Instrumental variables: IV, LATE, and Examples

Marianne Bitler UC Davis & NBER

Good references

- Intuition: Angrist, J. and Pischke, S. Mostly Harmless Econometrics: An Empiricist's Companion. 2009.

- Basics: SI 2007 Methods Lectures: What's New In Econometrics?

- Weak IV: Andrews I, Stock J, Sun L. Weak Instruments in IV Regression: Theory and Practice. Annual Review of Economics. 2019;11 :727-753.
  SI 2018 Methods Lectures: Weak IV

Lots of jargon, please stop me if you don't know what I mean

- Internal validity: Results like a RCT would produce
- External validity: Results extend to other settings, time periods, groups
- Omitted variable bias: More below
- Sample selection bias: More below
- Simultaneous causality bias

- Errors in variables/measurement error: Mostly focused on error in the RHS/independent variable

- Wrong functional form: Form of regression doesn't match the data generating process, examples using OLS for a 0-1 variable, or a count

Potential outcomes model, and experiments

Y_i is the outcome.
$D_i \in \{0, 1\}$ is treatment.
Potential outcomes Y_{0i} is outcome when $D_i = 0$, Y_{1i} is outcome when $D_i = 1$.
i's treatment effect is $Y_{1i} - Y_{0i}$.
$Y_i = Y_{0i} + (Y_{1i} - Y_{0i}) \cdot D_i$.

Missing data problem, never see a person in both states. Y can be health and D can be health insurance.

Note there is an i in the treatment effect, so this allows it to vary in the population.

What can we recover with observational data?
We can estimate the mean difference, $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$.
This is equal to: (1)$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$.

The first part–$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$–is the Average Treatment Effect on the Treated, ATET. We can see an estimate for $E[Y_{1i}|D_i = 1]$, it is the mean in the treated group. We can't see the other half of this expression.

The second part–$E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$–is selection bias. It is the difference in the untreated outcome for those treated and the untreated outcome for those who don't get treated. The second term $E[Y_{0i}]|D_i = 0]$ we can get an estimate for in the data, as the mean outcome for the untreated. If the treatment is having health insurance and the outcome is health, we think health for those who can't afford group health insurance might be worse than for those who can afford group insurance. So, the selection bias term would be negative, and even if having health insurance is good for you, the real effect might be offset.

Why does randomization help with this? If we randomize D, then $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$, and the second difference in (1) goes away, and we can replace $E[Y_{0i}|D_i = 1]$ with $E[Y_{0i}|D_i = 1]$. So, we can estimate the ATET in equation 1.

So, (1) simplifies to $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0_i}]$, as the people treated are randomly drawn from the population.

Constant treatment effects
Suppose there is a constant treatment effect, ρ.

Then $Y_i = \alpha + \rho \cdot D_i + \eta_i$.

Then $\alpha = E[Y_{0i}]$, $\rho = Y_{1i} - Y_{0i}$, and $\eta_i = Y_{0i} - E[Y_{0i}]$.

So (2)$E[Y_i|D_i = 1] = \alpha + \rho + E[\eta_i|D_i = 1]$, and
(3)$E[Y_i|D_i = 0] = \alpha + E[\eta_i|D_i = 0]$

The difference (2) $-$ (3) $= \rho + E[\eta_i|D_i = 1] - E[\eta_i|D_i = 0]$.
ρ is the treatment effect and the second term is selection bias, which is caused by a correlation between η_i (the error term) and D.

If η_i is uncorrelated (or mean independent) of D, then conditional mean independence holds, and the OLS estimate is consistent for β.

Can add exogenous Xs to the regression. If there is an experiment, they should not affect the estimate of the ATE, but improve precision if the Xs are correlated with Y_i (explain Y_i).

Problems where conditional independence of the error and the key independent variable D does not hold, and you get bias.

- Omitted variable bias (more below): Some other factor is associated with both D and Y and not controlled for. Many come to mind for the example: Income, employment, etc.

- Simultaneous causation (reverse causation). Past bad health causes me to sign up for health insurance. Or supply and demand with a shock to one.

- Not today but IV for selection can help: Sample selection: Some types of people are systematically missing based on their value for Y or for η_i, the error term. Suppose you only looked at people who are alive at age 80 to understand how health insurance at age 45 mattered for health. Least healthy people are not in the data.

- Not today but IV helps with classical measurement error: Wrong functional form/errors in variables.

Signing OVB Formula for magnitude of OVB An omitted variable is left out which is associated with X and Y. Then it is in the error term, η, if I leave it out of the regression.

