Some more notes on IV estimation and Diff-in-Diff

We saw that the 2SLS estimator is consistent if we have a valid instrument. There are however some pitfalls of which you should be aware. First and foremost of course that your instrument is not really valid, that is one of the assumptions (correlated with the endogenous regressor, uncorrelated with the error term) might be invalid. An implication of the IV assumptions is that your instrument cannot have an impact on the outcome by itself, i.e. the only reason why your instrument is correlated with the outcome of interest is because it is correlated with the endogenous regressor.

Another problem might be that while the instrument might be correlated with the endogenous regressor, this correlation is weak. This will not only result in very large standard errors for your estimate but might also heavily bias your results ("weak instruments" is the buzzword for this discussion).

On the other hand an instrument that is too highly correlated might not be a good thing either. We showed that 2SLS is consistent, but not unbiasedness. In fact 2SLS is biased and the bias increases with the goodness of fit of your first stage (think of it this way: the better the instrument explains your endogenous regressor, the more likely you are to pick some of the "bad variation" up as well).

But despite these problems, instrumental variable strategies are a crucial and invaluable tool in establishing causal links in economics. We have to be careful though. Not only do we have to circumnavigate some pitfalls, but we also have to be modest about the interpretation of our results.

Why do we have to be careful? So far we assumed that the impact of our endogenous regressor is the same for all units of observation (the coefficient is just a fixed parameter). But in practice it is much more likely that the impact is heterogeneous. Are we in trouble? Well, OLS will give us the average impact of a regressor, so that is not too bad and probably something we are interested in anyway. 2SLS on the other hand will deliver a local average, rather than the overall average impact.

To see what this means let us re-evaluate the IV estimator from a different point of view. We are still interested in the return to education, but now let us focus to the return to college. We use the specification

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where x is a dummy that is equal to one if someone attended college, y denotes the (log) wage and ε is the error term.

The fundamental problem we face is that for each person in our sample we can only observe one outcome, either the outcome that prevails if they attend college, or the outcome if they did not attend. If we would know the outcome in both states of the world, the estimation would be easy. We denote the outcome (wage) if someone goes to college as y_{1i} and the wage for the same person without college attendance as y_{0i}. From the underlying model we have:

$$y_{1i} = \beta_0 + \beta_1 + \varepsilon_i$$
$$y_{0i} = \beta_0 + \varepsilon_i$$

How to estimate β_1 should be obvious.

But since we can only observe one outcome that is not feasible. We observe y_{1i} if $x_i = 1$ and y_{0i} if $x_i = 0$. So what we are estimating with OLS will be

$$b_1 = E(y_{1i}|x_i = 1) - E(y_{0i}|x_i = 0)$$

Now is this what we want? $E(y_{1i}|x_i = 1) - E(y_{0i}|x_i = 1)$ gives us the college premium for those who went to college $E(y_{1i}|x_i = 0) - E(y_{0i}|x_i = 0)$ gives us the college premium for those who didn't go to college. Assuming that the premium is the same for both groups (will will consider heterogeneous returns later) we can rewrite the difference:

$$\begin{aligned} b_1 &= E(y_{1i}|x_i = 1) - E(y_{0i}|x_i = 1) + E(y_{0i}|x_i = 1) - E(y_{0i}|x_i = 0) \\ &= E(y_{1i} - y_{0i}|x_i = 1) + E(y_{0i}|x_i = 1) - E(y_{0i}|x_i = 0) \end{aligned}$$

That means we get the causal effect only if the wage that someone who went to college would have gotten if the person didn't go to college is the same as the wage that someone who decided not to go to college gets. But that seems a rather dubious assumption, last time we discussed "ability" to matter, so if people with high ability attend college, but high ability also influences earnings in absence of attending college, $E(y_{0i}|x_i = 1)$ and $E(y_{0i}|x_i = 0)$ will differ and we are back to the case of omitted variable bias.

An example where we can assume that $E(y_{0i}|x_i = 1) = E(y_{0i}|x_i = 0)$ is if we randomly assign x, that is we run an experiment. If we could randomly assign college education to a population, we would not have the selection problem through ability.

