Part D: Instrumental Variables

D2: IV with Heterogeneous Effects

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2024

D2 Outline

1. Basic LATE result
2. Characterizing compliers
3. Extensions

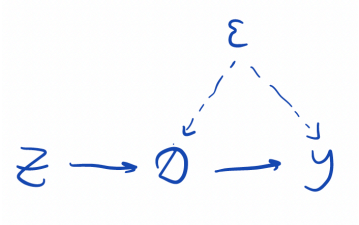
Readings: MHE Chapters 4.4, 4.5; IW Lecture 5

Internal and external validity

- In Angrist (1990), does the draft lottery IV identify a causal effect? (**Internal validity**)
- Is it the ATE for the entire population or for which subpopulation? (**External validity**)
  - Is this subpopulation of interest and policy relevance?
  - How can we describe it theoretically?
  - What can we say about it in the data?

Recap of general IV assumptions

Consider $J = K = 1$ and a causal first-stage with potential outcomes $D(z)$:



- **Exclusion**: in writing $Y_i(d)$ rather than $Y_i(d, z)$
- **Independence**: $Z_i \perp\!\!\!\perp Y_i(d)$ for all d and $Z_i \perp\!\!\!\perp D_i(z)$ for all z
- **Relevance**: $D_i(z) \neq D_i(z')$ for some z, z' with positive probability

The four groups

- Consider binary D and Z, e.g. Vietnam draft lottery
- $D_i(0)$ and $D_i(1)$ are binary $\Longrightarrow$ four latent groups:
  - $D_i(0) = D_i(1) = 0$: **never-takers**
  - $D_i(0) = 0 < D_i(1) = 1$: **compliers**
  - $D_i(0) = 1 > D_i(1) = 0$: **defiers**
  - $D_i(0) = D_i(1) = 1$: **always-takers**
- First-stage: $\pi = \mathbb{E}\left[D_i(1) - D_i(0)\right] = Pr(\text{Complier}_i) - Pr(\text{Defier}_i)$

Reduced-form

- Reduced-form (a.k.a. **intent-to-treat** effect, **ITT**):

$$
\begin{aligned}
\rho &= \mathbb{E}\left[Y_i \mid Z_i = 1\right] - \mathbb{E}\left[Y_i \mid Z_i = 0\right] \\
&= Pr(NT_i) \cdot \mathbb{E}\left[Y_i(0) - Y_i(0) \mid NT_i\right] \\
&\quad + Pr(Complier_i) \cdot \mathbb{E}\left[Y_i(1) - Y_i(0) \mid Complier_i\right] \\
&\quad + Pr(Defier_i) \cdot \mathbb{E}\left[Y_i(0) - Y_i(1) \mid Defier_i\right] \\
&\quad + Pr(AT_i) \cdot \mathbb{E}\left[Y_i(1) - Y_i(1) \mid AT_i\right] \\
&= Pr(\text{Complier}_i)\mathbb{E}\left[\tau_i \mid \text{Complier}_i\right] - Pr\left(\text{Defier}_i\right)\mathbb{E}\left[\tau_i \mid \text{Defier}_i\right]
\end{aligned}
$$

Ratio (τ_{IV}): doesn't have an intuitive interpretation, except constant effects or...

Monotonicity assumption

- Assume **monotonicity**: the instrument affects the treatment weakly in the same direction for everyone
  - Either no defiers (when $\pi > 0$) or no compliers (when $\pi < 0$)
  - Without loss, consider positive first-stage
- **LATE theorem** (Angrist and Imbens 1994): Under exclusion, independence, and monotonicity, IV identifies the (internally valid) average causal effect among compliers:

$$\tau_{IV} = \mathbb{E}\left[\tau_i \mid \text{Complier}_i\right] \equiv \text{Local average treatment effect}$$

Is monotonicity plausible?

