Lecture 3

Today I introduce instrumental variables in the context of Angrist and Evan's AER paper looking at the effects labor supply of women from additional children.

As in the example presented last class, we can frame today's question in terms of wanting to know counterfactual outcomes, this time from having an additional child.

Let $D_i = 1$ indicate woman i has 'kids' and $D_i = 0$ if not. I'm going to stick with a binary conditional variable here, which leads to a more intuitive way of interpreting the results as causal (or not). Later, I'll come back to the case when D_i can take on more than 2 values. For each individual, there are 2 possibilities. Y_{1i} is woman i's circumstances (or universe) with 'kids', and Y_{0i} is her circumstances without. To be concrete, suppose this outcome variable Y is weeks worked in the year i is age 40. So we are looking at the effect of children on women's labor supply.

$$Y_{0i} = \beta_0 + e_i$$
$$Y_{1i} = \beta_0 + \beta_{1i} + e_i$$

It would be nice to know the causal effect of 'having kids' for each woman: $Y_{1i} - Y_{0i} = \beta_{1i}$. This difference tells us how much more would i work if a she has kids. Having kids, here, is independent of any other circumstance. Think of it like a stork randomly dropping off kids off at different households: the woman has no choice in this allocation for the purpose of considering the independent effect). We don't want to mistake the analysis with reverse causality: a change in circumstances leads also to a change in the decision to have kids.

The causal effect is individual specific. There is no reason for the effect to be the same for everyone and, indeed, it likely is not. Our static model from last lecture suggests different preferences, opportunity costs, and reservation wages will lead to different changes in behaviour.

The fundamental problem of causal inference is that we can never observe the counterfactuals: if i has kids, for example, we never observe what her circumstances would have been if she didn't have them.[1] We only observe $Y_i = Y_{0i}(1-D_i) + Y_{1i}D_i$ (one equation, two unknowns).

The most common identification strategy for predicting values of Y_{0i} and Y_{1i} uses ordinary least squares (OLS) A single variable OLS regression model in this context is

(2) $Y_i = \beta_0 + \beta_1 D_i + e_i$,

where e_i is a statistical error term. Choosing values for β_0 and β_1 to minimizing the sum of squared errors in the case where D_i is binary is done by choosing β_1 as the difference in means between those in the sample with kids and those without:

$$\begin{aligned}
\hat{\beta}_1 &= E(Y_i \mid D_i = 1) - E(Y_i \mid D_i = 0) \\
&= E(Y_{1i} \mid D = 1) - E(Y_{0i} \mid D_i = 0) \\
&= E(Y_{1i} - Y_{0i} \mid D_i = 1) \\
&\qquad + [E(Y_{0i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 0] \\
&= E(\beta_{1i}) + [E(e_i \mid D_i = 1) - E(e_i \mid D_i = 0]
\end{aligned}$$

$\hat{\beta}_1$ provides an estimate for the average causal effect from having kids only if $[E(Y_{0i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 0] = 0$. In words, only if mean hours worked for women with no kids is equal to the mean hours worked for women with kids had instead they not had any. Women who don't have kids may have better job opportunities or prefer working over women who do, and would work more anyway, even a stork made a visit to the house. In this example,

[1] Bill Murray in the movie 'Groundhog Day' could observe counterfactuals because he kept waking up on the same day. He used this ability for testing out different pick up lines to attract women.

$[E(Y_{0i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 0]$ would be negative. A nonzero amount for this expression is called 'omitted variables bias'. Other variables that affect both Y_i and D_i will bias our estimate of the causal effect.

More generally, we can express omitted variables bias (OVB) using the OLS formula for estimating β_1, even when D_i is continuous (e.g. number of kids)

$$p\lim \hat{\beta}_1 = p\lim \frac{\sum (Y_i - \bar{Y})(D_i - \bar{D})}{\sum (D_i - \bar{D})^2} = \frac{\text{cov}(Y_i, D_i)}{\text{var}(D_i)} = \beta_1 + \frac{\text{cov}(e_i, D_i)}{\text{var}(D_i)}$$

The omitted variables bias is $\frac{\text{cov}(e_i, D_i)}{\text{var}(D_i)}$: the estimate for the causal effect of having kids is biases upwards (downwards) if factors that positively relate to Y_i (e_i) also positive (negatively) relate to having more children.

Note that random assignment of D_i ensures that $\frac{\text{cov}(e_i, D_i)}{\text{var}(D_i)}$ is zero. That's why experiments with random assignment generate the most convincing analyses for estimating causal effects. Of course, most people choose not to participate in an experiment involving random assignment of kids.

