Testing the LATE Assumptions

Francis DiTraglia

Oxford Economics Summer School 2023

Review of "Textbook" Instrumental Variables (IV) Model

Observed

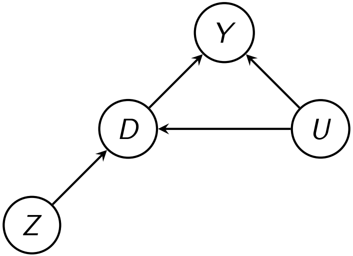
- Y = Outcome (Wage)
- D = Treatment (Attend Uni)
- Z = IV (Live Nearby)

Unobserved

- U = Confounders (Ability)

Assumptions

- Model: $Y = \alpha + \beta D + U$
- Relevance: $\text{Cov}(Z, D) \neq 0$
- Exogeneity: $\text{Cov}(Z, U) = 0$



A Relevant, Exogenous Instrument Identifies β

Assumptions

$$Y = \alpha + \beta D + U, \quad \text{Cov}(Z, D) \neq 0, \quad \text{Cov}(Z, U) = 0$$

OLS

$$\beta_{\text{OLS}} = \frac{\text{Cov}(D, Y)}{\text{Var}(D)} = \frac{\beta \text{Var}(D) + \text{Cov}(D, U)}{\text{Var}(D)} = \beta + \frac{\text{Cov}(D, U)}{\text{Var}(D)} \neq \beta$$

IV

$$\beta_{\text{IV}} = \frac{\text{Cov}(Z, Y)}{\text{Cov}(Z, D)} = \frac{\beta \text{Cov}(Z, D) + \text{Cov}(Z, U)}{\text{Cov}(Z, D)} = \beta + \frac{\text{Cov}(Z, U)}{\text{Cov}(Z, D)} = \beta$$

```r
library(mvtnorm)
library(tidyverse)
set.seed(587103)
n <- 1e4
sims <- rmvnorm(n, sigma = matrix(c(1, 0.5,
                                     0.5, 1), 2, 2, byrow = TRUE))
U <- sims[,1]
V <- sims[,2]
Z <- rbinom(n, size = 1, prob = 0.3)
D <- (-0.5) + 0.3 * Z + V
beta <- (-0.3)
Y <- 1 + beta * D + U
c(OLS = cov(D, Y) / var(D), IV = cov(Z, Y) / cov(Z, D),
  truth = beta) |> round(2)
```

```
##   OLS    IV truth
##  0.20 -0.28 -0.30
```

Which assumptions are testable in the textbook IV model?

Instrument Relevance

- Since D and Z are observed, directly estimate $\text{Cov}(D, Z)$.
- Beware of weak instruments!

Instrument Exogeneity

- Since U is unobserved, can't directly estimate $\text{Cov}(Z, U)$.
- Could we use the IV residuals?

Simulation with a Bad Instrument

It has a direct effect on Y separate from its effect on D!

```
library(mvtnorm); library(tidyverse); library(broom); library(AER)
set.seed(587103)
n <- 1e5
sims <- rmvnorm(n, sigma = matrix(c(1, 0.5,
                                    0.5, 1), 2, 2, byrow = TRUE))
U <- sims[,1]
V <- sims[,2]
Z <- rbinom(n, size = 1, prob = 0.3)
D <- -0.5 + 0.3 * Z + V
beta <- 0
Y <- 1 + beta * D - Z + U # Instrument isn't excluded!
```

Bad Instrument Is Uncorrelated with IV Residuals!

```r
iv_results <- ivreg(Y ~ D | Z)
tidy(iv_results) |> knitr::kable(digits = 2)
```

term	estimate	std.error	statistic	p.value
(Intercept)	-0.72	0.04	-17.31	0
D	-3.45	0.10	-35.90	0

```r
cov(residuals(iv_results), Z)
```

```
## [1] -1.534378e-16
```

Z Is Uncorrelated with the IV Residuals *By Construction*

- Let U be the **structural error** and V be the **IV residual**: $V \equiv Y - \alpha_{IV} - \beta_{IV} D$.

