Machine Learning for Economists
(and Business Analysts)

Optimal Policy Learning

Anthony Strittmatter

Literature

- Athey and Wager (2018): "Efficient Policy Learning", download.
- Kitagawa and Tetenov (2018): "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice", Econometrica, 86(2), pp. 591-616, download.

What are Policy Rules?

- Policy rules determine the allocation of treatments to individuals based on observable covariates.
- Policy rules are often labelled assignment rules, individualized treatment rules (ITR), personalized treatment rules, etc.

$\Rightarrow$ Policy rules can potentially improve the allocation of limited resources.

Scope of Applications

Targeted allocation of publicly sponsored programs:

- Assignment of unemployed to training programs.
- Preventive medical counselling (e.g. midwife or nutritionist).

$\Rightarrow$ Increase the effectiveness of publicly sponsored programs.

Targeted campaigns to foster public opinions/behavior:

- Get-out-the-vote campaigns.
- Information campaigns for organ donations.
- Marketing of charitable organisations.

$\Rightarrow$ Reduce the costs of the campaign.
$\Rightarrow$ Improve the desired response to the campaign.

Solicitation Letters

She's known nothing but abject poverty her entire life. Why on earth should Sebastiana have hope now? After forty-two years of toil in the unforgiving land of the high Andes, Sebastiana looks much older than her years. She has borne nine children and is alone to care for them after losing her husband six years ago. But a few months ago, Sebastiana joined a women's group sponsored by Freedom from Hunger. There she received a loan of $64 and training on how to grow her small, home-based business.

Treatment Definition

Treatment (efficacy story):
But does she really have a right to hope for something different? According to studies on our programs in Peru that used rigorous scientific methodologies, women who have received both loans and business education saw their profits grow, even when compared to women just received loans for their business. But the real difference comes when times are slow. The study showed that women in Freedom from Hunger's Credit with Education program kept their profits strong-ensuring that their families would not suffer, but thrive.

Control (emotional story):
But does she really have a right to hope for something different? Like Sophia and Carmen before her, the good news is, yes! Because of caring people like you, Freedom from Hunger was able to offer Sebastiana a self-help path toward achieving her dream of getting "a little land to farm" and pass down to her children. As Sebastiana's young son, Aurelio, runs up to hug her, she says, "I do whatever I can for my children."

Potential Effects

How to increase fundraising for Freedom for Hunger's Credit with Education program?

- **Altruistic donation motive:**
  Altruistic donors gain utility through the increased social welfare generated by the donation
- **Alternative donation motives:**
  Warm-glow, social norms, social pressure, own benefit, casually, participation (e.g., DellaVigna, List, and Malmendier, 2012)
- Mixtures between different motives possible. I will not distinguish sharply between different donation motives.

Potential Outcome Framework

- Treatment dummy:

$$D_i = \begin{cases} 1 & \text{mailer with efficacy story,} \\ -1 & \text{mailer with emotional story.} \end{cases}$$

- $Y_i(1)$ potential donations under the efficacy story.
- $Y_i(-1)$ potential donations under the emotional story.

Observed Donations:
In the absence of spillover effects,

$$Y_i = Y_i(-1) + \frac{1 + D_i}{2} \left(Y_i(1) - Y_i(-1) \right).$$

Fundamental Identification Problem

Individual Causal Effects:

$$\delta_i = Y_i(1) - Y_i(-1).$$

Optimal Policy Rule:

$$\pi_i^* = 1\{\delta_i > 0\} - 1\{\delta_i \leq 0\}.$$

- Mailer with efficacy story if $\delta_i > 0$.
- Mailer with emotional story if $\delta_i \leq 0$.

$\Rightarrow$ **Infeasible to identify and estimate individual causal effects!**

Objective Function

- The estimated policy rule should maximize the expected utility of the policy
$$\hat{\pi}_i = \max_{\pi} E[Y_i(\pi_i)].$$
- Equivalent to minimizing the expected maximum regret
$$\hat{\pi}_i = \min_{\pi} R(\pi_i),$$
where the expected maximum regret
$$R(\pi_i) = E[Y_i(\pi_i^*) - Y_i(\pi_i)]$$
is the expected gap between the optimal and estimated policy.

