

# Part C: Instrumental Variables

## C2: IV with Heterogeneous Effects

Kirill Borusyak

ARE 213 Applied Econometrics

UC Berkeley, Fall 2025

## Questions when effects are heterogeneous

- In Angrist (1990), does the draft lottery IV identify *a* causal parameter?
- Is it the ATE or the affect for some subpopulation?
- Is this subpopulation interesting and policy relevant?
- How can we describe it theoretically?
- What can we say about it in the data?

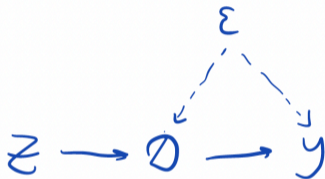
# C2 Outline

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

Readings: MHE Chapters 4.4, 4.5; IW Lecture 5

## 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  $\implies$  (up to) 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:  $\pi = \mathbb{E}[D_i(1) - D_i(0)] = Pr(\text{Complier}_i) - Pr(\text{Defier}_i)$

# Understanding the groups: True/False

- “A defier is someone who gets the encouragement ( $Z = 1$ ) but doesn't take the treatment ( $D = 0$ )”
- “Someone who took the treatment but had the lowest possible outcome is a defier”

## Reduced-form

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

$$\begin{aligned}\rho &= \mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0] \\ &= Pr(NT_i) \cdot \mathbb{E}[Y_i(0) - Y_i(0) | NT_i] \\ &\quad + Pr(Complier_i) \cdot \mathbb{E}[Y_i(1) - Y_i(0) | Complier_i] \\ &\quad + Pr(Defier_i) \cdot \mathbb{E}[Y_i(0) - Y_i(1) | Defier_i] \\ &\quad + Pr(AT_i) \cdot \mathbb{E}[Y_i(1) - Y_i(1) | AT_i] \\ &= Pr(Complier_i) \mathbb{E}[\tau_i | Complier_i] - Pr(Defier_i) \mathbb{E}[\tau_i | Defier_i]\end{aligned}$$

Ratio

$$\tau_{IV} = \frac{Pr(Complier_i) \mathbb{E}[\tau_i | Complier_i] - Pr(Defier_i) \mathbb{E}[\tau_i | Defier_i]}{Pr(Complier_i) - Pr(Defier_i)}$$

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** (Imbens and Angrist (1994)): Under exclusion, independence, and monotonicity, IV identifies the (internally valid) average causal effect among compliers:

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

# Is monotonicity plausible?

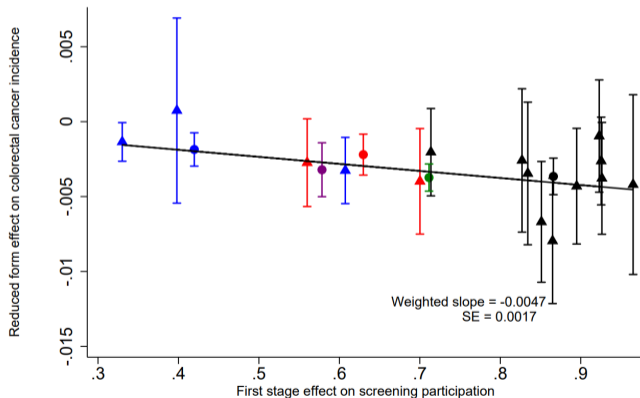
- In the new pill example?
  - ▶ Does the treated group = complier population?
  - ▶ How does LATE relate to the ATT?
- In Angrist (1990) Vietnam draft study?
- In Angrist and Evans (1998) with same-sex IV?

