Experimental Methods: Lecture 4

Heterogeneous Treatment Effects

Raymond Duch

May 17, 2022

University of Oxford

Road Map to Lecture 4: Heterogeneity

- Non-interference re-visited
- Attrition
- RAND case study
- Combating Attrition
- MIPO

Constant Treatment Effects

- Recall the fundamental assumption about treatment effects
- What does "constant treatment effects" really mean?
- More importantly, is the average treatment effect the same for every single observation in the sample?
- Furthermore, we are often interested in the "generalizability" of experimental findings and their policy relevance
- Treatment effect heterogeneity is one way to address these issues

Ferraz and Finan 2008

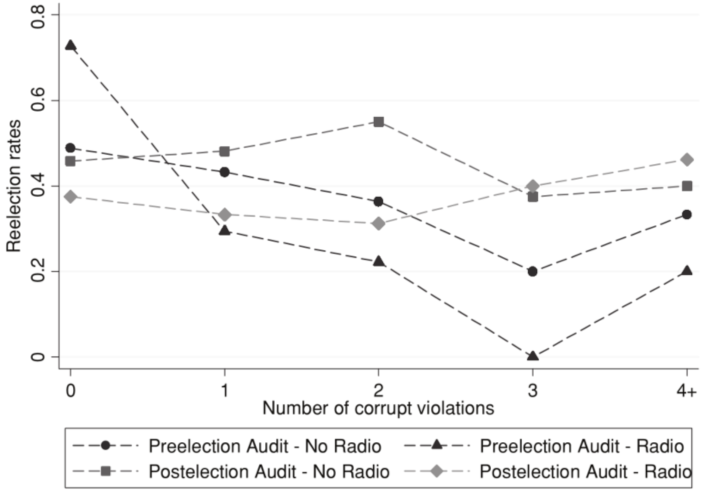

FIGURE IV
Relationship between Reelection Rates and Corruption Levels

Theory

We move away from constant treatment effects and therefore define

$$\tau_i \equiv Y_i(1) - Y_i(0) \tag{1}$$

The fundamental interest under treatment effect heterogeneity is in

$$\begin{aligned} Var(\tau_i) &= Var(Y_i(1) - Y_i(0)) \\ &= Var(Y_i(1)) + Var(Y_i(0)) + 2Cov(Y_i(1), Y_i(0)) \end{aligned} \tag{2}$$

Informally, we define treatment effect heterogeneity as *variance of the treatment effect* τ_i *across subjects*.

What is the problem with Eq. 2?

Theory

- This is an old and now for us very familiar problem:
- Any experiment does not allow us to estimate every component of $Var(\tau_i)$
- We have information about the marginal distributions of $Y_i(1)$ and $Y_i(0)$, but not about the joint distribution of these potential outcomes
- So what should we do?

Bounding $Var(\tau_i)$

- Recall that by randomization, $E[Y_i(0)|D_i = 1] = E[Y_i(0)|D_i = 0]$
- We can pair each observed $Y_i(1)$ with one of the observed $Y_i(0)$
- But which one? Many combinations possible
- We place bounds suggesting how large or small $Var(\tau_i)$ may be
- Pair values of $Y_i(0)$ and $Y_i(1)$ such that implied $Cov(Y_i(0), Y_i(1)$ is as large (upper bound) or as small (lower bound) as possible
- Sort values in ascending-ascending / ascending-descending order

Testing for heterogeneity

Suppose $H_0 : Var(\tau_i) = 0$ What if we compared $Var(Y_i(1))$ and $Var(Y_i(0))$?

Note that

$$\begin{aligned} Var(Y_i(1)) &= Var(Y_i(0) + \tau_i) \\ &= Var(Y_i(0)) + Var(\tau_i) + 2Cov(Y_i(0), \tau_i) \end{aligned} \tag{3}$$

Then, the Null of constant τ_i implies that

$$Var(\tau_i) = -2Cov(Y_i(0), \tau_i) = 0 \tag{4}$$

These two terms therefore cancel in Eq. 3 and we have shown that testing $H_0 : Var(\tau_i) = 0$ is the same as testing $Var(Y_i(1)) = Var(Y_i(0))$

Observed Outcome Local Budget

We can test this with randomization inference

	Budget share if village head is male	Budget share if village head is female
Village 1	?	15
Village 2	15	?
Village 3	20	?
Village 4	20	?
Village 5	10	?
Village 6	15	?
Village 7	?	30
Mean	16	22.5
Variance	17.5	112.5

