Lecture 4 – Instrumental Variables

We will focus on using Instrumental Variables (IV) to get at causal estimates. IV is a framework that allows us to establish the causal effect of X on Y given certain assumptions. It requires that you have a variable, which we call an instrument.

What is Instrumental Variables good for?

Imperfect compliance for an experiment

Measurement error

Solving endogeneity problems

More generally, getting causal estimates

What are the requirements for a good instrument?

1. First Stage: Instrument is strongly correlated with the treatment
2. Independence: Instrument is as good as randomly assigned
3. Exclusion Restriction: The instrument only affects the outcome through the treatment variable

With these requirements in mind, let's talk through IV in a common setting. Imperfect compliance in policy experiments. We can always make the comparison of people who were randomized into treatment versus control. That comparison is valid and the resulting difference in means is the effect of being assigned to treatment. However, we are often interested in the effect of treatment rather than the effect of being assigned to treatment.

In MM, the authors consider a charter school in Massachusetts that is oversubscribed called KIPP. Because there are fewer slots than applicants, admissions is based on a lottery. Some students are offered a seat and some are not.

Is the comparison of students offered a seat to students who are not a valid comparison? (Yes) What does it reveal? (The effect of being offered a seat). This is often referred to as the *Reduced Form*. It is also referred to as the *Intent to Treat (ITT)*.

Is that what we want to know? (Probably not, more interesting to know the effect of *attending* KIPP)

In the KIPP case, students may receive an offer to attend KIPP but do not enroll for a variety of reasons (moving, changing their mind, etc) and some students will sneak into KIPP even if they haven't be offered a seat via lottery. In fact about 78% of students attend KIPP if they get an offer and 4% attend KIPP if they don't. As a result, there are some students who should see no treatment effect but we include them in the "reduced form" estimate.

Some notation: Y_i is the outcome, Z_i is the instrument and X_i is the treatment

The first stage equation is:

$$X_i = \gamma_0 + \gamma_1 \cdot Z_i + \nu_i$$

A strong first stage will change the probability of treatment by a lot.

The reduced form equation is :

$$Y_i = \beta_0 + \beta_1 \cdot Z_i + \varepsilon_i$$

The IV estimator we learned is then just:

$$\frac{\beta_1}{\gamma_1}$$

Which is the sample analog of

$$= \frac{E[Y_i|Z=1] - E[Y_i|Z=0]}{E[X_i|Z=1] - E[X_i|Z=0]}$$

This has an intuitive interpretation of how much did the outcome change when the instrument changed versus how much the treatment changed. This scales up the estimate to account for imperfect compliance. With no controls, this is called a Wald Estimator

What is the interpretation of this estimator? We need to introduce some terminology first.

Always Taker This person will always get treatment, regardless of the value of the instrument

Never Taker This person will never get treatment, regardless of the value of the instrument

Complier This person will get treatment only if the instrument indicates they should. That is $X_=1$ if $Z_i = 1$ and $X = 0$ if $Z_i = 0$

Defiers This person will be treated if the instrument says they should not be.

We often assume monotonicity, that is, there are no defiers. This makes sense in many settings but doesn't make sense in all settings.

If we assume no defiers, our IV estimate is the estimate of the *Local Average Treatment Effect (LATE)*. This is the average effect among compliers. As a contrast the

Average Treatment Effect (ATE). ATE is the average treatment effect for every unit in the sample.

Example

Group	$Y_{X=0}$	$Y_{X=1}$	Proportion
Always Takers	3	4	.25
Compliers	0	2	.5
Never Takers	0	1	.25
Defiers	0	4	0

What is the estimated treatment effect?

For $Z = 1$ Average Outcome $= .25 * 4 + .5 * 2 + .25 * 0 + 0 * 0 = 2$

For $Z = 0$ Average Outcome $= .25 * 4 + .5 * 0 + .25 * 0 + 0 * 4 = 1$

So the ITT=1, The first stage is .5. So the IV would be $\frac{1}{.5} = 2$...which is the effect for compliers! If we assume monotonicity, or no defiers, then we get a LATE or the treatment effect for compliers.

Is the LATE useful? Is it about who we want to know about? Yes and No. It is the policy relevant population. However, we may be interested in the treatment effects for a broader set of people.

An often levied criticism of "reduced form" econometrics is that just reveals a LATE. This criticism is 100% correct. Ultimately this is a statement about external validity. As with all statements about external validity, you should avoid "cheap" comments.

If the treatment variable is continuous (e.g. dollars) then there is a generalization of LATE called the Average Causal Response (Angrist Imbens 1995 JASA)

Would comparing people who were actually treated to people who were not be appropriate in a randomized controlled trial with imperfect compliance? (No) Why not? (Because compliance is not random, and so there would be selection bias)

Another parameter you will hear about is the *Treatment on the Treated (TOT)*. This is the treatment effect among people who are treated. This can contain selection bias if there is imperfect compliance. However, if there are no always takers than TOT=LATE

As we covered, the IV estimator can be generalized to *Two Stage Least Squares (2SLS)*.

In two stage least squares you obtain IV estimates in two stages (hence the name).

1. First you estimate the first stage equation:

$$X_i = \beta_0 + \beta_1 Z_i + \mathbf{X_i}\beta + \varepsilon_i$$

2. Second, you take the fitted values from the first stage, $\widehat{X_i}$ and use them in the second stage:

$$Y_i = \alpha_0 + \alpha_1 \widehat{D_i} + \mathbf{X_i}\alpha + \eta_i$$

The IV estimate is then α_1. If the assumptions of IV are met, this can be interpretted as the causal effect of treatment on the outcome Y, for compliers. That is, this is the LATE.

Notes on 2SLS:

1. Do not do this in two steps, you will get the standard errors wrong if you do
2. This is why we call Z_i the excluded instrument, it is excluded from the second stage
3. You must include all control variables in both the first stage and the second stage–if you only include them in the second stage you're treating them like instruments. You probably don't think they satisfy the assumptions of an instrument
4. There are other methods to get IV estimates with differnet properties. We've talked about two: 2SLS and Wald. Other examples include LIML and GMM
5. 2SLS is nice because it allows you to control for things. This can help with the independence assumption. For example, we think Z is independent only after we control for something.
6. 2SLS is also nice because you can use multiple instruments easily. You just include them in the first stage (and exclude them from the second stage).

IV essentially focuses the attention of the estimator on certain types of variation. You can see this with how 2SLS is computed. If you're careful this will be "good" variation. This means that it "throws out" lots of variation and hence has larger standard errors.

Examples:

Effect of more siblings on test scores (from MM)

Why is this interesting?

Why is this hard to answer?

Twins Instrument: Asses the following

First Stage–Do twins cause some people to have more siblings? Which people?

Exclusion Restriction–Can twins have an impact on test scores in a way other than having an additional sibling?

Independence–Is it random who has twins?

Monotonicity–Does having twins always increase the number of children?

Same Sex Instrument: Asses the following

First Stage–Does two kids of the opposite sex have an effect on the total number of children?

Exclusion Restriction–Is there any way that having siblings of the same gender could have a direct effect on grades? (The authors mention sharing a room) How do they test this? (Consider unaffected by the instrument)

Independence–Is it random which families have same-sex siblings versus opposite sex siblings?

Monotonicity–Does having same sex siblings always increase the number of children?