Instrumental Variables

EC 425/525, Set 8

Edward Rubin

13 May 2019

Prologue

Schedule

Last time

Matching and propensity-score methods

- Conditional independence
- Overlap

Today

Instrumental variables (and two-stage least squares)

Upcoming

- Assignment due Sunday
- Proposal due Wednesday 5/22
- Midterm?

Research designs

Selection on observables and/or unobservables

We've been focusing on ***selection-on-observables* designs**, *i.e.*,

$$(\mathrm{Y}_{0i},\ \mathrm{Y}_{1i}) \perp\!\!\!\perp \mathrm{D}_i | \mathrm{X}_i$$

for **observable** variables X_i.

***Selection-on-unobservable* designs** replace this assumption with two new (but related) assumptions

1. $(\mathrm{Y}_{0i},\ \mathrm{Y}_{1i}) \perp \mathrm{Z}_i$

2. $\mathrm{Cov}(\mathrm{Z}_i,\ \mathrm{D}_i) \neq 0$

Selection on observables and/or unobservables

Our main goal in causal-inference minded (applied) econometrics boils down to isolating **"good" variation** in $\mathbf{D}_i$ (exogenous/as-good-as-random) from **"bad" variation** (the part of $\mathbf{D}_i$ correlated with $\mathbf{Y}_{0i}$ and $\mathbf{Y}_{1i}$).

(We want to avoid selection bias.)

- **Selection-on-observables designs** assume that we can control for all *bad variation* (selection) in $\mathbf{D}_i$ through a known (observed) $\mathbf{X}_i$.

- **Selection-on-unobservables designs** assume that we can extract part of the *good variation* in $\mathbf{D}_i$ (generally using some $\mathbf{Z}_i$) and then use this *good* part of $\mathbf{D}_i$ to estimate the effect of $\mathbf{D}_i$ on $\mathbf{Y}_i$. We throw away the *bad variation* in $\mathbf{D}_i$ (it's bad).

Which route?

So set of research designs is more palatable?

1. There are plenty of bad applications of both sets.
   Violated assumptions, bad controls, *etc.*

1. **Selection on observables** assumes we know *everything* about selection into treatment—we can identify *all* of the good (or bad) variation in $\mathbf{D}_i$.
   Tough in non-experimental settings. Difficult to validate in practice.

1. **Selection on unobservables** assumes we can isolate *some* good/clean variation in $\mathbf{D}_i$, which we then use to estimate the effect of $\mathbf{D}_i$ on $\mathbf{Y}_i$.
   Seems more plausible. Possible to validate. May be underpowered.

Instrumental variables

Introduction

Instrumental variables (IV)[†] is the canonical selection-on-unobservables design—isolating *good variation* in $\mathbf{D}_i$ via some magical instrument $\mathbf{Z}_i$.

Consider some model (structural equation)

$$\mathbf{Y}_i = \beta_0 + \beta_1 \mathbf{D}_i + \varepsilon_i \tag{1}$$

To guarantee consistent OLS estimates for β_1, want $\mathbf{Cov}(\mathbf{D}_i,\, \varepsilon_i) = \mathbf{0}$.
In general, this is a heroic assumption.

Alternative: Estimate β_1 via instrumental variables.

† For the moment, we're lumping together IV and two-stage least squares (2SLS) together—as many people do—even though they are technically different.

Instrumental variables

Definition

For our model

$$\mathrm{Y}_i = \beta_0 + \beta_1 \mathrm{D}_i + \varepsilon_i \tag{1}$$

A valid **instrument** is a variable Z_i such that

1. $\mathbf{Cov}(\mathrm{Z}_i,\, \mathrm{D}_i) \neq 0$
   our instrument correlates with treatment (so we can keep part of D_i)

1. $\mathbf{Cov}(\mathrm{Z}_i,\, \varepsilon_i) = 0$
   our instrument is uncorrelated with other (non-D_i) determinants of Y_i ,
   i.e., Z_i is excludable from equation (1). **(exclusion restriction)**

Instrumental variables

Example

Back to the returns to a college degree,

$$\text{Income}_i = \beta_0 + \beta_1 \text{Grad}_i + \varepsilon_i$$

OLS is likely biased.

