Lecture 8: Instrumental Variables
Part I

Wages are a function of schooling:

$$Y_{si} \equiv f(s_i)$$

$$Y_{si} = \alpha + \rho s_i + \eta_i \quad \text{model we estimate}$$

$$\eta_i = A_i' \gamma + v_i$$

omitted variable: ability.

OLS:

(1) $\varepsilon_i \sim iid(0, \sigma^2) \Rightarrow E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2 \cdot I_{N \times N}$

~~(2)~~ $E(\varepsilon_i | X_{ik}) = 0$ for all K

(3) X has full column rank K

↳ not satisfied. $Corr(\eta_i, s_i) \neq 0$ because $Corr(A_i, s_i)$ likely > 0.

We could determine causal effect of s_i on Y_i using OLS only if we had all the variables that determined Y_i and could be correlated with s_i.

Propensity Score Matching

- Valid when X_i is binary $\{0,1\}$.
  Cannot use with a continuous variable like years of schooling

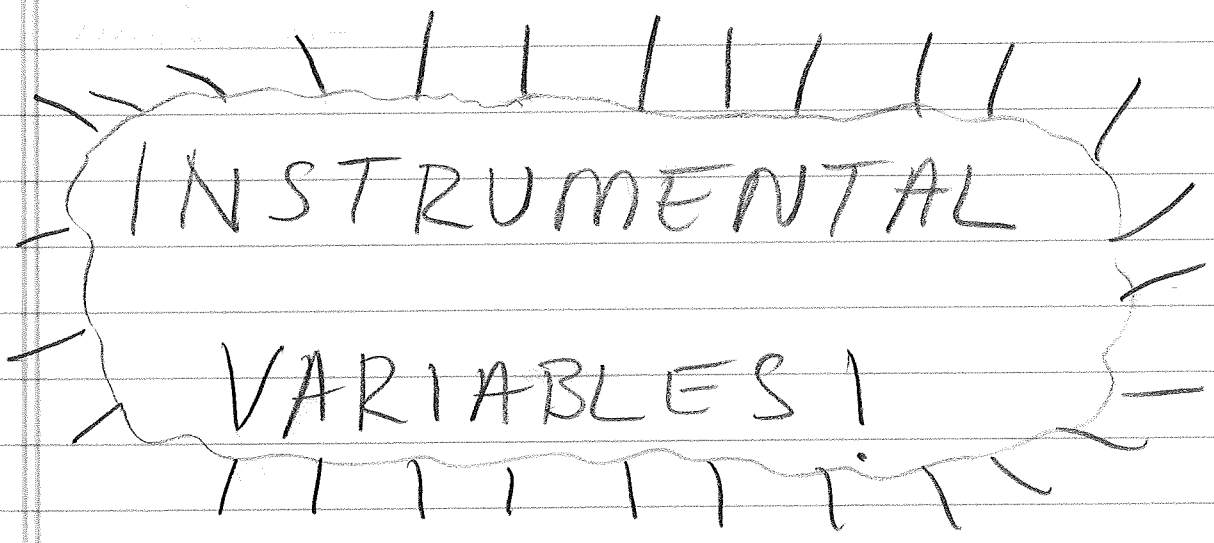
Regression Discontinuity:

- Valid when X_i is binary $\{0,1\}$ and there's some discontinuous rule:

$$X_j \geq \bar{X}_j \Rightarrow X_i = 1$$
$$X_j < \bar{X}_j \Rightarrow X_i = 0.$$

- Cannot use with a continuous variable and when there's no known rule that affects continuous variable.

What's next?



Instrumental Variables:

$$y_i = X'\beta + \varepsilon$$

What is an I.V.?

A variable (or variables) z that:

"strong" { (1) are strongly correlated with endogenous X

"valid" { (2) do not otherwise affect Y except through X.

