Advanced Econometrics I

Jürgen Meinecke

Lecture 5 of 12

Research School of Economics, Australian National University

Roadmap

We have discussed two types of linear models:

linear projection:	$Y_i = X_i'\beta^* + u_i$	$\mathrm{E}(u_i X_i) = 0$
linear regression:	$Y_i = X_i'\beta + e_i$	$\mathrm{E}(e_i \vert X_i) = 0$

If we are only interested in β^*, then OLS is best

The linear regression model can also be estimated via OLS because $\mathrm{E}(e_i|X_i) = 0$ implies $\mathrm{E}(e_i X_i) = 0$

But it is more restrictive: it says that the $\mathrm{E}(Y_i|X_i)$ is linear

It defines a structural model that is very simplistic:
it basically states that the projection *is* the structural relation of interest

In practice, we are often dealing with this model instead:

$$Y_i = X_i'\beta + e_i \qquad \mathrm{E}(e_i X_i) \neq 0$$

It is often called **structural model** to emphasize that β is the coefficient of interest

Correspondingly, β is called **structural parameter**

I find this terminology unfortunate, because there is nothing inherently *structural* about these “models”

They are simple regression equations with the complication that β should not be estimated via OLS
(because $\mathrm{E}(e_i X_i) \neq 0$, and so β isn’t the projection coefficient)

When $\mathrm{E}(e_i X_i) \neq 0$ we say that X_i is **endogenous**

Clearly, $\mathrm{E}(e_i | X_i) \neq 0$

The three textbook examples of endogeneity are

- measurement error
- simultaneity, simultaneous equations, and
- omitted variable bias

Let’s have a look

Measurement Error

Let's say the "true" model is

$$Y_i = X_i'\beta + e_i, \qquad \mathsf{E}(e_i|X_i) = 0$$

If you had data on (X_i, Y_i) then OLS would be best

But let's say you only observe $(\tilde{X}_i, Y_i)$ with $\tilde{X}_i = X_i + r_i$ where r_i is a measurement error statistically independent of e_i and X_i

Despite its randomness, this error causes serious problems:

$$Y_i = \tilde{X}_i'\beta + v_i, \qquad \text{where } v_i := e_i - r_i'\beta$$

Can you safely use OLS here? Assuming $\mathsf{E}r_i = 0$,

$$\mathsf{E}(\tilde{X}_i v_i) = \mathsf{E}\left((X_i + r_i)(e_i - r_i'\beta)\right) = -\mathsf{E}(r_i r_i')\beta \neq 0$$

No you cannot! (unless $\beta = 0$ or $\mathsf{E}(r_i r_i') = 0$)

Simultaneity, Simultaneous Equations

Consider the following two equation model

$$Y_{i1} = X_{i1}'\beta_1 + \theta_1 Y_{i2} + e_{i1}$$
$$Y_{i2} = X_{i2}'\beta_2 + \theta_2 Y_{i1} + e_{i2}$$

Let X_{i1} and X_{i2} be well behaved in the sense:
$\mathsf{E}(e_{i1}X_{i1}) = \mathsf{E}(e_{i1}X_{i2}) = \mathsf{E}(e_{i2}X_{i1}) = \mathsf{E}(e_{i2}X_{i2}) = 0$

Further assume $\mathsf{E}(e_{i1}e_{i2}) = 0$ to keep things simple

Using Y_{i1} and Y_{i2} as regressors is problematic:

$$\begin{aligned}
\mathsf{E}(e_{i1}Y_{i2}) &= \mathsf{E}\left(e_{i1}(X_{i2}'\beta_2 + \theta_2 Y_{i1} + e_{i2})\right) = \theta_2 \mathsf{E}(e_{i1}Y_{i1}) \\
&= \theta_2 \mathsf{E}\left(e_{i1}(X_{i1}'\beta_1 + \theta_1 Y_{i2} + e_{i1})\right) \\
&= \theta_1\theta_2 \mathsf{E}\left(e_{i1}Y_{i2}\right) + \sigma_1^2 \\
&= \frac{\theta_2}{1-\theta_1\theta_2}\sigma_1^2 \neq 0
\end{aligned}$$

where $\sigma_1^2 := \mathsf{Var}\left(e_{i1}\right)$

Therefore, in the two equation model

$$Y_{i1} = X'_{i1}\beta_1 + \theta_1 Y_{i2} + e_{i1}$$
$$Y_{i2} = X'_{i2}\beta_2 + \theta_2 Y_{i1} + e_{i2}$$

the errors are not uncorrelated with all regressors

$$\mathrm{E}(e_{i1}Y_{i2}) = \frac{\theta_2}{1-\theta_1\theta_2}\sigma_1^2 \qquad \mathrm{E}(e_{i2}Y_{i1}) = \frac{\theta_1}{1-\theta_1\theta_2}\sigma_2^2$$

