w241: Experiments and Causality

Heterogeneous Treatment Effects

David Reiley, David Broockman, D. Alex Hughes
UC Berkeley, School of Information
Updated: 2021-06-23

Introduction to Heterogeneous Treatment Effects

Main Topics

Heterogeneous treatment effects

- Does the same same treatment have different effects on different subjects?
- Use of regression to measure HTEs: Interaction between covariates of interest and treatment variable

Data analysis with non-experimental control

- Multiple comparisons problem: Fishing expeditions can cause overstatement of true statistical significance.

Reading: *Field Experiments*

Please read

- Section 9.0 (the Introduction), and,
- Section 9.1

Motivating Examples of Heterogeneous Treatment Effects

Quiz Question Discussion

- Potential outcomes are hypothetical population parameters.
- In real-world samples, we get to measure only
  - Treatment outcomes for the treatment group
  - Control outcomes for the control group
- We can measure only treatment outcomes or control outcomes for a single person.

Reading

Reading: *Field Experiments*, Section 9.3.1

While reading, consider, "Do different groups have different treatment effects?"

Example: Electricity Consumption

Test Groups

1. Below-average; and,
2. Above-average electricity consumers

Treatment

- Social comparison
- How much electricity are your neighbors using?

HTE

- Are people who use a *lot* of energy more responsive than people who aren't using much energy?
- Are people who are using less than their neighbors going to start using more?

Example: Congressional Responsiveness

Test groups

1. Members of Congress from the North; and,
2. Members of Contress from the South

Treatment

- Being informed of a meeting with a donor versus a constituent

HTE

- Responsiveness of different congressmen

Example: eBay Shipping Policies

Test groups

- Buyers of high-priced items
- Buyers of low-priced items

Treatment

- Charging (or not charging) a shipping price on an auction Some measure of value should be predetermined before the experiment

HTE

- Responsiveness to shipping costs on high-value versus low-value items

Example: eBay Seller Reputation

Test Groups

- Buyers of high-priced items
- Buyers of low-priced items

Treatment

- Selling from a high- or low-reputation account

HTE

- Are people differently reactive to reputation when they are buying high-priced, rather than low-priced items?

Example: Donation Matching

Test Groups

- People who live in *Blue States*
- Peolpe who live in *Red States*

Treatment

- Informing individuals that a donation they make to the ACLU will be matched

HTE

- Because the ACLU is perceived to be a "Liberal" non-profit organization, do individuals who live in areas that are more liberal react more strongly to the donor-matching treatment?

Treatment-by-Covariate Interactions

Quiz Review

Answer

- Students whose parents' literacy was above the median

Discussion

- Treatment effect not statistically significantly different between students whose parents have above-median versus below-median literacy
- We need to know standard errors in order to evaluate how much to believe a point estimate

Estimating HTEs

Etimating with two samples

- Split data into two separate samples
- Compute means and standard errors of the treatment effect in each group
  - Do a two-sample test of means.
  - **Group one**: Students whose parents have above-median literacy
  - **Group two**: Students whose parents have below-median literacy

Estimating HTEs (cont'd)

Estimating with regression

- Estimate a regression with dummy variables
  - I indicator for teacher incentive treatment
  - P indicator for parents with above-average literacy
- Include an interaction between $I \times P$

$$Y_i = \beta_0 + \beta_1 I_i + \beta_2 P_i + \beta_3 (I_i \times P_i) + \epsilon_i$$

Inferences

- Test whether there is evidence that the interaction term is different from zero
  - H_0: The treatment effects are no different between the groups
  - H_A: The treatment effects are different between the groups

Reading Clarification

- Always present standard errors with point estimates.
- Always show the number of observations.
- Reports the coefficient on one row, and the standard error in parentheses on the row below
- Includes a reporting of:
  - The number of observations
  - The R^2 and *F-Test* vs. an intercept-only model

