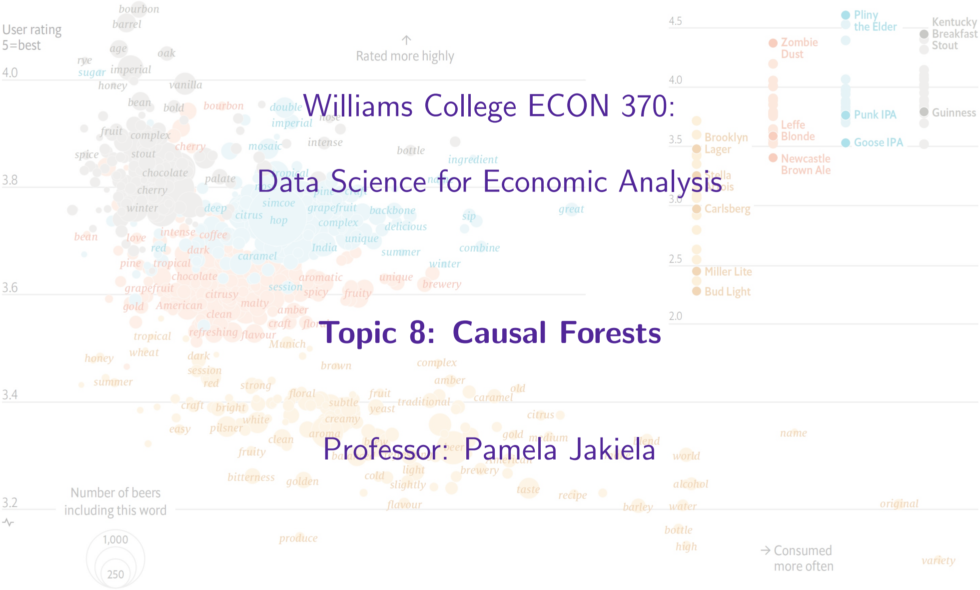


Williams College ECON 370:

Data Science for Economic Analysis

Topic 8: Causal Forests

Professor: Pamela Jakiela

Outline

- Review of the potential outcomes framework
- Causal forests
- Lab: buried treasure

The Causal Impact of Treatment

We are interested in the relationship between some "**treatment**" (e.g. going to the hospital) and some outcome that may be impacted by the treatment (eg. self-assessed health status)

Each individual is either treated or not:

- $W_i =$ is a **treatment dummy** equal to 1 if i is treated and 0 otherwise

Outcome of interest:

- $Y =$ outcome we are interested in studying (e.g. health status)
- $Y_i =$ value of outcome of interest *for individual i*

Potential Outcomes

For each individual, there are two **potential outcomes**:

- $Y_{0,i} = i$'s outcome if she **doesn't** receive treatment
- $Y_{1,i} = i$'s outcome if she **does** receive treatment

The **causal impact** of treatment on individual i is: $\tau_i = Y_{1,i} - Y_{0,i}$

- How much does treatment change outcome of interest for i?
- We are often interested in **average treatment effect** – average of τ_i across people
- We are also interested in understanding variation in τ_i across people and subgroups

Potential Outcomes: Example

Alejandro has a broken leg.

- $Y_{0,a}$ = If he doesn't go to the hospital, his leg won't heal properly
- $Y_{1,a}$ = If he goes to the hospital, his leg heals completely

Benicio doesn't have any broken bones. His health is fine.

- $Y_{0,b}$ = If he doesn't go to the hospital, his health is still fine
- $Y_{1,b}$ = If he goes to the hospital, his health is still fine

Potential Outcomes: Example

	Yes Hospital	**No Hospital**
Alejandro	$Y_{1,a}$	$Y_{0,a}$
Benicio	$Y_{1,b}$	$Y_{0,b}$

The Fundamental Problem of Causal Inference

The fundamental problem of causal inference:

We never observe both potential outcomes for the same individual

$\Rightarrow$ Creates a missing data problem whenever we try to compare treated to untreated

For any individual, we can only observe one potential outcome:

$$Y_i = \begin{cases} Y_{0,i} & \text{if } W_i = 0 \\ Y_{1,i} & \text{if } W_i = 1 \end{cases}$$

Potential outcomes without treatment (i.e. values of $Y_{0,i}$) may differ between those who choose to take-up treatment (Alejandro with a broken leg) and those who do not (healthy Benicio)

Average Treatment Effect (ATE)

The quantity of interest is the **average treatment effect** (ATE), or average causal effect, or intent-to-treat effect, or average impact, or treatment effect...

$$E[Y_{1,i} - Y_{0,i}|W_i = 1] = E[Y_{1,i}|W_i = 1] - E[Y_{0,i}|W_i = 1]$$

- ATE is average difference in potential outcomes (usually) across treated population
- Fundamental problem of causal inference: we never observe $Y_{0,i}$ for treatment group
  - $\bar{Y}_T$ is an unbiased estimator of $E[Y_i|W_i = 1] = E[Y_{1,i}|W_i = 1]$
  - We need an unbiased estimator of $E[Y_{0,i}|W_i = 1]$

Selection Bias

When we compare (many) participants to (many) non-participants:

$$E[\bar{Y}_T - \bar{Y}_C] = E[Y_i|W_i = 1] - E[Y_i|W_i = 0]$$

$$= E[Y_{1,i}|W_i = 1] - E[Y_{0,i}|W_i = 0]$$

Adding in $\underbrace{-E[Y_{0,i}|W_i = 1] + E[Y_{0,i}|W_i = 1]}_{=0}$, we get:

Difference in group means

$$= \underbrace{E[Y_{1,i}|W_i = 1] - E[Y_{0,i}|W_i = 1]}_{\text{average causal effect on participants}} + \underbrace{E[Y_{0,i}|W_i = 1] - E[Y_{0,i}|W_i = 0]}_{\text{selection bias}}$$

Random Assignment Eliminates Selection Bias

Experimental approach:

- **Random assignment to treatment:** eligibility for program is determined at random, e.g. via pulling names out of a hat, or using a computer pseudo-random number generator

When treatment status is randomly assigned,
treatment, control groups are random samples of a single population
(e.g. the population of all eligible applicants for the program)

$$\Rightarrow E[Y_{0,i}|W_i = 1] = E[Y_{0,i}|W_i = 0] = E[Y_{0,i}]$$

Expected outcomes are equal in the absence of the program

Random Assignment Eliminates Selection Bias

$\bar{Y}_T - \bar{Y}_C$ provides an unbiased estimate of the (casual) average treatment effect (or ATE):

$$= E[Y_i|W_i = 1] - E[Y_i|W_i = 0]$$

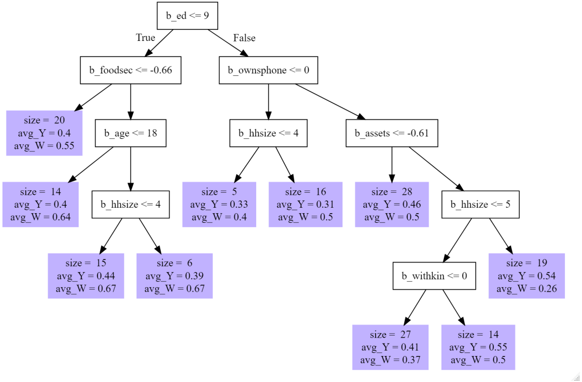
$$= E[Y_{1,i}|W_i = 1] - E[Y_{0,i}|W_i = 0]$$

$$= E[Y_{1,i}|W_i = 1] - E[Y_{0,i}|W_i = 1] + E[Y_{0,i}|W_i = 1] - E[Y_{0,i}|W_i = 0]$$

$$= \underbrace{E[Y_{1,i}|W_i = 1] - E[Y_{0,i}|W_i = 1]}_{\text{average treatment effect on participants}} + \underbrace{E[Y_{0,i}] - E[Y_{0,i}]}_{=0}$$

$$= \underbrace{E[Y_{1,i}] - E[Y_{0,i}]}_{\text{ATE}}$$

Causal Forests



A **causal forest** is a machine learning algorithm for identifying treatment effect heterogeneity (Athey and Imbens 2016, Davis and Heller 2017, Wager and Athey 2018)

Causal Forests Are Built from Regression Trees

- A **regression tree** is an algorithm that predicts Y_i through recursive binary splitting
- A **random forest** predicts Y_i by averaging many regression trees
- A **causal tree** is a regression tree where successive partitions are chosen to minimize expected mean squared error of estimated treatment effects (instead of MSE of Y)
  - Identifying sub-samples with less treatment effect heterogeneity
- A **causal forest** is a random forest of causal trees
- **Honest trees** use split sample methods for partitioning, estimation, inference
  - Allows for the construction of standard errors

Causal Forests in R and Python

- Causal forests are implemented as double ML (relevant when treatment is not random)
  - Use X variables to predict outcome Y and treatment W
  - Train the causal forest on residualized $\tilde{Y} = Y - \hat{Y}$ and $\tilde{W} = W - \hat{W}$
    - In R (`grf`), we can set the predicted values of to 0 or the mean
    - In Python (`econml`), we have to choose an ML model to predict Y, W
  - Tuning parameters are similar to regression trees and random forests
    - Number of trees, minimum leaf size, depth (Python)
    - Proportion treated within a leaf is also relevant with trees

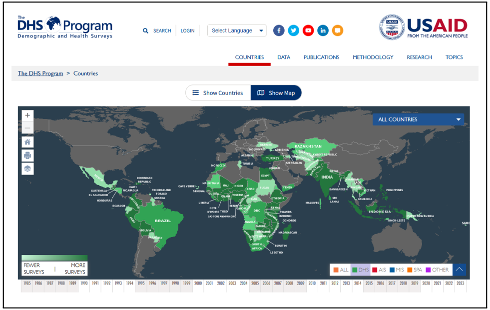
Interpreting Causal Forests

Causal forests (can) tell us two things:

- Is there (meaningful) treatment effect heterogeneity that is explained by observables?
  - Machine learning is noisy in small samples!
  - Algorithm looks for heterogeneity that is predicted by covariates
- Which baseline covariates predict the observed treatment effect heterogeneity?

Relevant outputs from causal forests:

- What are the observation-level out-of-bag (OOB) predicted conditional ATEs (CATEs)?
- Does the variation in OOB CATES explain variation in Y (test calibration in R)?
- Are estimated treatment effects higher for observations w/ high OOB CATEs?
- How often was a covariate was used to split tree?



Objective: find the treatment effect heterogeneity hidden in the DHS