If $\theta_2 = 0$, then the first equation doesn't have an endogeneity problem and OLS is fine (similarly $\theta_1 = 0$ for second equation)

Omitted Variables Bias

A simple model illustrates the main idea

$$Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + u_i$$

where $\mathrm{E}(u_i|X_{i1}) = 0$, $\mathrm{E}(X_{i2}|X_{i1}) \neq 0$ and you don't observe X_{i2}

You have to omit X_{i2} from the regression

Effectively you are facing the model

$$Y_i = X_{i1}\beta_1 + e_i \qquad e_i := X_{i2}\beta_2 + u_i$$

where $\mathrm{E}(e_i|X_{i1}) \neq 0$

Is this a problem?

Only if $\mathrm{E}(e_i X_{i1}) \neq 0$ which may well be the case

Roadmap

Let's say you have got three scalar rvs X_i, Y_i, Z_i and you have the model

$$Y_i = X_i\beta + e_i \qquad \text{(structural equation)}$$

$$X_i = Z_i\pi + v_i, \qquad \mathsf{E}(v_i Z_i) = 0 \qquad \text{(first stage regression)}$$

Notice: first stage is simply a projection

Your research interest is β

Should you use OLS to estimate it? Yes if

- $\mathsf{E}(e_i|X_i) = 0$
  (then you don't really need Z_i at all)
- $\mathsf{E}(e_i|X_i) \neq 0$ but $\mathsf{E}(e_i X_i) = 0$
  (case of *omitted variable non-bias*)

In short: you need $\mathsf{E}(e_i X_i) = 0$ for OLS to make sense

What if $\mathrm{E}(e_i X_i) \neq 0$

Then the existence of Z_i will be helpful as long as $\mathrm{E}(e_i Z_i) = 0$

Notice that $\mathrm{E}(e_i Z_i) = 0$ implies that $\mathrm{E}(e_i v_i) \neq 0$, that is, the error terms of both equations must be correlated

How does Z_i help?

Combine the two equations to get

$$\begin{aligned} Y_i &= Z_i \pi \beta + (e_i + v_i \beta) \\ &= Z_i \pi \beta + w_i, \end{aligned}$$

where $\mathrm{E}(w_i Z_i) = 0$

Therefore you can consistently estimate $\pi\beta$

Of course you can also consistently estimate π

Simple idea: divide the estimator of $\pi\beta$ by the estimator of π

It follows that you can back out a consistent estimator of β

Alternative motivation: estimate β in two stages

(i) Estimate π via OLS in the first stage regression, create $\hat{X}_i = \hat{\pi} Z_i$

(ii) Regress Y on $\hat{X}_i$ using OLS

The estimator from stage (ii) is numerically identical to the one from the procedure explained on the preceding slide

Why should this make sense?
Why can you use $\hat{X}_i$ but not X_i in the structural equation?

Intuition: writing $X_i = \hat{X}_i + \hat{v}_i$ we see that

- $\hat{X}_i$ captures the variation of X_i that is *exogenous*
- $\hat{v}_i$ captures the variation of X_i that is *endogenous*

This little example provides a lot of the main ideas about IV estimation already

Unfortunately, however, things get considerably more intricate and complicated once the setup is generalized

It is very important to discuss this extensively in the lecture

IV and 2SLS estimation are pervasive in economics

I'm not sure you can publish a paper only based on OLS

People always cry "endogeneity!" and ask for an instrument

Let's properly understand the pros and cons, and provide best practices

Roadmap

Instrumental Variables Estimation

- Motivation
- Main Idea in a Nutshell
- General Setup
- Identification
- Definition of IV Estimator
- Large Sample Properties of IV Estimator
- Potential Outcomes Framework, Treatment Effects
- Treatment Heterogeneity: What is IV Estimating?
- Example

Starting point is the following partition of the linear model

$$\begin{aligned} Y_i &= X_i'\beta + e_i \\ &= X_{i1}'\beta_1 + X_{i2}'\beta_2 + e_i \end{aligned}$$

where $\dim \beta_1 = \dim X_{i1} = K_1 \times 1$

$\dim \beta_2 = \dim X_{i2} = K_2 \times 1$ with $K_1 + K_2 = K$

The two types of regressors are characterized by

$\mathrm{E}(e_i X_{i1}) = 0$ (exogenous regressors)

$\mathrm{E}(e_i X_{i2}) \neq 0$ (endogenous regressors)

This immediately tells you that $\beta \neq \beta^*$

Should we use OLS to estimate β?

