Chapter 7: IV Identification

Peter Hull

Applied Econometrics II
Brown University
Spring 2024

Outline

1. The LATE Theorem
2. Generalizations
3. Characterizing Compliers
4. Marginal Treatment Effects
5. Application: Agan et al. (2023)

Whose Treatment Effect is it Anyway?

We motivated IV with a constant-effects model, $Y_i = \beta X_i + \varepsilon_i$, but in practice of course the effects β_i may vary across individuals (etc.)

- Recall: with selection-on-observables, which estimator you choose (e.g. regression vs. IPW) can affect the weighting of het effects

It turns out IV often identifies a convex average of heterogeneous effects under *first-stage monotonicity*: i.e. if Z_i only affects X_i in one direction

- This was first famously established in the Imbens and Angrist (1994) "LATE theorem," for simple IV setups
- We now understand it to hold fairly generally (though there are a few notable exceptions – including multiple treatments, as before)

The IA setup also helps clarify other key facets of IV identification

- E.g. difference between instrument "independence" and "exclusion"

Basic IA '94 Setup

Binary treatment D_i, binary instrument Z_i, some outcome Y_i

- Usual potential outcomes: $Y_i = Y_i(0)(1 - D_i) + Y_i(1)D_i$
- New potential treatments: $D_i = D_i(0)(1 - Z_i) + D_i(1)Z_i$

Four assumptions:

1. *Exclusion*: $Y_i(d)$ is only indexed by d, not by z (already implicit)
2. *Independence*: $(Y_i(0), Y_i(1), D_i(0), D_i(1)) \perp Z_i$ (we'll relax this)
3. *Relevance*: $E[D_i \mid Z_i = 1] > E[D_i \mid Z_i = 0]$ (sign is without loss)
4. *Monotonicity*: $D_i(1) \geq D_i(0)$ almost-surely (sign, again, without loss)

Monotonicity says there are three types of individuals:

1. "Always-takers," with $D_i(1) = D_i(0) = 1$
2. "Never-takers," with $D_i(1) = D_i(0) = 0$
3. "Compliers," with $1 = D_i(1) > D_i(0) = 0$

The Local Average Treatment Effect Theorem

IA prove the IV (Wald) estimand identifies the average effect on compliers:

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

Proof:

- By independence, $E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0] = E[D_i(1) - D_i(0)]$
- Similarly,

$$\begin{aligned} & E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0] \\ =& E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(1) \mid Z_i = 1] \\ & - E[Y_i(0) + (Y_i(1) - Y_i(0))D_i(0) \mid Z_i = 0] \\ =& E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] \end{aligned}$$

- By monotonicity, $D_i(1) - D_i(0) \in \{0, 1\}$. Thus:

$$\frac{E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))]}{E[D_i(1) - D_i(0)]} = E[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0)]$$

Implications

Potential outcomes notation makes clear independence vs. exclusion

- A lottery'd Z_i can ensure independence, but exclusion can still fail

First-stage monotonicity becomes important under heterogeneous effects

- Otherwise, β^{IV} identifies a non-convex average:

$$\begin{aligned}\frac{E[(Y_i(1)-Y_i(0))(D_i(1)-D_i(0))]}{E[D_i(1)-D_i(0)]} =& \frac{c}{c-d} E[Y_i(1)-Y_i(0) \mid D_i(1) > D_i(0)] \\ &+ \frac{-d}{c-d} \underbrace{E[Y_i(1)-Y_i(0) \mid D_i(1) < D_i(0)]}_{\text{Avg. effect for defiers}}\end{aligned}$$

The LATE result formalizes a key limitation of overid. tests:

- Two valid IVs can identify different LATEs
- E.g. lottery and grandfather IV estimate of UP effects could differ

Oregon Medicaid Experiment Revisited

	Control mean (1)	ITT (2)	LATE (3)	p-values (4)
Panel A: Extensive margin				
All hospital admissions	0.067	0.0054	0.021	[0.004]
	(0.250)	(0.0019)	(0.0074)	
Admissions through ER	0.048	0.0018	0.0070	[0.265]
	(0.214)	(0.0016)	(0.0062)	
Admissions not through ER	0.029	0.0041	0.016	[0.002]
	(0.167)	(0.0013)	(0.0051)	
Panel B: All hospital admissions				
Days	0.498	0.026	0.101	[0.329]
	(3.795)	(0.027)	(0.104)	{0.328}
List charges	2,613	258	1,009	[0.077]
	(19,942)	(146)	(569)	{0.106}
Procedures	0.155	0.018	0.070	[0.031]
	(1.08)	(0.0083)	(0.032)	{0.059}
Standardized treatment effect		0.012	0.047	[0.073]
		(0.0067)	(0.026)	

Recall: first stage ≈ 0.25

Better LATE Than Nothing?

