PHS2000B: Instrumental Variables Lab

March 23, 2023

Endogeneity, a brief review

Although many economists realize the limitation of describing the world using only linear functions, many basic ideas in Econometrics are still expressed using linear models. One such idea is that of **endogeneity**.

Suppose we are interested in estimating the effect of treatment schooling, S, on income, Y, and suppose we write out the following model:

$$y_i = \beta_0 + \beta_1 s_i + \epsilon_i$$

where $s \in S$, $y \in Y$, i indexes each individual, and ϵ represents the population error term.

We define s to be endogenous if $cov(s_i, \epsilon_i) \neq 0$. When s is endogenous, our estimates for β_1 will be biased and inconsistent. In an Econometrics text book, you will find three reasons cited for endogeneity:

1) One or more confounders are not included in the model (i.e., we have **omitted variables**). So far, we've been thinking of this as "unmeasured confounding".
2) The observed values of the exposure is a function of the true value of the exposure and some **measurement error**.
3) The outcome is a function of the exposure but the exposure is also a function of the outcome (i.e., we have **reverse causality**). This is something we don't really seem to focus on in Epi studies. However, reverse causality is a source of great concern in the Econ literature where studies might be trying to understand people's preferences for goods or services, and current preferences may be highly influenced by our ideas about future outcomes.

Keep in mind that ideas of omitted variable bias, measurement error, and reverse causality apply to non-linear settings as well. For instance, we've been worried about unmeasured confounding even in cases where we believe the exposure-outcome relationship needs to be modeled using, for instance, the exponential function.

Overcoming endogeneity

1) **Randomized treatment assignment**: Randomized controlled trials are considered to be the "gold standard" method for overcoming endogeneity. As we've discussed before in this course, randomization helps to ensure exchangeable treatment and control groups in expectation, which allows us to consistently estimate the causal effect of the treatment on the outcome. In Econometrics parlance, we would say that randomization gets rid of "selection bias", which is just another way of saying that randomization ensures lack of confounding. For more on selection bias and why randomization solves this problem, please see **Appendix 1**. For a discussion on randomized controlled trials, please see **Appendix II**.
2) **Closing all back-door paths from exposure to outcome:** Randomization is, of course, not always possible for various reasons. It might be too expensive, it might be unethical, it might be too slow. Another strategy is to control for variables that block all back-door paths from the exposure to the outcome. This is perfectly valid approach grounded in very sound theory. Practically, however,

it requires two things: first, it requires us knowing all the confounders of the exposure-outcome relationship; second, it requires us being able to measure all of these confounders. Economists have been historically skeptical of the idea that we can know every possible confounder of the exposure-outcome relationship we are interested in. Perhaps this has to do with the types of problems that have been of interest to them: for instance, one of the key questions in Labor Economics is what we discussed above - the impact of schooling on income. Can you construct a set of all possible confounders of this relationship?

3) **Instrumental variables:** A third approach involves using a separate variable, called an "instrument", to identify the effect of the exposure on the outcome. Intuitively, the instrument helps us isolate variation in the exposure which is "exogenous", i.e., variation which is not affected by omitted variable bias, measurement error, and reverse causality. The idea then is to use the exogenous variation in the exposure to try and identify the effect of the exposure on the outcome overall.

What is an instrument?

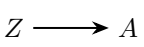
Suppose we are interested in identifying the effect of a well-defined treatment A on an outcome Y. A random variable Z is an instrument if it meets the following three criteria:

1) Z is associated with the treatment A. This is also known as the **relevance** criterion.
2) Z affects the outcome Y only through A. This is also known as the **exclusion restriction**.
3) There are no common causes of the $Z - Y$ relationship. This is known as the **independence** criterion.

Relevance criterion

In Econometrics textbooks, you will find the relevance criterion described as follows. For a given candidate instrument Z and an exposure A, Z satisfies the relevance criterion if $cov(z, a) \neq 0$ where $z \in Z$, $a \in A$, and $cov(.)$ represents the covariance operator. Note that covariance is a claim about linear dependence.

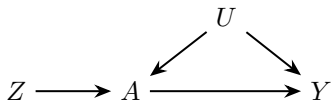
In Epidemiology and Computer Science (and, these days, also in Econometrics), you will find that the relevance criteria has been expanded to focus not just on non-linear relationships. We can, in fact, encode the relevance criterion in a DAG, the simplest form of which is:



Recall that a DAG does not make any functional form assumptions about the relationship between variables encoded in the graph.

Exclusion restriction

The exclusion restriction implies that the instrument Z is related with the outcome Y only through the exposure A. This can be easily encoded in a DAG, the simplest form of which is:



Note that there are no arrows from Z to Y other than through A.

Relatedly, we can also express the exclusion restriction using counterfactual notations. Suppose $Y_{(z,a)}$ encodes the value outcome Y takes when the instrument Z is set to z and the outcome A is set to a. Further, suppose that Z is a binary variable taking the values 1 and 0. The exclusion restriction can then be written as:

$$Y_{(z=1,a)} = Y_{(z=0,a)} = Y_a$$

In other words, conditional on fixing A to a, toggling the instrument from $z = 1$ to $z = 0$ has no effect on Y. Recall that this is basically saying that the Controlled Direct Effect of the instrument on the outcome is 0.