No, we shouldn't use OLS to estimate β

$\hat{\beta}^{\text{OLS}}$ will consistently estimate β^*, but $\beta \neq \beta^*$

We need something new

Enter the instrumental variable:

Definition (Instrumental Variable (IV))

A $L \times 1$ vector Z_i is called an **instrumental variable (IV)** if

(i)	$\mathrm{E}(Z_i e_i) = 0$	instrument exogeneity
(ii)	$\text{rank } \mathrm{E}(Z_i X_i') = K$	instrument relevance

Notice that X_{i1} does satisfy (i) and will *always* be included in Z_i

Intuition for (ii): nonzero correlation between X_i and Z_i

A necessary condition for (ii) is $L \geq K$
(at least as many equations as unknowns)

Think of Z_i as partitioned like so:

$$Z_i := \begin{pmatrix} Z_{i1} \\ Z_{i2} \end{pmatrix} = \begin{pmatrix} X_{i1} \\ Z_{i2} \end{pmatrix}$$

Let $\dim Z_{i2} = L_2$; it is clear that $\dim Z_{i1} = K_1$

In other words, the first component of Z_i is always X_{i1} and the second component of Z_i are genuinely *new* instrumental variables that were not included in the model in the first place

The existence of Z_{i2} is crucial to be able to estimate β

Depending on the dimension of Z_{i2} we call the system

$$\dim Z_{i2} = \dim X_{i2} \quad \text{(exactly identified)}$$

$$\dim Z_{i2} > \dim X_{i2} \quad \text{(over identified)}$$

$$\dim Z_{i2} < \dim X_{i2} \quad \text{(under identified)}$$

Roadmap

Now we turn our attention to two **reduced form** regressions:

(i) regressing X_i on Z_i (this is the first stage regression)

(ii) regressing Y_i on Z_i

Think of the reduced form as an auxiliary regression that you are using merely as a means to an end

You are not typically interested in the reduced form itself, you are only using it as a tool

The reduced form is usually free of any economic meaning

Reduced form for X_i as dependent variable

Consider the *multivariate* regression model

$$X_i = \pi' Z_i + v_i$$

This notation comprises K regressions with each element of X_i as a dependent variable

Notice that $\dim \pi = L \times K$

Let $\pi = \mathsf{E}(Z_i Z_i')^{-1}\mathsf{E}(Z_i X_i')$, implying $\mathsf{E}(Z_i v_i') = 0$
(this means that the π are the projection coefficients)

Reduced form for Y_i as dependent variable

The reduced form for X_i can be plugged into the original regression:

$$\begin{aligned} Y_i &= X_i'\beta + e_i \\ &= (\pi' Z_i + v_i)'\beta + e_i \\ &= Z_i'\lambda + w_i, \end{aligned}$$

with $\lambda := \pi\beta$ and $w_i := v_i'\beta + e_i$

Notice that $\mathsf{E}(Z_i w_i) = \mathsf{E}(Z_i v_i')\beta + \mathsf{E}(Z_i e_i) = 0$

This means that λ is a projection coefficient, that is,
$\lambda = \mathsf{E}(Z_i Z_i')^{-1}\mathsf{E}(Z_i Y_i)$

Collecting results: for the two reduced form coefficients we have

$$\lambda = \mathsf{E}(Z_i Z_i')^{-1} \mathsf{E}(Z_i Y_i)$$

$$\pi = \mathsf{E}(Z_i Z_i')^{-1} \mathsf{E}(Z_i X_i')$$

The rhs expressions are population moments which are uniquely determined by the distribution that generates the observed data

This implies that λ and π are uniquely determined too

They are *identified*

Great! But wait: we're not interested in λ and π

Instead we want to know about β

Identification of β is not so straightforward, recall:

$$\pi\beta = \lambda$$

Our goal: solve for β

Can't simply divide by π

Let's think about the dimensions

- $\dim \pi = L \times K$
- $\dim \beta = K \times 1$
- $\dim \lambda = L \times 1$

So $\pi\beta = \lambda$ is a system of L equations for K unknowns

Linear algebra tells you that there

- are no solutions or infinitely many solutions if $L < K$
- is hope for unique solution only if $L \geq K$

