ARE 213 **Applied Econometrics**

UC Berkeley Department of Agricultural and Resource Economics

Selection On Unobservables Designs:

Part 5, IV with Treatment Effect Heterogeneity

All of our previous IV discussion concentrates on IV in the context of homogeneous treatment effects. This was the focus of IV estimation for the first 50 years, but it doesn't fit in with our discussion of heterogeneous treatment effects at the beginning of the course. Recall the distinction between ATE – the average treatment effect for a randomly drawn individual in our sample – and TOT – the average treatment effect for a randomly drawn treated individual in our sample. With IV, these distinctions become more interesting. We have sidestepped this discussion so far by assuming homogeneous treatment effects, so ATE is equal to TOT, and both are equal to the average treatment effect for any other subpopulation one might think of. If we allow for heterogeneous treatment effects, however, what is it that IV actually estimates? ATE? TOT? The answer, presented in Angrist, Imbens, and Rubin's seminal 1996 paper (henceforth AIR 1996), is "neither."

1 Intuition

I will proceed somewhat unconventionally by first explaining intuitively what IV estimates in the context of heterogeneous treatment effects and then presenting the mathematical proof. My hope is that understanding the terms and concepts intuitively will make the math easier to interpret. What IV generally estimates is the "local average treatment effect," or LATE. LATE is the average treatment effect of d_i on y_i for the units for whom changing the instrument (changing z_i) changes their treatment status (changes d_i). This is somewhat abstract, but it should become clearer in the context of our two examples, the medical trial and the quarter of birth instrument.

What does it mean to say that IV estimates the average treatment effect of d_i on y_i for

the units for whom changing z_i changes d_i? In practice, this is best illustrated in the medical trial example. In this example, there are four potential types of people. Note that not all of these types need exist in practice; in fact, we will explicitly rule out one type by assumption when we do the proof. The first type are people who always take the pill, regardless of whether they are assigned to the treatment group or the control group.[1] In the language of AIR 1996, we call these people "always-takers." The second type are people who never take the pill, regardless of whether they are assigned to the treatment group or the control group. We call these people "never-takers." The third type are people that take the pill if and only if they are assigned to the treatment group. We call these people "LATE-compliers." Finally, the fourth type are people who take the pill if and only if they are in the control group. We call this perverse group the "LATE-defiers," and we rule them out by assumption.

The people "for whom changing z_i changes their value of d_i" are the people who take the pill if and only if they are in the treatment group, i.e. the LATE-compliers (recall that assignment to treatment versus control group is the instrument in this example). The always-takers are unaffected by the instrument, because they take the treatment regardless of whether they are in the treatment or control group. Likewise, the never-takers are also unaffected by the instrument, because they eschew the treatment regardless of whether they are in the treatment or control group. The defiers are ruled out by assumption. Therefore, the IV estimator estimates the effect of the pill on blood pressure for the people who take the pill if they are in the treatment group but do not take it if they are in the control group. If the effect is homogeneous, then this distinction is irrelevant, but if the effect varies across individuals, then this distinction can become important.

Suppose that there are two types of people: people who respond to the pill and people who do not respond to the pill. This is not a far-fetched assumption – most medical trials find that the treatment is successful in treating some cases, but unsuccessful in treating other cases. So β_1 is negative for people who respond to the pill (remember that we think the pill should lower blood pressure), and β_1 is zero for people who do not respond to the

[1] You might wonder how the control group could get the pill. Think about terms like "black market" or "prescription abuse."

pill. Further suppose that people who respond to the pill know that they will respond to it (don't ask me how), so they always take it, regardless of whether they are in the treatment or the control group. However, the people for whom the treatment has no effect take the pill only if they are in treatment group (when they are given the pill for free), and not if they are in the control group. We know that IV estimates the effect of the treatment on the LATE-compliers, i.e. the people that take it if and only if they are in the treatment group. Therefore, in this case, IV estimates the effect of the treatment on the people for whom the treatment has no effect, because they are the only ones for whom the instrument changes whether or not they take the pill. So IV will estimate $\beta_1 = 0$ in this example, despite the fact that the average treatment effect is negative.[2]