If this term is not zero, our estimate of β_1 has no causal interpretation.

Drawing causal inferences from data is the central focus in applied labor economics: so central, that most labor economists that make causal conclusions from a set of results use more than half the paper or presentation trying to convince us that the research design is credible and in fact the omitted variables bias is negligible. Most papers are scrutinized on the basis of whether or not this argument seems correct. Failing to account for omitted variables bias or not accounting for it well enough comes up time and time again, almost everywhere drawing causal conclusions. It's good to be sceptical, because it's so difficult to generate conditions in social sciences when we can draw causal conclusions. Freeman's shoe leather paper provides an excellent critique of this and you are encouraged to read it.

One of the most common approaches to dealing with the omitted variables bias is to add more control variables to the regression. Without loss of generality, consider one additional control. Suppose we had a measure of woman i's years of schooling S_i. Let's again suppose, for now, that schooling is a binary variable for the purposes of simplifying what is being estimated and for causal interpretation. Schooling is likely related to the possible wage a woman could earn if she did work, and so may affect Y_i positively and D_i negatively. The omitted variable bias from leaving S_i out of the regression is negative. The multivariate regression equation is:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_1 S_i + e_i$$

With both binary variables, $\hat{\beta}_1$ is:

$$\hat{\beta}_1 = \mathrm{f}_{\mathrm{S_i}=1} E\left[(Y_i \mid D_i = 1, S_i = 1) - E(Y_i \mid D_i = 0, S_i = 1)\right] + \mathrm{f}_{\mathrm{S_i}=0} E\left[(Y_i \mid D_i = 1, S_i = 0) - E(Y_i \mid D_i = 0, S_i = 0)\right]$$

$$= \mathrm{f}_{\mathrm{S_i}=1} E(Y_{1i} - Y_{0i} \mid D_i = 1, S_i = 1) + \mathrm{f}_{\mathrm{S_i}=0} E(Y_{1i} - Y_{0i} \mid D_i = 1, S_i = 0)$$

$$+ \mathrm{f}_{\mathrm{S_i}=1}\left[E(Y_{0i} \mid D_i = 1, S_i = 1) - E(Y_{0i} \mid D_i = 0, S_i = 1)\right] + \mathrm{f}_{\mathrm{S_i}=0}\left[E(Y_{0i} \mid D_i = 1, S_i = 0) - E(Y_{0i} \mid D_i = 0, S_i = 0)\right]$$

$$= E(Y_{1i} - Y_{0i} \mid D_i = 1)$$

$$+ \mathrm{f}_{\mathrm{S_i}=1}\left[E(Y_{0i} \mid D_i = 1, S_i = 1) - E(Y_{0i} \mid D_i = 0, S_i = 1)\right] + \mathrm{f}_{\mathrm{S_i}=0}\left[E(Y_{0i} \mid D_i = 1, S_i = 0) - E(Y_{0i} \mid D_i = 0, S_i = 0)\right]$$

where $\mathrm{f}_{\mathrm{S_i}=0}$ is the fraction in the sample with schooling equal to zero. If there is no omitted variables bias, $\hat{\beta}_1$ can be interpreted as the average causal effect of having kids.[2]

[2] This causal effect will be weighted by the distribution of S if the conditional expectations function is not actually linear (see Angrist and Kruger, 1999).

The now weighted omitted variable bias is conditional on the additional control variable. A bias remains in estimating the (weighted) average causal effect from kids if any additional factors are related to having kids and hours worked, among women with the same level of education.

A key question in any regression study is whether selection on observables is enough to make the remaining OVB zero. The approach would clearly be acceptable if D_i is randomly assigned conditional on S_i. Whether you win the Green Card lottery, for example depends on your nationality, since some nationalities are not eligible. But conditional on nationality, winning the Green Card is supposed to be random. Even without random assignment, a multivariate regression approach might be plausible if we know a lot about the process generating the regressor of interest, and we have accurately measured variables that capture factors behind this process. Often, however, it is not realistic to believe we understand human behaviour enough to map out an accurate functional form for D_i. In this case, without actually having detailed knowledge of the process that determines whether a woman has kids, we can never be sure we've controlled for other factors by adding more covariates.[3]

