Instrumental Variables

Reading: Angrist and Pischke Chapter 1,
Wooldridge Chapter 15

Obstacles to Randomized Treatment

Randomizing Treatment

- Last week, we talked a lot about experiments.
- True experiments are the gold standard for establishing causal effects.
- The key is for our treatment variable to be randomly assigned (and therefore uncorrelated with the error), so OLS consistently estimates causal effects.
  - The treatment could be randomly assigned by us, or by nature.
- Unfortunately, randomizing the treatment is often not possible.
  - This is famously true in an academic setting: we have limited resources and a human subjects review board
  - Even in data savvy companies, experiments are much harder to run than you might expect.
- In fact, there are a wide range of obstacles that prevent researchers in both academia and industry from randomizing the treatment variable they are investigating.

Obstacle 1: Cost

- For a long time, researchers wanted to know how much smaller classroom sizes helped learning outcomes.
- The Tennessee Star experiment was a major research effort that randomly placed students into smaller classrooms.
  - Teachers were hired, students were shuffled around and tested, Follow-ups were conducted years later.
  - Participants: 6,500 students, about 80 schools
  - The cost was $12 million.
- Experiments are often cheaper in the realm of online services.
  - If you want to how customers respond to a new user layout, you can simply change the code.
  - But reconfiguring systems can be expensive and may disrupt normal operations
  - There’s also the time spent by the experiment designer.

Obstacle 2: Undelivered Treatments

- Sometimes, a you may try to administer a treatment to a group of people, but not everyone receives the treatment.
  - An online service company may select customers to receive ad impressions, but not all of them may visit your website.
  - The problem is especially well-known in clinical trials. Not all patients may take the medicine you give them.
- One solution is to change your x variable from treatment to intend-to-treat (which is random) and measure the intend-to-treat effect.
  - There are times when the intend-to-treat effect is what really matters. If we make a malaria medication available to a tropical population, we may balance the cost of making it available against the overall health outcomes.
  - But much of the time, it's the actual treatment effect that we care about – does a medicine work on people who take it? How many customers that see an ad click on it?

Obstacle 3: Treatment not under Direct Control

- Sometimes, one can only manipulate the treatment variable indirectly.
  - A company may be hesitant to change prices, and may experiment with promotions instead.
  - In a well-known (controversial) experiment by facebook, the company was interested in how negative posts would affect their users.
    - They couldn't make up negative posts themselves, so they had to alter the newsfeed algorithm.
- These are cases where a company can take an action that affects the treatment variable, but there are other factors that also affect the treatment.

Obstacle 4: Manipulation Unethical or Disruptive to Product.

- In many cases, companies rely heavily on user-generated data and won't be able to manipulate it without breaching the trust of their users.
  - The facebook experiment on negative and positive posts is an example.
  - A dating site may want to understand the dating preferences of its users, but it can't create fake profiles to see how users react
  - A movie distributor might be interested in understanding the effects of movie characteristics on user ratings, but can't create new movies.
  - Wikipedia may be interested in what type of writing to promote to improve the editing process, but can't generate fake articles for the live site.

Identification Strategies

- For all these reasons, you'll often find that you can't randomly assign treatment.
- This week, we'll continue to explore strategies we can use to measuring causal effects when experiments are impossible.

Questions about Questions

Approaching a Causal Investigation

- Let’s say that we’ve decided that a true experiment is impossible.
- How do we approach the problem of measuring causal effects?
- Angrist and Pischke provide a set of 4 frequently asked questions (FAQs) that can help guide a research effort
  - These are the types of questions that you should ask before undertaking a study
    - Many researcher don’t, but they probably should
  - They are designed to clarify pitfalls before you gather data.

Question 1

- What is the causal effect of interest?
  - What goes into our x-variable?
    - And what is held constant and should go in to other variables or the error term?
  - Our x could be something we're thinking of manipulating directly
    - And we want to choose the optimal value
  - Or it could be a variable that we can't directly manipulate
    - Perhaps we want to understand our customer's behavior.
- The causal effect will be the centerpiece that we build our investigation around

Question 2

- What is the ideal experiment?
  - We just explained that we are often unable to execute a true experiment.
  - Even so, you should try to imagine one.
    - Imagine (just for a moment) you have an infinite budget and no legal or ethical constraints.
  - Why is this important?
    - It helps us choose interesting research directions
      - Think about the famous Stanford Prison Experiments conducted by Phillip Zimbardo.
        - » Zimbardo randomly assigned subjects to act as prisoners or prison guards, in order to see how they would respond to their roles.
        - » The experiment was traumatizing to the participants, and it would never be allowed by modern academic review committees.
        - » Even so, the questions raised by the study, about human responses to authority, are fascinating as well as pressing – and we may want to study them using more ethical methods.
    - It helps us clarify what our x variable is.
      - In designing an experiment, we have to state what we're manipulating, and this becomes our x variable.
    - If you can't answer a question with an ideal experiment, you probably can't do it at all.

