Lecture 10: Instrumental Variables II

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

March 7, 2021

Last time:

- Basics of IV from linear regression perspective.
- Constant effects assumption to derive straightforward results.

Last time:

- Basics of IV from linear regression perspective.
- Constant effects assumption to derive straightforward results.

For today:

- IV from more general potential outcomes perspective.
- Allows for effect heterogeneity.

Last time:

- Basics of IV from linear regression perspective.
- Constant effects assumption to derive straightforward results.

For today:

- IV from more general potential outcomes perspective.
- Allows for effect heterogeneity.

Some key concepts for today:

- “Local” average treatment effect (LATE).
- “Principal strata.”
- “Non-compliance” and “intention to treat.”

Motivating Example

ESTIMATING THE IMPACT OF THE HAJJ: RELIGION AND TOLERANCE IN ISLAM'S GLOBAL GATHERING*

DAVID CLINGINGSMITH
ASIM IJAZ KHWAJA
MICHAEL KREMER

...

Among the successful Hajj applicants we surveyed, 99% went on the Hajj. Some unsuccessful lottery applicants secure a place with a private Hajj operator or through the special quota. Thus, 11% of those who were unsuccessful in the government lottery still performed the Hajj that year.

Because compliance with the lottery is not perfect, we use success in the lottery as an instrumental variable to estimate the effect of performing the Hajj.

(QJE, 2009)

Motivating Example

Attendance behavior for different types

Attendance type	Lottery outcome: Unsuccessful	Successful	Explanation
1. Always go	Go	Go	Rich? Very pious?
2. Go only if successful	Stay	Go	Not rich?
3. Never go	Stay	Stay	Not rich? Busy?
4. Go only if unsuccessful	Go	Stay	Crazy?

- Lottery assignment is randomly assigned.
- Attendance (stay or go) is endogenous:

Motivating Example

Attendance behavior for different types

Attendance type	Lottery outcome: Unsuccessful	Successful	Explanation
1. Always go	Go	Go	Rich? Very pious?
2. Go only if successful	Stay	Go	Not rich?
3. Never go	Stay	Stay	Not rich? Busy?
4. Go only if unsuccessful	Go	Stay	Crazy?

- Lottery assignment is randomly assigned.
- Attendance (stay or go) is endogenous:
  - Those who go are mix of 1,2, 4.
  - Those who stay are mix of 2,3, 4.
  - $\Rightarrow$ those who go non-comparable to those who stay (avg. potential outcomes probably different).

Motivating Example

Attendance behavior for different types

Attendance type	Lottery outcome: Unsuccessful	Lottery outcome: Successful	Explanation
1. Always go	Go	Go	Rich? Very pious?
2. Go only if successful	Stay	Go	Not rich?
3. Never go	Stay	Stay	Not rich? Busy?
4. Go only if unsuccessful	Go	Stay	Crazy?

- Lottery assignment is randomly assigned.
- Attendance (stay or go) is endogenous:
  - Those who go are mix of 1,2, 4.
  - Those who stay are mix of 2,3, 4.
  - $\Rightarrow$ those who go non-comparable to those who stay (avg. potential outcomes probably different).
- But: randomization $\Rightarrow$ proportion of 1,2,3,and 4 types are equal over *lottery* outcomes.
- Also, type 4 seems implausible (monotonicity).

Motivating Example

Attendance behavior for different types

Attendance type	Lottery outcome: Unsuccessful	Successful	Explanation
1. Always go	Go	Go	Rich? Very pious?
2. Go only if successful	Stay	Go	Not rich?
3. Never go	Stay	Stay	Not rich? Busy?
4. Go only if unsuccessful	Go	Stay	Crazy?

- Lottery assignment is randomly assigned.
- Attendance (stay or go) is endogenous:
  - Those who go are mix of 1,2, 4.
  - Those who stay are mix of 2,3, 4.
  - $\Rightarrow$ those who go non-comparable to those who stay (avg. potential outcomes probably different).
- But: randomization $\Rightarrow$ proportion of 1,2,3,and 4 types are equal over *lottery* outcomes.
- Also, type 4 seems implausible (monotonicity).
- These conditions open possibility that effect for *type 2* identified.

Setting

- Sample S indexed by $i = 1, .., N$ drawn from a large population P.
- Suppose binary $Z_i = 0, 1$ that denotes treatment *assigned*.

Setting

- Sample S indexed by $i = 1, .., N$ drawn from a large population P.
- Suppose binary $Z_i = 0, 1$ that denotes treatment *assigned*.
- Binary potential treatment responses to Z_i,
  - $D_{1i} = 0, 1$ and $D_{0i} = 0, 1$,
  - $\Rightarrow$ treatment *actually taken up* by i is
    $D_i = Z_i D_{1i} + (1 - Z_i) D_{0i} = D_{0i} + (D_{1i} - D_{0i}) Z_i$.

Setting

- Sample S indexed by $i = 1, .., N$ drawn from a large population P.
- Suppose binary $Z_i = 0, 1$ that denotes treatment *assigned*.
- Binary potential treatment responses to Z_i,
  - $D_{1i} = 0, 1$ and $D_{0i} = 0, 1$,
  - $\Rightarrow$ treatment *actually taken up* by i is $D_i = Z_i D_{1i} + (1 - Z_i) D_{0i} = D_{0i} + (D_{1i} - D_{0i}) Z_i$.
- Potential outcomes as functions of both Z_i and D_i: $Y_i(d, z)$ for $D_i = d$ and $Z_i = z$.
- (Keeping things general for now, "exclusion restriction" will come soon.)