The analysis for the continuous variable case is analogous. From our original regression equation $Y_i = \beta_0 + \beta_1 D_i + \beta_1 S_i + e_i$, let $\hat{u}_i$ be the estimated residual after regression D_i on S_i. $\hat{u}_i$ is, by construction, the residual portion of D_i that is uncorrelated with S_i after fitting a linear relationship. Now we have:

$$p\lim \hat{\beta}_1 = \frac{\text{cov}(\hat{u}_i, Y_i)}{\text{var}(\hat{u}_i)} = \beta_1 + \frac{\text{cov}(\hat{u}_i, e_i)}{\text{var}(\hat{u}_i)}$$

[3] Showing that the estimate of β_1 does not change substantially after adding another control variable is sometimes used as evidence that the OVB is negligible because umobserved factors correlated with this additional variable will be partially accounted for (if the correlation was 1, the left out variable would be fully accounted for). If unobserved variables correlated with the control contributed to a significant omitted variables bias, we might expect the estimate of β_1 to change after adding the observed control. Not seeing this is reassuring, but there could still be other factors not correlated with the added control or only weakly correlated that lead to OVB. For more on this, see Altonjii, Elder, and Taber: "Selection on observed and unobserved variables" (unpublished).

The omitted variables bias is now independent of the linear relationship between D_i and S_i. Any other factors that affect Y_i and D_i will lead to a biased causal interpretation, as discussed in the discrete case above. If the relationship in nonlinear, the estimate may still be biased from S_i, with the direction of bias dependent on how $\hat{u}_i$ over predicts D_i and under predicts, (draw figure)

Causal Interpretation when regressor is not binary

When there is no omitted variables bias, we saw above the OLS estimate for D_i can be interpreted as the estimate for the average effect of switching from $D_i = 0$ to $D_i = 0$. The interpretation differs a little when there are more than two possible cases of interest. Let D be the total number of children woman i could have, and $Y_{D,i}$ be woman i's circumstances (or the state of i's universe) with D children. We are interested in knowing the underlying functional relationship that describes how an individual's weeks worked would differ if she had a different number of children. This relationship is person specific, so we write:

(1) $$Y_{D,i} = f_i(D).$$

$Y_{D,i}$ describes the potential (or latent) weeks that person i would work after having D children. The function $f_i(D)$ tells us what i would earn for any value of D. The most common identification strategy for predicting values of $Y_{D,i}$ uses ordinary least squares (OLS) A single variable OLS regression model in this context is

(2) $$Y_i = \beta_0 + \beta_1 D_i + e_i,$$

where Y_i and D_i are actual weeks worked and children observed for individual i respectively.

If the OLS assumption that $Cov(D_i, e_i) = 0$ is satisfied, the estimate of β_1 from this OLS regression has the interpretation as the 'average derivative' of the causal relationship: $E\left[f_i^{'}(D)\right]$.

Proof:

The OLS coefficients minimize: $E\left[Y_i - (\hat{\beta}_0 + \hat{\beta}_1 D_i)\right]^2$

This is the same as minimizing: $E\left[\left(E(Y_i \mid D_i) - (\hat{\beta}_0 + \hat{\beta}_1 D_i)\right) + \left(Y_i - E(Y_i \mid D_i)\right)\right]^2$

Which is the same as minimizing $E\left(E(Y_i \mid D_i) - (\hat{\beta}_0 + \hat{\beta}_1 D_i)\right)^2 + E\left(Y_i - E(Y_i \mid D_i)\right)^2$

Which is the same as minimizing: $E\left(E(Y_i \mid D_i) - (\hat{\beta}_0 + \hat{\beta}_1 D_i)\right)^2$

So $\hat{\beta}_1$ is a linear approximation to for: $E(Y_i \mid D_i = D) - E(Y_i \mid D_i = D-1)$

$= E(f(D)_i \mid D_i = D) - E(f(D-1)_i \mid D_i = D-1)$

$= E(f(D)_i \mid D_i = D) - E(f(D-1)_i \mid D_i = D-1) + [E(f(D-1) \mid D_i = D] - [E(f(D-1) \mid D_i = D]$

$= E(f_i(D) - f_i(D-1) \mid D_i = D) + [E(f(D-1) \mid D_i = D] - E(f(D-1)_i \mid D_i = D-1)$

$= E(f_i(D) - f_i(D-1) \mid D_i = D)$ if $Cov(D_i, e_i) = 0$

$\cong E(f_i^{'}(D)$

Instrumental Variables

Let's return to the single variable regression model of interest:

$$Y_i = \beta_0 + \beta_1 D_i + e_i$$

And let me define D_i here as the total number of children woman i has. Angrist and Evans (1998 AER) adopt an instrumental variables approach to estimating the causal impact of having kids on women's labor supply. An 'instrument' is a variable that affects the independent variable of interest (in this case, D_i), but that variable is unrelated to any unobserved factor that we are worried about which could affect Y_i. Angrist and Evans note that when parents often desire to raise boys and girls. Parents with two sons or two daughters are more likely to have a third child (with the hope that the third one will be of the opposite sex). This can easily be verified and is done in their Table 3.

