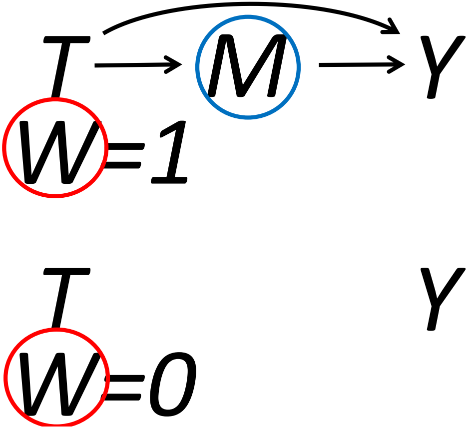
Lecture 16: Moderators, Mediators, and Causal Explanation II

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

April 4, 2022

Motivation



- M =“Mediator”
- W =“Moderator”

Motivation

- Analysis of “moderators” = characterizing effect heterogeneity.
- Useful for
  - Adjudicating between rival explanations/theories of an effect.
  - Defining optimal treatment regimes.
  - Assessing generalizability to other contexts.

Motivation

- Essentially we are talking about interaction effects:

$$\text{“}Y_i = \alpha + \rho D_i + \gamma W_i + \lambda D_i W_i + \varepsilon_i\text{”}$$

- If W binary or otherwise has a linear moderating effect, this specification is adequate and can test for moderator effects with usual test on λ.
- But if W is not binary, then we need to consider possible non-linearities.

Nonlinear Moderator Effects

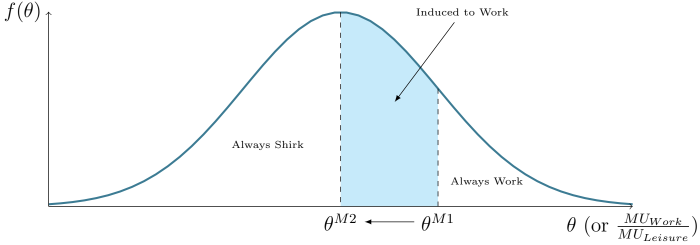


Figure A.1: Effect of an Increase in Detection Probability on the Decision to Work or Shirk

(Callen et al. 2018)

Nonlinear Moderator Effects

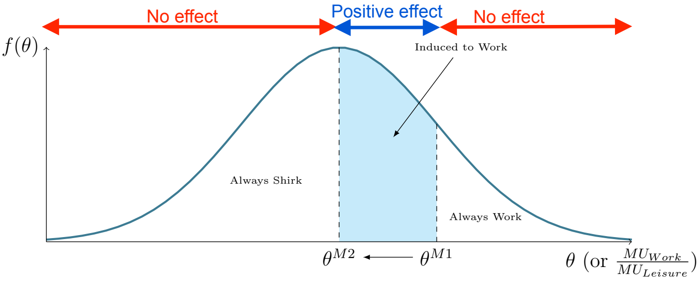


Figure A.1: Effect of an Increase in Detection Probability on the Decision to Work or Shirk

(Callen et al. 2018)