Setting

- Sample S indexed by $i = 1, .., N$ drawn from a large population P.
- Suppose binary $Z_i = 0, 1$ that denotes treatment *assigned*.
- Binary potential treatment responses to Z_i,
  - $D_{1i} = 0, 1$ and $D_{0i} = 0, 1$,
  - $\Rightarrow$ treatment *actually taken up* by i is $D_i = Z_i D_{1i} + (1 - Z_i) D_{0i} = D_{0i} + (D_{1i} - D_{0i}) Z_i$.
- Potential outcomes as functions of both Z_i and D_i: $Y_i(d, z)$ for $D_i = d$ and $Z_i = z$.
- (Keeping things general for now, "exclusion restriction" will come soon.)
- Observed outcome:

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

Setting

- Four conceivable potential outcomes,

$$Y_i(D_i, Z_i) \in \{Y_i(1,1), Y_i(1,0), Y_i(0,1), Y_i(0,0)\}$$

- For i, we have (D_{1i}, D_{0i}), so only two of the four are ever realized for i.

Setting

- Four conceivable potential outcomes,

$$Y_i(D_i, Z_i) \in \{Y_i(1,1), Y_i(1,0), Y_i(0,1), Y_i(0,0)\}$$

- For i, we have (D_{1i}, D_{0i}), so only two of the four are ever realized for i.
- We can define various unit-level causal effects:
  - Unit level causal effect of treatment *taken up* on outcome:

$$Y_i(1,1) - Y_i(0,1) \quad \text{or} \quad Y_i(1,0) - Y_i(0,0).$$

  - Reduced-form unit level causal effect of treatment *assigned* on outcome:

$$Y_i(D_{1i}, 1) - Y_i(D_{0i}, 0)$$

  - Finally, unit level causal effect of treatment *assigned* on treatment *taken up*:

$$D_{1i} - D_{0i}$$

Identification

Identification

(Constructive proof of LATE Theorem, Angrist et al. 1996)

Identification

(Constructive proof of LATE Theorem, Angrist et al. 1996)

- Assumption 1: Z_i is randomly assigned:

$$(\{Y_i(d,z)\forall d,z\}, D_{1i}, D_{0i}) \perp\!\!\!\perp Z_i.$$

- Based on assignment mechanism:
  - Known for experiments,
  - presumed known for natural experiments.

Identification

(Constructive proof of LATE Theorem, Angrist et al. 1996)

- Assumption 1: Z_i is randomly assigned:

$$(\{Y_i(d,z) \forall d,z\}, D_{1i}, D_{0i}) \perp\!\!\!\perp Z_i.$$

- Based on assignment mechanism:
  - Known for experiments,
  - presumed known for natural experiments.
- Identifies reduced-form average causal effect of Z_i:

$$\begin{aligned} \mathrm{E}[Y_i|Z_i=1] - \mathrm{E}[Y_i|Z_i=0] &= \mathrm{E}[Y_i(D_{1i},1)|Z_i=1] - \mathrm{E}[Y_i(D_{0i},0)|Z_i=0] \\ &= \mathrm{E}[Y_i(D_{1i},1) - Y_i(D_{0i},0)] \end{aligned}$$

- "Intention to treat effect" (ITT).
- (Sub CIA for Assn. 1 to generalize.)

ITT

Brief aside:

- You are hired to study effects of a program.
- You design a study that randomly assigns access.
- Take-up among those assigned to access is 20%.
- Take-up among those assigned to control is 0.
- You go to your boss worried that the study was a failure.
- Your boss says, "don't worry. The ITT is actually the estimand of interest. Just use treatment assigned, not treatment taken up."
- Does this make sense to you? Why or why not?

Identification

- Assn. 1 identifies average effect of Z_i on D_i:

$$\begin{aligned} \mathrm{E}[D_i|Z_i=1]-\mathrm{E}[D_i|Z_i=0] &= \mathrm{E}[D_{1i}|Z_i=1]-\mathrm{E}[D_{0i}|Z_i=0] \\ &= \mathrm{E}[D_{1i}-D_{0i}]. \end{aligned}$$

Identification

- Assn. 1 identifies average effect of Z_i on D_i:

$$\begin{aligned} \mathrm{E}[D_i|Z_i = 1] - \mathrm{E}[D_i|Z_i = 0] &= \mathrm{E}[D_{1i}|Z_i = 1] - \mathrm{E}[D_{0i}|Z_i = 0] \\ &= \mathrm{E}[D_{1i} - D_{0i}]. \end{aligned}$$

- If = 1 $\Rightarrow$ "perfect compliance."
- If $\neq 1 \Rightarrow$ some "non-compliance."
- Measures rate of "compliance" versus "defiance," "never-taking," and "always-taking."

Identification

- Assn. 1 identifies average effect of Z_i on D_i:

$$\begin{aligned} \mathrm{E}[D_i|Z_i=1]-\mathrm{E}[D_i|Z_i=0] &= \mathrm{E}[D_{1i}|Z_i=1]-\mathrm{E}[D_{0i}|Z_i=0] \\ &= \mathrm{E}[D_{1i}-D_{0i}]. \end{aligned}$$

- If = 1 $\Rightarrow$ "perfect compliance."
- If $\neq 1 \Rightarrow$ some "non-compliance."
- Measures rate of "compliance" versus "defiance," "never-taking," and "always-taking."
- Assumption 2: Treatment assignment has some effect on take-up outcomes:

$$\mathrm{E}[D_{1i}-D_{0i}] \neq 0.$$

- Testable (first stage).

