Labor Economics, Week 1
Causality, Identification, Instrumental variables

Maximilian Kasy

Department of Economics, Oxford University

Takeaways for week 1 and 2

1. fundamental notions of causal inference:
   - causality
   - identification
2. identification approaches:
   - randomized experiments
   - instrumental variables
   - conditional independence
   - difference in differences
   - regression discontinuity
3. analog estimators

Roadmap

Basic concepts

Treatment effects

Instrumental variables

Causal objects in economics

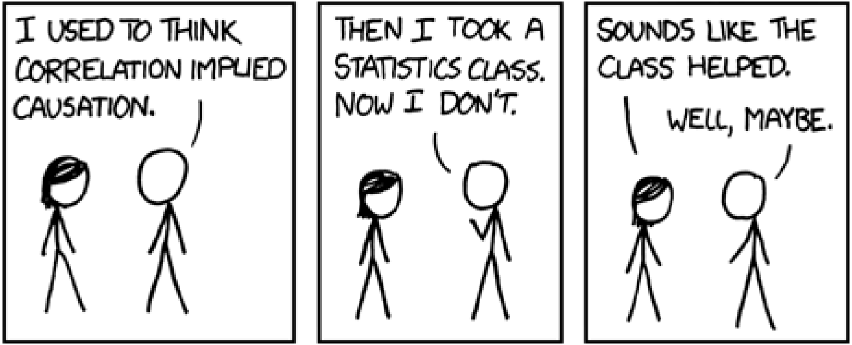
- Returns to schooling
- Elasticity of the tax base with respect to tax rates
- Effect of minimum wage on employment
- Effect of deworming pills on school attendance
- Price elasticity of demand for gasoline
- ...

Correlation and causality

Do observable distributions tell us something about causality?

- College graduates earn $x\%$ more than high school graduates
- Countries with higher GDP have higher tax rates on average
- Minimum wage levels seem uncorrelated with unemployment levels across time and space
- Gasoline consumption and gasoline price are negatively correlated over time
- ...

Figure: correlation and causation



Causality

Practice problem

How would you define causality?

“Pure” statistics

- causality is meaningless
- observations only tell us about **correlations**
- more generally, joint distributions
- disclaimer: few statisticians today would say this!

Sciences

- Galileo Galilei: one of the first to follow experimental ideal
- full **control of experimental circumstances.**
- do the same thing
  $\Rightarrow$ same thing happens to the outcomes you measure
- variation in experimental circumstances
  $\Rightarrow$ difference in observed outcomes $\approx$ causal effect
- example:
  - dropping a ball from different floors of the tower of Pisa
  - different time till the ball hits the ground
- crucial component:
  external intervention (“exogenous variation”)
  $\Rightarrow$ allows to interpret correlation as causation

Social and biological sciences

- economics is not physics
- this version of the experimental ideal is not very useful
- reason: many unobserved, and unknown, factors which we cannot hope to control
- $\Rightarrow$ **can never replicate** experimental circumstances fully
- there is "**unobserved heterogeneity**."

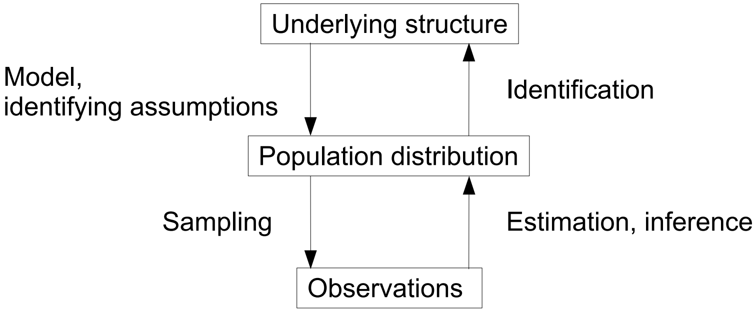
Social and biological sciences ctd

- not all is lost, however
- can still hope create experimental circumstances which are the same **on average**
- this is the idea of a **randomized experiment**!
- randomly pick treatment and control groups
  $\Rightarrow$ they are identical on average.
- “compare apples with apples.”
- many settings of interest in economics:
  not possible to run experiments
- but: definition of causality is intimately tied to randomized experiment, hypothetical or actual

