Canonical Research Designs IV: Instrumental Variables II

Paul Goldsmith-Pinkham

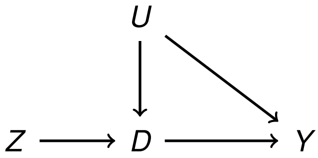
March 30, 2021

Roadmap for Today / Next Class

1. (identification) Reiterating why it's easy to screw up exclusion. Discuss two examples:
   - lottery
   - Weather
2. Marginal treatment effects
3. Discuss why better LATE than never
4. (identification) How monotonicity can fail
5. (identification) montoonicity in more complicated settings (Fuzzy DD)
6. (estimation) Characterizing compliers
7. Next class: (estimation) overid bias (jackknife)
8. Next class: (estimation) weak iv bias (pre-test)
9. Next class: (estimation) IVQTE
10. Next class: (estimation) Lasso IV

Why is the exclusion restriction challenging?

- Recall the key (untestable) feature for IV: exclusion restriction
- In the context of the DAG, the intuition is that Z only affects Y through D
- Intuitively, it feels like something randomly assigned or nearly random should satisfy this, so long as it affects D
- *This is not sufficient*
  - You need to think critically about the IV



Why is the exclusion restriction challenging?

- Consider two examples. First, using Vietnam war lottery numbers as an IV for military service, studying the impact on mortality.
  - Y: death, D: vietnam vet, Z: lottery number
- Lottery number was randomly assigned as a function of birthdate
  - Well-defined design-based view of Z allocation!
- Does that necessarily satisfy exclusion restriction? Seems like a pretty slam dunk IV
  - Clearly affects veteran status
  - Clearly random!

struments are rarely based on actual randomization. A major reason for using this example was to stress that randomization alone does not make a candidate instrument a valid one because randomization does not make the exclusion restriction more plausible. The fact that economists do not always make a clear distinction between ignorability and exclusion restrictions is evidenced by Moffitt's incorrect comment that randomization makes the draft lottery "by *necessity* an obvious and convincing instrument" (italics ours) for the effect of the military service. In fact, one contribution of our approach is to provide a framework that clearly separates ignorability and exclusion assumptions. Both statisticians and economists should find this separation useful and clarifying.

Why is the exclusion restriction challenging?

- Does that necessarily satisfy exclusion restriction?
  - Not necessarily!
- Why? Consider one simple example: being drafted induces you to change your behavior to *avoid* the draft
  - Stay in school
  - Flee to Canada
- This would violate the exclusion restriction!

tery number has no effect on health outcome. Next, consider someone with $D_i(0) = D_i(1) = 0$, who would have managed to avoid military service with a high or low lottery number. For someone exempted from military service for medical reasons, it seems plausible that there was no effect of the draft lottery number. But a draftee who managed to avoid military service by staying in school or moving abroad could experience an effect of Z on future life outcomes that would violate the exclusion restriction. For both these groups of noncompliers, the exclusion restriction requires the researcher to consider a difference in outcomes that were potentially observable, even though after the population was randomly allocated to treatment and control groups, only one of the outcomes was actually observed. In fact, if one could identify compliers and noncompliers, then it would be possible to test the exclusion restriction by comparing average outcomes for noncompliers by assignment status.

Why is the exclusion restriction challenging?

- Second, consider rainfall as an instrument for income in agriculture environments (many crops are heavily dependent on it)
  - This is not uncommon in development papers, as Sarsons (2015) points out
  - Y: conflict, D: income, Z: rainfall

- Exclusion restriction is that rainfall has no effect on conflict *beyond* income
  - While the logic seems reasonable, Sarsons (2015) shows that places with dams (which protect against the income shocks due to rain) have similar conflict to those without dams

- Plausible that while rain is "*random*", it might have many channels

As income is endogenous to conflict, several researchers have relied on rainfall as a source of exogenous variation for income. Miguel et al. (2004) instrument GDP growth in sub-Saharan Africa with rainfall growth and find that lower economic growth increases the probability of civil war. Bohlken and Sergenti (2010) conduct the same analysis in India, instrumenting for state-level GDP with rainfall. They too find that low rainfall growth increases the number of riots that a state experiences in a given year. Rainfall is a plausible candidate instrument if the country or region in question is economically dependent on rain-fed agriculture. Low levels of rainfall result in crop failure, thereby depressing rural income.

dams. These dams protect against weather shocks, providing districts downstream of the dam with water during droughts and holding excess rainwater during periods of heavy rainfall. I identify districts that are downstream of dams (dam-fed districts) and find that while agricultural production in rain-fed districts (those upstream of a dam) is dependent on rainfall, production in dam-fed districts is uncorrelated with the amount of rain. Yet despite having little influence on production in these districts, rainfall still predicts riot incidence, suggesting that rainfall affects conflict through some other channel.