Question 2

- Examples of finding ideal experiments:
  - Political scientists are interested in whether democratic governments are better for economic growth than authoritarian ones.
    - We can't answer this with a naïve regression, of course, because a country's type of government is a highly endogenous variable.
      - It could be correlated with past growth, natural resources, war, income inequality…
    - We might still be able to imagine an experiment. Angrist and Pischke suggest going back in time to install random government types in former colonies on their independence day.
      - Since we can imagine this experiment, however improbable it is, there's at least a chance we can find another way to answer the question.
  - We might want to know what effect being biologically Male has on your personality at adulthood.
    - The problem is that we can't just switch a person's chromosomes, without also changing their upbringing, the way people treat them, and so forth.
    - It seems we haven't properly identified a treatment variable.
    - Angrist and Pischke have a (cheeky) name for a question that we can't address with an ideal experiment: a Fundamentally Unanswerable Question (FUQ)
      - While these may seem interesting, you can't answer them with any dataset

Question 3

- What is the identification strategy?
- An ***identification strategy*** is what we call our plan to consistently measure a causal effect.
  - It's our plan to overcome endogeneity.
  - A randomized experiment would be an identification strategy.
  - Another one that we looked at last week was difference-in-difference
    - This identification strategy works when confounding factors are constant over time and we have two measurements for each individual.
- This step in analysis is where we need to be especially clever.
  - We'll usually need domain knowledge, and a fair bit of creativity to find an identification strategy.
  - If you can find the right identification strategy, questions that have eluded us for years may suddenly become answerable.
  - It's one of the most exciting aspects of empirical work.
- This week, we'll cover the most important tool for identification in modern empirics – instrumental variables.

Question 4

- What is the mode of statistical inference?
- These are nuts-and-bolts questions
  - What is the population?
  - How do we choose our sample?
  - What assumptions are we making in our model?
    - Based on these, what properties will our estimates have?

- By answering these four FAQs, we think that you'll be well positioned to undertake your investigation

Defining Instrumental Variables

Introducing Instrumental Variables

- This week, we'll look at the most important tool for identification: instrumental variables (IV)
- To use IV estimation, we don't need the treatment itself to be random or exogenous.
- All we need is some factor that contributes exogenous or random variation to our treatment variable.
  - It could be variation that comes from nature, or it could come from the researcher.
  - So we know that part of the variation in our treatment variable is exogenous.
  - This is what we call an instrument.
- It turns out that this is enough to estimate our causal effect
  - IV estimation tremendously expands the set of scenarios in which we can measure causal effects.

Instrument Definition

- Let's start with a model that has one endogenous variable.

  $y = \beta_0 + \beta_1 x + u$ where $cov(x,u) \neq 0$

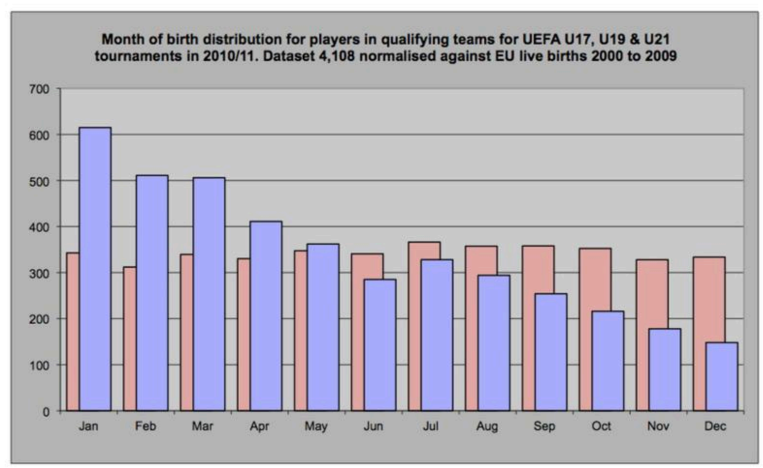
- A variable z is an instrument for x if:
  - 1) It does not appear in the regression
  - 2) $cov(z,x) \neq 0$
    - We want it to be highly correlated with x
  - 3) $cov(z,u) = 0$
    - z is uncorrelated with the error. This means that if we were to add z to our regression, its coefficent would be zero.
    - For this reason, we call this an exclusion restriction.

