Econ 210 - Instrumental Variables

Sidharth Sah[1]

September 8, 2023

[1]Thanks to Azeem Shaikh and Max Tabord-Meehan for useful material

Introduction

- Suppose we have a *causal* model

$$Y = \beta_0 + \beta_1 X + U \tag{1}$$

  and we want to estimate the causal parameter, β_1
- So far we've talked about two ways we might do so:
  - If we have $E[XU] = 0$ (such as in an experiment), we can consistently estimate equation (1) using OLS
  - If we have a control variable, D, such that $E[U|X, D] = E[U|D]$, we can consistently estimate:

$$Y = \tilde{\beta}_0 + \tilde{\beta}_1 X + \tilde{\beta}_2 D + \tilde{U}$$

    using OLS to recover $\tilde{\beta}_1 = \beta_1$

Instrumental Variables

- Suppose that $E[XU] \neq 0$ and we can't find an appropriate control variable D. We might be able to proceed if we can find a variable, Z, to serve as an Instrument for X
- A valid instrument satisfies two conditions:
  - Instument Exogeneity: $Cov(Z, U) = E[ZU] = 0$
  - Instument Relevance: $Cov(Z, X) \neq 0$
- Basic idea: find a Z that *only* affects Y *through* X - then use the variation in X induced by Z to get at relationship between X and Y

Instrumental Variable Example

- Suppose we're interested in the effect of sentencing of convicted felons on recidivism. Ie, we're interested in:

$$R = \beta_0 + \beta_1 P + U$$

with

$$R = \begin{cases} 1 \text{ if commit another crime} \\ 0 \text{ if not} \end{cases} \qquad P = \begin{cases} 1 \text{ if go to prison} \\ 0 \text{ if not} \end{cases}$$

- Many reasons why $E[PU] \neq 0$ - lower-income defendants can't afford representation and may be more likely to get sent to prison (and also commit future crimes), etc

Instrumental Variable Example

- Might be able to use $J =$ judge severity as an instrument. If judges are assigned to cases "randomly" and judges don't interact with defendants in any way other than the sentencing decision, it's reasonable to assume:

$$Cov(J, U) = 0$$

- However, if some judges are stricter than others, it would also be the case that:

$$Cov(J, P) \neq 0$$

- Thus J is a valid instrument for P

Calculating β

- In order to guide estimation, we'll derive expressions for β_0 and β_1 in terms of moments of X, Y, and Z. Using $E[U] = 0$ (can be assumed for same reasons as in causal linear regression):

$$E[Y - \beta_0 - \beta_1 X] = 0$$
$$\Rightarrow \beta_0 = E[Y] - \beta_1 E[X]$$

- Now, where we made use of the $E[XU] = 0$ assumption in the linear regression case, we analogously make use of the $E[ZU]$ assumption

Calculating β

$$
\begin{aligned}
E[Z(Y-\beta_0-\beta_1 X)] &= 0 \\
E[Z(Y-(E[Y]-\beta_1 E[X])-\beta_1 X)] &= 0 \\
\underbrace{E[Z(Y-E[Y])]}_{=Cov(Y,Z)} &= \beta_1 \underbrace{E[Z(X-E[X])]}_{=Cov(X,Z)}
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \beta_1 &= \frac{Cov(Y,Z)}{Cov(X,Z)} \\
\Rightarrow \beta_0 &= E[Y] - \frac{Cov(Y,Z)}{Cov(X,Z)} E[X]
\end{aligned}
$$

Calculating β for Binary Z

- When Z is binary-valued, β_1 simplifies particularly nicely:

$$\beta_1 = \frac{Cov(Y,Z)}{Cov(X,Z)} = \frac{\frac{Cov(Y,Z)}{Var(Z)}}{\frac{Cov(X,Z)}{Var(x)}} = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[X|Z=1] - E[X|Z=0]}$$

- The final equality follows using the same argument we used to simplify the SLR β_1 for binary X

Heterogeneous Treatment Effects

- The previous expression will allow us to form a more general interpretation of IV while allowing for *heterogeneous treatment effects*
- Assuming a causal model

$$Y = \beta_0 + \beta_1 X + U$$

  implies that an additional unit of X has the same causal effect on everyone's Y: β_1
- We can use potential outcomes to think about people having different treatment effects

