

# MS&E 228: Discussion and FAQ

Vasilis Syrgkanis  
MS&E, Stanford

Max Schuessler 2d

Could we go through a summary of all the different settings/methods we have seen in class and make a tree-like map? I have seen a professor do this for an introductory class to ML in which she summarized in which settings which ML models work particularly well. To do this, she started off by splitting into regression versus classification first etc, then discussed settings such as high "p (predictors) low n (observations)" and the role of regularization. Perhaps something similar would be possible for causal inference and ML specifically, starting off with RCT versus non-RCT setting and then making branches going downstream. I think this could be incredibly helpful as a future reference.

♥ Reply Edit Delete ...

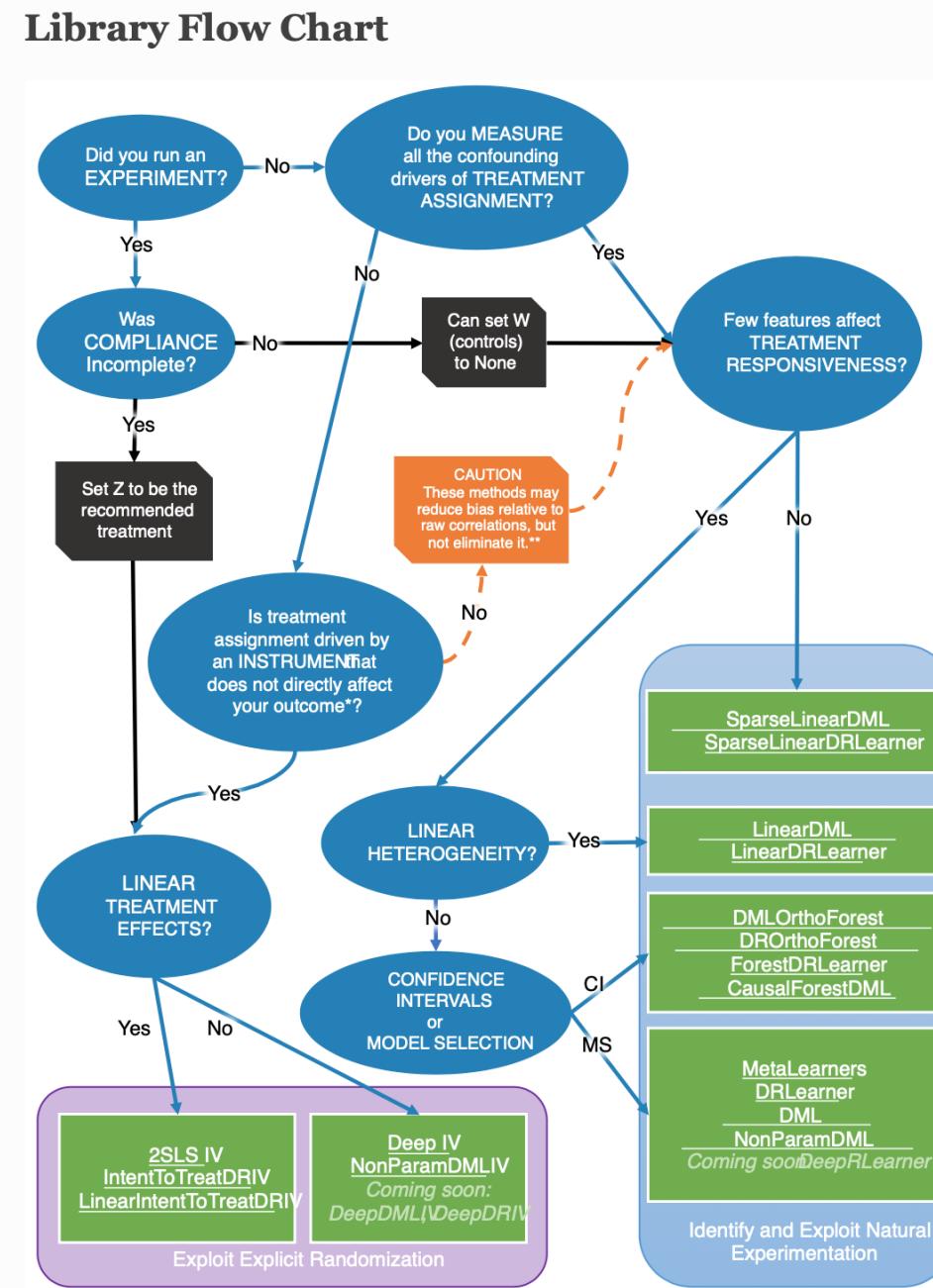
Daniel Jenson 2d

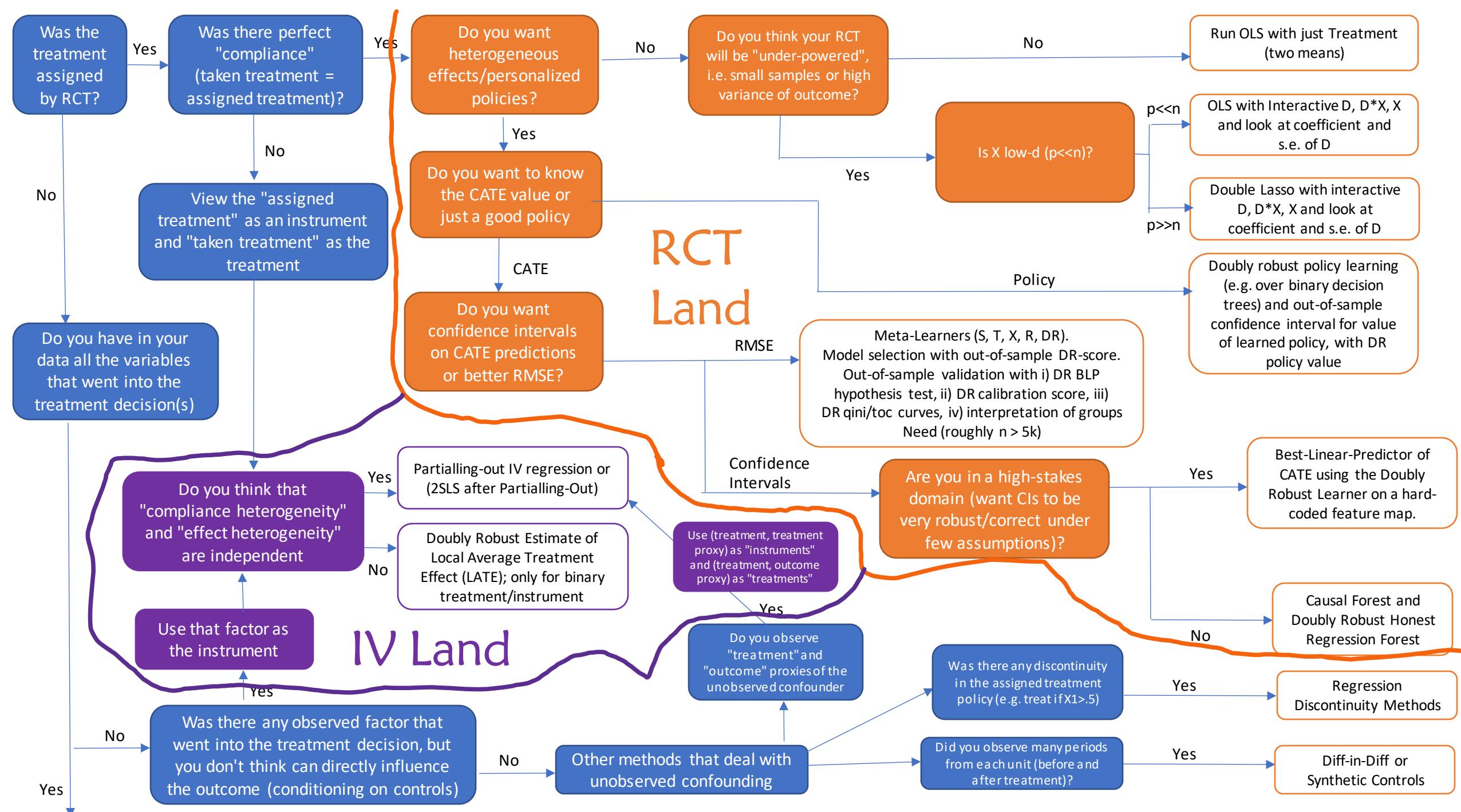
Can we build a mind map of the techniques we've used?