Example Instrument

- Let's start with the wage equation.
  $$\log(wage_i) = \beta_0 + \beta_1 educ_i + u_i$$
- One instrument for education from a famous paper by Angrist and Krueger is quarter of birth.
- Assumptions:
  - Quarter of birth is not in our model
  - Is quarter of birth correlated with educ?
    - Turns out that it is. The reason is that many states require students to enroll in school during the year they turn 6.
      - A student born Dec 31 enters school almost a year younger than a student born Jan 1.
    - Additionally, states usually require students to stay in school until they turn 16.
    - This means that students born in the first quarter tend to get less education than students born in the fourth quarter
      - You can confirm this with a naïve regression, the difference is about .15 years on the average.
  - Is quarter of birth uncorrelated with u?
    - The idea is that the quarter you were born in seems pretty random from your perspective.
    - It shouldn't be too correlated with genetics, parental achievement, wealth, or many other confounding covariates we can think of.

Doubt on the Quarter-of-Birth Instrument

- The Angrist and Krueger paper was groundbreaking.
- Since 1991, the evidence has been thoroughly examined and debated
- People have questioned whether qob is a valid instrument.
- One problem is that quarter of birth influences outcomes in many ways we might not expect.
    - A student that enters school at a higher age is more developed, may have an easier time at school.
    - Moreover early differences can build up as teacher devote more energy to students that do better.
    - By the time students are leave school, the differences may be dramatic.
- Here, you can see the distribution of birth month for kids in European football leagues



Derivation of IV Estimator

10 minute Lightboard

- How can we use an instrumental variable to estimate our coefficients?
- First, let‘s see a simple proof that OLS is consistent when x is exogenous

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$Cov(x_i, u_i) = 0 \qquad \text{(Exogeneity)}$$

$$\Leftrightarrow \; 0 = Cov(x_i, y_i - \beta_0 - \beta_1 x_i) = Cov(x_i, y_i) - \beta_1 Var(x_i)$$

$$\Leftrightarrow \; \beta_1 = Cov(x_i, y_i) / Var(x_i)$$

$$\Rightarrow \; \hat{\beta}_1 = \widehat{Cov}(x_i, y_i) / \widehat{Var}(x_i) \; \to \; Cov(x_i, y_i) / Var(x_i) = \beta_1$$

This holds as long as the data are such that sample variances and cova- riances converge to their theoretical counterparts as n $\to 1$, i.e. if a LLN holds. <u>OLS will basically be consistent if, and only if, exogeneity holds.</u>

- **Assume existence of an instrumental variable z:**

$\boxed{Cov(z_i, u_i) = 0}$ (but $Cov(x_i, u_i) \neq 0$)

$$\Leftrightarrow \; 0 = Cov(z_i, y_i - \beta_0 - \beta_1 x_i) = Cov(z_i, y_i) - \beta_1 Cov(z_i, x_i)$$

$$\Leftrightarrow \; \beta_1 = Cov(z_i, y_i)/Cov(z_i, x_i)$$

$$\Rightarrow \; \widehat{\beta}_{IV} = \widehat{Cov}(z_i, y_i)/\widehat{Cov}(z_i, x_i) \; \to \; Cov(z_i, y_i)/\boxed{Cov(z_i, x_i)} = \beta_1$$

The instrumental variable is un-correlated with the error term

<u>IV-estimator</u>: $\widehat{\beta}_{IV} = \dfrac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}$

The instrumental variable is correlated with the explanatory variable

IV Estimation

Example of IV estimation

- Example from Wooldridge: Father's education as an IV for education

$$\log(wage_i) = \beta_0 + \beta_1 educ_i + u_i$$

- First, look at the regular (naive) OLS esimate:

<u>OLS</u>: $$\widehat{\log}(wage) = \underset{(.185)}{-\ .185} + \underset{(.014)}{.109}\ educ$$

- Because of omitted variable bias, we supect our coefficient is too high.
- Can we use father's education as a IV?
- First see if it's correlated to educ by regressing educ on fatheduc:

$$\widehat{educ} = \underset{(.28)}{-\ 10.24} + \underset{(.029)}{.269}\ fatheduc$$

- Looks good, there's a strong relationship, over 25%.

