x^2}
\end{aligned}
$$

$\widehat{\beta}_1 \stackrel{p}{\rightarrow} \beta_1 + \frac{c}{\sigma_x^2}$

It converges to the wrong thing.

- $\widehat{\beta}_1$ is inconsistent.
- For large samples it could be almost anything, depending on the value of c, the covariance between X_i and ϵ_i.
- Small sample estimates could be accurate, but only by chance.
- The only time $\widehat{\beta}_1$ behaves properly is when $c = 0$.
- Test $H_0 : \beta_1 = 0$: Probability of making a Type I error goes to one as $n \rightarrow \infty$.

All this applies to multiple regression

Of course

When a regression model fails to include all the explanatory variables that contribute to the response variable, and those omitted explanatory variables have non-zero covariance with variables that are in the model, the regression coefficients are inconsistent.

Estimation and inference are almost guaranteed to be misleading, especially for large samples.

Correlation-Causation

- The problem of omitted variables is a technical aspect of the correlation-causation issue.
- The omitted variables are "confounding" variables.
- With random assignment and good procedure, x and ϵ have zero covariance.
- But random assignment is not always possible.
- Most applications of regression to observational data provide very poor information about the regression coefficients.
- Is bad information better than no information at all?

How about another estimation method?

Other than ordinary least squares

- Can *any* other method be successful?
- This is a very practical question, because almost all regressions with observational data have the disease.

For simplicity, assume normality

$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

- Assume (X_i, ϵ_i) are bivariate normal.
- This makes (X_i, Y_i) bivariate normal.
- $(X_1, Y_1), \ldots, (X_n, Y_n) \stackrel{i.i.d.}{\sim} N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

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

and

$$\boldsymbol{\Sigma} = \left(\begin{array}{cc} \sigma_{11} & \sigma_{12} \\ & \sigma_{22} \end{array} \right) = \left(\begin{array}{cc} \sigma_x^2 & \beta_1 \sigma_x^2 + c \\ & \beta_1^2 \sigma_x^2 + 2\beta_1 c + \sigma_\epsilon^2 \end{array} \right).$$

- All you can ever learn from the data are the approximate values of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
- Even if you knew $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ exactly, could you know β_1?

Five equations in six unknowns

The parameter is $\theta = (\mu_x, \sigma_x^2, \sigma_\epsilon^2, c, \beta_0, \beta_1)$. The distribution of the data is determined by

$$\left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array}\right) = \left(\begin{array}{c} \mu_x \\ \beta_0 + \beta_1\mu_x \end{array}\right) \quad \text{and} \quad \left(\begin{array}{cc} \sigma_{11} & \sigma_{12} \\ & \sigma_{22} \end{array}\right) = \left(\begin{array}{cc} \sigma_x^2 & \beta_1\sigma_x^2 + c \\ & \beta_1^2\sigma_x^2 + 2\beta_1 c + \sigma_\epsilon^2 \end{array}\right)$$

- $\mu_x = \mu_1$ and $\sigma_x^2 = \sigma_{11}$.
- The remaining 3 equations in 4 unknowns have infinitely many solutions.
- So infinitely many sets of parameter values yield the *same distribution of the sample data*.
- This is serious trouble – lack of parameter identifiability.
- *Definition*: If a parameter is a function of the distribution of the observable data, it is said to be *identifiable*.

Showing identifiability

Definition: If a parameter is a function of the distribution of the observable data, it is said to be identifiable.

- How could a parameter be a function of a distribution?
- $d \sim F_\theta$ and $\theta = g(F_\theta)$
- Usually g is defined in terms of moments.
- Example: $F_\theta(x) = 1 - e^{-\theta x}$ and $f_\theta(x) = \theta e^{-\theta x}$ for $x > 0$.

$$
\begin{aligned}
f_\theta(x) &= \frac{d}{dx} F_\theta(x) \\
E(X) &= \int_0^\infty x f_\theta(x)\, dx = \frac{1}{\theta} \\
\theta &= \frac{1}{E(X)}
\end{aligned}
$$

Sometimes people use moment-generating functions or characteristic functions instead of just moments.

Showing identifiability is like Method of Moments Estimation

- The distribution of the data is always a function of the parameters.
- The moments are always a function of the distribution of the data.
- If the parameters can be expressed as a function of the moments,
  - Put hats on to obtain MOM estimates,
  - Or observe that the parameter is a function of the distribution, and so is identifiable.