$$\beta_{IV} = \frac{\text{Cov}(Z,Y)}{\text{Cov}(Z,D)} = \beta + \frac{\text{Cov}(Z,U)}{\text{Cov}(Z,D)}, \quad \alpha_{IV} = \mathbb{E}(Y) - \beta_{IV}\mathbb{E}(D).$$

- $V = U \iff Z$ is exogenous: the only way to obtain $\beta_{IV} = \beta$ and $\alpha_{IV} = \alpha$.

$$\begin{aligned}\text{Cov}(Z,V) = \text{Cov}(Z, Y - \alpha_{IV} - \beta_{IV} D) &= \text{Cov}(Z,Y) - \beta_{IV}\text{Cov}(Z,D) \\ &= \text{Cov}(Z,Y) - \frac{\text{Cov}(Z,Y)}{\text{Cov}(Z,D)}\text{Cov}(Z,D) = 0.\end{aligned}$$

- $\text{Cov}(Z,V) = 0$ *by construction* even when $\text{Cov}(Z,U) \neq 0$

Multiple Instruments and Over-identification

Assumptions

- $Y = \alpha + \beta D + U$
- $\text{Cov}(Z_1, D) \neq 0, \ \text{Cov}(Z_2, D) \neq 0$
- $\text{Cov}(Z_1, U) = \text{Cov}(Z_2, U) = 0$

Implications

- Both IVs identify *same* effect: β
- If not, **at least one is endogenous**

Over-identifying Restrictions Test

- Test of null that all MCs identify same parameters.

$$\beta_{IV}^{(1)} \equiv \frac{\text{Cov}(Z_1, Y)}{\text{Cov}(Z_1, D)} = \beta + \frac{\text{Cov}(Z_1, U)}{\text{Cov}(Z_1, D)}$$

$$\beta_{IV}^{(2)} \equiv \frac{\text{Cov}(Z_2, Y)}{\text{Cov}(Z_2, D)} = \beta + \frac{\text{Cov}(Z_2, U)}{\text{Cov}(Z_2, D)}$$

$$\beta_{IV}^{(1)} - \beta_{IV}^{(2)} = \frac{\text{Cov}(Z_1, U)}{\text{Cov}(Z_1, D)} - \frac{\text{Cov}(Z_2, U)}{\text{Cov}(Z_2, D)}$$

Beyond the Textbook IV Model

Heterogenous Treatment Effects

- $Y = \alpha + \beta D + U$ implies that everyone has the same treatment effect: β.
- In reality, treatment effects differ across people.

Overidentifying restrictions?

- Out the window! Different instruments may identify different causal parameters.

Local Average Treatment Effects (LATE) Model

- What does IV tell us when treatment effects are heterogeneous?

Review of the LATE Model

- Suppose that both D and Z are binary

$$\beta_{\text{IV}} \equiv \frac{\text{Cov}(Z,Y)}{\text{Cov}(Z,D)} = \frac{\frac{\text{Cov}(Y,Z)}{\text{Var}(Z)}}{\frac{\text{Cov}(D,Z)}{\text{Var}(Z)}} = \frac{\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]}{\mathbb{E}[D|Z=1] - \mathbb{E}[D|Z=0]} \equiv \text{Wald Estimand}$$

Intent-to-treat Effect: $\mathbb{E}[Y|Z=1] - \mathbb{E}[Y|Z=0]$ (ITT)

- E.g. randomized experiment with treatment **offer** Z and treatment **take-up** D
- **Non-compliance** / randomized encouragement design: D may not equal Z
- In this setting the ITT is the ATE of *offering* treatment.

