Impact Evaluation Methods

Topic 6: Instrumental Variables

Alex Alekseev

University of Regensburg, Department of Economics

Previously on *Impact Evaluation Methods*...

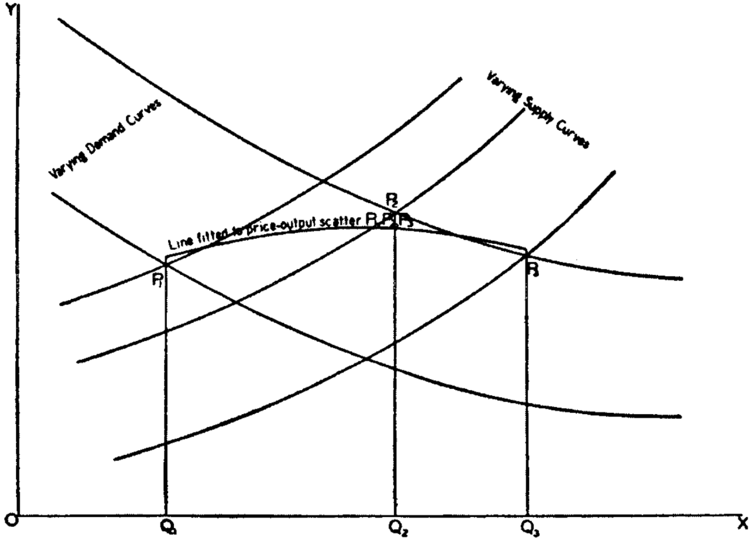
- Matching practice
- Regression

Background

History of IV

- First developed in the 1920s by father and son Phillip and Sewall Wright who were studying agricultural markets
- A 1928 book about animal and vegetable oils and tariffs
- The challenge: how to estimate the supply and demand curves
- The equilibrium prices and quantities are determined simultaneously
- When we see a scatterplot of prices and quantities, how can we use it to infer the supply and demand curves?
- Without further assumptions, we cannot say much

FIGURE 4. PRICE-OUTPUT DATA FAIL TO REVEAL EITHER SUPPLY OR DEMAND CURVE.



Instrumental Variables

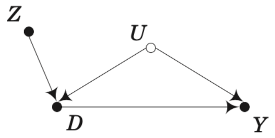
- The method proposed by Phillip Wright (and inspired by his son's work on path analysis) solves the simultaneity problem
- It uses the variables that appear in one equation to "shift" this equation and trace out the other
- These variables are now called the **instrumental variables**
- Stock and Trebbi (2003) provide a historical perspective on the development of IV by Phillip and Sewall Wright
- The modern use of IV still uses the language developed for the simultaneous equations models application
- However, the actual use is different

Measurement Error

- IV methods were used to solve the problem of bias from measurement error in regression models (Wald, 1940)
- In linear models, a regression coefficient is biased towards zero when the regressor of interest is measured with random errors
- IV methods can eliminate this bias

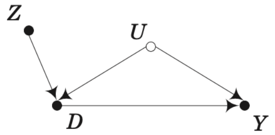
Homogeneous Treatment Effects

IV DAG



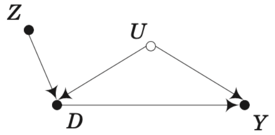
- We are interested in the causal effect of D on Y
- There is a backdoor path $D \leftarrow U \rightarrow Y$
- U is unobserved so we cannot close this backdoor path
- This is a situation of **selection on unobservables**
- There is no conditioning strategy that will satisfy the backdoor criterion.

Isolating Variation



- D varies because of Z and U
- Under TE homogeneity, it is not necessary to relate all of the variation in D to all of the variation in Y
- The covariation in D and Y that is generated by U can be ignored...
- ...if we can find a way of isolating the variation in D and Y that is causal

Using IV



- There is a mediated path from Z to Y via D
- When Z varies, D varies, which causes Y to vary
- Importantly, Y is only varying because D has varied
- There is another path from Z to Y via U
- But D is a collider along that path
- For this to work, Z has to cause Y only through D

Closing Back-Door Paths for IV

- In general, any paths between the instrument Z and the outcome Y must either pass through the treatment D or be closed
- There might be some common causes of Z and Y or mediated paths between Z and Y that do not go through D...
- ...but they must be closed
- IV moves that responsibility of closing the back doors from the treatment to the instrument