Standard Format of Results

```
                    ##
## ========================================
##                  Dependent variable:
##              ---------------------------
##                         Score
## ----------------------------------------
## Treatment               1.06***
##                         (0.03)
##
## Constant                0.80***
##                         (0.20)
##
## ----------------------------------------
## Observations             100
## R2                       0.91
## Adjusted R2              0.91
## F Statistic    991.19*** (df = 1; 98)
## ========================================
## Note:        *p<0.1; **p<0.05; ***p<0.01
```

Reading Clarification (cont'd)

$$\widehat{Y}_i = \beta_0 + \beta_1 I_i + \beta_2 P_i + \beta_3 (I_i \times P_i)$$

- If a test for β_3 rejects the null hypothesis, then treatment effects differ along with levels of the covariate
- Possible to estimate this with regression or randomization inference
  - Gerber and Green use randomization inference,
  - With data that is well-behaved, we can rely on the Central Limit Theorem and use regression

Teacher Incentives: Evidence from India

Example: Teacher Incentives

Muralidharan and Sundararaman (2011)

We present results from a randomized evaluation of a teacher performance pay program implemented across a large representative sample of government-run rural primary schools in the Indian state of Andhra Pradesh. At the end of 2 years of the program, students in incentive schools performed significantly better than those in control schools by 0.27 and 0.17 standard deviations in math and language tests, respectively. We find no evidence of any adverse consequences of the program. The program was highly cost effective, and incentive schools performed significantly better than other randomly chosen schools that received additional schooling inputs of a similar value.

School Enrollments

	Log School Enrollment (1)	School Proximity (8–24) (2)	School Infrastructure (0–6) (3)	Household Affluence (0–7) (4)	Parental Literacy (0–4) (5)	Scheduled Caste/Tribe (6)	Male (7)	Normalized Baseline Score (8)
	Two-Year Effect							
Incentive	−.198	−.019	.28**	.09	.224***	.226***	.233***	.219***
	(.354)	(.199)	(.130)	(.073)	(.054)	(.049)	(.049)	(.047)
Covariate	−.065	−.005	.025	.017	.068***	−.066	.029	.448***
	(.058)	(.010)	(.038)	(.014)	(.015)	(.042)	(.027)	(.024)
Interaction	.083	.018	−.02	.038**	−.003	−.013	−.02	.006
	(.074)	(.014)	(.040)	(.019)	(.019)	(.056)	(.034)	(.031)
Observations	29,760	29,760	29,760	25,231	25,226	29,760	25,881	29,760
R^2	.244	.244	.243	.272	.273	.244	.266	.243
	One-Year Effect							
Incentive	−.36	−.076	.032	.004	.166***	.164***	.157***	.149***
	(.381)	(.161)	(.110)	(.060)	(.047)	(.045)	(.044)	(.042)
Covariate	−.128**	−.016*	−.001	.017	.08***	.007	.016	.502***
	(.061)	(.008)	(.025)	(.013)	(.012)	(.035)	(.020)	(.021)
Interaction	.103	.017	.041	.042**	−.013	−.06	.002	.000
	(.081)	(.011)	(.031)	(.017)	(.016)	(.048)	(.025)	(.026)
Observations	42,145	41,131	41,131	38,545	38,525	42,145	39,540	42,145
R^2	.31	.32	.32	.34	.34	.31	.33	.31