We saw that using an instrument can help us. With a binary regressor and a binary instrument (let's use a dummy indicating whether someone lived near a college, as in David Card's study) we can rewrite the IV estimator. Remember the simple IV estimator from last session (y_i now refers to the observed dependent variable):

$$b_1^{IV} = \frac{\frac{1}{N}\sum_{i=1}^{N}(z_i - \bar{z})(y_i - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(z_i - \bar{z})(x_i - \bar{x})}$$

We can rewrite this expression, first the numerator (the denominator can be rewritten analogously). Let N_j denote the number of observations with $z = j$ and $\bar{y}_j$the average of y for the observations where $z = j$, both for $j = 1,2$:[1]

$$\sum_{i=1}^{N}(z_i - \bar{z})(y_i - \bar{y}) = \sum_{i=1}^{N} z_i y_i - \sum_{i=1}^{N} z_i \bar{y} - \sum_{i=1}^{N} \bar{z} y_i + \sum_{i=1}^{N} \bar{z}\bar{y} = \sum_{i=1}^{N} z_i y_i - N_1 \bar{y} - N_1 \bar{y} + N_1 \bar{y}$$

[1] Since the instrument z is just a binary variable, its average $\bar{z}$ is equal to the sample share of observations with the instrument equal to one $\bar{z} = \frac{N_1}{N}$. Also we can represent the total average as the weighted sum of the average for those with the instrument equal to one and those with instrument equal to zero. Where the weights are the sample shares: $\bar{y} = \frac{N_1}{N}\bar{y}_1 + \frac{N_0}{N}\bar{y}_0$.

$$= \sum_{i|z_i=1} y_i - N_1\bar{y} = N_1\bar{y}_1 - N_1\bar{y} = N_1\left(\bar{y}_1 - \frac{N_1}{N}\bar{y}_1 - \frac{N_0}{N}\bar{y}_0\right) = N_1\left(\frac{N-N_1}{N}\bar{y}_1 - \frac{N_0}{N}\bar{y}_0\right)$$

$$= \frac{N_1 N_0}{N}(\bar{y}_1 - \bar{y}_0)$$

The factor in front of the parentheses cancel out (the same factor appears in the numerator and the denominator) and we are left with: [2]

$$b_1^{IV} = \frac{(\bar{y}_1 - \bar{y}_0)}{(\bar{x}_1 - \bar{x}_0)} = \frac{\frac{1}{N_1}\sum_{i|z_i=1} y_i - \frac{1}{N_0}\sum_{i|z_i=0} y_i}{\frac{1}{N_1}\sum_{i|z_i=1} x_i - \frac{1}{N_0}\sum_{i|z_i=0} x_i}$$

Taking the probability limit shows you what we estimate:

$$plim\ b_1^{IV} = \frac{E(y_i|z_i=1) - E(y_i|z_i=0)}{E(x_i|z_i=1) - E(x_i|z_i=0)}$$

The difference in the expected outcome for those people that get treated by the instrument and those that do not get treated. In our example it is the difference in the expected wage for men who lived closed to a college and those who did not live close to a college. The difference is scaled by the difference in expected shares of men going to college in each of the two instrument groups. We can plug in the true model for y_i and get consistency. But more interestingly we can extend the model such that we allow for heterogeneous returns, i.e. not everyone gains the same from college, but there are differences among individuals:

$$y_i = \beta_0 + \beta_{1i}x_i + \varepsilon_i$$

Note that we cannot hope to identify β_{1i}, but something we can estimate is $E(\beta_{1i})$, this is often referred to as the average treatment effect (ATE). Back to our model, plugging in gives us

$$plim\ b_1^{IV} = \frac{E(\beta_{1i}x_i|z_i=1) - E(\beta_{1i}x_i|z_i=0)}{E(x_i|z_i=1) - E(x_i|z_i=0)}$$

So what we estimate is a weighted average of the heterogeneous treatment effect β_{1i}. But we need another assumption for the weights to be valid. Weights may not be negative. That means we need $E(x_i|z_i=1) - E(x_i|z_i=0) > 0$

What does this assumption mean? Well there are four groups of people in our example:

Always takers, never takers, compliers and defiers. Always takers choose to go to college ($x=1$) no matter what the instrument is an never takers will never go to college, i.e. $x=0$, no matter the instrument. Only compliers and defiers are affected by the instrument. Compliers go to college ($x=1$) if they live near a college ($z=1$) but not if they don't ($x=0, z=0$), lastly defiers do exactly the opposite. They choose to go to college if they live far away and they do not choose to go to college if they live close.

[2] This IV estimator is usually referred to as the Wald estimator.

So why could $E(x_i|z_i = 1) - E(x_i|z_i = 0)$ be negative? Well if there are sufficiently many defiers. $E(x_i|z_i = 1)$ is the share of always takers and compliers and $E(x_i|z_i = 0)$ is the share of always takers and defiers. So the difference will be negative if the share of defiers is larger than the share of compliers. So we need to make the assumption that there are no defiers.

Assuming there are no defiers what we identify is the treatment effect for compliers. This is not necessarily the same as the ATE, but rather a local average treatment effect (LATE), i.e. the average treatment for those individuals who will go to college because they live close but wouldn't otherwise. It is crucial to keep this in mind when interpreting one's results.