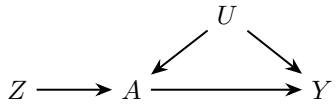
Independence criterion

The independence criterion implies that there are no open back-door paths from the instrument to the outcome. The DAG which we drew for the exclusion restriction demonstrates the independence criterion as well since there are no back-door paths from Z to Y.

We can express this in potential outcomes notation as well: $Z \perp\!\!\!\perp Y_a$ where Y_a refers to the potential outcome Y when A, possibly contrary to fact, is set to a.

Overall structure of the instrument-exposure-outcome relationship

Putting it all together, the simplest DAG which encodes an instrument is:



Identifying causal effects using instrumental variables

To identify the causal effect of A on Y using Z, we need to make an additional assumption about the nature of the treatment effect. Within the IV framework, we're going to think of A's effect on Y as being either **homogenous** or **heterogenous**.

Homogenous treatment effect means that the effect of A on Y is the same for *every single individual in the study population.* If we assume A is binary, then homogenous treatment effects is the assumption that $Y_{i(a=1)} - Y_{i(a=0)}$ is the same for all i's in the study population. In contrast, heterogenous treatment effect is an assumption that the effect of A on Y is different for different groups of individuals.

Identifying assumptions in a world where the treatment effect is homogenous

If we assume that the effect of A on Y is homogenous, then we can use Z to idenfity the treatment effect so long as Z is a valid instrument. In other words, the treatment effect is identified under the following assumptions:

1) **Relevance:** Z is associated with the treatment A.
2) **Exclusion restriction:** Z affects the outcome Y only through A for all individuals in the study population.
3) **Independence:** There are no common causes of the $Z - Y$ relationship.

If these identifying assumptions hold, then the instrument Z helps us identify the **Average Treatment Effect** (ATE) in the study population. This is because whatever effect you identify using IV is the effect for everyone in the population by the assumption of homogenous treatment effects.

Idenfitying assumptions in a world where the treatment effect is heterogenous

It's extremely unlikely that the effect of any exposure A on an outcome Y is the same for everyone in the population. Thus, if we assume that the treatment effect is heterogenous, then a candidate instrument Z can help us identify a treatment effect under the following assumptions:

1) **Relevance:** Z is associated with the treatment A.

2) **Exclusion restriction:** Z affects the outcome Y only through A for all individuals in the study population.

3) **Independence:** There are no common causes of the $Z - Y$ relationship.

4) **Monotonicity:** The instrument affects treatment status for *everyone* in the same direction.

Note, the first three assumptions are simply what we need for a candidate instrument Z to be a valid instrument. The additional assumption here is that of monotonicity. Consider a binary instrument Z and a binary exposure A. The assumption of monotonicity implies that $A_i^{z=1} \geq A_i^{z=0}$ for all i or $A_i^{z=1} \leq A_i^{z=0}$ for all i. In other words, in this binary instrument-exposure world, monotonicity implies that the instrument either weakly increases or weakly decreases the probability of being exposed for everyone in the study population. Whether or not monotonicity holds is never quite known, but we can make educated guesses about it in the context of whatever study we are using IVs for.

Assuming the four identifying assumptions hold, then in the heterogenous treatment effect world, IV analysis helps us identify the **Local Average Treatment Effect** (LATE). If you are interested in why IV identifies the LATE in this context, please see **Appendix 3** for the proof. The LATE is, as the name suggests, an average treatment effect, but is "local" in the sense that it applies to a very specific population sub-group. To understand which sub-group the LATE applies to, let us briefly discuss the notion of **principal strata**.

Principal strata

Within the heterogeneous treatment effects framework, we divide our population into four groups or principal strata. Assume we have a binary instrument Z and a binary treatment A. Assume further that we interpret $Z = 1$ as a form of inducement to take up treatment $A = 1$. Finally, let $A_{i(z)}$ represent the value of the treatment for individual i when Z is set to z, possibly contrary to fact.

We can then define these four principle strata as:

1) **Always takers**: Always takers are individuals who take the treatment regardless of the value of the instrument. In counterfactual notation, an individual i is an always taker if $A_{i(z=1)} = A_{i(z=0)} = 1$.

2) **Never takers**: Never takers are individuals who do not take the treatment regardless of the value of the instrument. In counterfactual notation, an individual i is a never taker if $A_{i(z=1)} = A_{i(z=0)} = 0$.

3) **Compliers**: Compliers are individuals who take the treament when induced to do so by the instrument and do not take the treatment when not induced to do so by the instrument. In counterfactual notation, an individual i is a complier if $A_{i(z=1)} = 1$ and $A_{i(z=0)} = 0$.

4) **Defiers**: Defiers are individuals who do not take the treatment when induced to take the treatment by the instrument and who take the treatment when induced to not take the treatment by the instrument. In counterfactual notation, an individual i is a defier if $A_{i(z=1)} = 0$ and $A_{i(z=0)} = 1$.

