Lecture 10: Instrumental variables III

PPHA 34600
Prof. Fiona Burlig

Harris School of Public Policy
University of Chicago

From last time: applications of IV

Z_i is a valid instrument when the following are satisfied:

1. **First stage**: $Cov(Z_i, D_i) \neq 0$
2. **Exclusion restriction**: $Cov(Z_i, \varepsilon_i) = 0$

When we have these two conditions, we can...:

- Handle OVB
- Handle measurement error

Opening Pandora's box

With $\tau_i = \tau$ for all i, life is "easy":

- All we need is a first stage...
- ... and an exclusion restriction (☠)...
- ... and we are in business!

Opening Pandora's box

With $\tau_i = \tau$ for all i, life is "easy":

- All we need is a first stage...
- ... and an exclusion restriction (☠)...
- ... and we are in business!

What happens with heterogeneous treatment effects?

$\rightarrow$ What are we actually recovering with $\hat{\tau}^{IV}$?

Opening Pandora's box

With $\tau_i = \tau$ for all i, life is "easy":

- All we need is a first stage...
- ... and an exclusion restriction (☠)...
- ... and we are in business!

What happens with heterogeneous treatment effects?

→ What are we actually recovering with $\hat{\tau}^{IV}$?

- Recall that we set up IV with the following decomposition:

$$D_i = B_i \varepsilon_i + C_i$$

Opening Pandora's box

With $\tau_i = \tau$ for all i, life is "easy":

- All we need is a first stage...
- ... and an exclusion restriction (☠)...
- ... and we are in business!

What happens with heterogeneous treatment effects?

→ What are we actually recovering with $\hat{\tau}^{IV}$?

- Recall that we set up IV with the following decomposition:

$$D_i = B_i \varepsilon_i + C_i$$

- For the IV to work, it must be correlated with C_i but not $B_i \varepsilon_i$
- But! Z_i is just generating variation in *part* of C_i

Opening Pandora's box

With $\tau_i = \tau$ for all i, life is "easy":

- All we need is a first stage...
- ... and an exclusion restriction (☠)...
- ... and we are in business!

What happens with heterogeneous treatment effects?

$\rightarrow$ What are we actually recovering with $\hat{\tau}^{IV}$?

- Recall that we set up IV with the following decomposition:

$$D_i = B_i \varepsilon_i + C_i$$

- For the IV to work, it must be correlated with C_i but not $B_i \varepsilon_i$
- But! Z_i is just generating variation in *part* of C_i
- If this part affects Y_i differently than the non-moved bit, $\hat{\tau} \neq \tau^{ATE}$

A more general IV setup

Let's generalize our setup a bit:

- $Y_i(D_i, Z_i)$ is the outcome as a function of both treatment and the instrument

A more general IV setup

Let's generalize our setup a bit:

- $Y_i(D_i, Z_i)$ is the outcome as a function of both treatment and the instrument
- $Y_i(D_i = 1, Z_i) - Y_i(D_i = 0, Z_i)$:
  Causal effect of treatment given your instrument
- $Y_i(D_i, Z_i = 1) - Y_i(D_i, Z_i = 0)$:
  Causal effect of your instrument given your treatment status

A more general IV setup

In our intended causal chain, $Z_i \to D_i \to Y_i$:

- We want notation to think about Z_i having a causal effect on D_i. Define:
  - $D_i(Z_i = 1)$ or just $D_i(1)$ is treatment status when $Z_i = 1$
  - $D_i(Z_i = 0)$ or just $D_i(0)$ is treatment status when $Z_i = 0$

A more general IV setup

In our intended causal chain, $Z_i \rightarrow D_i \rightarrow Y_i$:

- We want notation to think about Z_i having a causal effect on D_i. Define:
  - $D_i(Z_i = 1)$ or just $D_i(1)$ is treatment status when $Z_i = 1$
  - $D_i(Z_i = 0)$ or just $D_i(0)$ is treatment status when $Z_i = 0$
- Observed treatment status is just:

$$D_i = D_i(0) + (D_i(1) - D_i(0))Z_i = \alpha + \gamma_i Z_i + \nu_i$$

A more general IV setup

In our intended causal chain, $Z_i \rightarrow D_i \rightarrow Y_i$:

- We want notation to think about Z_i having a causal effect on D_i. Define:
  - $D_i(Z_i = 1)$ or just $D_i(1)$ is treatment status when $Z_i = 1$
  - $D_i(Z_i = 0)$ or just $D_i(0)$ is treatment status when $Z_i = 0$
- Observed treatment status is just:

$$D_i = D_i(0) + (D_i(1) - D_i(0))Z_i = \alpha + \gamma_i Z_i + \nu_i$$

(This should look familiar!)

- As before, $\alpha = E[D_i(0)]$
- But now $\gamma_i \equiv (D_i(1) - D_i(0))$: the *i*-specific causal effect of Z_i on D_i

A more general IV setup

In our intended causal chain, $Z_i \to D_i \to Y_i$:

- We want notation to think about Z_i having a causal effect on D_i. Define:
  - $D_i(Z_i = 1)$ or just $D_i(1)$ is treatment status when $Z_i = 1$
  - $D_i(Z_i = 0)$ or just $D_i(0)$ is treatment status when $Z_i = 0$
- Observed treatment status is just:

$$D_i = D_i(0) + (D_i(1) - D_i(0))Z_i = \alpha + \gamma_i Z_i + \nu_i$$

(This should look familiar!)

- As before, $\alpha = E[D_i(0)]$
- But now $\gamma_i \equiv (D_i(1) - D_i(0))$: the i-specific causal effect of Z_i on D_i

$\rightarrow$ We can't observe both $D_i(1)$ and $D_i(0)$ (why?)

A more general IV setup

In our intended causal chain, $Z_i \to D_i \to Y_i$:

- We want notation to think about Z_i having a causal effect on D_i. Define:
  - $D_i(Z_i = 1)$ or just $D_i(1)$ is treatment status when $Z_i = 1$
  - $D_i(Z_i = 0)$ or just $D_i(0)$ is treatment status when $Z_i = 0$
- Observed treatment status is just:

$$D_i = D_i(0) + (D_i(1) - D_i(0))Z_i = \alpha + \gamma_i Z_i + \nu_i$$

(This should look familiar!)

- As before, $\alpha = E[D_i(0)]$
- But now $\gamma_i \equiv (D_i(1) - D_i(0))$: the *i*-specific causal effect of Z_i on D_i

$\rightarrow$ We can't observe both $D_i(1)$ and $D_i(0)$ (why?)

$\rightarrow$ We can hope for the *average* causal effect of Z_i on $D_i = E[\gamma_i]$

With this framework, we need some (new) assumptions

We'll make four assumptions:

1. **First stage:** $E[D_i|Z_i = 1] \neq E[D_i|Z_i = 0]$ for some i
    - This is the same as before: $Cov(D_i, Z_i) \neq 0$
2.
3.
4.

With this framework, we need some (new) assumptions

We'll make four assumptions:

1. **First stage:** $E[D_i|Z_i = 1] \neq E[D_i|Z_i = 0]$ for some i
   - This is the same as before: $Cov(D_i, Z_i) \neq 0$
2. **Independence:** $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$
3. **Exclusion restriction:** $Y_i(Z_i = 1, D_i) = Y_i(Z_i = 0, D_i)$
4.