What if that state conducts a (random) **lottery** for scholarships?

Let Lottery_i denote an indicator for whether i won a lottery scholarship.[†]

1. $\text{Cov}(\text{Lottery}_i, \text{Grad}_i) \neq 0 \; (> 0)$ if scholarships increase grad. rates.

1. $\text{Cov}(\text{Lottery}_i, \varepsilon_i) = 0$ since the lottery is randomized.

† We'll have to focus on families who were eligible/who applied.

The IV estimator

The IV estimator for our model

$$\mathrm{Y}_i = \beta_0 + \beta_1 \mathrm{D}_i + \varepsilon_i \tag{1}$$

with (valid) instrument Z_i is

$$\hat{\beta}_{\mathrm{IV}} = \left(\mathrm{Z}'\mathrm{D}\right)^{-1} \left(\mathrm{Z}'\mathrm{Y}\right)$$

If you have no covariates, then

$$\hat{\beta}_{\mathrm{IV}} = \frac{\mathrm{Cov}(\mathrm{Z}_i,\ \mathrm{Y}_i)}{\mathrm{Cov}(\mathrm{Z}_i,\ \mathrm{D}_i)}$$

The IV estimator

The IV estimator for our model

$$\mathrm{Y}_i = \beta_0 + \beta_1 \mathrm{D}_i + \varepsilon_i \tag{1}$$

with (valid) instrument Z_i is

$$\hat{\beta}_{\mathrm{IV}} = \left(\mathrm{Z}'\mathrm{D}\right)^{-1}\left(\mathrm{Z}'\mathrm{Y}\right)$$

If you have additional (exogenous) covariates $\mathbf{X}_i$, then

$$\mathbf{Z} = \begin{bmatrix} \mathrm{Z}_i & \mathbf{X}_i \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} \mathrm{D}_i & \mathbf{X}_i \end{bmatrix}$$

Proof: Consistency

With a valid instrument $\mathbf{Z}_i$, $\hat{\beta}_{\text{IV}}$ is a consistent estiamtor for β_1 in

$$\mathbf{Y}_i = \beta_0 + \beta_1 \mathbf{X}_i + \varepsilon_i \tag{1}$$

$$\begin{aligned}
&\text{plim}\left(\hat{\beta}_{IV}\right) \\
&\quad = \text{plim}\left(\left(\text{Z}'\text{D}\right)^{-1}\left(\text{Z}'\text{Y}\right)\right) \\
&\quad = \text{plim}\left(\left(\text{Z}'\text{D}\right)^{-1}\left(\text{Z}'\text{D}\beta + \text{Z}'\varepsilon\right)\right) \\
&\quad = \text{plim}\left(\left(\text{Z}'\text{D}\right)^{-1}\left(\text{Z}'\text{D}\right)\beta\right) + \text{plim}\left(\frac{1}{N}\text{Z}'\text{D}\right)^{-1}\text{plim}\left(\frac{1}{N}\text{Z}'\varepsilon\right) \\
&\quad = \beta \quad ✔
\end{aligned}$$

Two-stage least squares

Two-stage least squares

Setup

You'll commonly see IV implemented as a two-stage process known as **two-stage least squares** (2SLS).

First stage Estimate the effect of the instrument Z_i on our endogenous variable $\mathbf{D}_i$ and (predetermined) covariates $\mathbf{X}_i$. Save $\widehat{\mathrm{D}}_i$.

$$\mathbf{D}_i = \gamma_1 \mathrm{Z}_i + \gamma_2 \mathbf{X}_i + u_i$$

Second stage Estimate model we wanted—but only using the variation in $\mathbf{D}_i$ that correlates with Z_i, *i.e.*, $\widehat{\mathrm{D}}_i$.

$$\mathbf{Y}_i = \beta_1 \widehat{\mathrm{D}}_i + \beta_2 \mathbf{X}_i + \varepsilon_i$$

Note The controls $\mathbf{X}_i$ must match in the first and second stages.