Binary Treatment

- Consider now a simple regression model with homogeneous treatment effects, in which both D and Z are binary
- Z causes D but is independent of ϵ

$$Y = \alpha + \delta D + \epsilon.$$

- We assume that D is correlated with ϵ and that we cannot fix this issue by any conditioning strategy
- Let's compute the expected value of Y conditional on Z:

$$\mathbb{E}[Y \mid Z = 1] = \alpha + \delta \mathbb{E}[D \mid Z = 1] + \mathbb{E}[\epsilon \mid Z = 1]$$
$$\mathbb{E}[Y \mid Z = 0] = \alpha + \delta \mathbb{E}[D \mid Z = 0] + \mathbb{E}[\epsilon \mid Z = 0].$$

The Wald Estimator

- Taking the difference between the two, we get

$$\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0] =$$
$$\delta\left(\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0]\right) + \mathbb{E}[\epsilon \mid Z = 1] - \mathbb{E}[\epsilon \mid Z = 0].$$

- Notice that the difference $\mathbb{E}[\epsilon \mid Z = 1] - \mathbb{E}[\epsilon \mid Z = 0] = 0$ since Z and ϵ are independent
- Hence, we have

$$\delta = \frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0]}.$$

- The sample analogue of this quantity is called the **Wald estimator**

Reduced Form and First Stage

$$\delta = \frac{\mathbb{E}[Y \mid Z = 1] - \mathbb{E}[Y \mid Z = 0]}{\mathbb{E}[D \mid Z = 1] - \mathbb{E}[D \mid Z = 0]}.$$

- We can estimate the causal effect δ by simply dividing the NTE of Z on Y by the NTE of Z on D
- The NTE of Z on Y is called the **reduced form**
- The NTE of Z on D is called the **first stage**
- This is due to the language of a two-stage least squares (2SLS) estimator we will introduce later

Math time: Private schools and vouchers

- Causal effect of attending a private school (D) on academic achievement (Y)
- A simple regression of Y on D - biased estimate - unobserved confounders, like family background, that affect both going to a private school and achievement
- A voucher program that randomly assigns vouchers for attending a private school
- We can use vouchers as an IV (Z)

Math time: Private schools and vouchers

- Assume the following joint distribution of Z and D, $\mathbb{P}(D, Z)$

	$D = 0$	$D = 1$	$\mathbb{P}(Z)$
$Z = 0$	0.8	0.1	0.9
$Z = 1$	0.08	0.02	0.1
$\mathbb{P}(D)$	0.88	0.12	

- Overall, only 12% of the students attend a private school and vouchers are assigned to 10% of the students
- Notice that not everyone who receives a voucher goes to a private school

Math time: Private schools and vouchers

	$D = 0$	$D = 1$	$\mathbb{P}(Z)$
$Z = 0$	0.8	0.1	0.9
$Z = 1$	0.08	0.02	0.1
$\mathbb{P}(D)$	0.88	0.12	

- It will be helpful to compute the conditional distribution of $\mathbb{P}(D \mid Z)$ using the Bayes rule:

$$\mathbb{P}(D \mid Z) = \frac{\mathbb{P}(D, Z)}{\mathbb{P}(Z)}$$

	$Z = 0$	$Z = 1$
$\mathbb{P}(D = 0 \mid Z)$	$8/9$	$4/5$
$\mathbb{P}(D = 1 \mid Z)$	$1/9$	$1/5$

Math time: Private schools and vouchers

	$Z = 0$	$Z = 1$
$\mathbb{P}(D = 0 \mid Z)$	$8/9$	$4/5$
$\mathbb{P}(D = 1 \mid Z)$	$1/9$	$1/5$

- Now assume the following expected outcomes, $\mathbb{E}[Y \mid D, Z]$

	$D = 0$	$D = 1$
$Z = 0$	50	60
$Z = 1$	50	58

Math time: Private schools and vouchers

	$Z = 0$	$Z = 1$
$\mathbb{P}(D = 0 \mid Z)$	8/9	4/5
$\mathbb{P}(D = 1 \mid Z)$	1/9	1/5

- Now assume the following expected outcomes, $\mathbb{E}[Y \mid D, Z]$

	$D = 0$	$D = 1$
$Z = 0$	50	60
$Z = 1$	50	58

- $$\begin{aligned} \mathbb{E}[Y \mid Z = 1] &= \mathbb{E}[Y \mid Z = 1, D = 0]\mathbb{P}(D = 0 \mid Z = 1) \\ &\quad + \mathbb{E}[Y \mid Z = 1, D = 1]\mathbb{P}(D = 1 \mid Z = 1) \\ &= 50 \times 4/5 + 58 \times 1/5 = 51.6 \end{aligned}$$