Random assignment and the exclusion restriction

What used to just be the exclusion restriction, $Cov(Z_i, \varepsilon_i) = 0$ is now:

(A) Independence: $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$

Random assignment and the exclusion restriction

What used to just be the exclusion restriction, $Cov(Z_i, \varepsilon_i) = 0$ is now:

(A) Independence: $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$

- The reduced form, a regression of Y_i on Z_i, is identified:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

$$= E[Y_i(D_i(Z_i = 1), Z_i = 1)|Z_i = 1] = E[Y_i(D_i(Z_i = 0), Z_i = 0)|Z_i = 0]$$
$$= E[Y_i(D_i(Z_i = 1), 1) - Y_i(D_i(Z_i = 0), 0)]$$

Random assignment and the exclusion restriction

What used to just be the exclusion restriction, $Cov(Z_i, \varepsilon_i) = 0$ is now:

(A) Independence: $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$

- The reduced form, a regression of Y_i on Z_i, is identified:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

$$= E[Y_i(D_i(Z_i = 1), Z_i = 1)|Z_i = 1] = E[Y_i(D_i(Z_i = 0), Z_i = 0)|Z_i = 0]$$
$$= E[Y_i(D_i(Z_i = 1), 1) - Y_i(D_i(Z_i = 0), 0)]$$

- The first stage, a regression of D_i on Z_i, is identified:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$

$$= E[D_i(Z_i = 1)|Z_i = 1] - E[D_i(Z_i = 0)|Z_i = 0]$$
$$= E[D_i(Z_i = 1) - D_i(Z_i = 0)]$$

Random assignment and the exclusion restriction

What used to just be the exclusion restriction, $Cov(Z_i, \varepsilon_i) = 0$ is now:

(A) Independence: $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$

- The reduced form, a regression of Y_i on Z_i, is identified:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$$

$$= E[Y_i(D_i(Z_i = 1), Z_i = 1)|Z_i = 1] = E[Y_i(D_i(Z_i = 0), Z_i = 0)|Z_i = 0]$$
$$= E[Y_i(D_i(Z_i = 1), 1) - Y_i(D_i(Z_i = 0), 0)]$$

- The first stage, a regression of D_i on Z_i, is identified:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$

$$= E[D_i(Z_i = 1)|Z_i = 1] - E[D_i(Z_i = 0)|Z_i = 0]$$
$$= E[D_i(Z_i = 1) - D_i(Z_i = 0)]$$

(B) Exclusion restriction: $Y_i(Z_i = 1, D_i) = Y_i(Z_i = 0, D_i)$

Random assignment and the exclusion restriction

What used to just be the exclusion restriction, $Cov(Z_i, \varepsilon_i) = 0$ is now:

(A) Independence: $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$

(B) Exclusion restriction: $Y_i(Z_i = 1, D_i) = Y_i(Z_i = 0, D_i)$

Random assignment and the exclusion restriction

What used to just be the exclusion restriction, $Cov(Z_i, \varepsilon_i) = 0$ is now:

(A) Independence: $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$

(B) Exclusion restriction: $Y_i(Z_i = 1, D_i) = Y_i(Z_i = 0, D_i)$

- Z_i only affects Y_i through D_i

Random assignment and the exclusion restriction

What used to just be the exclusion restriction, $Cov(Z_i, \varepsilon_i) = 0$ is now:

(A) Independence: $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$

(B) Exclusion restriction: $Y_i(Z_i = 1, D_i) = Y_i(Z_i = 0, D_i)$

- Z_i only affects Y_i through D_i
- This lets us write:

$$Y_i(1) = Y_i(D_i = 1, Z_i = 1) = Y_i(D_i = 1, Z_i = 0)$$

$$Y_i(0) = Y_i(D_i = 0, Z_i = 1) = Y_i(D_i = 0, Z_i = 0)$$

Random assignment and the exclusion restriction

What used to just be the exclusion restriction, $Cov(Z_i, \varepsilon_i) = 0$ is now:

(A) Independence: $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$

(B) Exclusion restriction: $Y_i(Z_i = 1, D_i) = Y_i(Z_i = 0, D_i)$

- Z_i only affects Y_i through D_i
- This lets us write:

$$Y_i(1) = Y_i(D_i = 1, Z_i = 1) = Y_i(D_i = 1, Z_i = 0)$$

$$Y_i(0) = Y_i(D_i = 0, Z_i = 1) = Y_i(D_i = 0, Z_i = 0)$$

We can combine these two to express:

$$\begin{aligned} Y_i &= Y_i(D_i = 0, Z_i) + (Y_i(D_i = 1, Z_i) - Y_i(D_i = 0, Z_i))D_i \\ &= Y_i(0) + (Y_i(1) - Y_i(0))D_i \end{aligned}$$