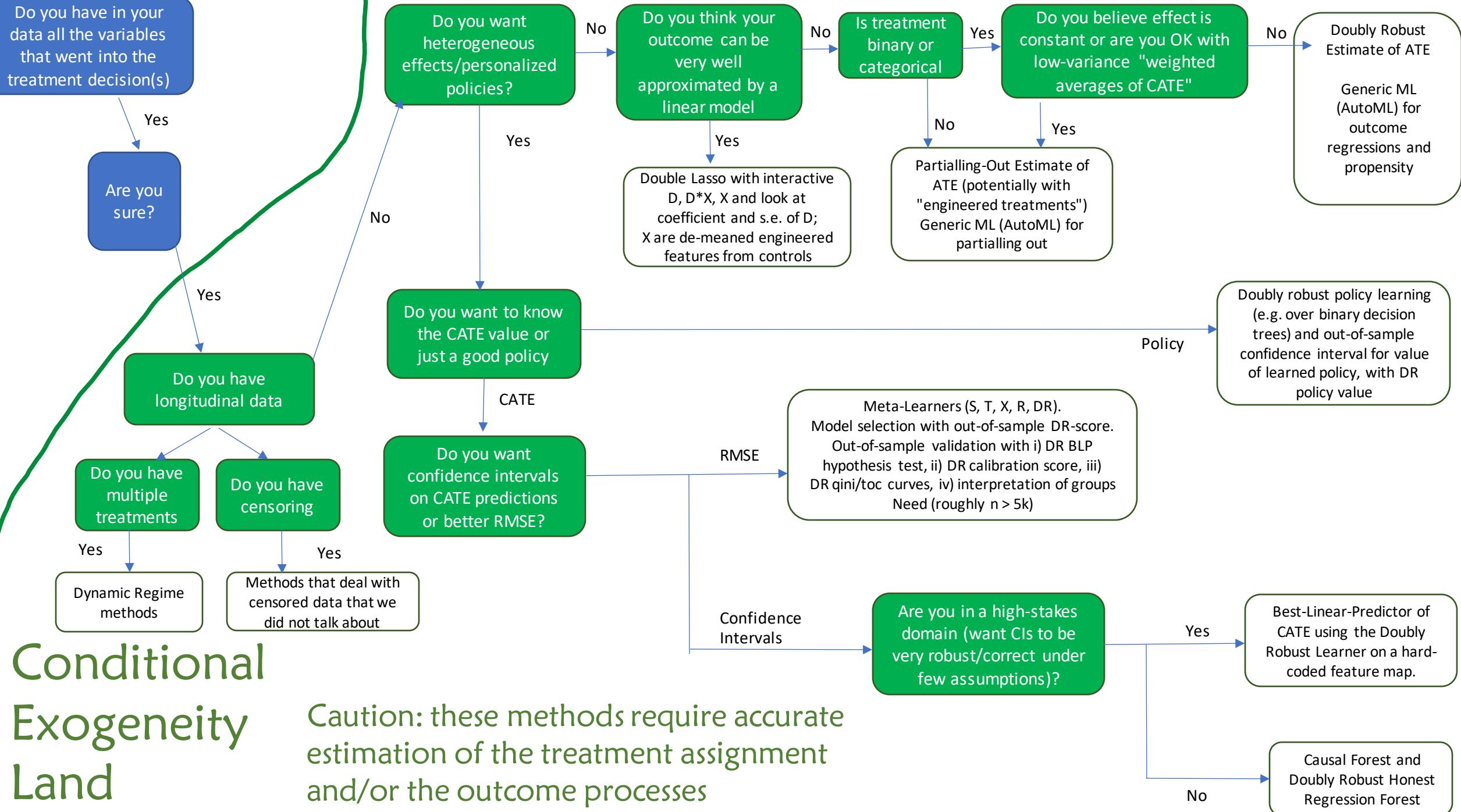
♥ 3 Reply Edit Delete ...