$\Rightarrow$ Minimax regret criterion (Manski, 2004, Savage, 1951).

CATE Based Approaches

Conditional Average Treatment Effects (CATEs):

$$\delta(x) = E[\delta_i | X_i = x] = E[Y_i(1) - Y_i(-1) | X_i = x]$$

X_i contains exogenous pre-treatment covariates/features/attributes that are potentially responsible for effect heterogeneity.

Conventional Practice: Test based approach

$$\hat{\pi}_i = \hat{\pi}(X_i) = \begin{cases} 1 & \text{if } \hat{\delta}(X_i) \text{ significant positive,} \\ -1 & \text{otherwise.} \end{cases}$$

- Fixes the probability of type I errors (mistakenly rejecting the null).
- Unnecessarily many type II errors (mistakenly assuming the null).

$\Rightarrow$ Imbalance between type I and II errors.

Empirical Success Rule (Manski, 2004)

Empirical Success Rule:

$$\hat{\pi}_i = \hat{\pi}(X_i) = \begin{cases} 1 & \text{if } \hat{\delta}(X_i) > 0, \\ -1 & \text{if } \hat{\delta}(X_i) \leq 0. \end{cases}$$

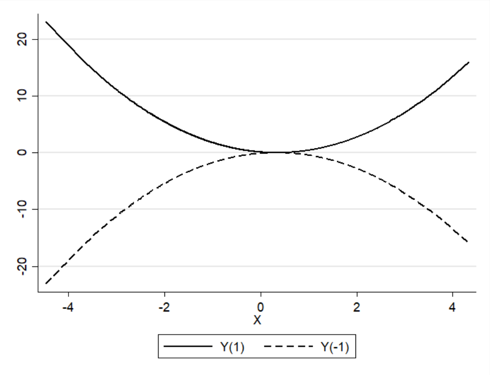
- Mailer with efficacy story if $\hat{\delta}(X_i) > 0$.
- Mailer with emotional story if $\hat{\delta}(X_i) \leq 0$.

Caveats:

- The selection of a policy rule is a classification problem.
- CATEs estimators are not targeted at this classification problem.

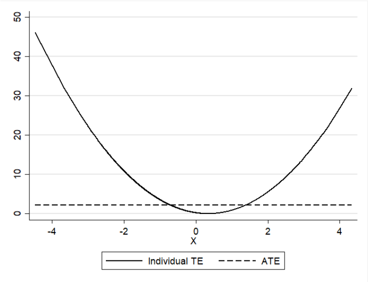
Simple Example

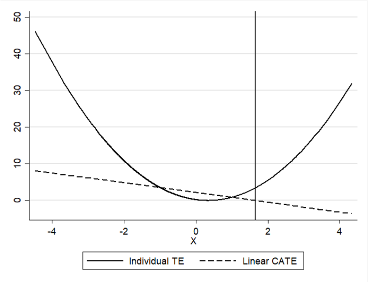
- $X \sim N(0,1)$
- $Y(1) = (X - 1/3)^2$
- $Y(-1) = -(X - 1/3)^2$



Reference: Qian and Murphy (2011)

CATEs Not Suited for Policy Rules





- Treating everybody is optimal
- ATEs find optimal policy rule ($MSE_{ATE} \approx 9.4$), even though linear prediction of CATEs approximate the individual treatment effects better ($MSE_{ATE} > MSE_{CATE} \approx 7.8$)

Policy Value Function

Define the policy value function of policy rule π_i:

$$\begin{aligned} Q(\pi_i) =& E[Y_i(\pi_i)] - \frac{1}{2} E[Y_i(1) + Y_i(-1)], \\ =& \frac{1}{2} E[\pi_i \cdot (Y_i(1) - Y_i(-1))], \\ =& \frac{1}{2} E[\pi_i \cdot \delta_i], \end{aligned}$$

where

- $E[Y_i(\pi_i)]$ is the expected utility under policy rule π_i and
- $\frac{1}{2} E[Y_i(1) + Y_i(-1)]$ is the expected utility under randomization.