$y_i = \beta X_i + \eta_i$.
Suppose that CORR$(X_i, \eta_i) = \rho_{X\eta} \neq 0$ and that the other regression assumptions hold. Then
$\widehat{\beta} \rightarrow^p \beta + \rho_{X\eta} \cdot \sigma_\eta / \sigma_X$. (Remember σ_η / σ_X) has to be positive.)
So, the sign of the bias depends on the sign of $\rho_{X\eta}$, and the magnitude of the bias on how correlated they are.

There are 4 key facts about OVB.

1) It is not a function of sample size, and won't go away with larger samples.

2) The magnitude of the bias is a function of $|\rho_{X\eta}|$. That is, the more highly correlated the potential omitted variable is with X, the larger (in magnitude) the bias.

3) The direction of the bias depends on the sign of $\rho_{X\eta}$.

4) It matters how the other covariates are correlated with η and X_i if there are other covariates. it is not a simple story then.

5) It is always worth thinking through which way the bias would go.

Examples

- Example 1: Education and wages. Skill is omitted. Schooling and skill are positively correlated. So the bias term is positive ($\widehat{\beta}$ is biased up). Since we expect the true β to be positive, this means we are biased away from zero.

- Example 2: Private HI and health. Income and resources are omitted, and are positively correlated with Health. So the bias term is positive ($\widehat{\beta}$ is biased up). Since we expect the true β to be positive, this means we are biased away from zero.

Simultaneous causation

There is a causal effect of Y on X,

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \mu_i$$

and there is a causal effect of X on Y,

$$X_i = \gamma_0 + \gamma_1 \cdot Y_i + \epsilon_i.$$

A large value of μ_i means a large Y_i which implies a large X_i (if $gamma_1 > 0$). So, then $CORR(X_i, \mu_i) \neq 0$, and $\widehat{\beta}_1$ is biased and inconsistent.

- Example 1: Health and having a job with private health insurance are simultaneously determined. People with poor health can't work and don't have private HI. But those with private HI have better health. So, HI and μ_i are correlated.

IV can be a solution to both OVB and simultaneous causality

- Instrumental variables is a way of finding a variable—the instrument, Z—that only affects the outcome Y through the endogenous (correlated with the error term) variable D through D, and affects D.

- Simplest example, an RCT where the variable Z is an random offer of treatment and D is a potentially endogenous measure of taking the offer.

- Concrete example: Oregon health experiment. D is Medicaid. Z is getting an offer of Medicaid. Y is a host of things.

- Concern, D is correlated with the error term in the equation $Y_i = \beta_0 + \beta_1 \cdot D_i + \eta_i$.

- But Z is random so, $E[\eta_i | Z_i] = 0$. So, the part of D that is associated with Z can be taken out and used to look at how D affects Y, and the bad part of D_i that is correlated with η_i is taken out.

- If the causal effect of D is constant, this is the effect for everyone. Else, it is the effect for folks who Z changes D for (compliers).

Basic Linear IV

- Y, D endogenous, $X_2 - X_k$ exogenous
- D and error term ϵ are correlated
- Instrument(s) Z

– Relevance: Check first stage (regress D on other Xs, Z) easy to do, make sure Z affects Z.

– Excludability: Z only affects Y through D, can only argue for–key piece of paper.

- Example $Y_{si} = f_i(s) = \alpha + \rho \cdot S_i + \eta$, earnings as a function of schooling.
  $\eta = A_i'\gamma + \nu$. A is ability. The only reason η and S are correlated is because of ability ($E[S_i \cdot \nu] = 0$).
  Selection on observables (ability), if had it would get consistent estimate of ρ from OLS.
  But we don't see it.

- Z_i an instrument. The Wald estimate
  $\rho = CORR(Y,Z)/CORR(S,Z) = \frac{Cov(Y,Z)/Var(Z)}{Cov(S,Z)/Var(Z)} = \frac{Cov(Y,Z)}{Cov(X,Z)}$.

- This is the ratio of the first stage (coefficient on Z when regress D on Z (and other Xs)) to the reduced form (coefficient on Z when regress Y on Z (and other Xs). It is causal if and only if 2 things hold. It is relevant and it is excludable. Excludability has 2 parts. 1) The potential outcomes given Xs are independent of the IV. $(Y_0, Y_1) \perp Z|X$. 2) The IV only affects Y via the first stage.
  Relevance means Z affects X.

- Instruments
  - Experiments
  - Program rules (especially when eligibility is a function of age)
  - Simulated eligibilty: Net out demographics and focus on program rules. Take a national sample of people for a set time period (ideally before the time period of the policy change). Run them through each state's rules for eligibility for each year. This provides a simulated eligiblity that only contains the policy effects.
  - True randomness (birth timing, gender mix before infertility treatment without stopping rules, twinning before infertility treatment).