The LATE theorem says what is possible, not necessarily what is desirable

- Very intuitive that Z_i can't tell us anything about the effects of always/never-takers
- The complier sub-population can be seen to be "policy relevant" (e.g. Kline and Walters 2016)
- We don't know exactly who compliers are, but as we'll soon see we can characterize them in certain ways

Important special case: when there are no always-takers LATE=ATT

$$\begin{aligned} E[Y_i(1) - Y_i(0) \mid D_i = 1] =& (1-\omega) E[Y_i(1) - Y_i(0) \mid D_i(1) = D_i(0) = 1] \\ &+ \omega \underbrace{E[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0), Z_i = 1]}_{= E[Y_i(1) - Y_i(0) \mid D_i(1) > D_i(0)], \text{ by independence}} \end{aligned}$$

Similarly, LATE=ATU when there are no never-takers

Outline

1. The LATE Theorem ✓
2. Generalizations
3. Characterizing Compliers
4. Marginal Treatment Effects
5. Application: Agan et al. (2023)

Multivalued Treatments

Angrist & Imbens '95 show that when $D_i \in \{0, \ldots, \overline{d}\}$ and a generalization of the LATE assumptions hold, IV identifies an "Average Causal Response"

$$\frac{E[Y_i|Z_i=1]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]} = \sum_{d=1}^{\overline{d}} \omega_d E[Y_i(d) - Y_i(d-1) \mid D_i(1) \geq d > D_i(0)]$$

where $\omega_d = \frac{Pr(D_i(1) \geq d > D_i(0))}{\sum_{d'=1}^{\overline{d}} Pr(D_i(1) \geq d' > D_i(0))}$ gives a convex weighting scheme

Averages unit causal responses $E[Y_i(d) - Y_i(d-1) \mid D_i(1) \geq d > D_i(0)]$

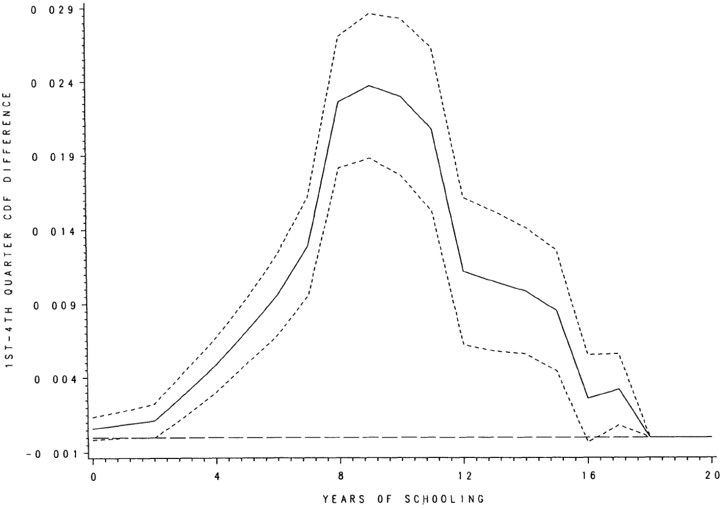
- More weight on margins d with more first-stage "action"
- Note: "Compliers" with $D_i(1) > D_i(0)$ can be double-counted at different margins
- The weights are identified by the difference in treatment CDF when $Z_i = 1$ vs. when $Z_i = 0$

Angrist-Krueger QOB Revisited

	(1) Born in 1st quarter of year	(2) Born in 4th quarter of year	(3) Difference (std. error) (1) − (2)
Panel A: Wald Estimates for 1970 Census—Men Born 1920–1929[a]			
ln (weekly wage)	5.1485	5.1578	−.00935 (.00374)
Education	11.3996	11.5754	−.1758 (.0192)
Wald est. of return to education			.0531 (.0196)
OLS est. of return to education[b]			.0797 (.0005)

Here the treatment counts years of completed schooling

Where the QOB Instrument Has "Bite"



Reassuringly, no FS effects at advanced years of schooling

Continuous Treatments

Angrist, Graddy, and Imbens (2000) extend things further, in the classic IV context of supply and demand

- Basic idea: use stormy weather as a supply shock to estimate demand at the Fulton Fish Market in NYC
- Let $Q_i(p)$ be log quantity demanded in market i at log price p, let P_i be observed prices, and let Z_i be an indicator for stormy weather

Under appropriate generalizations of the LATE assumptions:

$$\frac{E[Q_i \mid Z_i = 1] - E[Q_i \mid Z_i = 0]}{E[P_i \mid Z_i = 1] - E[P_i \mid Z_i = 0]} = \int_p E[q_i'(p) \mid P_i(1) \geq p > P_i(0)]\omega(p)dp$$

where $P_i(z)$ are potential prices and $\omega(p) = \frac{Pr(P_i(1) \geq p' > P_i(0))}{\int_{p'} Pr(P_i(1) \geq p' > P_i(0))dp'}$

Again, the weights are identified by first-stage CDF effects

Continuous Instruments

Borusyak and Hull (2024, AEA P&P) gives the most general version of LATE I know of, including continuous Z_i:

$$\frac{Cov(Z_i, Y_i)}{Cov(Z_i, X_i)} = E\left[\int_x Y_i'(x)\omega_i(x)dx\right]$$

where $\omega_i(x) \geq 0$ under the appropriate first-stage monotonicity condition

- Follows from weighted-average-effect interpretation of the reduced form and first stage, and chain rule: $\frac{d}{dz}Y_i(X_i(z)) = Y_i'(x)X_i'(z)$

Notation becomes complicated, but basic logic still applies:

- More weight given to margins where Z_i has more first-stage "bite"
- Weights are identified by first-stage effects

Necessary Controls

So far we've assumed Z_i is unconditionally as-good-as-randomly assigned. But what if we need to control for some W_i?