Since all four types of individuals are defined in terms of counterfactuals, we *generally* cannot identify which of these four categories an individual falls into given our data.

	$Z_i = 1$	$Z_i = 0$
$A_i = 1$	Complier/Always taker	Defier/Always taker
$A_i = 0$	Defier/Never taker	Complier/Never taker

What is the LATE?

The LATE is the effect of A on Y among the *compliers* in a study population. For a binary intrument Z and binary treatment A, we can express the LATE in counterfactual terms as:

$$LATE = E[Y_{i(a=1)} - Y_{i(a=0)} | A_{i(z=1)} = 1, A_{i(z=0)} = 0]$$

Why does IV isolate the treatment effect amongst the compliers only? Let's think back to the identifying assumptions:

1) The monotonicity assumption implies that there is no one in the study population who does not take up treatment when induced to do so by the instrument, i.e., the instrument inducing treatment take-up only (weakly) positively increases the likelihood of treatment. In other words, **monotonicity implies that there are no defiers in the population**.

   If there are no defiers, then the table above is reduced to:

	$Z_i = 1$	$Z_i = 0$
$A_i = 1$	Complier/Always taker	Always taker
$A_i = 0$	Never taker	Complier/Never taker

2) The exlcusion and independence assumptions ensure that the only way Z can affect Y is through A. This means, that IV analysis involves estimating the causal effect of A on Y by focusing on the variation in A which is affected by Z. By definition, the instrument does not affect always takers and never takers in terms of treatment uptake. Therefore, the only principal strata for which IV can identify the treatment effect are the compliers.

Better LATE than nothing?

While IV helps us identify the LATE, one question remains: how useful is the impact of the exposure on the outcome among compliers? There has been much debate about this in the Economics literature.

One reason why we might be concerned about LATE is that, in general, it does not equal the ATE. Note that in the principal strata framework and assuming no defiers, the ATE is a weighted average of the LATE, the treatment effect among always takers, and the treatment effect among never takers. It could be that the population of always takers and never takers are quite different from the population of compliers and perhaps what we are truly concerned about is the treatment effect in the overall population – LATE, in general, will not give us this effect.

However, assuming the four IV assumptions hold, the LATE is a treatment effect that has high internal validity. In addition, if we were to scale up the treatment, perhaps the only group we could ever induced to take up treatment are people like the compliers; in such a case, LATE might have some intrinsic value as well!

Estimating the causal effect using instrumental variables

Wald Estimator

This simplest form of the IV estimator is known as the Wald Estimator. Suppose we have a binary instrument Z and a binary treatment A. The canonical IV estimator, also known as the Wald estimator, is defined as:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[A_i|Z_i = 1] - E[A_i|Z_i = 0]}$$

The IV estimator scales the effect of the instrument on the outcome by the effect of the instrument on the treatment. The numerator of this expression is known as the **reduced form** while the denominator of this expression is known as the **first stage**.

Two Stage Least Squares

When both the instrument-exposure relationship and the exposure-outcome relationship is (or can be) written as linear functions, then we can get the IV estimate using an estimator called Two Stage Least Squares (2SLS).

Suppose for a valid instrument Z, exposure A, and outcome Y, we can write the following two equations:

$$a_i = \gamma_0 + \gamma_1 z_i + \epsilon_i \tag{1}$$

and

$$y_i = \beta_0 + \beta_1 a_i + \mu_i \tag{2}$$

Equation 1 is known as the first stage model while Equation 2 is known as the second stage model. In the models, ϵ and μ refer to the error terms. Assume that the zero-conditional mean holds in both cases. Further assume that $cov(\mu_i, z_i) = 0$, which is a classic Econometrics method of referring to the exclusion restriction and independence criterion simultaneously (although it is usually only presented as the exclusion restriction in classic texts).

We could use Equation 1 to substitute a_i in Equation 2 such that:

$$y_i = (\beta_0 + \beta_1\gamma_0) + (\beta_1\gamma_1)z_i + (\beta_1\epsilon_i + \mu_i)$$

2SLS uses this logic: it involves first estimating the first stage model, then predicting the values of a_i from this model, and substituting these predicted values into the second stage model to estimate the $\beta_1\gamma_1$ term. From here, we can isolate our estimate for β_1 by simply dividing our estimate for $\beta_1\gamma_1$ by our estimate for γ_1. Mathematically, what 2SLS does is the RHS of the following equation:

$$\hat{\beta_1} = \frac{\hat{\beta_1\gamma_1}}{\hat{\gamma_1}}$$

The 2SLS method can easily accommodate more than one instrument. Can you write out what the algebraic expression for $\hat{\beta_1}$ would look like with Z_1 and Z_2 as two instruments?

Standard error of the IV estimate

We won't derive this here, but it's worth knowing that

$$se(\hat{\beta_{IV}}) = \frac{\sigma_z\sigma_\mu}{\sigma_{z,a}} \tag{3}$$

In words, the standard error of the IV estimate is the ratio of the product of the standard deviations of the instrument and the error term of the second stage and the covariance of the instrument with the exposure.

