# Econ 2148, fall 2019
# Instrumental variables I, origins and binary treatment case

Maximilian Kasy

Department of Economics, Harvard University

# Agenda instrumental variables part I

- Origins of instrumental variables: Systems of linear structural equations
- Strong restriction: Constant causal effects.
- Modern perspective: Potential outcomes, allow for heterogeneity of causal effects
- Binary case:
  1. Keep IV estimand, reinterpret it in more general setting:
     Local Average Treatment Effect (LATE)
  2. Keep object of interest average treatment effect (ATE):
     Partial identification (Bounds)

# Agenda instrumental variables part II

- Continuous treatment case:
    1. Restricting heterogeneity in the structural equation:
       Nonparametric IV (conditional moment equalities)
    2. Restricting heterogeneity in the first stage:
       Control functions
    3. Linear IV:
       Continuous version of LATE

## Takeaways for this part of class

- Instrumental variables methods were invented jointly with the idea of economic equilibrium.
- Classic assumptions impose strong restrictions on heterogeneity: same causal effect for every unit.
- Modern formulations based on potential outcomes relax this assumption.
- With effect heterogeneity, average treatment effects are not point-identified any more.
- Two solutions:
    1. Re-interpret the classic IV-coefficient in more general setting.
    2. Derive bounds on the average treatment effect.

# Origins of IV: systems of structural equations

- econometrics pioneered by “Cowles commission” starting in the 1930s
- they were interested in demand (elasticities) for agricultural goods
- introduced systems of simultaneous equations
  - outcomes as equilibria of some structural relationships
  - goal: recover the slopes of structural relationships
  - from observations of equilibrium outcomes and exogenous shifters

# System of structural equations

$$Y = A \cdot Y + B \cdot Z + \varepsilon,$$

- $Y$: $k$-dimensional vector of equilibrium outcomes
- $Z$: $l$-dimensional vector of exogenous variables
- $A$: unknown $k \times k$ matrix of coefficients of interest
- $B$: unknown $k \times l$ matrix
- $\varepsilon$: further unobserved factors affecting outcomes

# Example: supply and demand

$$Y = (P, Q)$$
$$P = A_{12} \cdot Q + B_1 \cdot Z + \varepsilon_1 \text{ demand}$$
$$Q = A_{21} \cdot P + B_2 \cdot Z + \varepsilon_2 \text{ supply}$$

- demand function: relates prices to quantity supplied and shifters $Z$ and $\varepsilon_1$ of demand
- supply function relates quantities supplied to prices and shifters $Z$ and $\varepsilon_2$ of supply.
- does not really matter which of the equations puts prices on the “left hand side.’
- price and quantity in market equilibrium: solution of this system of equations.

# Reduced form

- solve equation $Y = A \cdot Y + B \cdot Z + \varepsilon$
  for $Y$ as a function of $Z$ and $\varepsilon$
- bring $A \cdot Y$ to the left hand side,
  pre-multiply by $(I - A)^{-1} \Rightarrow$

$$Y = C \cdot Z + \eta \text{ “reduced form”}$$
$$C := (I - A)^{-1} \cdot B \text{ reduced form coefficients}$$
$$\eta := (I - A)^{-1} \cdot \varepsilon$$

- suppose $E[\varepsilon|Z] = 0$ (ie., $Z$ is randomly assigned)
- then we can **identify** $C$ from

$$E[Y|Z] = C \cdot Z.$$

# Exclusion restrictions

- suppose we know $C$
- what we want is $A$, possibly $B$
- problem: $k \times l$ coefficients in $C = (I - A)^{-1} \cdot B$
  $k \times (k + l)$ coefficients in $A$ and $B$
- $\Rightarrow$ further assumptions needed
- exclusion restrictions: assume that some of the coefficients in $B$ or $A$ are $= 0$.
- Example: rainfall affects grain supply but not grain demand

# Supply and demand continued

- suppose $Z$ is (i) random, $E[\varepsilon|Z] = 0$
- and (ii) "excluded" from the demand equation
  $\Rightarrow B_{11} = 0$
- by construction, $\mathrm{diag}(A) = 0$
- therefore

$$\mathrm{Cov}(Z, P) = \mathrm{Cov}(Z, A_{12} \cdot Q + B_1 \cdot Z + \varepsilon_1) = A_{12} \cdot \mathrm{Cov}(Z, Q),$$