Back to the five equations in six unknowns

$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$

$$\mathbf{d}_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ where}$$

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

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \cdot & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_x^2 & \beta_1 \sigma_x^2 + c \\ \cdot & \beta_1^2 \sigma_x^2 + 2\beta_1 c + \sigma_\epsilon^2 \end{pmatrix}$$

We have expressed the moments in terms of the parameters, but we can't solve for $\theta = (\mu_x, \sigma_x^2, \sigma_\epsilon^2, c, \beta_0, \beta_1)$.

Skipping the High School algebra

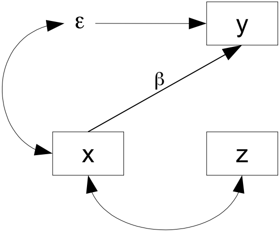
$\theta = (\mu_x, \sigma_x^2, \sigma_\epsilon^2, c, \beta_0, \beta_1)$

- For *any* given $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, all the points in a one-dimensional subset of the 6-dimensional parameter space yield $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and hence the same distribution of the sample data.
- In that subset, values of β_1 range from $-\infty$ to $-\infty$, so $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ could have been produced by *any* value of β_1.
- There is no way to distinguish between the possible values of β_1 based on sample data.
- The problem is fatal, if all you can observe is X and Y.

Instrumental Variables (Wright, 1928)

A partial solution

- An instrumental variable is a variable that is correlated with an explanatory variable, but is not correlated with any error terms and has no direct connection to the response variable.



- An instrumental variable is often not the main focus of attention; it's just a tool.
- The usual definition is that conditionally on the x variables, the instrumental variables are independent of all the other variables in the model.
- In Econometrics, the instrumental variable usually *influences* the explanatory variable.

Model One: A Simple Example

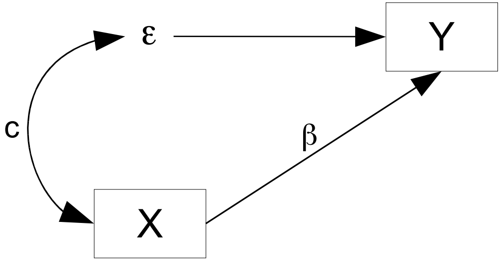
What is the contribution of income to credit card debt?

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where $E(X_i) = \mu_x$, $Var(X_i) = \sigma_x^2$, $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma_\epsilon^2$, and $Cov(X_i, \epsilon_i) = c$.

A path diagram of Model One

$Y_i = \alpha + \beta X_i + \epsilon_i$, where $E(X_i) = \mu$, $Var(X_i) = \sigma^2_x$, $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2_\epsilon$, and $Cov(X_i, \epsilon_i) = c$.



The least squares estimate of β is inconsistent, and so is every other possible estimate. (This is strictly true if the data are normal.)

Model Two: Add an instrumental variable

An instrumental variable for an explanatory variable is another random variable that has non-zero covariance with the explanatory variable, and *no direct connection with any other variable in the model.*

Focus the study on real estate agents in many cities. Include median price of resale home.

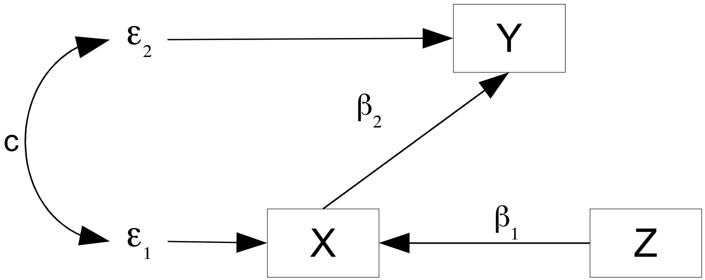
- X is income.
- Y is credit card debt.
- Z is median price of resale home.

$$
\begin{aligned}
X_i &= \alpha_1 + \beta_1 Z_i + \epsilon_{i1} \\
Y_i &= \alpha_2 + \beta_2 X_i + \epsilon_{i2}
\end{aligned}
$$

Picture of Model Two

Z_i is median price of resale home, X_i is income, Y_i is credit card debt.

$$
\begin{aligned}
X_i &= \alpha_1 + \beta_1 Z_i + \epsilon_{i1} \\
Y_i &= \alpha_2 + \beta_2 X_i + \epsilon_{i2}
\end{aligned}
$$



Main interest is in β_2.