The screenshot shows the EconML 0.14.0 documentation homepage. The top navigation bar includes the EconML logo, version 0.14.0, and a search bar labeled "Search docs". The main content area is titled "EconML User Guide" and contains several sections:

- Overview**: Machine Learning Based Estimation of Heterogeneous Treatment Effects.
- Motivating Examples**: Introduction to Causal Inference.
- Problem Setup and API Design**.
- Library Flow Chart** (highlighted in grey):
  - Detailed estimator comparison
  - Estimation Methods under Unconfoundedness
  - Estimation Methods with Instruments
  - Estimation Methods for Dynamic Treatment Regimes
  - Inference
  - Interpretability
  - References
  - Frequently Asked Questions (FAQ)
- Public Module Reference**
- Private Module Reference**







Roberto Lobato Lopez 2d

I would like to know the intuition of what to do when we have don't have Stable Unit Value Assumptions or we have Spill over effects. For example, in networks.

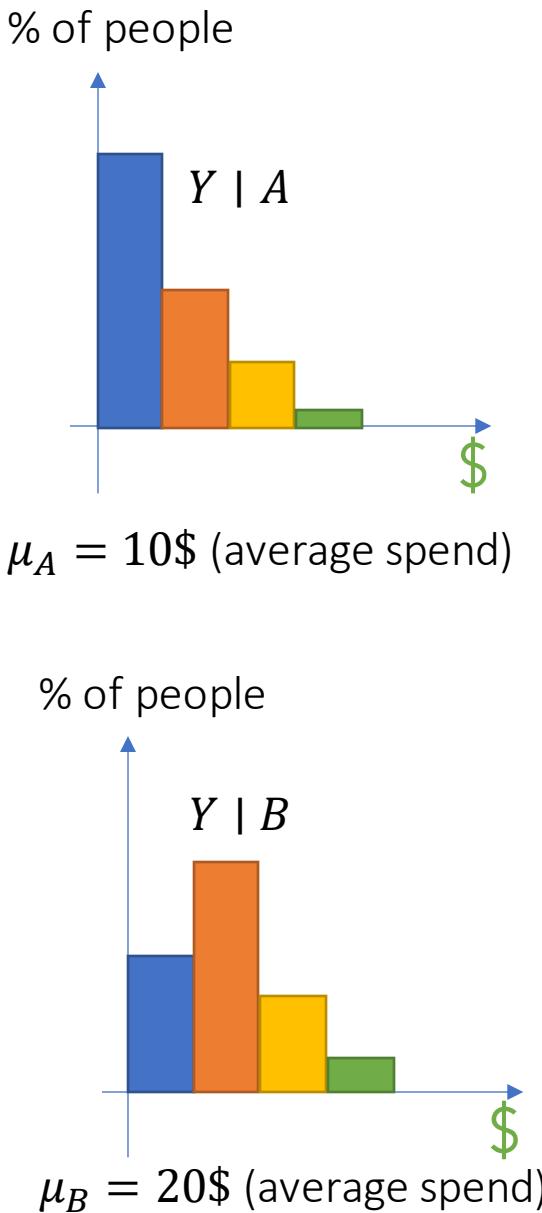
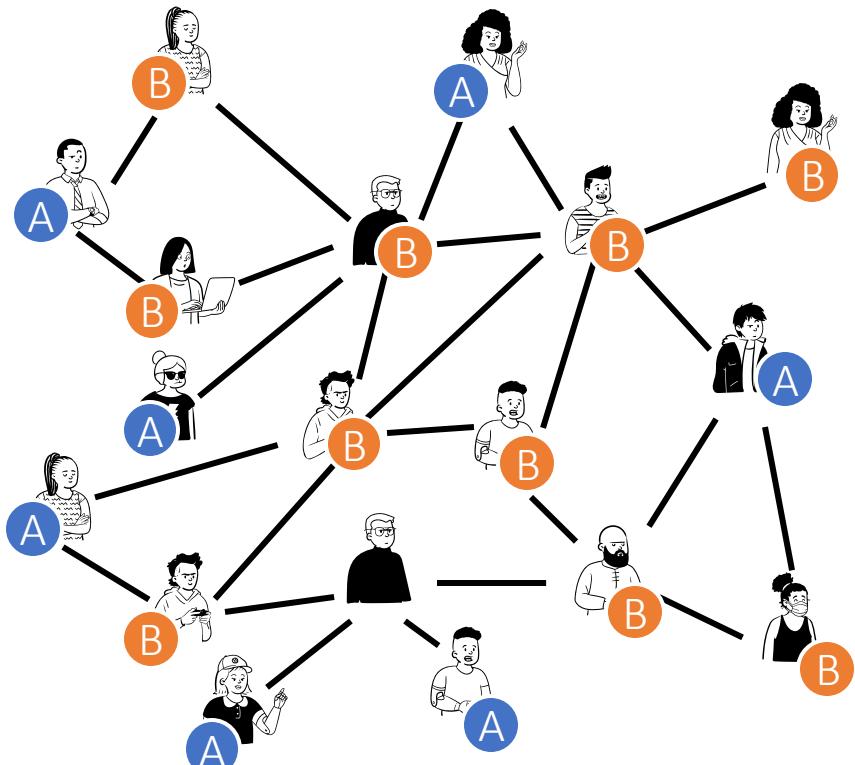
 Reply Edit Delete ...



