Session IV
Sources of Heterogeneity

Thomas J. Leeper

Government Department
London School of Economics and Political Science

1 Other Survey Experimental Designs

2 Attention and Satisficing

3 Moderators and Effect Heterogeneity
- Manipulate the Moderator
- Blocking/Block Randomization
- Post-hoc Approaches

Beyond One-shot Designs

- Surveys can be used as a measurement instrument for a field treatment or a manipulation applied in a different survey panel wave
    1. Measure effect duration in two-wave panel
    2. Solicit pre-treatment outcome measures in a two-wave panel
    3. Measure effects of field treatment in post-test only design
    4. Randomly encourage field treatment in pre-test and measure effects in post-test

Beyond One-shot Designs

- Surveys can be used as a measurement instrument for a field treatment or a manipulation applied in a different survey panel wave
  1. Measure effect duration in two-wave panel
  2. Solicit pre-treatment outcome measures in a two-wave panel
  3. Measure effects of field treatment in post-test only design
  4. Randomly encourage field treatment in pre-test and measure effects in post-test
- Problems? Compliance & nonresponse

I. Effect Duration

- Use a two- (or more-) wave panel to measure duration of effects
    - T1: Treatment and outcome measurement
    - T2+: Outcome measurement
- Two main concerns
    - Attrition
    - Panel conditioning

II. Within-Subjects Designs

- Estimate treatment effects as a difference-in-differences
- Instead of using the post-treatment mean-difference in Y to estimate the causal effect, use the difference in pre-post differences for the two groups:

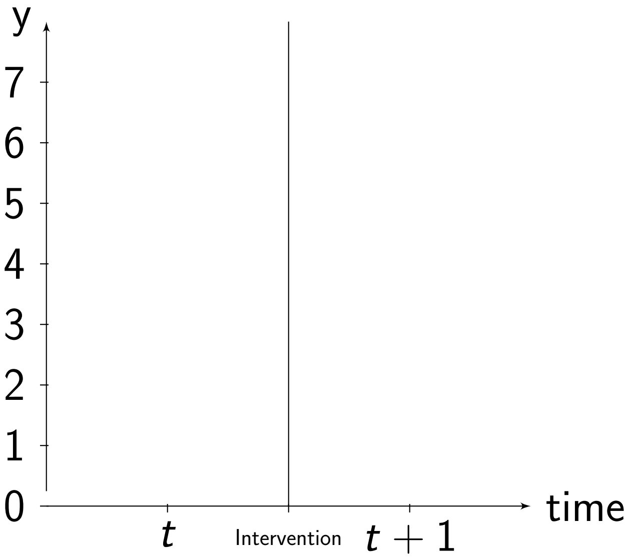
$$(\hat{Y}_{0,t+1} - \hat{Y}_{0,t}) - (\hat{Y}_{j,t+1} - \hat{Y}_{j,t})$$

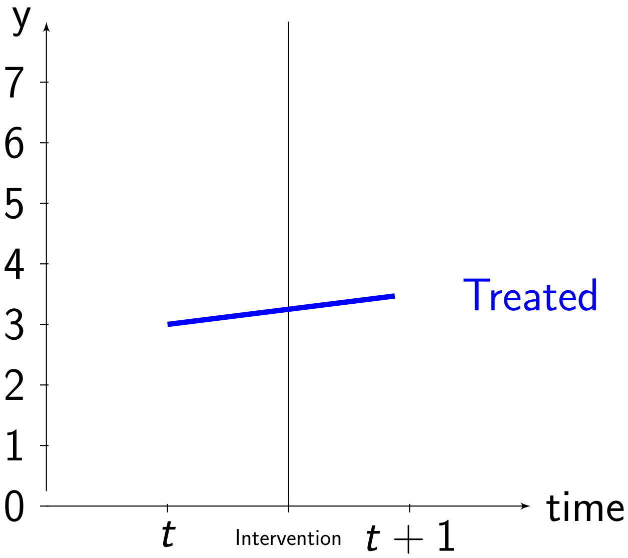
II. Within-Subjects Designs

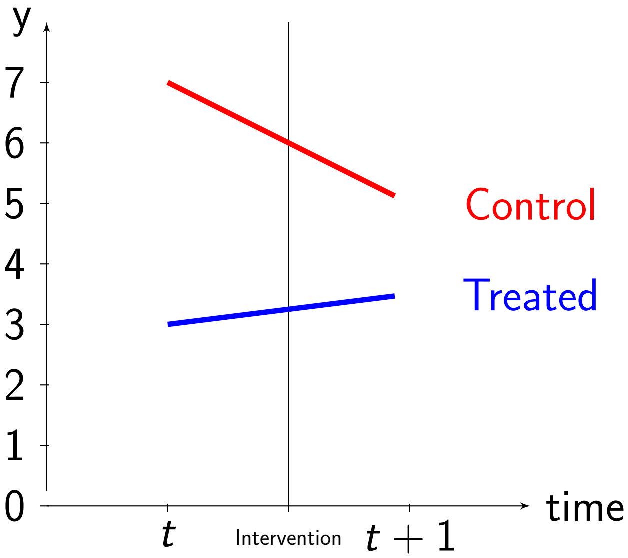
- Estimate treatment effects as a difference-in-differences
- Instead of using the post-treatment mean-difference in Y to estimate the causal effect, use the difference in pre-post differences for the two groups:

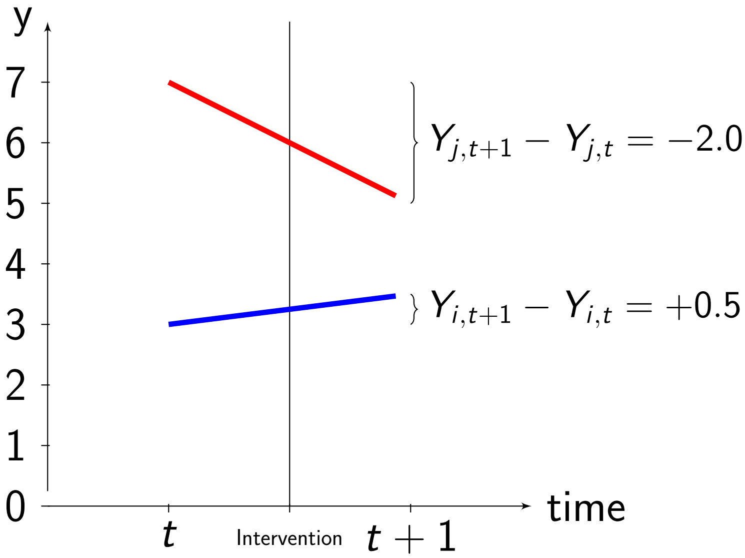
$$(\hat{Y}_{0,t+1} - \hat{Y}_{0,t}) - (\hat{Y}_{j,t+1} - \hat{Y}_{j,t})$$