Classification Problem

Estimate the policy $\hat{\pi}_i$ that maximizes

$$\hat{\pi}_i = \arg\max_{\pi} \frac{1}{2} E\left[\pi_i \cdot sign(\delta_i) \cdot |\delta_i|\right].$$

- Classification of $sign(\delta_i)$ with weights $|\delta_i|$.
- **Intuitively:**
  - Misclassifications hurts more when the (absolute) treatment effects are large.
  - Misclassifications of individuals with (almost) zero effects is not very costly.

Estimation Strategy

Apply sample analogy principle

$$\hat{\pi}(X_i) = \arg\max_{\pi(X_i)} \frac{1}{2 \cdot N} \sum_{i=1}^{N} \pi(X_i) \cdot sign(\hat{\Gamma}_i) \cdot |\hat{\Gamma}_i|$$

with

- $\hat{\Gamma}_i$ being an approximation score of δ_i and
- $\hat{\pi}(X_i)$ is the estimated policy rule based on X_i.

Augmented Inverse Probability Weighting (AIPW)

Efficient Score:

$$\hat{\Gamma}_i = \hat{\mu}_1(X_i) - \hat{\mu}_{-1}(X_i) + D_i \frac{Y_i - \hat{\mu}_{D_i}(X_i)}{\hat{p}_{D_i}(X_i)}$$

with the nuisance parameters

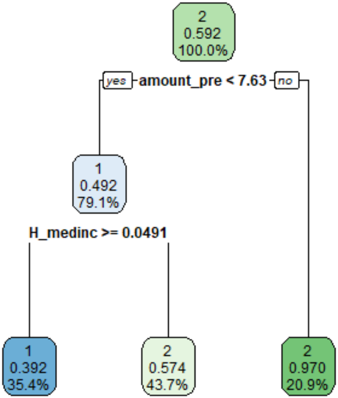
$$\hat{p}_1(x) = \hat{Pr}(D_i = 1|X_i = x) = 1 - \hat{p}_{-1}(x) \text{ and}$$
$$\hat{\mu}_d(x) = \hat{E}[Y_i(d)|X_i = x] = \hat{E}[Y_i|D_i = d, X_i = x].$$

- Nuisance parameters can be estimated with machine learning methods.
- Allows for high-dimensional confounding.
- Double robustness property.
- Cross-fitting to avoid overfitting.

Classification Methods

- **Classification Trees**
  - In contrast to regression trees, classification trees use different performance measures
  - These measures are targeted to minimise the impurity (instead of the regression fit)
  - Entropy or Gini index
- **Logistic LASSO**
- **Support Vector Machines**

Example of a Classification Tree



If the tree has bounded complexity:

- Estimated policy rule is consistent and semi-parametric efficiency.
- Maximum regret is bounded at attractive rates.

Policy Learning Algorithm

1. Partition the data into K-folds ($k = 1, ..., K$).
2. Repeat for all $k = 1, ..., K$ folds:
   2.1 Estimate $\hat{\mu}_{+1}^{-k}(X_i^{-k})$, $\hat{\mu}_{-1}^{-k}(X_i^{-k})$, and $\hat{p}_1^{-k}(X_i^{-k})$, which are the nuisance parameters in samples that disregards the kth-fold.
   2.2 Calculate the efficient score $\hat{\Gamma}_i^k$ for the kth-fold,

$$\hat{\Gamma}_i^k = \hat{\mu}_1^{-k}(X_i^k) - \hat{\mu}_{-1}^{-k}(X_i^k) + D_i^k \frac{Y_i^k - \hat{\mu}_{D_i}^{-k}(X_i^k)}{\hat{p}_{D_i}^{-k}(X_i^k)}.$$