Potential Outcomes Recap

- As previously mentioned in the class, potential outcomes can be used when we have some binary treatment, $X \in \{0, 1\}$, and for *each* person in the population, for a causal model, $Y = g(X, U)$ we define:

$$Y_1 = g(1, U)$$
$$Y_0 = g(0, U)$$

- That is, Y_1 and Y_0 represent the two possible outcomes a person would have *if* they got each of the two possible treatments
- We only observe one outcome, Y, per person:

$$Y = Y_1X + Y_0(1 - X)$$

Potential Outcomes Recap

- For each individual person, we can imagine an individual treatment effect - how would that person's outcome change from being treated to being untreated:

$$Y_1 - Y_0$$

- With SLR for experiments, we said that we were identifying the average of the individual treatment effect across the population, aka the ATE:

$$ATE = E[Y_1 - Y_0]$$

Potential Treatments

- We can analogously define Potential Treatments. For a binary instrument, $Z \in \{0, 1\}$, because we have assumed that the treatment, X, is influenced by the instrument, Z, we could create a causal model of X, $X = h(Z, V)$:

$$X_1 = h(1, V)$$
$$X_0 = h(0, V)$$

- That is, X_1 and X_0 represent the two possible treatments a person would have *if* they got each of the two possible values of Z
- We only observe one outcome, X, per person:

$$X = X_1 Z + X_0(1 - Z)$$

Local Average Treatment Effect

- Under the following conditions, the estimand of an IV regression can be more precisely interpreted as a LATE. Those conditions are:
  (a) $(Y_1, Y_0, X_1, X_0) \perp Z$ (implies instrument exogeneity)
  (b) $X_1 \neq X_0$ sometimes (analogous to instrument relevance)
  (c) $X_1 \geq X_0$ always - called Monotonicity
- For our purposes, we'll also assume that both $X \in \{0, 1\}$ and $Z \in \{0, 1\}$

Local Average Treatment Effect

- Under the monotonicity assumption, we can divide the population into three distinct groups, based on the values of their X_1 and X_0:
  - Always-takers: People for whom $X_1 = 1,\ X_0 = 1$
  - Never-takers: People for whom $X_1 = 0,\ X_0 = 0$
  - Compliers: People for whom $X_1 = 1,\ X_0 = 0$
- The monotonicity condition rules out the possibility of anyone having $X_1 = 0,\ X_0 = 1$ (Defiers)

Local Average Treatment Effect

- Consider an equation:

$$Y = \beta_0 + \beta_1 X + U$$

  where U is defined causally, and we have an instrument Z that meets the LATE assumptions, but we're allowing for heterogeneous treatment effects (β_1 is not a homogeneous treatment effect, it is left undefined for now)
- Our LATE assumption (a) ensures instrument exogeneity and assumptions (b)+(c) ensure instrument relevance, so, for binary Z, we can express β_1 as

$$\beta_1 = \frac{Cov(Y, Z)}{Cov(X, Z)} = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[X|Z = 1] - E[X|Z = 0]}$$

Local Average Treatment Effect

- Re-express the denominator using our assumptions:

$$
\begin{aligned}
& E[X|Z=1] - E[X|Z=0] = E[X_1|Z=1] - E[X_0|Z=0] \\
& = E[X_1] - E[X_0] \qquad ((X_1, X_0) \perp Z) \\
& = E[X_1 - X_0] \\
& = E[1]P\{X_1 > X_0\} + E[0]P\{X_1 = X_0\} + E[-1]P\{X_1 < X_0\} \\
& = P\{X_1 > X_0\} \qquad \text{(Monotonicity)}
\end{aligned}
$$