Nonlinear Moderator Effects

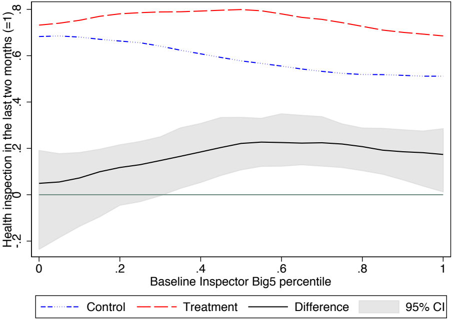


Figure 8: Nonparametric treatment effects

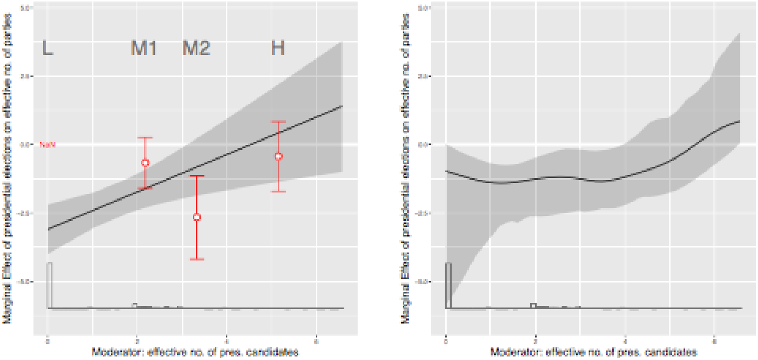
Notes: This figure plots a kernel-weighted local polynomial regression of whether a clinic had a health inspection in the last two months on every 5th percentile of baseline Big Five index separately for treatment and control districts, as well as the difference at each 5th percentile of baseline scores. The confidence intervals of the treatment effects are constructed by drawing 1,000 bootstrap samples of data that preserve the within-district correlation structure in the original data and plotting the 95 percent range for the treatment effect at each 5th percentile of baseline scores.

(Callen et al. 2018)

Nonlinear Moderator Effects

- Hainmueller et al. (2018) discuss estimation with `interflex` package for Stata and R:

FIGURE 7. NONLINEARITY: CLARK AND GOLDER (2006)



(a) (b)

Note: The above plots examine the marginal effects plot in Clark and Golder (2006): (a) marginal effects estimates from the replicated model (black line) and the binning estimator (red dots); (b) marginal effects estimates from the kernel estimator.

Extended applications

- But there is more we can do than just test for interaction effects.

Extended applications

American Economic Journal: Applied Economics 2013, 5(4): 1–27
http://dx.doi.org/10.1257/app.5.4.1

Explaining Charter School Effectiveness†

By Joshua D. Angrist, Parag A. Pathak, and Christopher R. Walters*

Estimation of Heterogeneous Treatment Effects from Randomized Experiments, with Application to the Optimal Planning of the Get-Out-the-Vote Campaign

Kosuke Imai
Department of Politics, Princeton University, Princeton, NJ 08544
e-mail: kimai@princeton.edu (corresponding author)

Aaron Strauss
The Mellman Group, 1023 31st St NW, 5th Floor, Washington, DC 20007
e-mail: aaronbs@gmail.com

American Economic Journal: Applied Economics 2013, 5(4): 1–27
http://dx.doi.org/10.1257/app.5.4.1

Explaining Charter School Effectiveness†

By Joshua D. Angrist, Parag A. Pathak, and Christopher R. Walters*

Unpacking effect heterogeneity

- Angrist et al. study the puzzle of the urban/non-urban gap in charter school effectiveness.
- They use a variety of techniques:
  1. Plot potential outcome means.
  2. Decompose effects by student characteristics.
  3. Study moderating effects of school characteristics.

The puzzle

Table 4: Lottery Results for Massachusetts Charter Schools

				2SLS	
		First Stage	Reduced Form	Just identified	Overidentified
School level	Subject	(1)	(2)	(3)	(4)
Middle	ELA	0.987***	0.065**	0.066**	0.062**
		(0.043)	(0.029)	(0.029)	(0.028)
	N			12126	
	Math	0.984***	0.211***	0.214***	0.175***
		(0.043)	(0.034)	(0.033)	(0.031)
	N			12346	
High	ELA	0.509***	0.113**	0.221***	0.190**
		(0.101)	(0.050)	(0.076)	(0.074)
	N			3303	
	Math	0.510***	0.164**	0.322***	0.269***
		(0.101)	(0.064)	(0.090)	(0.093)
	N			3255	
	Writing Topic	0.514***	0.156***	0.303***	0.290***
		(0.101)	(0.057)	(0.087)	(0.080)
	N			3268	
	Writing Composition	0.514***	0.140**	0.271***	0.227***
		(0.101)	(0.058)	(0.092)	(0.085)
	N			3268	

Notes: This table reports estimates of the effects of years in charter schools on test scores. The sample is restricted to students with baseline demographic characteristics who attended a Massachusetts public school when tested, and excludes students with sibling priority and late applicants. Columns (1)-(3) are produced by a 2SLS procedure using a lottery offer dummy as an instrument for years spent in charter schools. Column (4) uses risk set and offer interactions as instruments. All models control for race, sex, special education, limited English proficiency, subsidized lunch status, and a female by minority dummy. Year of birth, year of test, and risk set dummies are also included. Middle school regressions pool post-lottery outcomes from 4th through 8th grade and cluster by student identifier as well as school-grade-year. High school regressions include only scores for 10th grade and cluster by school-grade-year.