Identification

- Assumption 3: D fully mediates effect of Z (exclusion restriction):

$$Y_i(d,0) = Y_i(d,1) = Y_{di} \text{ for } d = 0,1.$$

- Z_i has no effect on outcomes except through D_i.

Identification

- Assumption 3: D fully mediates effect of Z (exclusion restriction):

$$Y_i(d,0) = Y_i(d,1) = Y_{di} \text{ for } d = 0,1.$$

- Z_i has no effect on outcomes except through D_i.
- Testable implications along the lines of Pearl's instrumental inequalities (remember from last time).

Identification

- Assumption 3: D fully mediates effect of Z (exclusion restriction):

$$Y_i(d,0) = Y_i(d,1) = Y_{di} \text{ for } d = 0,1.$$

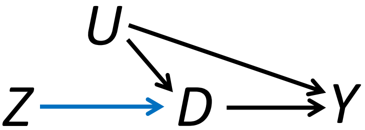
- Z_i has no effect on outcomes except through D_i.
- Testable implications along the lines of Pearl's instrumental inequalities (remember from last time).
- Assn. 3 allows dimension reduction:

$$\begin{aligned}
Y_i =& D_i[Z_i Y_i(1,1) + (1-Z_i)Y_i(1,0)] \\
&+ (1-D_i)[Z_i Y_i(0,1) + (1-Z_i)Y_i(0,0)] \\
&= D_i Y_{1i} + (1-D_i)Y_{0i} \\
&= Y_{0i} + (Y_{1i} - Y_{0i})D_i.
\end{aligned}$$

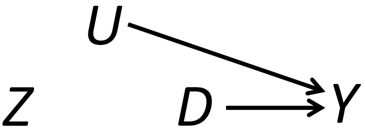
- (Remember D_i not randomly assigned though.)

Identification

Checking assumptions graphically:



Assn 2: 1^{st} stage,
Z not indt. of D.

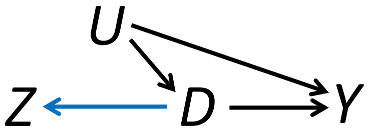


Assn 3: exclusion,
Z indt. of Y after
removing path to D

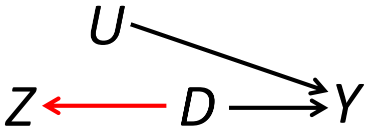
Standard case

Identification

Checking assumptions graphically:



Assn 2: 1st stage, Z not indt. of D.

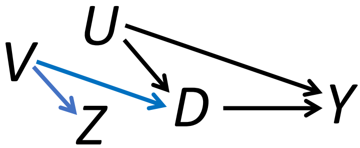


Assn 3: exclusion, **fails**

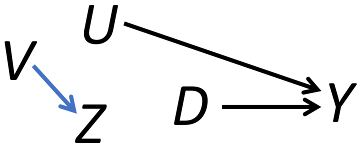
Reverse causal order for Z and D

Identification

Checking assumptions graphically:



Assn 2: 1^{st} stage,
Z not indt. of D.



Assn 3: exclusion,
Z indt. of Y after
removing path to D

Common cause Z and D

Identification

- Assumption 4: Effect of assignment on take-up is *monotonic*:

$$D_{1i} - D_{0i} \geq 0 \ \forall i.$$

- No one "defies" their assignment.
- (The $\geq$ is WLOG. If $D_{1i} - D_{0i} \leq 0$ under current labeling of D_{1i} and D_{0i}, then just relabel.)

Identification

- Define "principal strata" based on treatment take-up:

Principal Stratum	(D_{1i}, D_{0i})	$D_{1i} - D_{0i}$
Always-takers	$(1,1)$	0
Never-takers	$(0,0)$	0
Compliers	$(1,0)$	1
Defiers	$(0,1)$	-1

- Assn 4: no Defiers.

Identification

- Define "principal strata" based on treatment take-up:

Principal Stratum	(D_{1i}, D_{0i})	$D_{1i} - D_{0i}$
Always-takers	$(1,1)$	0
Never-takers	$(0,0)$	0
Compliers	$(1,0)$	1
Defiers	$(0,1)$	-1

- Assn 4: no Defiers.
- Assn 1: expected proportion of Always-takers, Never-takers, and Compliers equal over Z_i (two random samples).

Identification

- Define "principal strata" based on treatment take-up:

Principal Stratum	(D_{1i}, D_{0i})	$D_{1i} - D_{0i}$
Always-takers	$(1,1)$	0
Never-takers	$(0,0)$	0
Compliers	$(1,0)$	1
Defiers	$(0,1)$	-1

- Assn 4: no Defiers.
- Assn 1: expected proportion of Always-takers, Never-takers, and Compliers equal over Z_i (two random samples).
- Assn 1, 2, 4: expected proportion of Compliers equals

$$\mathrm{E}[D_{1i} - D_{0i}] \neq 0.$$

- Assn 3: potential outcomes for Always-takers and Never-takers are equal over Z_i.