Note that this is one reason why we usually don't want to try and manually fit the 2SLS model by estimating the first stage, predicting the values of the exposure, and then estimating the second stage using these predicted values. We'll usually get the standard errors on our estimate for β_1 from Equation 2 wrong because we wouldn't have asked our regression estimators to take into consideration the standard deviation of the instrument and the covariance of the instrument with the exposure.

IV analysis gone wrong

Weak instruments

Relevance implies that Z, the instrument, is associated with A, the exposure. However, with IV analysis, what we really want is for Z to be strongly associated with A. Suppose we estimate the first stage model:

$$a_i = \gamma_0 + \gamma_1 z_i + \epsilon_i$$

Z is considered to be a strong instrument if the F-statistic associated with γ_1 is greater than or equal to 10.[1]

Any instrument that's not strong is a **weak instrument**. A weak instrument is one which is not strongly correlated with the exposure. You can probably see why this is problematic immediately:

1) Recall that $se(\hat{\beta_{IV}}) = \frac{\sigma_z \sigma_\mu}{\sigma_{z,a}}$. A weak instrument implies that $\sigma_{z,a}$ is small, which means that the standard error of the IV estimate will be large.

2) Suppose the exclusion restriction or the independence criterion are violated. In terms of Equation 1 and Equation 2, this means that $cov(z_i, \mu_i) \neq 0$. In other words, Z in the second stage model is endogenous, which implies that any coefficient estimated for the $Z - Y$ relationship is biased and inconsistent. Note that the IV estimator involves estimating the second stage relationship using Z in place of Y (effectively estimating the $Z - Y$ relationship) and then scaling this by the first stage relationship. If our estimate of the $Z - Y$ relationship is biased, a weak instrument amplies this leading to biased and inconsistent IV estimates.

3) A more techincal issue which we will not get into in this lab is the idea that the 2SLS estimator is biased in small samples but consistent. The weaker the instrument, the more biased the 2SLS will be in small samples.

Recommended readings

1) Chapter 4: Instrumental Variables in Action: Sometimes You Get What You Need *in* Mostly Harmless Econometrics: An by Joshua D. Angrist and Jorn-Steffen Pischke (2009)

2) Hernan and Robins (2006). Instruments for Causal Inference: An Epidemiologist's Dream? *in* Epidemiology

3) Martens et. al. (2006). Instrumental Variables: Applications and Limitations *in* Epidemiology

4) Chapter 16: Instrumental variable estimation *in* Causal Inference: What If by Miguel Hernan and Jamie Robins (2020)

5) Andrews et. al. (2019). Weak Instruments in IV Regression: Theory and Practice *in* Annual Review of Economics

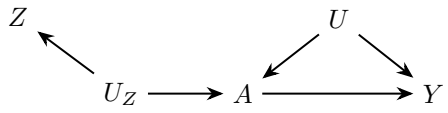
6) Huntington-Klein (.). Causal Inference Animated Plots. *Very nice animated ggplots explaining IV and other Econometric methods.*

[1] Depending on the number of instruments in your model, you can assess the F-statistic as follows:
1) If you have one instrument, you can compute the F-statistic by squaring the t-statistic associated with $\hat{\gamma}$ 2) If you have multiple instruments, you can determine the F-statistic by conducting a joint test of the null hypothesis that the coefficients on the instruments are zero

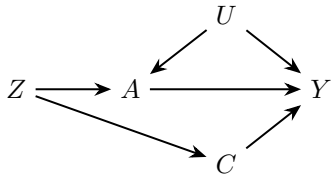
Exercise 1

Let Z be a candidate instrument, A the exposure, Y the outcome, C with any subscript a measured variable, and U with any subscript an unmeasured variable. In the following DAGs, determine if Z is a valid instrument and briefly explain why.

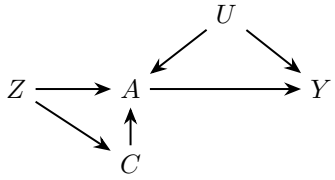
1) DAG 1: **Answer:** Yes, Z is a valid instrument. Z is associated with A through U_z. Z is associated with Y only through A. There are no open back-door paths from Z to Y.



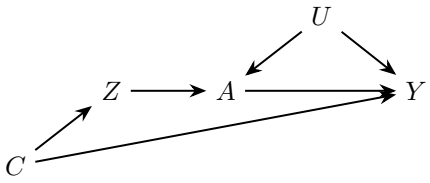
2) DAG 2: **Answer:** If we don't control for C, then Z is not a valid instrument. Z violates the exclusion restriction because $Z \rightarrow C \rightarrow Y$. By controlling for C, we can ensure that Z is a valid instrument. In practice, however, if there are common causes of both C and Y (which is likely), conditioning on C results in collider stratification and a violation of the exclusion restriction.



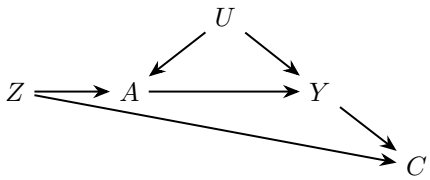
3) DAG 3: **Answer:** Yes, Z is a valid instrument. Z satisfied the relevance criterion. There is no direct path from Z to Y which does not go through A. There are no open back-door paths from Z to Y.