Local Average Treatment Effect

Similarly, for the numerator:

$$
\begin{aligned}
&E[Y|Z=1] - E[Y|Z=0] \\
&= E[Y_1 X + Y_0(1-X)|Z=1] - E[Y_1 X + Y_0(1-X)|Z=0] \\
&= E[Y_1 X_1 + Y_0(1-X_1)|Z=1] - E[Y_1 X_0 + Y_0(1-X_0)|Z=0] \\
&= E[Y_1 X_1 + Y_0(1-X_1)] - E[Y_1 X_0 + Y_0(1-X_0)] \\
&\qquad\qquad\qquad\qquad\qquad\qquad ((Y_1, Y_0, X_1, X_0) \perp Z) \\
&= E[Y_1 X_1 + Y_0(1-X_1) - Y_1 X_0 - Y_0(1-X_0)] \\
&= E[(Y_1 - Y_0)(X_1 - X_0)] \\
&= E[(Y_1 - Y_0)|X_1 > X_0]P\{X_1 > X_0\} + E[0|X_1 = X_0]P\{X_1 = X_0\} \\
&\quad - E[(Y_1 - Y_0)|X_1 < X_0]P\{X_1 < X_0\} \\
&= E[(Y_1 - Y_0)|X_1 > X_0]P\{X_1 > X_0\} \qquad\qquad \text{(Monotonicity)}
\end{aligned}
$$

Local Average Treatment Effect

- Thus, under the LATE assumptions:

$$\beta_1 = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[X|Z=1] - E[X|Z=0]}$$
$$= \frac{E[(Y_1 - Y_0)|X_1 > X_0]P\{X_1 > X_0\}}{P\{X_1 > X_0\}}$$
$$= E[(Y_1 - Y_0)|X_1 > X_0]$$

- This is the Local Average Treatment Effect (LATE) - the average of the treatment effect specifically among the compliers!

LATE Example

- Return to the example of the effect of sentencing of convicted felons on recidivism, but allow for heterogeneous treatment effects, $R = g(P, U)$ with

$$g = \text{causal model of recidivism}$$

$$R = \begin{cases} 1 \text{ if commit another crime} \\ 0 \text{ if not} \end{cases} \qquad P = \begin{cases} 1 \text{ if go to prison} \\ 0 \text{ if not} \end{cases}$$

- We'll use judges as the instrument, assuming there are only two judges:

$$J = \begin{cases} 1 \text{ if mean judge} \\ 0 \text{ if nice judge} \end{cases}$$

LATE Example

- Using potential treatment notation, let P_1 and P_0 represent your sentences *if* you got the mean judge vs. the nice judge, respectively, and R_1 and R_0 represent if you commit future crimes *if* you go to prison or not, respectively
- Further assume there are three possible crimes, C, you can be convicted of:

$$C = \begin{cases} 2 \text{ if grand theft auto} \\ 1 \text{ if shoplifting} \\ 0 \text{ if littering} \end{cases}$$

LATE Example

- Whether you go to prison depends on the combination of your judge and offense,
  - Car thieves: $P_1(C = 2) = 1,\ P_0(C = 2) = 1$
  - Shoplifters: $P_1(C = 1) = 1,\ P_0(C = 1) = 0$
  - Litterers: $P_1(C = 0) = 0,\ P_0(C = 0) = 0$
- The above implies that monotonicity and $P_1 \neq P_0$ sometimes are satisfied
- Assume also that,

$$(R_1, R_0, P_1, P_0) \perp\!\!\!\perp J$$

LATE Example

- This means that all of the LATE assumptions are satisfied. If we calculate β_1 using IV, we'll get:

$$\beta_1 = E[R_1 - R_0 | P_1 > P_0] = E[R_1 - R_0 | C = 1]$$

- We end up with the average effect of prison on recidivism among shoplifters specifically! IV won't say anything about the effect on car thieves or on litterers
- This LATE is probably policy-relevant - we can change laws regarding punishment for shoplifting. Other LATEs might not be. LATEs (like all parameters) need to be assessed for their usefulness case-by-case

Requirements to Estimate β_1

- We now discuss how to estimate parameters using IV from finite samples. We'll maintain our basic IV assumptions:
  (a) $E[U] = 0$
  (b) $E[ZU] = 0$
  (c) $\text{Cov}[X, Z] \neq 0$

  where (Y, X, Z, U) satisfy:

$$Y = \beta_0 + \beta_1 X + U$$

- Also let $(Y_1, X_1, Z_1), ..., (Y_n, X_n, Z_n)$ be iid $\sim (Y, X, Z)$

IV Estimators

- We discussed that, under the basic assumptions,

$$\beta_1 = \frac{Cov(Y,Z)}{Cov(X,Z)}$$
$$\beta_0 = E[Y] - \frac{Cov(Y,Z)}{Cov(X,Z)}E[X]$$