Does this mean that IV is inconsistent? Not really – it is simply providing a consistent estimate of the local average treatment effect (the average effect for the people for whom changing the instrument changed d_i), not the average treatment effect for the entire population or sample. As long as you interpret IV correctly, then it is not inconsistent. Of course, it may not estimate what you want to estimate (which might be ATE or TOT), but that's the way the cookie crumbles. So the lesson here is that IV is consistent, but that you have to be careful in thinking about exactly what it is estimating. Importantly, IV estimates the average treatment effect for individuals that "comply" with the instrument. Since different instruments will have different sets of compliers, it follows that different instruments can plim to different values, even if all the instruments under consideration meet the two criteria for valid instruments. This result basically invalidates overidentification tests as a valid scientific testing procedure and has implications for instrumenting for multiple endogenous variables simultaneously.

Why does IV estimate LATE in our example? As I have reiterated many times, the IV estimator is the reduced form divided by the first stage. So if the IV estimate is 0 in the example I discussed above, that means that the reduced form must be 0. In the medical

[2]Of course, we could alternatively construct a scenario in which the individuals with no treatment effect are the never-takers and the individuals with a negative treatment effect are the LATE-compliers. In that scenario, IV would produce a negative estimate of β_1, but the magnitude would be larger than ATE.

trial example, the reduced form is the mean blood pressure for the treatment group minus the mean blood pressure for the control group. Since the always-takers take the pill when they are in the treatment group and when they are in the control group, their mean blood pressure will not be any different when they are in the treatment group than it is when they are in the control group. So those people will never contribute anything to moving the reduced form away from zero. The people who can potentially move the reduced form away from zero are the people who take the treatment when they're in the treatment group but do not take it when they are in the control group. But we assumed that those were the people for whom the pill had no effect, so of course their mean blood pressure in the treatment group is not any different than their mean blood pressure in the control group. Thus we get a reduced form of 0 in our example.

If the pill did have an effect for these people, then the reduced form would be capturing that effect, and we would get a nonzero coefficient estimate. That coefficient would represent the total effect of the pill averaged over all of the individuals in the treatment group. In fact, however, only the LATE-compliers were affected. The IV thus rescales the reduced form by the first stage because the first stage estimates, in our example, the fraction of the sample that are LATE-compliers (i.e., the fraction that changed their value of d_i in response to being assigned to the treatment group).

To reiterate, the always-takers and the never-takers do not, in expectation, contribute anything to moving the reduced form away from zero, because for them the treatment indicator is always the same in the treatment group and the control group (and the random assignment procedure balances them, on average, across treatment and control). Thus their mean blood pressure is no different in the treatment group than it is in the control group. Therefore, the only group of people who can move the reduced form away from zero is the group of LATE-compliers, because for them the treatment level actually varies depending on whether they are in the treatment group or in the control group. So if the treatment has an effect for them, then their mean blood pressure will be different in the treatment group than it is in the control group. But by definition, the LATE-compliers are the people for

whom changing z_i changes d_i. Thus IV estimates the average treatment effect for the people for whom changing z_i changes d_i, because those are the people who drive the reduced form, and IV is just the reduced form rescaled by the first stage.