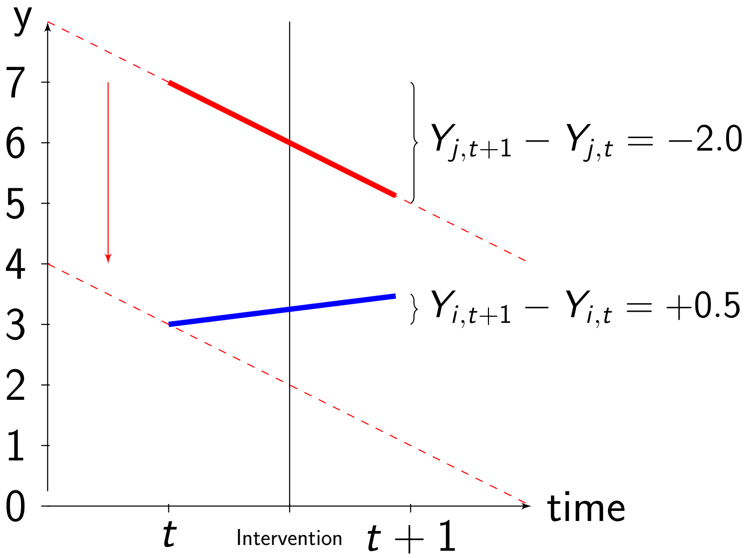
- Advantageous because variance for paired samples decreases as correlation between t_0 and t_1 observations increases

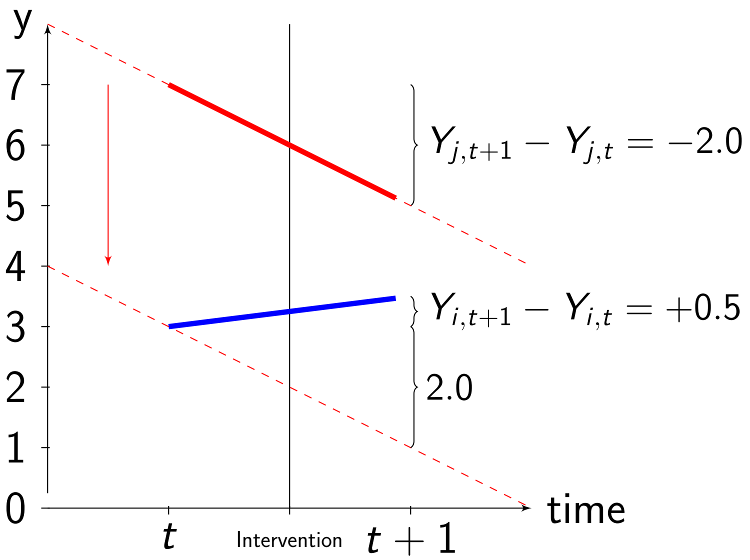

y
7
6
5
4
3
2
1
0
time
t
Intervention
t + 1

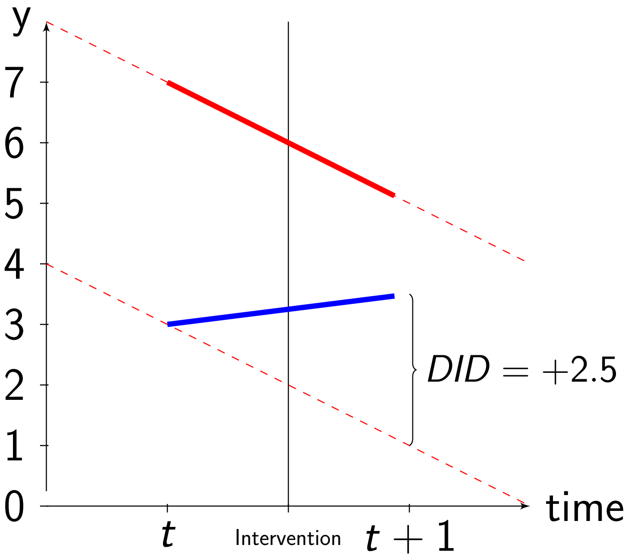

y
7
6
5
4
3
2
1
0
Treated
time
t
Intervention
$t+1$


y
7
6
5
4
3
2
1
0
time
t
Intervention
t + 1
Control
Treated


y
7
6
5
4
3
2
1
0
time
t
Intervention
$t+1$
$Y_{j,t+1} - Y_{j,t} = -2.0$
$Y_{i,t+1} - Y_{i,t} = +0.5$


y
7
6
5
4
3
2
1
0
t
Intervention
$t+1$
time
$Y_{j,t+1} - Y_{j,t} = -2.0$
$Y_{i,t+1} - Y_{i,t} = +0.5$


y
7
6
5
4
3
2
1
0
$Y_{j,t+1} - Y_{j,t} = -2.0$
$Y_{i,t+1} - Y_{i,t} = +0.5$
2.0
time
t
Intervention
$t+1$


y
7
6
5
4
3
2
1
0
time
t
Intervention
$t+1$
$DID = +2.5$

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[1]

[1] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[1]

1. History (simultaneous cause)

[1] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[1]

1. History (simultaneous cause)
2. Maturation (time trends)

[1] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[1]

1. History (simultaneous cause)
2. Maturation (time trends)
3. Testing (observation changes respondents)

[1] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[1]

1. History (simultaneous cause)
2. Maturation (time trends)
3. Testing (observation changes respondents)
4. Instrumentation (changing operationalization)

[1] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[1]

1. History (simultaneous cause)
2. Maturation (time trends)
3. Testing (observation changes respondents)
4. Instrumentation (changing operationalization)
5. Instability (measurement error)

[1] Shadish, Cook, and Campbell (2002)

Threats to Validity

As soon as time comes into play, we have to worry about threats to validity.[1]

1. History (simultaneous cause)
2. Maturation (time trends)
3. Testing (observation changes respondents)
4. Instrumentation (changing operationalization)
5. Instability (measurement error)
6. Attrition

[1] Shadish, Cook, and Campbell (2002)

III. Randomized Field Treatment

- Examples:

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage
  2. Different local media markets randomly assigned to receive different advertising

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage
  2. Different local media markets randomly assigned to receive different advertising
- Survey is used to measure outcomes, when treatment assignment is already known

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage
  2. Different local media markets randomly assigned to receive different advertising
- Survey is used to measure outcomes, when treatment assignment is already known
- Issues

III. Randomized Field Treatment

- Examples:
  1. Citizens randomly sent a letter by post encouraging them to reduce water usage
  2. Different local media markets randomly assigned to receive different advertising
- Survey is used to measure outcomes, when treatment assignment is already known
- Issues
  - Nonresponse
  - Noncompliance

IV. Treatment Encouragement

- Design:
  - T1: Encourage treatment
  - T2: Measure effects
- Examples:
  1. Albertson and Lawrence

IV. Treatment Encouragement

- Design:
    - T1: Encourage treatment
    - T2: Measure effects
- Examples:
    1. Albertson and Lawrence
- Issues

IV. Treatment Encouragement

- Design:
  - T1: Encourage treatment
  - T2: Measure effects
- Examples:
  1. Albertson and Lawrence
- Issues
  - Nonresponse
  - Noncompliance

Treatment Noncompliance

- Definition:

  "when subjects who were assigned to receive the treatment go untreated or when subjects assigned to the control group are treated" [2]

[2] Gerber & Green. 2012. *Field Experiments*, p.132.