- Principal strata proportions independent of Z.
- When $Z = 0$ you can figure out p_A.
- When $Z = 1$ you can figure out p_N.
- Then you can back out p_C.

Identification

Let's put it all together. Start with,

$$\begin{aligned} Y_i &= Y_{0i} + (Y_{1i} - Y_{0i})D_i && \text{(Assumption 3)} \\ &= Y_{0i} + (Y_{1i} - Y_{0i})[D_{0i} + (D_{1i} - D_{0i})Z_i] \end{aligned}$$

Therefore,

$$\begin{aligned} \mathrm{E}[Y_i|Z_i = 1] &= \mathrm{E}[Y_{0i} + (Y_{1i} - Y_{0i})D_{1i}|Z_i = 1] \\ &= \mathrm{E}[Y_{0i} + (Y_{1i} - Y_{0i})D_{1i}] && \text{(Assumption 1)} \end{aligned}$$

By symmetry,

$$\mathrm{E}[Y_i|Z_i = 0] = \mathrm{E}[Y_{0i} + (Y_{1i} - Y_{0i})D_{0i}]$$

Thus,

$$\mathrm{E}[Y_i|Z_i = 1] - \mathrm{E}[Y_i|Z_i = 0] = \mathrm{E}[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})].$$

Identification

By total probability,

$$
\begin{aligned}
\mathrm{E}\left[(Y_{1i}-Y_{0i})(D_{1i}-D_{0i})\right] =& \mathrm{E}\left[(Y_{1i}-Y_{0i})*0|D_{1i}-D_{0i}=0\right]\Pr[D_{1i}-D_{0i}=0] \\
& + \mathrm{E}\left[(Y_{1i}-Y_{0i})(D_{1i}-D_{0i})|D_{1i}-D_{0i}\neq 0\right] \\
& \times \Pr[D_{1i}-D_{0i}\neq 0] \\
& = \mathrm{E}\left[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}\right]\Pr[D_{1i}>D_{0i}] \\
& \qquad \text{(Assumption 4)}
\end{aligned}
$$

We have already shown that Assumptions 1,2, and 4 imply,

$$
\mathrm{E}\left[D_i|Z_i=1\right]-\mathrm{E}\left[D_i|Z_i=0\right]=\mathrm{E}\left[D_{1i}-D_{0i}\right]=\Pr[D_{1i}>D_{0i}]\neq 0.
$$

Stack the two and you get,

$$
\frac{\mathrm{E}\left[Y_i|Z_i=1\right]-\mathrm{E}\left[Y_i|Z_i=0\right]}{\mathrm{E}\left[D_i|Z_i=1\right]-\mathrm{E}\left[D_i|Z_i=0\right]}=\mathrm{E}\left[Y_{1i}-Y_{0i}|D_{1i}>D_{0i}\right]\equiv\rho_C
$$

Identification

Let's study ρ_C:

$$\rho_C = \frac{\mathrm{E}[Y_i|Z_i=1] - \mathrm{E}[Y_i|Z_i=0]}{\mathrm{E}[D_i|Z_i=1] - \mathrm{E}[D_i|Z_i=0]} = \mathrm{E}[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

Recall our principal strata:

Principal Stratum	(D_{1i}, D_{0i})	$D_{1i} - D_{0i}$
Always-takers	$(1,1)$	0
Never-takers	$(0,0)$	0
Compliers	$(1,0)$	1
~~Defiers~~	$\cancel{(0,1)}$	$\cancel{-1}$

$\rho_C =$ *average causal effect of D_i for Compliers*
"Local average treatment effect" (LATE).
"LATE Theorem" (Angrist et al. 1996).

Identification

Comparing ρ_C to the ATT and ATC:

- ρ_C is *not* ATT when Always-takers present nor ATC when Never-takers present.
- But if compliance is one-sided, it is either ATT or ATC:
  - E.g., suppose no Always-takers.
  - Then, $D_i = 1$ only for Compliers.
  - So ρ_C is in this case the ATT.
- cf. MHE pp. 158-166.

Estimation

Estimation

The left-hand side indicates how to estimate ρ_C:

$$\rho_C = \frac{\mathrm{E}[Y_i|Z_i = 1] - \mathrm{E}[Y_i|Z_i = 0]}{\mathrm{E}[D_i|Z_i = 1] - \mathrm{E}[D_i|Z_i = 0]} = \mathrm{E}[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$$

- Numerator: ITT or “reduced-form” effect of Z_i on Y_i.
- Denominator: compliance rate or “first stage” effect of Z_i on D_i.
- Under Assns, denominator btwn 0 and 1 $\Rightarrow$ “inflates” ITT.
- Denominator close to 0 $\Rightarrow$ “weak instrument” problems.