- Thus, natural estimators are:

$$\hat{\beta}_1^{\text{IV}} = \frac{\hat{\sigma}_{Y,Z}}{\hat{\sigma}_{X,Z}}$$
$$\hat{\beta}_0^{\text{IV}} = \bar{Y}_n - \hat{\beta}_1^{\text{IV}}\bar{X}_n$$

- These are the Instumental Variables estimators of β_0 and β_1

Residuals

- We can again form predicted values and residuals:

$$\hat{Y}_i = \hat{\beta}_0^{\mathrm{IV}} + \hat{\beta}_0^{\mathrm{IV}} X_i$$

are the predicted values. The amounts these are off by are the IV residuals:

$$\hat{U}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0^{\mathrm{IV}} - \hat{\beta}_0^{\mathrm{IV}} X_i$$

Consistency of IV Estimators

- Along with the normal maintained assumptions, assume that $E[Y^4], E[X^4], E[Z^4] < \infty$. Then, the IV estimators are consistent:

$$\hat{\beta}_1^{\text{IV}} \xrightarrow{p} \beta_1$$
$$\hat{\beta}_0^{\text{IV}} \xrightarrow{p} \beta_0$$

- we'll show this for β_1^{IV} using the CMT

Consistency of IV Estimators

- β_1^{IV} is composed of $\hat{\sigma}_{Y,Z}$ and $\hat{\sigma}_{Z,X}$. We know that the sample covariance is consistent under our conditions

$$\hat{\sigma}_{Y,Z} \xrightarrow{p} \sigma_{Y,Z} \text{ and } \hat{\sigma}_{Z,X} \xrightarrow{p} \sigma_{Z,X}$$

- Because we've assumed that $\sigma_{Z,X} \neq 0$, $\frac{\sigma_{Y,Z}}{\sigma_{X,Z}}$ is a continuous function, so, by CMT:

$$\hat{\beta}_1^{\text{IV}} = \frac{\hat{\sigma}_{Y,Z}}{\hat{\sigma}_{X,Z}} \xrightarrow{p} \frac{\sigma_{Y,Z}}{\sigma_{X,Z}} = \beta_1$$

Limiting Distribution of IV Estimators

- Continue to assume that $E[Y^4], E[X^4], E[Z^4] < \infty$. Then,

$$\sqrt{n}(\hat{\beta}_0^{\mathrm{IV}} - \beta_0) \xrightarrow{d} N(0, \sigma^2_{0,\mathrm{IV}})$$
$$\sqrt{n}(\hat{\beta}_1^{\mathrm{IV}} - \beta_1) \xrightarrow{d} N(0, \sigma^2_{1,\mathrm{IV}})$$

  where

$$\sigma^2_{1,\mathrm{IV}} = \frac{\mathrm{Var}[(Z - E[Z])U]}{\mathrm{Cov}(X, Z)^2}$$

- We'll omit the proof, as it is nearly line-by-line identical to the derivation of the limiting distribution of SLR

Inference on IV Estimators

- We'll need a consistent estimator of the variance of the limiting distribution to make it useful. Under the same conditions we've been using,

$$\hat{\sigma}^2_{1,\mathrm{IV}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2 \hat{U}_i^2}{\hat{\sigma}^2_{X,Z}}$$

is a consistent estimator for $\sigma^2_{1,\mathrm{IV}}$

Inference on IV Estimators

- Assume further that $\sigma^2_{1,\text{IV}} > 0$. Then, by Slutsky,

$$\frac{\sqrt{n}}{\hat{\sigma}_{1,\text{IV}}}(\hat{\beta}_1^{\text{IV}} - \beta_1) \xrightarrow{d} N(0,1)$$