Angrist '98 for IV: when W_i gives a set of mutually exclusive strata,

$$Y_i = \alpha + \beta X_i + W_i'\gamma + \varepsilon_i$$
$$X_i = \mu + \pi Z_i + W_i'\phi + \nu_i$$

identifies $\beta = E[\omega(W_i)\beta^{IV}(W_i)]$, where $\omega(w) = \frac{Var(Z_i|W_i=w)\pi(w)}{E[Var(Z_i|W_i)\pi(W_i)]}$ for $\pi(w) = \frac{Cov(Z_i,X_i|W_i=w)}{Var(Z_i|W_i=w)}$ and where $\beta^{IV}(w) = \frac{Cov(Z_i,Y_i|W_i=w)}{Cov(Z_i,X_i|W_i=w)}$

- Thus if $\beta^{IV}(w)$ is causal (e.g. the LATE theorem holds conditionally) and $\pi(w) \geq 0$ (e.g. monotonicity holds unconditionally), β identifies a convex average of heterogeneous treatment effects

Same story for other controls, as long as they're flexible (span $E[Z_i \mid W_i]$)

Multiple Instruments

Recall: 2SLS identifies a weighted average of one-at-a-time IV estimands

- The key question then becomes: when are these weights convex?

Imbens + Angrist '94 show convexity under generalized monotonicity:

- For all $z, z' \in Supp(Z_i)$, either $D_i(z) \geq D_i(z')$ almost-surely or $D_i(z) \leq D_i(z')$ almost-surely
- Implies instruments are "nested" (e.g. immediate + waitlist offer IVs)

Mogstad, Torgovitsky, and Walters (2021) show convexity may (or may not!) fail under a weaker "partial monotonicity" condition

- If $z' \geq z$ component-wise, then $D_i(z') \geq D_i(z)$ almost-surely
- E.g. a "price" and "distance" instrument for college (non-nested)

Contamination bias may further kick in with controls

- Linearly controlling for W_i may lead to negative first-stage weights, even if conditional monotonicity holds (Goldsmith-Pinkham et al. '22)

Model-Based Identification

So far we've focused on "design-based" IV, where Z_i is (perhaps conditionally) as-good-as-randomly assigned

- Can we interpret DiD-IV (e.g. "grandfathering" IV) via "LATEs"?

Recall: IV is just a fancy way to divide two regression coefficients

- If a reduced-form TWFE regression identifies the ATE of Z_i on Y_i and the corresponding TWFE first-stage identifies the ATE of Z_i on D_i, the ratio has a LATE/ACR/etc interpretation under monotonicity
- See Hudson et al. (2017) for a formalization of this for DiD-IV

If regression fails (e.g. b/c of negative weights), we can still divide

- E.g. use Callaway-Sant'Anna to estimate reduced form / first stage ATEs of Z_i, then take the ratio to estimate LATE/ACR/etc
- Alternatively: "Fuzzy DiD" (de Chaisemartin & D'Haultfoeuille)

Multiple Treatments :(

Unfortunately, IV with multiple treatments generally fails to identify convex weighted averages of heterogeneous effects

- Consider a model with two treatments: $Y_i = \beta_{1i} D_{1i} + \beta_{2i} D_{2i} + \varepsilon_i$
- Unless we have an IV which *only* shifts i between two alternatives (in one direction), the RF will generally mix together β_{1i} and β_{2i}
- Under constant effects mixing is no problem (like w/ monotonicity violations + one treatment), but the mixing weights can be negative
- See Bhuller and Sigstad (2023) for a general roadmap

Of course, as with other negative weighting concerns, how much you should worry about this in practice is not really clear *a priori*

- Here effects and weights will typically be correlated under Roy (1951) -style *selection-on-gains*: maybe more of a concern?
- Overid. tests may be helpful (earlier caveats notwithstanding)

Outline

1. The LATE Theorem✓
2. Generalizations✓
3. Characterizing Compliers
4. Marginal Treatment Effects
5. Application: Agan et al. (2023)

Getting Under the IV Hood

It can often be useful to characterize the compliers of a given IV

- E.g. to hint at mechanisms, contextualize findings, or reconcile findings across different (quasi-)experiments

Of course we can't tell if $D_i(1) > D_i(0)$ for any given i. But it turns out we can still learn about compliers *on average*. We'll step through:

1. Average potential outcomes: $E[Y_i(d) \mid D_i(1) > D_i(0)]$ for $d \in \{0, 1\}$
2. Characteristics: $E[W_i \mid D_i(1) > D_i(0)]$ for baseline W_i
3. Fancier things: distributions, nonlinear estimands, etc.