Estimation

Note that for the numerator of ρ_C,

$$\begin{aligned}
&\mathrm{E}[Y_i|Z_i=1]-\mathrm{E}[Y_i|Z_i=0]\\
&\quad=\frac{\mathrm{E}[Y_i|Z_i=1]\mathrm{E}[Z_i]}{\mathrm{E}[Z_i]}\\
&\qquad-\frac{\mathrm{E}[Y_i|Z_i=0](1-\mathrm{E}[Z_i])+\mathrm{E}[Y_i|Z_i=1]\mathrm{E}[Z_i]-\mathrm{E}[Y_i|Z_i=1]\mathrm{E}[Z_i]}{(1-\mathrm{E}[Z_i])}\\
&\quad=\frac{\mathrm{E}[Y_iZ_i]}{\mathrm{E}[Z_i]}-\frac{\mathrm{E}[Y_i]-\mathrm{E}[Y_iZ_i]}{1-\mathrm{E}[Z_i]}=\frac{\mathrm{E}[Y_iZ_i]-\mathrm{E}[Y_i]\mathrm{E}[Z_i]}{\mathrm{E}[Z_i](1-\mathrm{E}[Z_i])}=\frac{\mathrm{Cov}[Y_i,Z_i]}{\mathrm{Var}[Z_i]}.
\end{aligned}$$

Similarly, for denominator,

$$\mathrm{E}[D_i|Z_i=1]-\mathrm{E}[D_i|Z_i=0]=\frac{\mathrm{Cov}[D_i,Z_i]}{\mathrm{Var}[Z_i]}.$$

Estimation

Note that for the numerator of ρ_C,

$$\begin{aligned}
&\mathrm{E}[Y_i|Z_i=1]-\mathrm{E}[Y_i|Z_i=0]\\
&\quad=\frac{\mathrm{E}[Y_i|Z_i=1]\mathrm{E}[Z_i]}{\mathrm{E}[Z_i]}\\
&\qquad-\frac{\mathrm{E}[Y_i|Z_i=0](1-\mathrm{E}[Z_i])+\mathrm{E}[Y_i|Z_i=1]\mathrm{E}[Z_i]-\mathrm{E}[Y_i|Z_i=1]\mathrm{E}[Z_i]}{(1-\mathrm{E}[Z_i])}\\
&\quad=\frac{\mathrm{E}[Y_iZ_i]}{\mathrm{E}[Z_i]}-\frac{\mathrm{E}[Y_i]-\mathrm{E}[Y_iZ_i]}{1-\mathrm{E}[Z_i]}=\frac{\mathrm{E}[Y_iZ_i]-\mathrm{E}[Y_i]\mathrm{E}[Z_i]}{\mathrm{E}[Z_i](1-\mathrm{E}[Z_i])}=\frac{\mathrm{Cov}[Y_i,Z_i]}{\mathrm{Var}[Z_i]}.
\end{aligned}$$

Similarly, for denominator,

$$\mathrm{E}[D_i|Z_i=1]-\mathrm{E}[D_i|Z_i=0]=\frac{\mathrm{Cov}[D_i,Z_i]}{\mathrm{Var}[Z_i]}.$$

ρ_C is identical to IV estimand. Z_i is instrument, D_i endogenous regressor. Use 2SLS to estimate ρ_C and do inference.

Interpretation subtle: comparing 2SLS to OLS estimates

TABLE 4.4.1
Results from the JTPA experiment: OLS and IV estimates of training impacts

	Comparisons by Training Status (OLS)		Comparisons by Assignment Status (ITT)		Instrumental Variable Estimates (IV)	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

Notes: Authors' tabulation of JTPA study data. The table reports OLS, ITT, and IV estimates of the effect of subsidized training on earnings in the JTPA experiment. Columns 1 and 2 show differences in earnings by training status; columns 3 and 4 show differences by random-assignment status. Columns 5 and 6 report the result of using random-assignment status as an instrument for training. The covariates used for columns 2, 4, and 6 are high school or GED, black, Hispanic, married, worked less than 13 weeks in past year, AFDC (for women), plus indicators for the JTPA service strategy recommended, age group, and second follow-up survey. Robust standard errors are shown in parentheses. There are 5,102 men and 6,102 women in the sample.

Estimation

Describing the compliers:

- We know their share in the sample as well as over Z_i, but cannot spot them in the data.
- Characterize the distribution of their covariates.
- E.g., suppose a binary covariate, $X_i = 0, 1$. By Bayes Rule,

$$\Pr[X_i = 1 | D_{1i} > D_{0i}] = \frac{\Pr[D_{1i} > D_{0i} | X_i = 1] \Pr[X_i = 1]}{\Pr[D_{1i} > D_{0i}]}$$
$$= \frac{\mathrm{E}[D_i | Z_i = 1, X_i = 1] - \mathrm{E}[D_i | Z_i = 0, X_i = 1]}{\mathrm{E}[D_i | Z_i = 1] - \mathrm{E}[D_i | Z_i = 0]} \Pr[X_i = 1]$$

which consists of terms that are identified in the data.

Estimation

- More generally, Abadie (2003) shows that the marginal distribution of any $f(Y, D, X)$ for compliers is identified.

Estimation

- More generally, Abadie (2003) shows that the marginal distribution of any $f(Y,D,X)$ for compliers is identified.
- By total probability,

$$
\begin{aligned}
\mathrm{E}[f(Y,D,X)] =& \mathrm{E}[f(Y,D,X)|D_1 > D_0]\Pr[D_1 > D_0] \\
&+ \mathrm{E}[f(Y,D,X)|D_1 = D_0 = 1]\Pr[D_1 = D_0 = 1] \\
&+ \mathrm{E}[f(Y,D,X)|D_1 = D_0 = 0]\Pr[D_1 = D_0 = 0].
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \mathrm{E}[f(Y,D,X)|D_1 > D_0] =& \frac{1}{\Pr[D_1 > D_0]}\{\mathrm{E}[f(Y,D,X)] \\
&- \mathrm{E}[f(Y,D,X)|D_1 = D_0 = 1]\Pr[D_1 = D_0 = 1] \\
&- \mathrm{E}[f(Y,D,X)|D_1 = D_0 = 0]\Pr[D_1 = D_0 = 0]\} \\
=& \frac{1}{\Pr[D_1 > D_0]}\{\mathrm{E}[f(Y,D,X)] \\
&- \mathrm{E}[f(Y,D,X)|D = 1, Z = 0]\Pr[D = 1|Z = 0] \\
&- \mathrm{E}[f(Y,D,X)|D = 0, Z = 1]\Pr[D = 0|Z = 1]\},
\end{aligned}
$$

by Z independent and monotonicity.