The Wald Estimand

- ITT is "diluted" by people who are offered $(Z=1)$ but do not take up $(D=0)$
- Divide ATE of offer on outcome $Z \to Y$ by that of offer on take-up $Z \to D$.
- **Under what assumptions does this give us a meaningful causal quantity?**

Decomposing the ITT Effect

- Example: moving to opportunity (MTO) experiment randomly offered housing vouchers to encourage families to move to a more affluent neighborhood.
- 50% of offered families ($Z = 1$) moved; 20% of non-offered families ($Z = 0$) moved

$$Y = (1 - D)Y_0 + DY_1, \quad p_z \equiv \mathbb{P}(D = 1|Z = z)$$

- $\mathbb{E}[Y|Z = 1]$ is a *mixture* of Y_0 and Y_1 for different types of families:

$$\mathbb{E}[Y|Z = 1] = \underbrace{(1 - p_1)}_{\approx 0.5} \mathbb{E}\left[Y_0|Z = 1, D = 0\right] + \underbrace{p_1}_{\approx 0.5} \mathbb{E}\left[Y_1|Z = 1, D = 1\right]$$

- $\mathbb{E}[Y|Z = 0]$ is a *mixture* of Y_0 and Y_1 for different types of families:

$$\mathbb{E}[Y|Z = 0] = \underbrace{(1 - p_0)}_{\approx 0.8} \mathbb{E}\left[Y_0|Z = 0, D = 0\right] + \underbrace{p_0}_{\approx 0.2} \mathbb{E}\left[Y_1|Z = 0, D = 1\right]$$

Compliance "Types" in the LATE Model

- Catalogue all possible treatment take-up "decision rules"

$$\begin{aligned} \text{Never-taker:} \quad & T = n \quad \iff \quad D(Z) = 0 \\ \text{Always-taker:} \quad & T = a \quad \iff \quad D(Z) = 1 \\ \text{Complier:} \quad & T = c \quad \iff \quad D(Z) = Z \\ \text{Defier:} \quad & T = d \quad \iff \quad D(Z) = (1 - Z). \end{aligned}$$

In the MTO Example

- Never-takers: families that refuse to move with or without a voucher
- Always-takers: families that will move with or without a voucher
- Compliers are families that will **only move if given a voucher**
- Defiers are families that will only move if **not** given a voucher

Assumption 1 - Unconfounded Type

For all compliance types $t \in \{a, c, n, d\}$

$$\mathbb{P}(T = t) = \mathbb{P}(T = t|Z = 0) = \mathbb{P}(T = t|Z = 1).$$

Assumption 2 - No Defiers: $\mathbb{P}(T = d) = 0$

Assumption 3 - Mean Exclusion Restriction

For all compliance types $t \in \{a, c, n, d\}$

$$\mathbb{E}\left[Y_0|Z = 0, T = t\right] = \mathbb{E}\left[Y_0|Z = 1, T = t\right] = \mathbb{E}\left[Y_0|T = t\right]$$
$$\mathbb{E}\left[Y_1|Z = 0, T = t\right] = \mathbb{E}\left[Y_1|Z = 1, T = t\right] = \mathbb{E}\left[Y_1|T = t\right]$$

Assumption 4 - Existence of Compliers: $\mathbb{P}(T = c) > 0$

Lemma 1: Assumptions 1–2 $\implies$

$$
\begin{aligned}
\mathbb{P}(D=1|Z=1) &= \mathbb{P}(T=a)+\mathbb{P}(T=c) \\
\mathbb{P}(D=0|Z=1) &= \mathbb{P}(T=n) \\
\mathbb{P}(D=1|Z=0) &= \mathbb{P}(T=a) \\
\mathbb{P}(D=0|Z=0) &= \mathbb{P}(T=c)+\mathbb{P}(T=n)
\end{aligned}
$$

Lemma 2: Assumptions 1–3 $\implies$

$$
\begin{aligned}
\mathbb{E}\left[Y|D=1,Z=1\right] &= \frac{\mathbb{P}(T=a)\mathbb{E}\left[Y_1|T=a\right]+\mathbb{P}(T=c)\mathbb{E}\left[Y_1|T=c\right]}{\mathbb{P}(T=a)+\mathbb{P}(T=c)} \\
\mathbb{E}\left[Y|D=0,Z=1\right] &= \mathbb{E}\left[Y_0|T=n\right] \\
\mathbb{E}\left[Y|D=1,Z=0\right] &= \mathbb{E}\left[Y_1|T=a\right] \\
\mathbb{E}\left[Y|D=0,Z=0\right] &= \frac{\mathbb{P}(T=n)\mathbb{E}\left[Y_0|T=n\right]+\mathbb{P}(T=c)\mathbb{E}\left[Y_0|T=c\right]}{\mathbb{P}(T=n)+\mathbb{P}(T=c)}
\end{aligned}
$$