# Is LATE more useful than ITT?

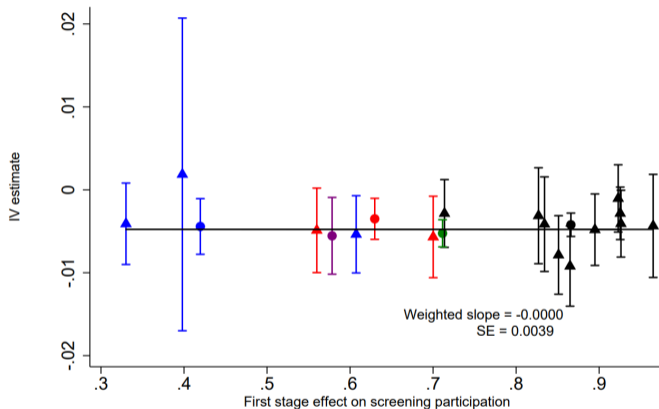
- Reduced-form (ITT) is always causal if independence holds
  - ▶ Always holds in RCTs, e.g. in encouragement designs
  - ▶ While LATE further requires exclusion and monotonicity
  - ▶ So why ever do IV?
- IV answers more interesting questions than ITT
  - ▶ Wage effects of serving in the army vs. of low lottery numbers
  - ▶ More interesting in part because hopefully more externally valid with respect to alternative implementations...

## Example: 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



*Open question:* but is this IV estimate policy relevant?

## 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 (2010))
  - ▶ E.g. better to focus on ATE: report partially identified set (“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

## Counting 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:
  - $n \equiv Pr(NT_i) = Pr(D_i = 0 \mid Z_i = 1)$
  - $a = Pr(AT_i) = Pr(D_i = 1 \mid Z_i = 0)$
  - $\pi = Pr(C_i) = Pr(D_i = 1 \mid Z_i = 1) - Pr(D_i = 1 \mid Z_i = 0) = \text{first stage}$
- 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

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

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

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

## Characterizing compliers (2)

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

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

Given  $a + \pi = Pr(D_i = 1 | Z_i = 1)$  and  $a = Pr(D_i = 1 | Z_i = 0)$ ,

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

Implement by IV:  $Y_i D_i$  on  $D_i$ , IVed with  $Z_i$

- *Intuition:* think potential outcomes of  $Y_i D_i$  and the causal effect on this outcome

## Characterizing compliers (3)

- 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  $X_i$  (that are  $\perp\!\!\!\perp Z$ ), 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, and Walters (2023))
  - ▶ *Note:* this trick works with all design-based OLS and IV specifications, even with non-binary  $D$  and  $Z$  (Hull, 2025)

# Characterizing compliers: Example

- Angrist, Pathak, and Walters (2013) and Angrist, Hull, and 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			Always-takers (4)	Never-takers (5)
	Untreated (1)	Treated (2)	Pooled (3)		
Female	0.506 (0.023)	0.510 (0.021)	0.508 (0.016)	0.539 (0.024)	0.463 (0.017)
Black	0.401 (0.022)	0.380 (0.021)	0.390 (0.016)	0.623 (0.023)	0.490 (0.017)
Hispanic	0.250 (0.02)	0.300 (0.018)	0.275 (0.013)	0.183 (0.019)	0.228 (0.014)
Asian	0.022 (0.007)	0.024 (0.005)	0.023 (0.004)	0.004 (0.003)	0.024 (0.005)
White	0.229 (0.018)	0.216 (0.016)	0.223 (0.012)	0.154 (0.016)	0.215 (0.014)
Special education	0.190 (0.018)	0.181 (0.016)	0.186 (0.012)	0.158 (0.018)	0.177 (0.013)
English language learner	0.143 (0.015)	0.148 (0.013)	0.145 (0.010)	0.054 (0.011)	0.088 (0.010)
Subsidized lunch	0.689 (0.021)	0.705 (0.019)	0.697 (0.014)	0.698 (0.022)	0.666 (0.016)
Baseline math score	-0.274 (0.047)	-0.312 (0.041)	-0.293 (0.032)	-0.394 (0.045)	-0.301 (0.036)
Baseline English score	-0.352 (0.050)	-0.349 (0.043)	-0.350 (0.033)	-0.362 (0.046)	-0.299 (0.038)
Share of sample			0.546	0.197	0.257

## Extrapolation via reweighting

- If compliers have different  $X_i$  than overall population or the treated group, we can reweight the IV estimator to make them look similar
  - ▶ If effect heterogeneity is fully explained by  $X_i$ , i.e. homogeneous effects given  $X_i$ , this yields ATE and ATT
- Aronow and Carnegie (2013): compliers have a higher female share  $\Leftrightarrow$  females have a higher complier share
  - ▶ Estimate  $Pr(\text{Complier}_i \mid X_i)$  via first stage heterogeneity
  - ▶ Use  $1/\hat{Pr}(\text{Complier}_i \mid X_i)$  as weights in IV
- Angrist and Fernandez-Val (2010): if two IVs give different answers (e.g., twin births vs. same sex), is it because compliers have different covariates?
  - ▶ Develop an overidentification-type test