*significant at 10%; **significant at 5%; ***significant at 1%

The puzzle

Table 5: Lottery Results for Urban and Non-urban Charter Schools

			Urban			Non-urban		
			First Stage	Reduced Form	2SLS	First Stage	Reduced Form	2SLS
School level	Subject		(1)	(2)	(3)	(4)	(5)	(6)
Middle	ELA		1.001***	0.141***	0.140***	0.978***	-0.155***	-0.156***
			(0.055)	(0.035)	(0.033)	(0.081)	(0.045)	(0.045)
		N		8762			3364	
	Math		0.990***	0.333***	0.336***	0.996***	-0.159***	-0.155***
			(0.054)	(0.038)	(0.036)	(0.081)	(0.050)	(0.051)
		N		9015			3331	
High	ELA		0.494***	0.117**	0.236***	1.082***	-0.014	-0.009
			(0.105)	(0.051)	(0.079)	(0.153)	(0.116)	(0.105)
		N		2954			349	
	Math		0.495***	0.178***	0.359***	1.088***	-0.274*	-0.246*
			(0.105)	(0.066)	(0.092)	(0.158)	(0.162)	(0.148)
		N		2910			345	
	Writing Topic		0.500***	0.166***	0.332***	1.082***	-0.157	-0.139
			(0.105)	(0.058)	(0.090)	(0.153)	(0.222)	(0.204)
		N		2920			348	
	Writing Composition		0.500***	0.149**	0.298***	1.082***	-0.155	-0.137
			(0.105)	(0.060)	(0.096)	(0.153)	(0.213)	(0.196)
		N		2920			348	

Notes: This table reports estimates of the effects of years in urban and non-urban charter schools on test scores. The sample is restricted to students with baseline demographic characteristics who attended a Massachusetts public school when tested, and excludes students with sibling priority and late applicants. Estimates are produced by a 2SLS procedure using urban and non-urban lottery offers as instruments for attendance at urban and non-urban charter schools. All models control for race, sex, special education, limited English proficiency, subsidized lunch status, and a female by minority dummy. Year of birth, year of test, and risk set dummies are also included. Middle school regressions pool post-lottery outcomes from 4th through 8th grade and cluster by student identifier as well as school-grade-year. High school regressions include only scores for 10th grade, and cluster by school-grade-year.

*significant at 10%; **significant at 5%; ***significant at 1%

The puzzle

- Differences due to differences in types of students served?
- That is, maybe we're just talking about radically different potential outcome distributions:

$$\tau_u - \tau_n = \underbrace{E_u[Y_{1i}|D_{1i} > D_{0i}] - E_n[Y_{1i}|D_{1i} > D_{0i}]}_{\gamma_1} \\ - \underbrace{(E_u[Y_{0i}|D_{1i} > D_{0i}] - E_n[Y_{0i}|D_{1i} > D_{0i}])}_{\gamma_0}. \tag{5}$$

Potential outcome differentials

- In the IV setting, we can estimate γ_0:

  Pooling urban and non-urban charter applicants, we estimate γ_0 using

  $$Y_i(1-D_i) = \psi(1-D_i) + \gamma_0(1-D_i)\cdot 1\{U_i = u\} + \sum_j \delta_j d_{ij} + \epsilon_i, \tag{6}$$

  with first stage

  $$1 - D_i = \sum_j \kappa_j d_{ij} + \sum_j \pi_j d_{ij} Z_i + \eta_i. \tag{7}$$

  The first stage equation for the interaction between $1 - D_i$ and urban status uses the same specification as equation (7).[8] For a model without covariates, Abadie (2003) shows that 2SLS

- Similar for γ_1.