Before showing the formal derivation of LATE, I will explain how it applies to the quarter of birth example. Recall that in the quarter of birth example, the instrument works because some people stay in school only as long as they are legally required to, and then they drop out as soon as they reach age 16. These are the people for whom the instrument z_i (quarter of birth) has an effect on d_i (years of school). If it helps, you could literally imagine a 15.5 year old potential dropout who was born in the third quarter thinking to himself, "If only I had been born in the first quarter, then I would be able to drop out of school right now, because I'd already be 16. But instead I have to stay in school until the third quarter and receive 11 years of schooling instead of 10.5 years of schooling!" These people are the equivalent of the LATE-compliers (they don't have to actually think in this manner though!). In contrast, however, for the vast majority of people the instrument (quarter of birth) has no effect on how long they stay in school, because they plan to stay in school long past the age at which they can legally dropout. They are the equivalent of the always-takers.[3]

Since IV estimates the causal effect of d_i (schooling) on y_i (wages) for the people for whom the instrument z_i (quarter of birth) changes their value of d_i, the IV estimate gives us the average effect of schooling on wages for people who drop out as soon as they are no longer legally required to stay in school. So the quarter of birth instrument is really estimating the average effect of an additional year of schooling on wages for high school dropouts. Is there any reason to believe that this is the same effect of schooling that the "average" person would have? Probably not. On the one hand, it may overestimate the "average" effect of schooling if we believe that wages are a concave function of schooling, so that the return to schooling falls as you get more schooling.[4] On the other hand, it may underestimate the "average"

[3]The never-takers would be the ones that disregard the law entirely and drop out of school long before they are legally allowed to.

[4]This phenomenon has been referred to as "discount rate bias" because a simple human capital model implies that an individual should stay in school until her return to schooling equals her discount rate. Students that drop out early do so because they have higher discount rates, and their marginal return to

effect of schooling if we believe that high school dropouts don't apply themselves in school anyway, so they don't get much out of being in school. Either way, the point is that the IV regression is estimating the average effect of schooling on wages for high school dropouts rather than for the entire population. It consistently estimates this effect, but this effect is probably different than the population average effect of schooling on wages. Thus we need to be careful about how we interpret the result. Finally, note that the reason IV estimates the average effect of schooling on wages for high school dropouts is not because our sample only consists of high school dropouts. The sample is taken from the entire population, but IV only estimates the average effect of d_i on y_i for the LATE-compliers (i.e., the high school dropouts), not the average effect for the entire population. However, if our policy interest pertains to students at risk of dropping out, the average effect for LATE-compliers may be very informative.

2 Proof

Let D_i be a binary treatment, Z_i a binary instrument, and Y_i an outcome. Let $\mathbf{Z}$ be an N-dimensional vector that contains the value of the instrument, Z_i, for each unit in the data set, and $\mathbf{D}$ be a similar vector for the treatment variable. We define the potential outcome $Y_i(\mathbf{Z}, \mathbf{D})$ as the potential outcome for unit i under a given vector of values for the instrument and a given vector of values for the treatment. Since D_i is assumed to be affected by Z_i, we also define the potential outcome $D_i(\mathbf{Z})$ as the potential outcome for the treatment under a given vector of values for the instrument.

As discussed above, there are four types of individuals: always-takers, never-takers, LATE-compliers, and LATE-defiers. Table 1 presents each type using the potential outcomes notation. $D_i(Z_i)$ is constant for the never-takers and always-takers – the instrument doesn't affect their choice to get treated or not get treated. $D_i(Z_i)$ changes positively with Z_i for the LATE-compliers – they "comply" with their intention to treat assignment. $D_i(Z_i)$

schooling is higher. However, the term "discount rate bias" is somewhat deceptive in the sense that it's not really an issue of bias but rather an issue of heterogeneous treatment effects and external validity.

changes negatively with Z_i for the LATE-defiers – they "defy" their intention to treat assignment.

Table 1: Types of Individuals by $D_i(0)$ and $D_i(1)$

		$D_i(0)$	
		0	1
$D_i(1)$	0	Never-taker	LATE-defier
	1	LATE-complier	Always-taker

We need to make several assumptions before proceeding. First, take as given the Stable Unit Treatment Value Assumption (SUTVA):

1. If $Z_i = Z_i^*$, then $D_i(\mathbf{Z}) = D_i(\mathbf{Z}^*)$.
2. If $Z_i = Z_i^*$ and $D_i = D_i^*$, then $Y_i(\mathbf{Z}, \mathbf{D}) = Y_i(\mathbf{Z}^*, \mathbf{D}^*)$.