# 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)  
 $\Rightarrow D_i = \pi_i Z_i + u_i$  for  $u_i = D_i(0)$
- Structural equation:  $Y_i(d) = Y_i(0) + \tau_i d$  (without loss if  $D_i$  in binary)  
 $\Rightarrow Y_i = \tau_i D_i + \varepsilon_i$  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  $\pi_i$ -weighted average of  $\tau_i$  if monotonicity holds ( $\pi_i$  have the same sign for all  $i$ )

# Proof

We know first-stage identifies  $\mathbb{E} [\pi_i]$ ; reduced-form identifies  $\mathbb{E} [\tau_i \pi_i]$  since

$$Y_i = \tau_i D_i + \varepsilon_i = \tau_i (\pi_i Z_i + u_i) + \varepsilon_i = \tau_i \pi_i Z_i + e_i$$

In more detail, let  $\mathbb{E} [Z_i] = \mu$ . Consider the numerator of  $\tau_{IV} = \text{Cov} [Y_i, Z_i] / \text{Cov} [D_i, Z_i]$ :

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

Similarly,  $\text{Cov} [D_i, Z_i] = \mathbb{E} [\pi_i] \text{Var} [Z_i]$  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  $z_0 \leq \dots \leq z_p$ 
  - ▶ E.g.  $D_i$  = attend college,  $Z_i$  = number of colleges nearby
- Monotonicity:  $D_i(z)$  never switches from 1 to 0 as  $z \uparrow$
- **Option 1:** select the sample of  $Z_i \in \{z_{p-1}, z_p\}$  and compare groups  $Z_i = z_p$  vs.  $Z_i = z_{p-1}$ 
  - ▶ Yields  $\tau_p = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i(z_p) > D_i(z_{p-1})]$
  - ▶ Non-overlapping complier groups for each  $p$

## Single multi-valued instrument (2)

- **Option 2:** Use linear IV with a single instrument  $Z_i$

- ▶ Yields a weighted average of  $\tau_p$ :

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

- ▶ 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  $Z_i$ 
  - ▶ E.g.  $Z_i$  = distance to nearest college (Card, 1993)
- Can compare  $Z_i = z + \epsilon$  to  $Z_i = z - \epsilon$  (for small  $\epsilon$ ) 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
  - ▶ *Intuition*: there are no always-takers or never-takers
  - ▶ See e.g. Heckman and Vytlacil (2005) and Mogstad and Torgovitsky (2024)

# Single multi-valued treatment (Angrist and Imbens (1995))

- Conversely, suppose  $Z_i$  is binary but  $D_i$  takes values  $d_0 \leq \dots \leq d_Q$ 
  - ▶ 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  $i$  (compliers:  $D_i(1) > D_i(0)$ )
- We can measure how  $Z_i$  shifts the CDF of  $D_i$ :

$$Pr(D_i \geq d_q \mid Z_i = 1) - Pr(D_i \geq d_q \mid Z_i = 0) = Pr(D_i(1) \geq d_q > D_i(0))$$

- ▶ Larger for values  $d_q$  where the IV pushes more people from below to above
- ▶ Complier groups  $D_i(1) \geq d_q > D_i(0)$  can overlap across different  $q$

## Single multi-valued treatment (2)

- IV (Wald) estimand:

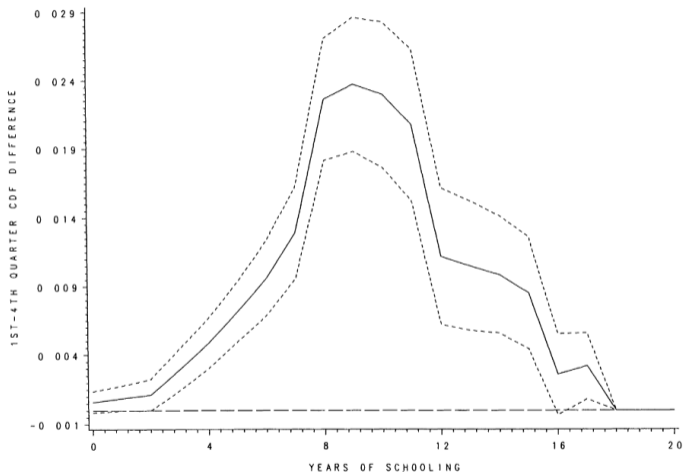
$$\frac{\text{Cov}[Y_i, Z_i]}{\text{Cov}[D_i, Z_i]} = \sum_{q=1}^Q \omega_q \mathbb{E} \left[ \frac{Y_i(d_q) - Y_i(d_{q-1})}{d_q - d_{q-1}} \mid D_i(1) \geq d_q > D_i(0) \right]$$

with weights that can be computed:

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

- Result generalizes immediately to continuous  $D_i$
- We cannot isolate individual margins by conditioning on  $D_i \in \{d_{q-1}, d_q\}$

# Weights in Angrist and Krueger (1991)



(From Angrist and Imbens (1995), Figure 3)

# Multiple instruments

- E.g. Angrist and Evans (1998):  $Z_1 = \text{two boys}$ ,  $Z_2 = \text{two girls}$
- 2SLS identifies a weighted average of individual IVs  $\rho_k/\pi_k$ 
  - ▶ Weight on  $Z_k = \text{slope of regression of } \pi_k Z_{ki} \text{ on the first-stage fitted value } \hat{D}_i = \sum_k \pi_k Z_{ki}$ ; typically convex but not guaranteed
  - ▶ Lee (2018): conventional SE of 2SLS are wrong!
    - ★ Moments  $\mathbb{E}[(Y_i - \tau_{2SLS} D_i) Z_i] = 0$  do not hold with heterogeneous effects!
    - ★ But the correction tends to be small

## Multiple instruments (2)

Joint monotonicity is trickier to define:

- Imbens and Angrist (1994): same ranking of  $D_i(\vec{z})$  across  $\vec{z} = (z_1, \dots, z_K)$  for all  $i$ 
  - ▶ Very restrictive: essentially  $\vec{z}$  is one multi-valued instrument
  - ▶ Can't have one family respond to two boys only and another to two girls only
- Mogstad, Torgovitsky, and Walters (2021) “partial monotonicity”: holding  $z_{-k}$  fixed, same ranking of  $D_i(z_k; z_{-k})$  across  $z_k$  for all  $i$ 
  - ▶ Binary  $D, Z_1, Z_2 \implies$  6 groups: e.g. “eager compliers”:  $D_i = \mathbf{1} [Z_{1i} = 1 \text{ or } Z_{2i} = 1]$
  - ▶ 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
    - ★  $\propto$  causal effects of  $D_{1i}$  on  $D_{i0}$  and  $D_{i2}$  (estimated by IV)
  - ▶ 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 does not produce causal estimands with heterogeneous effects (Kirkeboen et al. (2016))
  - ▶ Mountjoy develops a nonparametric IV approach to separate LATEs at each margin
  - ▶ See Bhuller and Sigstad (2024) for a different approach

## Including necessary covariates

- Suppose  $Z_i$  is randomly assigned only conditionally on  $X_i$
- If  $Z_i$  is scalar, pscore  $\mathbb{E}[Z_i | X_i]$  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  $\mathbb{E}[Z_i | X_i]$ : Blandhol, Bonney, Mogstad, and Torgovitsky (2025)
  - ▶ Multiple IVs (or OLS with multiple treatments): Goldsmith-Pinkham, Hull, and Kolesár (2024) “contamination bias”

## Non-causal first stage

- We introduced monotonicity in a casual way but it needn't be (Small et al. (2017))
- Independence + linear pscore:  $\mathbb{E}[Z_i \mid Y_i(0), Y_i(1), X_i] = \lambda'X_i$  (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) simplified to binary  $D_i$ : for IV with  $X_i$  controls,

