

Lecture 19: Randomization and Different Estimators

Pierre Biscaye

Fall 2022

Agenda

- 1 Review: potential outcomes framework
- 2 Randomization and the Average Treatment Effect
- 3 Encouragement designs
- 4 Intention to Treat and Treatment on the Treated estimators

Review: Potential outcomes framework

- When we analyze some policy/intervention/treatment T , we want to estimate $E[y_i | T_i = 1] - E[y_i | T_i = 0]$
- Can think of 4 potential outcomes we could observe

$$E[Y_i^T | T_i = 1]$$

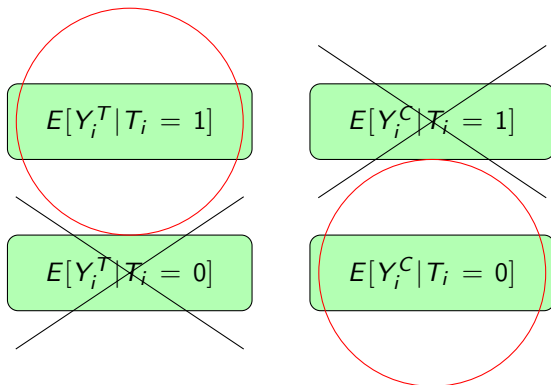
$$E[Y_i^C | T_i = 1]$$

$$E[Y_i^T | T_i = 0]$$

$$E[Y_i^C | T_i = 0]$$

Review: Potential outcomes framework

- But there's a big catch: we never observe the same person at the same time both receiving treatment and not.



What we observe and the *counterfactual*

- We can estimate $y_i = \beta_0 + \beta_1 T_i + u_i$
 - $\hat{\beta}_1$ estimates $E[Y_i^T | T_i = 1] - E[Y_i^C | T_i = 0]$
- We want to observe $E[Y_i^T | T_i = 1] - E[Y_i^T | T_i = 0]$ (or $E[Y_i^C | T_i = 1] - E[Y_i^C | T_i = 0]$)
 - But we don't observe the *counterfactuals*.
- When can we be confident that $E[Y_i^C | T_i = 0]$ is similar to the *counterfactual* $E[Y_i^T | T_i = 0]$?

What we observe and the *counterfactual*

- We can estimate $y_i = \beta_0 + \beta_1 T_i + u_i$
 - $\hat{\beta}_1$ estimates $E[Y_i^T | T_i = 1] - E[Y_i^C | T_i = 0]$
- We want to observe $E[Y_i^T | T_i = 1] - E[Y_i^T | T_i = 0]$ (or $E[Y_i^C | T_i = 1] - E[Y_i^C | T_i = 0]$)
 - But we don't observe the *counterfactuals*.
- When can we be confident that $E[Y_i^C | T_i = 0]$ is similar to the *counterfactual* $E[Y_i^T | T_i = 0]$?
 - Requires treatment and control group to be equivalent in expectation.
 - In other words, SLR4: $E[u | T] = 0$.

One condition: Randomization

- If treatment is *randomized*, and our sample is large enough:

$$E[Y_i^T | T_i = 1] = E[Y_i | T_i = 1] = E[Y_i^C | T_i = 1] \quad (1)$$

$$E[Y_i^T | T_i = 0] = E[Y_i | T_i = 0] = E[Y_i^C | T_i = 0] \quad (2)$$

- Randomization effectively ensures $E[u | T] = 0$.
- Thus, we observe the counterfactual, and we can estimate the effects of policy by

$$\bar{Y}_i^T - \bar{Y}_i^C \quad \text{or} \quad (3)$$

$$y_i = \beta_0 + \beta_1 T_i + u_i \quad (4)$$

- We call this the *Average Treatment Effect* (ATE).