4) DAG 4: **Answer:** Yes, Z is a valid instrument as long as we control for C. This DAG shows that the independence criterion can be valid even if we have conditional independence of the $Z-Y$ relationship.



5) DAG 5: **Answer:** Yes, Z is a valid instrument as long as we do not control for C. If we control for C, we will violate the independence assumption.



Exercise 2

The table below lists several exposure-outcome relationships of interest and candidate instruments for each exposure-outcome relationship. In your breakout groups, discuss the following questions:

1) What are some endogeneity concerns associated with each exposure-outcome relationship?

2) Do you think the candidate instrument satisfies the relevance criterion?

3) Do you think the candidate instrument satisfies the exclusion restriction?

4) Do you think the candidate instrument is (conditionally) independent of the outcome?

Example: Suppose we are interested in understanding the impact of public school closures on test scores of students displaced by the closure. This is a key question of interest in the field of Education Economics, especially in light of the fact that public school closures are quite common in the United States (Brummet, 2014).

One possible confounding factor might be that schools which are closed are schools without many teaching resources, and being exposed to low levels of teaching resources in the past may affect future test scores of displaced students.

Suppose someone proposes number of cases of COVID-19 in a school district as a possible instrument to help estimate the causal effect of school closures on test scores.

It is likely that this instrument is relevant in that it is potentially correlated with school closures. However, the exclusion restriction is almost certainly violated: one example is that higher number of COVID-19 cases may induce greater anxiety. If we consider the effect of anxiety to be relatively persistent over time, then this may affect future test scores. The number of COVID-19 cases is also not randomly assigned by school district: we might imagine that in a district with lower rates of community testing, we will have higher numbers of COVID-19 cases, but lower rates of community testing might reflect some socio-economic factor which is also associated with future student test scores.

Research question	**Study setting**	**Candidate instrument**	**Link to study**
What is the impact of GDP growth in time t on the risk of civil war in a country in time $t+1$?	A study across 40 sub-Saharan African countries	Change in rainfall intensity between time $t-1$ and t	Miguel et. al., 2003
What is the impact of preventive healthcare utilization on women's health outcomes?	Texas and Wisconsin	Planned Parenthood clinic closures	Lu and Slusky, 2016
Do hospitals that provide high cost, emergency medical care achieve better health outcomes for patients?	New York City	In areas served by multiple ambulance companies, the company dispatched for any given emergency is based on a rotational assignment system	Doyle et. al., 2015

Research question	Study setting	Candidate instrument	Link to study
What is the impact of Medicaid expansion on emergency department use?	Oregon	Lottery to give household option to enroll in Medicaid (lottery implies randomization but households who are selected know that they have been selected to enroll)	Finkelstein et. al., 2012

Exercise 3

Individual	$A^{z=0}$	$A^{z=1}$	$Y^{a=0}$	$Y^{a=1}$	$Y^{z=0}$	$Y^{z=1}$
1	0	1	5	10	5	10
2	0	1	5	10	5	10
3	0	1	5	10	5	10
4	0	1	5	10	5	10
5	0	1	5	10	5	10
6	1	1	5	10	10	10
7	1	1	5	10	10	10
8	1	1	5	10	10	10
9	0	0	5	10	5	5
10	0	0	5	10	5	5
11	1	0	5	10	10	5
12	1	0	5	10	10	5
Totals	5	8	60	120	85	100

1. Fill out the last two columns in the table above. First check what the value of A^z would be when $Z = z$, and then look at the corresponding counterfactual outcome Y^a
2. Which colors represent the compliers, defiers, always takers, and ever takers

Red (individuals 1-5) are compliers; Green (6-8) are always takers; Orange (9-10) are never takers; and blue (11-12) are defiers.

3. Calculate the effect of the instrument on the outcome: Z on Y $(E[Y^{z=1}] - E[Y^{z=0}])$

$$\frac{100 - 85}{12} = 1.25$$

4. Calculate the effect of the treatment on the outcome: A on Y $(E[Y^{a=1}] - E[Y^{a=0}])$

$$\frac{120 - 60}{12} = 5$$

5. Calculate the effect of the instrument on the treatment: Z on A $(E[A^{z=1}] - E[A^{z=0}])$

$$\frac{8-5}{12} = 0.25$$

6. Calculate the treatment effect using the IV estimator: $\frac{E[Y^{z=1}]-E[Y^{z=0}]}{E[A^{z=1}]-E[A^{z=0}]}$

$$\frac{1.25}{0.25} = 5$$

So we see that the IV estimator recovers the average treatment effect using just the effect of Z on Y and Z on A. Notice that the table of counterfactual outcomes indicate homoegenous treatment effects.