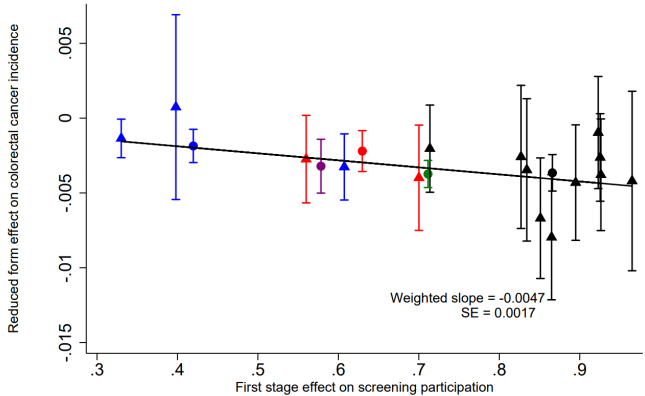
- In the new pill example?
  - Does the treated group = complier population?
  - How does LATE relate to the ATT?
- In Angrist (1990) Vietnam draft study?
- In Angrist and Evans (1998) with same-sex IV?

Is LATE more useful than ITT?

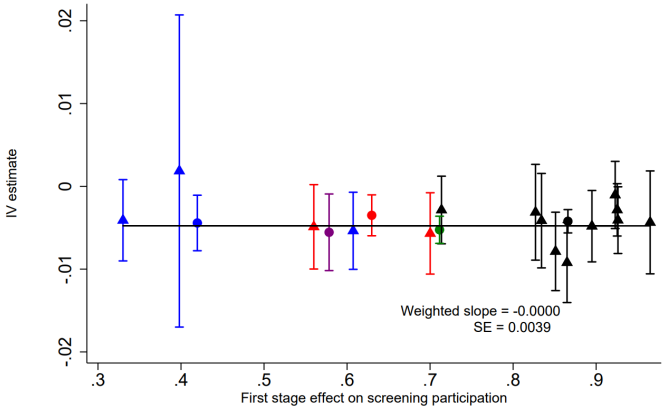
- Reduced-form (ITT) is always causal if independence holds
  - But to get LATE, we further need exclusion and monotonicity
- Independence holds with randomized instruments, e.g. in encouragement designs
  - Why ever do IV?
- IV answers more interesting questions than ITT
  - Wage effects of serving in the army vs. of low lottery numbers
- And hopefully it's more externally valid...

Angrist and Hull (2023)

- Does screening for prostate & colorectal cancer improve health outcomes?
  - A dozen of RCTs where screening is offered for free find very different ITTs



IV estimates of the effect of screening



External validity: Is LATE useful?

- Criticism: *"The LATE may, or may not, be a parameter of interest... and in general, there is no reason to suppose that it will be"* (Deaton 2009)
  - E.g. better to focus on ATE: report Manski bounds or use a structural model for point identification
- Imbens (2010) "Better LATE than nothing":
  - You should first report what the data directly gives, i.e. LATE. Then do whatever
  - E.g. *"I would prefer to keep [structural] assumptions separate and report both the LATE, with its high degree of internal but possibly limited external validity, and possibly add a set of estimates for the overall average effect with the corresponding additional assumptions"*

External validity: Is LATE useful? (2)

- Angrist and Pischke (2010):
  - *"A good structural model might tell us something about economic mechanisms as well as causal effects. But if the information about mechanisms is to be worth anything, the structural estimates should line up with those derived under weaker assumptions"*
- LATE sometimes has economic and policy relevance (e.g., Imbens (2010), Kline and Walters (2016))
- And we can learn a lot about the complier subpopulation...

Outline

1 Basic LATE result

2 Characterizing compliers

3 Extensions

The three groups

- Assume monotonicity (no defiers). Which groups occupy each cell?