Treatment Noncompliance

- Definition:

  "when subjects who were assigned to receive the treatment go untreated or when subjects assigned to the control group are treated" [2]
- Several strategies
  - "As treated" analysis
  - "Intention to treat" analysis
  - Estimate a LATE

[2] Gerber & Green. 2012. *Field Experiments*, p.132.

Analyzing Noncompliance

- If noncompliance only occurs in one group, it is *asymmetric* or *one-sided*
- We can ignore non-compliance and analyze the "intention to treat" effect, which will underestimate our effects because some people were not treated as assigned: $ITT = \overline{Y}_1 - \overline{Y}_0$

Analyzing Noncompliance

- If noncompliance only occurs in one group, it is *asymmetric* or *one-sided*
- We can ignore non-compliance and analyze the "intention to treat" effect, which will underestimate our effects because some people were not treated as assigned: $ITT = \overline{Y}_1 - \overline{Y}_0$
- We can use "instrumental variables" to estimate the "local average treatment effect" (LATE) for those that complied with treatment: $LATE = \frac{ITT}{\%Compliant}$

Local Average Treatment Effect

- IV estimate is *local* to the variation in X that is due to variation in D
- This matters if effects are *heterogeneous*
- LATE is effect for those who *comply*
- Four subpopulations:
  - Compliers: $X = 1$ only if $D = 1$
  - Always-takers: $X = 1$ regardless of D
  - Never-takers: $X = 0$ regardless of D
  - Defiers: $X = 1$ only if $D = 0$
- Exclusion restriction! Monotonicity!

Questions?

1 Other Survey Experimental Designs

2 Attention and Satisficing

3 Moderators and Effect Heterogeneity
- Manipulate the Moderator
- Blocking/Block Randomization
- Post-hoc Approaches

How should we deal with respondents that appear to not be paying attention, not "taking" the treatment, or not responding to outcome measures?

1. Keep them
2. Throw them away

Best Practice: Protocol

- Excluding respondents based on survey behavior is one of the easiest ways to "p-hack" an experimental dataset
  - Inattention, satisficing, etc. will tend to reduce the size of the SATE
- So regardless of how you handle these respondents, these should be decisions that are made *pre-analysis*

When are you excluding participants?

Pre-Treatment	Post-Treatment

When are you excluding participants?

Pre-Treatment	Post-Treatment
- Satisficing behaviors	

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention

Post-Treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection

Post-Treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection
- Pretreated

Post-Treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection
- Pretreated

Post-Treatment

- Speeding on treatment

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection
- Pretreated

Post-Treatment

- Speeding on treatment
- "Failing" a manipulation check

When are you excluding participants?

Pre-Treatment

- Satisficing behaviors
- Inattention
- Covariate-based selection
- Pretreated

Post-Treatment

- Speeding on treatment
- "Failing" a manipulation check
- Drop-off

Pre-Treatment Exclusion

- This is totally fine from a causal inference perspective

Pre-Treatment Exclusion

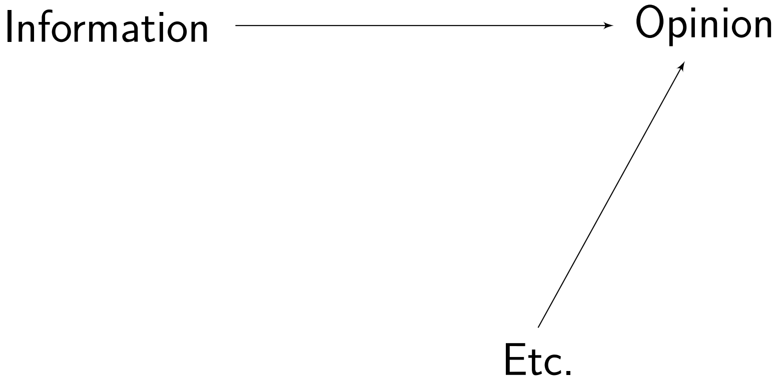
- This is totally fine from a causal inference perspective
- Advantages:
  - Focused on engaged respondents
  - Likely increase impact of treatment

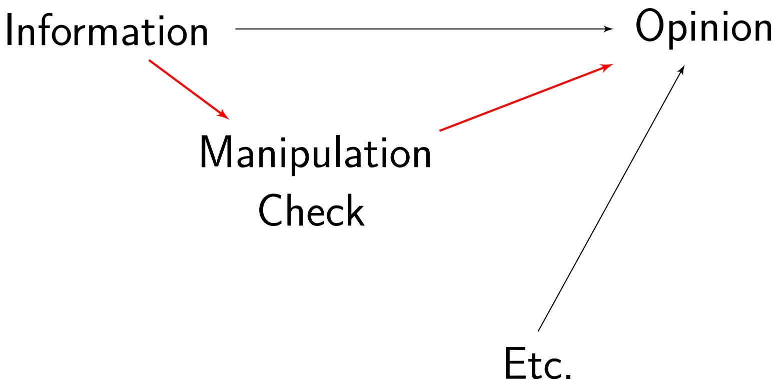
Pre-Treatment Exclusion

- This is totally fine from a causal inference perspective
- Advantages:
  - Focused on engaged respondents
  - Likely increase impact of treatment
- Disadvantages:
  - Changing definition of sample (and thus population)

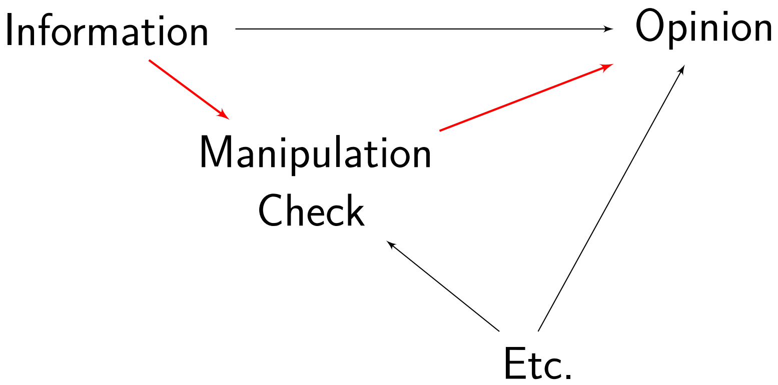
Post-Treatment Exclusion

This is much more problematic because it involves controlling for a *post-treatment* variable


Information
Opinion
Etc.



Risk that estimate of β_1 is diminished because effect is being carried through the manipulation check.



Introduction of “collider bias” wherein values of the manipulation check are affected by other factors.