   2.3 Estimate the probability

$$\hat{q}_i^k(X_i) = Pr(sign(\hat{\Gamma}_i^k) = 1 | X_i)$$

   using the weights $|\hat{\Gamma}_i^k|$.
   2.4 Implement the policy rule $\hat{\pi}^k(X_i) = 2 \cdot 1\{\hat{q}_i^k(X_i) > 0.5\} - 1$

Data

- In June 2007, Freedom from Hunger sent 11,259 mailers to previous donors who donated at least one gift since 2004.
- **Randomized Control Trial:**
  - Stratified randomization by most recent donation year and previous donation amount (above/below $100).
  - 5,631 mailers contained the insert with the efficacy story (treatment group).
  - 5,628 mailers contained the insert with the emotional story (control group).
- Experiment was repeated in October 2008 (3,173 mailers).

Variable Definitions

Outcomes:

- Donation amount in response to the mailer (in US-dollars).

Covariates:

- Amount given before mailer, amount of last gift before mailer, largest gift before mailer (in US-dollars).
- Date first donation, date last donation before mailer.
- # of gifts given in year prior to experiment, total # of gifts given, # of gifts per year.
- Median zip code income (in US-dollars).
- Average years of education in census tract.

Descriptive Statistics

	Mean	Std. Dev.	Min.	Max.
	(1)	(2)	(3)	(4)
Charitable Giving	10.87	209.59	0	20,251
# Donors	0.16	0.36	0	1
Total Previous Donations	286.87	1000.12	1	62,850
Last Previous Donation	44.86	150.39	1	10,000
Max. Previous Donation	60.22	252.52	1	15,000
# Gifts Last Year	0.76	1.19	0	21
Total # Gifts	6.81	11.09	1	245
# Gifts per Year	0.83	0.78	0.09	10.5
Median Income ZIP Region	51,816	23,957	0	200,001
Av. Years Educ. Cens. Reg.	13.98	2.99	0	21

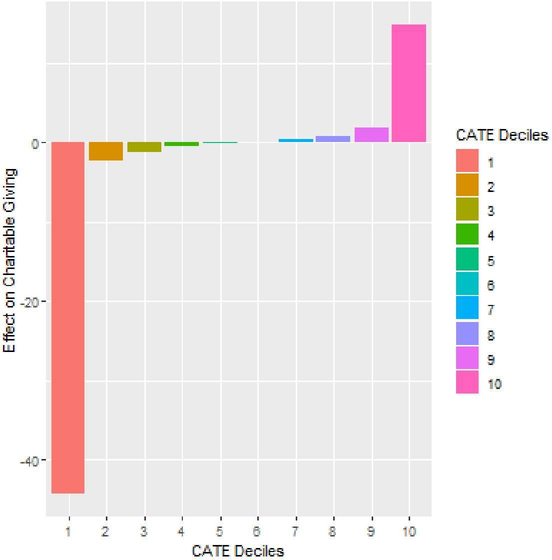
First Donation between May 5, 1955 and May 31, 2007
Last Donation between Jan 5, 2004 and Nov 15, 2007

Average Treatment Effects

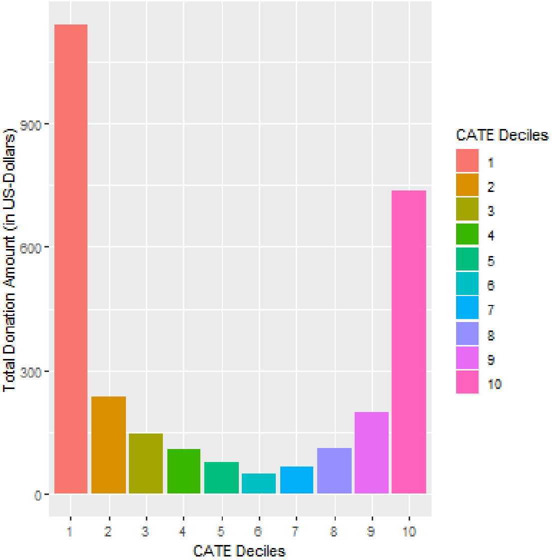
	Donation Amount (in US-dollars)	
	OLS	DML
	(1)	(2)
Efficacy Story	-1.52	-2.05
	(3.95)	(4.05)

- Total donations \$122,389 in response to the mailers.
- 1,814 donors (16%) give charity in response to the mailers.