With this framework, we need some (new) assumptions

We'll make four assumptions:

1. **First stage:** $E[D_i|Z_i = 1] \neq E[D_i|Z_i = 0]$ for some i
   - This is the same as before: $Cov(D_i, Z_i) \neq 0$
2. **Independence:** $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$
3. **Exclusion restriction:** $Y_i(Z_i = 1, D_i) = Y_i(Z_i = 0, D_i)$
4. **Monotonicity:** $D_i(Z_i = 1) - D_i(Z_i = 0) \geq 0$ for all i

Monotonicity

This new assumption says:

$$D_i(Z_i = 1) - D_i(Z_i = 0) \geq 0 \text{ for all } i$$

- While Z_i need not move everybody's treatment status...
- ... all affected units move in the same way

Monotonicity

This new assumption says:

$$D_i(Z_i = 1) - D_i(Z_i = 0) \geq 0 \text{ for all } i$$

- While Z_i need not move everybody's treatment status...
- ... all affected units move in the same way
- Either $D_i(Z_i = 1) \geq D_i(Z_i = 0)$ for all i
- Or $D_i(Z_i = 1) \leq D_i(Z_i = 0)$ for all i
- Moving from $Z_i = 0$ to $Z_i = 0$ doesn't move some units from $D_i = 0$ to $D_i = 1$ and others from $D_i = 1$ to $D_i = 0$

With this framework, we need some (new) assumptions

We'll make four assumptions:

1. **First stage:** $E[D_i|Z_i = 1] \neq E[D_i|Z_i = 0]$ for some i
   - This is the same as before: $Cov(D_i, Z_i) \neq 0$
2. **Independence:** $Y_i(D_i, Z_i), D_i(1), D_i(0) \perp Z_i$
3. **Exclusion restriction:** $Y_i(Z_i = 1, D) = Y_i(Z_i = 0, D)$ for $D \in \{0, 1\}$
4. **Monotonicity:** $D_i(Z_i = 1) - D_i(Z_i = 0) \geq 0$ for all i

What do these assumptions buy us?

As always, we'd (ideally) estimate the following regression:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$

Since D_i is not randomly assigned, we also need an instrument, Z_i
Recall that we can estimate $\hat{\tau}^{IV}$ using two regressions:

$$\underbrace{D_i = \alpha + \gamma Z_i + \eta_i}_{\text{first stage}}$$

and

$$\underbrace{Y_i = \alpha + \theta Z_i + \nu_i}_{\text{reduced form}}$$

Then

$$\hat{\tau}^{IV} = \frac{\hat{\theta}}{\hat{\gamma}} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

What do these assumptions buy us?

Let's decompose $\hat{\tau}^{IV} = \frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}$:

$$E[Y_i|Z_i = 1] = \underbrace{E[Y_i(0) + (Y_i(1) - Y_i(0))D_i|Z_i = 1]}_{\text{exclusion restriction}}$$
$$= \underbrace{E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(Z_i = 1)]}_{\text{independence}}$$

and

$$E[Y_i|Z_i = 0] = \underbrace{E[Y_i(0) + (Y_i(1) - Y_i(0))D_i|Z_i = 0]}_{\text{exclusion restriction}}$$
$$= \underbrace{E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(Z_i = 0)]}_{\text{independence}}$$

What do these assumptions buy us?

Taken together, these two yield

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))]$$
$$= \underbrace{E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]Pr(D_i(1) > D_i(0))}_{\text{monotonicity}}$$

where $E[Y_i(1) - Y_i(0)$ is some kind of treatment effect
$|D_i(1) > D_i(0)]$: for compliers only
$Pr(D_i(1) > D_i(0))$: share of compliers in the population.

What do these assumptions buy us?