Data Structure: Teacher Gender

Test Score	Male Teacher	Incentive	Interaction
10	1	1	1
12	1	0	0
8	0	1	0
14	0	0	0

Results: Teacher Gender

TABLE 6
HETEROGENOUS TREATMENT EFFECTS
A. HOUSEHOLD AND SCHOOL CHARACTERISTICS

	Log School Enrollment (1)	School Proximity (8–24) (2)	School Infrastructure (0–6) (3)	Household Affluence (0–7) (4)	Parental Literacy (0–4) (5)	Scheduled Caste/Tribe (6)	Male (7)	Normalized Baseline Score (8)
	Two-Year Effect							
Incentive	−.198	−.019	.28**	.09	.224***	.226***	.233***	.219***
	(.354)	(.199)	(.130)	(.073)	(.054)	(.049)	(.049)	(.047)
Covariate	−.065	−.005	.025	.017	.068***	−.066	.029	.448***
	(.058)	(.010)	(.038)	(.014)	(.015)	(.042)	(.027)	(.024)
Interaction	.083	.018	−.02	.038**	−.003	−.013	−.02	.006
	(.074)	(.014)	(.040)	(.019)	(.019)	(.056)	(.034)	(.031)
Observations	29,760	29,760	29,760	25,231	25,226	29,760	25,881	29,760
R^2	.244	.244	.243	.272	.273	.244	.266	.243

Estimating Equation

$$
\begin{aligned}
T_{ijkm}(Y_n) = \alpha &+ \gamma T_{ijkm}(Y_0) + \delta_1 Incentives_i \\
&+ \delta_2 Characteristic_i \\
&+ \delta_3 (Incentives_i \times Characteristic_i) \\
&+ \beta Z_m + \epsilon_k + \epsilon_{jk} + \epsilon_{ijk}
\end{aligned}
$$

- T is a test score taken in year Y
- Z_m are covariates that are used to increase efficiency of the estimate
- Subscripts are:
  - i indexes the student, j the grade, and k the school, and m mandal
  - n year of observation

Regressors of interest

- Incentives
- Characteristics

Reading Regression Results

- Each column represents a separate regression estimate
  - **Dependent Variable**: Students' test score after treatment, normalized by the standard deviation across students

TABLE 6
HETEROGENOUS TREATMENT EFFECTS
A. HOUSEHOLD AND SCHOOL CHARACTERISTICS

	Log School Enrollment (1)	School Proximity (8–24) (2)	School Infrastructure (0–6) (3)	Household Affluence (0–7) (4)	Parental Literacy (0–4) (5)	Scheduled Caste/Tribe (6)	Male (7)	Normalized Baseline Score (8)
	Two-Year Effect							
Incentive	−.198	−.019	.28**	.09	.224***	.226***	.233***	.219***
	(.354)	(.199)	(.130)	(.073)	(.054)	(.049)	(.049)	(.047)
Covariate	−.065	−.005	.025	.017	.068***	−.066	.029	.448***
	(.058)	(.010)	(.038)	(.014)	(.015)	(.042)	(.027)	(.024)
Interaction	.083	.018	−.02	.038**	−.003	−.013	−.02	.006
	(.074)	(.014)	(.040)	(.019)	(.019)	(.056)	(.034)	(.031)
Observations	29,760	29,760	29,760	25,231	25,226	29,760	25,881	29,760
R^2	.244	.244	.243	.272	.273	.244	.266	.243

- If a male student started at the 50th percentile, the treatment would have increased him to the 59th percentile.
- No compelling evidence that treatment works differently for male or female identifying students - no HTE

HTEs by Student Characteristics

Quiz Review

- Asterisks help reader pick out statistically significant effects
- Most columns had statistically significant treatment effects…
- … but no statistically significant HTEs

Household Wealth Interaction Effect

- Only the household affluence interaction term was statistically significant.
- Household affluence variable is difficult to interpret.
  - Interaction coefficient = `0.038 **`
  - Treatment effect increases by 0.038 for each additional point of household-affluence score.
  - *Household-affluence score* ranges from 0–7
  - Assumes all categories have equal treatment-effect benefits -- $\mathbf{1 \rightarrow 2 = 6 \rightarrow 7}$

Alternative Specification

Rather than a linear scale, allow multiple dimensions of wealth

- Seven separate dummy variables as covariates.
  - One for owning land
  - One for owning a house
  - One for having running water
  - One for owning a TV
- And then, seven different interaction terms.

One Regression to Rule Them All?

Several specific regressions or one big regression?

