# Instrumental Variables

Oscar Gálvez-Soriano

University of Houston

Department of Economics

November 2020

# Idea of IV

Our research question is: What is the effect of x on y?

$$y_i = \alpha + \beta x_i + ui$$

- Concern about endogeneity of $x$ in the relationship between $x$ and $y$.
- Suppose there exists $z$ satisfying two conditions:
  - $Cov(z_i, x_i) \neq 0$
  - $Cov(z_i, u_i) = 0$
- Then z can be used to instrument for x to provide a consistent estimate of the effect of x on y.

# Terminology

The **structural equation** is the one you wish to estimate:

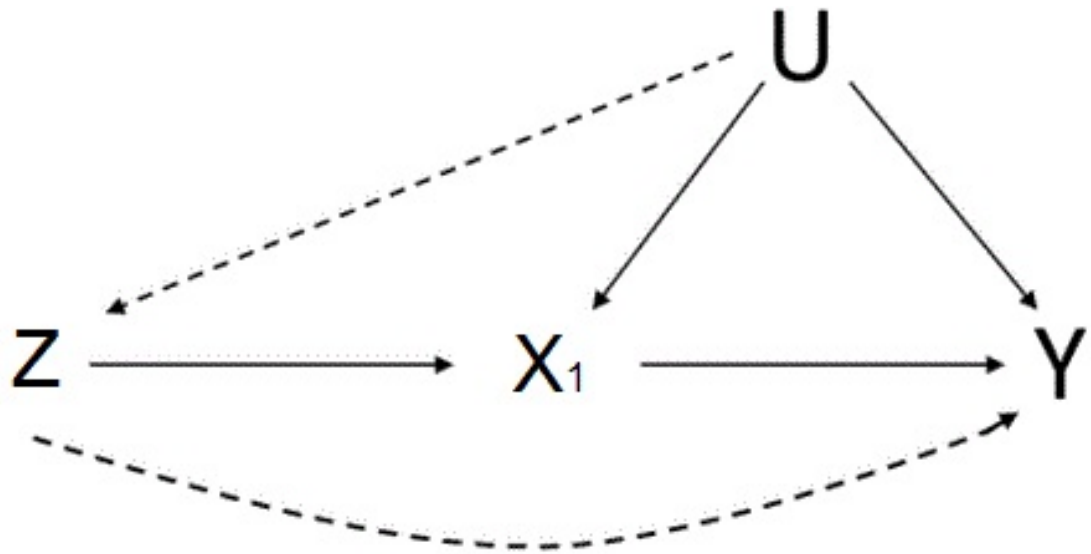
$$y_i = \alpha + \beta x_i + ui$$

The **first stage equation** is the equation with the endogenous regressor as the dependent variable and all the exogenous variables as the independent variables. Suppose part of $x$ can be partitioned into $\begin{bmatrix} x_1 & x_2 \end{bmatrix}$ where $x_1$ are thought to be endogenous and $x_2$ are thought to be exogenous. The first stage regression is:

$$x_{1i} = \pi_1 + \pi_2 x_{2i} + \pi_3 z_i + v_i$$

# Terminology

Your identification of the parameters in the structural model is from **exclusion restrictions**: you are assuming that $z$ does not affect $y$ other than through $x_1$.



# Terminology

The **reduced form equation** is the equation with the outcome of interest as the dependent variable and all the exogenous variables as the independent variables, i.e.,

$$y_i = \gamma_1 + \gamma_2 x_{2i} + \gamma_3 z_i + w_i$$

An equation is **just identified** if there are the same number of instruments (variables in $z$) as there are endogenous regressors (variables in $x_1$).

# Indirect Least Squares

$$y_i = \alpha + \beta_{1i} x_{1i} + \beta_{2i} x_{2i} + ui$$

Suppose you have only one instrument. Then the IV estimator is:

$$\beta_1 = \frac{\gamma_3}{\pi_3}$$

This ratio of reduced form to first stage coefficients is sometimes called the indirect least squares estimator.

# Instrumental Variables Estimation

The most efficient among IV estimators using instruments linear in $z$ is the two-stage least squares (2SLS) estimator. The 2SLS estimator takes advantage of variation in $x$ provided by all the instruments.

- Structural equation:

$$y_i = \alpha + \beta_{1i} x_{1i} + \beta_{2i} x_{2i} + ui$$

- First stage equation:

$$x_{1i} = \pi_1 + \pi_2 x_{2i} + \pi_3 z_i + v_i$$

$$\Rightarrow \quad \hat{x}_{1i}$$

# Instrumental Variables Estimation

- Second stage equation:

$$y_i = \alpha + \beta_{1i}\hat{x}_{1i} + \beta_{2i}x_{2i} + ui$$