Taken together, these two yield

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))]$$
$$= \underbrace{E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]Pr(D_i(1) > D_i(0))}_{\text{monotonicity}}$$

where $E[Y_i(1) - Y_i(0)$ is some kind of treatment effect
$|D_i(1) > D_i(0)]$: for compliers only
$Pr(D_i(1) > D_i(0))$: share of compliers in the population.
By the same logic:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = \underbrace{E[D_i(1) - D_i(0)]}_{\text{independence}}$$

$$= \underbrace{Pr(D_i(1) > D_i(0))}_{\text{monotonicity}}$$

What do these assumptions buy us?

Taken together, these two yield

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))]$$
$$= \underbrace{E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]Pr(D_i(1) > D_i(0))}_{\text{monotonicity}}$$

where $E[Y_i(1) - Y_i(0)$ is some kind of treatment effect
$|D_i(1) > D_i(0)]$: for compliers only
$Pr(D_i(1) > D_i(0))$: share of compliers in the population.
By the same logic:

$$E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = \underbrace{E[D_i(1) - D_i(0)]}_{\text{independence}}$$

$$= \underbrace{Pr(D_i(1) > D_i(0))}_{\text{monotonicity}}$$

$$\hat{\tau}^{IV} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]$$

What happens without monotonicity?

Monotonicity, $D_i(Z_i = 1) - D_i(Z_i = 0) \geq 0$ for all i, is a new assumption

- Without it, we have $D_i(Z_i = 1) - D_i(Z_i = 0) < 0$ for some i
- This breaks our ability to estimate τ^{LATE} using $\hat{\tau}^{IV}$

What happens without monotonicity?

Monotonicity, $D_i(Z_i = 1) - D_i(Z_i = 0) \geq 0$ for all i, is a new assumption

- Without it, we have $D_i(Z_i = 1) - D_i(Z_i = 0) < 0$ for some i
- This breaks our ability to estimate τ^{LATE} using $\hat{\tau}^{IV}$
- We had:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0)]$$

- But without monotonicity:

$$\begin{aligned} &E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] \\ &= E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]Pr(D_i(1) > D_i(0)) \\ &\quad -E[Y_i(1) - Y_i(0)|D_i(1) < D_i(0)]Pr(D_i(1) < D_i(0)) \end{aligned}$$

What happens without monotonicity?

Monotonicity, $D_i(Z_i = 1) - D_i(Z_i = 0) \geq 0$ for all i, is a new assumption

- Without it, we have $D_i(Z_i = 1) - D_i(Z_i = 0) < 0$ for some i
- This breaks our ability to estimate τ^{LATE} using $\hat{\tau}^{IV}$
- We had:

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0)]$$

- But without monotonicity:

$$\begin{aligned} &E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] \\ &= E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]Pr(D_i(1) > D_i(0)) \\ &\quad - E[Y_i(1) - Y_i(0)|D_i(1) < D_i(0)]Pr(D_i(1) < D_i(0)) \end{aligned}$$

$\rightarrow$ We can't deal with this

- τ^i could be > 0 for all i, but we could mistakenly estimate 0 effect

$\rightarrow$ We would have **defiers** (☠)

$\hat{\tau}^{IV} = E[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)]$

What is this "conditional on $D_i(1) > D_i(0)$" beast?

- $\hat{\tau}^{IV}$ estimates the (L)ATE, conditional on $D_i(1) > D_i(0)$
- $D_i(1) > D_i(0)$ means Z_i moves D_i

$\hat{\tau}^{IV} = E[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)]$

What is this "conditional on $D_i(1) > D_i(0)$" beast?

- $\hat{\tau}^{IV}$ estimates the (L)ATE, conditional on $D_i(1) > D_i(0)$
- $D_i(1) > D_i(0)$ means Z_i moves D_i
- We can divide the world into three groups:
  1. $D_i(1) > D_i(0)$: Compliers
  2. $D_i(1) = D_i(0) = 1$: Always-takers
  3. $D_i(1) = D_i(0) = 0$: Never-takers

$\hat{\tau}^{IV} = E[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)]$

What is this "conditional on $D_i(1) > D_i(0)$" beast?