Exclusion Restrictions

- Even with a variable that is near-random in its allocation, the exclusion restriction is not always satisfied
    - Worse yet, it's a fundamentally untestable restriction

- Using an IV requires thinking carefully about justifying the exclusion restriction
    - It can also be useful to think about what violations in the restriction implies

- $Y_i = Y_i(Z_i, D_i(Z_i))$. Let $H_i = Y_i(1, d) - Y_i(0, d)$, where d is 1 for an always-taker, and d is 0 for a never-taker.
    - Under monotonicity, (Angrist, Imbens, and Rubin (1996)):

$$\frac{E(Y_i(1, D_i(1)) - Y_i(0, D_i(0)))}{E(D_i(1) - D_i(0))} = E(Y_i(1, D_i(1)) - Y_i(0, D_i(0)) | i \text{ is a complier})$$

$$+ E(H_i | i \text{ is a noncomplier}) \cdot \frac{\Pr(i \text{ is a noncomplier})}{Pr(i \text{ is a complier})}$$

Knowable things about Exclusion Restriction violations

$$\frac{E(Y_i(1, D_i(1)) - Y_i(0, D_i(0)))}{E(D_i(1) - D_i(0))} = E(Y_i(1, D_i(1)) - Y_i(0, D_i(0)) | i \text{ is a complier})$$
$$+ E(H_i | i \text{ is a noncomplier}) \cdot \frac{\Pr(i \text{ is a noncomplier})}{Pr(i \text{ is a complier})}$$

- Key point is that the larger the complier group is, the less the bias from violations in the exclusion restriction

- If the effect of the exclusion is additive ($Y_i(1, 0) - Y_i(0, 0) = Y_i(1, 1) - Y_i(0, 1)$):

$$\frac{E(Y_i(1, D_i(1)) - Y_i(0, D_i(0)))}{E(D_i(1) - D_i(0))} = \tau_{IV,LATE} + \frac{E(H_i | i \text{ is a noncomplier})}{Pr(i \text{ is a complier})}$$

- See Angrist, Imbens and Rubin (1996) for more details

Modeling treatment choice

- Let's now revisit the choice of treatment. (Heckman and Vytlacil (1999, PNAS))
  - Let $D_i^* = \mu_D(Z_i) - \epsilon_{Di}$, $D_i = 1(D_i^* \geq 0)$
  - E.g. D_i^* is net utility gain from choosing D_i

- $Y_i = Y_{i1} D_i + Y_{i0}(1 - D_i) = \mu_1(\epsilon_{i1}) D_i + \mu_0(\epsilon_{i0})(1 - D_i)$
  - Hence if U_{Di} is correlated with $\epsilon_{1i}, \epsilon_{0i}$, this will cause sorting!
  - Omitting characteristics X_i for simplicity

- Finally, let $P(z) = Pr(D = 1 | Z = z) = F_{U_D}(\mu_D(z))$ and $\tilde{U}_D = F_{U_D}(U_D)$

- This latent index model captures a nice way to think about IV
  - We will assume exclusion restriction; the errors are absolutely continuous; and that Z is independent

Definition of parameters

- Under this setting, we can consider a number of estimands. Let $\Delta = Y_{i1} - Y_{i0}$
  1. $\Delta^{ATE} = E(\Delta)$
  2. $\Delta^{ATT}(D = 1) = E(\Delta | D = 1)$
  3. $\Delta^{LATE}(P(z), P(z')) = \frac{E(Y|p(z)) - E(Y|p(z'))}{p(z) - p(z')}$
     - Consider $E(Y|p(z)) = P(z)E(Y_1|P(z), D = 1) + (1 - P(z))E(Y_1|P(z), D = 1)$
     - This can be written as (by first fundamental theorem of calculus):

$$E(Y|p(z)) = \int_0^{P(z)} E(Y_1|\widetilde{U} = u)du + \int_{P(z)}^1 E(Y_1|\widetilde{U} = u)du$$

     - Hence:

$$E(Y|p(z)) - E(Y|p(z')) = \int_{P(z')}^{P(z)} E(Y_1|\widetilde{U} = u)du - \int_{P(z')}^{P(z)} E(Y_1|\widetilde{U} = u)du$$

$$\Delta^{LATE}(P(z), P(z')) = E(\Delta|P(z') \leq \widetilde{U}_D \leq P(z))$$

  4. $\Delta^{LIV}(P(z)) = \frac{\partial E(Y|P(Z) = P(z))}{\partial P(z)} = \lim_{P(z') \to P(z')} \Delta^{LATE}$