Post-Treatment Exclusion

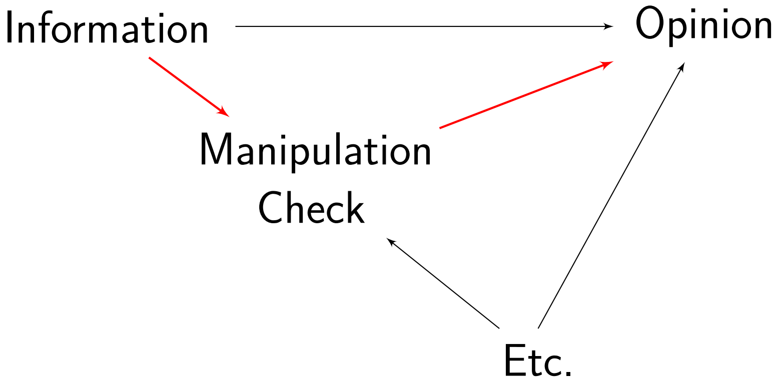
- Any post-treatment exclusion is problematic and should be avoided

Post-Treatment Exclusion

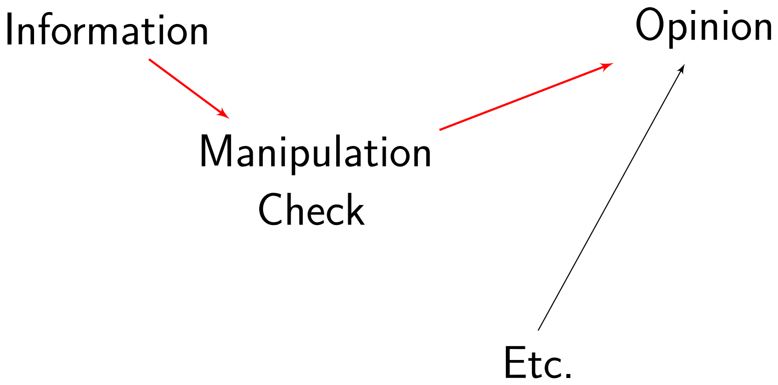
- Any post-treatment exclusion is problematic and should be avoided
- Can estimate a LATE
    - Interpretation: Effect of manipulation check among those whose value of the check can be changed by the treatment manipulation

Post-Treatment Exclusion

- Any post-treatment exclusion is problematic and should be avoided
- Can estimate a LATE
  - Interpretation: Effect of manipulation check among those whose value of the check can be changed by the treatment manipulation
- Non-response or attrition is the same as researcher-imposed exclusion
  - Not problematic if MCAR
  - Nothing really to be done if caused by treatment



Introduction of “collider bias” wherein values of the manipulation check are affected by other factors.


Information
Manipulation Check
Opinion
Etc.

Post-Treatment Exclusion

- Any post-treatment exclusion is problematic and should be avoided
- Can estimate a LATE
  - Interpretation: Effect of manipulation check among those whose value of the check can be changed by the treatment manipulation

Post-Treatment Exclusion

- Any post-treatment exclusion is problematic and should be avoided
- Can estimate a LATE
  - Interpretation: Effect of manipulation check among those whose value of the check can be changed by the treatment manipulation
- Non-response or attrition is the same as researcher-imposed exclusion
  - Not problematic if MCAR
  - Nothing really to be done if caused by treatment

Questions?

1 Other Survey Experimental Designs

2 Attention and Satisficing

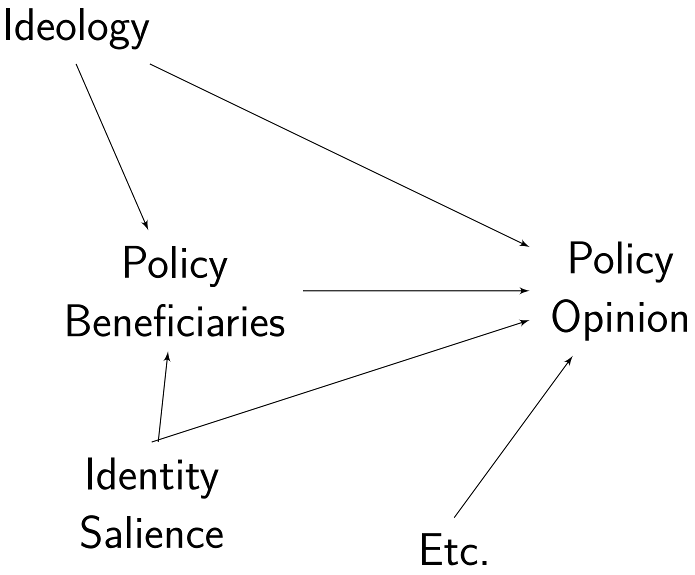
3 Moderators and Effect Heterogeneity
- Manipulate the Moderator
- Blocking/Block Randomization
- Post-hoc Approaches

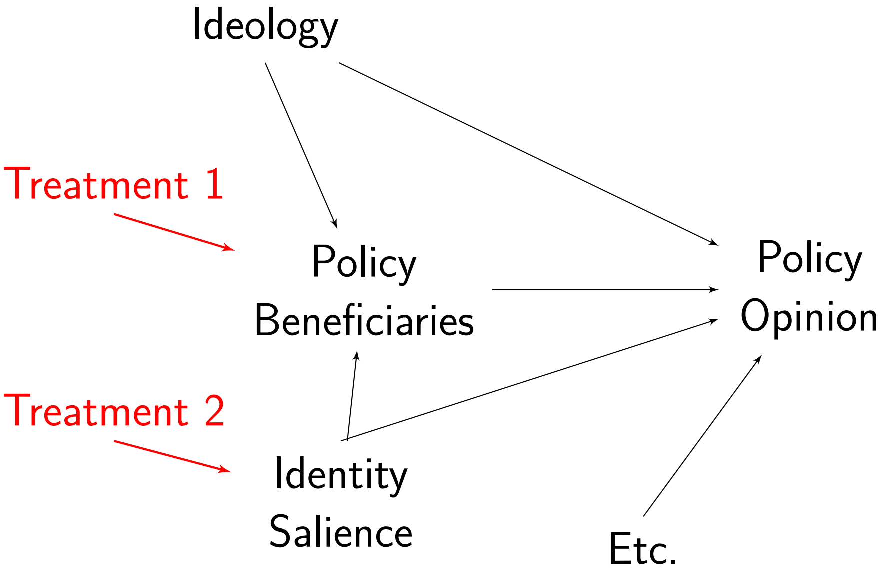
If we think there might be effect heterogeneity, what can we do?

- Best solution: manipulate the moderator
- Next best: block on the moderator
- Least best: post-hoc exploratory approaches

Simply: Manipulating the moderator variable is the best way to estimate a heterogeneous effect!

Why is this true?


Ideology
Policy Beneficiaries
Policy Opinion
Identity Salience
Etc.


Ideology
Treatment 1
Policy Beneficiaries
Policy Opinion
Treatment 2
Identity Salience
Etc.

Ex. Question-as-treatment[3]

- How close do you feel to your ethnic or racial group?
- Some people have said that taxes need to be raised to take care of pressing national needs. How willing would you be to have your taxes raised to improve education in public schools?

[3] Transue. 2007. "Identity Salience, Identity Acceptance, and Racial Policy Attitudes: American National Identity as a Uniting Force." *American Journal of Political Science* 51(1): 78–91.