Two-stage least squares

IV estimation

This two-step procedure, with a valid instrument, produces an estimator $\hat{\beta}_1$ that is consistent for β_1.

$$\hat{\beta}_{2\text{SLS}} = \left(\mathrm{D}'\mathrm{P}_{\mathrm{Z}}\mathrm{D}\right)^{-1}\left(\mathrm{D}'\mathrm{P}_{\mathrm{Z}}\mathrm{Y}\right)$$

$$\mathrm{P}_{\mathrm{Z}} = \mathrm{Z}\left(\mathrm{Z}'\mathrm{Z}\right)^{-1}\mathrm{Z}'$$

where $\mathbf{D}$ is a matrix of our treatment and predetermined covariates $(\mathbf{X}_i)$ and Z is a matrix of our instrument and our predetermined covariates.

Two-stage least squares

IV estimation

Important notes

- The controls (X_i) must match in the first and second stages.
- If you have exactly **one instrument** and exactly **one endogenous variable**, then 2SLS and IV are identical.
- Your second-stage standard errors are not correct.

The reduced form

In addition to the regressions within the two stages of 2SLS

1. $\mathbf{D}_i = \gamma_1 \mathrm{Z}_i + \gamma_2 \mathbf{X}_i + u_i$
2. $\mathbf{Y}_i = \beta_1 \widehat{\mathrm{D}}_i + \beta_2 \mathbf{X}_i + \varepsilon_i$

there is a third important and related regression: the reduced form.

The **reduced form** regresses the outcome $\mathbf{Y}_i$ (LHS of the second stage) on our instrument Z_i and covariates $\mathbf{X}_i$ (RHS of the first stage).

$$\mathbf{Y}_i = \pi_1 \mathrm{Z}_i + \pi_2 \mathbf{X}_i + u_i$$

Thus, the reduced form provides a consistent estimate of the causal effect of our instrument on the outcome.

Two-stage least squares

The reduced form, continued

While the reduced form estimates the causal effect of the instrument on our outcome, we're often actually interested in the effect of *treatment* $(\mathbf{D}_i)$.

That said, the reduced form is still incredibly helpful/important:

- Clarifies your source of identifying variation.
- Does not suffer from *weak instruments* problems.
- Only requires $\mathbf{Cov}(\mathbf{Z}_i,\, \varepsilon_i) = \mathbf{0}$.
- Offers insights into your estimates

$$\widehat{\beta}_1^{\text{2SLS}} = \frac{\widehat{\pi}_1}{\widehat{\gamma}_1}$$

when you have exactly one instrument

Two-stage least squares

The reduced form, intuition

This expression for the 2SLS (and IV) estimator can be very helpful.

$$\widehat{\beta}_1^{\text{2SLS}} = \frac{\widehat{\pi}_1}{\widehat{\gamma}_1} = \frac{\text{Reduced-form estimate}}{\text{First-stage estimate}}$$

What's the interpretation/intuition?

Back to our example: $\widehat{\beta}_1 =$ est. effect of college graduation on income.

$\widehat{\pi}_1$ gives the estimated causal effect of the scholarship lottery on income , but what share of lottery winners graduate? We need to rescale if $<$ 100%.

$\widehat{\gamma}_1$ estimates the effect of winning the scholarship lottery on graduation — the share of winners who graduated due to winning. We can scale with $\widehat{\gamma}_1$!

The reduced form, example

To see why this scaling makes sense, imagine that 50% of lottery winners graduate from college due to the lottery, *i.e.*, $\hat{\gamma}_1 = 0.50$.[†]

Our reduced-form estimate of $\hat{\pi}_1 = \$5{,}000$ says that lottery winners make \$5,000 more than the control group, on average.