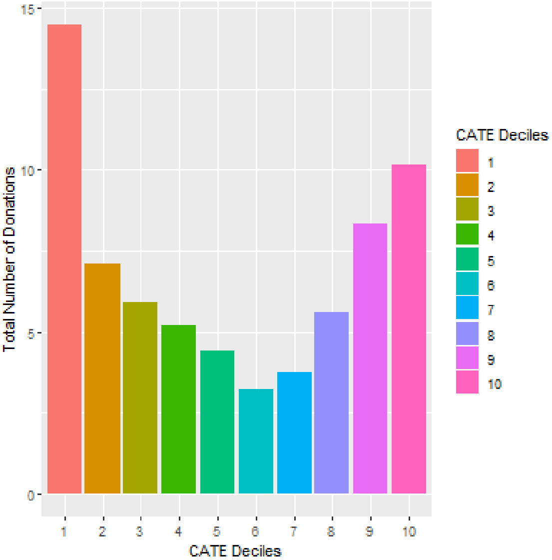
Deciles of Predicted CATEs



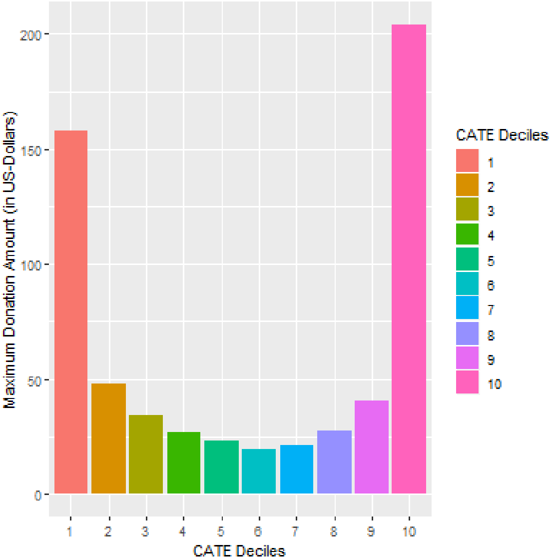
Previous Donation Amount



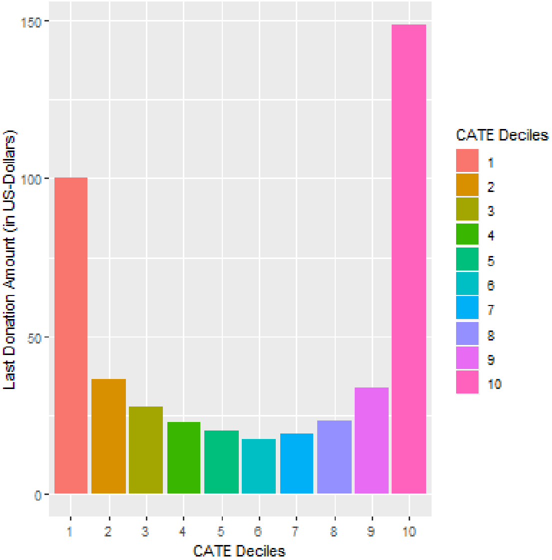
Previous Number of Gifts



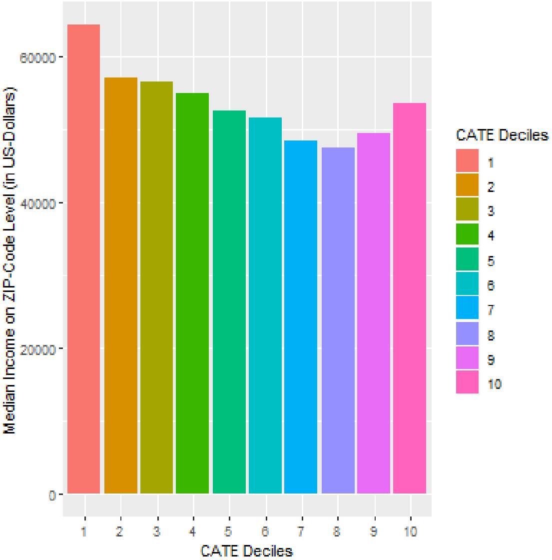
Previous Maximum Donation Amount



Amount Last Donation



Median ZIP-Code Level Income



Main Results In-Sample

	ES	Policy rule of classification tree with maximum depth					
	rule	2	4	6	9	12	15
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: In-Sample 2007			
Policy value	9.09***	3.21	3.23	3.48	4.71**	5.33**	5.52**
	(2.329)	(2.331)	(2.331)	(2.331)	(2.330)	(2.330)	(2.330)
Share efficacy story	0.50	0.001	0.01	0.01	0.02	0.03	0.03
Average number of TL		2.0	3.2	4.9	8.0	10.8	12.6
Min. number of TL		1	1	1	1	1	1
Max. number of TL		4	12	25	54	72	87

Notes: TL is the abbreviation for terminal leaves. ES is the abbreviation for empirical success. Standard errors are in parentheses.

Main Results Out-of-Sample

	ES	Policy rule of classification tree with maximum depth					
	rule	2	4	6	9	12	15
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel B: Out-of-Sample 2008			
Policy value	1.01	4.01*	4.24**	4.77**	4.56**	4.01*	4.12*
	(2.162)	(2.161)	(2.161)	(2.160)	(2.161)	(2.161)	(2.161)
Share efficacy story	0.47	0.01	0.02	0.02	0.03	0.05	0.05
Average number of TL		2.0	3.2	4.9	8.0	10.8	12.6
Min. number of TL		1	1	1	1	1	1
Max. number of TL		4	12	25	54	72	87

Notes: TL is the abbreviation for terminal leaves. ES is the abbreviation for empirical success. Standard errors are in parentheses.

