Lecture Note 10: IV

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

$\uparrow$ homogeneous effect

$$\begin{tikzcd}
Z_i \arrow[r] \arrow[dr, dotted, leftrightarrow] & X_i \arrow[r] & Y_i \\
& U_i \arrow[u] \arrow[ur] &
\end{tikzcd}$$

Assumptions:

1. Relevance: $\text{cov}(Z_i, X_i) \neq 0$
2. Exogeneity: $\text{cov}(Z_i, U_i) = 0$
   a) Z_i has no direct effect on Y_i
   b) Z_i is "as good as random"

$$\text{cov}(Z_i, Y_i) = \text{cov}(Z_i, \beta_0 + \beta_1 X_i + U_i) = \cancel{\text{cov}(Z_i, \beta_0)} + \text{cov}(Z_i, \beta_1 X_i) + \cancel{\text{cov}(Z_i, U_i)}$$

$$\text{cov}(Z_i, Y_i) = \beta_1 \text{cov}(Z_i, X_i)$$

$$\beta_1 = \frac{\text{cov}(Z_i, Y_i) / V[Z_i]}{\text{cov}(Z_i, X_i) / V[Z_i]}$$

1. Regress Y_i on Z_i (reduced form)
2. Regress X_i on Z_i (1st stage)
3. Ratio $\frac{RF}{1^{st}}$

Two-stage least squares (TSLS)

① Regress X_i on Z_i: $X_i = \pi_0 + \pi_1 Z_i + V_i$

Form predicted vals: $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$

② Regress Y_i on $\hat{X}_i$: $Y_i = \beta_0 + \beta_1 \hat{X}_i + \varepsilon_i$

$\hat{\beta}_1^{TSLS}$ is same as ratio $\frac{RF}{1^{st}}$ and consistent estimator for β_1

In R using fixest, regress Y on X, W_1, W_2, instrumenting for X using Z

```
feols(y ~ w1+w2 | x ~ z, data = df)
```

Wald estimator: Z_i is binary

→ $\bar{Y}_0, \bar{Y}_1, \bar{X}_0, \bar{X}_1$

→ $\hat{\beta}_1^{wald} = \dfrac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0}$

$= 0$ in eligibility experiment

Heterogeneous effects

- single binary Z_i, single binary X_i
- potential outcome: $Y_i(x,z)$ for treatment level x and instrument level z
- potential treatment status: $X_i(z) \begin{cases} X_i(1) \\ X_i(0) \end{cases}$
- to interpret IV/TSLS, 3 assumptions:

① Independence: $\{Y_i(x,z), X_i(z)\} \perp Z_i$ ← Z_i is "as good as random" ← allows us to estimate 1st-stage and RF regressions

② Exclusion restriction: $Y_i(x,0) = Y_i(x,1)$ ← $Z_i \rightarrow X_i \rightarrow Y_i$

so $Y_i(x,z) = Y_i(x)$

③ Monotonicity: either $X_i(1) \geq X_i(0)$ for all i

or $X_i(0) \geq X_i(1)$ for all i

		$Z_i=1$	
		$X_i=0$	$X_i=1$
$Z_i=0$	$X_i=0$	never takers	compliers
	$X_i=1$	~~defiers~~ monotonicity	always takers

TOT → never, compliers

LATE → never, always, compliers

Under assumptions 1-3:

$$\hat{\beta}_1^{TSLS} \xrightarrow{p} LATE = \text{avg effect of } X \text{ on } Y \text{ among compliers}$$

Generalizing to non-binary Z_i and X_i:

$$X_i = \pi_{0i} + \pi_{1i} Z_i + V_i \quad 1^{st}$$

$$Y_i = \beta_{0i} + \beta_{1i} Z_i + U_i \quad 2^{nd}$$

Then:

$$\hat{\beta}_1^{TSLS} \xrightarrow{p} E\left[\frac{\overbrace{\pi_{1i}}^{\text{ind. sensitivity to } Z_i}}{\underbrace{E[\pi_{1i}]}_{\text{avg sensitivity to } Z_i}} \underbrace{\beta_{1i}}_{\text{ind. causal effect of } X_i \text{ on } Y_i}\right]$$