Statement of Model Two

Z_i is median price of resale home, X_i is income, Y_i is credit card debt.

$$\begin{aligned} X_i &= \alpha_1 + \beta_1 Z_i + \epsilon_{i1} \\ Y_i &= \alpha_2 + \beta_2 X_i + \epsilon_{i2}, \end{aligned}$$

where

- $E(Z_i) = \mu_z$, $Var(Z_i) = \sigma_z^2$.
- $E(\epsilon_{i1}) = 0$, $Var(\epsilon_{i1}) = \sigma_1^2$.
- $E(\epsilon_{i2}) = 0$, $Var(\epsilon_{i2}) = \sigma_2^2$.
- $Cov(\epsilon_{i1}, \epsilon_{i2}) = c$.
- Z_i is independent of ϵ_{i1} and ϵ_{i2}.

Calculate the covariance matrix of the observable data for Model Two. Call it $\boldsymbol{\Sigma} = [\sigma_{ij}]$

From $X_i = \alpha_1 + \beta_1 Z_i + \epsilon_{i1}$ and $Y_i = \alpha_2 + \beta_2 X_i + \epsilon_{i2}$, get the symmetric matrix

$\boldsymbol{\Sigma} =$

	X	Y	Z
X	$\beta_1^2\sigma_z^2 + \sigma_1^2$	$\beta_2(\beta_1^2\sigma_z^2 + \sigma_1^2) + c$	$\beta_1\sigma_z^2$
Y	$\cdot$	$\beta_1^2\beta_2^2\sigma_z^2 + \beta_2^2\sigma_1^2 + 2\beta_2 c + \sigma_2^2$	$\beta_1\beta_2\sigma_z^2$
Z	$\cdot$	$\cdot$	σ_z^2

$$\beta_2 = \frac{\sigma_{23}}{\sigma_{13}}$$

Parameter Estimation for Model Two

$X_i = \alpha_1 + \beta_1 Z_i + \epsilon_{i1}$ and $Y_i = \alpha_2 + \beta_2 X_i + \epsilon_{i2}$

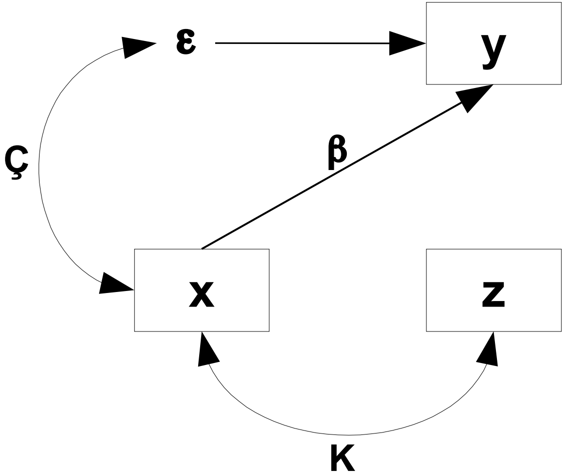
$\boldsymbol{\Sigma} =$

	Z	X	Y
Z	σ_w^2	$\beta_1\sigma_w^2$	$\beta_1\beta_2\sigma_w^2$
X	$\cdot$	$\beta_1^2\sigma_w^2 + \sigma_1^2$	$\beta_2(\beta_1^2\sigma_w^2 + \sigma_1^2) + c$
Y	$\cdot$	$\cdot$	$\beta_1^2\beta_2^2\sigma_w^2 + \beta_2^2\sigma_1^2 + 2\beta_2 c + \sigma_2^2$

- $\widehat{\beta}_2 = \frac{\widehat{\sigma}_{23}}{\widehat{\sigma}_{13}}$
- All the other parameters are identifiable too.
- The instrumental variable saved us.
- There are 9 model parameters, and 9 moments in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.
- The invariance principle yields explicit formulas for the MLEs.
- If the data are normal, MLEs equal the Method of Moments estimates because they are both 1-1 with the moments.

Model Three: Matrix Version of Instrumental Variables

The usual rule is at least one instrumental variable for each explanatory variable.