	$D_i = 0$ (didn't serve)	$D_i = 1$ (served)
$Z_i = 0$ (high lottery #)	NT + C	AT
$Z_i = 1$ (low lottery #)	NT	AT + C

- Counting the groups:
  - $n \equiv Pr(NT_i) = Pr(D_i = 0 \mid Z_i = 1)$
  - $a = Pr(AT_i) = Pr(D_i = 1 \mid Z_i = 0)$
  - $\pi = Pr(C_i) = Pr(D_i = 1 \mid Z_i = 1) - Pr(D_i = 1 \mid Z_i = 0) =$ first stage
- E.g. in Angrist (1990), for white men born in 1950, the first-stage $= 0.16$
  - By Bayes rule, compliers are 32% of veterans and 10% of non-veterans

Characterizing compliers

- Are compliers unusual...
  - on potential outcomes: $Y_i(1)$ (relative to AT) and $Y_i(0)$ (relative to NT)?
  - on predetermined characteristics X_i (relative to NT and AT)?
- Start with $\mathbb{E}[Y_i(1) \mid AT_i]$ and $\mathbb{E}[Y_i(1) \mid C_i]$:

$$\mathbb{E}[Y_i \mid Z_i = 0, D_i = 1] = \mathbb{E}[Y_i(1) \mid AT_i],$$

$$\mathbb{E}[Y_i \mid Z_i = 1, D_i = 1] = \mathbb{E}[Y_i(1) \mid AT_i] \frac{a}{a+\pi} + \mathbb{E}[Y_i(1) \mid C_i] \frac{\pi}{a+\pi} \implies$$

$$\mathbb{E}[Y_i(1) \mid C_i] = \frac{\mathbb{E}[Y_i \mid Z_i = 1, D_i = 1] \cdot (a+\pi) - \mathbb{E}[Y_i \mid Z_i = 0, D_i = 1] \cdot a}{\pi}$$

$$= \frac{\mathbb{E}[Y_i D_i \mid Z_i = 1] - \mathbb{E}[Y_i D_i \mid Z_i = 0]}{\mathbb{E}[D_i \mid Z_i = 1] - \mathbb{E}[D_i \mid Z_i = 0]}$$

- Implement by IV: $Y_i D_i$ on D_i, IVed with Z_i (think about potential outcomes of $Y_i D_i$)

Characterizing compliers (2)

- Similarly, for $\mathbb{E}[Y_i(0) \mid C_i]$: run $Y_i \cdot (1 - D_i)$ on $1 - D_i$, instrumented with Z_i
- Can get entire distributions $Y_i(0)$ and $Y_i(1)$ for compliers (Abadie 2002)
  - See Kitagawa (2015) for a test of IV validity based on this
- For predetermined characteristics X_i, have multiple options:
  - $X_i D_i$ on D_i, instrumented with Z_i
  - $X_i(1 - D_i)$ on $1 - D_i$, instrumented with Z_i
  - A combination of the two (see Angrist, Hull, Walters 2023)

Characterizing compliers: Example

- Angrist, Pathak, Walters (2013) and Angrist, Hull, Walters (2023): D_i = studying in a charter school, Z_i = getting an offer in a lottery, Y_i = test scores

Table 3: Characteristics of Lottery Compliers at Massachusetts Urban Charter Schools

	Compliers				
	Untreated (1)	Treated (2)	Pooled (3)	Always-takers (4)	Never-takers (5)
Female	0.506	0.510	0.508	0.539	0.463
	(0.023)	(0.021)	(0.016)	(0.024)	(0.017)
Black	0.401	0.380	0.390	0.623	0.490
	(0.022)	(0.021)	(0.016)	(0.023)	(0.017)
Hispanic	0.250	0.300	0.275	0.183	0.228
	(0.02)	(0.018)	(0.013)	(0.019)	(0.014)
Asian	0.022	0.024	0.023	0.004	0.024
	(0.007)	(0.005)	(0.004)	(0.003)	(0.005)
White	0.229	0.216	0.223	0.154	0.215
	(0.018)	(0.016)	(0.012)	(0.016)	(0.014)
Special education	0.190	0.181	0.186	0.158	0.177
	(0.018)	(0.016)	(0.012)	(0.018)	(0.013)
English language learner	0.143	0.148	0.145	0.054	0.088
	(0.015)	(0.013)	(0.010)	(0.011)	(0.010)
Subsidized lunch	0.689	0.705	0.697	0.698	0.666
	(0.021)	(0.019)	(0.014)	(0.022)	(0.016)
Baseline math score	-0.274	-0.312	-0.293	-0.394	-0.301
	(0.047)	(0.041)	(0.032)	(0.045)	(0.036)
Baseline English score	-0.352	-0.349	-0.350	-0.362	-0.299
	(0.050)	(0.043)	(0.033)	(0.046)	(0.038)
Share of sample			0.546	0.197	0.257