Individual	$A^{z=0}$	$A^{z=1}$	$Y^{a=0}$	$Y^{a=1}$	$Y^{z=0}$	$Y^{z=1}$
1	0	1	5	10	5	10
2	0	1	5	15	5	15
3	0	1	5	15	5	15
4	0	1	5	10	5	10
5	0	1	5	10	5	10
6	1	1	5	15	15	15
7	1	1	5	15	15	15
8	1	1	5	15	15	15
9	0	0	5	10	5	5
10	0	0	5	10	5	5
11	1	0	5	10	10	5
12	1	0	5	10	10	5
Totals	5	8	60	145	100	125

7. Calculate the effect of the instrument on the outcome: Z on Y $(E[Y^{z=1}] - E[Y^{z=0}])$

$$\frac{125-100}{12} = 2.08333$$

8. Calculate the effect of the treatment on the outcome: A on Y $(E[Y^{a=1}] - E[Y^{a=0}])$

$$\frac{145-60}{12} = 7.08333$$

This is the average treatment effect (ATE) in the total population. According to IV assumptions this cannot be identified using the IV estimator because we have heterogeneity (see how each individual has different effect in response to change in A) on the outcome.

9. Calculate the effect of the instrument on the treatment: Z on A $(E[A^{z=1}] - E[A^{z=0}])$

$$\frac{8-5}{12} = 0.25$$

10. Calculate the treatment effect using the IV estimator: $\frac{E[Y^{z=1}]-E[Y^{z=0}]}{E[A^{z=1}]-E[A^{z=0}]}$

$$\frac{2.08333}{0.25} = 8.333$$

We can see this is not the ATE. According to IV assumptions, this effect is biased because we have defiers in our population.

Ignoring the last two individuals:

We are ignoring the defiers.

11. Calculate the effect of the instrument on the outcome: Z on Y $(E[Y^{z=1}] - E[Y^{z=0}])$

We are ignoring the defiers.

$$\frac{115 - 80}{10} = 3.5$$

12. Calculate the effect of the treatment on the outcome: A on Y $(E[Y^{a=1}] - E[Y^{a=0}])$

$$\frac{125 - 50}{10} = 7.5$$

This is the average treatment effect in the population if we are saying that the total population is just the first 10 individuals in the population (i.e. no defiers.)

13. Calculate the effect of the instrument on the treatment: Z on A $(E[A^{z=1}] - E[A^{z=0}])$

$$\frac{8 - 3}{10} = 0.5$$

14. Calculate the treatment effect using the IV estimator: $\frac{E[Y^{z=1}] - E[Y^{z=0}]}{E[A^{z=1}] - E[A^{z=0}]}$

$$\frac{3.5}{0.5} = 7$$

This is NOT the ATE even in the population with the defiers. This is what we would expect because the IV assumptions tell us that we cannot recover the ATE with treatment heterogeneity , we can only estimate the LATE among compliers. Let us see if the treatment effect among compliers only is equal to 7.

For the compliers ONLY:

15. Calculate the effect of the treatment on the outcome: A on Y $(E[Y^{a=1}] - E[Y^{a=0}])$

$$\frac{60 - 25}{5} = 7$$

So we can see that the IV estimator is the average treatment effect among the compliers, when there are no defiers, in the presence of heterogenous treatment effects. However, in the first table, with homogenous treatment effects, the IV estimator gives us the average treatment effects. This is true in spite of the fact that we have counterfactual data.

Appendix 1: Randomization solves the problem of selection bias

Suppose we are interested in estimating the effect of an exposure A on an outcome Y. Let A be defined as a binary variable which takes the value 1 if an individual i receives treatment and 0 otherwise. On the difference scale, we can estimate:

$$E[Y_i|A=1] - E[Y_i|A=0],$$

which we can further decompose as:

$$E[Y_i|A=1] - E[Y_i|A=0] = E[Y_i^1|A=1] - E[Y_i^0|A=0] = E[Y_i^1|A=1] - E[Y_i^0|A=1] + E[Y_i^0|A=1] - E[Y_i^0|A=0].$$

Where Y^a is the outcome Y when A is set to a, possibly contrary to fact. Note, we've used the assumption of Consistency in this decomposition in the second step. Also observe that in this decomposition, $E[Y_i^1|A=1] - E[Y_i^0|A=1]$ represents the Average Treatment Effect on the Treated (ATT) while $E[Y_i^0|A=1] - E[Y_i^0|A=0]$ represents selection bias, i.e., bias in the ATT as a result of individuals selecting into treatment. Recall from Lab 1 of PHS 2000B that selection bias is simply another term for confounding. It is called selection bias in Economics to represent the idea that people of different background characteristics select into the treatment group versus the control group. Selection bias is used differently in Epidemiology, where it refers to the idea of selecting into the study and <u>not</u> selecting into treatment.

Randomization ensures that $Y^a \perp\!\!\!\perp A$, which implies that we can manipulate the selection bias as follows:

$$E[Y_i^0|A=1] - E[Y_i^0|A=0] = E[Y_i^0|A=1] - E[Y_i^0|A=1] = 0.$$

Therefore, randomization gets rid of selection bias in expectation. Also note that randomization ensures that the ATT = ATE, where ATE means the Average Treatment Effect. Can you show why this is?