The marginal treatment effect (MTE) as a building block

- Each estimand can be constructed from the underlying local effects (now referred to as Marginal Treatment Effects (MTE))

- These MTE identify the effect for an individual who is shifted by the change in the instrument on the margin

- Hence if Z increases the incentive of participating in a program, the local average treatment effect exploiting this will integrate over the MTE of the compliers

1. $\Delta^{LIV}(P(z)) = E(\Delta | \widetilde{U}_D = P(z))$
2. $\Delta^{ATE} = \int_0^1 E(\Delta | \widetilde{U}_D = u) du$
3. $\Delta^{ATT}(D = 1, P(z)) = \int_0^{P(z)} E(\Delta | \widetilde{U}_D = u) du / P(z)$
4. $\Delta^{ATT}(D = 1) = \int_0^1 \Delta^{ATT}(D = 1, P(z)) dF_{P(Z)|D=1}$
5. $\Delta^{LATE}(P(z), P(z')) = \int_{P(z')}^{P(z)} E(\Delta | \widetilde{U}_D = u) du / (P(z) - P(z'))$

The marginal treatment effect (MTE) as a building block

- Formally, the MTE can be estimated by fitting Y on functions of $\hat{p}(Z)$. Then take the derivative of that function
- Crucially, needs a lot of values to this instrument!
  - With only a binary instrument, or discrete instrument, can't really estimate a derivative
- Can use these MTE to try to reweight and construct potentially more policy relevant treatment parameters
- This view is driven by the idea that LATE is just not a policy relevant piece, b/c it reflects the self-selection choices of a particular group

The marginal treatment effect (MTE) as a building block

- Formally, the MTE can be estimated by fitting Y on functions of $\hat{p}(Z)$. Then take the derivative of that function

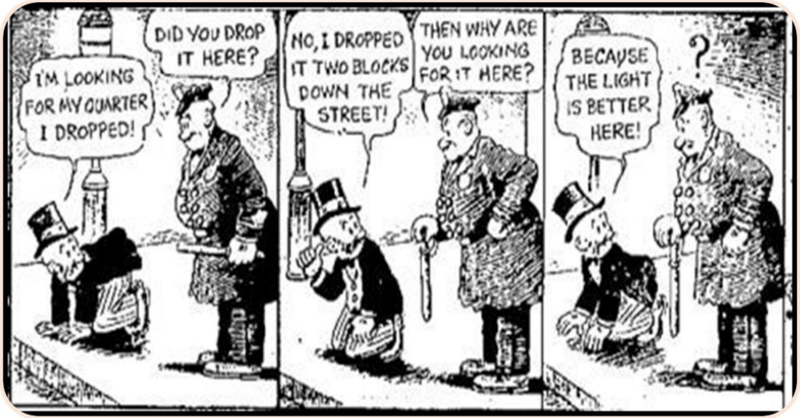
- Crucially, needs a lot of values to this instrument!
  - With only a binary instrument, or discrete instrument, can't really estimate a derivative

- Can use these MTE to try to reweight and construct potentially more policy relevant treatment parameters

- This view is driven by the idea that LATE is just not a policy relevant piece, b/c it reflects the self-selection choices of a particular group

Is LATE great? The criticism

- Correctly done, IV gives a very internally valid estimate
- But external validity is worrisome
  - Is the range $[P(z'), P(z)]$ special?
- Is it informative?
- Argument confounded with poor IV usage (exclusion restrictions)



Better LATE than never

- A well-identified design gives us a real set of facts
    - We can debate the merits of each design, but establishes a gold standard
- Concern is that IV for settings of interest is impossible
    - Evidence suggests this is not true. Creative researchers have found many good examples
    - Innovations in structural methods have incorporated credible designs into structural models (e.g. sufficient statistics)
- Even if there are not experiments design for the counterfactual of interest, an internally valid estimate can give important grounding for a structural model that attempts to extrapolate
- Key point: using *poorly* identified estimates is **not better**
    - No clarity on what is causal
    - The LATE literature is useful because it highlights *what is knowable*

A visual for compliers / non-compliers

Table 1. *Causal Effect of Z on Y, $Y_i(1, D_i(1)) - Y_i(0, D_i(0))$, for the Population of Units Classified by $D_i(0)$ and $D_i(1)$*