- Muralidharan and Sundararaman (2011) Used one column for each covariate
  - Each covariate could have been put into the same regression.
  - But, might be difficult to identify *each* effect, since covariates are not randomly-assigned
  - As a result, many of the covariates might covary with others, and so lead to "unstable" estimates
  - This instability is because one measure is picking up many, related effects

Colinearity isn't a problem with treatment variables

- Because the experiment has randomly assigned treatment, it should be independent from all sets of covariates

Teacher Incentives: Household Affluence

TABLE 6
HETEROGENOUS TREATMENT EFFECTS
A. HOUSEHOLD AND SCHOOL CHARACTERISTICS

	Log School Enrollment (1)	School Proximity (8–24) (2)	School Infrastructure (0–6) (3)	Household Affluence (0–7) (4)	Parental Literacy (0–4) (5)	Scheduled Caste/Tribe (6)	Male (7)	Normalized Baseline Score (8)
	Two-Year Effect							
Incentive	−.198	−.019	.28**	.09	.224***	.226***	.233***	.219***
	(.354)	(.199)	(.130)	(.073)	(.054)	(.049)	(.049)	(.047)
Covariate	−.065	−.005	.025	.017	.068***	−.066	.029	.448***
	(.058)	(.010)	(.038)	(.014)	(.015)	(.042)	(.027)	(.024)
Interaction	.083	.018	−.02	.038**	−.003	−.013	−.02	.006
	(.074)	(.014)	(.040)	(.019)	(.019)	(.056)	(.034)	(.031)
Observations	29,760	29,760	29,760	25,231	25,226	29,760	25,881	29,760
R^2	.244	.244	.243	.272	.273	.244	.266	.243

- Adding household affluence leads the reported effect on `Incentive` to shrink
  - Estimated overall ATE = 3.5 (average affluence) $\times 0.038 = 0.133$
  - (Note, this would only be the case if there was a uniform distribution, or symmetric distribution across the affluence scale)
  - Incentive + estimated overall ATE $0.09 + 0.133 = 0.223$
- In an interaction model, the effect of treatment depends on levels for interacted variables