So let's only consider $L \geq K$

Today we'll focus on the case $L = K$

This case is usually called the *exactly identified* case

(Aside: $L > K$ is called the *over-identified* case)

With $L = K$, to solve $\pi\beta = \lambda$ for β
we need rank $\pi = K$ (full rank) to ensure a unique solution

Notice that π is an upper triangular block matrix (see assignment 5)
for which rank $\pi = K_1 + \text{rank } \mathrm{E}(Z_{i2}X'_{i2})$

So it only boils down to whether or not rank $\mathrm{E}(Z_{i2}X'_{i2}) = K_2$

The IV relevance condition makes this happen

Recall the IV relevance condition: rank $\mathsf{E}(Z_i X_i') = K$

This condition implies rank $\mathsf{E}(Z_{i2} X_{i2}') = K_2$

The IV relevance condition therefore ensures that π has full rank, so that we can use matrix inversion to solve $\pi\beta = \lambda$:

$$
\begin{aligned}
\beta &= \pi^{-1}\lambda \\
&= \mathsf{E}(Z_i X_i')^{-1}\mathsf{E}(Z_i Z_i')\mathsf{E}(Z_i Z_i')^{-1}\mathsf{E}(Z_i Y_i) \\
&= \mathsf{E}(Z_i X_i')^{-1}\mathsf{E}(Z_i Y_i)
\end{aligned}
$$

(we have used the fact that $(AB)^{-1} = B^{-1}A^{-1}$)

This solution for β motivates the IV estimator

Roadmap

For the case $L = K$, we've just obtained this solution:

$$\beta = \mathrm{E}(Z_i X_i')^{-1} \mathrm{E}(Z_i Y_i)$$

Applying the analogy principle delivers the estimator

Definition (Instrumental Variable Estimator)

$$\hat{\beta}^{\mathrm{IV}} = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i X_i' \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} Z_i Y_i \right) = (Z'X)^{-1} Z'Y$$

Aside: when there is only one endogenous variable and one instrumental variable, then the IV estimator is simply

$$\hat{\beta}^{\mathrm{IV}} = \frac{s_{ZY}}{s_{XZ}}$$

that is, sample covariance between Z_i and Y_i over the sample covariance between X_i and Z_i
(we need this in week 9 when we look at the Wald estimator)

Roadmap

In assignment 5 you are asked to show:

Proposition (Consistency of $\hat{\beta}^{\text{IV}}$)

$\hat{\beta}^{IV} = \beta + o_p(1)$.

Proposition (Asymptotic Distribution of $\hat{\beta}^{\text{IV}}$)

$$\sqrt{N}(\hat{\beta}^{IV} - \beta) \xrightarrow{d} N\left(0, E(Z_i X_i')^{-1} E(e_i^2 Z_i Z_i') E(X_i Z_i')^{-1}\right).$$

Roadmap

Some new terminology

You are interested in the effect of a *treatment* X_i a person receives on an outcome Y_i

To keep things simple, the treatment is binary: $X_i \in \{0, 1\}$

The outcome is a function of the treatment: $Y_i(X_i)$

Only two *potential outcomes* per person: $Y_i(0)$ and $Y_i(1)$

The individual treatment effect (ITE) is: $Y_i(1) - Y_i(0)$,
that is: the difference in potential outcomes

This is the effect of the treatment on person i

Problem:
The individual treatment effect is never observed
(bc only one of the potential outcomes is observed per person)

How can we solve the problem of the missing counterfactual?

One idea would be to find an otherwise identical person $j \neq i$ who did not receive the treatment

For that person, the observed factual would be $Y_j(0)$

The individual treatment effect would be $Y_i(1) - Y_j(0)$

We do not observe if $Y_i(p) = Y_j(p)$ for $p \in \{0, 1\}$

What else can be done?

We are moving goal posts:
Instead of studying the *individual treatment effect* we look at the *average treatment effect*

The ATE is given by $\mathrm{E}\left(Y_i(1) - Y_i(0)\right)$

It is the effect on the average person in the population

While it would be great to know about the ITE, learning about the ATE also is immensely important

People who like to run regression want to know:
Can I estimate the ATE using OLS?

How does ATE relate to our regression β?

The potential outcomes framework is closely related to the study of *randomized controlled trials (RCT)*

RCTs have their roots in medical literature

A typical example is that of a medication that is randomly offered to some part of a sample and a placebo treatment to the other part

What are examples of treatments in economics?