Average donation in 2008 is $12.65.
Policy rules increase donation by more than 35%.

Budget constraints

- Substract cost (e.g., Kitagawa and Tetenov, 2018):

$$\hat{\Gamma}_i^{Budget} = \hat{\Gamma}_i - c_i$$

- Fix number of participants (e.g., Bhattacharya and Dupas, 2012):

$$\hat{\pi}_i^{Budget} = 2 \cdot 1\{\hat{\pi}(X_i) \geq \bar{\pi}\} - 1$$

Batch vs. Bandit Algorithms

Batch

- Historical dataset
- Potentially optimal policy rules change over time
- Then findings cannot be extrapolated to the future

Bandit

- Data arrives sequentially (typically online data)
- Treatment decisions are made sequentially
- Presence of exploration vs. exploitation trade-off
- Example: targeted online advertisements
- Reference: Dimakopoulou, Zhou, Athey, and Imbens (2018)

Further Extensions

- **Multiple treatments**
  (e.g., Frölich, 2008, Kallus, 2017, Zhou, Athey, Wager, 2018)
- **Ordered treatments**
  (e.g., Chen, Fu, He, Kosorok, and Liu, 2018)
- **Dynamic treatments**
  (e.g., Zhang and Zhang, 2018,
  Zhao, Zheng, Laber, and Kosorok, 2015)
- **Continuous treatments**
  (e.g., Chen, Zheng, and Kosorok, 2016, Athey and Wager, 2018)

Ethical Concerns?

- Statistical discrimination even if we omit critical variable (e.g., gender, migration, etc.)
- Examples: hiring decisions, flight prices, program assignments
- More or less than discrimination than humans?
- Targeting rules also have the potential to reduce discrimination, but it has to be used appropriately
- Current scandals: Cambridge Analytica, Amazons' unethical hiring algorithm