- $\hat{\tau}^{IV}$ estimates the (L)ATE, conditional on $D_i(1) > D_i(0)$
- $D_i(1) > D_i(0)$ means Z_i moves D_i
- We can divide the world into three groups:
  1. $D_i(1) > D_i(0)$: Compliers
  2. $D_i(1) = D_i(0) = 1$: Always-takers
  3. $D_i(1) = D_i(0) = 0$: Never-takers

→ Note that Z_i doesn't affect D_i for never-takers or always-takers

→ The instrument is useless for them

→ We can't learn about their treatment effects!

→ (They essentially have no first stage)

→ We can estimate LATEs for compliers only

Non-compliance throwback

We looked at several scenarios of non-compliance:

- If only T can non-comply, we can show:

$$\hat{\tau}^{IV} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1]} = E[Y_i(1) - Y_i(0)|D_i = 1] = \tau^{LATE}$$

- If only C can non-comply, we can show:

$$\hat{\tau}^{IV} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1]} = E[Y_i(1) - Y_i(0)|D_i = 0] = \tau^{LATE}$$

Non-compliance throwback

If both T and C can non-comply:

$$\hat{\tau}^{IV} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]}$$

$$= E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] = \tau^{LATE}$$

Why does this work?

- We have an as-good-as-random estimate, $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]$
- We need to scale this by the complier proportion

Counting compliers

The fraction of compliers is just:

$$\pi^C = Pr(D_i(1) > D_i(0)) = E[D_i(1) - D_i(0)]$$

$$= E[D_i(1)] - E[D_i(0)]$$

$$= E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$

Counting compliers

The fraction of compliers is just:

$$\pi^C = Pr(D_i(1) > D_i(0)) = E[D_i(1) - D_i(0)]$$

$$= E[D_i(1)] - E[D_i(0)]$$

$$= E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$$

We can also count the fraction of the treatment group which complies:

$$Pr(D_i(1) > D_i(0)|D_i = 1) = \frac{Pr(D_i = 1|D_i(1)) > D_i(0))Pr(D_i(1) > D_i(0)}{Pr(D_i = 1)}$$

$$= \frac{Pr(Z_i = 1)(E[D_i|Z_i = 1] - E[D_i|Z_i = 0])}{Pr(D_i = 1)}$$

Who are the LATE compliers?

- We can't pick out individual compliers
- We can just count them
- But we can actually learn something more about them!

Let's ask: are compliers more likely to be men?

Who are the LATE compliers?

- We can't pick out individual compliers
- We can just count them
- But we can actually learn something more about them!

Let's ask: are compliers more likely to be men?

$$\frac{Pr(Male_i = 1 | D_i(1) > D_i(0)}{Pr(Male_i = 1)}$$

Who are the LATE compliers?

- We can't pick out individual compliers
- We can just count them
- But we can actually learn something more about them!

Let's ask: are compliers more likely to be men?

$$\frac{Pr(Male_i = 1 | D_i(1) > D_i(0))}{Pr(Male_i = 1)}$$

$$= \frac{Pr(D_i(1) > D_i(0) | Male_i = 1)}{Pr(D_i(1) > D_i(0))}$$

Who are the LATE compliers?

- We can't pick out individual compliers
- We can just count them
- But we can actually learn something more about them!

Let's ask: are compliers more likely to be men?

$$\frac{Pr(Male_i = 1 | D_i(1) > D_i(0))}{Pr(Male_i = 1)}$$

$$= \frac{Pr(D_i(1) > D_i(0) | Male_i = 1)}{Pr(D_i(1) > D_i(0))}$$

$$= \frac{E[D_i | Z_i = 1, Male_i = 1 - E[D_i | Z_i = 0, Male_i = 1]}{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]}$$

Who are the LATE compliers?

- We can't pick out individual compliers
- We can just count them
- But we can actually learn something more about them!