Ex. Question-as-treatment[3]

- How close do you feel to other Americans?
- Some people have said that taxes need to be raised to take care of pressing national needs. How willing would you be to have your taxes raised to improve education in public schools?

[3] Transue. 2007. "Identity Salience, Identity Acceptance, and Racial Policy Attitudes: American National Identity as a Uniting Force." *American Journal of Political Science* 51(1): 78–91.

Ex. Question-as-treatment[3]

- How close do you feel to your ethnic or racial group?
- Some people have said that taxes need to be raised to take care of pressing national needs. How willing would you be to have your taxes raised to improve educational opportunities for minorities?

---

[3] Transue. 2007. "Identity Salience, Identity Acceptance, and Racial Policy Attitudes: American National Identity as a Uniting Force." *American Journal of Political Science* 51(1): 78–91.

Ex. Question-as-treatment[3]

- How close do you feel to other Americans?
- Some people have said that taxes need to be raised to take care of pressing national needs. How willing would you be to have your taxes raised to improve educational opportunities for minorities?

---

[3] Transue. 2007. "Identity Salience, Identity Acceptance, and Racial Policy Attitudes: American National Identity as a Uniting Force." *American Journal of Political Science* 51(1): 78–91.

2x2 Factorial Design

Condition	
Educ. for Minorities	Y_1
Schools	Y_0

2x2 Factorial Design

Condition	Americans	Own Race
Educ. for Minorities	$Y_{1,0}$	$Y_{1,1}$
Schools	$Y_{0,0}$	$Y_{0,1}$

Two ways to estimate this

Dummy variable regression:

$$Y = \beta_0 + \beta_1 X_{0,1} + \beta_2 X_{1,0} + \beta_3 X_{1,1} + \epsilon$$

Interaction effect:

$$Y = \beta_0 + \beta_1 X1_1 + \beta_2 X2_1 + \beta_3 X1_1 * X2_1 + \epsilon$$

Considerations

- Need to have hypotheses about heterogeneity a priori
- Factorial designs can quickly become unwieldy and expensive

Probably obvious, but. . .

Factors	Conditions per factor	Total Conditions	n
1	2	2	400
1	3	3	600
1	4	4	800
2	2	4	800
2	3	6	1200
2	4	8	1600
3	3	9	1800
3	4	12	2400
4	4	16	3200

Assumes power to detect a relatively small effect, but no consideration of multiple comparisons.

Considerations

- Need to have hypotheses about heterogeneity a priori
- Factorial designs can quickly become unwieldy and expensive

Considerations

- Need to have hypotheses about heterogeneity a priori
- Factorial designs can quickly become unwieldy and expensive
- Need to consider what CATEs are of theoretical interest
  - Treatment–control
  - Treatment–treatment

Considerations

- Need to have hypotheses about heterogeneity a priori
- Factorial designs can quickly become unwieldy and expensive
- Need to consider what CATEs are of theoretical interest
  - Treatment–control
  - Treatment–treatment
- One solution may be conjoint designs

Conjoint Designs I

- "Classic vignettes" taken to an extreme
- Example: Judge whether to admit an immigrant to your country

Conjoint Designs I

- "Classic vignettes" taken to an extreme
- Example: Judge whether to admit an immigrant to your country
- Respondents see a series of vignettes that are fully randomized along any number of dimensions
  - Sex
  - Education
  - Language proficiency
  - etc.

Conjoint Designs I

- "Classic vignettes" taken to an extreme
- Example: Judge whether to admit an immigrant to your country
- Respondents see a series of vignettes that are fully randomized along any number of dimensions
    - Sex
    - Education
    - Language proficiency
    - etc.
- Outcome is judgment (binary or rating scale)

Conjoint Designs II

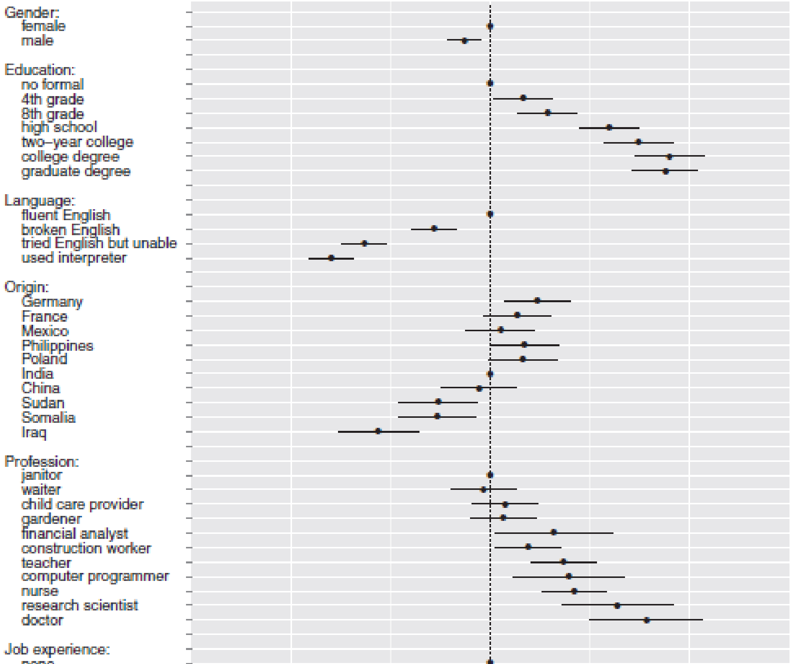
Why is this useful?

- Understand complex decision-making
- Within-subjects comparisons
- Heterogeneous effects across versions of treatment
- Pilot testing: Sensitivity of design to specification of *compound* vignette

Please read the descriptions of the potential immigrants carefully. Then, please indicate which of the two immigrants you would personally prefer to see admitted to the United States.

	Immigrant 1	Immigrant 2
Prior Trips to the U.S.	Entered the U.S. once before on a tourist visa	Entered the U.S. once before on a tourist visa
Reason for Application	Reunite with family members already in U.S.	Reunite with family members already in U.S.
Country of Origin	Mexico	Iraq
Language Skills	During admission interview, this applicant spoke fluent English	During admission interview, this applicant spoke fluent English
Profession	Child care provider	Teacher
Job Experience	One to two years of job training and experience	Three to five years of job training and experience
Employment Plans	Does not have a contract with a U.S. employer but has done job interviews	Will look for work after arriving in the U.S.
Education Level	Equivalent to completing two years of college in the U.S.	Equivalent to completing a college degree in the U.S.
Gender	Female	Male

	Immigrant 1	Immigrant 2
If you had to choose between them, which of these two immigrants should be given priority to come to the United States to live?	○	○


Gender:
female
male
Education:
no formal
4th grade
8th grade
high school
two–year college
college degree
graduate degree
Language:
fluent English
broken English
tried English but unable
used interpreter
Origin:
Germany
France
Mexico
Philippines
Poland
India
China
Sudan
Somalia
Iraq
Profession:
janitor
waiter
child care provider
gardener
financial analyst
construction worker
teacher
computer programmer
nurse
research scientist
doctor
Job experience:

Conjoint Designs III

Conjoint Designs III

- As long as profiles are randomized, this is just a complex factorial design where we can estimate *marginal effect* of each attribute
  - Treatment–control SATE, conditional on all other randomized factors

Conjoint Designs III

- As long as profiles are randomized, this is just a complex factorial design where we can estimate *marginal effect* of each attribute
    - Treatment–control SATE, conditional on all other randomized factors
- Assumptions:
    - Fully randomized profiles
    - No "carry-over" effects
    - No profile order effects

Questions?