Potential outcome differentials

Table 8: Urban Gaps in Treatment and No-treatment Counterfactuals

					Differences in potential outcomes	
		Urban effect	Non-urban effect	Effect difference	Υ_0	Υ_1
School level	Subject	(1)	(2)	(3)	(4)	(5)
Middle	ELA	0.154**	-0.218***	0.371***	-0.705***	-0.333***
		(0.074)	(0.054)	(0.092)	(0.081)	(0.065)
	N	3817	1851	5668	5668	5668
	Math	0.468***	-0.252***	0.720***	-0.628***	0.092
		(0.084)	(0.072)	(0.111)	(0.085)	(0.071)
	N	4127	1768	5895	5895	5895

Notes: This table estimates the components of the difference in charter treatment effects by urban status due to differences in non-charter "fallback" and differences in treated outcomes. Outcomes are test scores the year after the lottery. Columns (1) and (2) display urban and non-urban charter treatment effects, respectively, and column (3) gives the difference. Column (4) shows an estimate of the difference in average non-treated outcomes between urban and non-urban compliers, computed as described in the text. Column (5) shows an estimate of the difference in treated outcomes between these two groups.

*significant at 10%; **significant at 5%; ***significant at 1%

Potential outcome differentials

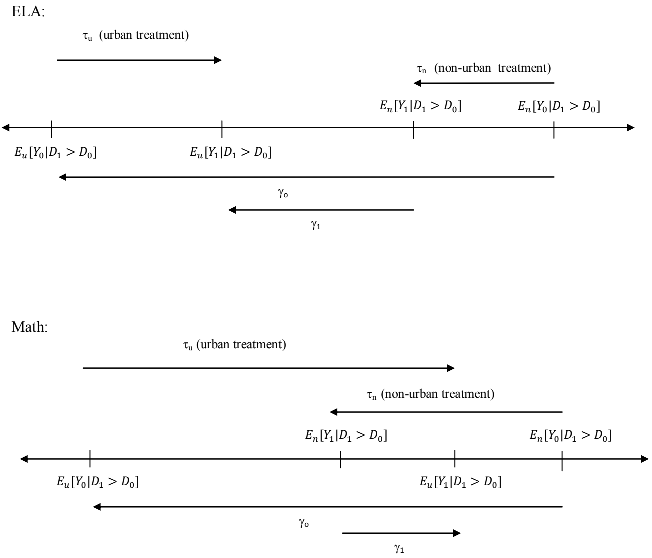


Figure 1: Gaps in Treatment and No-Treatment Counterfactuals, Urban vs. Non

Effect decomposition

How much of the urban/non-urban gap in *effect sizes* is due to urban/non-urban differences in *student characteristics*?

- Oaxaca-Blinder linear decomposition.
- Start with model that interacts student characteristics with treatment in each location type:

$$E_l[Y_i|D_{1i} > D_{0i}, D_i, X_i, d_{ij}] = X_i'\theta_l + \omega_l D_i + D_i X_i'\rho_l + \sum_j \delta_j d_{ij}, \quad l \in \{u, n\}. \tag{10}$$

$$\Rightarrow \tau_l = \omega_l + \mathrm{E}\left[X_l|D_{1i} > D_{0i}\right]'\rho_l$$

Then,

$$\tau_u - \tau_n = (\omega_u - \omega_n) + (\bar{X}_u'\rho_u - \bar{X}_n'\rho_n),$$

where $\bar{X}_l = \mathrm{E}_l[X_i|D_{1i} > D_{0i}]$ (complier means in $l = u, n$).

Effect decomposition

How much of the urban/non-urban gap in *effect sizes* is due to urban/non-urban differences in *student characteristics*?

