I. Types of Treatment Estimates

i. Average Treatment Effect

As we have seen, a really nice thing about randomized controlled trials (RCTs) is that they solve MLR4, our biggest enemy in this class. Sample variation can still lead to some odd results, but this can be controlled for. Solving MLR4 means that we can be pretty confident that the estimate we find for the impact of our treatment (program, intervention, policy, etc.) on a given outcome is *causal*.

For a given treatment T where $T = 1$ for those who received the treatment and $T = 0$ for those who did not, we can estimate the causal effect of the treatment on an outcome Y by regressing

$$Y_i = \alpha + \beta_1 T_i + u_i$$

The coefficient β_1 on treatment is referred to as the **Average Treatment Effect (ATE)**. This tells you exactly what it sounds like: what is the difference in the outcome between the treatment and control groups, on average. Whenever we successfully randomize treatment, we can estimate the ATE, and we will almost always be interested in this estimate.

Note: Recall that randomization ensures balance between treatment and control *in expectation*. In any given sample, we might observe some spurious correlations between treatment and other characteristics. We will want to check for balance on variables we measure, and may choose to control for any characteristics differ significantly by treatment status.

ii. Intention to Treat

Sometimes we don't have perfect **compliance** among the treatment and control groups: some observations in the treatment group do not take the treatment; and some observations in the control group find a way to take the treatment. This decision to "not comply" with the research design is almost certainly correlated with other unobservable charactersitics. In this case, if we decide to compare those in the treatment group who took the treatment, to those in the control group who did not, we are no longer comparing the randomly assigned groups, and we won't get a causal estimate.

	Assigned to Treatment	Assigned to Control
Treated	T Complier	Always-Taker
Not Treated	Never-Taker	C Complier

Think of this in the context of a medical randomized control trial. I assign 50 people to control, and 50 people to treatment. The treatment group is given a pill that is supposed to help with energy levels. Among the 50 people in the treatment group, 40 people comply (*T Compliers*) and take the pill but 10 fail to comply and refuse to take the pill (*Never-Takers*). Among the 50 people in the control group, 10 people find a way to get the pill (*Always-Takers*) and 40 people comply and don't take pill (*C Compliers*).

If I were to compare the outcomes of the 40 treatment individuals who took the pill (like they were supposed to), to the 40 control individuals who did not take the pill (like they were supposed to) - then we can no longer say that we are observing the causal effect of the treatment. This is precisely

because certain people "selected" into these treatment and control groups based on unobservables we may or may not see, and hence can't always control for. In our case, maybe those who didn't comply in the treatment group by not taking the pill are generally people with greater energy levels. Similarly maybe those in the control group who took the pill are hypochondriacs and are always complaining of illnesses.

What do we do in this case? We acknowledge that we don't have perfect compliance, but we continue to "naively" compare those who were *assigned* to the treatment group, to those who were *assigned* to the control group. So if you are in the treatment group but didn't comply, you will still be considered treated, and if you are in the control group and didn't comply, you will still be considered control. In keeping with the initial allocation of treatment, we call the estimator "**Intention to Treat**" (ITT). We estimate the ITT using the same specification as the ATE - the only difference is that we acknowledge that our treatment assignment was not perfectly successful, so we no longer estimate the ATE.

We can always estimate the ITT with an RCT. If the treatment only affects the outcome through its effect in varying a characteristic of interest (taking a pill in the example above), the ITT tells us the effect of that characteristic on the outcome, diluted by the fact that not every treated observation complied with the treatment (e.g., took the pill).