The LATE Theorem: Wald = ATE for Compliers

Theorem: Assumptions 1–4 $\implies$

$$\frac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=0)}{\mathbb{E}(D|Z=1) - \mathbb{E}(D|Z=0)} = \mathbb{E}\left[Y_1 - Y_0 | T = c\right]$$

Proof

- Algebra and of Iterated Expectations, using the two lemmas. (See lecture notes)

MTO Example

- ITT is the average treatment effect of *offering* a housing voucher.
- Wald = LATE is the average treatment effect of *moving to opportunity* for families who can be induced to move by a housing voucher.
- Different IV $\implies$ different compliers $\implies$ different LATE. It's a **local** effect!

Are the LATE Assumptions Testable?

LATE Assumptions

1. Unconfounded Type
2. No Defiers
3. Mean Exclusion Restriction
4. Existence of Compliers

At Least One is Testable!

- Assumptions 1–3 $\implies \mathbb{P}(T = c) = \mathbb{E}[D|Z = 1] - \mathbb{E}[D|Z = 0]$
- Thus, Assumption 4 is just **instrument relevance**, hence testable.
- What about the others?

Even Nobel Laureates Make Mistakes

Angrist & Imbens (1994)

Part (i) is similar to an exclusion restriction in a regression model. It is not testable and has to be considered on a case by case basis.

Pearl (1995)

exogeneity . . . can be given an empirical test. The test is not guaranteed to detect all violations of exogeneity, but it can, in certain circumstances, screen out very bad would-be instruments.

This Lecture

- Testable implications LATE assumptions from above: Huber & Mellace (2015)

Closely-related Work

- Kitagawa (2015)
- Mourifié & Wan (2017)

Huber & Mellace (2015) – The Big Picture

- Assumptions 1–3 imply four inequalities: $\theta_1 \leq 0$, $\theta_2 \leq 0$, $\theta_3 \leq 0$, $\theta_4 \leq 0$
- $\theta \equiv (\theta_1, \theta_2, \theta_3, \theta_4)$ depend only on (Y, D, Z); we'll define them shortly.
- If any element of θ is *positive* at least one of Assumptions 1–3 must be false.
- In practice: compare estimate $\widehat{\theta}$ to appropriate standard errors.
- Not all violations of the LATE assumptions lead to a positive value for θ
- **Necessary but not sufficient** for validity of LATE assumptions.
- The four inequalities come in *pairs*. We'll look at each pair in turn.

Example: Card (1995)[1]

- $Y = \log(Wage)$, $D =$ College, $Z =$ Live Nearby

```
library(tidyverse); library(wooldridge); library(estimatr)
card1995 <- as_tibble(card) |>
  mutate(Y = lwage,    # log(wage)
         Z = nearc4,   # Live near 4 year college?
         D = 1 * (educ >= 16)) |> # Attend college? (>=16 years educ)
  select(Y, Z, D)

iv <- iv_robust(Y ~ D | Z, card1995)
ols <- lm_robust(Y ~ D, card1995)
```

[1] Using geographic variation in college proximity to estimate the return to schooling

IV Estimate is Implausibly Large

```r
library(modelsummary)
modelsummary(list(OLS = ols, IV = iv), output = 'latex', fmt = 2,
             gof_omit = 'Num.Obs.|R2|R2 Adj.|AIC|BIC|RMSE',
             coef_omit = '(Intercept)')
```

	OLS	IV
D	0.23	2.27
	(0.02)	(0.55)

Remember: this is on the log scale!