- Now, $\rho_u = \rho_n + (\rho_u - \rho_n)$ and $\bar{X}_u = \bar{X}_n + (\bar{X}_u - \bar{X}_n)$.
- By substitution,

$$\begin{aligned}\bar{X}_u'\rho_u &= [\bar{X}_n + (\bar{X}_u - \bar{X}_n)]'[\rho_n + (\rho_u - \rho_n)]. \\ &= \bar{X}_n'\rho_n + \bar{X}_n'(\rho_u - \rho_n) + (\bar{X}_u - \bar{X}_n)'\rho_n \\ &\quad + (\bar{X}_u - \bar{X}_n)'\rho_u - (\bar{X}_u - \bar{X}_n)'\rho_n,\end{aligned}$$

- And so,

$$\tau_u - \tau_n = \underbrace{(\omega_u - \omega_n)}_{A} + \underbrace{\bar{X}_n'(\rho_u - \rho_n)}_{B} + \underbrace{(\bar{X}_u - \bar{X}_n)'\rho_u}_{C}$$

Effect decomposition

- $$\tau_u - \tau_n = \underbrace{(\omega_u - \omega_n)}_{A} + \underbrace{\bar{X}_n'(\rho_u - \rho_n)}_{B} + \underbrace{(\bar{X}_u - \bar{X}_n)'\rho_u}_{C}$$
- $A =$ unexplained differences in charter effectiveness.
- $B =$ differences in effectiveness due to urban students *responding or being treated differently* than non-urban students. "If non-urban students were treated as if they were in an urban school, how would their results differ?"
- $C =$ differences in effectiveness due to *differences in measured characteristics* of urban students relative to non-urban students. "If urban schools were to serve the non-urban instead of urban population, how would the results differ?"
- Decomposition is with respect to urban schools (ρ_u).
- Can also do it with respect to non-urban schools.

Effect decomposition

- Estimation follows these steps:
- Estimate coefficients in interacted model via 2SLS.
- Estimate $\bar{X}_l$ via kappa-weighting.
- X_i includes sex, race, special ed. status, English proficiency, free lunch status, and baseline test scores.
- See Ding et al. (2017) for related methods and randomization-based tests.

Effect decomposition

Table 11: Decomposition of Urban Differences in Impact

School level	Subject		Urban vs. non-urban difference in TE (1)	Decomposition 1 (urban loading): Due to diffs in cov. levels (2) (C)	Decomposition 1 (urban loading): Due to diffs in cov-specific TE (3) (A+B)	Decomposition 2 (non-urban loading): Due to diffs in cov. levels (4)	Decomposition 2 (non-urban loading): Due to diffs in cov-specific TE (5)
Middle	ELA		0.403***	0.252***	0.151	0.057	0.345
			(0.079)	(0.086)	(0.104)	(0.399)	(0.436)
		N	4523				
	Math		0.675***	0.329***	0.346***	0.248	0.428
			(0.074)	(0.081)	(0.093)	(0.333)	(0.353)
		N	4521				

Notes: This table decomposes the difference between urban and non-urban charter treatment effects. Outcomes are test scores the year after the lottery. The treatment is a dummy for charter attendance. Column (1) shows the difference in urban vs. non-urban treatment effects, computed as described in the text. Columns (2) and (3) report the components of the urban/non-urban difference due to differences in covariate levels and differences in covariate-specific effects, respectively, weighting the difference in covariate levels by the urban treatment effects. Columns (4) and (5) report a decomposition that weights the difference in covariate levels by the non-urban treatment effects. The covariates used in the decompositions are race, sex, special education, limited English proficiency, free/reduced price lunch, and baseline score categories (advanced, proficient, needs improvement, warning) in math and ELA.
*significant at 10%; **significant at 5%; ***significant at 1%

Meta-regression

How much of the urban/non-urban differential is due to differences in school characteristics?

- Angrist et al. expand the dataset to non-lottery schools.
- Pair-match each charter school student to a non-charter school student. Estimate charter-school-specific effects, τ_s:

$$\tau_s = \mathrm{E}_s[Y_{1i} - Y_{0i}],$$

  where $\mathrm{E}_s[Y_{0i}]$ given by matched pair outcomes.
- Then regress these *effect estimates* on school characteristics:

$$\hat{\tau}_s = \phi_0 + \phi_1 U_s + \phi_2 L_s + \phi_3 H_s + P_s' \phi_4 + u_s, \tag{15}$$

- First estimate baseline model with urban, lottery, and high school dummies.
- Introduce school-level policy variables.
- Interpret with caution (remember last week...)