Activity!

- Work in groups of 2–3
- Consider a research question of interest to you
  - Opinions on policies
  - Product purchasing decisions
  - Information selection
  - Attitudes toward out-groups
  - etc.
- Try to describe a basic conjoint design
  - What is the outcome?
  - What are the factors/features?

Block Randomization I

Stratification:Sampling::Blocking:Experiments

Block Randomization I

Stratification:Sampling::Blocking:Experiments

- Basic idea: randomization occurs within strata defined before treatment assignment

Block Randomization I

Stratification:Sampling::Blocking:Experiments

- Basic idea: randomization occurs within strata defined before treatment assignment
- CATE is estimate for each stratum; aggregated to SATE

Block Randomization I

Stratification:Sampling::Blocking:Experiments

- Basic idea: randomization occurs within strata defined before treatment assignment
- CATE is estimate for each stratum; aggregated to SATE
- Why?
  - Eliminate chance imbalances
  - Optimized for estimating CATEs
  - More precise SATE estimate

Exp.	Control				Treatment			
1	M	M	M	M	F	F	F	F
2	M	M	M	F	M	F	F	F
3	M	M	F	F	M	M	F	F
4	M	F	F	F	M	M	M	F
5	F	F	F	F	M	M	M	M

```
# population of men and women
pop <- rep(c("Male", "Female"), each = 4)

# randomly assign into treatment and control
split(sample(pop, 8, FALSE), c(rep(0,4), rep(1,4)))
```

Obs.	X_{1i}	X_{2i}	D_i
1	Male	Old	0
2	Male	Old	1
3	Male	Young	1
4	Male	Young	0
5	Female	Old	1
6	Female	Old	0
7	Female	Young	0
8	Female	Young	1

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*
- When is blocking *statistically* useful?

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*
- When is blocking *statistically* useful?
  - If those covariates affect values of potential outcomes, blocking reduces the variance of the SATE

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*
- When is blocking *statistically* useful?
  - If those covariates affect values of potential outcomes, blocking reduces the variance of the SATE
  - Most valuable in small samples

Block Randomization II

- Blocking ensures ignorability of all covariates used to construct the blocks
- Incorporates covariates explicitly into the *design*
- When is blocking *statistically* useful?
    - If those covariates affect values of potential outcomes, blocking reduces the variance of the SATE
    - Most valuable in small samples
    - Not valuable if all blocks have similar potential outcomes

Statistical Properties I

Complete randomization:

$$SATE = \frac{1}{n_1}\sum Y_{1i} - \frac{1}{n_0}\sum Y_{0i}$$

Block randomization:

$$SATE_{blocked} = \sum_{1}^{J}\left(\frac{n_j}{n}\right)(\widehat{CATE}_j)$$

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	
2	Male	Old	1	10	
3	Male	Young	1	4	
4	Male	Young	0	1	
5	Female	Old	1	6	
6	Female	Old	0	2	
7	Female	Young	0	6	
8	Female	Young	1	9	

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	5
2	Male	Old	1	10	
3	Male	Young	1	4	
4	Male	Young	0	1	
5	Female	Old	1	6	
6	Female	Old	0	2	
7	Female	Young	0	6	
8	Female	Young	1	9	

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	5
2	Male	Old	1	10	
3	Male	Young	1	4	3
4	Male	Young	0	1	
5	Female	Old	1	6	
6	Female	Old	0	2	
7	Female	Young	0	6	
8	Female	Young	1	9	

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	5
2	Male	Old	1	10	
3	Male	Young	1	4	3
4	Male	Young	0	1	
5	Female	Old	1	6	4
6	Female	Old	0	2	
7	Female	Young	0	6	
8	Female	Young	1	9	

Obs.	X_{1i}	X_{2i}	D_i	Y_i	CATE
1	Male	Old	0	5	5
2	Male	Old	1	10	
3	Male	Young	1	4	3
4	Male	Young	0	1	
5	Female	Old	1	6	4
6	Female	Old	0	2	
7	Female	Young	0	6	3
8	Female	Young	1	9	

SATE Estimation

$$
\begin{aligned}
SATE &= \left(\frac{2}{8} * 5\right) + \left(\frac{2}{8} * 3\right) + \left(\frac{2}{8} * 4\right) + \left(\frac{2}{8} * 3\right) \\
&= 3.75
\end{aligned}
$$

SATE Estimation

$$
\begin{aligned}
SATE &= \left(\frac{2}{8} * 5\right) + \left(\frac{2}{8} * 3\right) + \left(\frac{2}{8} * 4\right) + \left(\frac{2}{8} * 3\right) \\
&= 3.75
\end{aligned}
$$

The blocked and unblocked estimates are the same here because $Pr(Treatment)$ is constant across blocks and blocks are all the same size.

SATE Estimation

- We can use weighted regression to estimate this in an OLS framework
- Weights are the inverse prob. of being treated w/in block
    - Pr(Treated) by block: $p_{ij} = Pr(D_i = 1|J = j)$
    - Weight (Treated): $w_{ij} = \frac{1}{p_{ij}}$
    - Weight (Control): $w_{ij} = \frac{1}{1 - p_{ij}}$

Statistical Properties II

Complete randomization:

$$\widehat{SE}_{SATE} = \sqrt{\frac{\widehat{Var}(Y_0)}{n_0} + \frac{\widehat{Var}(Y_1)}{n_1}}$$

Block randomization:

$$\widehat{SE}_{SATE_{blocked}} = \sqrt{\sum_1^J \left(\frac{n_j}{n}\right)^2 \widehat{Var}(SATE_j)}$$

Statistical Properties II

Complete randomization:

$$\widehat{SE}_{SATE} = \sqrt{\frac{\widehat{Var}(Y_0)}{n_0} + \frac{\widehat{Var}(Y_1)}{n_1}}$$

Block randomization:

$$\widehat{SE}_{SATE_{blocked}} = \sqrt{\sum_{1}^{J} \left(\frac{n_j}{n}\right)^2 \widehat{Var}(SATE_j)}$$

When is the blocked design more efficient?

Practicalities

- Blocked randomization only works in exactly the same situations where stratified sampling works
  - Need to observe covariates pre-treatment in order to block on them
  - Work best in a panel context
- In a single cross-sectional design that might be challenging
  - Some software can block "on the fly"

Questions?

Three Post-hoc Approaches

- Suggestive evidence
- Regression using treatment-by-covariate interactions
- Automated approaches

Three Post-hoc Approaches

- Suggestive evidence
- Regression using treatment-by-covariate interactions
- Automated approaches
- (Replication and meta-analysis)

Suggestive Evidence

We can never know $Var(TE_i)$!