		$D_i(0)$	
		0	*1*
$D_i(1)$	0	$Y_i(1, 0) - Y_i(0, 0) = 0$ Never-taker	$Y_i(1, 0) - Y_i(0, 1) = -(Y_i(1) - Y_i(0))$ Defier
	1	$Y_i(1, 1) - Y_i(0, 0) = Y_i(1) - Y_i(0)$ Complier	$Y_i(1, 1) - Y_i(0, 1) = 0$ Always-taker

How monotonicity can fail

- Three examples of possible failure (from de Chaisemartin (2017)):
    1. Examiner designs: Imagine you have two judges who decide guilty/not guilty, and you randomly assign them. If monotonicity means that one judge is always stricter (for every person), then monotonicity holds. However, easy to envision failures of this (e.g. strictness on different crimes, different types of people)

$$P(z) > P(z')$$
vs.
$$D_i(z) \geq D_i(z') \forall i$$

How monotonicity can fail

- Three examples of possible failure (from de Chaisemartin (2017)):
  1. Examiner designs: Imagine you have two judges who decide guilty/not guilty, and you randomly assign them. If monotonicity means that one judge is always stricter (for every person), then monotonicity holds. However, easy to envision failures of this (e.g. strictness on different crimes, different types of people)
  2. Sibling-sex composition: Angrist Evan (1998) uses two siblings of same sex as instrument for third child, b/c more likely to have a third child. But, some families may want two boys, vs. two girls. Same sex composition could generate defiers

$$P(z) > P(z')$$

vs.

$$D_i(z) \geq D_i(z') \forall i$$

How monotonicity can fail

- Three examples of possible failure (from de Chaisemartin (2017)):
    1. Examiner designs: Imagine you have two judges who decide guilty/not guilty, and you randomly assign them. If monotonicity means that one judge is always stricter (for every person), then monotonicity holds. However, easy to envision failures of this (e.g. strictness on different crimes, different types of people)
    2. Sibling-sex composition: Angrist Evan (1998) uses two siblings of same sex as instrument for third child, b/c more likely to have a third child. But, some families may want two boys, vs. two girls. Same sex composition could generate defiers
    3. Encouragement designs may backfire if the nudge is too heavy-handed (Duflo and Saez (2003))

$$P(z) > P(z')$$

vs.

$$D_i(z) \geq D_i(z') \forall i$$

Avoiding monotonicity assumption

- There are a papers that attempt to characterize alternative assumptions (instead of monotonicity)
- These may work better for your particular application! See, e.g.,
  - de Chaisemartin "Tolerating defiers" (2017)
  - Frandsen et al. "Judging Judge Fixed Effects" (2019)
- In the end, however, the problem is that defiers create a fundamental mismeasurement problem
  - Any solution will attempt to alleviate this by adding extra assumptions
  - These solutions need to be situational!

IV is a useful tool for estimating causal effects in many settings

- Note that IV is simply a tool for evaluating a causal impact using an instrument

- An example: imagine we use diff-in-diff to induce random variation in a policy. This can be combined with IV to construct a causal estimate:

$$Y_i = \alpha_i + \gamma_t + D_{it}\beta + \epsilon_{it} \tag{1}$$
$$D_{it} = \alpha_i + \gamma_t + Z_{it}\beta + \epsilon_{it} \tag{2}$$
$$\tag{3}$$

- The same issues apply for both DinD and IV, but can be a powerful way to convert a DinD evaluation of a policy into a structural parameter of interest
  - Need exclusion to hold, and monotonicity as well, if there are heterogeneous effects
  - See discussion in “Fuzzy difference-in-differences” by de Chaisemartin and d’Haultfoeuille and “Interpreting Instrumented Difference-in-Differences” by Hudson, Hull and Liebersohn

- The “fuzzy” label will come again with regression discontinuity

Understanding compliers

- Under the LATE assumptions, we can know a decent amount about the compliers.

- First, if D_i is binary, the difference in propensity scores (first stage) is exactly the complier share:

$$Pr(D_i(1) > D_i(0)) = E(D_i(1) - D_i(0)) = E(D_i|Z_i = 1) - E(D_i|Z_i = 0)$$

- We can even know the share treated, using Bayes' rule:

$$\begin{aligned} Pr(D_i(1) > D_i(0)|D_i = 1) &= \frac{P(D_i = 1|D_i(1) > D_i(0)) \times Pr(D_i(1) > D_i(0))}{Pr(D_i = 1)} \\ &= \frac{P(Z_i = 1) \times Pr(D_i(1) > D_i(0))}{Pr(D_i = 1)} \end{aligned}$$

- This identifies the share of compliers

Understanding compliers

- Second, we can actually know average characteristics of compliers using the same logic, if the characteristic is discrete.