- Shift-share or Bartik, recent Goldsmith/Pinkham, Sorkin, and Swift paper. Is equivalent to using as IVs the baseline shares, the national growth rates only affect the relevance (first stage). Several other recent papers. Often used in immigration studies.

- **Added at end** Judge or adjudicator designs, where the instrument for my ruling is the average ruling by my judge. Need random assignment of people to judges or adjudicators.

Classics: Vietnam draft lottery (Angrist); Compulsory Schooling (Angrist and Krueger)

Angrist and Krueger 1991

- Look at effects of compulsory schooling on men born 1930-1950. You start school in the year you turn 6, and you can leave when you turn 16 in many states. Thus, if you are younger (born later in the year), you end up starting school later, and can drop out with fewer years of school.

- Leads to IV quarter of birth (QOB). Most places you need to be age 6 by Q4, so the Q1 kids start school later. 12/31 cutoff, 4Q start near when you are 6, Q1 start when you are 6.5. Need to stay til you are 16. The Q4 births get more years of schooling before they can drop out.

- Z_i is QOB dummies, or Q1 versus the rest. $\tilde{Z}_i$ is residual after take out some Xs. $\rho = cov(Y, \tilde{Z})/cov(S, \tilde{Z})$. ILS.

- 2SLS $\hat{S} = X'\hat{\pi}_{10} + \hat{\pi}_{11}Z$. Plug $\hat{S}$ into regression with Y, coefficient is 2SLS.

- Wald AK Q1 vs. Q4: $\Delta Y = -.0135, \Delta S = -.151$, Wald ILS: IV = 0.089. OLS = .070.

- Can't verify excludability, but can see Z and X uncorrelated (actually are not, Hungerman and Buckles show that the mothers of first quarter births are different in Xs), Z and pre-experiment outcomes are uncorrelated.

- Identifies the effect of leaving school around 10th grade (dropping out) with a little more or less schooling, not informative higher up the education distribution.

- Powerful to see raw data supporting identifying variation without lots of Xs.

Example, RD where age is relevant for eligibility: Card, Dobkin, & Maestas, 2008 (see other subset of slides)

- Eligibilty for Medicare kicks in at age 65
- Regression Discontinuity: Compare those just over 65 to those just under
- If no other changes there, then being at or over 65 is an instrument. Control for age flexibly, then thought experiment is any change at 65 is about Medicare eligibility. First show no other changes.

- See following slides, increase in coverage. No other changes (not here). Changes in care. Changes in hospital admissions. Some urgent, some not.

- Other stuff about this.

Oregon Health Insurance Experiment

- Y is ED use, in Science 2014, D is Medicaid takeup. Z is the lottery.
- First stage: Lottery winners had 0.246 increase in the probability of being on Medicaid.
- Reduced form: Lottery winners were 1.7 percentage points more likely to use the ED.
- IV estimate: $0.017/0.246 = 0.07$. Seven percentage point increase in using the ED with Medicaid.

Example: My own work on cash-out of the SSI program

- SSI is a cash transfer program for low-income disabled, blind and elderly persons.

- As of 5/2019, every other state made SSI recipients eligible for SNAP (Food Stamps), but California did not. This changed 6/2019, and I am part of a research team trying to evaluate this change.

- Instrument: Counties run SNAP in California. Use county outreach plans interacted with new policy as possible instrument for participation.

- Relevance-works in first stage.

- Exclusion-hard to see how outreach about this would affect anything except through the SNAP program.

Summary

Concerns about correlations not being causation. IV would enable getting around this. Need an instrument that drives D (relevant) but doesn't directly affect Y (exclusion). Now what if effects vary by person?

Local Average Treatment Effect (LATE)

- Series of papers Angrist, Imbens, Rubin; Imbens and Angrist.

- Example, draft lottery is z, treatment d is veteran status, Y is earnings.

  Next few slides I did not go through but will leave here for completeness

- Potential outcomes $Y_i(d, z)$, where $Z_i = z$ is the instrument, $D_i = d$ is the treatment. The causal effect of being treated given one's draft lottery status is $Y_i(1, Z_i) - Y_i(0, Z_i)$.

- Causal effect of getting a low draft number given one's veteran status is $Y_i(D_i, 1) - Y_i(D_i, 0)$.

- D_{1i} is i's treatment if $Z = 1$, and D_{0i} is i's treatment if $Z = 0$.