Example of IV estimation

- What about our exclusion restriction?
  - Is fatheduc uncorrelated with u?
  - Equivalently, can we safely exclude it from the regression?
- This is very questionable. You might argue that fatheduc is unrelated to ability at birth.
  - But the way one is brought up surely depends on father's education in many ways.
  - Fatheduc is correlated with father's ability, and highly able fathers might motivate their children to be high achievers.
  - I'd call this a poor instrument, and we'll discuss the implications in a moment.
- Here are the results from IV estimation:

$$\underline{\text{IV}}:\ \widehat{\log}(wage) = \underset{(.446)}{.441} + \underset{(.035)}{.059}\ educ$$

$$n = 428, R^2 = 1 - RSS_{IV}/TSS = .093$$

- Notice 1) effect is smaller than before.
- 2) standard error is larger
- 3) R^2 is smaller – it has to be, since OLS maximizes R^2

Other Instruments for Education

- **Other IVs for education that have been used in the literature:**
- The number of siblings
  - Correlated with education because of household resource constraints
  - Uncorrelated with innate ability, at least at birth. (but what about upbringing? Parent's characteristics?)
  - More advanced papers use twin births as instrumental variables – twin births are essentially random, so at least uncorrelated with parent's traits.
- College proximity when 18 years old
  - Correlated with education because more education if lived near college
  - Uncorrelated with error? Problem if areas near college are richer, etc…
- Month of birth
  - Correlated with education because of compulsory school attendance laws
  - Uncorrelated with most factors we can think of at birth.

Poor Instruments

- What happens if we have a less-than-perfect instrument?
  - Let's see what the bias of IV estimation is. Solution from Wooldridge:

$$plim\ \widehat{\beta}_{1,IV} = \beta_1 + \frac{Corr(z,u)}{Corr(z,x)} \cdot \frac{\sigma_u}{\sigma_x} \quad = \beta_1 + \frac{\mathrm{cov}(z,u)}{\mathrm{cov}(z,x)}$$

  - Numerator: correlation between the instrument and the error
  - Denominator: correlation between instrument and x
  - So the ratio matters.
  - Even a small correlation between z and u can be bad if the instrument is only weakly correlated with x.
- This is a potential problem with the season-of-birth instrument.
  - Only a small number of students are affected by mandatory schooling laws.
  - So even a small correlation with the error could significantly bias the estimate.

Choosing IV or OLS

- If we have a poor instrument, we need to choose between IV and OLS.
- We know that our error will be larger for IV. Let's compare the biases:

$$plim\ \widehat{\beta}_{1,IV} = \beta_1 + \frac{Corr(z,u)}{Corr(z,x)} \cdot \frac{\sigma_u}{\sigma_x}$$

$$plim\ \widehat{\beta}_{1,OLS} = \beta_1 + Corr(x,u) \cdot \frac{\sigma_u}{\sigma_x}$$

<u>IV worse than OLS if</u>: $\frac{Corr(z,u)}{Corr(z,x)} > Corr(x,u)$

- On the left, we have the ratio we just talked about. On the right, we have correlation between x and u - how much endogeneity do we have?
- We never know these correlations, but at least you can reason about them.
  - It's common practice to report both IV and OLS estimates.

Question

- Think about a dataset or research project you're familiar with, and describe a situation in which endogeneity bias prevents you from measuring a causal effect.
- If you can, suggest a potential instrument for your endogenous variable

- <2U: this should be a free text or video response from students, and they should be able to read each other's responses>

IV Estimation in Multiple Regression

IV in Multiple Regression

- Here's a multiple regression model. For now, let's assume only one variable is endogenous, y_2.

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + u_1$$

- The conditions for an instrumental variable, z_k, are similar to the simple regression case:
  - 1) Does not appear in regression equation
  - 2) Is uncorrelated with error term
  - 3) Is correlated with endogenous explanatory variable - after controlling for the other exogenous variables.
    - In a regression of the endogenous explanatory variable on all exogenous variables, the instrumental variable must have a non-zero coefficient.

$$y_2 = \pi_0 + \pi_1 z_1 + \cdots + \pi_k z_{k-1} + \pi_k z_k + v_2$$

This is called the reduced form regression.

