Economics 696F, Causal Inference and Program Evaluation

Lecture Note 8: Instrumental Variables Part I

1 Wald Estimator

Suppose we are interested in estimating the effect of a treatment T on an outcome Y, and we believe the treatment does *not* satisfy the random assignment assumption, nor the unconfoundedness assumption. Then the approaches we have talked about up to now will not work.

In some cases, we might have an additional variable, call it Z, that is known to be randomly assigned, and is thought to be related to the treatment.

Example: suppose that Y is wages, the treatment T is whether or not the individual goes to college, and suppose that some individuals are randomly assigned a tuition subsidy. So Z would be the indicator for the tuition subsidy.

We call such a variable Z an <u>instrumental variable</u>, or <u>instrument</u> for short, although perhaps a more descriptive word might be an <u>encouragement</u>. Later, we will define the term instrumental variable more precisely.

It seems plausible that the instrument Z could affect the decision whether or not to go to college. Through the college decision, it might therefore also have an impact on wages. This suggests the following simple statistic, called the <u>Wald estimator</u>:

$$b = \frac{E(\widehat{Y|Z=1}) - E(\widehat{Y|Z=0})}{E(\widehat{T|Z=1}) - E(\widehat{T|Z=0})}$$

where $E(\widehat{Y|Z=1})$ is the sample average of Y among observations with $Z = 1$, and so on. So we can also write

$$b = \frac{\sum Y_i Z_i / \sum Z_i - \sum Y_i(1-Z_i)/\sum(1-Z_i)}{\sum T_i Z_i / \sum Z_i - \sum T_i(1-Z_i)/\sum(1-Z_i)}.$$

To interpret this quantity, consider first the denominator. This is roughly the effect of the subsidy on the treatment (college). We might expect this to be some number between 0 and 1: the subsidy increases college attendance for some fraction of the individuals.

The numerator is roughly the effect of the subsidy on wages. We expect that subsidies increase college attendance, which increases wages, but since the subsidy only makes a difference for some fraction of the individuals, the overall effect on wages will be "diluted."

Dividing the numerator by the denominator "blows up" the overall effect on wages, in proportion to the fraction of people influenced by the subsidy. So perhaps b might be a reasonable attempt to estimate the return to college. Can we fit this into our causal framework?

2 Causal IV Model with Constant Treatment Effects

We will continue to work with *potential outcomes* $Y(1)$ and $Y(0)$, where $Y(1)$ is the outcome that would obtain if $T = 1$, and $Y(0)$ is the outcome that would obtain if $T = 0$. We are interested in averages of individual-level treatment effects $Y(1)-Y(0)$. As before, we only observe one of the two potential outcomes:

$$Y \equiv TY(1) + (1 - T)Y(0).$$

Now, we have a binary instrument, that could affect the treatment received. So we will use a potential outcomes notation for the treatment under different values of the instrument:

$$T(1) = \text{ Treatment status if } Z = 1$$
$$T(0) = \text{ Treatment status if } Z = 0.$$

The idea is that different values of the instrument might lead to different treatment regimes for a given individual. As with the potential outcomes, we only observe one of the potential treatments:

$$T \equiv ZT(1) + (1 - Z)T(0).$$

We will make the following assumptions about the instrument Z:

Random Assignment of Instrument:

$$Z \perp Y(0), Y(1), T(0), T(1)$$

We also assume that Z_i is correlated with the actual treatment received:

Instrument Correlated with Treatment Received:

$$Pr(T(1) = 1) \neq Pr(T(0) = 1).$$

Suppose we assume that the treatment effect is constant across individuals: for all i:

Constant Treatment Effect:

$$Y_i(1) - Y_i(0) = \alpha.$$

(So the average treatment effect is also α.)

Under these assumptions, does the Wald estimator consistently estimate α? By the law of large numbers, we have

$$b \xrightarrow{p} \frac{E(Y|Z=1) - E(Y|Z=0)}{E(T|Z=1) - E(T|Z=0)}.$$

Looking at the numerator, we can write:

$$\begin{aligned} E[Y|Z=1] - E[Y|Z=0] &= E[T(1)Y(1) + (1-T(1))Y(0)|Z=1] \\ &\quad -E[T(0)Y(1) + (1-T(0))Y(0)|Z=0] \\ &= E[T(1)Y(1) + (1-T(1))Y(0)] \\ &\quad -E[T(0)Y(1) + (1-T(0))Y(0)] \qquad \text{(by independence of } Z\text{)} \\ &= E[(T(1)-T(0))(Y(1)-Y(0))] \\ &= \alpha E[T(1)-T(0)]. \end{aligned}$$

Also,

$$\begin{aligned} E[T|Z=1] - E[T|Z=0] &= E[T(1)|Z=1] - E[T(0)|Z=0] \\ &= E[T(1)-T(0)]. \end{aligned}$$

So,

$$b \xrightarrow{p} \alpha = E[Y(1) - Y(0)].$$

We conclude that the treatment effect is *identified*, and that the Wald estimator is a consistent estimator for it.