Recap

- Framework of classic probability theory:
  - Does not allow to talk about causality,
  - only joint distributions.
- Causality in the sciences:
  - Additional concept:
  - External intervention / exogeneous variation
  - $\Rightarrow$ experiments.
- Causality in econometrics, biostatistics,...:
  - Additional concept:
  - Unobserved heterogeneity.
  - $\Rightarrow$ randomized experiments
  - (or “quasi-experiments”)

Identification vs inference



Identification vs inference

- goal of econometrics:
  learning interesting things (hopefully)
  about economic phenomena from observations.
- two separate components of econometrics:
  1. identification
  2. estimation and inference

Estimation and Inference

1. learning about a population distribution
2. from a finite number of observations.
3. examples:
   - estimate a difference in expectations using a difference in means
   - perform inference using a t-test.

Identification

1. learning about underlying structures, causal mechanisms
2. from a population distribution.
3. example:
   identify a causal effect
   by a difference in expectations
   if we have a randomized experiment.

- identification inverts the mapping
- from underlying structures to a population distribution
- implied by a model and identifying assumptions.

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

$$
\begin{aligned}
E[Y|D=1] = E[Y^1|D=1] &\neq E[Y^1] \\
E[Y|D=0] = E[Y^0|D=0] &\neq E[Y^0] \\
E[Y|D=1] - E[Y|D=0] &\neq E[Y^1 - Y^0] = ATE.
\end{aligned}
$$

Randomization

- no selection $\Leftrightarrow$ D is random
$$(Y^0, Y^1) \perp D.$$
- in this case,
$$\begin{aligned} E[Y|D=1] &= E[Y^1|D=1] = E[Y^1] \\ E[Y|D=0] &= E[Y^0|D=0] = E[Y^0] \\ E[Y|D=1] - E[Y|D=0] &= E[Y^1 - Y^0] = ATE. \end{aligned}$$
- 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**!

Empirical example

- **Bertrand, M. and Mullainathan, S. (2004).** Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review*, 94(4):991–1013.
- Randomly assign names which are statistically “white” or “black” to resumes which are sent out as job applications.
- Estimate causal effect on likelihood of getting invited to a job interview.

TABLE 4—AVERAGE CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES AND RESUME QUALITY

	Low	High	Ratio	Difference (*p*-value)
	Panel A: Subjective Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (*p*-value)
White names	8.50	10.79	1.27	2.29
	[1,212]	[1,223]		(0.0557)
African-American names	6.19	6.70	1.08	0.51
	[1,212]	[1,223]		(0.6084)
	Panel B: Predicted Measure of Quality (Percent Callback)			
	Low	High	Ratio	Difference (*p*- value)
White names	7.18	13.60	1.89	6.42
	[822]	[816]		(0.0000)
African-American names	5.37	8.60	1.60	3.23
	[819]	[814]		(0.0104)

Notes: Panel A reports the mean callback percents for applicant with a White name (row 1) and African-American name (row 2) depending on whether the resume was subjectively qualified as a lower quality or higher quality. In brackets is the number of resumes sent for each race/quality group. The last column reports the *p*-value of a test of proportion testing the null hypothesis that the callback rates are equal across quality groups within each racial group. For Panel B, we use a third of the sample to estimate a probit regression of the callback dummy on the set of resume characteristics as displayed in Table 3. We further control for a sex dummy, a city dummy, six occupation dummies, and a vector of dummy variables for job requirements as listed in the employment ad (see Section III, subsection D, for details). We then use the estimated coefficients on the set of resume characteristics to estimate a predicted callback for the remaining resumes (two-thirds of the sample). We call "high-quality" resumes the resumes that rank above the median predicted callback and "low-quality" resumes the resumes that rank below the median predicted callback. In brackets is the number of resumes sent for each race/quality group. The last column reports the *p*-value of a test of proportion testing

Instrumental variables

- Instrument Z, treatment D, outcome Y.
- Three numerically equivalent estimands:
  1. The slope
  $$\mathrm{Cov}(Z, Y)/\,\mathrm{Cov}(Z, D).$$
  2. The two-stage least squares slope from the regression
  $$Y = \alpha_0 + \alpha_1 \widehat{D} + \tilde{U},$$
  where $\widehat{D}$ is the first stage predicted value $\widehat{D} = \gamma_0 + \gamma_1 Z$.
  3. The slope of the regression with control
  $$Y = \delta_0 + \delta_1 D + \delta_2 V + W,$$
  where the control function V is given by the first stage residual, $V = D - \gamma_0 - \gamma_1 Z$.