IV in Multiple Regression

- The IV estimator can be computed much like in the simple regression case
- We write down the exogeneity and zero-mean conditions for each z:

$Cov(z_j, u_1) = 0, \ j = 1, \ldots, k$ as well as $E(u_1) = 0$

- Applying the method of moments, we write down the sample analogs

$$n^{-1} \sum_{i=1}^{n} z_{ij} \hat{u}_{i1} = \widehat{Cov}(z_j, \hat{u}_1) = 0, \ j = 1, \ldots, k$$
$$n^{-1} \sum_{i=1}^{n} (y_{i1} - \widehat{\beta}_0 - \widehat{\beta}_1 y_{i2} - \widehat{\beta}_2 z_{i1} - \cdots - \widehat{\beta}_k z_{ik-1}) = n^{-1} \sum_{i_1}^{n} \hat{u}_{i1} = 0$$

- At this point, we have k+1 equations, with k+1 parameters, so we can solve for them.
  - Of course, we'll do this automatically in R, so will be much easier.
- Next we'll look at another way to compute the IV estimator…

Two-Stage Least Squares

Two-Stage Least Squares

- It turns out that the IV estimator is equivalent to the following procedure:
- Here's our population model with one endogenous variable.

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + u_1$$

- As we did before, regress the endogenous variable on all exogenous variables, including the instrument
  - This is called the first stage, or the reduced form.

$$\widehat{y}_2 = \widehat{\pi}_0 + \widehat{\pi}_1 z_1 + \cdots + \widehat{\pi}_k z_{k-1} + \widehat{\pi}_k\, z_k$$

- In the second stage, we take the predicted values from the first stage and put them into our main regression:

$$y_1 = \beta_0 + \beta_1 \widehat{y}_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + error$$

Two-Stage Least Squares

- Why does Two Stage Least Squares work?
  - In the first stage, we regress y_2 on all exogenous variables. What we're left with is part of the variation in y_2, but we know that all of this variation is exogenous.
    - In other words, we're purging y_2 of all endogenous information
    - None of the variation left is related to the error term.
  - When we put the predicted values of y_2 into the second stage, we have a regression with all exogenous variables, so OLS is consistent.
  - We've replaced y_2 by the part of y_2 we know is exogenous, then we see how the outcome responds to this variation.
    - We have some exogenous variation in y_2, so this is what we use to estimate the effect.
- Properties of Two Stage Least Squares
  - The standard errors from the OLS second stage regression are wrong. However, it is not difficult to compute correct standard errors.
  - If there is one endogenous variable and one instrument then 2SLS = IV
  - The 2SLS estimation is quite general: It can be used if there is more than one endogenous variable and at least as many instruments

TSLS Example

- Wooldridge gives an example of estimating returns to education using both father's eduation and mother's education as instruments
  - These are poor instruments, but it does illustrate the point.
- In the first stage, we regress educ on all exogenous variables.

$$\widehat{educ} = \underset{(.27)}{8.37} + \underset{(.026)}{.085}\ exper - \underset{(.001)}{.002}\ exper^2$$

$$+ \underset{(.024)}{.185}\ \boxed{fatheduc} + \underset{(.026)}{.186}\ \boxed{motheduc}$$

Education is significantly partially correlated with the education of the parents

- In the second stage, we regress wage, using our predicted values of educ.

$$\widehat{log}(wage) = \underset{(.400)}{0.48} + \underset{(.031)}{.061}\ educ + \underset{(.013)}{.044}\ exper - \underset{(.0004)}{.0009}\ exper^2$$

- The results are similar to what we had before, but errors have decreased a bit.