We can also see the magic of randomization using linear regression models, the workhorse model for all economists. Let's go back to our schooling and income example to see why this is the case. Suppose the "true" causal regression (i.e. the way the world works) is:

$$y_i = \alpha + \beta s_i + \sigma b_i + \epsilon_i$$

where y_i is individual i's income; s_i is the number of years individual i attended school, and b_i is some measure of their ability. Now imagine we are trying to estimate this equation, but we haven't collected any data on ability. If ability is associated with an individual's income and years of schooling, then in the following model

$$y_i = \alpha + \beta s_i + v_i$$

the exposure variable s_i is endogenous because $cov(s_i, v_i) \neq 0$. If we were to estimate β, we know that our estimate would be biased because:

$$\hat{\beta} = \frac{cov(y_i, s_i)}{var(s_i)} = \frac{cov(\alpha + \beta s_i + \sigma b_i + \epsilon_i, s_i)}{var(s_i)} = \beta + \sigma \frac{cov(s_i, b_i)}{var(s_i)}$$

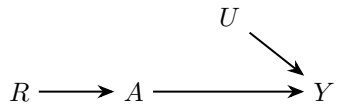
If schooling were randomized, however, note that $cov(s_i, b_i)$, the very term that is introducing bias into our estimate of β would be 0, and therefore $\hat{\beta} = \beta$.

Appendix 2: A brief note on randomized controlled trials

Ideal randomized controlled trials

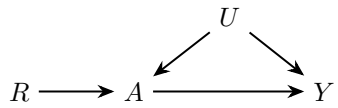
The ideal randomized controlled trial (RCT) is considered to be the "gold standard" of causal inference methodologies. An ideal RCT is one which does not suffer from loss to follow-up (at least not from differential loss to follow-up), full adherance to the assigned treatment throughout the study, a well-defined treatment, and double blind assignment. The ideal RCT is the gold standard for causal inference because it solves the *selection problem*. Please see **Appendix I** for a discussion on why this is the case.

The ideal RCT is very easy to encode as a DAG. Suppose A represents some well-defined treatment that is assigned by random assignment R, Y represents the outcome of interest, and U represents all other covariates which is related to the outcome. The DAG which represents this scenario is:



Randomized controlled trials with non-compliance

One deviation from the ideal RCT is a RCT with non-compliance. This means that treatment take-up or adherence by treatment arm is not 100%. A RCT with non-compliance could be represented by the DAG below, where R represents the randomization into treatment groups, A represents the treatment, Y represents the outcome. Note that treatment take-up may be affected by unmeasured factors U which are also associated with the outcome. This implies that given this structure of the relationship between A and Y, we cannot identify the effect of A on Y by simply estimating some regression model involving those two variables.



Note, however, that the randomization itself (R) is unconfounded which implies that we can identify the effect of being randomized into one treatment group versus the other on both treatment take-up A and the outcome Y. The idea behind an adherence-adjusted estimate is exactly this: to use these two unconfounded estimates to try and estimate the causal impact of A on Y. Alternatively, we could also estimate the impact of R on Y directly, which we call the Intention to Treat effect (ITT).

Appendix 3: The Local Average Treatment Effect Theorem

Recall that if the effect of a treatment A on an outcome Y differs for different groups of individuals, we say that the treatment effect is heterogenous. Further recall that when the treatment effect is heterogenous and we assume relevance, exclusion, independence, and monotonicity, the IV analysis helps us identify the Local Average Treatment Effect (LATE), i.e., the treatment effect among compliers. The following presents a proof for the binary instrument, binary treatment case.

Let Z be a binary instrument, A a binary treatment variable, and Y the outcome. The IV estimator can be written as:

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[A_i|Z_i = 1] - E[A_i|Z_i = 0]} \tag{4}$$

Let's focus on the numerator. Let $Y^{z,a}$ represent the outcome Y when Z is set to z and A is set to a, possibly contrary to fact. $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i^{z=1,a=0} + (Y_i^{z=1,a=1} - Y_i^{z=1,a=0})A_i^{z=1}] - E[Y_i^{z=0,a=0} + (Y_i^{z=0,a=1} - Y_i^{z=0,a=0})A_i^{z=0}]$ (by the exclusion restriction, independence, and the hidden but ever present assumption of consistency).

Since expectation is a linear operator, $E[Y_i^{z=1,a=0} + (Y_i^{z=1,a=1} - Y_i^{z=1,a=0})A_i^{z=1}] - E[Y_i^{z=0,a=0} + (Y_i^{z=0,a=1} - Y_i^{z=0,a=0})A_i^{z=0}] = E[Y_i^{z=1,a=0} + (Y_i^{z=1,a=1} - Y_i^{z=1,a=0})A_i^{z=1} - Y_i^{z=0,a=0} - (Y_i^{z=0,a=1} - Y_i^{z=0,a=0})A_i^{z=0}] = E[(Y_i^{z=1,a=1} - Y_i^{z=1,a=0})A_i^{z=1} - Y_i^{z=0,a=1} - Y_i^{z=0,a=0})A_i^{z=0}] = E[(Y_i^{z=1,a=1} - Y_i^{z=1,a=0})(A_i^{z=1} - A_i^{z=0})]$.