Formally, in the example above with pills to increase energy levels, suppose we have a measure of individual energy levels (y) and we also know rates of take-up of the pill (P) by treatment status (T or C). Then we have

$$\bar{y}_T = P_T * E[y|P = 1, T = 1] + (1 - P_T) * E[y|P = 0, T = 1]$$
$$\bar{y}_C = P_C * E[y|P = 1, T = 0] + (1 - P_C) * E[y|P = 0, T = 0]$$

We are told $P_T = 0.8$ and $P_C = 0.2$, but we need to also consider the breakdown of compliers, never-takers, and always-takers. Since 20% of control individuals took the pill and treatment was randomized, we assume 20% of the population overall are always-takers. Similarly, since 20% of treatment individuals didn't take the pill, we assume 20% of the population are never-takers. Then we can write

$$\bar{y}_T = 0.2 * E[y|Always - Taker, T = 1] + 0.6 * E[y|Complier, T = 1] + 0.2 * E[y|Never - Taker, T = 1]$$
$$\bar{y}_C = 0.2 * E[y|Always - Taker, T = 0] + 0.6 * E[y|Complier, T = 0] + 0.2 * E[y|Never - Taker, T = 0]$$

Noting that randomization of treatment means expected outcomes should be the same for always-takers and never-takers (since these groups differ only in their treatment assignment but not in their treatment take-up), we end up with

$$\begin{aligned}\bar{y}_T - \bar{y}_C &= 0.6 * (E[y|Complier, T = 1] - E[y|Complier, T = 0]) \\ &= E[Complier] * (E[y_T|Complier] - E[y_C|Complier])\end{aligned}$$

The ITT estimate thus gives us the impact of the treatment among compliers, weighted by the share of compliers. Observe that if everyone complies with the treatment, this is the same as the ATE.

iii. Treatment on the Treated

When compliance with random assignment is not perfect, how close the estimate of the ITT is to the ATE will depend on a few things, most importantly what the take-up (share of treatment compliers) is (and any differences in treatment effects).

In the example above with pills to increase energy levels, we ended up with $\bar{y}_T - \bar{y}_C = E[Complier] * (E[y_T|Complier] - E[y_C|Complier])$. Then, if the only way the treatment impacts energy levels is through the likelihood of taking the pill, we can estimate

$$(\bar{y}_T - \bar{y}_C)/E[Complier] = E[y_T|Complier] - E[y_C|Complier] = TOT$$

This gives us the "**Treatment on the Treated**" (TOT) estimate, and tells us the average impact of the treatment among compliers. Note that the TOT will always be larger than the ITT, since it is not diluted by the null impact of the treatment on individuals assigned to treatment that did not actually get treated.

If the characteristics of compliers are very different from those of the full population, the TOT may look very different from the ATE. For example, if the treatment individuals who took the energy pill were those who thought they could benefit most from a boost in their energy levels, the TOT might be greater than the ATE. The only situation when the TOT will equal the ATE is if compliers experience the same treatment effects as the population would on average, i.e., if

$$E[y_T|Complier] - E[y_C|Complier] = E[y_T] - E[y_C]$$

So anytime we don't have perfect compliance with treatment we need to be concerned about heterogeneity in treatment effects and how these could create differences between the TOT and ATE.

iv. Encouragement Designs

Despite the fact that ITT estimates may not recover average treatment effects, they are still very useful. We cannot always successfully randomize some characteristic of interest to policy-makers. For example, if you wanted to ask what the effects were of education on wages, it would not be possible to randomly force some people to get more or less education. What can we do in this context? A common approach is to randomize the provision of *encouragement* with the goal of pushing (or pulling) the treatment group to vary (usually increase) their use or level of the variable whose impact we want to evaluate. If the encouragement is randomly assigned and only affects the outcome through its effect on the variable of interest, then this randomly-generated variation in the variable of interest allows us to recover an estimate of the causal effect of that variable on the outcome. More on this if we get to *Instrumental Variables* in the class.

For example, to elicit random variation in years of education, we might consider a program that pays parents if their children stay in school (e.g., Progresa in Mexico). This randomly *encourages* some households to increase their children's years of education, without varying years of education directly. To elicit random variation in fertilizer use, the study in Big Assignment 4 uses a program that provides vouchers for subsidized fertilizer. Random assignment of this program leads to random variation in fertilizer use, allowing us to estimate ITT impacts of fertilizer on maize yields.

II. Regression Discontinuity

i. Motivation

Sometimes randomizing treatment by a program or policy is impossible because the intervention has actually already happened, or because it's unethical or otherwise infeasible to withhold the treatment or intervention from a group in order to have a control.