- Consider X_i binary:

$$
\begin{aligned}
Pr(X_i | D_i(1) > D_i(0)) &= \frac{P(D_i(1) > D_i(0) | X_i) \times Pr(X_i)}{Pr(D_i(1) > D_i(0))} \\
&= \frac{(E(D_i | Z_i = 1, X_i) - E(D_i | Z_i = 0, X_i)) \times Pr(X_i)}{E(D_i | Z_i = 1, X_i) - E(D_i | Z_i = 0, X_i)}
\end{aligned}
$$

- Note that if we scale by $Pr(X_i)$, we get the relative probability of X_i compared to the overall pop
  - Just the ratio of the first stages for each group!

Understanding compliers

- Finally, Abadie (2002, JASA) shows how to construct the potential outcomes for the compliers

- Let $g(Y)$ be any measurable function. Then,

$$E(g(Y_i(1))|D_i(1) > D_i(0))) = \frac{E(D_i g(Y_i)|Z_i = 1) - E(D_i g(Y_i)|Z_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)}$$

$$E(g(Y_i(0))|D_i(1) > D_i(0))) = \frac{E((1 - D_i)g(Y_i)|Z_i = 1) - E((1 - D_i)g(Y_i)|Z_i = 0)}{E(1 - D_i|Z_i = 1) - E(1 - D_i|Z_i = 0)}$$

- Simplest case of $g(\cdot)$ as the identity gives the means for the two marginals

- Can identify distributional effects by the dummy functions for compliers
    - $F_1(y) = E(1(Y_i(1) \leq y|D_i(1) > D_i(0))$
    - $F_0(y) = E(1(Y_i(0) \leq y|D_i(1) > D_i(0))$

Abadie (2002) OLS distributions

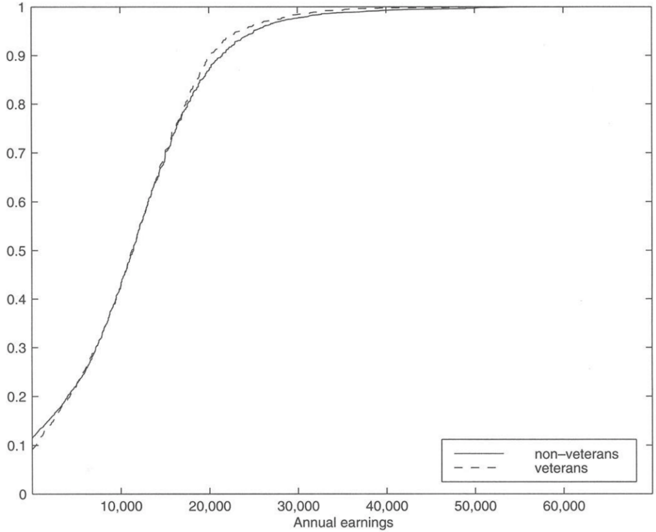


Figure 1. Empirical Distributions of Earnings for Veterans and Nonveterans.

Abadie (2002) complier distributions

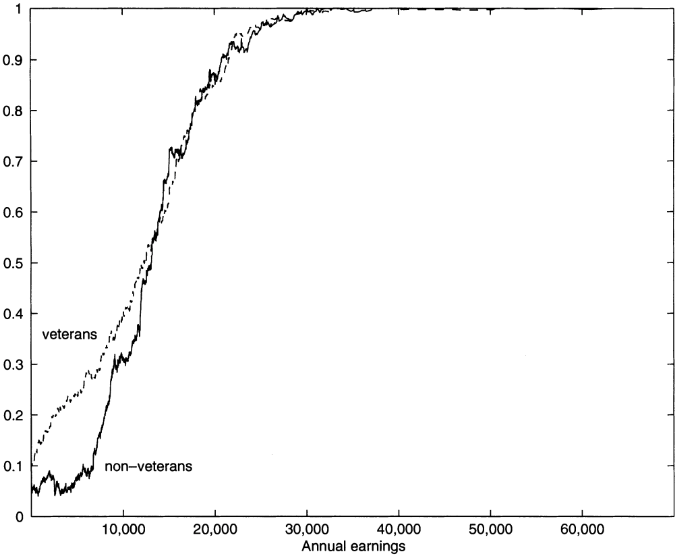


Figure 2. Estimated Distributions of Potential Earnings for Compliers.