Outline

1 Basic LATE result

2 Characterizing compliers

3 Extensions

Extensions

- Non-binary treatments and instruments with linear heterogeneity
- Single multi-valued or continuous instrument
- Single multi-valued or continuous treatment
- Multiple instruments
- Multiple treatments
- Including necessary covariates
- Non-causal first stage

Linear heterogeneity

Consider arbitrary D_i and Z_i but assume linear heterogeneity of causal effects and first-stage coefs:

- First-stage: $D_i(z) = D_i(0) + \pi_i z$ (without loss if Z_i is binary)

  $\Rightarrow \quad D_i = \pi_i Z_i + u_i$ for $u_i = D_i(0)$

- Structural equation: $Y_i(d) = Y_i(0) + \tau_i d$ (without loss if D_i in binary)

  $\Rightarrow \quad Y_i = \tau_i D_i + \varepsilon_i$ for $\varepsilon_i = Y_i(0)$

- Reduced-form: $Y_i = \tau_i \pi_i Z_i + e_i$ for $e_i = \tau_i u_i + \varepsilon_i$
- By independence, $Z_i \perp\!\!\!\perp (\tau_i, \varepsilon_i, \pi_i, u_i)$
- *Claim*: $\tau_{IV} = \frac{\mathbb{E}[\tau_i \pi_i]}{\mathbb{E}[\pi_i]}$, i.e. a π_i-weighted average of τ_i if monotonicity holds (π_i have the same sign for all i)

Proof

Let $\mathbb{E}[Z_i] = \mu$. Consider the numerator of $\tau_{IV} = \mathrm{Cov}[Y_i, Z_i] / \mathrm{Cov}[D_i, Z_i]$:

$$\begin{aligned}\mathrm{Cov}[Y_i, Z_i] &= \mathbb{E}[Y_i (Z_i - \mu)] \\ &= \mathbb{E}[(\tau_i \pi_i Z_i + e_i)(Z_i - \mu)] \\ &= \mathbb{E}[\tau_i \pi_i] \mathrm{Var}[Z_i].\end{aligned}$$

Similarly, $\mathrm{Cov}[D_i, Z_i] = \mathbb{E}[\pi_i] \mathrm{Var}[Z_i]$ and thus

$$\tau_{IV} = \frac{\mathbb{E}[\tau_i \pi_i]}{\mathbb{E}[\pi_i]}.$$

Single multi-valued instrument

- Suppose D_i is binary but Z_i takes values $z_0 \leq \cdots \leq z_P$
- Monotonicity: $D_i(z)$ never switches from 1 to 0 as $z \uparrow$
- Comparing groups $Z_i = z_p$ and $Z_i = z_{p-1}$ yields

$$\tau_p = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid D_i(z_p) > D_i(z_{p-1})\right]$$

  - Non-overlapping complier groups for each p

- Linear IV estimates a weighted average of τ_p:

$$\frac{\mathrm{Cov}\left[Y_i, Z_i\right]}{\mathrm{Cov}\left[D_i, Z_i\right]} = \sum_{p=1}^{P} \omega_p \tau_p, \quad \omega_p = \frac{Pr(D_i(z_p) > D_i(z_{p-1})) \cdot \mathrm{Cov}\left[\mathbf{1}\left[Z_i \geq z_p\right], Z_i\right]}{\sum_{\ell=1}^{P} Pr(D_i(z_\ell) > D_i(z_{\ell-1})) \cdot \mathrm{Cov}\left[\mathbf{1}\left[Z_i \geq z_\ell\right], Z_i\right]}$$