3 Wald Estimator as an M-Estimator

Continue using the constant treatment effect assumption in the previous section. We can also write this as

$$Y_i(1) = Y_i(0) + \alpha.$$

Then we can express the observed outcome as

$$\begin{aligned} Y_i &= Y_i(0) + T_i\alpha \\ &= \gamma + T_i\alpha + U_i, \end{aligned}$$

where $\gamma = E[Y_i(0)]$ and $U_i = Y_i(0) - \gamma$.

At first glance, one might think to estimate γ and α by OLS, since they are parameters in a simple linear equation. With a binary regressor, the slope coefficient will just be equal to

$$\widehat{E[Y|T=1]} - \widehat{E[Y|T=0]},$$

i.e. the difference in means between treated and controls. This would be fine if the *treatment* were randomly assigned, but in this section, we want to allow the treatment to be correlated with the potential outcomes. So we would not expect this to be a good estimator.

Can we make this more precise? Think more generally of a linear model:

$$y_i = x_i'\beta + u_i.$$

The least squares estimator is

$$\hat{\beta} = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i.$$

Under what conditions does $\hat{\beta} \xrightarrow{p} \beta$? By the law of large numbers,

$$\hat{\beta} \xrightarrow{p} E[x_i x_i']^{-1} E[x_i y_i] = E[x_i x_i']^{-1} E[x_i(x_i'\beta + u_i)].$$

Assume that $E[x_i x_i']$ is invertible, and that $E[x_i u_i] = 0$. Then $\hat{\beta} \xrightarrow{p} \beta$.

If $E[u_i|x_i] = 0$, so that $x_i'\beta$ is the conditional mean of y_i given x_i, then

$$E[x_i u_i] = E[x_i E[u_i|x_i]] = 0.$$

So, we conclude that it is essential to have the *orthogonality* condition $E[x_i u_i]$ for OLS to consistently estimate the parameter θ.

Now, return to the linear model arising from the potential outcomes specification:

$$Y_i = \gamma + T_i\alpha + U_i.$$

We could set $x_i = (1, T_i)$ and $\theta = (\gamma, \alpha)$, and be in the standard linear framework. However, since $U_i = Y_i(0) - E[Y_i(0)]$, we expect that T_i will generally be correlated with U_i, so that

$$E[U_i T_i] \neq 0.$$

Thus, we do *not* have orthogonality, and therefore OLS will not be consistent.

However, we do know that the instrument Z_i is independent of the potential outcomes. So we would have that

$$E[Z_i U_i] = 0,$$

or substituing in for U_i,

$$E[Z_i(Y_i - \gamma - T_i\alpha)] = 0$$

We should also have that

$$E[(Y_i - \gamma - T_i\alpha)] = 0.$$

This suggests estimating γ and α using a GMM estimator[1], based on the moment function:

$$\psi(Y_i, T_i, Z_i, \gamma, \alpha) = \begin{pmatrix} Y_i - \gamma - T_i\alpha \\ Z_i(Y_i - \gamma - T_i\alpha) \end{pmatrix}.$$

The GMM estimator $\hat{\gamma}, \hat{\alpha}$ solves:

$$\frac{1}{n}\sum_{i=1}^{n} \psi(Y_i, T_i, Z_i, \hat{\gamma}, \hat{\alpha}) = 0.$$

Using standard GMM results, and some algebra, it is not hard to show that

$$\hat{\alpha} = \frac{\sum(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum(Z_i - \bar{Z})(T_i - \bar{T})} = \frac{\widehat{Cov(Z,Y)}}{\widehat{Cov(Z,T)}},$$

and that when T_i and Z_i are binary,

$$\hat{\alpha} = \frac{E(\widehat{Y|Z=1}) - E(\widehat{Y|Z=0})}{E(\widehat{T|Z=1}) - E(\widehat{T|Z=0})}$$

In other words, the estimator $\hat{\alpha}$ is identical to the Wald estimator. This is nice, because we can use standard GMM (or M-estimator) results to obtain the asymptotic variance of $\hat{\alpha}$.