- job training
- changes of legislation
- reducing class size in primary school
- sending out fake CVs to employers

The point here is: you have found a credible way to assign treatment *randomly*, that is X_i can be viewed as random

The potential outcomes framework can be mapped into a regression model

In the data you observe (X_i, Y_i)

Still, to keep things simple, binary treatment: $X_i \in \{0, 1\}$

Observed outcome is given by

$$
\begin{aligned}
Y_i &:= Y_i(1) \cdot X_i + Y_i(0) \cdot (1 - X_i) \\
&= (Y_i(1) - Y_i(0)) X_i + Y_i(0) \\
&= \underbrace{\mathrm{E}\left(Y_i(0)\right)}_{\beta_0} + \underbrace{\left(Y_i(1) - Y_i(0)\right)}_{\beta_{1i}} X_i + \underbrace{\left(Y_i(0) - \mathrm{E}\left(Y_i(0)\right)\right)}_{\tilde{u}_i} \\
&= \beta_0 + \beta_{1i} X_i + \tilde{u}_i
\end{aligned}
$$

The last line looks like a regression

Careful though:

$$Y_i = \beta_0 + \beta_{1i} X_i + \tilde{u}_i$$

The slope coefficient is *indidvidual specific* (it has an i-subscript)

The coefficient β_{1i} is the individual treatment effect

You wouldn't use OLS here unless you think that β_{1i} is constant

Even then you would still need $E(X_i \tilde{u}_i) = 0$, which will be implied by random treatment (as we show soon)

Let's turn $Y_i = \beta_0 + \beta_{1i} X_i + \tilde{u}_i$ into a regression model in which the slope coefficient does not have an i-subscript

$$
\begin{aligned}
Y_i &= \mathsf{E}(Y_i(0)) + (Y_i(1) - Y_i(0))\, X_i + (Y_i(0) - \mathsf{E}(Y_i(0))) \\
&= \mathsf{E}(Y_i(0)) + (Y_i(1) - Y_i(0))\, X_i + (Y_i(0) - \mathsf{E}(Y_i(0))) \\
&\qquad + \mathsf{E}\,(Y_i(1) - Y_i(0)) \cdot X_i - \mathsf{E}(Y_i(1) - Y_i(0)) \cdot X_i \\
&= \mathsf{E}(Y_i(0)) + \mathsf{E}(Y_i(1) - Y_i(0)) \cdot X_i \\
&\qquad + \Big((Y_i(0) - \mathsf{E}(Y_i(0))) \\
&\qquad + \quad (Y_i(1) - Y_i(0))\, X_i - \mathsf{E}(Y_i(1) - Y_i(0)) \cdot X_i \Big) \\
&= \beta_0 + \beta_1 \cdot X_i + u_i,
\end{aligned}
$$

where $\beta_0 := \mathsf{E}(Y_i(0))$ and $\beta_1 := \mathsf{E}(Y_i(1) - Y_i(0))$ and everything in big parentheses is u_i

Notice that $\beta_1 := \mathsf{E}\,(Y_i(1) - Y_i(0))$ is equal to the ATE

For OLS go yield a good estimator of the ATE, need $\mathsf{E}(X_i u_i) = 0$

Let's investigate how we can obtain $\mathsf{E}(X_i u_i) = 0$

$$
\begin{aligned}
\mathsf{E}(u_i|X_i) &= \mathsf{E}\Big((Y_i(0) - \mathsf{E}(Y_i(0))) \\
&\qquad + (Y_i(1) - Y_i(0))\, X_i - \mathsf{E}(Y_i(1) - Y_i(0)) \cdot X_i | X_i \Big) \\
&= \mathsf{E}(Y_i(0)|X_i) - \mathsf{E}(\mathsf{E}(Y_i(0))|X_i) \\
&\qquad + \mathsf{E}(Y_i(1)|X_i) \cdot X_i - \mathsf{E}(Y_i(0)|X_i) \cdot X_i \\
&\qquad - \mathsf{E}(\mathsf{E}(Y_i(1))|X_i) \cdot X_i + \mathsf{E}(\mathsf{E}(Y_i(0))|X_i) \cdot X_i \\
&= \mathsf{E}(Y_i(0)|X_i) - \mathsf{E}(Y_i(0)) \\
&\qquad + \mathsf{E}(Y_i(1)|X_i) \cdot X_i - \mathsf{E}(Y_i(0)|X_i) \cdot X_i \\
&\qquad - \mathsf{E}(Y_i(1)) \cdot X_i + \mathsf{E}(Y_i(0)) \cdot X_i \\
&\stackrel{?}{=} \mathsf{E}(Y_i(0)) - \mathsf{E}(Y_i(0)) \\
&\qquad + \mathsf{E}(Y_i(1)) \cdot X_i - \mathsf{E}(Y_i(0)) \cdot X_i \\
&\qquad - \mathsf{E}(Y_i(1)) \cdot X_i + \mathsf{E}(Y_i(0)) \cdot X_i \\
&= 0
\end{aligned}
$$