Example of the Huber & Mellace (2015) Approach

- Suppose that all of the LATE assumptions hold and define:

$$r \equiv \frac{\mathbb{P}(T = n)}{\mathbb{P}(T = c) + \mathbb{P}(T = n)} = \frac{\mathbb{P}(D = 0|Z = 1)}{\mathbb{P}(D = 0|Z = 0)} \qquad \text{(by Lemma 1)}$$

- Distribution of $Y|(D = 0, Z = 0)$ is a mixture of Y_0 for compliers and never-takers.
- The mixture contains $r \times 100\%$ never-takers and $(1 - r) \times 100\%$ compliers.
- Let's calculate r in the Card (1995) example:

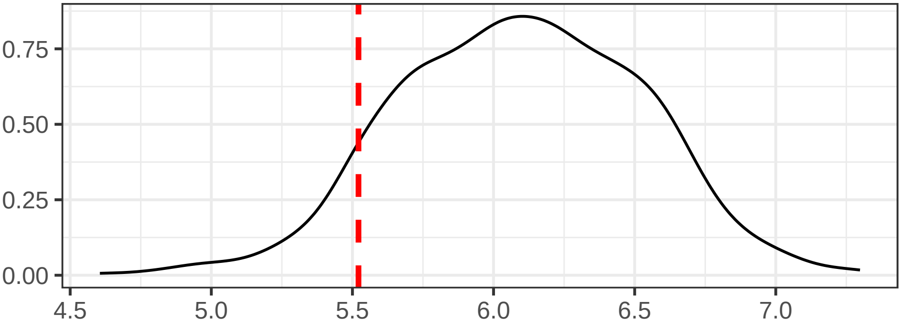
```
r <- card1995 |>
  summarize(p01 = sum(D == 0 & Z == 1) / sum(Z == 1),
            p00 = sum(D == 0 & Z == 0) / sum(Z == 0),
            r = p01 / p00) |> pull(r)
r
```

```
## [1] 0.9115626
```

Density of $Y|(D = 0, Z = 0)$ from Card (1995)

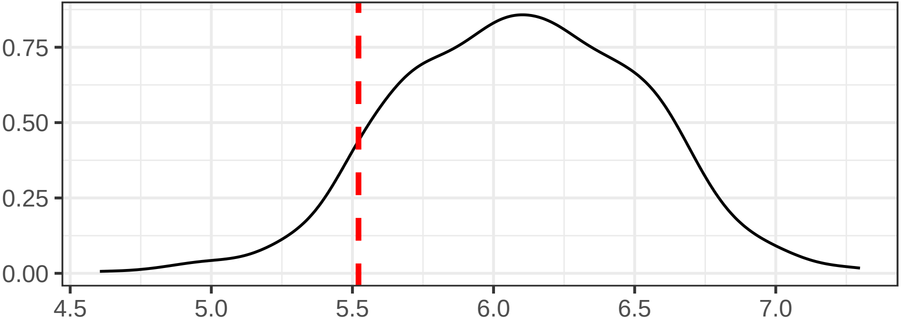
```
y00_density <- card1995 |> filter(D == 0, Z == 0) |>
  ggplot(aes(x = Y)) +
  geom_density() +
  geom_vline(aes(xintercept = quantile(Y, 1 - r)),
             color = 'red', linetype = 'dashed', size = 1) +
  xlab('') + ylab('') +
  theme_bw()
```

Density of $Y|(D = 0, Z = 0)$ from Card (1995)



- This is the density of Y_0 for a mix of never-takers and compliers.
- The mix contains 91% never-takers. But we don't know where they are.
- Dashed red line: 9th %-tile of the density.
- If *all* never-takers are at the top of the distribution, they're above this line.

Density of Y_0 for a mixture containing 91% never-takers, 9% compliers



- If all never-takers are at the top of the distribution, they're above the red line.
- Mean of all observations *above* red line bounds $\mathbb{E}[Y_0|T=n]$ **from above**
- But Lemma 2 shows that $\mathbb{E}(Y_0|T=n) = \mathbb{E}(Y|D=0, Z=1)$.
- If this contradicts the upper bound **something must be wrong**.