Variance in control:

$\frac{1}{7-2-1}2(15-16)^2+2(20-16)^2+(10-16)^2=17.5$

Variance in treatment: $\frac{1}{2-1}(15-22.5)^2+(30-22.5)^2=112.5$

Interaction

- These approaches test *whether* τ_i varies
- But we want to know more: conditions under which τ_i varies
- We are interested in a different estimand: Conditional Average Treatment Effect (CATE) = ATE for a defined subset of subjects $\tau_i(x) = E[Y_i(1) - Y_i(0)|X_i = x]$ (individual), and, if distribution of X_i is known, $E[\tau_i(X_i)]$ is identified (average)
- Change in treatment effect that occurs from one subgroups to the next is the difference between 2 CATEs
- These subgroups can either be defined by covariate values (*treatment-by-covariate interactions*) or by design (*treatment-by-treatment interactions*)

Treatment-by-covariate interactions

- What is the H_0 here?
- We can test the difference in CATEs with randomization inference or in a regression framework

$$Y_i = a + bI_i + cP_i + dI_iP_i + u_i \tag{5}$$

When $P_i = 0$, the CATE is b:

$$Y_i = a + bI_i + u_i \tag{6}$$

When $P_i = 1$, the CATE is $b + d$:

$$Y_i = a + bI_i + c + dI_i + u_i = (a + c) + (b + d)I_i + u_i \tag{7}$$

where d yields the change in CATEs that occurs when P_i changes

Treatment-by-covariate interactions

- An alternative is to conduct an F test using randomization inference
- Compares sum of squared residuals from from the two nested models (alternative model is Eq. 5 and null model is $Y_i = a + bI_i + u_i$)
- If there are interaction affects, Eq. 5 should reduce SSR
- Simulate random assignments and calculate fraction of F-statistics at least as large as the observed F-statistic
- H_0 is that 2 CATEs are the same

Caveats

- Multiple comparisons problem:
  - With 20 covariates, the probability of finding at least 1 that significantly interacts with the treatment at $\alpha = 0.05$ is $1 - (1 - 0.05)^{20} = 0.642$
  - Bonferroni correction (divide target p-value by number of hypothesis tests h)
  - Pre-register your design! (lab)
- Subgroup analysis is non-experimental: groups that are not formed by random assignment, but pre-assignment
- Teacher incentives and teacher education

Treatment-by-treatment interactions

- Manipulate treatment *and* contextual factor / personal characteristic (e.g. COVID and community infection levels)
- Define a factorial experiment as an experiment involving factors 1 and 2, with factor 1 conditions being A and B, and factor 2 conditions being C and D and E
- Then, allocate subjects at random to every possible combination of experimental conditions
- $\{AC, AD, AE, BC, BD, BE\}$

Gottlieb et al. 2018: EGAP Metaketa II: Taxation

Jessica Gottlieb, Adrienne LeBas, Nonso Obikili: “Formalization, Tax Appeals, and Social Intermediaries in Lagos, Nigeria”

T1. Control condition, not encouraged

T2. Encouraged, but not receiving a follow-up visit

T3. Encouraged, and receiving one of the following four follow-up visit combinations:

- T3a. Public goods message from state representative
- T3b. Enforcement message from state representative
- T3c. Public goods message from marketplace representative
- T3d. Enforcement message from marketplace representative

Figure 2: Research Design and Assignment Probabilities

				Message Type	
				Public Goods	Enforcement
Control	Formalization Intervention only	Delivery Type	State Rep.	T3a: 5/36	T3b: 5/36
T1: 1/6	T2: 5/18		Market Association	T3c: 5/36	T3d: 5/36

Multiple treatment arms

From Rosen 2010

	Colin		Jose	
	Good grammar	Bad grammar	Good grammar	bad grammar
% Received reply	52	29	37	34
(N)	(100)	(100)	(100)	(100)

This design requires us to be especially careful with defining the causal estimand – what quantity are we interested in in this application?

Multiple treatment arms

Quiz: Why would these two models estimate the same quantities from the Rosen 2010 experiment?

$\{NG, HG, NB, HB\}$ are indicator variables for each of the 4 treatment groups

$J_i = 1$ if Jose Ramirez; $G_i = 1$ if good grammar

$$Y_i = b_1 CG + b_2 JG + b_3 CB + b_4 JB + u_i$$
$$Y_i = a + bJ_i + cG_i + d(J_i G_i) + u_i$$

What quantity in the table do each of the coefficients represent?