Math time: Private schools and vouchers

	$Z = 0$	$Z = 1$
$\mathbb{P}(D = 0 \mid Z)$	$8/9$	$4/5$
$\mathbb{P}(D = 1 \mid Z)$	$1/9$	$1/5$

- Now assume the following expected outcomes, $\mathbb{E}[Y \mid D, Z]$

	$D = 0$	$D = 1$
$Z = 0$	50	60
$Z = 1$	50	58

- $$\begin{aligned}\mathbb{E}[Y \mid Z = 0] &= \mathbb{E}[Y \mid Z = 0, D = 0]\mathbb{P}(D = 0 \mid Z = 0) \\ &\quad + \mathbb{E}[Y \mid Z = 0, D = 1]\mathbb{P}(D = 1 \mid Z = 0) \\ &= 50 \times 8/9 + 60 \times 1/9 \approx 51.11\end{aligned}$$

Math time: Private schools and vouchers

	$Z=0$	$Z=1$
$\mathbb{P}(D=0 \mid Z)$	8/9	4/5
$\mathbb{P}(D=1 \mid Z)$	1/9	1/5

- Now assume the following expected outcomes, $\mathbb{E}[Y \mid D, Z]$

	$D=0$	$D=1$
$Z=0$	50	60
$Z=1$	50	58

-

$$\mathbb{E}[D \mid Z=1] = \mathbb{P}(D=1 \mid Z=1) = 1/5$$
$$\mathbb{E}[D \mid Z=0] = \mathbb{P}(D=1 \mid Z=0) = 1/9$$

Math time: Private schools and vouchers

	$Z = 0$	$Z = 1$
$\mathbb{P}(D = 0 \mid Z)$	$8/9$	$4/5$
$\mathbb{P}(D = 1 \mid Z)$	$1/9$	$1/5$

- Now assume the following expected outcomes, $\mathbb{E}[Y \mid D, Z]$

	$D = 0$	$D = 1$
$Z = 0$	50	60
$Z = 1$	50	58

- $$\delta = \frac{51.6 - 51.11}{1/5 - 1/9} \approx 5.51.$$

Homework

1. Show that the naive treatment effect is 9.67.

Non-Binary Treatment

- When the treatment variable is non-binary, we can derive a similar result
- Assume that the treatment variable is generated according to

$$D = \gamma + \beta Z + \eta.$$

- Recall that

$$\beta = \frac{Cov(D, Z)}{V(Z)}.$$

- Let $\hat{D} = \mathbb{E}[D \mid Z] = \gamma + \beta Z$ be the conditional expectation function (or predicted/fitted values) of D
- Notice that $V(\hat{D}) = V(\beta Z)$ and $Cov(\hat{D}, X) = Cov(\beta Z, X)$ (for any random variable X).

IV Formula in Non-Binary Case

- Now consider the covariance between Y and Z:

$$Cov(Y,Z) = Cov(\alpha + \delta D + \epsilon, Z) = \delta Cov(D,Z) + Cov(\epsilon, Z).$$

- Since ϵ and Z are independent by assumption, we get

$$\delta = \frac{Cov(Y,Z)}{Cov(D,Z)}.$$

- This is a generalization of the Wald estimator to the case when D is non-binary
- Interestingly, this formula also generalizes a regular OLS estimator
- If you substitute D for Z, we are back to the usual expression for the regression coefficient
- In other words, OLS treats D as its own IV

- If we divide both part by $V(Z)$, we get

$$\delta = \frac{Cov(Y, Z)/V(Z)}{Cov(D, Z)/V(Z)}.$$

- The numerator is then the coefficient on Z in the simple regression of Y on Z (reduced form) and the denominator is the coefficient on Z in the simple regression of D on Z (first stage)

2SLS

- We can rewrite the expression for δ one more time.

$$\delta = \frac{Cov(Y,Z)/V(Z)}{\beta} = \frac{\beta Cov(Y,Z)}{\beta^2 V(Z)} = \frac{Cov(Y,\beta Z)}{V(\beta Z)} = \frac{Cov(Y,\hat{D})}{V(\hat{D})}.$$

- This means that, effectively, we can estimate δ as a coefficient from a simple regression of Y on the **predicted** values of D
- This is how the 2SLS estimator works
- First, you get the predicted values of D by regressing it on the instrument Z
- Second, you regress Y on the predicted values of D
- So in estimating the effect of D on Y using IV, we are only using the **exogenous** variation in D caused by Z.