↳ $Cov(\varepsilon_i, z_i) = 0$

What is our structural relation of interest:

(1) $Y_i = \alpha + \rho s_i + \eta_i$ (structural)

→ we cannot estimate this by OLS, but we have an instrument z_i that affects s_i but does not otherwise affect Y_i

(2) $s_i = \pi_{10} + \pi_{11} z_i + u_i$ (first stage)

Substitute (2) into (1):

(3) $Y_i = \alpha + \rho(\pi_{10} + \pi_{11} z_i + u_i) + v_i$

$= (\alpha + \rho\pi_{10}) + \rho\pi_{11} z_i + (\rho u_i + v_i)$

$Y_i = \pi_{20} + \pi_{21} z_i + \xi_i$ (reduced form)

Note that:

$$\frac{\text{red. form}}{\text{first stage}} = \frac{\pi_{21}}{\pi_{11}} = \frac{\dfrac{\text{Cov}(Y_i, z_i)}{\text{Var}(z_i)}}{\dfrac{\text{Cov}(s_i, z_i)}{\text{Var}(z_i)}} = \frac{\text{Cov}(Y_i, z_i)}{\text{Cov}(s_i, z_i)} = \frac{\rho\pi_{11}}{\pi_{11}} = \boxed{\rho}$$

We obtained our causal estimate of ρ!

[This is known as "indirect least squares: ILS]

What happens when more than one instrument?

(2a) $S_i = \pi_{10} + \pi_{11} z_{i1} + \pi_{12} z_{i2} + u_i$

(3a) $Y_i = \alpha + \rho(\pi_{10} + \pi_{11} z_{i1} + \pi_{12} z_{i2} + u_i) + \eta_i$

$= \underbrace{(\alpha + \rho \pi_{10})}_{} + \underbrace{\rho(\pi_{11} z_{i1} + \pi_{12} z_{i2})}_{} + \underbrace{(\rho u_i + \eta_i)}_{}$

$= \pi_{20} + \rho[\text{a weighted avg. of instruments.}] \quad \varepsilon_i$

Can no longer do ICS: $\rho = \frac{\pi_{21}}{\pi_{11}} = \frac{\rho \pi_{11}}{\pi_{11}}$

This is where 2SLS comes in. From (2a), calculate $\hat{S}$.

Estimate: $Y_i = \tilde{\alpha} + \tilde{\rho} \hat{S}_i + \tilde{\eta}_i$

Output from 2SLS:

F-statistic.

We need to make sure that $\text{corr}(x_i, \tilde{z}_i) \neq 0$.
Why?

$\rho = \frac{\pi_{21}}{\pi_{11}} = \frac{\rho \pi_{11}}{\pi_{11}}$ If $\pi_{11} \approx 0$, then we have $\frac{0}{0}$.

undefined.

What is an F-statistic?

$$s_i = \pi_0 + \pi_{11} z_{i1} + \pi_{21} z_{i2} + u_i$$

Joint test that π_{11} or (not and) π_{21} not equal to zero.

Rule of thumb: F-statistic ≥ 10.
The larger, the better.

Underidentification Test

Suppose K endog. variables: $X_1 \ldots X_K$.

Need at least K instruments, one for each endog. variable.

F-stat will tell you that $\pi_{11}, \ldots, \pi_{1K}$ jointly different from zero, but we also want to know that each $\pi_{11}, \ldots, \pi_{1K}$ are different from zero and that one of z_K is not a linear combination of the others.

Hansen J-statistic. (Sargan when errors not robust).

For over-identified case. (more instruments than endogenous regressors).

(1) Estimate equation using I.V.

(2) If I.V. valid, should obtain "good" parameter estimates and "good" estimates of original errors.

(3) Instruments should be uncorrelated with errors. Regress:

$$\hat{\varepsilon}_i = \alpha + \beta_1 z_1 + \beta_2 z_2 + \underbrace{X's}_{\text{exogenous}}$$

Perform F-test to see if β_1 and β_2 are jointly different from zero.

$NR^2 \sim$ chi sq (z's − endog. x's)

degrees of freedom.