- $\Rightarrow$ the slope of demand is identified by

$$A_{12} = \frac{\mathrm{Cov}(Z, P)}{\mathrm{Cov}(Z, Q)}.$$

- $Z$ is an **instrumental variable**

# Remarks

- historically, applied researchers have not been very careful about choosing $Z$ for which (i) randomization and (ii) exclusion restriction are well justified.
- since the 1980s, more emphasis on credibility of identifying assumptions
- some additional problematic restrictions we imposed:
  1. linearity
  2. constant (non-random) slopes
  3. heterogeneity $\varepsilon$ is $k$ dimensional and enters additively
- $\Rightarrow$ causal effects assumed to be the same for everyone
- next section: framework which does not impose this

# Modern perspective: Treatment effects and potential outcomes

- coming from biostatistics / medical trials
- potential outcome framework: answer to “what if” questions
- two “treatments:” $D = 0$ or $D = 1$
- eg. placebo vs. actual treatment in a medical trial
- $Y_i$ person $i$’s outcome
  eg. survival after 2 years
- potential outcome $Y_i^0$:
  what if person $i$ would have gotten treatment 0
- potential outcome $Y_i^1$:
  what if person $i$ would have gotten treatment 1
- question to you: is this even meaningful?

- causal effect / treatment effect for person $i$ :
  $Y_i^1 - Y_i^0$.
- average causal effect / average treatment effect:

$$ATE = E[Y^1 - Y^0],$$

- expectation averages over the population of interest

# The fundamental problem of causal inference

- **we never observe both $Y^0$ and $Y^1$ at the same time**
- one of the potential outcomes is always missing from the data
- treatment $D$ determines which of the two we observe
- formally:

$$Y = D \cdot Y^1 + (1 - D) \cdot Y^0.$$

# Selection problem

- distribution of $Y^1$ among those with $D = 1$
  need not be the same as the distribution of $Y^1$ among everyone.
- in particular

$$E[Y|D=1] = E[Y^1|D=1] \neq E[Y^1]$$
$$E[Y|D=0] = E[Y^0|D=0] \neq E[Y^0]$$
$$E[Y|D=1] - E[Y|D=0] \neq E[Y^1 - Y^0] = ATE.$$

# Randomization

- no selection $\Leftrightarrow$ $D$ is random
$$(Y^0, Y^1) \perp D.$$
- in this case,
$$E[Y|D=1] = E[Y^1|D=1] = E[Y^1]$$
$$E[Y|D=0] = E[Y^0|D=0] = E[Y^0]$$
$$E[Y|D=1] - E[Y|D=0] = E[Y^1 - Y^0] = ATE.$$
- can ensure this by actually randomly assigning $D$
- independence $\Rightarrow$ comparing treatment and control actually compares “apples with apples”
- this gives **empirical content to** the “metaphysical” notion of **potential outcomes**!

# Instrumental variables

- recall: simultaneous equations models with exclusion restrictions
- $\Rightarrow$ instrumental variables

$$\beta = \frac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, D)}.$$

- we will now give a new interpretation to $\beta$
- using the potential outcomes framework, allowing for heterogeneity of treatment effects
- "Local Average Treatment Effect" (LATE)

# 6 assumptions

1. $Z \in \{0, 1\}$, $D \in \{0, 1\}$
2. $Y = D \cdot Y^1 + (1 - D) \cdot Y^0$
3. $D = Z \cdot D^1 + (1 - Z) \cdot D^0$
4. $D^1 \geq D^0$
5. $Z \perp (Y^0, Y^1, D^0, D^1)$
6. $\text{Cov}(Z, D) \neq 0$

# Discussion of assumptions

Generalization of randomized experiment

- $D$ is "partially randomized"
- instrument $Z$ is randomized
- $D$ depends on $Z$, but is not fully determined by it

1. **Binary treatment and instrument:**
   both $D$ and $Z$ can only take two values
   results generalize, but things get messier without this
2. **Potential outcome equation for $Y$:** $Y = D \cdot Y^1 + (1 - D) \cdot Y^0$
   - *exclusion restriction*: $Z$ does not show up in the equation determining the outcome.
   - *"stable unit treatment values assumption"* (SUTVA): outcomes are not affected by the treatment received by other units.
     excludes general equilibrium effects or externalities.