- We can now proceed to do inference in the typical ways

Biasedness of IV Estimators

- What happened to unbiasedness?
- During the demonstration of unbiasedness of OLS, we exploited an assumption that $E[U|X] = 0$
- To do an analogous thing in the case of IV, we'd need to say $E[U|X, Z] = 0$. But, the LIE would then imply that $E[XU] = 0$, which is exactly what we are *avoiding* assuming when we use IV
- Does this matter? All the large sample properties are intact, so, if have "large" n, probably not

IV With Controls

- Suppose we have a causal model:

$$Y = \beta_0 + \beta_1 X + U$$

  where $E[XU] \neq 0$ and $E[ZU] \neq 0$ for potential instrument Z (assume Z is relevant, so $Cov(X, Z) \neq 0$)
- Then we can't use OLS or IV to recover β_1
- Now suppose we have a vector of controls C such that,

$$E[U|C, Z] = E[U|C] = \gamma' C$$

  for a vector of coefficients γ
- U is mean independent of Z *conditional* on C. Now we *can* get at causal β_1

IV With Controls

- Define a new error:

$$\begin{aligned} V &= U - E[U|C, Z] \\ &= U - E[U|C] \\ &= U - \gamma' C \end{aligned}$$

- Note that

$$\begin{aligned} E[V|C, Z] &= E[U - E[U|C, Z]|C, Z] \\ &= E[U|C, Z] - E[U|C, Z] \\ &= 0 \end{aligned}$$

so V is mean ind. of (C, Z), so it is also uncorrelated

IV With Controls

- Then, $U = V + \gamma' C$, so,

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + V + \gamma' C \\ &= \beta_0 + \beta_1 X + \gamma' C + V \end{aligned}$$

  and we have that V is uncorrelated with (C, Z)
- This will allow us to estimate a new version of IV with multiple regressors and an instrument

IV With Multiple Regressors (One Endogenous) and One Instrument

- Consider an equation:

$$Y = X'\beta + U$$

where $X = (1, X_1, ..., X_k)'$

- We have that $E[X_1 U] \neq 0$, but $E[X_j U] = 0$ for all $j \neq 1$ (say X_1 is *endogenous* and the other X_j's are *exogenous*)
- We have an instrument, Z, such that $E[ZU] = 0$
- Thus, we can say that for $W = (1, Z, X_2, ..., X_k)'$, $E[WU] = 0$ (Instr. Exog.)

IV With Multiple Regressors (One Endogenous) and One Instrument

- We need a slightly different version of instrument relevance in this context
- For the best linear predictor of X_1 given $(Z, X_2, ..., X_k)$:

$$X_1 = \pi_0 + \pi_1 Z + \pi_2 X_2 + ... + \pi_k X_k + V$$

  we need that $\pi_1 \neq 0$
- This means that Z still has some predictive value "controlling for" the other X_j's
- Along with assuming no perfect colinearity in W, this ensures that $E[WX']$ is invertible

IV With Multiple Regressors (One Endogenous) and One Instrument

- We have that $U = Y - X'\beta$ and $E[WU] = 0$. Combining these:

$$\begin{aligned} E[W(Y - X'\beta)] &= 0 \\ E[WY - WX'\beta] &= 0 \\ E[WY] - E[WX']\beta &= 0 \\ E[WY] &= E[WX']\beta \\ \Rightarrow \beta &= E[WX']^{-1}E[WY] \end{aligned}$$

- The last line is assured to be possible because of the new version of instrument relevance and no perfect colinearity in W

Estimating IV With Multiple Regressors (One Endogenous) and One Instrument

- We now discuss how to estimate parameters using IV from finite samples. We'll call these our maintained multiple regressor IV assumptions:
  (a) $E[WU] = 0$
  (b) No perfect colinearity in W
  (c) $E[WX']$ is invertible

  where (Y, X, Z, U) satisfy:

$$Y = X'\beta + U$$

- Also let $(Y_1, X_1, Z_1), ..., (Y_n, X_n, Z_n)$ be iid $\sim (Y, X, Z)$

Multivariate IV Estimator

- We've said that

$$\beta = E[WX']^{-1}E[WY]$$

so natural estimator of β is:

$$\hat{\beta}_n^{IV} = (\frac{1}{n}\sum_{i=1}^{n} W_i X_i')^{-1}\frac{1}{n}\sum_{i=1}^{n} W_i Y_i$$