The fourth equality follows if X_i is assigned randomly (treatment effect literature typically writes: $(Y_i(1), Y_i(0)) \perp\!\!\!\perp X_i$)

The conclusion follows because
$\mathrm{E}(X_i u_i) = \mathrm{E}\left(\mathrm{E}(X_i u_i | X_i)\right) = \mathrm{E}\left(X_i \cdot \mathrm{E}(u_i | X_i)\right) = 0$

So, given random assignment, we obtain the desired result that OLS results in a good estimator of the average treatment effect:

Theorem (OLS in Randomized Controlled Trial)

Suppose you have available data (X_i, Y_i) from a randomized controlled trial. In particular, X_i is a randomly assigned treatment dummy variable. Then the OLS estimator of β_1 in the model $Y_i = \beta_0 + \beta_1 X_i + u_i$ is a consistent estimator of the average treatment effect $E\left(Y_i(1) - Y_i(0)\right)$.

Roadmap

What if you cannot effectively randomize treatment?

Earlier I said that job training is a treatment

In practice, there's no way you can randomly assign job training as a treatment and expect full compliance

If you cannot effectively randomize treatment then X_i may not be independent of potential outcomes

There's a clever work around:
randomize *eligibility* for treatment instead

Let Z_i be eligibility for treatment, with $Z_i \in \{0, 1\}$

It is not a coincidence that we are using the letter Z_i here:
eligibility will play the role of an instrumental variable

We conjecture that $\hat{\beta}^{\text{IV}}$ could be a good estimator in this setting

Is this true? Does $\hat{\beta}^{\text{IV}}$ estimate the ATE?

Like before, let's study a model in which treatment effect is heterogeneous

$$Y_i = \beta_{0i} + \beta_{1i}X_i + u_i \qquad \text{(equation of interest)}$$
$$X_i = \pi_{0i} + \pi_{1i}Z_i + v_i \qquad \text{(first stage)}$$

where

$$\beta_{1i} = Y_i(1) - Y_i(0)$$
$$\pi_{1i} = X_i(1) - X_i(0),$$

with $Y_i(p) = Y_i(X_i = p)$ and $X_i(p) = X_i(Z_i = p)$ for $p \in \{0, 1\}$

Using a little bit of math, it can be shown that

$$\hat{\beta}^{\mathsf{IV}} = \frac{\mathsf{E}\left(\beta_{1i} \cdot \pi_{1i}\right)}{\mathsf{E}\left(\pi_{1i}\right)} + \mathsf{o}_p(1) \neq \mathsf{E}\left(\beta_{1i}\right) + \mathsf{o}_p(1)$$

(you will show this in assignment 9)

Two results here:

- IV estimator does not converge to the ATE (bad news?)
- Instead it converges to $\mathrm{E}(\beta_{1i} \cdot \pi_{1i}) / \mathrm{E}(\pi_{1i})$ (looks complicated)

Let's take a closer look at the probability limit

For no apparent reason, let's call it LATE

$$\text{LATE} := \frac{\mathrm{E}(\beta_{1i} \cdot \pi_{1i})}{\mathrm{E}(\pi_{1i})}$$

What is LATE and how does it relate to ATE?

Here is a useful way to contrast them:

$$\text{ATE} = \mathrm{E}\left(\beta_{1i}\right)$$

$$\text{LATE} = \frac{\mathrm{E}\left(\beta_{1i} \cdot \pi_{1i}\right)}{\mathrm{E}\left(\pi_{1i}\right)} = \mathrm{E}\left(\beta_{1i} \cdot \frac{\pi_{1i}}{\mathrm{E}\left(\pi_{1i}\right)}\right)$$

- interpret $\frac{\pi_{1i}}{\mathrm{E}(\pi_{1i})}$ as weights
- then the rhs is equal to the expected value of β_{1i} adjusted for these weights
- in other words: the rhs is a weighted average of β_{1i}
- ideally, we would not want any weights in there
(because we are after the ATE, which is the simple average)
- some intuition for the weights:
when π_{1i} is large relative to $\mathrm{E}(\pi_{1i})$ then the weight is large;
therefore people with large π_{1i} influence the IV estimator more
(their Z_i have a strong impact on X_i)