Example: effect of job portal in India

- Research question: Does enrollment in a web-based job portal improve employment outcomes?
- Collected a sample frame of recent vocational institute graduates
- Randomly enrolled some of them in a web-based portal that shares job information (group T) and others to be enrolled and be sent a lot of job information (group TP), while the remainder were controls (group C : $T = 0$ and $TP = 0$).
 - Randomization by computer program.
 - Implies $E[Y_i^T | T_i = 0] = E[Y_i^C | T_i = 0]$, so we can compare mean outcomes for control and treatment groups to estimate portal impacts.
- Collected data three times:
 - Baseline before treatment assignment
 - Midline 6 months after treatment
 - Endline 12-18 months after treatment

Specifications with and without controls

Focus on employment outcomes

$$y_i = \beta_0 + \beta_1 Access_i + \beta_2 Priority_i + u_i \quad (5)$$

$$y_i = \beta_0 + \beta_1 Access_i + \beta_2 Priority_i + \gamma_1 x_{1i} + \dots + \gamma_k x_{ki} + u_i \quad (6)$$

- Why add controls?

Specifications with and without controls

Focus on employment outcomes

$$y_i = \beta_0 + \beta_1 Access_i + \beta_2 Priority_i + u_i \quad (5)$$

$$y_i = \beta_0 + \beta_1 Access_i + \beta_2 Priority_i + \gamma_1 x_{1i} + \dots + \gamma_k x_{ki} + u_i \quad (6)$$

- Why add controls?
- Rarely concerned about omitted variables.
 - After all, treatment assignment is randomized.
- But, controls can still reduce σ_u^2 : increase estimate precision.
 - Recall: $var(\hat{\beta}_j) = \frac{\sigma_u^2}{SST(1-R_j^2)}$

To Jupyter!

What could influence our estimates?

- We find that job portal access significantly *decreases* the probability of being employed at endline by 5 percentage points, ceteris paribus.
 - Holds with and without controls.
- Priority job portal access has no significant effect.
- Could there be omitted variables (violating MLR4)?

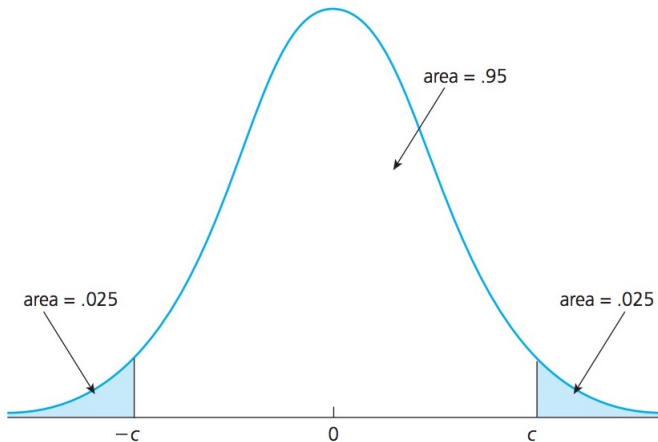
What could influence our estimates?

- We find that job portal access significantly *decreases* the probability of being employed at endline by 5 percentage points, ceteris paribus.
 - Holds with and without controls.
- Priority job portal access has no significant effect.
- Could there be omitted variables (violating MLR4)?
- Randomization guarantees that $E[u|T] = 0$.
- Even more than that, for any variable x (that cannot be affected by treatment), $E[x|T] = E[x]$.
- But what about in an *individual* sample?

What could influence our estimates?

- We find that job portal access significantly *decreases* the probability of being employed at endline by 5 percentage points, *ceteris paribus*.
 - Holds with and without controls.
- Priority job portal access has no significant effect.
- Could there be omitted variables (violating MLR4)?
- Randomization guarantees that $E[u|T] = 0$.
- Even more than that, for any variable x (that cannot be affected by treatment), $E[x|T] = E[x]$.
- But what about in an *individual* sample?
 - Randomization only guarantees independence in expectation.
 - In a given sample, *some* variables will be correlated with T by chance.