Suggestive Evidence

We can never know $Var(TE_i)$! But...

- Quantile-quantile plots
- Equality of variance tests

Suggestive Evidence

We can never know $Var(TE_i)$! But...

- Quantile-quantile plots
  - Compare the distribution of Y_0's to distribution of Y_1's
  - If homogeneity, a vertical shift in Y_1's
  - If heterogeneity, a slope $\neq 1$
- Equality of variance tests

Suggestive Evidence

We can never know $Var(TE_i)$! But...

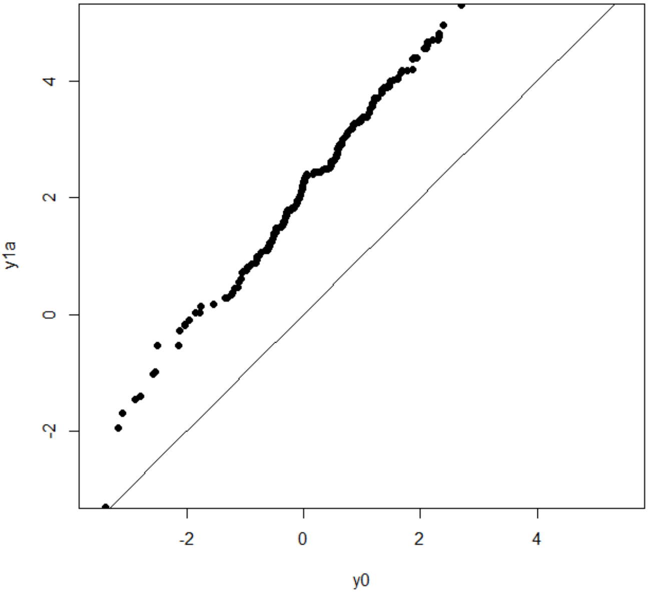
- Quantile-quantile plots
  - Compare the distribution of Y_0's to distribution of Y_1's
  - If homogeneity, a vertical shift in Y_1's
  - If heterogeneity, a slope $\neq 1$
- Equality of variance tests
  - If homogeneity, variance should be equal
  - If heterogeneity, variances should differ

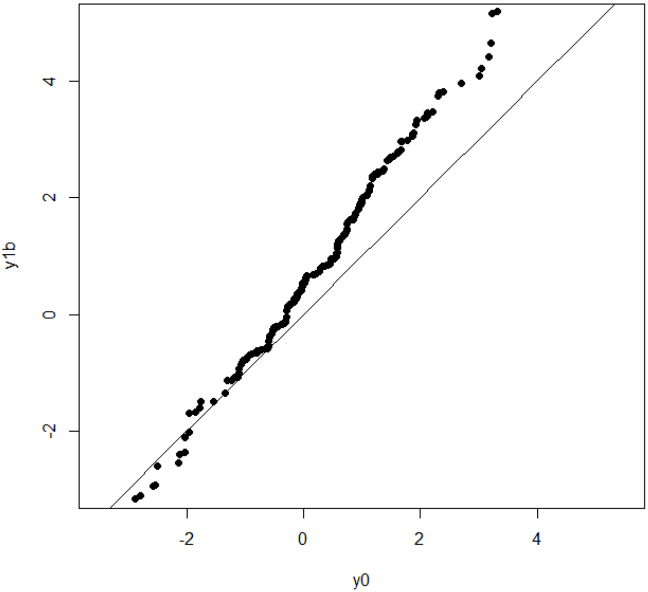
QQ Plots

```
# y_0 data
set.seed(1)
n <- 200
y0 <- rnorm(n) + rnorm(n, 0.2)

# y_1 data (homogeneous effects)
y1a <- y0 + 2 + rnorm(n, 0.2)
# y_1 data (heterogeneous effects)
y1b <- y0 + rep(0:1, each = n/2) + rnorm(n, 0.2)

qqplot(y0, y1a, pch=19, xlim=c(-3,5), ylim=c(-3,5), asp=1)
curve((x), add = TRUE)
qqplot(y0, y1b, pch=19, xlim=c(-3,5), ylim=c(-3,5), asp=1)
curve((x), add = TRUE)
```


y1a
y0
-2
0
2
4
-2
0
2
4



y1b
y0
-2
0
2
4

Equality of Variance tests

```
> var.test(y0, y1a)

        F test to compare two variances

data:  y0 and y1a
F = 0.60121, num df = 199, denom df = 199,
  p-value = 0.0003635
alternative hypothesis:
  true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4549900 0.7944289
sample estimates:
ratio of variances
         0.6012131
```

Equality of Variance tests

```
> var.test(y0, y1b)

        F test to compare two variances

data:  y0 and y1b
F = 0.53483, num df = 199, denom df = 199,
  p-value = 1.224e-05
alternative hypothesis:
  true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4047531 0.7067133
sample estimates:
ratio of variances
         0.5348312
```

Questions?

Regression Estimation

Aside: Regression Adjustment in Experiments, Generally

- Recall the general advice that we do not need covariates in the regression to "control" for omitted variables (because there are none)
- Including covariates can reduce variance of our SATE by explaining more of the variation in Y

Scenario

Imagine two regression models. Which is correct?

1. Mean-difference estimate of SATE is "not significant"
2. Regression estimate of SATE, controlling for sex, age, and education, is "significant"

Scenario

Imagine two regression models. Which is correct?

1. Mean-difference estimate of SATE is "not significant"
2. Regression estimate of SATE, controlling for sex, age, and education, is "significant"

This is a small-sample dynamic, so make these decisions pre-analysis!

Treatment-Covariate Interactions

- The regression paradigm allows us to estimate CATEs using interaction terms
  - X is an indicator for treatment
  - M is an indicator for possible moderator

Treatment-Covariate Interactions

- The regression paradigm allows us to estimate CATEs using interaction terms
  - X is an indicator for treatment
  - M is an indicator for possible moderator
- SATE: $Y = \beta_0 + \beta_1 X + e$

Treatment-Covariate Interactions

- The regression paradigm allows us to estimate CATEs using interaction terms
  - X is an indicator for treatment
  - M is an indicator for possible moderator
- SATE: $Y = \beta_0 + \beta_1 X + e$
- CATEs:

$$Y = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 X * M + e$$

Treatment-Covariate Interactions

- The regression paradigm allows us to estimate CATEs using interaction terms
  - X is an indicator for treatment
  - M is an indicator for possible moderator
- SATE: $Y = \beta_0 + \beta_1 X + e$
- CATEs:

$$Y = \beta_0 + \beta_1 X + \beta_2 M + \beta_3 X * M + e$$

  - Homogeneity: $\beta_1 = \beta_3$
  - Heterogeneity: $\beta_1 \neq \beta_3$

Considerations

Considerations

- Coefficients on moderators have no causal interpretation without further conditioning on observables

Considerations

- Coefficients on moderators have no causal interpretation without further conditioning on observables
- Nearly unlimited potential moderators
    - First-order interactions with every covariate in dataset
    - Second-, third-order, etc. interactions
- Thus, multiple comparisons problem!