Let's ask: are compliers more likely to be men?

$$\frac{Pr(Male_i = 1 | D_i(1) > D_i(0)}{Pr(Male_i = 1)}$$

$$= \frac{Pr(D_i(1) > D_i(0) | Male_i = 1)}{Pr(D_i(1) > D_i(0))}$$

$$= \frac{E[D_i | Z_i = 1, Male_i = 1 - E[D_i | Z_i = 0, Male_i = 1]}{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]}$$

$\rightarrow$ This is just the first stage for men divided by the overall first stage!

What happens if we have multiple instruments?

Heterogeneous τ_i makes things interesting:

- With homogenous τ_i, all instruments should yield the same τ^{LATE}
- With heterogeneous τ, this need not be true!

What happens if we have multiple instruments?

Heterogeneous τ_i makes things interesting:

- With homogenous τ_i, all instruments should yield the same τ^{LATE}
- With heterogeneous τ, this need not be true!

With multiple instruments, we get multiple estimates of

$$E[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)]$$

Each instrument $Z_i^1, ... Z_i^K$ will have its own compliers where $D_i(1) > D_i(0)$

What happens if we have multiple instruments?

With some small tweaks, we can still estimate $\hat{\tau}^{IV}$:

What happens if we have multiple instruments?

With some small tweaks, we can still estimate $\hat{\tau}^{IV}$:

- We now estimate the first stage as:

$$D_i = \alpha + \pi_1 Z_i^1 + \pi_2 Z_i^2 + \nu_i$$

What happens if we have multiple instruments?

With some small tweaks, we can still estimate $\hat{\tau}^{IV}$:

- We now estimate the first stage as:

$$D_i = \alpha + \pi_1 Z_i^1 + \pi_2 Z_i^2 + \nu_i$$

- As usual, the second stage is just:

$$Y_i = \alpha + \tau \hat{D}_i + \varepsilon_i$$

What happens if we have multiple instruments?

With some small tweaks, we can still estimate $\hat{\tau}^{IV}$:

- We now estimate the first stage as:

$$D_i = \alpha + \pi_1 Z_i^1 + \pi_2 Z_i^2 + \nu_i$$

- As usual, the second stage is just:

$$Y_i = \alpha + \tau \hat{D}_i + \varepsilon_i$$

- And the 2SLS estimator will be:

$$\hat{\tau}^{2SLS} = \frac{Cov(Y_i, \hat{D}_i)}{Cov(D_i, \hat{D}_i)} = \frac{\pi_1 Cov(Y_i, Z_i^1)}{Cov(D_i, \hat{D}_i)} + \frac{\pi_2 Cov(Y_i, Z_i^2)}{Cov(D_i, \hat{D}_i)}$$

What happens if we have multiple instruments?

With some small tweaks, we can still estimate $\hat{\tau}^{IV}$:

- We now estimate the first stage as:

$$D_i = \alpha + \pi_1 Z_i^1 + \pi_2 Z_i^2 + \nu_i$$

- As usual, the second stage is just:

$$Y_i = \alpha + \tau \hat{D}_i + \varepsilon_i$$

- And the 2SLS estimator will be:

$$\hat{\tau}^{2SLS} = \frac{Cov(Y_i, \hat{D}_i)}{Cov(D_i, \hat{D}_i)} = \frac{\pi_1 Cov(Y_i, Z_i^1)}{Cov(D_i, \hat{D}_i)} + \frac{\pi_2 Cov(Y_i, Z_i^2)}{Cov(D_i, \hat{D}_i)}$$

$\rightarrow$ This is just a weighted average of each instrument's $\hat{\tau}^{IV}$

Non-binary treatments

What happens with non-binary treatment?

- Binary treatment: there is only $Y_i(1)$ and $Y_i(0)$

Non-binary treatments

What happens with non-binary treatment?