Meta-regression

Table 13: Effects of School Characteristics

Variable	Math (1)	Math (2)	Math (3)	Math (4)	ELA (5)	ELA (6)	ELA (7)	ELA (8)
Urban	0.198***	0.072	0.008	-0.041	0.120***	0.062*	0.011	0.014
	(0.057)	(0.082)	(0.062)	(0.053)	(0.036)	(0.037)	(0.033)	(0.042)
Total minutes per day/100	-	0.154*		0.095	-	0.080*	-	0.055
		(0.090)		(0.078)		(0.042)		(0.038)
Minutes in relevant subject/100	-	0.203	-	0.207	-	0.023	-	0.007
		(0.211)		(0.168)		(0.075)		(0.068)
Per-pupil expenditure/1000	-	-0.002	-	-0.009	-	0.004	-	-0.001
		(0.014)		(0.010)		(0.008)		(0.009)
School is No Excuses	-	-	0.306***	0.231***	-	-	0.169***	0.117**
			(0.082)	(0.060)			(0.045)	(0.048)
Lottery	0.154**	0.086*	0.051	0.038	0.101**	0.055	0.047	0.033
	(0.069)	(0.050)	(0.052)	(0.041)	(0.043)	(0.035)	(0.036)	(0.033)
High School	0.039	0.078	0.035	0.087	0.069*	0.076*	0.062*	0.078*
	(0.071)	(0.065)	(0.052)	(0.057)	(0.036)	(0.040)	(0.032)	(0.040)
Constant	-0.131*	-0.835**	-0.064	-0.490	-0.085*	-0.445**	-0.047	-0.267
	(0.067)	(0.375)	(0.043)	(0.299)	(0.045)	(0.176)	(0.033)	(0.183)
N	30	28	30	28	30	28	30	28

Notes: This table reports regressions of school-specific treatment effects on school characteristics. The sample includes only schools that completed the charter survey. Regressions weight by the inverse of the standard error of the coefficient estimates and cluster at the school level.

*significant at 10%; **significant at 5%; ***significant at 1%

Unpacking effect heterogeneity

- Thus, urban charter schools have a much larger impact than non-urban charter schools.
- The nature of the urban charter schools impact is to bring students who would otherwise perform very poorly up toward the average performance of students elsewhere.
- Thus, much of this effect differential can be associated with student demographics,
- Differences in disciplinary practices also seem to play an important role.

Estimation of Heterogeneous Treatment Effects from Randomized Experiments, with Application to the Optimal Planning of the Get-Out-the-Vote Campaign

Kosuke Imai
Department of Politics, Princeton University, Princeton, NJ 08544
e-mail: kimai@princeton.edu (corresponding author)

Aaron Strauss
The Mellman Group, 1023 31st St NW, 5th Floor, Washington, DC 20007
e-mail: aaronbs@gmail.com

Optimal treatment regimes

- Field experimental studies show modest effectiveness of various get-out-the-vote strategies.
- A campaign manager with a limited budget may want to know how to get the most out of these strategies.
- Imai & Strauss address this problem.

Optimal treatment regimes

Strategy:

- Characterize effect heterogeneity agnostically but not statistically recklessly.
- Determine for whom treatment is most effective.
- Use that to determine optimal treatment regime for situations when you can't treat everyone.

