IN4402: Applied statistics for management and economics
Endogeneity

ANDRÉS FERNÁNDEZ

www.sistemaspublicos.cl

- Formally, one of the **main assumptions** of OLS is the unobserved errors (u_i) is not correlated with the independent variables (X_i):
  - Under this assumption $\boldsymbol{\beta_1}$ is *unbiased* and *consistent*

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

- The expression of **endogeneity**, then :

$$E(u_i | X_i) \neq 0$$

- And X_i will be an **endogenous regressor.** Instead, when it is not correlated is **exogenous**

- One concrete expression:

- Omitted Variables - if variable O_i is omitted, then it is part of unobserved error (u_i), and if it correlates with X_i, then:

$$\widehat{\beta_1} = \beta_1 + \beta_2 \frac{Cov(X_i, O_i)}{Var(X_i)}$$

Omitted variable bias: direction of bias

	$Corr(x_1, x_2) > 0$	$Corr(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

- **Endogeneity**, $E(u_i|X_i) \neq 0$, implies:

  - An independent variables correlate with unobserved error

  - There is a correlation that we cannot identify

  - The expected value of errors, given X_i, is not 0

  - The estimators are biased

IN4402: Applied statistics for management and economics

Instrumental Variables - Models

ANDRÉS FERNÁNDEZ

www.sistemaspublicos.cl

- We have an *endogenous* regressor (i.e. omitted variable)?

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

- What if we have information that we expect to be *exogenous?*

- How can we use it?

- So, a good **instrument** Z_i will be a variable that is:

  relevant: $$Cov(X_i, Z_i,) \neq 0$$

  exogenous: $$Cov(u_i, Z_i,) = 0$$

- What can we say about these assumptions?
  - The first one **can** be tested
  - The second **cannot** be tested, so good arguments are needed

- To capture this, we run a **first stage:**

first stage: $\qquad X_i = \pi_0 + \pi_1 Z_i + v_i$

second stage: $\qquad Y_i = \beta_0 + \beta_1 \widehat{X_i} + u_i$

- This method is called **two-stages least square (2SLS).**

- Let's see it in another way:

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

$$Cov(Y_i, Z_i) = Cov(\beta_0 + \beta_1 X_i + u_i, Z_i)$$

$$Cov(Y_i, Z_i) = Cov(\beta_{0,}, Z_i) + Cov(\beta_1 X_i, Z_i) + Cov(u_i, Z_i)$$

$$Cov(Y_i, Z_i) = \beta_1 Cov(X_i, Z_i) + \boldsymbol{Cov(u_i, Z_i)}$$

$$\frac{Cov(Y_i, Z_i)}{Cov(X_i, Z_i)} = \beta_1$$

- Without **relevance,** the estimator is overestimated *(**weak** instrument).*
- Without **exogeneity** is biased.

- How to look for a good instrument?

$$\frac{Cov(Y_i, Z_i)}{Cov(X_i, Z_i)} = \beta_1$$

IN4402: Applied statistics for management and economics

Instrumental Variables - Application

ANDRÉS FERNÁNDEZ

www.sistemaspublicos.cl

- Does government social program improve income?

$$Income_i = \beta_0 + \beta_1 Attended_i + u_i$$

- Voluntary participation:
  - What variable is omitted?

- How it could relate to $Income_i$ and $Attended_i$?

CSP

Vanguardia
Inicio | Bucaramanga | Santander | Deportes | Judicial | Colombia | Mundo | Política | Economía | Opinión

domingo 29 de septiembre de 2019 - 12:00 AM

Hogares comunitarios recibieron implementos

¿De qué se trata?
Los Hogares Comunitarios ofrecen atención integral a las niñas y los niños menores de cinco (5) años pertenecientes a familias clasificadas de acuerdo con los criterios definidos por el ICBF, dando prioridad a aquellos que requieren atención diaria. En total cada Hogar atiende entre doce (12) y (14) niñas y niños.

¿A dónde dirigirse?
La familia interesada podrá dirigirse al Hogar Comunitario Integral o centro zonal del ICBF, más cercano a su lugar de residencia.

Puntos de Atención ICBF

Any ideas about instrumental variables??

From: https://www.icbf.gov.co/portafolio-de-servicios-icbf/hogares-comunitarios-integrales

- Attanasio & Vera-Hernández (2004) idea:
  - Distance to government office

First stage regressions
Children 0-6

	Exposure		Currently attending	
	OLS	Tobit	OLS	Probit
Distance from the nearest *HC*	-0.17 (0.03)	-0.54 (0.08)	-0.36 (0.06)	-1.88 (0.36)
(Distance from the nearest *HC*)^2	0.05 (0.01)	0.08 (0.05)	0.10 (0.02)	0.37 (0.21)
Average min. dist. from an *HC* in town.	-0.65 (0.16)	-1.34 (0.21)	-1.16 (0.26)	-4.27 (1.06)
(Average min. dist. from an *HC* in town.) ^2	0.54 (0.16)	1.18 (0.26)	1.08 (0.26)	3.95 (1.25)
N	2445	2445	2554	2554
(Joint significance for the four instruments)	(0.000)	(0.000)	(0.000)	(0.000)

From: Attanasio, O., & Vera-Hernandez, M. (2004). Medium-and long run effects of nutrition and child care: evaluation of a community nursery programme in rural Colombia.
Bernal, R., & Peña, X. (2011). *Guía práctica para la evaluación de impacto*. Ediciones Uniandes-Universidad de los Andes.

The sign of the bias

	OLS	IV
Height according to age	-0,059 (0,050)	0,486 (0,156)
Weight according to age	0,006 (0,046)	0,274 (0,170)
Standard errors in parenthesis		

From: Attanasio, O., & Vera-Hernandez, M. (2004). Medium-and long run effects of nutrition and child care: evaluation of a community nursery programme in rural Colombia.
Bernal, R., & Peña, X. (2011). *Guía práctica para la evaluación de impacto*. Ediciones Uniandes-Universidad de los Andes.