This is all easier than they might seem...

Separating Treated and Untreated Outcomes

Suppose we want to know $E[Y_i(1) \mid D_i(1) > D_i(0)]$ in the basic IA setup

Trick: Consider IV on the modified outcome $\tilde{Y}_i = Y_i D_i$ instead of Y_i

- Potential outcomes: $\tilde{Y}_i(1) = Y_i(1)$, and $\tilde{Y}_i(0) = 0$
- Hence "treatment effects:" $\tilde{Y}_i(1) - \tilde{Y}_i(0) = Y_i(1) - 0 = Y_i(1)$
- Check: exclusion, independence, and monotonicity still hold with $\tilde{Y}_i$
- Hence IV identifies LATE: $E[Y_i(1) \mid D_i(1) > D_i(0)]$

Similar logic shows that IV of $\tilde{Y}_i = Y_i(1 - D_i)$ on $\tilde{D}_i = 1 - D_i$ identifies a LATE of $E[\tilde{Y}_i(1) - \tilde{Y}_i(0) \mid D_i(1) > D_i(0)] = E[Y_i(0) \mid D_i(1) > D_i(0)]$

Such an easy bonus analysis to any *ivreg*!

Illustration: Angrist, Pathak, and Walters (2013)

	Urban				Nonurban			
Subject	Treatment effect (1)	$E_u[Y_0 \mid D=0]$ (2)	λ_0^u (3)	λ_1^u (4)	Treatment effect (5)	$E_n[Y_0 \mid D=0]$ (6)	λ_0^n (7)	λ_1^n (8)
Panel A. Middle school								
Math	0.483***	−0.399***	0.077	0.560***	−0.177**	0.236***	0.010	−0.143***
	(0.074)	(0.011)	(0.049)	(0.054)	(0.074)	(0.007)	(0.061)	(0.042)
N	4,858				2,239			
ELA	0.188***	−0.422***	0.118**	0.306***	−0.148***	0.260***	0.102**	−0.086***
	(0.064)	(0.012)	(0.054)	(0.049)	(0.048)	(0.007)	(0.050)	(0.030)
N	4,551				2,323			

Decomposing

$$LATE = \underbrace{E[Y_i(1) \mid D_i(1) > D_i(0)] - E[Y_i(0) \mid D_i = 0]}_{\lambda_1}$$
$$- \underbrace{(E[Y_i(0) \mid D_i(1) > D_i(0)] - E[Y_i(0) \mid D_i = 0])}_{\lambda_0}$$

shows that charter compliers have typical counterfactual achievement

Complier Summary Statistics

We can use the same trick to estimate $E[W_i \mid D_i(1) > D_i(0)]$:

- IV of $W_i D_i$ on D_i, instrumenting with Z_i
- IV of $W_i(1 - D_i)$ on $(1 - D_i)$, instrumenting with Z_i
- Some weighted average of the two

Testing that these two approaches indeed estimate the same thing can be shown to be equivalent to a balance regression of W_i on Z_i

- "Stacking" the two specifications up, and estimating a single $E[W_i \mid D_i(1) > D_i(0)]$ w/two IVs, automates the weights+overid. test

Fun to compare with $E[W_i \mid D_i(1) = D_i(0) = 1] = E[W_i \mid D_i = 1, Z_i = 0]$ and $E[W_i \mid D_i(1) = D_i(0) = 0] = E[W_i \mid D_i = 0, Z_i = 1]$

Illustration: Angrist, Hull, and Walters (2023)

	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

Fancier Things

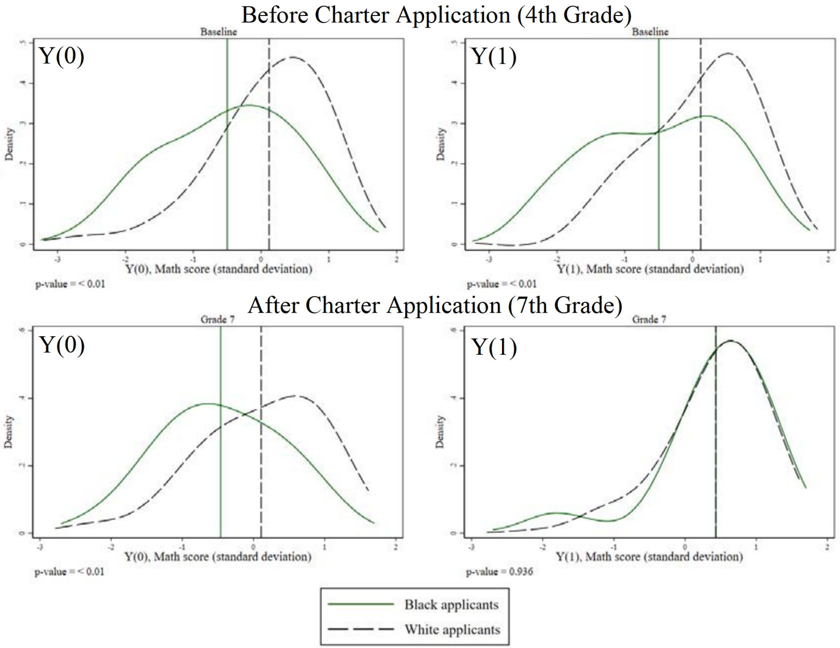
General result: $E[g(W_i, Y_i(d)) \mid D_i(1) > D_i(0)]$ for any $g(\cdot)$ and any $d \in \{0,1\}$ is identified by β in the IV regression:

$$g(W_i, Y_i) \times \mathbf{1}[D_i = d] = \alpha + \beta \mathbf{1}[D_i = d] + \varepsilon_i$$
$$\mathbf{1}[D_i = d] = \mu + \pi Z_i + \nu_i$$

Lots of fun stuff we can do here. E.g.: distributions

- $g(W_i, Y_i) = \mathbf{1}[Y_i \leq y]$ for $y \in Supp(Y_i)$ estimates the CDF of complier potential outcomes, $F(y) = Pr(Y_i(d) < y \mid D_i(1) > D_i(0))$
- $g(W_i, Y_i) = \frac{1}{h} K(\frac{Y_i - y}{h})$ for $y \in Supp(Y_i)$ estimates the corresponding PDF, where $K(\cdot)$ is a kernel function and h is a bandwidth

Illustration: Angrist, Hull, and Walters (2023)



Outside the Basic Setup

The same logic applies to IV regressions with flexible controls

- E.g. *ivreg*ing $Y_i D_i$ on D_i instrumenting by Z_i and controlling for cell FE ID's a variance-weighted avg of within-cell $E[Y_i(1) \mid D_i(1) > D_i(0)]$
- See Abadie (2003) for an alternative weighting estimator for unconditional averages (and other things)

Logic also goes through with continuous instruments

- E.g. can always mechanically decompose an IV on a binary D_i into implied "$Y(1)$" and "$Y(0)$" terms

Unfortunately, the logic gets trickier with non-binary treatments

- Let's chat if you ever want to do this...

Outline

1. The LATE Theorem✓
2. Generalizations✓
3. Characterizing Compliers✓
4. Marginal Treatment Effects
5. Application: Agan et al. (2023)

Doing More With Continuous IVs

If we have a valid Z_i that varies continuously (and a binary D_i), we can potentially learn more about how effects vary with compliance

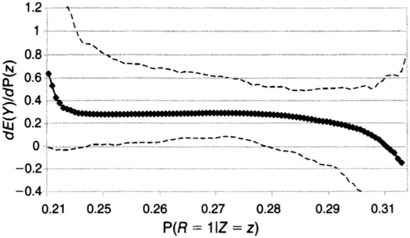
- Individuals with more "resistance" to treatment may respond at different margins of Z_i and have different treatment effects

Heckman & Vytlicil (2005, 2007, 2010, 2013...) write $D_i = \mathbf{1}[p(Z_i) \geq U_i]$

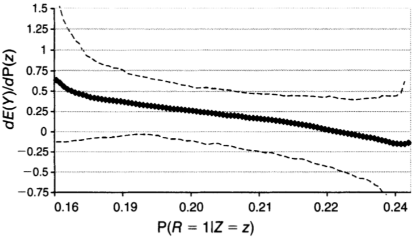
- $p(z) = Pr(D_i \mid Z_i = z)$ is the propensity score with respect to the IV
- $U_i \mid Z_i$ is uniformly distributed (without loss), and represents individual i's treatment "resistance"
- Vytlicil (2005) shows this model is equivalent to IA's monotonicity

Marginal treatment effect: $\beta(u) = E[Y_i(1) - Y_i(0) \mid U_i = u]$

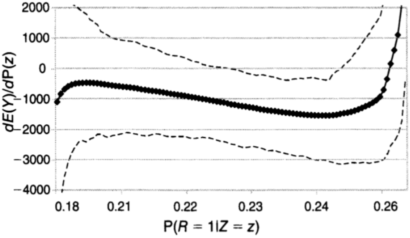
Illustration: Doyle (2007) MTEs of Foster Care Removal



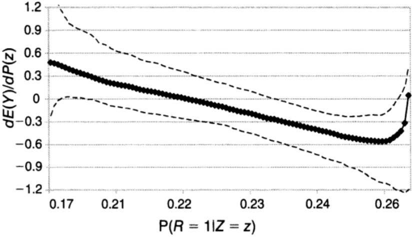
A. Delinquency MTE



B. Teen Motherhood MTE



C. Earnings MTE



D. Employment MTE

Identification: Local IV

Heckman (2000) shows how MTEs are identified:

$$E[Y_i(1) - Y_i(0) \mid U_i = u] = \frac{\partial E[Y_i \mid p(Z_i) = u]}{\partial u}$$