- This is the (multivariate) instrumental variables estimator of β

Consistency of Multivariate IV

- Along with the maintained assumptions, say that $E[Y^4], E[Z^4] < \infty$ and $E[X_i^4] < \infty \; \forall \; i$. Then, the IV estimator is consistent,

$$\hat{\beta}_n^{IV} \xrightarrow{p} \beta$$

Consistency of Multivariate IV

- We have an iid sample and appropriate moment conditions, so, by WLLN

$$\frac{1}{n}\sum_{i=1}^{n} W_i X_i' \xrightarrow{p} E[WX'] \;\;\&\;\; \frac{1}{n}\sum_{i=1}^{n} W_i Y_i \xrightarrow{p} E[WY]$$

- We've assumed that $E[WX']$ is invertible (instr. rel.), so, by CMT

$$\hat{\beta}_n^{IV} = \Big(\frac{1}{n}\sum_{i=1}^{n} W_i X_i'\Big)^{-1} \frac{1}{n}\sum_{i=1}^{n} W_i Y_i \xrightarrow{p} E[WX']^{-1} E[WY] = \beta$$

Limiting Distribution of Multivariate IV

- $\sqrt{n}(\hat{\beta}_n^{IV} - \beta)$ has a limiting distribution as $n \to \infty$. That limiting distribution has a variance that we can estimate consistently. Using Slutsky, we can combine the limiting distribution of $\hat{\beta}_n^{IV}$ with the estimator of its variance to do inference
- All of the above looks nearly identical to inference for multivariate linear regression

IV With Controls Example

- Say we're interested in the effect of

$$X = \begin{cases} 1 \text{ go to charter HS} \\ 0 \text{ go to typical public HS} \end{cases}$$

on

$$Y = \begin{cases} 1 \text{ go to college} \\ 0 \text{ don't} \end{cases}$$

- For a causal model

$$Y = \beta_0 + \beta_1 X + U$$

we might thing $E[XU] \neq 0$ for a variety of reasons discussed before

IV With Controls Example

- Now we've got a lottery for charter schools, notated:

$$Z = \begin{cases} 1 \text{ win the lottery and have the option of charter school} \\ 0 \text{ lose the lottery} \end{cases}$$

- We can also see whether or not people enter the lottery, notated

$$C = \begin{cases} 1 \text{ enter the lottery} \\ 0 \text{ don't} \end{cases}$$

IV With Controls Example

- This looks a lot like our selection on observables example from MLR. However, there we made the (weird) assumption that *every* student who wins the lottery goes to charter school. Here, we're just saying winning the lottery gives you the *option* of charter school, but individual students might still opt out
- We'll now treat the lottery outcome as an instrument, conditioning on entering the lottery

IV With Controls Example

- Under the general discussion of IV with controls, we said that if we have

$$E[U|Z,C] = E[U|C] = \gamma_0 + \gamma_2 C$$

then for the new regression equation:

$$Y = \beta_0 + \gamma_0 + \beta_1 X + \gamma_2 C + V$$

it will be the case that

$$E[\begin{bmatrix} 1 \\ Z \\ C \end{bmatrix} V] = 0$$

- This is saying the knowing whether or not you win the lottery doesn't tell us anything about you, if we know you entered the lottery - seems reasonable

IV With Controls Example

- The above is one of the requirements that we need to estimate β^{IV} with multiple regressors consistently
- We also need that $\pi_1 \neq 0$ for a BLP,

$$X = \pi_0 + \pi_1 Z + \pi_2 C + \varepsilon$$

- This would be true - winning the lottery will help predict whether or not you enter the charter school even after "controlling" for entering the lottery

IV With Controls Example

- The last requirement is no perfect collinearity in $(1, Z, C)'$ - this is true so long as at least some students who enter the lottery lose
- Thus, we can estimate the vector of parameters $((\beta_0 + \gamma_0), \beta_1, \gamma_2)'$ consistently, including the causal parameter β_1