Teacher Incentives: Parental Literacy

TABLE 6
HETEROGENOUS TREATMENT EFFECTS
A. HOUSEHOLD AND SCHOOL CHARACTERISTICS

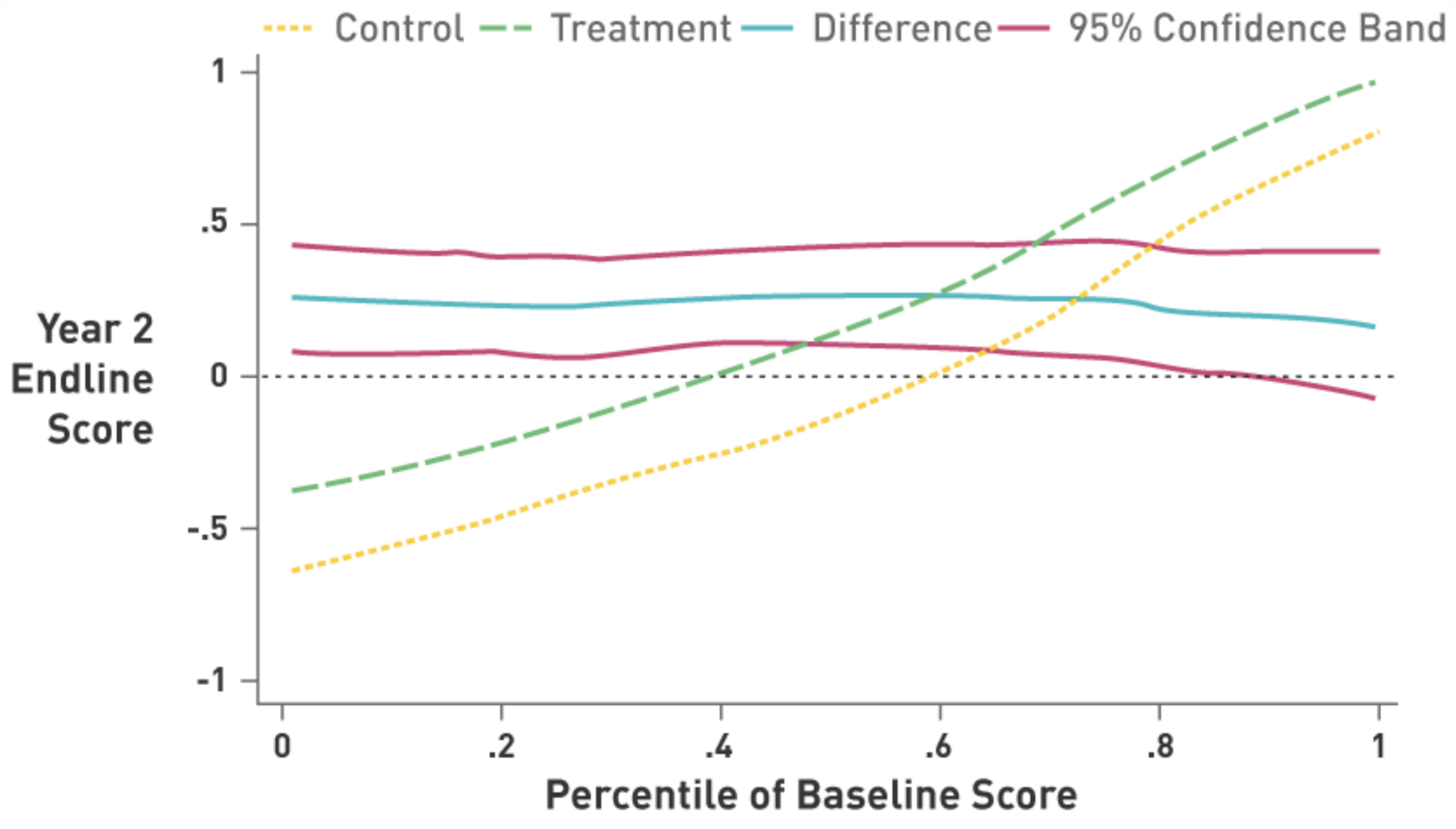
	Log School Enrollment (1)	School Proximity (8–24) (2)	School Infrastructure (0–6) (3)	Household Affluence (0–7) (4)	Parental Literacy (0–4) (5)	Scheduled Caste/Tribe (6)	Male (7)	Normalized Baseline Score (8)
				Two-Year Effect				
Incentive	−.198	−.019	.28**	.09	.224***	.226***	.233***	.219***
	(.354)	(.199)	(.130)	(.073)	(.054)	(.049)	(.049)	(.047)
Covariate	−.065	−.005	.025	.017	.068***	−.066	.029	.448***
	(.058)	(.010)	(.038)	(.014)	(.015)	(.042)	(.027)	(.024)
Interaction	.083	.018	−.02	.038**	−.003	−.013	−.02	.006
	(.074)	(.014)	(.040)	(.019)	(.019)	(.056)	(.034)	(.031)
Observations	29,760	29,760	29,760	25,231	25,226	29,760	25,881	29,760
R^2	.244	.244	.243	.272	.273	.244	.266	.243

- Always be aware of the amount of statistical uncertainty in estimates
  - Confidence interval: -0.003 ± 0.038 per unit of literacy
  - The largest absolute value in that interval is $-0.003 - 0.038 = -0.041$.
- Even using the coefficient estimate with the largest magnitude -0.041 and evaluating it at the largest parental literacy value, 4, the estimated total interaction effect confidence interval is $-0.041 \times 4 = -0.16$
- This is still smaller than the baseline treatment effect, 0.224 for those with totally parents marked as a zero on literacy

Always look at the magnitude of the estimate!

Always look at size of the confidence interval!

Broad based Treatment Effect



HTEs by Teacher Characteristics

Review: Student Characteristic HTE

- Broad-based benefits from teacher incentive program were found.
  - No HTEs by past student test score
  - Household wealth was the only significant HTE interaction.

Teacher Characteristics

Review the table. Answer the quiz questions.