  - Weights can be computed; larger for big complier groups, around the median of Z_i
  - *Exercise*: prove using result from part A1 of the course

Single continuous instrument

- Result on linear IV weights extends naturally to continuous Z_i
- Can compare $Z_i = z + \varepsilon$ to $Z_i = z - \varepsilon$ to get **marginal treatment effects** (MTE)
  - If variation in Z_i is super rich, such that $Pr(D_i = 1 \mid Z_i)$ takes all values $\in [0, 1]$, weighted averages of MTEs identify the ATE, ATT, and other estimands
  - See e.g. Heckman and Vytlacil (2005)

Single multi-valued treatment (Angrist and Imbens 1995)

- Conversely, suppose Z_i is binary but D_i takes values $d_0 \leq \cdots \leq d_Q$
  - E.g. Angrist and Krueger (1991): $D_i =$ years of schooling, $Z_i =$ born in 1st quarter
- Monotonicity: $D_i(1) \geq D_i(0)$ for all i
- Then IV (Wald) estimand is the "average causal response"

$$\frac{\mathrm{Cov}\left[Y_i, Z_i\right]}{\mathrm{Cov}\left[D_i, Z_i\right]} = \sum_{q=1}^{Q} \omega_q \mathbb{E}\left[\frac{Y_i(d_q) - Y_i(d_{q-1})}{d_q - d_{q-1}} \mid D_i(1) \geq d_q > D_i(0)\right]$$

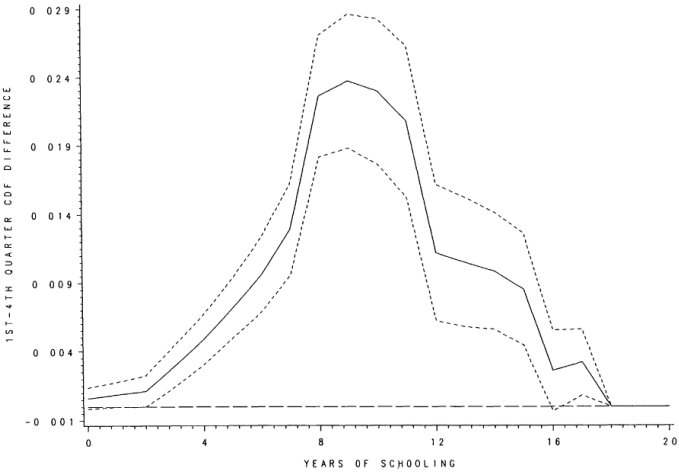
with weights

$$\omega_q = \frac{Pr\left(D_i(1) \geq d_q > D_i(0)\right)}{\sum_{\ell=1}^{Q} Pr\left(D_i(1) \geq d_\ell > D_i(0)\right)}.$$

Single multi-valued treatment (2)

- Weights ω_q can be computed from $Pr(D_i \geq d_q \mid Z_i = 1) - Pr(D_i \geq d_q \mid Z_i = 0)$
  - I.e., how Z_i shifts in the CDF of D_i
  - Larger for treatment values where the IV generates higher response
- Complier groups $D_i(1) \geq d_q > D_i(0)$ can overlap across different q
- Result generalizes immediately to continuous D_i

Weights in Angrist and Krueger (1991)



(From Angrist and Imbens 1995, Figure 3)