If we haven't been able to do an RCT, and haven't been able to assign the treatment (or program, or policy, etc.) randomly ourselves, then we have to assume that the treatment was not randomly assigned: that it depends on either observable or unobservable characteristics of the people, firms, cities under consideration. In this case, there are important differences between our treated and untreated units that we cannot control for in a regression. Leaving these variables out in u_i will cause OVB, which totally foils our attempts at estimating the causal effect.

One class of methods for dealing with this is called **selection on observables**. This class assumes that if we can control for observed differences between groups that did and did not receive a particular treatment, this will allow us to recover the causal impact of the treament. We have talked about the most basic version of this method already in the context of Omitted Variables Bias. Successfully controlling for all variables in u_i that are associated with treatment will allow us to estimate unbiased treatment effects. Another approach involves "matching estimators," where we compare outcomes among individuals/firms/cities that vary in their treatment status but are otherwise similar in observable characteristics. This approach "matches" pairs of treated and control observations that look similar, and estimate the average difference in the outcome variable across these pairs. A major problem with these approaches is the one we already discussed: it is very hard to argue convincingly that you have measured and controlled for everything in u_i that could be associated with treatment.

Regression discontinuity (RD) is an example of an estimation strategy that attempts to sidestep the problems with non-random treatment assignment and with selection on observables. The RD approach is applied in situations where the treatment is assigned only to eligible units and eligibility is determined by a threshold value of (what's called) a running variable.

When eligibility to a treatment is determined by meeting a threshold value (such as a poverty line, an age limit, a geographical boundary, or a score in a standardized test), we can compare the outcome variable for observations just below and just above the threshold after the treatment has been implemented. The argument is that such thresholds mean that very similar units will get very different access to treatment.

For example, some of you may be all too aware of GPA requirements for majors at Cal. For example, UC Santa Cruz implemented a minimum GPA of 2.8 to declare a major in Economics. If we wanted to study the causal effect of majoring in economics on early-career earnings, it would be hard (and unethical) to force some students to major in econ and others not to. However, we might believe that students with a GPA of 2.79 are pretty much identical to students with a GPA of 2.81, except that the latter group was able to major in economics.[1] We would therefore consider GPA the "running variable" and 2.80 as the threshold for determining whether an individual is (eligibile for) treatment.

[1] This is a real paper (by a former Berkeley student) forthcoming in the *American Economic Journal: Applied Economics*.

ii. Causal Effect

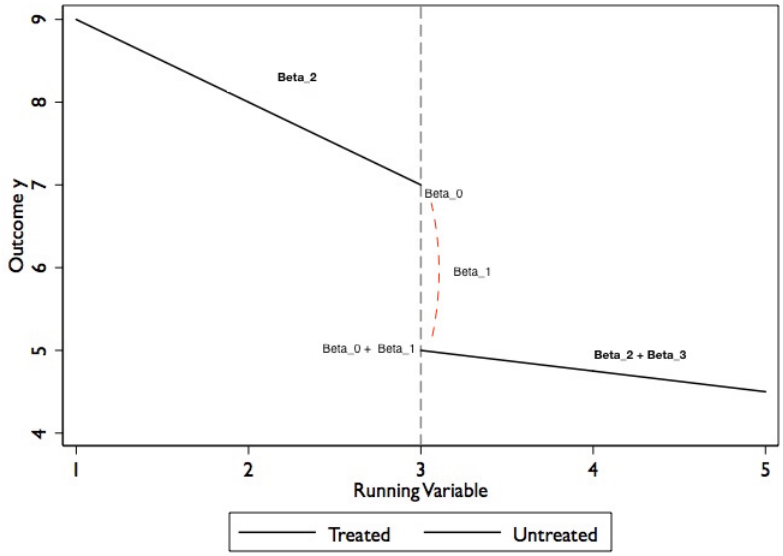
The expectation is that people just below and above the threshold for the running variable are identical in all observable and non-observable characteristics and conditions, except for program participation. In this case, we measure a **Local Average Treatment Effect (LATE)** that is valid around the threshold.

So if we only look at units just slightly above the threshold of the running variable and got the treatment, then a good comparison group are the units just slightly below the threshold and did not get the treatment. But the "Local" in LATE is us explicitly admitting is that our effect is only valid for people close to the threshold – the effects of an economics major for someone with a GPA close to 2.8 might be quite different for a student with a 3.8 GPA or a 1.8 GPA, for example.