As we discussed earlier, SUTVA basically amounts to assuming that the treatment is well-defined and that there is no interference between units. The causal effect of Z_i on D_i is $D_i(1) - D_i(0)$. The causal effect of Z_i on Y_i is $Y_i(1, D_i(1)) - Y_i(0, D_i(0))$.

We assume that Z_i is randomly assigned. We also assume that the exclusion restriction holds, i.e.,

$$Y(\mathbf{Z}, \mathbf{D}) = Y(\mathbf{Z}', \mathbf{D}) \; \forall \; \mathbf{Z}, \mathbf{Z}', \mathbf{D}$$

In other words, for a given value of D_i, it doesn't matter what the value of Z_i is – the instrument only matters insofar as it affects the treatment. Given the exclusion restriction, we can then define the causal effect of D_i on Y_i as $Y_i(1) - Y_i(0)$ (we no longer need to include Z_i as an argument in Y_i because it is irrelevant conditional on D_i). This is the causal effect of interest.

We assume that the instrument has a nonzero effect on the treatment:

$$E[D_i(1) - D_i(0)] \neq 0$$

This is equivalent to our normal IV assumption that the instrument is correlated with the treatment. Finally, we impose a "monotonicity assumption" stating that the instrument does not change treatment status in opposite directions for different units:

$$D_i(1) \geq D_i(0) \; \forall \; i = 1, ..., N$$

The monotonicity assumption rules out the possibility of LATE-defiers, i.e., individuals that change from $D_i = 1$ (treated) to $D_i = 0$ (untreated) when their instrument changes from $Z_i = 0$ (intended to not treat) to $Z_i = 1$ (intended to treat).

Under these assumptions (which are basically the typical IV assumptions, except for the monotonicity assumption), what does IV estimate? To answer this question, we leverage the fact that the IV estimator can be written as a ratio of the reduced form over the first stage. First, consider the causal effect of Z on Y for unit i:

$$Y_i(1, D_i(1)) - Y_i(0, D_i(0)) =$$

$$[Y_i(1)D_i(1) + Y_i(0)(1 - D_i(1))] - [Y_i(1)D_i(0) + Y_i(0)(1 - D_i(0))] =$$

$$(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))$$

If the equality of the first and second lines is not clear, recall in our first lectures that we defined $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$. We are doing exactly the same thing here when moving from the first line to the second line. The only difference is that we now have an additional layer of complexity from introducing the instrument Z_i, as D_i is now a function of Z_i.

The result above implies that the causal effect of Z on Y for unit i equals $(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))$. This formulation is convenient in part because it implies that Z has no causal effect on Y for the always-takers and the never-takers. For these two groups, $D_i(1) = D_i(0)$, so $(Y_i(1) - Y_i(0))(D_i(1) - D_i(0)) = 0$. This confirms my earlier claim that

"[the always-takers and never-takers] never contribute anything to moving the reduced form away from zero."

Table 2: Causal Effect of Z on Y by $D_i(0)$ and $D_i(1)$

		$D_i(0)$	
		0	1
$D_i(1)$	0	$Y_i(1,0) - Y_i(0,0) = 0$ Never-taker	$Y_i(1,0) - Y_i(0,1) = -(Y_i(1) - Y_i(0))$ LATE-defier
	1	$Y_i(1,1) - Y_i(0,0) = Y_i(1) - Y_i(0)$ LATE-complier	$Y_i(1,1) - Y_i(0,1) = 0$ Always-taker

Source: Angrist, Imbens, and Rubin (1996).

Table 2 summarizes the causal effect of Z on Y for each type. For always-takers and never-takers, there is no effect of Z on Y, as argued above. For LATE-compliers, the effect of Z on Y is $Y_i(1) - Y_i(0)$, i.e., the difference in their potential outcomes under $D_i = 1$ and $D_i = 0$. For LATE-defiers, the effect of Z on Y is $Y_i(0) - Y_i(1)$, i.e., the difference in their potential outcomes under $D_i = 0$ and $D_i = 1$. It is the opposite of the effect for the LATE-compliers because the instrument (Z) affects the treatment (D) in the opposite manner for defiers vis a vis compliers.