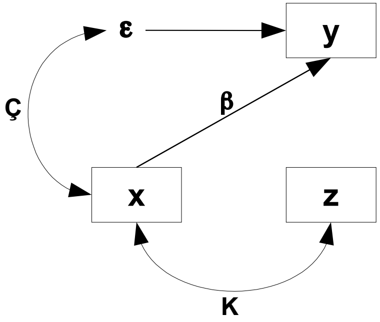


Model Three is a Multivariate Regression Model

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

With these additional stipulations

- $cov(\mathbf{x}_i, \boldsymbol{\epsilon}_i) = \text{Ç}$, a $p \times q$ matrix of covariances.
- There are at least p instrumental variables. Put the best p in the random vector $\mathbf{z}_i$.
- $cov(\mathbf{x}_i, \mathbf{z}_i) = \boldsymbol{\mathcal{K}}$, $p \times p$ matrix of covariances. Assume $\boldsymbol{\mathcal{K}}$ has an inverse.
- $cov(\mathbf{z}_i) = \boldsymbol{\Phi}_z$.



Moments for Model Three

$\mathbf{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \mathbf{x}_i + \boldsymbol{\epsilon}_i, \quad cov(\mathbf{x}_i, \boldsymbol{\epsilon}_i) = \text{Ç}, \quad cov(\mathbf{x}_i, \mathbf{z}_i) = \boldsymbol{\kappa}, \quad cov(\mathbf{z}_i) = \boldsymbol{\Phi}_z$

$$\boldsymbol{\mu} = E\left(\begin{array}{c} \mathbf{x}_i \\ \hline \mathbf{y}_i \\ \hline \mathbf{z}_i \end{array}\right) = \left(\begin{array}{c} \boldsymbol{\mu}_x \\ \hline \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 \boldsymbol{\mu}_x \\ \hline \boldsymbol{\mu}_z \end{array}\right)$$

$$\boldsymbol{\Sigma} = cov\left(\begin{array}{c} \mathbf{x}_i \\ \hline \mathbf{y}_i \\ \hline \mathbf{z}_i \end{array}\right) = \left(\begin{array}{c|c|c} \boldsymbol{\Phi}_x & \boldsymbol{\Phi}_x \boldsymbol{\beta}_1^\top + \text{Ç} & \boldsymbol{\kappa} \\ \hline \boldsymbol{\beta}_1 \boldsymbol{\Phi}_x + \text{Ç}^\top & \boldsymbol{\beta}_1 \boldsymbol{\Phi}_x \boldsymbol{\beta}_1^\top + \boldsymbol{\beta}_1 \text{Ç} + \text{Ç}^\top \boldsymbol{\beta}_1^\top + \boldsymbol{\Psi} & \boldsymbol{\beta}_1 \boldsymbol{\kappa} \\ \hline \boldsymbol{\kappa}^\top & \boldsymbol{\kappa}^\top \boldsymbol{\beta}_1^\top & \boldsymbol{\Phi}_z \end{array}\right)$$

Solve for the parameters from the moments

For Model Three

- Proving identifiability, so consistent estimation is possible.
- Obtain method of moments estimators.

Start with the covariance matrix $\boldsymbol{\Sigma}$

Parameters in the covariance matrix are $\boldsymbol{\Phi}_x$, $\boldsymbol{\beta}_1$, $\boldsymbol{\kappa}$, Ç, $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}_z$

$$\left(\begin{array}{c|c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \hline \cdot & \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \\ \hline \cdot & \cdot & \boldsymbol{\Sigma}_{33} \end{array}\right) = \left(\begin{array}{c|c|c} \boldsymbol{\Phi}_x & \boldsymbol{\Phi}_x\boldsymbol{\beta}_1^\top + \text{Ç} & \boldsymbol{\kappa} \\ \hline \cdot & \boldsymbol{\beta}_1\boldsymbol{\Phi}_x\boldsymbol{\beta}_1^\top + \boldsymbol{\beta}_1\text{Ç} + \text{Ç}^\top\boldsymbol{\beta}_1^\top + \boldsymbol{\Psi} & \boldsymbol{\beta}_1\boldsymbol{\kappa} \\ \hline \cdot & \cdot & \boldsymbol{\Phi}_z \end{array}\right)$$

Six matrix equations in six unknowns.