Assumptions and Weaknesses

Relevance and Validity

- We need two assumptions for IV to identify a homogenous TE: **relevance** and **validity**
- Relevance means that our IV should have a causal effect on the treatment variable
- Technically, if relevance is not satisfied, then the denominators in the IV formulas are zero and the estimators are undefined
- Validity means that the instrument should affect the outcome only through the treatment variable
- Technically, this assumption is what leads to the correlation between ϵ and Z to be zero
- The validity assumption is also called the **exclusion restriction**

Weak Instruments

- A bigger issue is that the covariance between the instrument and treatment is small
- This is know as a **weak instrument** problem (Bound, Jaeger, and Baker, 1995)
- In general, weak instruments tend to produce biased estimates
- The argument is the following
    - In finite samples, IV point estimates can always be computed because sample covariances are never exactly equal to zero
    - An IV point estimate can be computed even for an instrument that is not relevant
    - The formulas for calculating the standard errors of IV estimates fail in such situations, giving artificially small standard errors
    - The bias due to small violations of the validity assumption can explode

Bias Due to Weak Instruments

- For example, for a non-binary treatment variable, the bias is equal to

$$\frac{Cov(\epsilon, Z)}{Cov(D, Z)}$$

- A weak instrument will blow up the bias
- The expected bias of the 2SLS estimator is inversely related to the F statistic about the joint significance from the first stage
- If F is large, the bias of the 2SLS goes to zero, as F gets smaller, the bias increases
- The F statistics decreases if the instrument becomes weaker or if you increase the number of weak instruments
- A common rule of thumb is that it should be greater than 10, but Stock and Yogo (2005) provide more specific guidance.

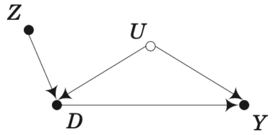
How to Deal with Weak Instruments

- One option is to use a single strongest IV instead of many weak IVs
- Another option is to use use alternative IV estimators, such as a limited-information maximum likelihood estimator (LIML)
- One can use Anderson-Rubin confidence intervals (Anderson and Rubin, 1949)
- They provide valid standard errors even if the instruments are weak
- But the ultimate solution would be to just get better instruments, if possible

Validity

- The validity assumption (or exclusion restriction) is **untestable**
- But suppose you run a regression of Y on Z and D
- If there is an association between Z and Y after conditioning on D, then the instrument must be invalid (right?)

Can We Test It?



- If the only association between Z and Y is through D, then there should be no association between Z and Y after conditioning on D
- However, this logic is false
- If the IV is invalid, then Z and Y will be associated after conditioning on D
- But the converse is **not true**
- Z and Y will always be associated after conditioning on D when validity holds, because D is a collider that is mutually caused by both Z and U

Good Instruments Should Feel Weird

- Tests for overidentification in case you have more than one instrument: Durbin-Wu-Hausman test and Sargan test
- These tests can tell you if some instruments are likely to be invalid.
- An informal test for the validity is that "good instruments should feel weird."
- Without knowledge of the treatment variable, the relationship between the instrument and the outcome should not make much sense
- Because a valid instrument should be irrelevant to the determinants of the outcome except for its effect on the treatment

Example

- Suppose you tell someone that mothers whose first two children were the same gender were employed outside the home less than those whose two children were a boy and a girl?
- What does the gender composition have to do with whether a woman works outside the home?

Example

- Suppose you tell someone that mothers whose first two children were the same gender were employed outside the home less than those whose two children were a boy and a girl?
- What does the gender composition have to do with whether a woman works outside the home?
- Empirically if the first two children are the same gender, families are more likely to have a third compared to those who had a boy and a girl first
- So you have an instrument (gender composition) that only changes the outcome (labor supply) through changing a treatment variable (family size)
- This allows us to identify the causal effect of family size on labor supply

IV is Noisy

- IV estimators are also noisy (have large standard errors)
- By using only a portion of the available covariation in the treatment and outcome, IV estimators use only a portion of the information in the data
- This represents a loss in statistical power
- IV estimators tend to exhibit substantially more sampling variance than other estimators