Some improbable values occur naturally



Frequentist significance

- Our standard of proof of a statistical relationship is that it is less than $\alpha\%$ likely that we would see a relationship due to chance.
- We can flip the $\alpha\%$ around: it is exactly $\alpha\%$ likely that *any* x variable will be correlated with T in our sample.
- If we test lots of variables, we will certainly find some:
 - Suppose we test 10 variables. The probability that none are correlated with T at the 5% level is $0.95^{10} = 0.6$.
 - So the probability that at least one is is $1 - 0.6 = 0.4$.
 - If we test 20 variables, the probability that at least one is correlated with T is $1 - 0.95^{20} = 0.65$
 - There are *lots* of variables that might be in u . *Some* be correlated with T in a given sample.
- We can use baseline data (from before treatment) to detect which x variables are associated with T just due to chance.
 - Important to only consider as controls variables that cannot be affected by T . Why?

Balance at baseline

Table 17: Balance

	(1) Control	(2) Treatment	(3) Priority Treat- ment	(4) (1) vs. (2), p-value	(5) (1) vs. (3), p-value	(6) (2) vs. (3), p-value	(7) Joint test
=1 if male	1.26	0.02	0.01	0.19	0.52	0.53	0.42
Age	26.31	-0.20	0.03	0.36	0.89	0.29	0.51
Married Y/N	1.46	-0.01	0.01	0.66	0.60	0.31	0.60
Religion=Hindu	0.10	0.02	0.03	0.07	0.05	0.82	0.10
Religion=Muslim	0.90	-0.02	-0.03	0.07	0.06	0.85	0.11
=1 if ST/SC caste	0.81	-0.02	-0.02	0.23	0.26	1.00	0.42
=1 if OBC caste	0.01	0.05	0.06	0.02	0.01	0.58	0.01
=1 if general caste	0.19	-0.03	-0.04	0.17	0.07	0.56	0.17
Father's education>0	0.04	0.03	0.02	0.23	0.47	0.66	0.48
Mother's education>0	1.05	0.03	-0.02	0.34	0.39	0.06	0.18
=1 if live in village	-0.17	-0.01	0.01	0.61	0.74	0.38	0.67
Received formal skills training	1.94	0.03	0.02	0.17	0.40	0.64	0.39
=1 if currently employed	0.87	0.03	0.02	0.16	0.35	0.69	0.37
=1 if looking for job	0.33	0.02	0.01	0.40	0.80	0.57	0.69
Access to Internet Y/N (clean)	0.67	0.04	0.02	0.03	0.28	0.33	0.10
Reservation wage (winsorized)	19376.31	-398.80	753.96	0.21	0.03	0.00	0.00

Standard errors are clustered at the respondent level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

What does this mean for MLR4?

- Randomization guaranteed that on average, there would be no x variables in u correlated with T .
 - $E[u|T] = 0$, and in particular, $E[x|T] = E[x]$ for x variables that cannot be affected by treatment.
- In practice, we know that in any one sample, *some* variables will be correlated with T .
- We've just shown this happened in this RCT.
 - What if these correlations influence $\widehat{\beta}_{OLS}$?
- One option: control for x variables that you know about that are correlated with T .
 - One reason collecting baseline data can still be valuable in an RCT.
- $y_i = \beta_0 + \beta_1 T_i + \beta_2 TP_i + \gamma_1 x_{1i} + \dots + \gamma_k x_{ki} + u_i$
 - In this case we would want to add religion dummies, access to internet, and reservation wage to the model along our existing controls.

Challenges with implementing RCTs

- Because of MLR4, randomization (through an RCT) is an increasingly popular tool in economic research.
- Requires a lot of preparation to randomize treatment.
- Ethical considerations are important.
 - Key principles: informed consent and minimal harm.
 - With the web portal: obtained informed consent to be enrolled in the portal, likely no harm of portal access.
 - In other cases: informed consent may be impossible, or treating (or not treating) people may expose to more than minimal harm.

Challenges with implementing RCTs

- Because of MLR4, randomization (through an RCT) is an increasingly popular tool in economic research.
- Requires a lot of preparation to randomize treatment.
- Ethical considerations are important.
 - Key principles: informed consent and minimal harm.
 - With the web portal: obtained informed consent to be enrolled in the portal, likely no harm of portal access.
 - In other cases: informed consent may be impossible, or treating (or not treating) people may expose to more than minimal harm.
- Logistical and practical considerations also key.
 - Randomization often impractical.
 - Governments would (likely) never implement a randomized minimum wage.
 - We could never randomly choose how much schooling people receive.
 - Can RCTs help us learn about the importance of the minimum wage or education?