This is just a continuous version of the Wald IV:

$$\frac{\partial E[Y_i \mid p(Z_i) = u]}{\partial u} \approx \frac{E[Y_i \mid Z_i = p^{-1}(u+\delta)] - E[Y_i \mid Z_i = p^{-1}(u)]}{E[D_i \mid Z_i = p^{-1}(u+\delta)] - E[D_i \mid Z_i = p^{-1}(u)]}$$

for small δ, since $E[D_i \mid Z_i = p^{-1}(u^*)] = E[D_i \mid p(Z_i) = u^*] = u^*$

In practice, estimation of these local IVs often employs parametric specifications for $p(Z_i)$ and $E[Y_i \mid p(Z_i)]$, and “plugging in”

Practical Challenges

Finding truly continuous instruments is hard, especially ones which plausibly satisfy monotonicity

- Need individuals to be effectively "ranked" in one dimension of treatment resistance (recall Mogstad et al. (2021))

Estimation can also be tricky – how much are the parametrics doing?

- A recent literature, starting with Brinch et al (2017), studies this by considering multivalued Z_i & parsimonious specifications (more soon)

These days the most common place you'll see MTEs is with so-called "judge"/"examiner" IVs...

Outline

1. The LATE Theorem✓
2. Generalizations✓
3. Characterizing Compliers✓
4. Marginal Treatment Effects✓
5. Application: Agan et al. (2023)

Motivation:

Most initial contacts with the criminal justice system is via misdemeanor charges (e.g. drug possession, disorderly conduct, petty theft...)

- Misdemeanor prosecution may generally have deterrence effects, but may also increase later criminal activity of the prosecuted
- District attorneys (DAs) and policymakers thus face a challenging question of when to prosecute what crimes

Agan et al (ADH) leverage quasi-random DA assignment to estimate the effects of nonprosecution on subsequent criminal justice involvement

- Setting: Suffolk County, MA, 2004-2018
- Main finding: nonprosecution reduces future criminal complains by ≈ 30 percentage points (!)
- Seeming mechanism: the mark of a criminal record

Identification: Quasi-Random DA Assignment

One DA is assigned to a courtroom each day, to arraign all cases

- Key decision is whether/how to advance a case (ADH do not directly observe nonprosecution; instead proxy for it by no further case events)

ADH assume the "leniency" of case c's assigned DA is as-good-as-random controlling for court-by-year-month and court-by-day-of-week FE, in X_c

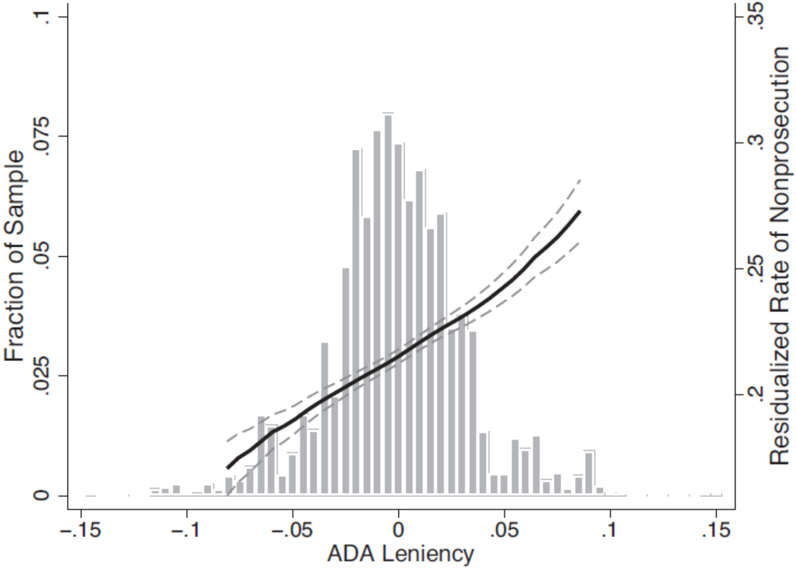
- Specifically, they use a residualized leave-out instrument

$$Z_c = \frac{1}{N_{a(c)} - N_{a(c),i(c)}} \left(\sum_{k:a(k)=a(c)} \widetilde{NP}_k - \sum_{k:a(k)=a(c),i(k)=i(c)} \widetilde{NP}_k \right)$$

where $\widetilde{NP}_c$ is the residual from regressing non-prosecution on X_c, $a(c)$ is the DA assigned to case c, and $i(c)$ is the individual in case c

- This kind of construction is common for judge/examiner IVs...