IV in action

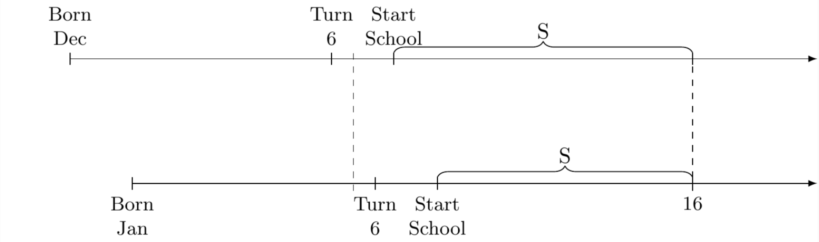
Education and Earnings

- IVs have been used to estimate the causal effect of education
- A variety of IVs: proximity to college, regional and temporal variation in school construction, tuition at local colleges, temporal variation in the minimum school-leaving age, and quarter of birth
- Each of these variables predicts educational attainment but has no direct effect on earnings

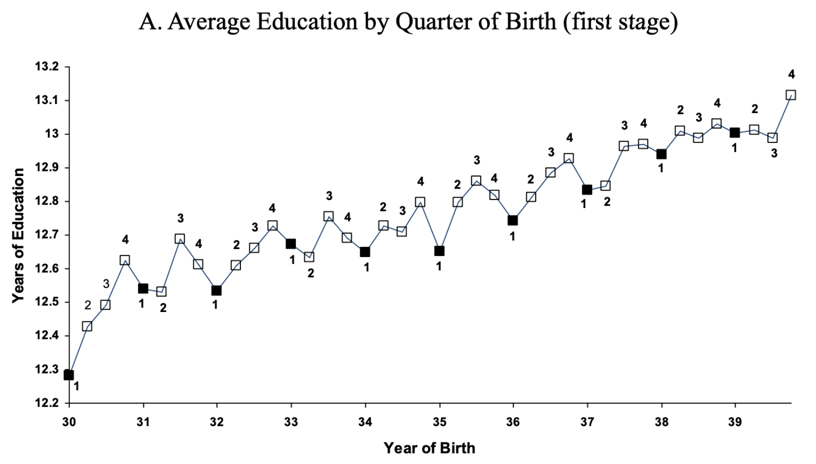
Quarter of Birth

- Angrist and Krueger (1991)
- In the US, a child enters a grade on the basis of his or her birthday
- For a long time, that cutoff was late December
- If children were born on or before December 31, then they were assigned to the first grade
- But if their birthday was on or after January 1, they were assigned to kindergarten
- Two people—one born on December 31 and one born on January 1—were exogenously assigned different grades
- Compulsory schooling laws that forced a person to remain in high school until age 16

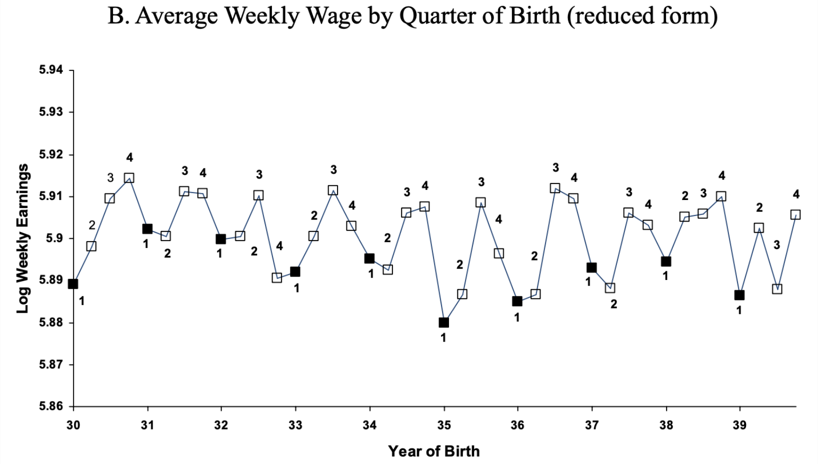
Illustration



First Stage


A. Average Education by Quarter of Birth (first stage)
Years of Education
Year of Birth
13.2
13.1
13
12.9
12.8
12.7
12.6
12.5
12.4
12.3
12.2
30
31
32
33
34
35
36
37
38
39

Reduced Form



Next Time on *Impact Evaluation Methods*...

IV and LATE