Properties of IV Methods

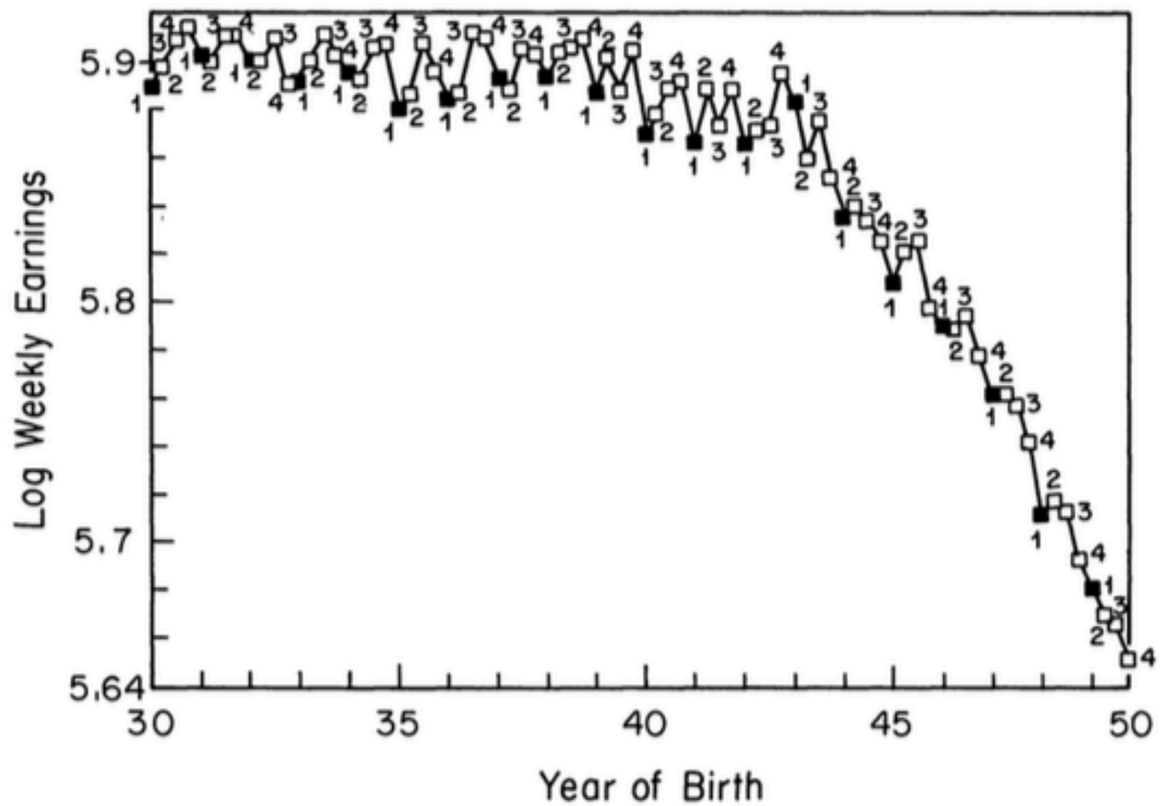
- **Statistical properties of 2SLS/IV-estimation**
  - Under assumptions completely analogous to OLS, but conditioning on z_i rather than on x_i, 2SLS/IV estimators are consistent and asymptotically normal
  - 2SLS/IV is typically much less precise because we're working with much less variation in our explanatory variable
    - R will automatically compute our standard errors for us.
  - Corrections for heteroskedasticity are available much as in OLS.
  - 2SLS/IV easily extends to time series and panel data situations

Quarter of Birth Example

Quarter of Birth

- A better example comes from the Angrist and Krueger quarter of birth study.
  - This is a famous study that highlights many of the concepts we've been talking about
- They use census data, which includes which quarter of the year each individual was born in.
  - Quarter 1 is January through March, Quarter 2 is April through June, and so on.
  - A problem is that this is an ordinal variable, so we wouldn't want to put it in directly.
- Instead, we can create three dummy variables, for quarter 2, quarter 3, and quarter 4, and use each of these as a separate instrumental variable.
  - Angrist and Krueger actually create a separate dummy variable for each quarter of each year, but we'll just use 3 to keep the presentation simple
- Is quarter of birth a good instrument?
  - First we have to check if it is actually related to education
  - And we need to argue that it doesn't affect wages in any other way.

- In this figure, we can see the weekly wage of individuals by quarter of birth.
  - The black squares are all first quarter.
  - Notice that these are consistently lower than the others.
- Also notice the declining trend for men born in the 40's.
  - This could be because these men have less experience in the work force.
  - To deal with this complication, Angrist and Krueger focus on men born in the 30's.



- Does qob affect earnings in other ways?
- Potential is clearly there.
- But angrist and krueger look at the subset of men who graduate college.
  - These are a natural control group, because quarter of birth doesn't influence how much education they receive.
  - The authors note that qob doesn't seem to influence earnings for this group.

First Stage Regression

- To see how quarter of birth influences earnings, we can run the first stage regression
  - This is the regression of years of education on the quarter of birth dummies.
  - Here are the results for the 30’s cohort

$$\text{Educ} = \underset{(0.011)}{12.7} + \underset{(.016)}{.057}\,Q2 + \underset{(.016)}{.117}\,Q3 + \underset{(.016)}{.151}\,Q4$$

- Notice that men born in the 4^{th} quarter get over a tenth of a year more education, on the average.

IV Estimation

- Now we could take the predicted values from our first stage regression, and use them to estimate the effect of educ on wages.

Log(wage) = 4.58 + 0.10 educ^

(0.25) (0.02)

- Suprisingly, this is not too different from estimates we get from a naïve ols regression, which should include endogeneity bias.
  - We'll touch on possible reasons for this later.