- Potential outcomes for treatment too. $D_i = D_{0i} + (D_{1i} - D_{0i})Z_i = \pi_0 + \pi_{1i}Z_i + \epsilon_i$. $\pi_0 = E[D_{0i}]$. $\pi_{1i} = D_{1i} - D_{0i}$.

- Average causal effect of z_i on d_i is $E[\pi_{1i}]$.
  π_{1i} = heterogeneous effects because there is an i..

- A1: Independence assumption. The IV is as good as random. The instrument is independent of potential . outcomes AND potential treatment assignment.
  $\{Y_i(d,z); \forall d,z\}, D_{1i}, D_{0i} \perp z_i$.
  This means regressing Y_i on Z_i yields $E[Y_i(D_{1i},1) - Y_i(D_{0i},0)]$, and the reduced form captures the causal effect of the IV on Y, and the first stage is causal for the effects of Z_i on D_i.

- A2: Exclusion restriction. $Y_i(d,0) = Y_i(d,1)\ \forall d \in 0,1$. Linear constant effects model this is $E[Z_i \nu_i] = 0$. Fails in draft lottery if low number guys are more likely to do something else (like schooling). Not enough that IV is random.

- Exclusion means, we can define potential outcomes without regard to value of Z. $Y_{1i} \equiv Y_i(1,1) = Y_i(1,0)$. $Y_{0i} \equiv Y_i(0,1) = Y_i(0,0)$.

- Obs. variable Y_i can be written $Y_i = Y_i(0,z) + [Y_i(1,z) - Y_i(0,z)] \cdot D_i = Y_{0i} + (Y_{1i} - Y_{0i}) \cdot D_i$

- Random coefficient $Y_i = \alpha_0 + \rho_i \cdot D_i + \nu_i$.

- A3: First stage. Need $E[D_{1i} - D_{0i}] \neq 0$.

- A4: **Monotonicity - the new assumption to LATE**x: Assume either $\pi_{1i} \geq \pi_{0i}$ or $\pi_{1i} \leq \pi_{0i}$ for all i. Imbens and Angrist.

- A4 implies $D_{1i} \geq D_{0i}$ for all i or $D_{1i} \leq D_{0i}$ for all i.

- A4 in draft lottery. Some might not have had their probability changed by draft lottery—always and never takers, but there is no one who would have served if they got a high draft number who failed to serve if they got a low one (no defiers).

- Then, IV gives a weighted estimate of the individual causal effect. (Else, would be average of that minus average for defiers.)

- 4 types: Compliers ($D = 1$ if $Z = 1, D = 0$ if $Z = 0$), defiers ($D = 0$ if $Z = 1, D = 1$ if $Z = 0$), always takers ($D = 1$ if $Z = 1, D = 1$ if $Z = 0$) and never takers ($D = 0$ if $Z = 1, D = 0$ if $Z = 0$).
  LATE is mix of complier and defier effects unless assume no defiers.

- $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[(Y_{1i} - Y_{0i}) \cdot (D_{1i} - D_{0i})] = E[Y_{1i} - Y_{0i}|D_{1i} \geq D_{0i}] \cdot P(D_{1i} \geq D_{0i}) - E[Y_{1i} - Y_{0i}|D_{1i} \leq D_{0i}] \cdot P(D_{1i} \leq D_{0i})$. Rule out 1 with defiers being gone. Else might have all positive effects, some cancel each other out.

Back into part I discussed

- Note with constant effects defiers don't matter because $E[Y_{1i} - Y_{0i}|D_{1i} \geq D_{0i}] \cdot P(D_{1i} \geq D_{0i}) - E[Y_{1i} - Y_{0i}|D_{1i} \leq D_{0i}] \cdot P(D_{1i} \leq D_{0i}) = \rho \cdot (P(D_{1i} \geq D_{0i}) - P(D_{1i} \leq D_{0i}))$.

- Key additional assumption: No defiers (monotonicity), some object this is very restrictive. There are tests for this.

- Get internal but not external validity.

- If there are no never takers, then the LATE is the same as the ATE on the untreated. If there are no always takers, then the LATE is the same as the ATE on the treated. Can arrange in some RCTs.

- Example, twins instrument for having a twin on the second birth. There is no one who has a twin on the second birth who doesn't have 3 kids. Y_{0i} is mom's earnings if 2 kids, and

Y_{1i} is mom's earnings if 3 or more. (Not 100% true, some twins die.)

- Example, compulsory extension of schooling by 1 year in Britain by 1 year. Perfect compliance, so LATE is ATE for untreated as well-no never takers. (Not quite 100% true, but close.)