In a regression framework, we estimate the following:

$$y_i = \beta_0 + \beta_1 T_i + \beta_2(RunningVar_i - threshold) + \beta_3 T_i \times (RunningVar_i - threshold) + u_i$$

Where $RunningVar_i$ is the running variable, $threshold$ is the threshold value of the running variable that determines treatment assignment, and T_i is a dummy variable that equals 1 if a unit has a value of the running variable that indicates it receives the treatment.



In this model, β_0 is the intercept (mean outcome value) for the control group at the threshold (since we have $T = 0$ and $RunningVar = threshold$). β_1 is the difference in the intercept (mean outcome value) for the treatment group relative to the control group at the threshold. We refer to this difference in intercepts at the thresholds as a **discontinuity**. β_2 is the slope of the relationship between the running variable and the outcome variable for the control group (since $T = 0$ implies $Runningvar < threshold$). β_3 is the difference in the slope of the relationship between the running variable and the outcome variable for the treatment group relative to the control group (since $T = 1$ implies $Runningvar > threshold$). We include β_3 in order to allow for different relationships between the running variable and the outcome at different levels of the running variable. This is important, because otherwise we might not get the correct intercepts.

iii. Key Assumption

The key assumption for the validity of the RD method is that the outcome would be a continuous function of the running variable (variable that determines eligibility) around the threshold, if it were not for the program.

In other words we need to assume that the treatment is the only reason why there would be a discontinuity, i.e. without the treatment, there would be no discontinuity in the outcomes across the threshold. When we compare units just above and below the threshold, we assume that the only systematic differences in their observable and unobservable characteristics is that the ones above (below) the threshold were treated (untreated).

We worry about this assumption being violated if, for instance, some households could manipulate the measurement of their income so that they would be below a threshold for eligibility to receive a transfer. Or in the econ major case, people could try and convince their professors to be more lenient in grading so that they can get just above the GPA threshold (this obviously will not work in EEP). In this case, the people who were treated because of their strategic actions are probably pretty different front those that weren't treated – people that took the effort to lobby for higher grades would be overrepresented in the treatment group and this willingness to negotiate might also lead people to bargain for higher salaries at work. In otherwords, it would still be an omitted variable.

iv. Tests for Validity of Assumption

A test of the validity of the approach is that there are no discontinuities around the threshold for relevant variables other than the treatment and the outcome variable.

To make ourselves feel better about this assumption, we look at the averages of observable characteristics of units just above and below the threshold and make sure they're similar (kind of like testing for balance in an RCT). In other words, we test for the *absence* of discontinuity for any other variable.

$$x_i = \beta_0 + \beta_1 T_i + \beta_2(RunningVar_i - threshold) + \beta_3 T_i \times (RunningVar_i - threshold) + u_i$$

Here we want to find a coefficient of zero for our estimated $\hat{\beta}_1$.

We also might want to test the theory of people manipulating their treatment status by checking for "bunching at the threshold". For example, we might expect the distribution of grades between 2.7 and 2.9 to be fairly smooth. However, if we saw surprisingly few students just below 2.8 and suprisingly many students just above 2.9, we might have greated reason to suspect that people are able to affect which side of the discontinutity they fall on.

v. Example

If a youth who is less than 18 years old commits a criminal offense, the case is sent to the more lenient juvenile court. However, if a youth commits an offense after his/her 18th birthday, the case is sent to the much harsher adult criminal court. You have a cross-sectional data set of youths of ages 16-20 in Florida in 2005. This data set includes each youth's birthday, gender, family income, and whether or not they had been arrested for committing a criminal offense in the past year. (*Solutions at the end.*)

1. How would you estimate the causal effect of harsher punishments on the probability of committing a criminal offense[2]? Be sure to write down the exact regression you would run and define each variable in your regression. State which coefficient will give you the estimated causal effect.
2. What key assumption do you need to make for your regression in part 1) to estimate the causal effect of harsher punishments on the probability of committing a crime?
3. What test can we conduct in support of our assumption?

vi. Sharp vs. Fuzzy RD

Just as with a randomized controlled trial, compliance with treatment assignments under a threshold eligibility rule may not be perfect. In the above example, compliance would be perfect if once a youth turns 18 they are sent to the adult instead of the juvenile court without exceptions. We call situations of perfect compliance "Sharp" RD. We can use our RD approach to estimate the LATE.