What is the average causal effect of Z on Y?

$$E[Y_i(1, D_i(1)) - Y_i(0, D_i(0))] =$$

$$E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))] =$$

$$E[E[(Y_i(1) - Y_i(0))(D_i(1) - D_i(0))|D_i(1) - D_i(0)]] =$$

$$E[(D_i(1) - D_i(0))E[Y_i(1) - Y_i(0)|D_i(1) - D_i(0)]] =$$

$$1 \cdot E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \cdot P(D_i(1) - D_i(0) = 1)$$

$$-1 \cdot E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = -1] \cdot P(D_i(1) - D_i(0) = -1) =$$

$$= E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \cdot P(D_i(1) - D_i(0) = 1)$$

The equality between the fourth and fifth/sixth lines holds because we do not have to consider individuals for which $D_i(1) - D_i(0) = 0$ (i.e., the never-takers and always-takers). The last equality holds because we rule out LATE-defiers by assumption, i.e., we assume that $P(D_i(1) - D_i(0) = -1) = 0$. Thus we conclude that the average causal effect of Z on Y is $E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \cdot P(D_i(1) - D_i(0) = 1)$.

We know the IV estimator is equal to the reduced form divided by the first stage, so its limit must equal the ratio of the limits of those two estimators. The limit of the reduced form is average causal effect of Z on Y, or $E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \cdot P(D_i(1) - D_i(0) = 1)$. The limit of the first stage is the average causal effect of Z on D, or $E[D_i(1) - D_i(0)] = P(D_i(1) - D_i(0) = 1)$. Thus the IV estimand is:

$$\frac{E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1] \cdot P(D_i(1) - D_i(0) = 1)}{P(D_i(1) - D_i(0) = 1)} =$$

$$E[(Y_i(1) - Y_i(0))|D_i(1) - D_i(0) = 1]$$

But $D_i(1) - D_i(0) = 1$ if and only if an individual is a LATE-complier. Thus IV estimates the average effect of D on Y for LATE-compliers.

3 The Monotonicity Assumption

All of the assumptions we made above are standard textbook IV assumptions with the exception of the monotonicity assumption, i.e., the assumption that Z only changes D in one direction (or not at all). What happens if the monotonicity assumption is not met? In that case, we cannot drop out the last term in our derivation of the causal effect of Z on Y; the reduced form becomes $\tau_c P(\text{complier}) - \tau_d P(\text{defier})$, where τ_c and τ_d are the average treatment effects of D on Y for compliers and defiers respectively. The first stage becomes

$E[D_i(1) - D_i(0)] = 1 \cdot P(\text{complier}) - 1 \cdot P(\text{defier})$. Thus the IV estimand is:

$$\frac{\tau_c P(\text{complier}) - \tau_d P(\text{defier})}{P(\text{complier}) - P(\text{defier})}$$

This looks like a simple weighted average; the danger, however, is that the weight for defiers can take on negative values. There is thus no guarantee that the IV estimand need lie between τ_c and τ_d.[5] So in general it's probably best if you can make a case that the monotonicity assumption holds.

4 Multi-valued Treatments and Instruments

Angrist and Imbens (1995) discuss cases in which the treatment or instrument is not binary. We deal first with the case in which the treatment is not binary. Suppose the treatment, D_i, takes on $J+1$ values ($J > 1$). In that case, we can write the potential outcome $Y_i(D_i)$ as a quantity that has a different value for each value of D_i: $Y_i(0)$, $Y_i(1)$,...,$Y_i(J)$. Note that the notation is identical to the notation we use when D_i is binary, except that now $Y_i(D_i)$ can take on $J+1$ different values instead of just 2 different values. Our instrument is still binary, however, so $D_i(Z_i)$ still has only two possible values: $D_i(0)$ or $D_i(1)$.