Solutions

Equations

$$\left(\begin{array}{c|c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \hline \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \hline \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{array}\right) = \left(\begin{array}{c|c|c} \boldsymbol{\Phi}_x & \boldsymbol{\Phi}_x\boldsymbol{\beta}_1^\top + \text{Ç} & \boldsymbol{\kappa} \\ \hline \boldsymbol{\beta}_1\boldsymbol{\Phi}_x + \text{Ç}^\top & \boldsymbol{\beta}_1\boldsymbol{\Phi}_x\boldsymbol{\beta}_1^\top + \boldsymbol{\beta}_1\text{Ç} + \text{Ç}^\top\boldsymbol{\beta}_1^\top + \boldsymbol{\Psi} & \boldsymbol{\beta}_1\boldsymbol{\kappa} \\ \hline \boldsymbol{\kappa}^\top & \boldsymbol{\kappa}^\top\boldsymbol{\beta}_1^\top & \boldsymbol{\Phi}_z \end{array}\right)$$

Solutions

$$\begin{aligned} \boldsymbol{\Phi}_x &= \boldsymbol{\Sigma}_{11} \\ \boldsymbol{\kappa} &= \boldsymbol{\Sigma}_{13} \\ \boldsymbol{\Phi}_z &= \boldsymbol{\Sigma}_{33} \\ \boldsymbol{\beta}_1 &= \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{13}^{-1} \\ \text{Ç} &= \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{31}^{-1}\boldsymbol{\Sigma}_{32} \\ \boldsymbol{\Psi} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{13}^{-1}\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{31}^{-1}\boldsymbol{\Sigma}_{32} + \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{13}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{31}^{-1}\boldsymbol{\Sigma}_{32} \end{aligned}$$

Solve for intercepts and expected values

Using

$$E\left(\begin{array}{c}\mathbf{x}_i\\ \hline \mathbf{y}_i\\ \hline \mathbf{z}_i\end{array}\right)=\left(\begin{array}{c}\boldsymbol{\mu}_1\\ \hline \boldsymbol{\mu}_2\\ \hline \boldsymbol{\mu}_3\end{array}\right)=\left(\begin{array}{c}\boldsymbol{\mu}_x\\ \hline \boldsymbol{\beta}_0+\boldsymbol{\beta}_1\boldsymbol{\mu}_x\\ \hline \boldsymbol{\mu}_z\end{array}\right)$$

$$\begin{aligned}\boldsymbol{\mu}_x &= \boldsymbol{\mu}_1\\ \boldsymbol{\mu}_z &= \boldsymbol{\mu}_3\\ \boldsymbol{\beta}_0 &= \boldsymbol{\mu}_2-\boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{13}^{-1}\boldsymbol{\mu}_1\end{aligned}$$

Full Solution for Model Three

For the record

$$
\begin{aligned}
\boldsymbol{\Phi}_x &= \boldsymbol{\Sigma}_{11} \\
\boldsymbol{\kappa} &= \boldsymbol{\Sigma}_{13} \\
\boldsymbol{\Phi}_z &= \boldsymbol{\Sigma}_{33} \\
\boldsymbol{\beta}_1 &= \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{13}^{-1} \\
\text{Ç} &= \boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{31}^{-1}\boldsymbol{\Sigma}_{32} \\
\boldsymbol{\Psi} &= \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{13}^{-1}\boldsymbol{\Sigma}_{12} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{31}^{-1}\boldsymbol{\Sigma}_{32} + \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{13}^{-1}\boldsymbol{\Sigma}_{11}\boldsymbol{\Sigma}_{31}^{-1}\boldsymbol{\Sigma}_{32} \\
\boldsymbol{\mu}_x &= \boldsymbol{\mu}_1 \\
\boldsymbol{\mu}_z &= \boldsymbol{\mu}_3 \\
\boldsymbol{\beta}_0 &= \boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{13}^{-1}\boldsymbol{\mu}_1
\end{aligned}
$$