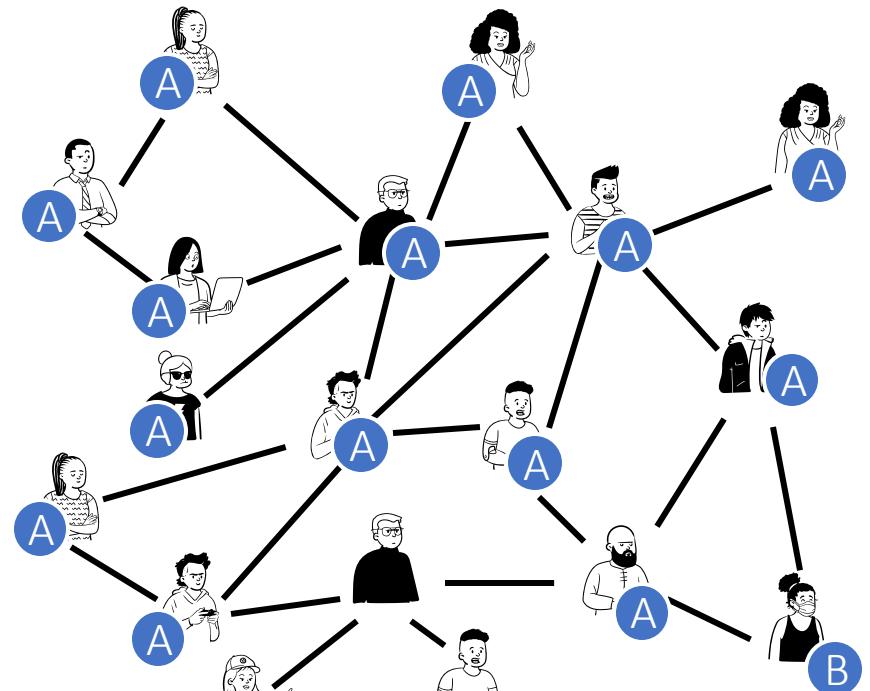
# Interference

- Social Network interference
- Equilibrium effects
- Stateful systems and time effects