Suppose we can write this relationship down as:

$$D_i = \delta_0 + \delta_1 Z_i + v_i,$$

where Z_i is equal to 1 if the mother has two boys or two girls,

For an instrument to be valid, the following must be true:

$\text{cov}(D_i, Z_i) \neq 0$, and $\text{cov}(Z_i, e_i) = 0$

In practice, we also would like a strong correlation between D_i and Z_i. If, for example fewer than 1% of the women in our sample had an additional child because their first two were both girls, it could be very hard to detect the effect from the additional child off of such a small sample. The estimated effect will tend to be imprecisely measured. Weak correlations lead to measurement error problems (see below).

Suppose we estimate:

$$\hat{D}_i = \hat{\delta}_0 + \hat{\delta}_1 Z_i$$

This step is called the First Stage. For example, if the portion of mothers with their first 2 children of the same sex that have more than 2 children is 10 percentage points higher than mothers with their first 2 children a boy and a girl, then $\hat{\delta}_1 = .1$. We thus attribute the average difference to having initial same sex children.

Let's substitute the predicted number of children for each individual, based only on whether the individual had initial same sex children, into the main equation:

$$Y_i = \beta_0 + \beta_1 \hat{D}_i + e_i$$

Note that $D_i = \hat{D}_i + \hat{v}_i$: We use only the variation from initial same sex composition to estimate the effect children have on labor supplied.

Note that if initial same sex composition perfectly predicted number of children, then $D_i = \hat{D}_i$. We could just use the original equation if this was the case, since child bearing is only determined by initial same sex composition. On the other hand, if the instrument did not predict child bearing at all, $\hat{\delta}_1$ would be zero, and so we would have no variation to work with.

Regressing Y_i on $\hat{D}_i$ we have:

$$\hat{\beta}_1 = \beta_1 + \frac{\text{cov}(\hat{D}_i, e_i)}{\text{var}(\hat{D}_i)}$$

This estimate is called the Second Stage. From the omitted variable bias formula, our estimate of B1 is unbiased if $\text{cov}(\hat{D}_i, e_i) = 0$, and by assumption it is.

The 'reduced form' is the regression equation between the outcome variable and the instrument:

$$Y_i = \lambda_0 + \lambda_1 Z_i + \varepsilon_i$$

Compare this regression equation of Y_i on $\hat{D}_i$:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 \hat{D}_i + e_i \\ &= \beta_0 + \beta_1 \left(\hat{\delta}_0 + \hat{\delta}_1 Z_i\right) + e_i \\ &= \beta_0 + \beta_1 \hat{\delta}_0 + \beta_1 \hat{\delta}_1 Z_i + e_i \end{aligned}$$

The OLS estimates are the same: $\hat{\lambda}_1 = \hat{\beta}_1 \hat{\delta}_1$.

$$\hat{\beta}_1 = \frac{\hat{\lambda}_1}{\hat{\delta}_1}$$

When Z_i is a binary variable, as it is here, we can rewrite this expression as:

$$\hat{\beta}_1 = \frac{[\overline{Y} \mid Z_i = 1] - [\overline{Y} \mid Z_i = 0]}{[\overline{D} \mid Z_i = 1] - [\overline{D} \mid Z_i = 0]}$$

This is called the Wald estimator, which provides a very intuitive way of thinking about the instrumental variables approach. It says the IV estimate for the β_1 is simply the difference in mean hours worked between women with their first 2 children of the same sex and those without, divided by the difference in the total number of children between women with their first

2 children of the same sex and those without. If having two same sex siblings is associated with a change in labor supply, and having same sex children only leads to an increase in the likelihood of having another child, then dividing this change in labor supply by the average increase in children associated with having initially same sex children will give us an estimate of the increase in labor supply from having one additional child.

Causal interpretation of IV

The IV estimate has an important causal interpretation. if we allow for one more additional assumption. The monotonicity assumption says the instrument works only in one direction. While an instrument may have no effect on some individuals, for those that do react, either the variable being instrumented becomes always larger, or always smaller. In our example, the monotonicity assumption clearly holds, because it’s not possible to have fewer children as a response to having the first 2 children the same sex.