- Experiments. Can sometimes rule out non-compliance of 1 kind-always takers, by not letting people in the control group get the treatment.. Then the LATE is the ATET. Still often other non-compliance (can't force people to take it up). Reminder ITT/first stage is the WALD/IV. Bloom 1984.

- Can use encouragement design or offers when unethical to do a RCT.

- Example, smoking while pregnant. Unethical to assign people to smoke. But can encourage not smoking.

- What does it mean if different LATEs are different? If not different and populations are different, might be more willing to believe constant effects.

Compliers

- Can get characteristics of people who comply. How?

- Abadie (03).

IV no nos

- Always put Xs in both stages. If don't, don't get consistency (first stage residuals are not uncorrelated with Xs not in first stage).

- Forbidden regression Endogenous RHS is 0/1. Want to use non-linear model. Do not put $\hat{D}$ estimated with a non-linear model in the second stage, again, the residuals in first stage need not be orthogonal to fitted values and Xs. Can use $\widehat{D}$ as an instrument for D. Implicitly using non-linearity in relationship of D to X and Z. If have S and S^2 in Y, do 2 first stages, one with Z, and one with Z^2.

Weak instruments: Challenge if Z only weakly affects D

- Bound Jaeger Baker
  - Weak IV then small correlation with IV and error in structural equation can lead to inconsistency

    Next bit on the form of the bias I did not go through but will leave here for completeness

  - Finite Samples,IV biased to OLS

  - x is endogenous variable, z is iv. Equation for inconsistency, $\frac{plim(\widehat{\beta}_{IV}-\beta)}{plim(\widehat{\beta}_{OLS}-\beta)} = \frac{\sigma_{\widehat{X},e}/\sigma_{X,e}}{R^2_{X,Z}}$, where $R^2_{X,Z}$ is the population R^2 (partial if other Xs) if regress x on z. $\widehat{X}$ is the

project of x onto z. Worse if z and the structural equation are more correlated, worse if first stage weak (denominator small).

– finite sample bit, estimate first stage. Even if first stage is 0, won't get 0 in estimates, get equation below in Flores-Lagunes. More than 2 IVs, bias inversely proportional to F-statistic in first stage.

Here is where I got back to covering the material

- Issues with exclusion, QOB tied to attendance, testing. Health differences by QOB. SES lower in Q1 QOB.

- Look at AK91, make up random IVs. In T1 of their paper, they replicate AK. See F from first stage declines across columns. Table 2 adds SOB * YOB. Even smaller Fs. Then in T3 BJB show results with randomly assigned QOB. See not only coefficients like AK91, but also SEs are similar!

- Angrist and Krueger respond to BJB, try to find way to show their estimates are OK once they do some new techniques to undo the issues with weak IVS above.

- 2SIV Angrist and Krueger, 1992: Have Z and Y in one data set, Z and X in another. Can get indirect least squares estimate from both. Inoue and Solon discuss issues with the distributions of the Xs being different in the data sets. Think of as 2 sample TSLS although there is a 2 sample IV way. Also see Inoue and Solon paper.

- Split Sample IV. Angrist and Krueger JBES 95.

- Jackknife IV. Angrist, Imbens, Krueger. 1999. Blomquist and Dahlberg 1999. Leave me out when calculating first stage.

to test over-reject by 5%, critical value is 16 for 1 iv (!!!). (Also consider case where you want to limit over rejection worst case.)

- Angrist suggests: 1) look at first stage, 2) do some tests, 3) use best IV, and report just-identified (median unbiased), 4) look at reduced form

Andrews, Stock and Sun: Weak IVs

- Point out previous stuff all assumes homoeskedasticity. Better to use Montiel Olea and Pflueger F_{EFF}, which adjusts for heteroskedasticity. $k = 1$, Robust F OK, reduces to Kleiberger and Paap. Else use their critical values.

- Show this is practically an issue with many IV papers from the AER.

- Practically an issue with selecting on the first stage (bias, size distortions, pre-test). Simulation shows more size distortions if you screen on the F.

- Solution. Existing CI and tests which stay valid if IVs are weak. "Test inversion:" where take set of nulls can't reject for a value β_0, that provides a $1-\alpha$ confidence set.

- Anderson-Rubin test-efficient in a just identified model with 1 IV.

- Under null $H_0 : \beta = \beta_0$, $\delta - \pi\beta_0 = 0$, quadratic form of that (in sample) with sample analogue of variance, get AR statistic, which is distributied χ^2_k.

- With 1 IV, can be bounded interval, the real line, or the real line minus a bounded interval. (Real line if can't reject $\pi = 0$

or unidentified. With more than 1 IV, can be empty set (if no β ST $\delta = \pi\beta$. Not most powerful test always but is efficient if k=1.