How to learn about returns to education with an RCT

- We could never randomly assign education levels.
- But what if we don't need to make education completely random?
 - What if we can randomly encourage some people to get *a little more* education?
 - How?

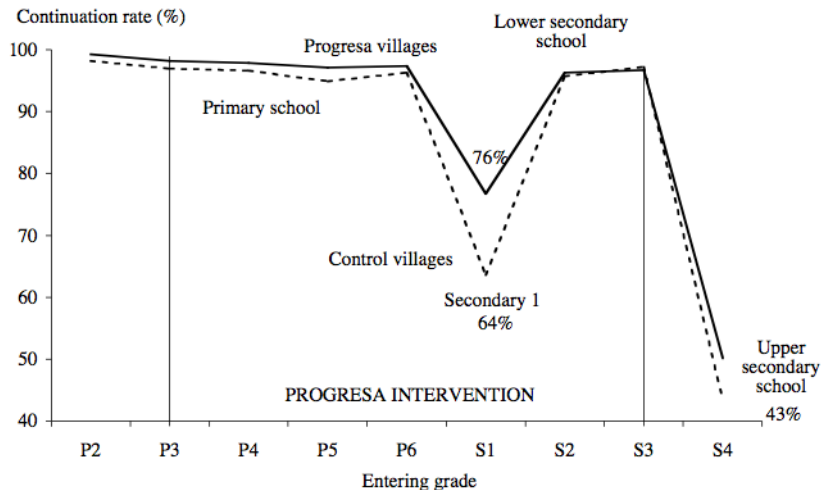
Encouragement designs

- Often we can randomly "encourage" people to change their behavior in an ethical, and logistically feasible way.
 - E.g., run promotional campaigns for health-seeking behaviors.
 - Offer scholarships for more schooling.
 - Health and education can't be randomized, but the encouragements can be.
- If the encouragement successful induces changes, this is random variation in treatment we can use to identify causal impacts.

PROGRESA program in Mexico

- PROGRESA (later Oportunidades; now PROSPERA) was a program in Mexico paying parents for sending kids to school.
 - A “Conditional Cash Transfer”: poor parents (based on an asset test) were paid a stipend if their children enrolled and attended school (and also met health benchmarks)
 - CCTs have become popular throughout developing world.
- PROGRESA was rolled out to randomly selected communities from 1997-2000.
 - For logistical reasons, could not enroll all communities at once.
 - Randomization allows a “fair” means of access to scarce (but popular) resources.
- Can look at data on wages (w_i) and Education (Ed_i) of adults in Mexico some years after PROGRESA was implemented.
- PROGRESA encouraged the poor to send their kids to additional schooling: random variation in Ed_i .

PROGRESA-induced variation in education



PROGRESA and the returns to secondary school

$$w_i = \beta_0 + \beta_1 \textit{Secondary}_i + u_i \quad (7)$$

- β_1 would be the ATE of completing secondary school. But what about MLR4?

PROGRESA and the returns to secondary school

$$w_i = \beta_0 + \beta_1 \text{Secondary}_i + u_i \quad (7)$$

- β_1 would be the ATE of completing secondary school. But what about MLR4?
- With or without PROGRESA, concerned about omitted variables: Secondary_i is not random.
- But PROGRESA is randomized. Can we use that to retrieve a causal estimate related to β_1 ?

$$w_i = \delta_0 + \delta_1 \text{Progresa}_i + u_i \quad (8)$$

$$\delta_1 = \bar{w}_P - \bar{w}_{NP} = E[w | \text{Progresa}] - E[w | \text{NoProgresa}] \quad (9)$$

- What is δ_1 ?