Putting things together: $\hat{\beta}^{\text{IV}}$ estimates the causal effect for those individuals for whom Z_i is most influential
(those with large π_{1i})

LATE is the acronym for *local average treatment effect*

The LATE can be understood as the ATE for the subpopulation whose treatment X_i is most heavily influenced by the instrument Z_i

LATE is an ATE only for this peculiar ("local") subpopulation; it is not equal to the ATE in the population

Actually, we can express LATE as a function in ATE:

Notice that $\text{Cov}(\beta_{1i}, \pi_{1i}) = \text{E}(\beta_{1i} \cdot \pi_{1i}) - \text{E}(\beta_{1i}) \cdot \text{E}(\pi_{1i})$

It follows

$$\begin{aligned} \text{LATE} &:= \frac{\text{E}(\beta_{1i} \cdot \pi_{1i})}{\text{E}(\pi_{1i})} \\ &= \frac{\text{E}(\beta_{1i})\text{E}(\pi_{1i}) + \text{Cov}(\beta_{1i}, \pi_{1i})}{\text{E}(\pi_{1i})} \\ &= \text{E}(\beta_{1i}) + \frac{\text{Cov}(\beta_{1i}, \pi_{1i})}{\text{E}(\pi_{1i})} \\ &= \text{ATE} + \frac{\text{Cov}(\beta_{1i}, \pi_{1i})}{\text{E}(\pi_{1i})} \end{aligned}$$

In words: LATE equals ATE plus “some stuff”

From previous slide

$$\text{LATE} = \text{ATE} + \frac{\text{Cov}(\beta_{1i}, \pi_{1i})}{\text{E}(\pi_{1i})}$$

But what exactly is “some stuff”?

It is the covariance between the two individual-specific parameters β_{1i} and π_{1i}

If the treatment effect β_{1i} tends to be large for individuals for whom the effect of the instrument π_{1i} is also large, then $\text{Cov}(\beta_{1i}, \pi_{1i}) > 0$ and therefore LATE > ATE
(supposing $\text{E}(\pi_{1i}) > 0$)

On the other hand, if the treatment effect β_{1i} tends to be small for individuals for whom the effect of the instrument π_{1i} is also large, then $\text{Cov}(\beta_{1i}, \pi_{1i}) < 0$ and therefore LATE < ATE

When does IV estimate the ATE?

- If $\beta_{1i} = \beta_1$ (no heterogeneity in equation of interest)
- If $\pi_{1i} = \pi_1$ (no heterogeneity in first stage equation)
- If β_{1i} and π_{1i} vary but are independently distributed

But these three are unrealistic

In general, $\hat{\beta}^{\text{IV}}$ does not estimate ATE

Whether this is important depends on the application

Define four exhaustive and mutually exclusive types based on their treatment response wrt to a particular value of $Z_i \in \{0, 1\}$

$$\begin{cases} \text{always taker} & X_i(0) = 1 \text{ and } X_i(1) = 1 \\ \text{complier} & X_i(0) = 0 \text{ and } X_i(1) = 1 \\ \text{defier} & X_i(0) = 1 \text{ and } X_i(1) = 0 \\ \text{never taker} & X_i(0) = 0 \text{ and } X_i(1) = 0 \end{cases}$$

This results in the following values of π_{1i} for these types:

$$\begin{aligned} \pi_{1i} &= X_i(1) - X_i(0) \\ &= \begin{cases} 0 & \text{always taker} \\ 1 & \text{complier} \\ -1 & \text{defier} \\ 0 & \text{never taker} \end{cases} \end{aligned}$$

Let's say the proportions of these four types are $\tau_{AT}, \tau_C, \tau_D, \tau_{NT}$, adding up to one

Furthermore, for simplicity claim that $\tau_D = 0$ (no defiers)

Then

$$\begin{aligned} \mathrm{E}(\pi_{1i}) &= \tau_{AT}\mathrm{E}(\pi_{1i}|AT) + \tau_C\mathrm{E}(\pi_{1i}|C) + \tau_{NT}\mathrm{E}(\pi_{1i}|NT) \\ &= \tau_C\mathrm{E}(\pi_{1i}|C) \\ &= \tau_C \end{aligned}$$

Likewise

$$\mathrm{E}\left(\beta_{1i} \cdot \pi_{1i}\right) = \tau_C\mathrm{E}\left(\beta_{1i}|C\right)$$