However, half of the winners did not graduate, so $\hat{\pi}_1$ "underestimates" the effect of college graduation by combining graduates by nongraduates.

Thus, we want to double $\hat{\pi}_1$, *i.e.*, divide by $\hat{\gamma}_1$: $\hat{\pi}_1 / \hat{\gamma}_1$ = \$5,000/0.5 = \$10,000.

† Imagine none of the applicants would have graduated otherwise

Two-stage least squares

Q How do we get this magical expression? $\left(\widehat{\beta}_1^{\text{IV}} = \dfrac{\widehat{\pi}_1}{\widehat{\gamma}_1}\right)$

Derivation

$$
\begin{aligned}
\widehat{\beta}_1^{\text{IV}} &= \left(\mathrm{Z}'\mathrm{D}\right)^{-1}\left(\mathrm{Z}'\mathrm{Y}\right) \\
&= \left(\widetilde{\mathrm{Z}}'\widetilde{\mathrm{D}}\right)^{-1}\left(\widetilde{\mathrm{Z}}'\mathrm{Y}\right) \qquad \text{applying FWL to reduce } \mathbf{D} \text{ and } \mathbf{Z} \text{ to vectors.} \\
&= \frac{\mathrm{Cov}\left(\widetilde{\mathrm{Z}}_i,\ \mathrm{Y}_i\right)}{\mathrm{Cov}\left(\widetilde{\mathrm{Z}}_i,\ \widetilde{\mathrm{D}}_i\right)} = \frac{\mathrm{Cov}\left(\widetilde{\mathrm{Z}}_i,\ \mathrm{Y}_i\right)/\,\mathrm{Var}\left(\widetilde{\mathrm{Z}}_i\right)}{\mathrm{Cov}\left(\widetilde{\mathrm{Z}}_i,\ \widetilde{\mathrm{D}}_i\right)/\,\mathrm{Var}\left(\widetilde{\mathrm{Z}}_i\right)} \\
&= \frac{\widehat{\pi}_1}{\widehat{\gamma}_1} \qquad \checkmark
\end{aligned}
$$

Let's push a bit deeper into IV's mechanics and intuition.

IV: Mechanics and intuition

Setup

In this section, we'll use medical trials as a working example.[†]

We are interested in the regression model for the effect of some treatment (*e.g.*, blood-pressure medication) on medical outcome $\mathbf{Y}_i$

$$\mathbf{Y}_i = \beta_0 + \beta_1 \mathbf{D}_i + \varepsilon_i$$

$\mathbf{D}_i$ indicates whether i *takes* the treatment (medication). ε_i captures all other factors that affect $\mathbf{Y}_i$. Or in potential-outcomes framework:

$$\begin{aligned} \mathbf{Y}_i &= \mathbf{Y}_{1i}\mathbf{D}_i + \mathbf{Y}_{0i}(1 - \mathbf{D}_i) \\ \mathbf{Y}_{0i} &= \beta_0 + \varepsilon_i \\ \mathbf{Y}_{1i} &= \mathbf{Y}_{0i} + \beta_1 \end{aligned}$$

† Credit/thanks go to Michael Anderson for this example—and much of these notes.

IV: Mechanics and intuition

Research design

Goal **Estimate the effect of blood-pressure medication** on blood pressure.

Challenge **Selection bias:** Even if treatment reduces blood pressure, selection bias will fights against the estimated effect.

Solution **Randomized medical trial:** Ask randomly chosen individuals in treatment group to take the pill. Control individual get placebo (or nothing).

Analysis 1 **Intention to treat** (**ITT**): $\widehat{\beta}_1^{\text{ITT}} = \overline{\mathbf{Y}}_{\text{Trt}} - \overline{\mathbf{Y}}_{\text{Ctrl}}$

ITT problem **Bias from noncompliance:** People don't always follow rules. *E.g.*, treated folks who don't take pills; control folks who take pills.