Estimated Leniency Distribution and First Stage



Checking Instrument Balance

	(1) Nonprosecution	(2) ADA Leniency (St Dev Units)
Number Counts	-0.017***	-0.002
	(0.003)	(0.009)
Number Misdemeanor Counts	0.016***	-0.001
	(0.004)	(0.013)
Number of Serious Misdemeanor Counts	-0.101***	-0.010
	(0.006)	(0.006)
Misd Conviction within Past Year	-0.061***	-0.014
	(0.005)	(0.010)
Felony Conviction within Past Year	-0.050***	-0.013
	(0.006)	(0.017)
Citizen	0.037***	-0.006
	(0.004)	(0.008)
Disorderly/Theft	-0.021**	-0.016
	(0.008)	(0.016)
Motor Vehicle	0.102***	-0.003
	(0.009)	(0.012)
Drug	-0.093***	-0.020
	(0.008)	(0.013)
Male	-0.060***	-0.030***
	(0.004)	(0.011)
Age 24-30	-0.019***	0.012
	(0.005)	(0.012)
Age 31-40	-0.024***	0.003
	(0.005)	(0.011)
Age $\geq$ 41	-0.011**	0.017
	(0.005)	(0.011)
Prob Hispanic	-0.075***	-0.025
	(0.014)	(0.038)
Prob Black	-0.073***	-0.055
	(0.012)	(0.045)
Prob White	-0.037***	-0.037
	(0.013)	(0.040)
Observations	67060	67060
Joint F-Test p-value	0	0.169

Assessing Monotonicity

"Strict" first-stage monotonicity requires DAs to have a common ordering of cases for their appropriateness of prosecution, while varying in leniency

- I.e. no variation in "skill" for determining what cases should be prosecuted (at least in terms of heterogeneous treatment effects)
- Seems tough... and in fact Imbens and Angrist '94 warn against monotonicity in such settings (see their "Example 2")

Frandsen et al. (2023) propose a test of this condition, as well as a weaker condition sufficient for judge-IV to recover a convex average of het FX

- "Average" monotonicity: covariance between DA leniency and heterogeneous effects is non-negative
- Similar to de Chaisemartin (2017): "tolerating defiance"

ADH show the Frandsen et al test passes in 6 out of 9 courtrooms, and that the first stage is positive in subgroups (Dobbie, Goldin, and Yang '18)

Assessing Exclusion

"Strict" exclusion requires DAs to only affect future criminal justice involvement through the decision to prosecute

- Fairly plausible: judges and other prosecutors / lawyers who make future decisions are independently assigned

Like monotonicity, this can be weakened to an "on average" condition:

- Kolesar et al. (2015): exclusion restriction violations are okay when uncorrelated with first-stage effects (here, DA leniency)

ADH probe exclusion by checking whether DA leniency predicts later case outcomes for prosecuted defendants

- Not perfect: conditioning on an "outcome" of DA assignment
- Another strategy could have been instrumenting for multiple treatments with other "leniency" measures

Estimated Effects of Misdemeanor (Non-)Prosecution

	OLS		IV		RF
	(1)	(2)	(3)	(4)	(5)
Panel A: Criminal complaint within two years					
Not prosecuted	-0.14^{***}	-0.10^{***}	-0.36^{***}	-0.29^{***}	
	(0.01)	(0.01)	(0.10)	(0.10)	
			[−0.57, −0.15]	[−0.49, −0.07]	
ADA leniency					-0.16^{**}
					(0.06)
Mean dep var prosecuted	0.37				
Mean dep var prosecuted compliers	0.55				
Panel B: Number criminal complaints within two years					
Not prosecuted	-0.73^{***}	-0.51^{***}	-2.27^{***}	-1.71^{**}	
	(0.04)	(0.03)	(0.66)	(0.67)	
			[−3.67, −1.00]	[−3.15, −0.42]	
ADA leniency					-0.93^{**}
					(0.36)
Mean dep var prosecuted	1.64				
Mean dep var prosecuted compliers	2.84				
Observations	67,060	67,060	67,060	67,060	67,060
Court × time FE	Yes	Yes	Yes	Yes	Yes
Case/def covariates	No	Yes	No	Yes	Yes

What's the Right Instrument?

A more subtle issue with judge IV: using a constructed instrument Z_c

- Estimating leniency as (non-leave-out) sample averages = using 2SLS with judge dummies as (many) IVs
- Simple leave-out leniency averages are similarly equivalent to Jacknife Instrumental Variables Estimation (Angrist et al. '99)
- Kolesar's UJIVE (for "unbiased") estimator may be better for handling many-weak bias given the large number of court-by-time FE controls

ADH show they get negative+significant (though smaller) estimates with these alternative IVs, as well as others (LIML + lasso)

- Paul GP, Kolesar, and I have a work-in-progress UJIVE Stata package, if you ever need to use it yourself!

Alternative IV Estimators

	(1) Main	(2) All Dummies	(3) LIML	(4) UJIVE	(5) UJIVE Interacted	(6) lasso
Not Prosecuted	-0.29***	-0.16***	-0.22**	-0.20**	-0.15***	-0.32***
	(0.10)	(0.05)	(0.09)	(0.08)	(0.04)	(0.11)
Observations	67060	67060	67060	67060	60554	67060
Court x Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Case/Def Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Mean Not Prosecuted	0.372					
Mean Not Prosecuted Compliers	0.550					

Interpreting the Results

Second-stage estimates are large: 60% decrease in criminal complaints relative to the complier control mean

- Compliers seem representative on baseline observables

MTEs are negative for almost all cases, and if anything slope downwards

- Interpretation: DAs do not seem to be trying to deter future criminal behavior by making nonprosecution decisions (maybe not surprising)