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

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

# References I

- ABADIE, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292.
- ANGRIST, J. AND I. FERNANDEZ-VAL (2010): "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework," *NBER Working Paper*.
- ANGRIST, J. AND P. HULL (2023): "IV Methods Reconcile Intention-to-Screen Effects Across Pragmatic Cancer Screening Trials," *Proceedings of the National Academy of Sciences*, 120.
- ANGRIST, J., P. HULL, AND C. R. WALTERS (2023): "Methods for Measuring School Effectiveness," in *Handbook of the Economics of Education*.
- ANGRIST, J. D. (1990): "Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records," *American Economic Review*, 80, 313–336.
- ANGRIST, J. D. AND W. N. EVANS (1998): "Children and Their Parents ' Labor Supply : Evidence from Exogenous Variation in Family," *American Economic Association*, 88, 450–477.
- ANGRIST, J. D. AND G. W. IMBENS (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442.

## References II

- ANGRIST, J. D. AND A. B. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*.
- ANGRIST, J. D., P. A. PATHAK, AND C. R. WALTERS (2013): "Explaining charter school effectiveness," *American Economic Journal: Applied Economics*, 5, 1–27.
- ANGRIST, J. D. AND J.-S. PISCHKE (2010): "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3–30.
- ARONOW, P. M. AND A. CARNEGIE (2013): "Beyond LATE: Estimation of the Average Treatment Effect with an Instrumental Variable," *Political Analysis*, 21, 492–506.
- BHULLER, M. AND H. SIGSTAD (2024): "2SLS with Multiple Treatments," *Journal of Econometrics*, 242, 105785.
- BLANDHOL, C., J. BONNEY, M. MOGSTAD, AND A. TORGOVITSKY (2025): "When is TSLS Actually LATE?" *Working Paper*.
- BORUSYAK, K. AND P. HULL (2024): "Negative weights are no concern in design-based specifications," *AEA Papers and Proceedings*, 114, 597–600.

# References III

- CARD, D. (1993): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," *NBER Working Paper*.
- DEATON, A. (2010): "Instruments, Randomization, and Learning about Development," *Journal of Economic Literature*, 48, 424–455, iSBN: 0022-0515.
- GOLDSMITH-PINKHAM, P., P. HULL, AND M. KOLESÁR (2024): "Contamination Bias in Linear Regressions," *American Economic Review*, 114, 4015–4051.
- HECKMAN, J. J. AND E. VYTLACIL (2005): "Structural equations, treatment effects, and econometric policy evaluation," *Econometrica*, 73, 669–738, iSBN: 4030008526.
- HULL, P. (2025): "One Weird Trick to Characterize Effective Populations in Design-Based Specifications," *Unpublished*.
- IMBENS, G. W. (2010): "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 48, 399–423.
- IMBENS, G. W. AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475, iSBN: 00129682.

## References IV

- KIRKEBOEN, L. J., E. LEUVEN, AND M. MOGSTAD (2016): “Field of Study, Earnings, and Self-Selection,” *The Quarterly Journal of Economics*, 131, 1057–1111.
- KITAGAWA, T. (2015): “A Test for Instrument Validity,” *Econometrica*, 83, 2043–2063.
- KLINE, P. AND C. R. WALTERS (2016): “Evaluating Public Programs with Close Substitutes: The Case of HeadStart\*,” *The Quarterly Journal of Economics*, 131, 1795–1848.
- LEE, S. (2018): “A Consistent Variance Estimator for 2SLS When Instruments Identify Different LATEs,” *Journal of Business & Economic Statistics*, 36, 400–410.
- MOGSTAD, M. AND A. TORGOVITSKY (2024): “Instrumental variables with unobserved heterogeneity in treatment effects,” in *Handbook of Labor Economics*, Elsevier, vol. 5, 1–114.
- MOGSTAD, M., A. TORGOVITSKY, AND C. R. WALTERS (2021): “The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables,” *American Economic Review*, 111, 3663–3698.
- MOUNTJOY, J. (2022): “Community Colleges and Upward Mobility,” *American Economic Review*, 112, 2580–2630.
- SMALL, D. S., Z. TAN, R. R. RAMSAHAI, S. A. LORCH, AND M. A. BROOKHART (2017): “Instrumental Variable Estimation with a Stochastic Monotonicity Assumption,” *Statistical Science*, 32.