Therefore

$$\mathrm{LATE} = \frac{\mathrm{E}\left(\beta_{1i} \cdot \pi_{1i}\right)}{\mathrm{E}\left(\pi_{1i}\right)} = \mathrm{E}\left(\beta_{1i}|C\right) \neq \mathrm{E}\left(\beta_{1i}\right)$$

So

$$\begin{aligned} \text{LATE} &= \mathrm{E}\left(\beta_{1i}|C\right) \\ &= \mathrm{E}\left(Y_i(1) - Y_i(0)|C\right) \end{aligned}$$

This is important because it says that LATE is the ATE for the subpopulation of compliers

The four types AT, NT, D, and C differ in how their outcomes respond to a treatment

We would not expect a *homogenous* treatment effect, that is, each of these four types would have the same treatment effect

LATE is the treatment effect for one particular type, the compliers

IV estimation successfuly estimates that local treatment effect

When Z_i is binary, there's a special form for the IV estimator

$$\hat{\beta}^{\text{IV}} = \frac{s_{ZY}}{s_{XZ}} = \frac{\hat{\mathrm{E}}(Y_i|Z_i = 1) - \hat{\mathrm{E}}(Y_i|Z_i = 0)}{\hat{\mathrm{E}}(X_i|Z_i = 1) - \hat{\mathrm{E}}(X_i|Z_i = 0)}$$

It is customary to write

$$\hat{\beta}^{\text{IV}} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0}$$

with

$$\bar{Y}_1 := \frac{\sum_{i=1}^N Z_i Y_i}{\sum_{i=1}^N Z_i} \qquad \bar{Y}_0 := \frac{\sum_{i=1}^N (1 - Z_i) Y_i}{\sum_{i=1}^N (1 - Z_i)}$$

$$\bar{X}_1 := \frac{\sum_{i=1}^N Z_i X_i}{\sum_{i=1}^N Z_i} \qquad \bar{X}_0 := \frac{\sum_{i=1}^N (1 - Z_i) X_i}{\sum_{i=1}^N (1 - Z_i)}$$

This represenation of $\hat{\beta}^{\text{IV}}$ is called the *Wald estimator*

More generally, the Wald estimator is really any estimator that compares averages in grouped data as portrayed here

Roadmap

Let me present an application taken from Angrist and Pischke, "*Mostly Harmless Econometrics*", (2008)

The actual underlying paper is Bloom et al., "*The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study*", (1997)

Note:
I will be presenting a much simplified version of the paper

Background for understanding the paper

Their research question:
Is job training beneficial for economically disadvantaged adults?

People were randomly *made eligible* for job training

This is an example where treatment was not randomly assigned but instead the *eligibility* for treatment was

Key variables:

- X_i: treatment dummy equal 1 if received job training
- Z_i: dummy equal 1 if offered job training (randomly assigned)
- Y_i: total earnings in the 30-months period after random assignment

Typical example of *one-sided compliance*:

$$Z_i = 0 \Rightarrow X_i = 0$$

$$Z_i = 1 \Rightarrow X_i \in \{0, 1\}$$

A person in the control group cannot access treatment

You might expect that $X_i = Z_i$

But many people refuse the offer of treatment (takes effort!)

In the job training example

$$\widehat{\Pr}(X_i = 1 | Z_i = 1) = \hat{\mathrm{E}}(X_i | Z_i = 1) \approx 0.6$$

$$\widehat{\Pr}(X_i = 1 | Z_i = 0) = \hat{\mathrm{E}}(X_i | Z_i = 0) \approx 0.02$$

More or less confirms one-sided compliance

The estimation results...

	(1)	(2)	(3)	(4)
	OLS	ITT		LATE
	$\hat{\mathrm{E}}(Y_i \mid X_i = 1)$ $-\hat{\mathrm{E}}(Y_i \mid X_i = 0)$	$\hat{\mathrm{E}}(Y_i \mid Z_i = 1)$ $-\hat{\mathrm{E}}(Y_i \mid Z_i = 0)$	$\hat{\mathrm{E}}(X_i \mid Z_i = 1)$ $-\hat{\mathrm{E}}(X_i \mid Z_i = 0)$	(2) / (3)
Men	\$3,970	\$1,117	0.61	\$1,825
Women	\$2,133	\$1,243	0.64	\$1,942

ITT: *intention-to-treat* effect; the effect you would have calculated under full compliance (that is, if it had been true that $X_i = Z_i$)

Here ITT gives you a sort of lower bound

But compliance was only around 60% therefore ITT underestimates LATE

LATE is the treatment effect for *compliers*:
the subpopulation who are willing to take the treatment if offered