# Social Network Interference



Counterfactual world

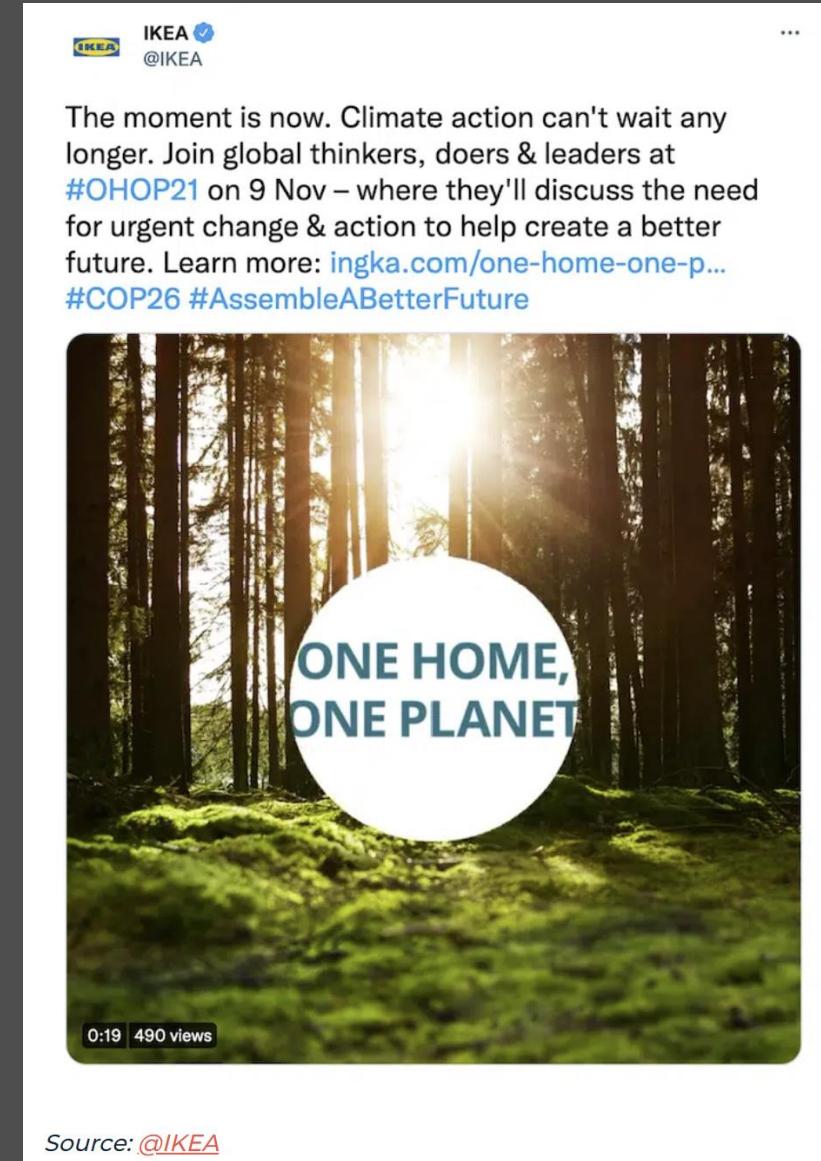
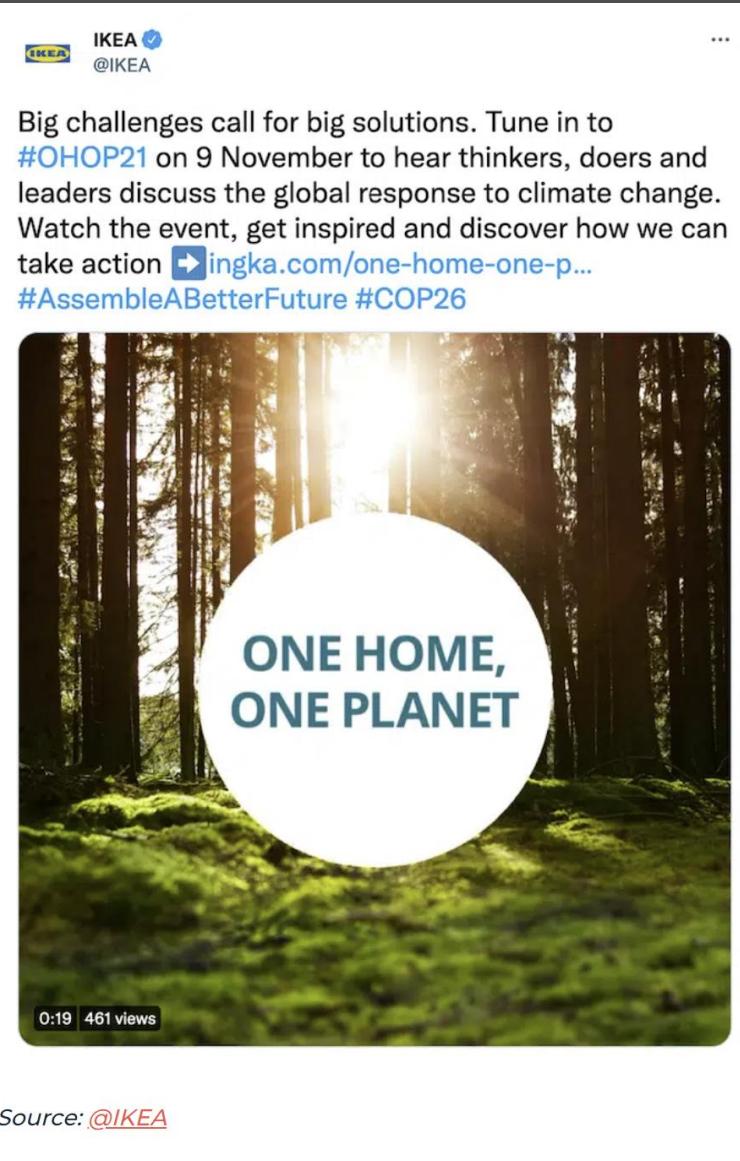