4 Heterogeneous Treatment Effects

The analysis so far has assumed "constant treatment effects". Now, we want to explicitly allow for *heterogeneity* in treatment effects across individuals, and see whether it is still possible to estimate treatment effects in an IV setting.

For now, we will stick with the simple setup with a binary treatment, and a binary instrument, and make the same assumptions about the instrument (random assignment and correlation with treatment).

We have *potential outcomes* $Y(1)$ and $Y(0)$, and we are interested in averages of individual-level treatment effects $Y(1) - Y(0)$.

[1] Formally, it would be more appropriate to think of this as an M-estimator, but the terminology GMM estimator is widely used.

Does the Wald estimator estimate a causal effect, such as $E[Y_i(1)-Y_i(0)]$? In general, the answer turns out to be no. Recall that the Wald estimator satisfies:

$$b \xrightarrow{p} \frac{E(Y|Z=1)-E(Y|Z=0)}{E(T|Z=1)-E(T|Z=0)}.$$

Looking at the numerator, we can write:

$$\begin{aligned}
E[Y|Z=1]-E[Y|Z=0] &= E[T(1)Y(1)+(1-T(1))Y(0)|Z=1] \\
&\quad -E[T(0)Y(1)+(1-T(0))Y(0)|Z=0] \\
&= E[T(1)Y(1)+(1-T(1))Y(0)] \\
&\quad -E[T(0)Y(1)+(1-T(0))Y(0)] \qquad \text{(by independence of } Z) \\
&= E[(T(1)-T(0))(Y(1)-Y(0))] \\
&= Pr(T(1)-T(0)=1)E[Y(1)-Y(0)|T(1)-T(0)=1] \\
&\quad -Pr(T(1)-T(0)=-1)E[Y(1)-Y(0)|T(1)-T(0)=-1].
\end{aligned}$$

Now, suppose that

$$Pr(T(1)-T(0)=-1)=Pr(T(1)-T(0)=1)\frac{E[Y(1)-Y(0)|T(1)-T(0)=1]}{E[Y(1)-Y(0)|T(1)-T(0)=-1]}.$$

Then, we will have

$$E[Y|Z=1]-E[Y|Z=0]=0$$

even if the treatment effect is positive for every individual. So, in the general model with heterogeneous treatment effects, b does not consistently estimate the average treatment effect. More generally, there is no consistent estimator; the average treatment effect is *not identified*.

5 Local Average Treatment Effect

Imbens and Angrist (1994) show that under an additional assumption on the instrument, a different average treatment effect can be identified. The assume:

Monotonicity of the Instrument: either

$$T_i(1) \geq T_i(0) \qquad \text{for all } i$$

or

$$T_i(1) \leq T_i(0) \qquad \text{for all } i.$$

This says that the instrument affects all individuals' treatment choices in the same direction. To see how this works, it may help to draw a matrix of potential treatments:

	$T(0) = 0$	$T(0) = 1$
$T(1) = 0$	$T(1) - T(0) = 0$	$T(1) - T(0) = -1$
$T(1) = 1$	$T(1) - T(0) = 1$	$T(1) - T(0) = 0$

Now, suppose that $T_i(1) \geq T_i(0)$ for all i. (The other case works similarly.) Then

$$Pr(T(1) - T(0) = -1) = 0.$$

It follows that

$$E[Y|Z = 1] - E[Y|Z = 0] = Pr(T(1) - T(0) = 1)E[Y(1) - Y(0)|T(1) - T(0) = 1].$$

Also, note that since $T(1) - T(0)$ is either 0 or 1,

$$E[T|Z = 1] - E[T|Z = 0] = E[T(1) - T(0)] = Pr(T(1) - T(0) = 1).$$

So we have

$$b \xrightarrow{p} E[Y(1) - Y(0)|T(1) - T(0) = 1].$$

The quantity $E[Y(1) - Y(0)|T(1) - T(0) = 1]$ is called the *local average treatement effect*. To understand what it means, notice that $T(1) - T(0) = 1$ implies that

$$T(1) = 1 \qquad \text{and} \qquad T(0) = 0.$$

In other words, these are individual who change their treatment in response to the instrument. This subgroup is sometimes called the *compliers* subgroup, because their "comply" with the instrument.

Therefore, under monotonicity, the instrumental variables estimator estimates the average effect of the treatment for those individuals who are influenced by the instrument to change their treatment status.

6 References

Angrist, J., and Imbens, G., (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica* 62, 467-475.

Angrist, J., Imbens, G., and Rubin, D., (1996), "Identification of Causal Effects using Instrumental Variables," *Journal of the American Statistical Association* 91.