If a valid instrument and the monotonicity assumption, IV estimates capture the average effect of treatment on those who change state in response to a change in the instrument. In this example, the IV approach estimates the average effect of having an additional child for those women who respond to having an additional child because their first 2 children were of the same sex. Imbens and Angrist (JASA 94) call this the Local Average Treatment Effect (LATE).

Note that the LATE may be very different from the average treatment effect (ATE) for the whole population. While we may be interested in the labor supply effects from children from a more general population, IV allows for causal estimates among a very particular subset of the population: those affected by the instrument. It is not correct to extrapolate that this average effect is also the same for other women in the whole population.

Often, however, the individuals we wish to know the treatment effect for are those affected by the instrument.

This might not be the case here. What group of women are we estimating the LATE for here? These are women who want both sex children and are willing enough to have an additional child solely for this reason. Angrist and Evans estimate that, at most, 7% of American women had an additional child as a result of sex preferences (table 3). We are not able to estimate the effect of children on labor supply, for other 93%. For example, the results will not identify the effect from 'accidental' additional children. And the results do not identify the effect from not having any children versus one child (this is obvious in this example, because Angrist and Evan's sample only looks at families with at least 2 children).

Proof of LATE

We only observe whether an individual has kids or not, and whether the binary instrument is zero or not. For example, for a woman with a third child and the first two children of the same sex, we don't know whether she would have had that third child if the first two children were of opposite sex. Let D_{1i} and D_{0i} denote potential outcomes for D_i depending on whether the instrument is assigned or not. We can express the observed outcomes as:

$$D_i = Z_i D_{1i} + (1 - Z_i) D_{0i}$$

4 possibilities for each individual:

$D_{0i} = 0, D_{1i} = 0$
$D_{0i} = 1, D_{1i} = 1$
$D_{0i} = 1, D_{1i} = 0$
$D_{0i} = 0, D_{1i} = 1$

Assume independence: differences in Z are not correlated with differences in D_{1i}, D_{0i}, Y_{1i}, and Y_{0i}

Assume first stage: D and Z are correlated

Assume monotonicity: either $D_{1i} \geq D_{0i} \forall i$ or vice versa.

Monotonicity narrows the set of possible outcomes for D_i from 4 to 3.

$$E(Y_i \mid Z_i = 1) = E[(1-D_i)Y_{0i} + D_i Y_{1i} \mid Z_i = 1]$$
$$= E[Y_{0i} + (Y_{1i} - Y_{0i})D_i \mid Z_i = 1]$$
$$= E[Y_{0i} \mid Z_i = 1] + E[(Y_{1i} - Y_{0i})D_i \mid Z_i = 1]$$
$$= E[Y_{0i}] + E[(Y_{1i} - Y_{0i}) \mid Z_i = 1]E[D_i \mid Z_i = 1]$$

since $E(Y_{1i}D_i) = E(Y_{1i})E(D_i) + \text{cov}(Y_{1i}, D_i) = E(Y_{1i})E(D_i)$ by independence

$$= E[Y_{0i} + (Y_{1i} - Y_{0i})D_{1i}]$$

since $E(D_i \mid Z_i = 1) = E(D_{1i})$

(by independence of Z and Y0i)

Similarly $E(Y_i \mid Z_i = 0) == E[Y_{0i} + (Y_{1i} - Y_{0i})D_{0i}]$

$$\frac{E[Y_{0i} + (Y_{1i} - Y_{0i})D_{1i}] - E[Y_{0i} + (Y_{1i} - Y_{0i})D_{0i}]}{E(D_i \mid Z_i = 1) - E(D_i \mid Z_i = 0)} = \frac{E[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})]}{E(D_{1i} - D_{0i})}$$

$$\frac{E[(Y_{1i} - Y_{0i})(D_{1i} - D_{0i})]}{E(D_{1i} - D_{0i})} = \frac{E[(Y_{1i} - Y_{0i}) \mid D_{1i} > D_{0i}]\Pr(D_{1i} > D_{0i})}{\Pr(D_{1i} > D_{0i})} = E[(Y_{1i} - Y_{0i}) \mid D_{1i} > D_{0i}]$$

if $D_{1i} \geq D_{0i}$, since this implies $D_{1i} - D_{0i}$ is either 0 or 1.