Notable heterogeneity by prior complaint/conviction: big effects are concentrated among those without a previous criminal record

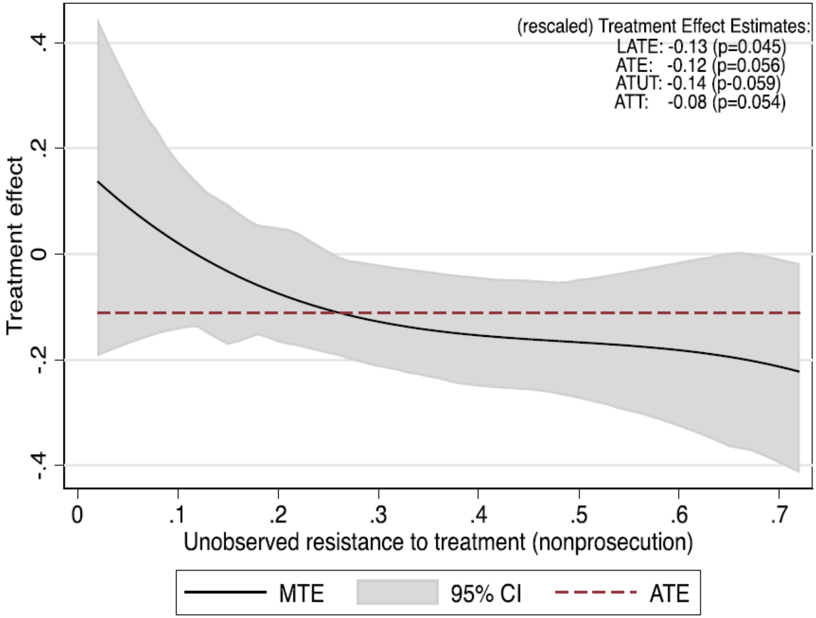
- Consistent with a large literature on the “mark of a criminal record” (e.g. Pager ‘03), though much of this focuses on felonies ...

Bottom line: increasing leniency in misdemeanor prosecution decisions could yield large net social benefits

Complier Characteristics

	(1) $\Pr[X = x]$	(2) $\Pr[X = x \mid Complier]$	(3) Ratio
Counts $= 1$	0.55	0.54	0.99
Counts > 1	0.45	0.42	0.94
Misd Counts $= 1$	0.72	0.74	1.02
Misd Counts > 1	0.28	0.22	0.78
No Serious Misd	0.53	0.57	1.06
Serious Misd	0.47	0.37	0.80
No Misd Conviction 1 Yr Prior	0.94	0.97	1.03
Misd Conviction 1 Yr Prior	0.06	0.03	0.56
No Felony Conviction 1 Yr Prior	0.97	0.98	1.01
Felony Conviction 1 Yr Prior	0.03	0.02	0.77
Not Citizen	0.24	0.10	0.41
Citizen	0.76	0.87	1.14
Any Disorderly/Theft Charge	0.33	0.39	1.18
Any Motor Vehicle Charge	0.45	0.42	0.92
Any Drug Charge	0.21	0.04	0.19
Any Other Charge	0.12	0.08	0.63
Age ≤ 23	0.23	0.29	1.25
Age 24-30	0.25	0.22	0.89
Age 31-40	0.22	0.18	0.83
Age ≥ 41	0.31	0.29	0.96
Male	0.80	0.75	0.94
Female	0.20	0.24	1.21
(Predicted) Black	0.40	0.43	1.06
(Predicted) White	0.36	0.34	0.94
(Predicted) Hispanic	0.22	0.22	1.00
(Admin) Black	0.39	0.34	0.88
(Admin) White	0.14	0.09	0.66
(Admin) Hispanic	0.31	0.28	0.90

MTEs



(rescaled) Treatment Effect Estimates:
LATE: -0.13 (p=0.045)
ATE: -0.12 (p=0.056)
ATUT: -0.14 (p-0.059)
ATT: -0.08 (p=0.054)
Treatment effect
-.4
-.2
0
.2
.4
0
.1
.2
.3
.4
.5
.6
.7
Unobserved resistance to treatment (nonprosecution)
MTE
95% CI
ATE

Heterogeneity by Prior Criminal Records

	Prev. complaint		Prev. DCJIS		
	No	Yes	No	Yes, no conv.	Yes, has conv.
	(1)	(2)	(3)	(4)	(5)
Not prosecuted	−0.18*	−0.04	−0.26***	0.55	0.15
	(0.10)	(0.23)	(0.10)	(0.58)	(0.47)
	[−0.39, 0.04]	[−0.49, 0.51]	[−0.46, −0.05]	[−0.55, 2.95]	[−0.74, 1.41]
Observations	33,367	33,562	38,472	12,172	16,168
Mean dep var prosecuted	0.20	0.52	0.22	0.46	0.61
Mean dep var prosecuted compliers	0.26	0.61	0.32	0.37	0.77