Estimation

- More generally, Abadie (2003) shows that the marginal distribution of any $f(Y,D,X)$ for compliers is identified.
- By total probability,

$$\begin{aligned}
\mathrm{E}[f(Y,D,X)] =& \mathrm{E}[f(Y,D,X)|D_1 > D_0]\Pr[D_1 > D_0] \\
&+ \mathrm{E}[f(Y,D,X)|D_1 = D_0 = 1]\Pr[D_1 = D_0 = 1] \\
&+ \mathrm{E}[f(Y,D,X)|D_1 = D_0 = 0]\Pr[D_1 = D_0 = 0].
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \mathrm{E}[f(Y,D,X)|D_1 > D_0] =& \frac{1}{\Pr[D_1 > D_0]}\{\mathrm{E}[f(Y,D,X)] \\
&- \mathrm{E}[f(Y,D,X)|D_1 = D_0 = 1]\Pr[D_1 = D_0 = 1] \\
&- \mathrm{E}[f(Y,D,X)|D_1 = D_0 = 0]\Pr[D_1 = D_0 = 0]\} \\
=& \frac{1}{\Pr[D_1 > D_0]}\{\mathrm{E}[f(Y,D,X)] \\
&- \mathrm{E}[f(Y,D,X)|D = 1, Z = 0]\Pr[D = 1|Z = 0] \\
&- \mathrm{E}[f(Y,D,X)|D = 0, Z = 1]\Pr[D = 0|Z = 1]\},
\end{aligned}$$

by Z independent and monotonicity.

- Terms are identified.

Estimation

- To take advantage of this result, construct the "kappa" weights,

$$\kappa = 1 - \frac{D_i(1-Z_i)}{\Pr[Z_i = 0]} - \frac{(1-D_i)Z_i}{\Pr[Z_i = 1]}.$$

- In that case,

$$\mathrm{E}\left[f(Y,D,X)|D_{1i} > D_{0i}\right] = \frac{\mathrm{E}\left[\kappa f(Y,D,X)\right]}{\Pr[D_{1i} > D_{0i}]} = \frac{\mathrm{E}\left[\kappa f(Y,D,X)\right]}{\mathrm{E}\left[D_{1i} - D_{0i}\right]}.$$

Estimation

- To take advantage of this result, construct the "kappa" weights,

$$\kappa = 1 - \frac{D_i(1 - Z_i)}{\Pr[Z_i = 0]} - \frac{(1 - D_i)Z_i}{\Pr[Z_i = 1]}.$$

- In that case,

$$\mathrm{E}\left[f(Y, D, X) | D_{1i} > D_{0i}\right] = \frac{\mathrm{E}\left[\kappa f(Y, D, X)\right]}{\Pr[D_{1i} > D_{0i}]} = \frac{\mathrm{E}\left[\kappa f(Y, D, X)\right]}{\mathrm{E}\left[D_{1i} - D_{0i}\right]}.$$

- So weighting by $\kappa / \Pr[D_{1i} > D_{0i}] = \kappa / \mathrm{E}\left[D_{1i} - D_{0i}\right]$ and then computing the usual sample statistics allows you to characterize compliers' attributes.
- Opens possibility of covariate conditioning, non-linear LATE estimators (e.g., glms fit via MLE), and characterizing distribution (e.g., quantile) functions for covariates or potential outcomes.

Identification checks

Identification checks

Kitagawa (2015)

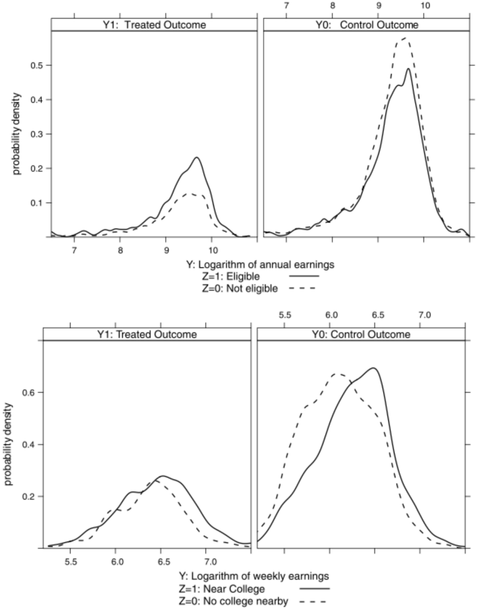
- Given LATE assns, the following must hold (Balke & Pearl 1997):

$$\underbrace{Pr(Y=y,D=1|Z=1)}_{a+c}-\underbrace{Pr(Y=y,D=1|Z=0)}_{a}=Pr(Y_1=y,D_1>D_0)\geq 0$$
$$\underbrace{Pr(Y=y,D=0|Z=0)}_{n+c}-\underbrace{Pr(Y=y,D=0|Z=1)}_{n}=Pr(Y_0=y,D_1>D_0)\geq 0.$$

- Holds more generally, e.g., for continuous Y: $P(y,D=1|Z=1)$ must nest $P(y,D=1|Z=0)$, and $P(y,D=0|Z=0)$ must nest $P(y,D=0|Z=1)$.
- Can test using Kolgomorov-Smirnov statistic.