Intention to Treat (ITT) estimators

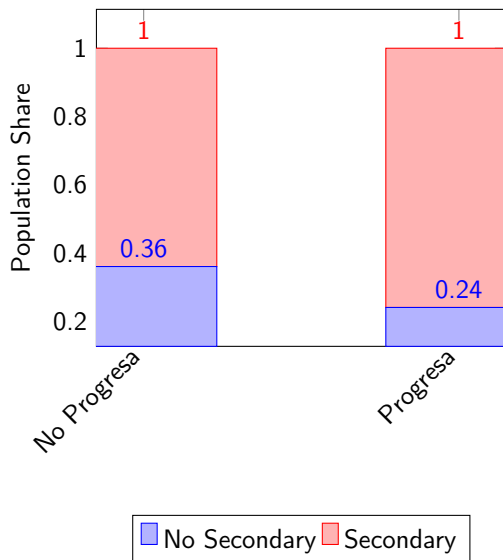
$$\delta_1 = \bar{w}_P - \bar{w}_{NP} = E[w|Progresa] - E[w|NoProgresa]$$

- We care about is the ATE of secondary school completion on wages. What we can do is get an unbiased estimate of δ_1 .
- Since PROGRESA is intended to increase education, we call this the *Intention to Treat* or ITT estimate.
- Also interpretable as the ATE of PROGRESA on wages.
 - With an RCT we can *always* calculate this.
 - Even if some people still do not complete secondary school with PROGRESA, as long as *some* people change their behavior in response to the encouragement, PROGRESA will have an impact.
 - Then, since PROGRESA is randomly assigned, can interpret the effect as causal.

What is the ITT?

- We might think that PROGRESA only impacted adult wages through extra education.
- $\bar{w}_P - \bar{w}_{NP}$ (the ITT) tells us the impact of PROGRESA on wages. It *does not* tell us the effect of secondary school on wages.
- If PROGRESA only impacts wages through the likelihood of completing secondary school, then the ITT tells us the effect of secondary school on wages, *diluted* by the fact that not everyone enrolled in PROGRESA gets more schooling.
- How close the ITT is to the ATE for secondary school on wages (β_1) will depend on the take-up rate (and any differences in treatment effects).

Group Means



ITT estimate

- We know that $\bar{w}_P - \bar{w}_{NP}$ is the ITT, but what exactly is that estimating?
- We can break the components down:

$$\bar{w}_P = 0.76(E[w|Sec = 1, P = 1]) + 0.24(E[w|Sec = 0, P = 1]) \quad (10)$$

$$\bar{w}_{NP} = 0.64(E[w|Sec = 1, P = 0]) + 0.36E[w|Sec = 0, P = 0]) \quad (11)$$

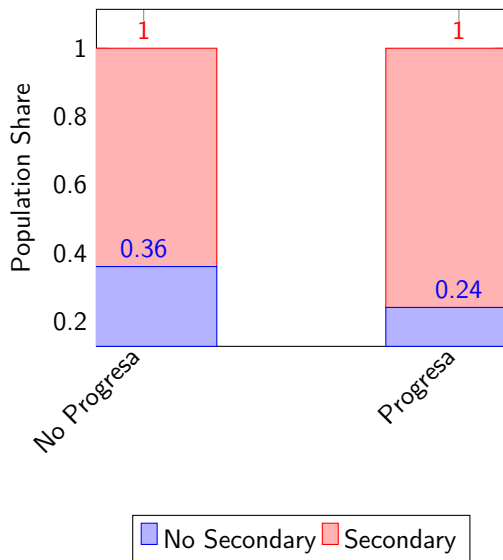
$$\bar{w}_P - \bar{w}_{NP} = ? \quad (12)$$

- Can we say anything more about these groups?