Analysis 2 **IV!** Instrument medication $\mathbf{D}_i$ with intention to treat $\mathbf{Z}_i$.

The IV solution

First question: Is $\mathbf{Z}_i$ a valid instrument for $\mathbf{D}_i$?

1. $\mathbf{Cov}(\mathbf{Z}_i,\, \varepsilon_i) = \mathbf{0}$ as $\mathbf{Z}_i$ was randomly assigned (exclusion restriction).

1. $\mathbf{Cov}(\mathbf{Z}_i,\, \mathbf{D}_i) \neq \mathbf{0}$ if assignment to treatment changes the likelihood you take the pills (first stage).

$\therefore$ $\mathbf{Z}_i$ is a valid instrument for $\mathbf{D}_i$ and IV consistently estimates β_1.

IV: Mechanics and intuition

Noncompliance

Noncompliant individuals do not abide by their treatment assignment.

Let's see how IV "solves" this problems.

First, assume noncompliance only affects treated individuals—*i.e.*, treated folks sometimes don't take their pills; control folks never take pills.

Noncompliance, continued

The **first stage** recovers the share of treatment individuals who take the pill

$$\mathbf{D}_i = \gamma_1 \mathbf{Z}_i + u_i$$

i.e., if 50% of treated individuals take the medication, $\widehat{\gamma} = 0.50$.

The **reduced form** estimates the *ITT*

$$\mathbf{Y}_i = \pi_1 \mathbf{Z}_i + v_i$$

which we know IV rescales using the first stage

$$\widehat{\beta}_1^{\text{IV}} = \frac{\widehat{\pi}_1}{\widehat{\gamma}_1} = \frac{\widehat{\pi}_1}{\mathbf{0.50}} = \mathbf{2} \times \widehat{\pi}_1$$

IV: Mechanics and intuition

Noncompliance, continued

IV solves the noncompliance issue by rescaling by the rate of compliance.

If everyone perfectly complies, then $\widehat{\gamma}_1 = \mathbf{1}$ and $\widehat{\beta}_1^{\text{IV}} = \widehat{\pi}_1/\mathbf{1} = \widehat{\beta}_1^{\text{ITT}}$.

Further example N_{Trt} = 10; trt. compliance = 50%; ctrl. compliance = 100%.

$\overline{\mathbf{Y}}_{\text{Trt}} = \dfrac{\mathbf{5}(\beta_0 + \beta_1) + \mathbf{5}(\beta_0)}{\mathbf{10}} = \beta_0 + \dfrac{\beta_1}{\mathbf{2}}$ and $\overline{\mathbf{Y}}_{\text{Ctrl}} = \beta_0$.

So our reduced-form estimate (the ITT) is $\widehat{\gamma}_1 = \dfrac{\beta_1}{\mathbf{2}}$ (half the true effect).

IV consistently estimates β_1 via rescaling the ITT by the rate of compliance

$$\widehat{\beta}_1^{\text{IV}} = \frac{\pi}{\gamma} = \frac{\beta_1/\mathbf{2}}{\mathbf{1}/\mathbf{2}} = \beta_1$$

IV: Mechanics and intuition

Takeaways

Main points

1. IV **rescales** the causal effect of $\mathbf{Z}_i$ on $\mathbf{Y}_i$ by the causal effect of $\mathbf{Z}_i$ on $\mathbf{D}_i$.

1. IV **does not** compare treated compliers to untreated compliers.
   Such a comparison/estimator would re-introduce selection bias.

Table of contents

Admin

1. Schedule

Instrumental variables

1. Research designs
2. Introduction
3. Definition
4. Example
5. IV estimator

Two-stage least squares

1. Setup
2. The reduced form
   - Defined
   - Intuition
   - Example
   - Derivation

IV: Intuition and mechanics

1. Setup
2. Noncompliance
3. Rescaling