Angrist and Krueger Results

TABLE V

OLS AND TSLS ESTIMATES OF THE RETURN TO EDUCATION FOR MEN BORN 1930–1939: 1980 CENSUS[a]

Independent variable	(1) OLS	(2) TSLS	(3) OLS	(4) TSLS	(5) OLS	(6) TSLS	(7) OLS	(8) TSLS
Years of education	0.0711	0.0891	0.0711	0.0760	0.0632	0.0806	0.0632	0.0600
	(0.0003)	(0.0161)	(0.0003)	(0.0290)	(0.0003)	(0.0164)	(0.0003)	(0.0299)
Race (1 = black)	—	—	—	—	−0.2575	−0.2302	−0.2575	−0.2626
					(0.0040)	(0.0261)	(0.0040)	(0.0458)
SMSA (1 = center city)	—	—	—	—	0.1763	0.1581	0.1763	0.1797
					(0.0029)	(0.0174)	(0.0029)	(0.0305)
Married (1 = married)	—	—	—	—	0.2479	0.2440	0.2479	0.2486
					(0.0032)	(0.0049)	(0.0032)	(0.0073)
9 Year-of-birth dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8 Region-of-residence dummies	No	No	No	No	Yes	Yes	Yes	Yes
Age	—	—	−0.0772	−0.0801	—	—	−0.0760	−0.0741
			(0.0621)	(0.0645)			(0.0604)	(0.0626)
Age-squared	—	—	0.0008	0.0008	—	—	0.0008	0.0007
			(0.0007)	(0.0007)			(0.0007)	(0.0007)
χ^2 [dof]	—	25.4 [29]	—	23.1 [27]	—	22.5 [29]	—	19.6 [27]

a. Standard errors are in parentheses. Sample size is 329,509. Instruments are a full set of quarter-of-birth times year-of-birth interactions. The sample consists of males born in the United States. The sample is drawn from the 5 percent sample of the 1980 Census. The dependent variable is the log of weekly earnings. Age and age-squared are measured in quarters of years. Each equation also includes an intercept.

The Wald Estimator

12 minute lightboard

The Wald Estimator

- In many cases, our instrumental variable is binary.
- Examples: Vietnam War Draft Status, Treatment versus placebo in a clinical trial, Twin births
- When this happens, our IV estimator takes on an elegant form.
- Suppose we have population model
- $y = \beta_0 + \beta_1 x + u$ where x is endogenous.
- Suppose z is a binary instrumental variable with values 0 and 1.

  $\text{cov}(z, x) \neq 0$, $\text{cov}(z, u) = 0$. This implies $E(u \mid z=0) = E(u \mid z=1) = 9$
- Alternate way to compute IV in this setting: take expectations of both sides,

  $E(y \mid z = 0) = E(\beta_0 + \beta_1 x + u \mid z = 0) = \beta_0 + \beta_1 E(x \mid z = 0) + E(u \mid z = 0)$

  $= \beta_0 + \beta_1 E(x \mid z = 0)$
- Similarly, $E(y \mid z = 1) = \beta_0 + \beta_1 E(x \mid z = 1)$
- Subtract to get, $E(y \mid z = 1) - E(y \mid z = 0) = \beta_1 E(x \mid z = 1) - \beta_1 E(x \mid z = 0)$
- Or $\beta_1 = [E(y \mid z = 1) - E(y \mid z = 0)] / [E(x \mid z = 1) - E(x \mid z = 0)]$
- You can check that this is exactly the IV estimator, but in this case we call it the Wald estimator.

The Wald Estimator

- β_1 = [E(y | z = 1) - E(y | z = 0)] / [E(x | z = 1) - E(x | z = 0)]
- The wald estimator is easy to understand. Z is an exogenous source of variation.
- In the denominator, we have the effect on x from changing z from 0 to 1.
- In the numerator, we have the effect on y from the same change.
  - And the exclusion restriction tells us that the only way z affects y is through x.
  - We have a causal chain z -> x -> y.
- The ratio of these effects is x’s effect on y.

Wald Estimator Example

- An example of the Wald Estimator comes from the Angrist and Krueger study on schooling.
- Model is ln(weekly wage) = $\beta_0 + \beta_1 educ + u$
- Let's compare men born in the first quarter, $q = 1$, to men born in other quarters, $q > 1$.
- We compute

$$[E(\ln(wage) \mid q = 1) - E(\ln(wage) \mid q > 1)] /$$
$$[E(educ \mid q = 1) - E(educ \mid q > 1)]$$
$$= -.011 / -.11 = .10199$$

- This is the Wald estimate of the wage caused by a year of schooling.
  - A year of schooling seems to increase wages by approximately 9%.