There are now J different causal effects of D on Y. There is the causal effect of changing D_i from 0 to 1, the causal effect of changing D_i from 1 to 2, the causal effect of changing D_i from 2 to 3, and so on, through the causal effect of changing D_i from $J-1$ to J. In fact, we could imagine even more causal effects by changing D_i by more than one unit (e.g., the effect of changing D_i from 0 to 5), but these additional effects will just be sums of the J causal effects that we already defined.

With a binary instrument and a multi-valued treatment, the IV estimator converges to:

$$\frac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]} = \sum_{j=1}^{J} w_j \cdot E[Y_i(j) - Y_i(j-1) \mid D_i(1) \geq j > D_i(0)]$$

[5] Consider, for example, a case in which $\tau_c = 3$, $\tau_d = 1$, $P(\text{complier}) = 2/3$, and $P(\text{defier}) = 1/3$.

where the weights in the sum are equal to $w_j = \frac{P(D_i(1) \geq j > D_i(0))}{\sum_{l=1}^{J} P(D_i(1) \geq l > D_i(0))}$.

To interpret the expression above, it's easiest to assume that the instrument induces no more than a single unit change in D_i for any individual i.[6] In that case, for any individual, $D_i(1) - D_i(0)$ is equal to zero or one. If $D_i(1) - D_i(0) = 0$, then an individual is a noncomplier, and he contributes nothing to the IV estimand. Indeed, you can see that the conditional expectation in the sum is conditioned on $D_i(1)$ being greater than $D_i(0)$. If $D_i(1) - D_i(0) = 1$, then the individual is a complier, but there are now J different types of compliers. There are compliers that move from $D_i(0) = 0$ to $D_i(1) = 1$, compliers that move from $D_i(1) = 1$ to $D_i(2) = 2$, and so on. Let us call a complier that moves from $D_i(0) = j$ to $D_i(1) = j + 1$ to be a "complier of type j." The sum above then takes a weighted average of the average treatment effects for each type of complier. When $j = 1$, the conditional expectation in the sum is the average treatment effect for individuals for whom $D_i = 1$ when assigned to the treatment group and $D_i = 0$ when assigned to the control group. When $j = 2$, the conditional expectation in the sum is the average treatment effect for individuals for whom $D_i = 2$ when assigned to the treatment group and $D_i = 1$ when assigned to the control group. And so on. The weights, w_j, are equal to the share of compliers that are of type j. IV therefore estimates a weighted average of J local average treatment effects (one local average treatment effect for each complier type), with weights equal to the share of compliers that are of type j. Angrist and Imbens describe IV as estimating "a weighted average of per-unit average causal effects along the length of an appropriately defined causal response function." They refer to this quantity as the "average causal response" (ACR).

In other words, changing the treatment by one unit has different average effects at different values of the treatment. IV estimates a weighted average of these different effects, and each effect is weighted by its share of the compliers. In the quarter-of-birth schooling example, IV estimates an average effect of moving from 10th to 11th grade or 11th to 12th grade for high school dropouts that comply with compulsory schooling laws. Although the

[6] It's not a problem if the instrument induces a multi-unit change in D_i – it just makes the expression more complicated to interpret because a single individual can now appear in the sum for multiple values of j.

causal effect of schooling on earnings varies from first grade to graduate study, all of the compliers in the quarter-of-birth example are individuals getting between 10 to 12 years of schooling (not counting kindergarten). Hence the weights in the sum above are zero for all j except $j = 11$ and $j = 12$.