% of people

$Y | A$

$\mu_A = 30\$$  (average spend)

\$



Google wool socks

Web Shopping Images Videos Maps More Search tools

About 15,700,000 results (0.27 seconds)

**Wool Socks at Walmart**  
www.walmart.com/Underwear\_&\_Socks • 4.3 ★★★★★ rating for walmart.com  
Save On Wool Socks at Walmart. Free Shipping Site to Store.

**Wool Socks Superstore - SocksAddict.com**  
www.socksaddict.com/Wool-Socks • 4.5 ★★★★★ rating for socksaddict.com  
Get Free S&H + 99% Ship Same Day. SmartWool, Darn Tough & More Wool!  
Guaranteed lowest prices - 99% ship same day - 180 day return policy  
Brands: Outdoor Series, Running Series, Kids Series, Darn Tough, ...  
Injinji Wool Socks - Wigwam Wool Socks - Women's Wool Socks - Shop All Socks

**Official Site: SmartWool® Clothing & Socks**  
www.smartwool.com/ • SmartWool • Extraordinarily comfortable SmartWool® socks & apparel: High performance Merino hiking, skiing, outdoor sport, running, walking, cycling & daily clothing.  
Hike - Run - Baselayers - Socks

**Amazon.com: Merino Wool Blend Socks: Clothing**  
www.amazon.com/Gilbin... • 3 Pairs Merino Wool Blend  
Merino Wool Blend Socks, perfect for wearing with hiking, hunting, skiing, ...

**SmartWool Socks at Sierra Trading Post**  
www.sierratradingpost.com/smartwool-socks~bs~19440~231/ • Improving upon nature's finest insulator, combine the best Merino wool and the latest advanced technology and SmartWool is born. It is the ultimate wool for ...

**REI Lightweight Merino Wool Hiking Crew Socks at REI.com**  
www.rei.com/.../rei-lightweight-merino-wool-hiking-crew-socks • REI • ★★★★★ Rating: 5 - Review by stringbreaker - Mar 21, 2014 - \$12.50