Empirical example

- **Aizer, A. and Doyle, J. J. (2015).** Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, 130(2):759–803.
- Judges (within days, courts) randomly assigned to cases.
- Use judge-specific incarceration rate as instrument for juvenile incarceration.
- Finding: Juvenile incarceration reduces high school graduation, increases adult crime.

TABLE III

FIRST STAGE

	(1)	(2)	(3)
Dependent variable: juvenile incarcerations		OLS	
First judge's leave-out mean incarceration rate among first cases	1.103	1.082	1.060
	(0.102)	(0.095)	(0.097)
Demographic controls	No	Yes	Yes
Court controls	No	No	Yes
Observations	37,692		
Mean of dependent variable	0.227		

Notes. This table reports the first-stage relationship between juvenile incarceration and the instrument: the judge's incarceration rate using the linked Chicago Public School–Juvenile Court of Cook County data including cases from 1990–2000 as described in the text. All models include community × weapons-offense × year-of-offense fixed effects. Demographic controls include indicators for four age-at-offense categories, four race/ethnicity categories, sex, special education status, and the 2000 U.S. census tract family poverty rate. Court controls include nine offense categories, indictors for seven risk-assessment index categories, and whether the first judge assigned was missing. Standard errors are reported in the parentheses and are clustered at the community level.

TABLE IV

JUVENILE INCARCERATION AND HIGH SCHOOL GRADUATION

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable: graduated high school						
	Full CPS sample			Juvenile court sample			
	OLS	OLS	Inverse propensity score weighting	OLS	OLS	2SLS	2SLS
Juvenile incarceration	–0.389	–0.292	–0.391	–0.088	–0.073	–0.108	–0.125
	(0.0066)	(0.0065)	(0.0055)	(0.0043)	(0.0041)	(0.044)	(0.043)
Demographic controls	No	Yes	Yes	No	Yes	No	Yes
Court controls	N/A	N/A	N/A	No	Yes	No	Yes
Observations	440,797	440,797	420,033	37,692			
Mean of dependent variable	0.428	0.428	0.433	0.099			

Notes. This table reports the relationship between juvenile incarceration and graduation from Chicago Public Schools. Columns (1)–(3) include all students in Chicago Public Schools in eighth grade during 1990–2006 and at least age 25 by 2008. Columns (1) and (2) include community fixed effects, while column (2) also includes indicators for race, sex, special education status, each year of birth, and the 2000 U.S. census tract family poverty rate. Column (3) used the same controls and community indicators to calculate the propensity score using a probit model, estimated on a subsample where probit estimation is possible (where there is variation in juvenile incarceration within cells). Columns (4)–(7) use the linked Chicago Public School–Juvenile Court of Cook County data including cases from 1990–2000 as described in the text. These models include community × weapons-offense × year-of-offense fixed effects. Demographic controls include those listed for column (2). Court controls include nine offense categories, indictors for seven risk-assessment index categories, and whether the first judge assigned was missing. Standard errors are reported in the parentheses and are clustered at the community level. The propensity score standard errors were calculated using 200 bootstrap replications.

TABLE V
JUVENILE INCARCERATION AND ADULT CRIME

	Dependent variable: entered adult prison by age 25						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full CPS sample			Juvenile court sample			
	OLS	OLS	Inverse propensity score weighting	OLS	OLS	2SLS	2SLS
Juvenile incarceration	0.407	0.350	0.219	0.200	0.155	0.260	0.234
	(0.0082)	(0.0064)	(0.013)	(0.0072)	(0.0073)	(0.073)	(0.076)
Demographic controls	No	Yes	Yes	No	Yes	No	Yes
Court controls	N/A	N/A	N/A	No	Yes	No	Yes
Observations	440797	440797	420033	37692			
Mean of dependent variable	0.067	0.067	0.057	0.327			