Standard error with IV

Note, that in calculating how precisely B1 is estimated, we have to adjust the variance estimate. It turns out the asymptotic variance of $\hat{\beta}_1$ is:

$$\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{N\sigma^2_{Kids}\rho^2_{Kids,samesex}},$$

where σ^2 is the population variance of v, σ^2_{Kids} is the population variance of Kids, N is the sample size, and $\rho^2_{Kids,samesex}$ is the square of the population correlation between Kids and Samesex. As with the OLS estimator, the variance decreases to zero as the sample size gets larger and larger. All values here can be estimated consistently. For completeness, the estimate of the variance is:

$$\text{var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{SST_{Kids}R^2_{Kids,samesex}},$$

where $\hat{\sigma}^2 = \frac{1}{N-2}\sum \hat{v}_i^2$, $R^2_{Kids,samesex}$ is the R-squared obtained after running the regression of Kids on samesex, and SST is the total sum of squares of the Kids variable.

Pitfalls with Instrumental Variables

1) Is the instrument valid?

The assumption that $\text{cov}(Z_i, e_i) = 0$ is just an assumption . Often in when using this approach, the credibility of the results will rest on this. Angrist and Evans show that same sex composition is uncorrelated with many observable background characteristics (Table 4). This doesn't show conclusively that other unobserved variables are uncorrelated with this instrument, but it helps.

2) a 'weak' instrument

An instrumental variable is said to be 'weak', if $corr(D_i, Z_i)$ is small. We can rewrite the asymptotic estimate of B1 as:

$$\hat{\beta}_1 = \beta_1 + \frac{\text{cov}(\hat{D}_i, e_i)}{\text{var}(\hat{D}_i)}$$

$$= \beta_1 + \frac{\text{cov}(\hat{\delta}_1 Z_i, e_i)}{\text{var}(\hat{D}_i)}$$

$$= \beta_1 + \frac{\hat{\delta}_1 \text{cov}(Z_i, e_i)}{\text{var}(\hat{D}_i)}$$

$$= \beta_1 + \frac{\hat{\delta}_1 \text{cov}(Z_i, e_i)}{\hat{\delta}_1^2 \text{var}(Z_i)}$$

$$= \beta_1 + \frac{\text{var}(Z_i)}{\text{cov}(D_i, Z_i)} \frac{\text{cov}(Z_i, e_i)}{\text{var}(Z_i)}$$

$$= \beta_1 + \frac{\text{var}(Z_i)}{\text{cov}(D_i, Z_i)} \frac{\text{cov}(Z_i, e_i)}{\text{var}(Z_i)}$$

$$= \beta_1 + \frac{corr(Z_i, e_i)}{corr(D_i, Z_i)} \frac{\sigma_e}{\sigma_D}$$

Compare this last term to the OLS estimate of B1:

$$\hat{\beta}_1 = \beta_1 + \frac{\text{cov}(D_i, e_i)}{\text{var}(D_i)}$$

$$= \beta_1 + corr(D_i, e_i) \frac{\sigma_e}{\sigma_D}$$

If there is ANY remaining correlation between the instrumental variable and unobserved factors affecting the outcome, then this omitted variables bias will be exacerbated by a weak instrument. That is, a weak correlation between the instrument and the independent variable will make the omitted variables bias (if it exists) larger. Compared to the OLS estimate, if there instrument is not completely valid, it is not clear whether the OLS estimate or the IV estimate will be less biased. In fact, OLS will give less biased results if:

$$corr(D_i, e_i) < \frac{corr(Z_i, e_i)}{corr(D, Z_i)}.$$

Thus, when using instrumental variables, it is always important to justify your instrument (convince us that it is uncorrelated with possible omitted variables) and show that the relationship between the instrument and the independent variable of interest is significant.

A quick note about using IV with multiple regression.

The IV estimator for the simple regression model is easily extended to the multiple regression case. One key thing to remember is that all of the independent variables (except the one you are instrumenting) must be used in the First Stage. For example, if the equation being estimated was:

$$H_i = \beta_0 + \beta_1 Kids_i + \beta_2 Z_i + v_i$$

where Z was some control variable, the First Stage would be:

$$Kids_i = \delta_0 + \delta_1 samesex_i + \delta_2 Z_i + e_i$$

We must include these controls when estimating the effect the instrument has on influencing the independent variable of interest.

What do Angrist and Evans find?

Table 5 shows results with no additional controls.

Table 7 shows results with controls, and compares OLS results with IV results.

Note, they also use having twins after the first child as an instrument for having 3 kids instead of 2, but I will not talk much about this. Feel free to read more about it in the paper.