Now consider the case in which the instrument, Z_i, can take on $K + 1$ distinct values. There are now K distinct average causal responses (ACRs), one for each point at which the instrument changes. The ACR at point $Z_i = k$ is:

$$\beta_{k,k-1} = \frac{E[Y_i|Z_i = k] - E[Y_i|Z_i = k-1]}{E[D_i|Z_i = k] - E[D_i|Z_i = k-1]}$$

In other words, we can think about changing the instrument Z_i by one unit at each point k, from $k-1$ to k. Changing the instrument by one unit at point k allows us to identify an average causal response at point k, and we label that ACR as $\beta_{k,k-1}$. You could alternatively think of each point k as representing a separate binary instrument (in fact, that is how Angrist and Imbens motivate this formula). The IV estimate is then a weighted average of K ACRs. Specifically, the IV estimate converges to:

$$\sum_{k=1}^{K} \mu_k \cdot \beta_{k,k-1}$$

with weights μ_k defined as:

$$\mu_k = (E[D_i|Z_i = k] - E[D_i|Z_i = k-1])$$
$$\cdot(E[D_i|Z_i \geq k] - E[D_i|Z_i < k]) \cdot P(Z_i \geq k) \cdot (1 - P(Z_i \geq k))$$

The first part of the expression for μ_k, $E[D_i|Z_i = k] - E[D_i|Z_i = k-1]$, implies that points along the instrument that induce larger changes in D_i (i.e., points that have a stronger first stage) will receive more weight. This should be intuitive since a stronger first stage

should give us more leverage in identifying the effect of D_i on Y_i. The second part of the expression for μ_k implies that points of Z_i near the median of Z_i receive more weight since $P(Z_i \geq k) \cdot (1 - P(Z_i \geq k))$ is maximized at the median of Z_i (points that split Z_i such that the average value of D_i is much larger in the upper part of Z_i than in the lower part of Z_i also receive more weight).

In summation, when the instrument Z_i can realize $K + 1$ multiple values, IV estimates a weighted average of the K average causal responses that correspond to increasing the instrument by one unit at points 1, 2, 3, ..., K. The weight used for the kth ACR is proportional to the strength of the first stage at $Z_i = k$.

5 Summary

We have seen in this section that IV estimates the "local average treatment effect," or LATE. This is the average treatment effect for units that are induced by the instrument to change their treatment status. The clear application of this finding is that it allows us to think more precisely about which group of individuals our treatment effect estimate applies to. There are, however, other important implications.

Most importantly, the LATE result implies that, in the presence of treatment effect heterogeneity, different instruments should produce different estimates, even in arbitrarily large samples. The choice of instrument defines the group of LATE-compliers; different instruments therefore estimate the average treatment effect for different groups of LATE-compliers. There is no reason why these averages need be equal for different groups.

The fact that different instruments can produce different treatment effect estimates (even absent sampling error) calls into question the general utility of overidentification tests. These tests compare coefficient estimates produced by different instruments – the idea is that if the instruments are all valid, all the estimates should be equal (up to sampling error). If some instruments are invalid, however, the estimates produced by different instruments may differ. In the context of heterogeneous treatment effects, however, we know that different instru-

ments can produce different coefficient estimates even if all of the instruments are internally valid. Thus it is impossible to ever "reject" the validity of the instruments, making the overidentification tests scientifically questionable. The same critique holds for the Hausman test, which compares the IV estimate to the OLS estimate. With heterogeneous treatment effects, there is no reason that OLS (which, under ideal conditions, will estimate ATE or TOT) need equal IV (which estimates LATE).

Heterogeneous treatment effects also complicate matters when you have multiple endogenous variables that you want to instrument for. Consider, for example, a simple case in which you wish to simultaneously estimate the effect of education (d_1) and experience (d_2) on earnings (y). The model might look like:

$$y_i = \beta_0 + \beta_1 d_{1i} + \beta_2 d_{2i} + \varepsilon_i$$

Both "treatments" are subject to selection issues and are endogenously determined. Instrumenting for education and controlling for experience as a covariate will not give consistent estimates of the effect of education on earnings – it is inappropriate to control for a variable that is affected by the treatment (in general, getting more education will mean getting less job experience). The correct way to estimate the causal effect of education on earnings is to instrument for education and include as covariates only predetermined variables.