Local Average Treatment Effect

Heterogeneous Treatment Effects

- There’s one detail that we’ve ignored up till this point.
- We’ve been computing treatment effects, for example, the effect of an extra year of schooling on wages.
- You might wonder, who’s wages, exactly?
  - Our population model assumes the increase in wage is the same for everybody.
  - More realistically, the increase in wages might be quite different from one person to the next
- This is known as a heterogeneous treatment effect.
- To compare different studies, and to understand the meaning of your own results, it’s important to understand how IV estimation works in this setting.

Binary Treatment, Binary Instrument

- Let's look at a simple case. Suppose our treatment variable is binary, D.
- Suppose we have a binary instrument z for D.
- Suppose both are 0-1 variables.
- Example:
  - z indicates giving a drug to a patient, D represents the patient actually taking the drug.
  - z represents being in the Vietnam war draft, D represents going to the Vietnam war.
  - z represents sending a coupon to a customer and D represents the customer showing up at your store.
- As some of these examples show, the instrument could come from nature, or it could be a case where you can deliberately manipulate something, but not the treatment directly.

- For each individual i, we write D_{0i} for the value of the treatment if z = 0, D_{1i} if z= 1.
- Let's add an assumption of monotonicity, so $D_{1i} \geq D_{0i}$ for all i.
- This means that we can divide people into three groups:

1. Always-takers: $D_{0i} = D_{1i} = 1$
2. Never-takers: $D_{0i} = D_{1i} = 0$
3. Compliers: $D_{0i} = 0$, $D_{1i} = 1$

- Notice that the monotonicity assumption means there are no defiers – people with $D_{0i} = 1$, $D_{1i} = 0$.

- The Wald estimator looks like

$[E(y_i \mid z = 1) - E(y_i \mid z = 0)] / [E(D_i \mid z=1) – E(D_i \mid z=0)]$

- Now let y_{0i} be the value of the outcome for individual i if treatment D were 0.
- Let y_{1i} be the value if D were 1.
  - These are potential outcomes, only one is every observed.
- The numerator of the Wald estimator can be written
- $E(y_i \mid z = 1) - E(y_i \mid z = 0) = E(y_{0i} + (y_{1i} – y_{0i})D_{1i}) – E(y_{0i} + (y_{1i} – y_{0i})D_{0i}) = E((y_{1i} – y_{0i})(D_{1i} - D_{0i}))$
  - This is the change in the treatment due to a change in z, times the change in y due to a change in the treatment.
- In fact, $D_{1i} - D_{0i}$ is only non-zero for compliers, so we can write this as
- $E(y_{1i} – y_{0i} \mid D_{1i} > D_{0i}) \Pr(D_{1i} > D_{0i})$
- Similarly, the denominator can be computed as $\Pr(D_{1i} > D_{0i})$
- Putting these together, we have a famous result…

LATE Theorem

- LATE Theorem:
  - Under the assumptions stated, the Wald estimator for the effect of D is $E(y_{1i} - y_{0i} \mid D_{1i} > D_{0i})$
- This says that the Wald estimator computes the average effect of treatment on the population of compliers.
- This is known as the Local Average Treatment Effect (LATE)
- So the never-takers and the always-takers don't affect the result. We only measure the people that are somehow affected by our instrument.
- LATE is not the same as the average causal effect of treatment on the treated.
- $E(y_{1i} - y_{0i} \mid D_i = 1) = E(y_{1i} - y_{0i} \mid D_{0i} = 1)\, Pr(D_{0i} = 1 \mid D_i = 1)$
  $\quad + E(y_{1i} - y_{0i} \mid D_{1i} > D_{0i})\, Pr(D_{1i} > D_{0i} \mid D_i = 1)$
- Here's an equation to explain this.
  - On the left, we have the average causal effect, given that a person is treated, $D_i = 1$.
  - On the right, we break this population into two groups. First, a person could be treated because they are an always-taker, second it could be because they are a complier and happen to get the instrument.
  - So we have to average the effect on the always-takers together with LATE
- If you are lucky enough to have a situation with no always-takers, LATE is the same as the effect on the treated.

LATE for Metric Treatments

- In the more general case, we have a metric treatment instead of a binary one.
- The math behind the late theorem is more complex here, but the intuition is the same.
- The effect we measure by IV is a weighted average of the effects on compliers – people that are affected by the instrument.
- Example: If the Angrist Krueger study of schooling, the compliers are the individuals that were affected by season of birth.
  - In other words, these are people that dropped out of school around their 16th birthday, so that the quarter of their birth influenced how much schooling they received.
  - One might argue that this is a very specific subpopulation.
  - How confident are you that the results generalize to the wider population of students?
- So local average treatment effect is an important concept to keep in mind when you see IV methods.

R example of IV Estimation