- Binary treatment: there is only $Y_i(1)$ and $Y_i(0)$
- Non-binary treatment: define $S_i \in \{0, 1, ..., \bar{S}\}$
- This has many potential outcomes $Y_i(0), Y_i(1), ..., Y_i(\bar{S})$
- And many causal effects: $Y_i(1) - Y_i(0), Y_i(2) - Y_i(1)...$

Non-binary treatments

<u>What happens with non-binary treatment?</u>

- Binary treatment: there is only $Y_i(1)$ and $Y_i(0)$
- Non-binary treatment: define $S_i \in \{0, 1, ..., \bar{S}\}$
- This has many potential outcomes $Y_i(0), Y_i(1), ..., Y_i(\bar{S})$
- And many causal effects: $Y_i(1) - Y_i(0), Y_i(2) - Y_i(1)...$
- In a linear model, these are all the same
- But that's unrealistic

$\rightarrow$ 2SLS to the rescue!

The average causal response

We can get a weighted average response with some assumptions:

- Independence + exclusion: $\{Y_i(0), Y_i(1), ...Y_i(\bar{S})\} \perp Z_i$
- First stage: $E[S_i(1) - S_i(0)] \neq 0$
- Monotonicity: $S_i(1) - S_i(0) \geq 0$ for all i (or vice versa)

The average causal response

We can get a weighted average response with some assumptions:

- Independence + exclusion: $\{Y_i(0), Y_i(1), ... Y_i(\bar{S})\} \perp Z_i$
- First stage: $E[S_i(1) - S_i(0)] \neq 0$
- Monotonicity: $S_i(1) - S_i(0) \geq 0$ for all i (or vice versa)

Then:

$$\hat{\tau}^{IV} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]}$$

$$= \sum_{s=1}^{\bar{S}} \omega_s E[Y_i(s) - Y_i(s-1)|S_i(1) \geq s \geq S_i(0)]$$

where

$$\omega_s = \frac{Pr(S_i(1) \geq s > S_i(0)}{\sum_{j=1}^{\bar{S}} Pr(S_i(1) \geq j > S_i(0))}$$

The average causal response

$$\hat{\tau}^{IV} = \sum_{s=1}^{\bar{S}} \omega_s E[Y_i(s) - Y_i(s-1)|S_i(1) \geq s \geq S_i(0)]$$

→ $\hat{\tau}^{IV}$ gives a weighted average of the unit causal response

→ The unit causal response, $E[Y_i(s) - Y_i(s-1)|S_i(1) \geq s \geq S_i(0)]$ is the average difference in potential outcomes for compliers at $S_i = s$

→ The size of the compliance group is $Pr(S_i(1) \geq s > S_i(0)$

What do we get from the IV?

We've talked through several cases

- Constant τ:
  - $\hat{\tau}^{IV} = \tau^{ATE}$
- Perfect compliance:
  - $\hat{\tau}^{IV} = \tau^{ATE}$
- Heterogeneous treatment effects, one IV:
  - $\hat{\tau}^{IV} = \tau^{LATE}$
- Heterogeneous treatment effects, multiple IVs:
  - $\hat{\tau}^{IV} = \frac{1}{K}\sum_k \omega_k \tau_k^{LATE}$
- Multiple values of treatment:
  - $\hat{\tau}^{IV} = \sum_{s=1}^{\bar{S}} \omega_s E[Y_i(s) - Y_i(s-1) | S_i(1) \geq s \geq S_i(0)]$

Taking stock of IV

We've come a long way from RCTs:

- Took a brief detour through the thicket of SOO
- Started our discussion of SOU

Taking stock of IV

We've come a long way from RCTs:

- Took a brief detour through the thicket of SOO
- Started our discussion of SOU

→ **IV** is our first SOU design

- IV helps us do causal inference with non-random treatment
- We just need some random leverage over treatment

Taking stock of IV

Under the right assumptions, we can use IV for...

- Eliminating bias due to measurement error
- Eliminating bias due to omitted variables
- Eliminating bias due to simultaneity
- Translating from ITT to LATE
- Estimating (L)ATEs

The trick is satisfying the exclusion restriction!

Recap

TL;DR:

1. Instrumental variables are very powerful
2. We can use them to handle non-compliance
3. More generally, the IV estimates LATE (not ATE) with heterogeneity