Optimal treatment regimes

- Simple example: Text-message mobilization. If we can only send a limited number of messages, to whom should we send so as to maximize impact on voter turnout?
- (Paper explains how to generalize to multiple treatment options and partisan campaigns,)

Optimal treatment regimes

Formal set-up (simplified relative to paper):

- Population of size N.
- We have at our disposal information on covariates, X_i.
- $\delta(x)$ is probability that people with $X_i = x$ are assigned to treatment ($1-\delta(x)$ is prob. to control). Choice variable.
- $\rho(t,x)$ is probability that people with $X_i = x$ vote when assigned to $t = 0, 1$. Needs to be estimated.
- We want to maximize expected turnout:

$$g(\delta,\rho) = N\mathrm{E}_X\{[\delta(x)\rho(1,x) + (1-\delta(x))\rho(0,x)]\}$$

- Subject to budget constraint:

$$B \equiv 1(\max_{x\in\mathscr{X}}\{\delta(x)\} > 0)\kappa + N\mathrm{E}_X[\delta(x)\xi(x)] \leq C$$

where κ is a fixed cost, and $\xi(x)$ are incremental costs.

Optimal treatment regimes

- Imai & Strauss use a Bayesian decision framework. (See Athey & Wager, 2018, for a non-Bayesian approach that uses a "regret" loss criterion.)
- Select δ^*, accounting for uncertainty about $\rho(.)$:

  $$\delta^* = \arg\max_{\delta \in \Delta} \int g(\delta, \rho) d\pi(\rho|D) \text{ s.t. } B \leq C,$$

  where Δ are available treatment regimes and $\pi(\rho|D)$ is the estimated (posterior) distribution of $\rho(t,x)$ given the data (D) available.
- To compute a posterior distribution for $\rho(t,x)$, they use a standard beta-binomial (conjugate) model:
  - Beta distribution prior,
  - $\rho(t,x)$ point estimates using a tree classifier.

Optimal treatment regimes

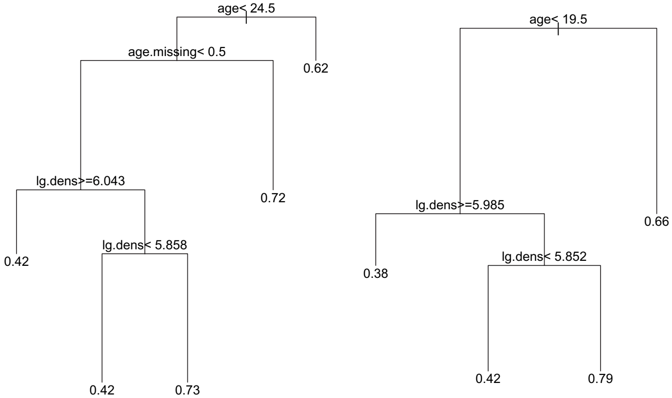


Fig. 2 Final classification trees for the control group (left panel) and treatment group (right panel). The complexity parameters are chosen from 10-fold cross validation using the algorithm described in Section 3.1 so that the resulting optimal turnout is maximized on the validation set. In this example, the planner budget allows treatment of at most 10% of the population. At each node, subjects who meet the node's criterion are filtered through the left branch of the tree. The leaves show the predicted probability of voting conditional on the leaves' covariates. Covariate abbreviations: age is the age in years of the subject, age.missing is whether the age of the participant is unknown, and lg.dens is the log of the subject's county population density.

Optimal treatment regimes

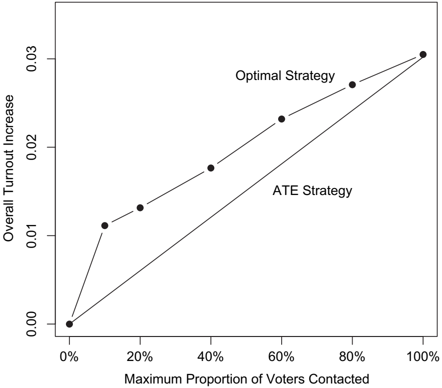


Fig. 3 Empirical evaluation of the performance of the proposed method for the text messaging experiment. The figure displays the overall turnout increase that results from two campaign strategies as a function of the maximum proportion of voters contacted. The first strategy is the ATE strategy (solid line), which contacts randomly selected voters. The second strategy is an optimal approach based on the methodology outlined in this paper, which uses covariate characteristics of voters to determine which voters receive the treatment. Solid circles represent the estimated optimal turnout using the difference-in-means estimator. The estimator is applied to test data that are not used for the derivation of the optimal strategy.