Method of Moments Estimators for Model Three

Just put hats on, and estimate population means with sample means

$$
\begin{aligned}
\widehat{\boldsymbol{\Phi}}_x &= \widehat{\boldsymbol{\Sigma}}_{11} \\
\widehat{\boldsymbol{\kappa}} &= \widehat{\boldsymbol{\Sigma}}_{13} \\
\widehat{\boldsymbol{\Phi}}_z &= \widehat{\boldsymbol{\Sigma}}_{33} \\
\widehat{\boldsymbol{\beta}}_1 &= \widehat{\boldsymbol{\Sigma}}_{23}\widehat{\boldsymbol{\Sigma}}_{13}^{-1} \\
\widehat{\text{Ç}} &= \widehat{\boldsymbol{\Sigma}}_{12} - \widehat{\boldsymbol{\Sigma}}_{11}\widehat{\boldsymbol{\Sigma}}_{31}^{-1}\widehat{\boldsymbol{\Sigma}}_{32} \\
\widehat{\boldsymbol{\Psi}} &= \widehat{\boldsymbol{\Sigma}}_{22} - \widehat{\boldsymbol{\Sigma}}_{23}\widehat{\boldsymbol{\Sigma}}_{13}^{-1}\widehat{\boldsymbol{\Sigma}}_{12} - \widehat{\boldsymbol{\Sigma}}_{21}\widehat{\boldsymbol{\Sigma}}_{31}^{-1}\widehat{\boldsymbol{\Sigma}}_{32} + \widehat{\boldsymbol{\Sigma}}_{23}\widehat{\boldsymbol{\Sigma}}_{13}^{-1}\widehat{\boldsymbol{\Sigma}}_{11}\widehat{\boldsymbol{\Sigma}}_{31}^{-1}\widehat{\boldsymbol{\Sigma}}_{32} \\
\widehat{\boldsymbol{\mu}}_x &= \overline{\mathbf{x}} \\
\widehat{\boldsymbol{\mu}}_z &= \overline{\mathbf{z}} \\
\widehat{\boldsymbol{\beta}}_0 &= \overline{\mathbf{y}} - \widehat{\boldsymbol{\Sigma}}_{23}\widehat{\boldsymbol{\Sigma}}_{13}^{-1}\overline{\mathbf{x}}
\end{aligned}
$$

Count the Parameters in Model Three

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

- $\boldsymbol{\mu}_x$ is a $p \times 1$ vector of expected values. That's p parameters.
- $\boldsymbol{\mu}_z$ is a $p \times 1$ vector of expected values. That's p more parameters.
- $\boldsymbol{\beta}_0$ is a $q \times 1$ vector of intercepts. That's q parameters.
- $\boldsymbol{\beta}_1$ is a $p \times q$ matrix of regression coefficients. That's pq parameters.
- $cov(\mathbf{x}_i) = \boldsymbol{\Phi}_x$ is a $p \times p$ covariance matrix, with $p(p+1)/2$ unique elements.
- $cov(\boldsymbol{\epsilon}_i) = \boldsymbol{\Psi}$ is a $q \times q$ covariance matrix, with $q(q+1)/2$ unique elements.
- $cov(\mathbf{z}_i) = \boldsymbol{\Phi}_z$ is a $p \times p$ covariance matrix, with $p(p+1)/2$ unique elements.
- $cov(\mathbf{x}_i, \boldsymbol{\epsilon}_i) = \boldsymbol{\c{C}}$ is a $p \times q$ matrix of covariances. That's pq more parameters.
- $cov(\mathbf{x}_i, \mathbf{z}_i) = \boldsymbol{\mathcal{K}}$ is $p \times p$ matrix of covariances. That's p^2 more parameters.

Counting

- Parameters

$$
\begin{aligned}
& 2p + q + pq + \frac{p(p+1)}{2} + \frac{q(q+1)}{2} + \frac{p(p+1)}{2} + pq + p^2 \\
= \quad & 3p + q + 2p^2 + 2pq + \frac{q^2}{2} + \frac{q}{2}
\end{aligned}
$$

- Moments
  - There are $2p + q$ expected values in $\boldsymbol{\mu} = (\boldsymbol{\mu}_1 | \boldsymbol{\mu}_2 | \boldsymbol{\mu}_3)^\top$.
  - $\boldsymbol{\Sigma}$ has $2p + q$ rows and $2p + q$ columns, for $(2p + q)(2p + q + 1)/2$ unique elements.
  - Total number of moments is $3p + q + 2p^2 + 2pq + \frac{q^2}{2} + \frac{q}{2}$.