Just like in the example of energy use pills, though, compliance with treatment may be imperfect. For example, Professor Magruder led a study that built irrigation canals on hillsides in Rwanda, which gave access to irrigation to all farms below the canal but none of the farms above (due to gravity). Despite this, about 5% of plots above the canal used irrigation, while only about 23% of plots under the canal used it. This type of imperfect compliance with a threshold-based treatment is sometimes called a "Fuzzy" RD. We can check whether we do in fact have a discontinuity around the threshold for treatment by running the RD regression with treatment take-up (in this example, use of irrigation) as the outcome variable.

If we want to know the effect of irrigation on yields in this context, what can we do? Just as in the above examples about imperfect compliance for an energy pills treatment, we can estimate ITT and TOT effects. A fuzzy RD regression with imperfect compliance yields a local ITT instead of LATE, and the TOT would be the LATE among compliers (those who take up irrigation, in this example).

vii. What do We Need to Worry About With RD?

1. Is the effect really there? Should be demonstrable with a figure: show the running variable on the x axis and the outcome on the y axis, and look for a discontinuity around the threshold.

2. Does the cutoff matter? Should be verified. Run the RD specification with treatment takeup as the outcome variable.

3. Can the running variable be manipulated? Want to check that there aren't more observations just above the cutoff for treatment eligibility than just below. Look for "bunching".

4. Does anything else change sharply at the cutoff? Depends on the type of discontinuity/threshold we are using. For example, if we are using geographic discontuities like state boundaries or elevations, what else might be changing across those boundaries or elevations?

[2]Given that we observe arrests rather than actual offenses, we might want to assume that the likelihood of being caught and arrested for such an offense is not associated with age

5. Who are we identifying the effect for? We know that we identify a local ATE (LATE). This might not be the same as the effect globally. For example, estimates of the impact of the harshness of punishment on crime among youths around age 18 might not apply for older adults. Or, farmers near the canal in Professor Magruder's study may have different impacts of irrigation than farmers at the bottom of the hill.

Solutions

1. How would you estimate the causal effect of harsher punishments on the probability of committing a criminal offense[3]? Be sure to write down the exact regression you would run and define each variable in your regression. State which coefficient will give you the estimated causal effect.

$$arrest_i = \beta_0 + \beta_1 Over18_i + \beta_2(age_i - 6574) + \beta_3 Over18_i \times (age_i - 6574) + u_i$$

   Where $arrest_i$ is a dummy variable equal to 1 if youth i was arrested for a criminal offense, age_i is age in days, and $Over18_i$ is a dummy variable equal to 1 if youth i is 18 or older. Note that 6574 is the number of days in 18 years. We could equivalently have set this up using age in months or weeks as the running variable rather than age in days. Age in years would likely be too broad of a range. The coefficient $\hat{\beta}_1$ will give us the LATE estimate of the effect of harsher punishments on the probability of committing a criminal offense.

2. What key assumption do you need to make for your regression in part 1) to estimate the causal effect of harsher punishments on the probability of committing a crime?

   We have to assume that without the "treatment" of harsher punishments after age 18, the probability of committing an offense is a continuous function of age. That is, we have to assume that there are no other discontinuities in observable or unobservable characteristics around age 18, making the group just under the threshold on average similar to the group just above the threshold—except for the treatment. This makes those about to turn 18 a suitable counterfactual for those who just turned 18.

3. What test can we conduct in support of our assumption?

   A test of the validity of the approach is that there are no discontinuities around the threshold for relevant variables other than the treatment and the outcome variable. With the data we have, we could run the same regression as in part 1) for family income.

[3] Given that we observe arrests rather than actual offenses, we might want to assume that the likelihood of being caught and arrested for such an offense is not associated with age