Contradiction $\implies$ LATE Assumptions Fail

Previous Slide: $\mathbb{E}(Y_0|T=n) \leq \mathbb{E}(Y|D=0, Z=0, Y \geq y_{1-r})$

```
card1995 |>  filter(D == 0, Z == 0) |>
  summarize(ninth_percentile = quantile(Y, 1 - r),
            upper_bound = mean(Y[Y >= ninth_percentile])) |>
  pull(upper_bound)
```

```
## [1] 6.154926
```

Lemma 2: $\mathbb{E}(Y_0|T=n) = \mathbb{E}(Y|D=0, Z=1)$

```
card1995 |> filter(D == 0, Z == 1) |> summarize(mean(Y)) |> pull()
```

```
## [1] 6.254177
```

This contradicts the upper bound! Something must be wrong!

First Pair of Inequalities

Define: $F_{11}(y) \equiv \mathbb{P}(Y \leq y | D = 1, Z = 1)$ and

$$y_q \equiv F_{11}^{-1}(q), \quad y_{1-q} \equiv F_{11}^{-1}(1-q), \quad q \equiv \frac{\mathbb{P}(D=1|Z=0)}{\mathbb{P}(D=1|Z=1)}$$

Under Assumptions 1–3:

$$\mathbb{E}(Y|D=1, Z=1, Y \leq y_q) \leq \mathbb{E}(Y|D=1, Z=0) \leq \mathbb{E}(Y|D=1, Z=1, Y \geq y_{1-q})$$

Key Points

- Lemma 1 $\implies \mathbb{E}(Y|D=1, Z=0) = \mathbb{E}(Y_1|T=a)$ so now we have two *partial identification bounds* as well.
- Why care? **Overidentifying Restrictions**
- At most one of the pair can be violated.

Second Pair of Inequalities

Define $F_{00}(y) \equiv \mathbb{P}(Y \leq y | D = 0, Z = 0)$ and

$$y_r \equiv F_{00}^{-1}(r), \quad y_{1-r} \equiv F_{00}^{-1}(1-r), \quad r \equiv \frac{\mathbb{P}(D=0|Z=1)}{\mathbb{P}(D=0|Z=0)}.$$

Under Assumptions 1–3:

$$\mathbb{E}(Y|D=0, Z=0, Y \leq y_r) \leq \mathbb{E}(Y|D=0, Z=1) \leq \mathbb{E}(Y|D=0, Z=0, Y \geq y_{1-r})$$

Key Points

- Lemma 1 $\implies \mathbb{E}(Y|D=0, Z=1) = \mathbb{E}(Y_0|T=n)$ so now we have two *partial identification bounds* as well.
- Why care? **Overidentifying Restrictions**
- At most one of the pair can be violated.

Theorem: Assumptions 1–3 $\implies$

$$\begin{bmatrix} \theta_1 \\ \\ \theta_2 \\ \\ \theta_3 \\ \\ \theta_4 \end{bmatrix} \equiv \begin{bmatrix} \mathbb{E}(Y|D=1,Z=1,Y\le y_q) - \mathbb{E}(Y|D=1,Z=0) \\ \\ \mathbb{E}(Y|D=1,Z=0) - \mathbb{E}(Y|D=1,Z=1,Y\ge y_{1-q}) \\ \\ \mathbb{E}(Y|D=0,Z=0,Y\le y_r) - \mathbb{E}(Y|D=0,Z=1) \\ \\ \mathbb{E}(Y|D=0,Z=1) - \mathbb{E}(Y|D=0,Z=0,Y\ge y_{1-r}) \end{bmatrix} \le \begin{bmatrix} 0 \\ \\ 0 \\ \\ 0 \\ \\ 0 \end{bmatrix}$$

Where: $F_{11}(y) \equiv \mathbb{P}(Y \le y|D=1,Z=1)$, $F_{00}(y) \equiv \mathbb{P}(Y \le y|D=0,Z=0)$, and

$$y_q \equiv F_{11}^{-1}(q), \quad y_{1-q} \equiv F_{11}^{-1}(1-q), \quad q \equiv \frac{\mathbb{P}(D=1|Z=0)}{\mathbb{P}(D=1|Z=1)}$$

$$y_r \equiv F_{00}^{-1}(r), \quad y_{1-r} \equiv F_{00}^{-1}(1-r), \quad r \equiv \frac{\mathbb{P}(D=0|Z=1)}{\mathbb{P}(D=0|Z=0)}.$$