If, however, you want to estimate a structural model that contains both education and experience, i.e., you want to know the effect of education when holding experience constant (even though we may not be able to imagine such a scenario in real life), you might find two instruments, one for education (call it z_1) and one for experience (call it z_2).[7] You can then identify β_1 and β_2 by running 2SLS, using both z_1 and z_2 as instruments. Intuitively, 2SLS is using z_2 to estimate β_2, and then using this estimate of β_2 to adjust for the fact that z_1 affects both d_1 and d_2 when estimating β_1 (i.e., the effect of education on earnings holding experience constant).

[7] Of course, the education instrument will invariably affect experience. In principle, however, the experience instrument need not affect education.

With homogenous treatment effects, this strategy is valid. With heterogenous treatment effects, however, we know that different instruments generally estimate different local average treatment effects. Assuming that z_1 estimates the same treatment effect for d_2 that z_2 estimates is therefore unjustified.[8] In principle, the effect of manipulating education while holding experience constant could be positive for all individuals, yet the 2SLS procedure could generate a negative estimate of β_1 (even ignoring sampling error).

6 Additional References

Angrist, J. and G. Imbens. "Two-stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association*, 1995, 90, 431-442.

[8]This is equivalent to assuming that z_1 and z_2 should both produce identical estimates of β_2.

$Y_i(0) = \mu_0(X_i) + u_{0i} \qquad Y_i(1) = \mu_1(X_i) + u_{1i}$

$\tau_i = Y_i(1) - Y_i(0) = \left(\mu_1(x) - \mu_0(x)\right) + \left(u_1 - u_0\right)$

$= X_i(\beta_1 - \beta_0) + (u_{1i} - u_{0i})$

$D_i^* = \mu_0(X_i, Z_i) - V_i \longrightarrow D_i = 1 \iff D_i^* > 0$

$\mu_D(X, Z) = X_i\gamma + Z_i\pi$

Assume $(u_1, u_0, V) \perp\!\!\!\perp Z \mid X$, $\pi > 0$, $\pi_i \geq 0$

$P(X, Z) = P\left(V < \mu_D(X, Z)\right) = F_V\left(\mu_D(X_i, Z_i)\right)$

so let $u_D = F_V(V)$ and $u_D \sim Unif($

$ATE(x) = E\left[\tau_i \mid X_i = x\right] = \mu_1(x) - \mu_0(x)$

$ATOT(x) = E\left[\tau_i \mid X_i = x, D_i = 1\right] = \mu_1(x) - \mu_0(x) + E\left[u_{1i} - u_{0i} \mid X_i = x, D_i = 1\right]$

Policy-Relevant Treatment Effect (PRTE) $= ATE(x) + E\left[u_{1i} - u_{0i} \mid X_i = x, D_i' = 1\right]E\left[D_i' \mid X_i\right] - \dfrac{E\left[u_{1i} - u_{0i} \mid X_i = x, D_i = 1\right]E\left[D_i \mid X_i = x\right]}{E\left[D_i' \mid X_i = x\right] - E\left[D_i \mid X_i = x\right]}$

effect on people moved into treatment due to the policy

$MTE(x, u_d) = E\left[Y_i(1) - Y_i(0) \mid X_i = x, U_D = u_d\right]$

$= ATE(x) + E\left[u_{1i} - u_{0i} \mid X_i = x, U_D = u_d\right]$

D_i' = your treatment status under the policy

D_i = your treatment status w/o the policy

$MTE(X = x, U_D = p) = \frac{\partial}{\partial p} E\left[Y \mid X = x, P(x, z) = p\right]$

Estimation:

① Estimate logit p-score $P(X, z)$

② Estimate $E[Y \mid X, \hat{p} \cdot X, \hat{p}]$ ⟶ for example, if non-parametric, could look at variation of $\hat{p}$ in each X cell — all variation must come from Zs, so boils down to an IV type approach