Optimal treatment regimes

- Imai & Strauss find substantial relative gains of targeting via this algorithm when budget constraints are very tight.
- Gains diminish as constraints loosen.
- They also present cases with multiple possible treatments and partisan campaigns.
- May be possible to improve on the selection of prescriptive covariates. Open question for future research.

Related approaches

- Manski (2004) provides more theory on welfare-maximizing treatment rules.
- Bhattacharya & Dupas (2012) analyze welfare-maximizing treatment allocation applied to the case of bed nets.
- Athey & Imbens (2015), Wager and Athey (2015), and Green & Kern (2012) examine tree-based methods for characterizing effect heterogeneity, Imai & Ratkovic (2012) use a LASSO-constrained support vector machine to hunt out interactions, and Grimmer et al. (2014) use a machine learning ensemble to estimate interaction effects.
- Athey & Wager (2017) develop methods for efficiently estimating optimal treatment regimes.

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:
  - Theories/explanations suggest predictions about types of units for which effects should be larger.

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:
  - Theories/explanations suggest predictions about types of units for which effects should be larger.
  - Exploratory analysis along the lines of Angrist et al.

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:
  - Theories/explanations suggest predictions about types of units for which effects should be larger.
  - Exploratory analysis along the lines of Angrist et al.
- Characterizing effect heterogeneity allows for *optimal treatment*:

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:
  - Theories/explanations suggest predictions about types of units for which effects should be larger.
  - Exploratory analysis along the lines of Angrist et al.
- Characterizing effect heterogeneity allows for *optimal treatment*:
  - Find out for whom effects are large and concentrate treatment on them.

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:
  - Theories/explanations suggest predictions about types of units for which effects should be larger.
  - Exploratory analysis along the lines of Angrist et al.
- Characterizing effect heterogeneity allows for *optimal treatment*:
  - Find out for whom effects are large and concentrate treatment on them.
  - Figure out whether a new person should be treated based on covariate profile.

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:
  - Theories/explanations suggest predictions about types of units for which effects should be larger.
  - Exploratory analysis along the lines of Angrist et al.
- Characterizing effect heterogeneity allows for *optimal treatment*:
  - Find out for whom effects are large and concentrate treatment on them.
  - Figure out whether a new person should be treated based on covariate profile.
- Some caveats:

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:
  - Theories/explanations suggest predictions about types of units for which effects should be larger.
  - Exploratory analysis along the lines of Angrist et al.
- Characterizing effect heterogeneity allows for *optimal treatment*:
  - Find out for whom effects are large and concentrate treatment on them.
  - Figure out whether a new person should be treated based on covariate profile.
- Some caveats:
  - Characterizing effect heterogeneity is predictive/descriptive. Randomization or natural experiments do not identify moderator effects (unless you also have identification for the moderator).

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:
  - Theories/explanations suggest predictions about types of units for which effects should be larger.
  - Exploratory analysis along the lines of Angrist et al.
- Characterizing effect heterogeneity allows for *optimal treatment*:
  - Find out for whom effects are large and concentrate treatment on them.
  - Figure out whether a new person should be treated based on covariate profile.
- Some caveats:
  - Characterizing effect heterogeneity is predictive/descriptive. Randomization or natural experiments do not identify moderator effects (unless you also have identification for the moderator).
  - Effect heterogeneity within an experiment is one thing, but experiment-level or context-level moderators require having experiments across contexts.

Discussion

- Characterizing effect heterogeneity aids in *interpreting* ATEs:
  - Theories/explanations suggest predictions about types of units for which effects should be larger.
  - Exploratory analysis along the lines of Angrist et al.
- Characterizing effect heterogeneity allows for *optimal treatment*:
  - Find out for whom effects are large and concentrate treatment on them.
  - Figure out whether a new person should be treated based on covariate profile.
- Some caveats:
  - Characterizing effect heterogeneity is predictive/descriptive. Randomization or natural experiments do not identify moderator effects (unless you also have identification for the moderator).
  - Effect heterogeneity within an experiment is one thing, but experiment-level or context-level moderators require having experiments across contexts.