Multiple instruments

- E.g. Angrist and Evans (1998): $Z_1 =$ two boys, $Z_2 =$ two girls
- 2SLS identifies a weighted average of individual IVs ρ_k/π_k
  - Weights $\propto \mathrm{Cov}\left[\pi_k Z_{ki}, \hat{D}_i\right]$ for first-stage fitted values $\hat{D}_i = \sum_k \pi_k Z_{ki}$
  - Lee (2018): conventional SE of 2SLS are wrong (moments $\mathbb{E}\left[\left(Y_i - \tau_{2SLS} D_i\right) Z_i\right] = 0$ do not hold with heterogeneous effects!)
- Joint monotonicity is trickier to define:
  - Imbens and Angrist (1994): same ranking of $D_i(\vec{z})$ across $\vec{z} = (z_1, \ldots, z_K)$ for all i
    - ⋆ Very restrictive: essentially $\vec{Z}$ is one multi-valued instrument
  - Mogstad, Torgovitsky, Walters (2021) "partial monotonicity": holding z_{-k} fixed, same ranking of $D_i(z_k; z_{-k})$ across z_k for all i
    - ⋆ Binary $D, Z_1, Z_2 \Longrightarrow$ 6 groups: e.g. "eager compliers": $D_i = \mathbf{1}\left[Z_{1i} = 1 \text{ or } Z_{2i} = 1\right]$
    - ⋆ 2SLS = weighted average of LATEs but some weights can be negative

Multiple treatments

- Literature has focused on exclusive treatments
- Kline and Walters (2016): D_{1i} = early childhood education program (Head Start), D_{2i} = alternative public programs, $D_{0i} = 1 - D_{1i} - D_{2i}$
  - Only one IV: Z_{1i} = randomized offer for the focal program
  - Monotonicity: no defiers & Z_{1i} doesn't affect choice between D_2 and D_0
  - Instrumenting D_{1i} with Z_{1i} identifies a weighted average of $Y_i(1) - Y_i(0)$ and $Y_i(1) - Y_i(2)$ among compliers at these margins
  - Weights are identified: how many kids switch to D_{1i} from each alternative
  - But "subLATEs" are not identified

Multiple treatments (2)

- Mountjoy (2022): D_{1i} = enrolling in a community college, D_{2i} = enrolling in a 4-year college
  - Z_{1i}, Z_{2i} = distance from nearest community college and 4-year college
  - IV with two endogenous vars doesn't produce causal estimands with heterogeneous effects (Kirkeboen, Leuven, Mogstad 2016)
  - Develops a nonparametric IV approach to separate LATEs at each margin
  - See Bhuller and Sigstad (2023) for a different approach

Including necessary covariates

- Suppose Z_i is randomly assigned only conditionally on X_i
- If Z_i is scalar, $\mathbb{E}[Z_i \mid X_i]$ is linear ("rich controls," e.g. X_i is saturated), and monotonicity holds unconditionally:
  - IV that controls for X_i identifies a convex average of conditional-on-X_i LATEs (see MHE Ch. 4.5.1)
- Otherwise, the estimand need not be causal
  - Misspecified $\mathbb{E}[Z_i \mid X_i]$: Blandhol, Bonney, Mogstad, Torgovitsky (2022)
  - Multiple IVs: Goldsmith-Pinkham, Hull, Kolesar (forthcoming) — "contamination bias"

Non-causal first stage

- We introduced monotonicity in a casual way but it needn't be (Small et al. 2017)
- Independence: $\mathbb{E}[Z_i \mid Y_i(0), Y_i(1), X_i] = \lambda' X_i$ (no $D_i(z)$!)
- Stochastic monotonicity: $Pr(D_i = 1 \mid Z_i = z, Y_i(0), Y_i(1), X_i)$ weakly increases in z
- Borusyak and Hull (2024, Proposition 1 simplified to binary D_i):

$$\tau_{IV} = \frac{\mathbb{E}[\phi_i \cdot (Y_i(1) - Y_i(0))]}{\mathbb{E}[\phi_i]}, \quad \text{for } \phi_i = \mathrm{Cov}[Z_i, D_i \mid Y_i(0), Y_i(1), X_i] \geq 0$$

  - "Negative weights are no concern in design-based specifications"
  - But compliers should not be interpreted as those whose treatment causally responds to the IV