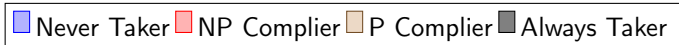
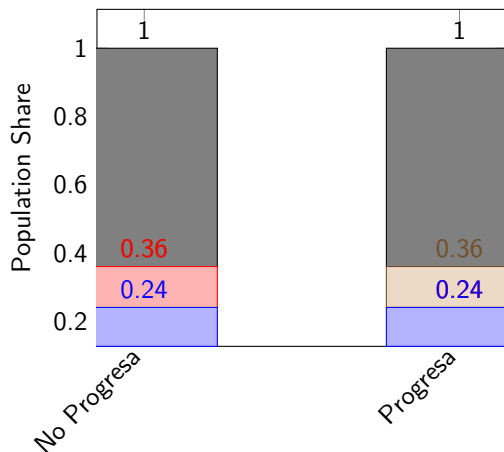
Possible responses under encouragement models

- Think about what these groups represent in a more general sense in the context of an encouragement treatment.
- *Always takers*: individuals that would get the treatment regardless of whether the encouragement is assigned to them.
- *Never takers*: individuals that would not get the treatment even if the encouragement is assigned to them.
- *Compliers*: individuals who only get the treatment if the encouragement is assigned to them.
- In this case, the treatment is completing secondary school and the encouragement is PROGRESA.

Group Means



Breaking down responses to encouragement



Apply these categories to ITT estimate $\bar{w}_P - \bar{w}_{NP}$

$$\bar{w}_P = 0.76(E[w|Sec = 1, P = 1]) + 0.24(E[w|Sec = 0, P = 1])$$

$$\begin{aligned}\bar{w}_P = & 0.64(E[w|AlwaysTaker, P = 1]) + 0.12(E[w|Complier, P = 1]) \\ & + 0.24(E[w|NeverTaker, P = 1])\end{aligned}$$

$$\bar{w}_{NP} = 0.64(E[w|Sec = 1, P = 0]) + 0.36E[w|Sec = 0, P = 0])$$

$$\begin{aligned}\bar{w}_{NP} = & 0.64(E[w|AlwaysTaker, P = 0]) + 0.12(E[w|Complier, P = 0]) \\ & + 0.24(E[w|NeverTaker, P = 0])\end{aligned}$$

- Since P is randomized, $E[w|AlwaysTaker, P = 1] = E[w|AlwaysTaker, P = 0] = E[w|AlwaysTaker]$
- $E[w|NeverTaker, P = 1] = E[w|NeverTaker, P = 0] = E[w|NeverTaker]$

So what does the ITT estimate?

$$\bar{w}_P = 0.64(E[w|AlwaysTaker, P = 1]) + 0.12(E[w|Complier, P = 1]) \\ + 0.24(E[w|NeverTaker, P = 1])$$

$$\bar{w}_{NP} = 0.64(E[w|AlwaysTaker, P = 0]) + 0.12(E[w|Complier, P = 0]) \\ + 0.24(E[w|NeverTaker, P = 0])$$

$$\bar{w}_P - \bar{w}_{NP} = 0.12 * (E[w_P|Complier] - E[w_{NP}|Complier])$$

$$\bar{w}_P - \bar{w}_{NP} = E[Complier] * (E[w_P|Complier] - E[w_{NP}|Complier])$$

- If Progresa only affects wages through schooling, then the ITT estimate tells you the return for completing secondary school for compliers, weighted by the share of compliers.

ITT and ATE

- We started from $w_i = \beta_0 + \beta_1 \text{Secondary}_i + u_i$
- $ATE = \hat{\beta}_1 = \bar{w}_S - \bar{w}_{NS}$ would be the ATE for the full population, if we could get an unbiased estimate.
- How does the ITT relate?
- If estimate $w_i = \delta_0 + \delta_1 \text{PROGRESA}_i + u_i$, get $ITT = \hat{\delta}_1 = \bar{w}_P - \bar{w}_{NP} = E[\text{Complier}] * (E[w_P | \text{Complier}] - E[w_{NP} | \text{Complier}])$.