Invariance

The MLE of a 1-1 function is that function of the MLE

- Pretend the data $\mathbf{d}_i$ are multivariate normal.
- Unrestricted multivariate normal MLE of $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is $(\overline{\mathbf{d}}, \widehat{\boldsymbol{\Sigma}})$.
- The moments $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are a function of the model parameters in $\boldsymbol{\theta}$.
- By solving equations, we have shown that the models parameters are also a function of the moments and there are the same number of moments and model parameters.
- The function is one-to-one (injective).
- By invariance, $(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}) \leftrightarrow \widehat{\boldsymbol{\theta}}$.
- And the MOM estimates are also the MLEs.

Explicit formulas for the Maximum Likelihood Estimators: Model Three

$$
\begin{array}{rcl}
\widehat{\boldsymbol{\beta}}_0 & = & \overline{\mathbf{y}} - \widehat{\boldsymbol{\Sigma}}_{23}\widehat{\boldsymbol{\Sigma}}_{13}^{-1}\overline{\mathbf{x}} \\
\widehat{\boldsymbol{\beta}}_1 & = & \widehat{\boldsymbol{\Sigma}}_{23}\widehat{\boldsymbol{\Sigma}}_{13}^{-1} \\
\widehat{\boldsymbol{\Psi}} & = & \widehat{\boldsymbol{\Sigma}}_{22} - \widehat{\boldsymbol{\Sigma}}_{23}\widehat{\boldsymbol{\Sigma}}_{13}^{-1}\widehat{\boldsymbol{\Sigma}}_{12} - \widehat{\boldsymbol{\Sigma}}_{21}\widehat{\boldsymbol{\Sigma}}_{31}^{-1}\widehat{\boldsymbol{\Sigma}}_{32} + \widehat{\boldsymbol{\Sigma}}_{23}\widehat{\boldsymbol{\Sigma}}_{13}^{-1}\widehat{\boldsymbol{\Sigma}}_{11}\widehat{\boldsymbol{\Sigma}}_{31}^{-1}\widehat{\boldsymbol{\Sigma}}_{32} \\
\widehat{\mathbf{Ç}} & = & \widehat{\boldsymbol{\Sigma}}_{12} - \widehat{\boldsymbol{\Sigma}}_{11}\widehat{\boldsymbol{\Sigma}}_{31}^{-1}\widehat{\boldsymbol{\Sigma}}_{32}
\end{array}
$$

$$
\begin{array}{rcl}
\widehat{\boldsymbol{\Phi}}_x & = & \widehat{\boldsymbol{\Sigma}}_{11} \\
\widehat{\boldsymbol{\kappa}} & = & \widehat{\boldsymbol{\Sigma}}_{13} \\
\widehat{\boldsymbol{\Phi}}_z & = & \widehat{\boldsymbol{\Sigma}}_{33} \\
\widehat{\boldsymbol{\mu}}_x & = & \overline{\mathbf{x}} \\
\widehat{\boldsymbol{\mu}}_z & = & \overline{\mathbf{z}}
\end{array}
$$

Multivariate Normal Likelihood

$\mathbf{d}_1, \ldots, \mathbf{d}_n \overset{iid}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

$$
\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}} (2\pi)^{\frac{p}{2}}} \exp\left\{ -\frac{1}{2} (\mathbf{d}_i - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{d}_i - \boldsymbol{\mu}) \right\} \\
&= |\boldsymbol{\Sigma}|^{-\frac{n}{2}} (2\pi)^{-\frac{np}{2}} \times \\
&\quad \exp -\frac{n}{2} \left\{ tr(\widehat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + (\overline{\mathbf{d}} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\overline{\mathbf{d}} - \boldsymbol{\mu}) \right\},
\end{aligned}
$$

where $\widehat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{d}_i - \overline{\mathbf{d}})(\mathbf{d}_i - \overline{\mathbf{d}})^\top$

Comments

- Instrumental variables are a great technical solution to the problem of omitted variables
- But good instrumental variables are not easy to find.
- They will not just happen to be in the data set, except by a miracle.
- They really have to come from another universe, but still have a strong and clear connection to the explanatory variable.
- Data collection has to be *planned*.
- Wright's original example was tax policy for cooking oil.

Copyright Information