

Discussion 6

Week of 03.09

Agenda

- Quiz review
- Instrumental Variables

Announcements

Quiz Review

Question 4

We now study the correlation between weight (lbs) and mpg, and weight and price, and find that price increases with weight and mpg decreases with weight. If we were to include weight into the regression above, which of the following will be true:

- () The new coefficient on MPG will be more negative than without controlling for weight
- () The new coefficient on MPG will be more positive than without controlling for weight
- () The standard errors will be larger
- () The coefficient on MPG will not change

Question 4

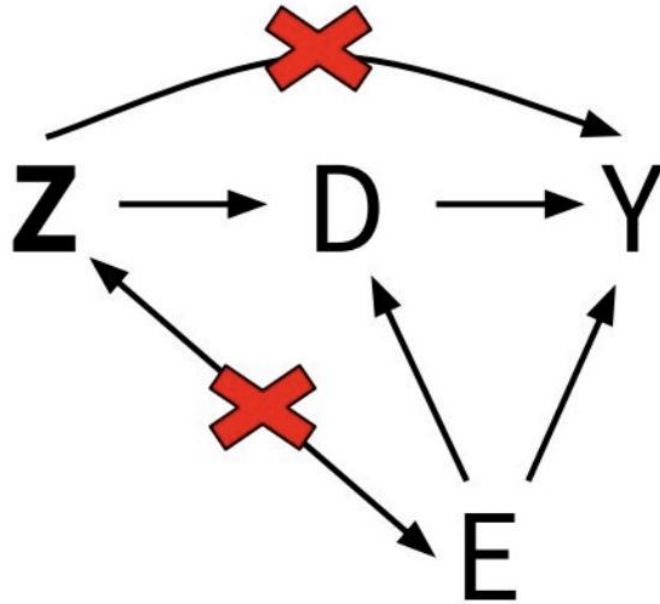
We now study the correlation between weight (lbs) and mpg, and weight and price, and find that price increases with weight and mpg decreases with weight. If we were to include weight into the regression above, which of the following will be true:

- () The new coefficient on MPG will be more negative than without controlling for weight
- (✓) The new coefficient on MPG will be more positive than without controlling for weight
- () The standard errors will be larger
- () The coefficient on MPG will not change

Instrumental Variables

Instrumental Variables (IV) use variation in our dependent variable **D** that can be explained by an exogenous instrument **Z** to examine the effect of **D** on our outcome **Y**.

Instrumental Variables (IV) use variation in our dependent variable **D** that can be explained by an exogenous instrument **Z** to examine the effect of **D** on our outcome **Y**.



Instrumental Variables: Reduced Form Method

- 1. First Stage:** effect of Z on D
 - a. Regress $D \sim Z$ → how much does our instrument affect D?
 - b. Coefficient on Z → first-stage estimate
- 2. Reduced Form/ITT:** effect of Z on Y
 - a. Regress $Y \sim Z$ → Does Y change as Z changes?
 - b. Coefficient on Z → reduced form estimate
- 3. Divide Effects:** Reduced Form / First Stage
- 4. Assumptions for Unbiased Estimate**
 - a. Relevance
 - b. Exclusion Restriction
- 5. Interpretation as local average treatment effect (LATE)**
 - a. To interpret as LATE, also need monotonicity: Z pushes D in same direction as Z (no defiers!)

Instrumental Variables and LATE

<p>Always Taker: Always takes up treatment, regardless of assignment.</p>	<p>Never Taker: Never takes up treatment, regardless of assignment.</p>
<p>Complier: Takes up treatment when assigned (and doesn't take up treatment when not assigned).</p>	<p>Defier: Takes up treatment when not assigned, does not take up treatment when assigned.</p>

LATE: the effect of treatment on your outcome, only for compliers.

Requirements for LATE:

1. Relevance
2. Exclusion restriction
3. No defiers (monotonicity); Z always pushes D in the same direction.

Instrumental Variables: Example 1

Let's say you want to estimate the effect of education on earnings. You start with the regression:

$$\mathbf{Earnings} = \beta_0 + \beta_1 \mathbf{Education} + \boldsymbol{\varepsilon}$$

But you don't believe that education is an exogenous variable. Instead, you might use distance to the nearest college as an instrument for education, because:

1. It satisfies the relevance assumption – that is, it's correlated with X
2. It satisfies the exclusion restriction – that is, it's correlated with Y but only through X

In the **first stage**, you regress X on Z , your instrument.

In this case, you would run the regression:

$$\mathbf{Education} = \beta_0 + \beta_1 \mathbf{Distance} + \boldsymbol{\varepsilon}$$

β_1 is your first-stage estimate, and tells you how much education varies with distance to the nearest college.

If the relevance assumption holds, β_1 should be non-zero.

In the **second stage**, you regress Y on Z.

Now, you would run the regression:

$$\mathbf{Earnings} = \alpha_0 + \alpha_1 \mathbf{Distance} + \epsilon$$

α_1 is your *reduced form* estimate, and tells you how much earnings vary with distance to the nearest college.

To get the local average treatment effect (**LATE**), you now divide your reduced form estimate by your first stage estimate:

$$\alpha_1 / \beta_1$$

The LATE tells you the relationship between distance to the nearest college and future earnings *for those who are more likely to attend college when they live nearby.*

Instrumental Variables: Example 2

Guiding Example: Knowledge Is Power Program, KIPP Lynn

KIPP Lynn is a charter school just above Boston, Massachusetts. Adheres to a “No Excuses” model. KIPP programs are often (intentionally) located in disadvantaged school districts, and are central in the debate about how to remedy achievement/opportunity gaps.

What concerns might we have about the following simple regression?

$$\text{TestScores} = \alpha + \beta_1 \text{KIPP} + \varepsilon$$

Instrumental Variables: Example 2

After 2005, demand for (or applications to) KIPP Lynn bypassed the seats they could offer and seats were offered based on a lottery system. We want to look at the relationship between enrolling in KIPP and test scores based on this lottery system.

What's our instrument?

What's our dependent variable?

What's our outcome variable?

Instrumental Variables: Example 2

Compliers: Students who win the lottery and attend KIPP; students who lose the lottery and do not attend KIPP

Deniers: Students who win the lottery and don't attend KIPP; students who lose the lottery and attend KIPP.

Always-Takers: Students who will always attend KIPP, even if they lose the lottery.

Never-Takers: Students who will never attend KIPP, even if they win the lottery.

Putting it all together:

	Lottery Loser	Lottery Winner
Attended KIPP	0.046	0.787
Test Score	-0.358	-0.003

Test score is in standard deviations

What's our first-stage estimate?

What's our reduced form estimate?

How do we find the LATE?

Putting it all together:

	Lottery Loser	Lottery Winner
Attended KIPP	0.046	0.787
Test Score	-0.358	-0.003

Test score is in standard deviations

What's our first-stage estimate? **0.74**

What's our reduced form estimate? **0.36**

How do we find the LATE? **$0.36/0.74 = 0.48$ standard deviations**

Conclusion:

	Lottery Loser	Lottery Winner
Attended KIPP	0.046	0.787
Test Score	-0.358	-0.003

Test score is in standard deviations

Attending KIPP Lynn increases student test scores by 0.48 standard deviations, for students who attend KIPP when assigned a seat in KIPP.