Notes. This table reports the relationship between juvenile incarceration and imprisonment in an adult facility by the age of 25. Columns (1)–(3) include all students in Chicago Public Schools in eighth grade during 1990–2006 and at least age 25 by 2008. Columns (1) and (2) include community fixed effects, while column (2) also includes indicators for race, sex, special education status, each year of birth, and the 2000 U.S. census tract family poverty rate. Column (3) used the same controls and community indicators to calculate the propensity score using a probit model, estimated on a subsample where probit estimation is possible (where there is variation in juvenile incarceration within cells). Columns (4)–(7) use the linked Chicago Public School–Juvenile Court of Cook County–Illinois Department of Corrections data including juvenile cases from 1990–2000 as described in the text. These models include community × weapons-offense × year-of-offense fixed effects. Demographic controls include those listed for column (2). Court controls include nine offense categories, indictors for seven risk-assessment index categories, and whether the first judge assigned was missing. Standard errors are reported in the parentheses and are clustered at the community level. The propensity score standard errors were calculated using 200 bootstrap replications.

Instrumental variables

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

6 assumptions

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.

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

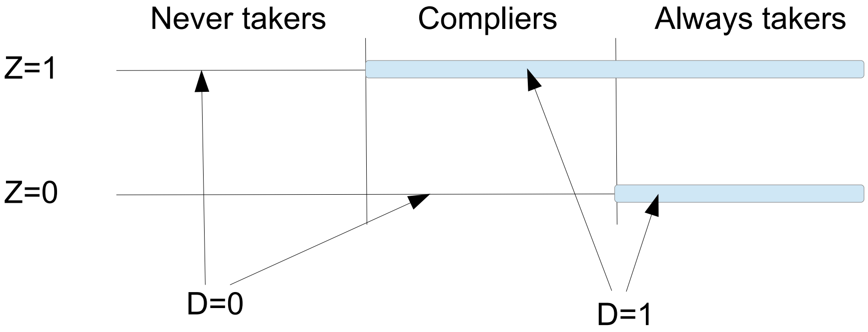


Illustration explained

- 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 40, 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

Try to 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

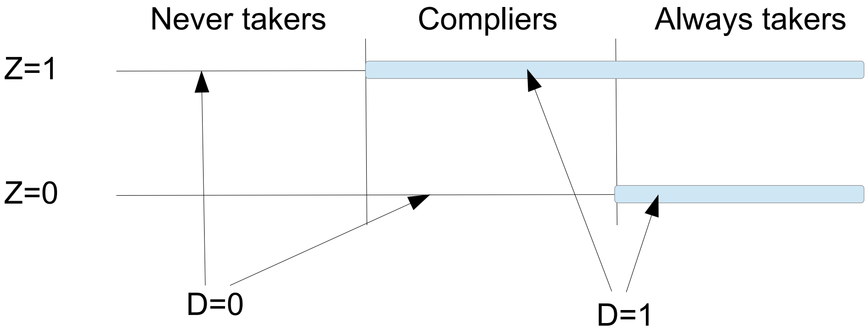
The "LATE" result

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

Practice problem

Try to 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:

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

- 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

Empirical example

- **Angrist, J.D. and Krueger, A.B. (1991).** Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014.
- compare individuals born in different quarters of the year
- school start age and structure compulsory schooling laws
- $\Rightarrow$ people born late in the year have to stay in school longer
- quarter of birth as an instrument for educational attainment in estimates of returns to schooling
- estimates effect for those affected by compulsory schooling laws

Practice problem

Who do you think are the compliers for the quarter of birth instrument?

TABLE III

PANEL A: WALD ESTIMATES FOR 1970 CENSUS—MEN BORN 1920–1929[a]

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.1484	5.1574	−0.00898
			(0.00301)
Education	11.3996	11.5252	−0.1256
			(0.0155)
Wald est. of return to education			$\boxed{0.0715}$
			$\boxed{(0.0219)}$
OLS return to education[b]			0.0801
			(0.0004)

Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.8916	5.9027	−0.01110 (0.00274)
Education	12.6881	12.7969	−0.1088 (0.0132)
Wald est. of return to education			$\boxed{0.1020 \ (0.0239)}$
OLS return to education			0.0709 (0.0003)

Supplementary readings

- Applied microeconomics perspective:
  Angrist, J. D. and Pischke, J. S. (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton Univ Press.
- Textbook focusing on binary treatments, experiments and conditional independence:
  Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- Principled treatment of (partial) identification:
  Manski, C. (2003). *Partial identification of probability distributions*. Springer Verlag.
- Theoretical computer scientist on the notion of causality:
  Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.