Identification checks

Kitagawa (2015) - top good, bottom bad:



Identification checks

Huber & Mellace (2015):

- Given LATE assumptions, we have

$$\begin{aligned} E(Y|D=1,Z=1) = & \frac{\pi_a}{\pi_a+\pi_c} E(Y_1|Type=a,Z=1) \\ & + \frac{\pi_c}{\pi_a+\pi_c} E(Y_1|Type=c,Z=1) \end{aligned}$$

- $E(Y_1|Type=a,Z=1)$ must thus reside between mean to two truncated distributions:
  - $P(y|D=1,Z=1)$ truncated above, chopping off mass $1-\frac{\pi_a}{\pi_a+\pi_c}$.
  - $P(y|D=1,Z=1)$ truncated below, chopping off mass $1-\frac{\pi_a}{\pi_a+\pi_c}$.
- Now, $E(Y_1|Type=a,Z=1) = E(Y_1|Type=a,Z=0)$, observable.
- So, we can check whether $E(Y_1|Type=a,Z=0)$ lies within bounds derived from truncated distributions.
- Other restrictions are also implied by LATE assumptions.

Simultaneous Equations

Simultaneous Equations

Canonical application of IV is for supply-demand response functions:

Simultaneous Equations

Canonical application of IV is for supply-demand response functions:

- Effect of prices on quantity sold cannot be estimated from equilibrium values: typically a positive correlation from market to market, but obviously this doesn't mean that higher prices cause more sales (classic correlation $\neq$ causation).

Simultaneous Equations

Canonical application of IV is for supply-demand response functions:

- Effect of prices on quantity sold cannot be estimated from equilibrium values: typically a positive correlation from market to market, but obviously this doesn't mean that higher prices cause more sales (classic correlation $\neq$ causation).
- IV needed for exogenous variation in either demand or supply.
- Classic analysis assumes linear system of equations with homogenous effects – classical regression – even though economic theory provides no basis for these assumption.

Simultaneous Equations

Angrist et al. (2000) "agnostic" reanalysis of effect of prices on quantity sold:

Simultaneous Equations

Angrist et al. (2000) “agnostic” reanalysis of effect of prices on quantity sold:

- Markets indexed by $t = 1, ..., T$ (could be same location on different days, for example).

Simultaneous Equations

Angrist et al. (2000) “agnostic” reanalysis of effect of prices on quantity sold:

- Markets indexed by $t = 1, ..., T$ (could be same location on different days, for example).
- Demand and supply functions: $q_t^d(p,z)$ and $q_t^s(p,z)$ given price p and “supply shock” instrument $z = 0, 1$.
- These are allowed to vary freely from market to market.

Simultaneous Equations

Angrist et al. (2000) "agnostic" reanalysis of effect of prices on quantity sold:

- Markets indexed by $t = 1, \ldots, T$ (could be same location on different days, for example).
- Demand and supply functions: $q_t^d(p,z)$ and $q_t^s(p,z)$ given price p and "supply shock" instrument $z = 0, 1$.
- These are allowed to vary freely from market to market.
- Let $p^e(z) \equiv p_t^e$ be the expected market clearing (equilibrium) price such that

$$q_t^d(p^e(z), z) = q_t^s(p^e(z), z) \equiv q_t^e,$$

  where q_t^e is shorthand for the expected equilibrium quantity sold given z.
- Thus our data are (z_t, p_t^e, q_t^e).

Simultaneous Equations

Angrist et al. (2000) “agnostic” reanalysis of effect of prices on quantity sold:

- Markets indexed by $t = 1, \ldots, T$ (could be same location on different days, for example).
- Demand and supply functions: $q_t^d(p,z)$ and $q_t^s(p,z)$ given price p and “supply shock” instrument $z = 0, 1$.
- These are allowed to vary freely from market to market.
- Let $p^e(z) \equiv p_t^e$ be the expected market clearing (equilibrium) price such that

$$q_t^d(p^e(z), z) = q_t^s(p^e(z), z) \equiv q_t^e,$$

  where q_t^e is shorthand for the expected equilibrium quantity sold given z.
- Thus our data are (z_t, p_t^e, q_t^e).
- Estimate $q_t^d(p,z) - q_t^d(p',z)$, for some p, p'.

Simultaneous Equations

- Assumption 1 (random assignment): Z_t randomly assigned.
- Identifies equilibrium response functions:

$$p^e(z) = \mathrm{E}\left[p_t^e | z_t = z\right] \text{ and } q^e(z) = \mathrm{E}\left[q_t^e | z_t = z\right].$$

- Assumption 2 (exclusion): $q_t^d(p,z) = q_t^d(p,z')$ for all p,z,z',t (effect of z runs solely through p).