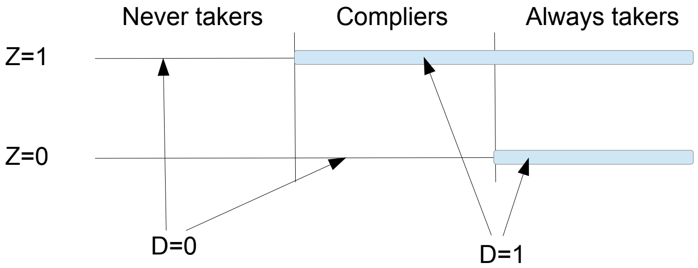
3. **Potential outcome equation for $D$:** $D = Z \cdot D^1 + (1 - Z) \cdot D^0$
   SUTVA; treatment is not affected by the instrument values of other units
4. **No defiers:** $D^1 \geq D^0$
   - four possible combinations for the potential treatments $(D^0, D^1)$ in the binary setting
   - $D^1 = 0, D^0 = 1$, is excluded
   - $\Leftrightarrow$ monotonicity

Table: No defiers

| | $D^0$ | $D^1$ |
|---|---|---|
| Never takers (NT) | 0 | 0 |
| Compliers (C) | 0 | 1 |
| Always takers (AT) | 1 | 1 |
| ~~Defiers~~ | 1 | 0 |

5. **Randomization:** $Z \perp (Y^0, Y^1, D^0, D^1)$
   - $Z$ is (as if) randomized.
   - in applications, have to justify both exclusion and randomization
   - no reverse causality, common cause!
6. **Instrument relevance:** $\text{Cov}(Z, D) \neq 0$
   - guarantees that the IV estimand is well defined
   - there are at least some compliers
   - testable
   - near-violation: weak instruments

# Graphical illustration



# Illustration explained

- 3 groups, never takers, compliers, and always takers
- by randomization of $Z$:
  each group represented equally given $Z = 0$ / $Z = 1$
- depending on group:
  observe different treatment values and potential outcomes.
- will now take the IV estimand
$$\frac{\mathrm{Cov}(Z, Y)}{\mathrm{Cov}(Z, D)}$$
- interpret it in terms of potential outcomes:
  average causal effects for the subgroup of compliers
- idea of proof:
  take the “top part” of figure 28, and subtract the “bottom part.”

# Preliminary result:

If $Z$ is binary, then

$$\frac{\mathrm{Cov}(Z,Y)}{\mathrm{Cov}(Z,D)} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]}.$$

**Practice problem**

Prove this.

# Proof

- Consider the covariance in the numerator:

$$\begin{aligned}\text{Cov}(Z, Y) &= E[YZ] - E[Y] \cdot E[Z] \\ &= E[Y|Z=1] \cdot E[Z] - (E[Y|Z=1] \cdot E[Z] + E[Y|Z=0] \cdot E[1-Z]) \cdot E[Z] \\ &= (E[Y|Z=1] - E[Y|Z=0]) \cdot E[Z] \cdot E[1-Z].\end{aligned}$$

- Similarly for the denominator:

$$Cov(Z, D) = (E[D|Z=1] - E[D|Z=0]) \cdot E[Z] \cdot E[1-Z].$$

- The $E[Z] \cdot E[1-Z]$ terms cancel when taking a ratio

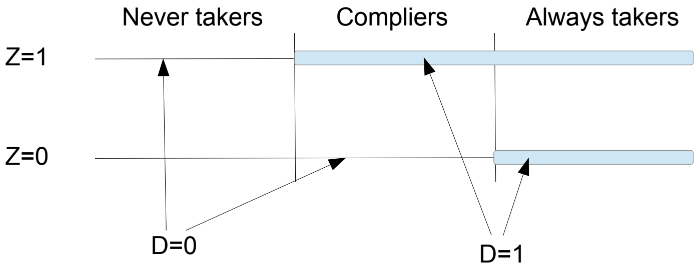
# The “LATE” result

$$\frac{E[Y|Z=1]-E[Y|Z=0]}{E[D|Z=1]-E[D|Z=0]} = E[Y^1 - Y^0|D^1 > D^0]$$

## Practice problem

Prove this.

Hint: decompose $E[Y|Z=1]-E[Y|Z=0]$ in 3 parts corresponding to our illustration


Never takers
Compliers
Always takers
Z=1
Z=0
D=0
D=1

# Proof

- "top part" of figure:

$$
\begin{aligned}
E[Y|Z=1] = {} & E[Y|Z=1, NT]\cdot P(NT|Z=1) \\
& + E[Y|Z=1, C]\cdot P(C|Z=1) \\
& + E[Y|Z=1, AT]\cdot P(AT|Z=1) \\
= {} & E[Y^0|NT]\cdot P(NT) + E[Y^1|C]\cdot P(C) + E[Y^1|AT]\cdot P(AT).
\end{aligned}
$$

  - first equation relies on the no defiers assumption
  - second equation uses the exclusion restriction and randomization assumptions.
- Similarly

$$
E[Y|Z=0] = E[Y^0|NT]\cdot P(NT) + E[Y^0|C]\cdot P(C) + E[Y^1|AT]\cdot P(AT).
$$

proof continued:

- Taking the difference, only the complier terms remain, the others drop out:

$$E[Y|Z=1]-E[Y|Z=0]=\left(E[Y^1|C]-E[Y^0|C]\right)\cdot P(C).$$

- denominator:

$$\begin{aligned}E[D|Z=1]-E[D|Z=0]=E[D^1]-E[D^0]&\\ =(P(C)+P(AT))-P(AT)=P(C).\end{aligned}$$

- taking the ratio, the claim follows. $\square$

# Recap

LATE result:

- take the **same statistical object** economists estimate a lot
- which used to be interpreted as average treatment effect
- **new interpretation** in a more general framework
- allowing for heterogeneity of treatment effects
- $\Rightarrow$ treatment effect for a subgroup

**Practice problem**

Is the LATE, $E[Y^1 - Y^0 | D^1 > D^0]$, a structural object?

# An alternative approach: Bounds

- keep the **old structural object** of interest: average treatment effect
- but analyze its identification in the more general framework with heterogeneous treatment effects
- in general: we can learn something, not everything
- $\Rightarrow$ bounds / “**partial identification**”

# Same assumptions as before

1. $Z \in \{0,1\}$, $D \in \{0,1\}$
2. $Y = D \cdot Y^1 + (1-D) \cdot Y^0$
3. $D = Z \cdot D^1 + (1-Z) \cdot D^0$
4. $D^1 \geq D^0$
5. $Z \perp (Y^0, Y^1, D^0, D^1)$
6. $\mathrm{Cov}(Z, D) \neq 0$

additionally:

7. $Y$ is bounded, $Y \in [0,1]$

# Decomposing ATE in known and unknown components

- decompose $E[Y^1]$:

$$E[Y^1] = E[Y^1|NT] \cdot P(NT) + E[Y^1|C \vee AT] \cdot P(C \vee AT).$$

- terms that are identified:

$$\begin{aligned} E[Y^1|C \vee AT] &= E[Y|Z=1, D=1] \\ P(C \vee AT) &= E[D|Z=1] \\ P(NT) &= E[1-D|Z=1] \end{aligned}$$

and thus

$$E[Y^1|C \vee AT] \cdot P(C \vee AT) = E[YD|Z=1].$$

- Data tell us nothing about $E[Y^1|NT]$.
  $Y^1$ is never observed for never takers.
- but we know, since $Y$ is bounded, that

$$E[Y^1|NT] \in [0,1]$$

- Combining these pieces, get upper and lower bounds on $E[Y^1]$:

$$\begin{aligned} E[Y^1] \in [&E[YD|Z=1], \\ &E[YD|Z=1] + E[1-D|Z=1]]. \end{aligned}$$

- For $Y^0$, similarly

$$E[Y^0] \in [E[Y(1-D)|Z=0], \\ E[Y(1-D)|Z=0] + E[D|Z=0]].$$

- Data are uninformative about $E[Y^0|AT]$.

**Practice problem**

Show this.

# Combining to get bounds on ATE

- lower bound for $E[Y^1]$, upper bound for $E[Y^0]$ $\Rightarrow$ lower bound on $E[Y^1 - Y^0]$

$$E[Y^1 - Y^0] \geq E[YD|Z = 1] - E[Y(1 - D)|Z = 0] - E[D|Z = 0]$$

- upper bound for $E[Y^1]$, lower bound for $E[Y^0]$
$\Rightarrow$ upper bound on $E[Y^1 - Y^0]$

$$E[Y^1 - Y^0] \leq E[YD|Z = 1] - E[Y(1 - D)|Z = 0] + E[1 - D|Z = 1]$$

# Between randomized experiments and nothing

- bounds on ATE:

$$E[Y^1 - Y^0] \in [E[YD|Z=1] - E[Y(1-D)|Z=0] - E[D|Z=0],\\ E[YD|Z=1] - E[Y(1-D)|Z=0] + E[1-D|Z=1]].$$

- length of this interval:

$$E[1-D|Z=1] + E[D|Z=0] = P(NT) + P(AT) = 1 - P(C)$$

- Share of compliers $\to 1$
  - interval ("identified set") shrinks to a point
  - In the limit, $D = Z$
  - thus $(Y^1, Y^0) \perp D$ – randomized experiment
- Share of compliers $\to 0$
  - length of the interval goes to 1
  - In the limit the identified set is the same as without instrument

# References

- Local average treatment effect:

*Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of causal effects using instrumental variables.* Journal of the American Statistical Association, *91(434):444–455.*

- Bounds on the average treatment effect:

*Manski, C. (2003).* Partial identification of probability distributions. *Springer Verlag, chapter 2 and 7.*