ITT and ATE

- We started from $w_i = \beta_0 + \beta_1 \text{Secondary}_i + u_i$
- $ATE = \hat{\beta}_1 = \bar{w}_S - \bar{w}_{NS}$ would be the ATE for the full population, if we could get an unbiased estimate.
- How does the ITT relate?
- If estimate $w_i = \delta_0 + \delta_1 \text{PROGRESA}_i + u_i$, get $ITT = \hat{\delta}_1 = \bar{w}_P - \bar{w}_{NP} = E[\text{Complier}] * (E[w_P | \text{Complier}] - E[w_{NP} | \text{Complier}])$.
- $E[w_P | \text{Complier}] - E[w_{NP} | \text{Complier}]$ is similar to β_1 but estimated only with the sub-population of compliers.
 - Could think of estimating $w_i = \tilde{\beta}_0 + \tilde{\beta}_1 \text{Secondary}_i + u_i$ for this sub-population, but can't usually identify them.
 - If PROGRESA affects wages other than through secondary school completion, these will no longer be comparable.
- The ITT is also weighed by $E[\text{Complier}]$: it will be smaller than the ATE for the sub-population of compliers since not everyone complies.
- The higher the share of compliers, the closer the ITT is to the ATE.

Treatment on the Treated (ToT) estimates

- This intuition allows us to say something about the returns to completing secondary school.

$$\bar{w}_P - \bar{w}_{NP} = (\bar{S}eC_P - \bar{S}eC_{NP})(\bar{w}_P^C - \bar{w}_{NP}^C) \quad (13)$$

$$\frac{\bar{w}_P - \bar{w}_{NP}}{\bar{S}eC_P - \bar{S}eC_{NP}} = \frac{\bar{w}_P - \bar{w}_{NP}}{0.12} = \bar{w}_P^C - \bar{w}_{NP}^C \quad (14)$$

- $\bar{w}_P^C - \bar{w}_{NP}^C$ is the returns to completing secondary school for compliers.
- This is known as the *Treatment on the Treated* (ToT) estimator.
 - Only those for whom PROGRESA actually affected their schooling decisions (Compliers) are *treated*.
 - $ToT = \frac{ITT}{E[Complier]} \Rightarrow$ will be larger than the ITT.
- When will the ToT estimator be the same as the Average Treatment Effect (ATE)?

ToT and ATE

- *If*

- 1 The program only impacted outcomes through direct participation (here, completing secondary school).

- How else could PROGRESA have changed future wages?

- 2 Compliers experience the same treatment effects as people would on average

- $E[w_P | Complier] - E[w_{NP} | Complier] = E[w_P] - E[w_{NP}]$

- then ToT = ATE

- Are these conditions likely to hold in this context?

Randomization Review

- When we can randomize a characteristic (treatment) of interest:
 - We are guaranteed that MLR4 holds.
 - We can produce ATE estimates and interpret them as causal.
 - Treatment will still be correlated with some variables in our sample
⇒ we will still need to examine potential spurious correlations.
- When we can generate random variation in a characteristic of interest:
 - We can produce ITT and ToT estimates.
 - For these estimators, MLR4 holds - however, what they estimate is slightly different than the ATE.
 - Need to be concerned about treatment effect heterogeneity. How do compliers compare to the rest of the population?
 - Need to be concerned about unintended effects (spillovers). Does the encouragement have any other effects?

Quasi-random impact evaluations

- For lots of questions we are interested in, randomization will be impossible, even via encouragement design
- What can we do to get a handle on MLR 4 if we can't randomize?
- 4 broad classes of approaches:
 - 1 Controlling for observables (did in first half); includes matching estimators - will not cover in this course
 - 2 Regression Discontinuity Design: matching on eligibility for treatment
 - 3 Panel data techniques: control for broader set of potential omitted variables
 - 4 Instrumental variable techniques: use a third variable to isolate quasi-random variation in independent variable of interest
- Call these 'quasi-random' because we attempt to identify "as good as random" variation in the independent variable to generate a causal estimate.

Next time: RDD

- Start by looking at interventions/programs with rules
 - Many programs have a threshold level for treatment.
 - E.g., poverty-related programs: may need to have income below some level.
 - True in PROGRESA too: needed to have an asset index below some level to be eligible.
- The use of thresholds means that there will be some similar people who get very different access to programs.
- Regression Discontinuity Designs (RDD) take advantage of this