Simultaneous Equations

- Assumption 1 (random assignment): Z_t randomly assigned.
- Identifies equilibrium response functions:

$$p^e(z) = \mathrm{E}[p_t^e | z_t = z] \text{ and } q^e(z) = \mathrm{E}[q_t^e | z_t = z].$$

- Assumption 2 (exclusion): $q_t^d(p,z) = q_t^d(p,z')$ for all p, z, z', t (effect of z runs solely through p).
- Assumption 3 (first stage): $p^e(z)$ is a non-trivial function of z.

Simultaneous Equations

- Assumption 1 (random assignment): Z_t randomly assigned.
- Identifies equilibrium response functions:

$$p^e(z) = \mathrm{E}\left[p_t^e | z_t = z\right] \text{ and } q^e(z) = \mathrm{E}\left[q_t^e | z_t = z\right].$$

- Assumption 2 (exclusion): $q_t^d(p,z) = q_t^d(p,z')$ for all p,z,z',t (effect of z runs solely through p).
- Assumption 3 (first stage): $p^e(z)$ is a non-trivial function of z.
- Assumption 4 (monotonicity): for all (z,z'), either $p_t^e(z) \geq p_t^e(z')$ or vice versa. (Generally downward-sloping demand and upward-sloping supply.)

Simultaneous Equations

Angrist et al. 2000, Thm. 1:

- Under these assumptions define the IV estimand,

$$\beta^* = \frac{q^e(1) - q^e(0)}{p^e(1) - p^e(0)}.$$

Simultaneous Equations

Angrist et al. 2000, Thm. 1:

- Under these assumptions define the IV estimand,

$$\beta^* = \frac{q^e(1) - q^e(0)}{p^e(1) - p^e(0)}.$$

- Then,

$$\beta^* = \int_{\mathscr{P}} \mathrm{E}\left[\frac{\partial q_t^d}{\partial p}(p) \middle| p_t^e(1) \geq p \geq p_t^e(0) \right] \omega(p) dp,$$

where

$$\omega(p) = \frac{\Pr[p_t^e(0) < p < p_t^e(1)]}{\int_{\mathscr{P}} \Pr[p_t^e(0) < r < p_t^e(1)] dr}$$

- A weighted average with two levels of averaging:
  - Across markets: For any p, only markets whose equil. prices, $p_t^e(0)$ and $p_t^e(1)$, bracket p enter into the expectation. More powerful instrument, wider brackets, more representative β^*.
  - Across prices: "The weight given to any particular price is proportional to the number of markets whose equilibrium prices bracket this price."

Simultaneous Equations

- Interplay of theory, identification, and specificity.
- Theory motivated target of inference and identifying assumptions.
- Agnostic analysis allowed us to assign the proper *specificity* to our estimand.

Discussion

Discussion

Compared to last class, what's different and what's the same?

Discussion

Compared to last class, what's different and what's the same?

- We *dropped* the constant effects assumption, which is usually unrealistic.

Discussion

Compared to last class, what's different and what's the same?

- We *dropped* the constant effects assumption, which is usually unrealistic.
- We *added* a weaker monotonicity assumption.

Discussion

Compared to last class, what's different and what's the same?

- We *dropped* the constant effects assumption, which is usually unrealistic.
- We *added* a weaker monotonicity assumption.
- We *clarify* that effects are identified only for a particular subpopulation — the "complier" subpopulation.

Discussion

Compared to last class, what's different and what's the same?

- We *dropped* the constant effects assumption, which is usually unrealistic.
- We *added* a weaker monotonicity assumption.
- We *clarify* that effects are identified only for a particular subpopulation — the "complier" subpopulation.
- If constant effects happens to hold, effects for compliers are by definition same as for entire population.

Discussion

Compared to last class, what's different and what's the same?

- We *dropped* the constant effects assumption, which is usually unrealistic.
- We *added* a weaker monotonicity assumption.
- We *clarify* that effects are identified only for a particular subpopulation — the "complier" subpopulation.
- If constant effects happens to hold, effects for compliers are by definition same as for entire population.
- With effect heterogeneity, different instruments *do not* identify the same causal effect. So, "over-identification" tests are *not* valid, *nor* is the use of multiple instruments to obtain more precise estimates of a given causal effect valid.

Discussion

The LATE theorem and its extensions are hugely important. Why?

- Non-parametric solution to non-compliance problem.
- Motivates the “encouragement design” that allows for non-parametric estimation of causal effects of things that are hard to randomize directly (breast feeding, smoking).
- Provided deep insights into consequences of effect heterogeneity and helped to clarify crucial concepts:“local” ATEs, “principal strata,” comparisons that are causal versus non-causal.
- Formalized and helped to clarify one class of differences between “internal” and “external” validity.

Discussion

Extensions:

- MHE (pp. 173-188) discusses extensions to multiple treatments, covariates, and continuous endogenous treatments. Basic intuitions about "LATE" carries over into these settings.
- From LATE to ATE:
  - Point identify ATE from LATE we think covariates are sufficient to identify differences between compliers' and the general population potential outcomes (Angrist & Fernandez-Val, 2010; Aronow & Carnegie, 2013; Bisbee et al. 2017),
  - Bound the ATE when we do not think this is plausible (Balke & Pearl 1997).
- Worth reading a debate between Deaton and Heckman/Urzua on the one hand and Imbens on the other (readings in Dropbox). Some important points about relating "agnostic" estimates to structural models to finesse our interpretation of LATEs.