Monotonicity implies that $A_i^{z=1} - A_i^{z=0}$ equals either 1 or 0 for all individuals. This implies that $E[(Y_i^{z=1,a=1} - Y_i^{z=1,a=0})(A_i^{z=1} - A_i^{z=0})] = E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0}|A_i^{z=1} > A_i^{z=0}] * Pr(A_i^{z=1} > A_i^{z=0})$.

Thus,

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0}|A_i^{z=1} > A_i^{z=0}] * Pr(D_i^{z=1} > D_i^{z=0}) \tag{5}$$

Now, let's focus on the denominator. $E[A_i|Z_i = 1] - E[A_i|Z_i = 0] = E[A_i^{z=1}] - E[A_i^{z=0}]$ (by independence and the hidden but ever present assumption of consistency). Given the linearity of the expectation operator, $E[A_i^{z=1}] - E[A_i^{z=0}] = E[A_i^{z=1} - A_i^{z=0}]$. Finally, given monotonicity, $E[A_i^{z=1} - A_i^{z=0}] = E[A_i^{z=1} > A_i^{z=0}] = Pr(A_i^{z=1} > A_i^{z=0})$.

Thus,

$$E[A_i|Z_i = 1] - E[A_i|Z_i = 0] = Pr(A_i^{z=1} > A_i^{z=0}) \tag{6}$$

The relevance assumption tells us that $Pr(A_i^{z=1} > A_i^{z=0}) \neq 0$. With the additional assumptions of exclusion, independence, and monotonicity, we now have

$$\begin{aligned}\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[A_i|Z_i = 1] - E[A_i|Z_i = 0]} &= \frac{E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0}|A_i^{z=1} > A_i^{z=0}] * Pr(A_i^{z=1} > A_i^{z=0})}{Pr(A_i^{z=1} > A_i^{z=0})} \\ &= E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0}|A_i^{z=1} > A_i^{z=0}]\end{aligned}$$

That is, the IV estimator identifies the impact of the treatment A on the outcome Y among compliers in the study population.

Appendix 4: What if monotonicity is an unreasonable assumption for your research question of interest?

If the effect of a treatment A on Y is different for different groups of individuals but monotonicity is an unreasonable assumption, then it is possible that the IV estimator will not recover a treatment effect even when there is one.

Consider the scenario with a binary instrument Z, a binary treatment variable A, and an outcome Y. We saw in **Appendix 3** that the numerator of the IV estimator can be written as: $E[(Y_i^{z=1,a=1} - Y_i^{z=1,a=0})(A_i^{z=1} - A_i^{z=0})]$ where $Y^{z,a}$ represents the outcome Y when Z is set to z and A is set to a, possibly contrary to fact. Suppose we cannot assume monotonicity. In this scenario, we have both compliers and defiers in our population (we may have always takers and never takers too but they are not relevant in the IV analysis since they are never induced to take up the treatment by the instrument).

If we have both compliers and defiers, note that $(A_i^{z=1} - A_i^{z=0}) = 1$ for compliers and $(A_i^{z=1} - A_i^{z=0}) = -1$ for defiers. Therefore, $E[(Y_i^{z=1,a=1} - Y_i^{z=1,a=0})(A_i^{z=1} - A_i^{z=0})] = E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0} | A_i^{z=1} > A_i^{z=0}] * Pr(A_i^{z=1} > A_i^{z=0}) - E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0} | A_i^{z=1} < A_i^{z=0}] * Pr(A_i^{z=1} < A_i^{z=0})$. The fact that the "reduced form" of the IV estimator is calculated as a difference of two terms means that we could, theoretically, have a situation where there exists a treatment effect of A on Y on average, but we do not recover it from the IV estimator because the effect on the complier is cancelled out by the effect on the defiers.

A natural question to ask is why monotonicity does not matter if we assume treatment effects to be homogenous. The fact that we do not explicitly have to make the monotonicity assumption in this scenario means that we are allowing for defiers to exist in our study population. Thus, the reduced form in the homogenous treatment effects framework can also be written as: $E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0} | A_i^{z=1} > A_i^{z=0}] * Pr(A_i^{z=1} > A_i^{z=0}) - E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0} | A_i^{z=1} < A_i^{z=0}] * Pr(A_i^{z=1} < A_i^{z=0})$.

However, note that because treatment effects are homogenous, $E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0} | A_i^{z=1} > A_i^{z=0}] = E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0} | A_i^{z=1} < A_i^{z=0}]$. Let ρ represent the treatment effect. Thus, we can re-write $E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0} | A_i^{z=1} > A_i^{z=0}] * Pr(A_i^{z=1} > A_i^{z=0}) - E[Y_i^{z=1,a=1} - Y_i^{z=1,a=0} | A_i^{z=1} < A_i^{z=0}] * Pr(A_i^{z=1} < A_i^{z=0})$ as $\rho * (Pr(A_i^{z=1} > A_i^{z=0}) - Pr(A_i^{z=1} < A_i^{z=0})) = \rho * E[A_i^{z=1} - A_i^{z=0}]$. Since the first stage of the IV estimator is $E[A_i^{z=1} - A_i^{z=0}]$, note that we recover the homogenous treatment effect in this setting unless the first stage is 0.