Considerations

- Coefficients on moderators have no causal interpretation without further conditioning on observables
- Nearly unlimited potential moderators
    - First-order interactions with every covariate in dataset
    - Second-, third-order, etc. interactions
- Thus, multiple comparisons problem!
- Power (esp. if M is continuous)

Questions?

BART

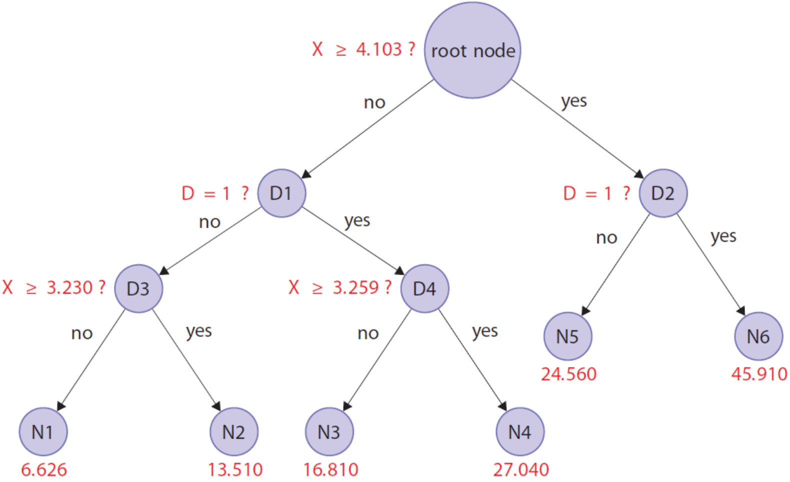
- Estimate CATEs in a fully automated fashion

BART

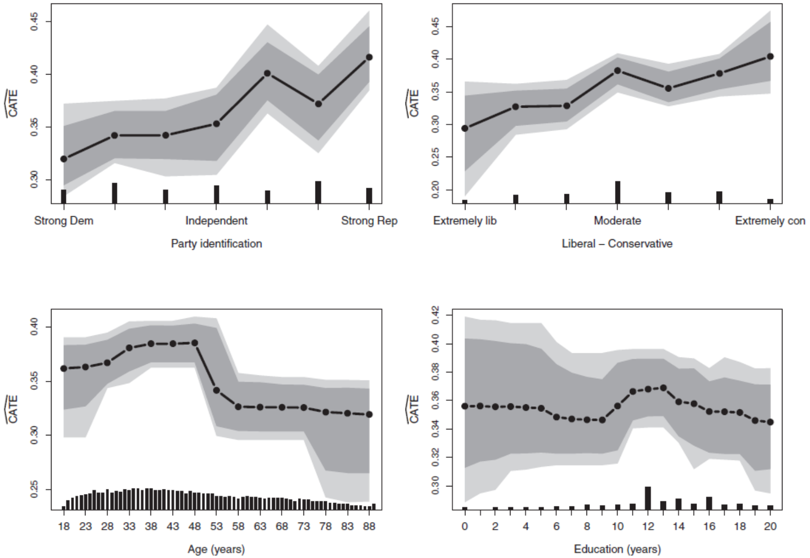
- Estimate CATEs in a fully automated fashion
- "Bayesian Additive Regression Trees"
    - Essentially an ensemble machine learning method

BART

- Estimate CATEs in a fully automated fashion
- "Bayesian Additive Regression Trees"
  - Essentially an ensemble machine learning method
- Iteratively split a sample into more and more homogeneous groups until some threshold is reached using binary (cutpoint) decisions
- Repeat this a bunch of times, aggregating across results



Green & Kern. 2012. "Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees." *Public Opinion Quarterly* 76(3): 491–511.



CATE
Strong Dem
Independent
Strong Rep
Party identification
Extremely lib
Moderate
Extremely con
Liberal – Conservative
Age (years)
Education (years)

Considerations

- BART is totally automated, conditional on the set of covariates used
- Only really works with dichotomous covariates
- Not widely used or tested
- Totally post-hoc and atheoretical

Replication!

- If we think effects are homogeneous (across SUTO), then replications in other SUTO conditions should provide us the same SATE (within sampling error)
- If we think effects are heterogeneous, then replications should give *systematically* different SATE (or CATE) estimates

Replication!

- If we think effects are homogeneous (across SUTO), then replications in other SUTO conditions should provide us the same SATE (within sampling error)
- If we think effects are heterogeneous, then replications should give *systematically* different SATE (or CATE) estimates
  - Identify those patterns of heterogeneity using meta-analysis

Replication!

- If we think effects are homogeneous (across SUTO), then replications in other SUTO conditions should provide us the same SATE (within sampling error)
- If we think effects are heterogeneous, then replications should give *systematically* different SATE (or CATE) estimates
    - Identify those patterns of heterogeneity using meta-analysis
    - Regress effect estimates from multiple studies on SUTO features of each study

Conclusion

- Do we want to know SATE, CATE(s), or both?

Conclusion

- Do we want to know SATE, CATE(s), or both?
- Decide in advance
  - Include in protocol
  - Design study to estimate CATE(s)

Conclusion

- Do we want to know SATE, CATE(s), or both?
- Decide in advance
  - Include in protocol
  - Design study to estimate CATE(s)
- Estimation of CATEs
  - Block randomization
  - Post-hoc procedures

Questions?

Apparent Satisficing

- Some common measures:
  - “Straightlining”
  - Non-differentiation
  - Acquiescence
  - Nonresponse
  - DK responding
  - Speeding
- Difficult to detect and distinguish from “real” responses

Metadata/Paradata

- Timing
  - Some survey tools will allow you to time page
  - Make a prior rules about dropping participants for speeding

Metadata/Paradata

- Timing
  - Some survey tools will allow you to time page
  - Make a prior rules about dropping participants for speeding
- Mousetracking or eyetracking
  - Mousetracking is unobtrusive
  - Eyetracking requires participants opt-in

Metadata/Paradata

- Timing
  - Some survey tools will allow you to time page
  - Make a prior rules about dropping participants for speeding
- Mousetracking or eyetracking
  - Mousetracking is unobtrusive
  - Eyetracking requires participants opt-in
- Record focus/blur browser events

Direct Measures

- How closely have you been paying attention to what the questions on this survey actually mean?

Direct Measures

- How closely have you been paying attention to what the questions on this survey actually mean?
- While taking this survey, did you engage in any of the following behaviors? Please check all that apply.
  - Use your mobile phone
  - Browse the internet
  - ...

Instructional Manipulation Check

We would like to know if you are reading the questions on this survey. If you are reading carefully, please ignore this question, do not select any answer below, and click "next" to proceed with the survey.

Strongly disagree
Somewhat disagree
Neither agree nor disagree
Somewhat agree
Strongly agree

Instructional Manipulation Check

Do you agree or disagree with the decision to send British forces to fight ISIL in Syria? We would like to know if you are reading the questions on this survey. If you are reading carefully, please ignore this question, do not select any answer below, and click "next" to proceed with the survey.

Strongly disagree
Somewhat disagree
Neither agree nor disagree
Somewhat agree
Strongly agree

Return