**Shop for wool socks on Google**

Sponsored

Merino Wool Ragg Sock, 12" Two-Pack...  
\$22.95 L.L.Bean

SmartWool Men's New Classic Rib So...  
\$18.95 Socks Addict

carhartt men's all terrain wool blend s...  
\$10.50 Sears

Hue Cable Knit Merino Wool Blend...  
\$10.00 Bloomingdale's

Dickies Men's 1pk Merino Wool Sock...  
\$11.00 Target

Men's Over-the-calf Wool Rib Dress Sock...  
\$9.99 Lands' End

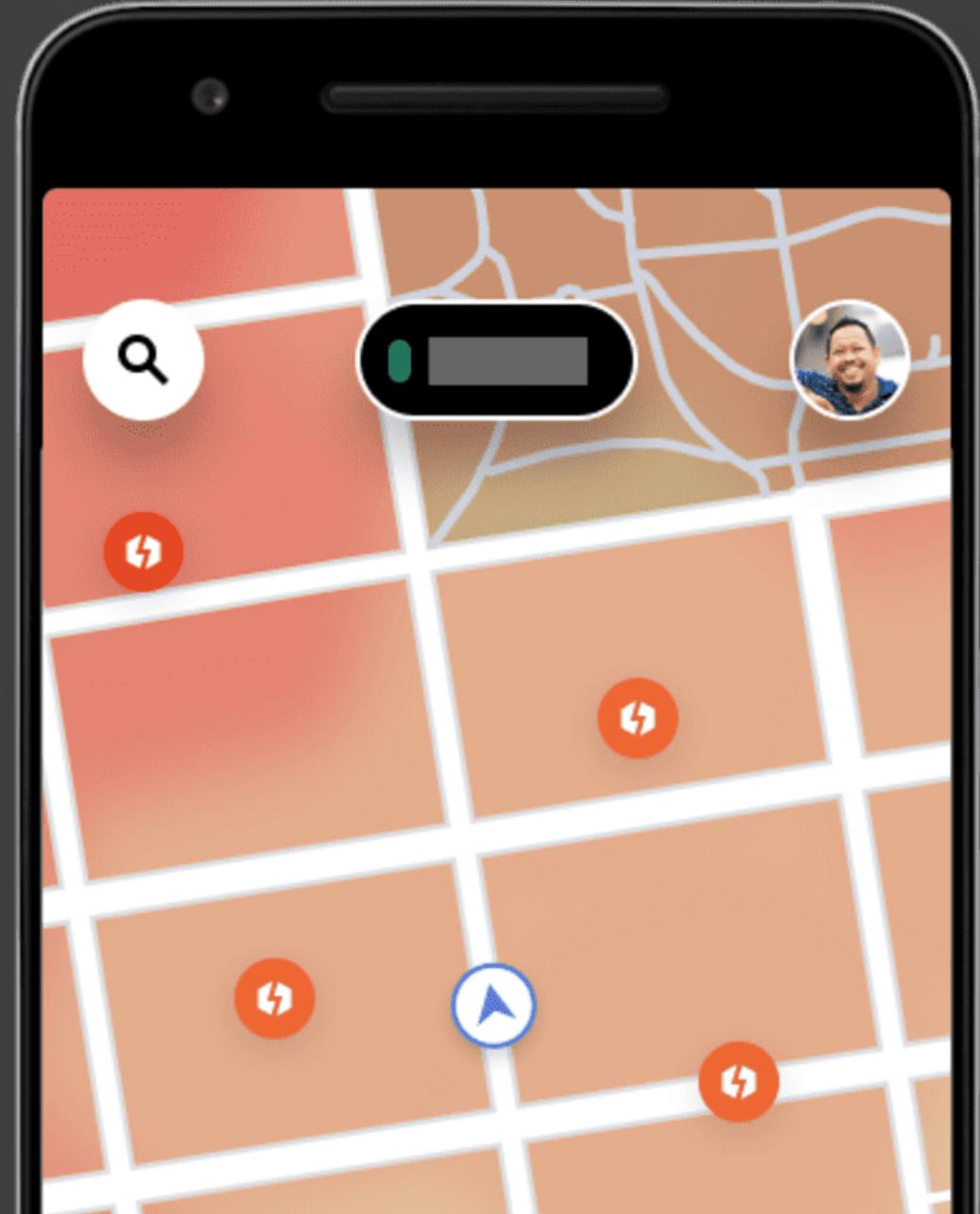
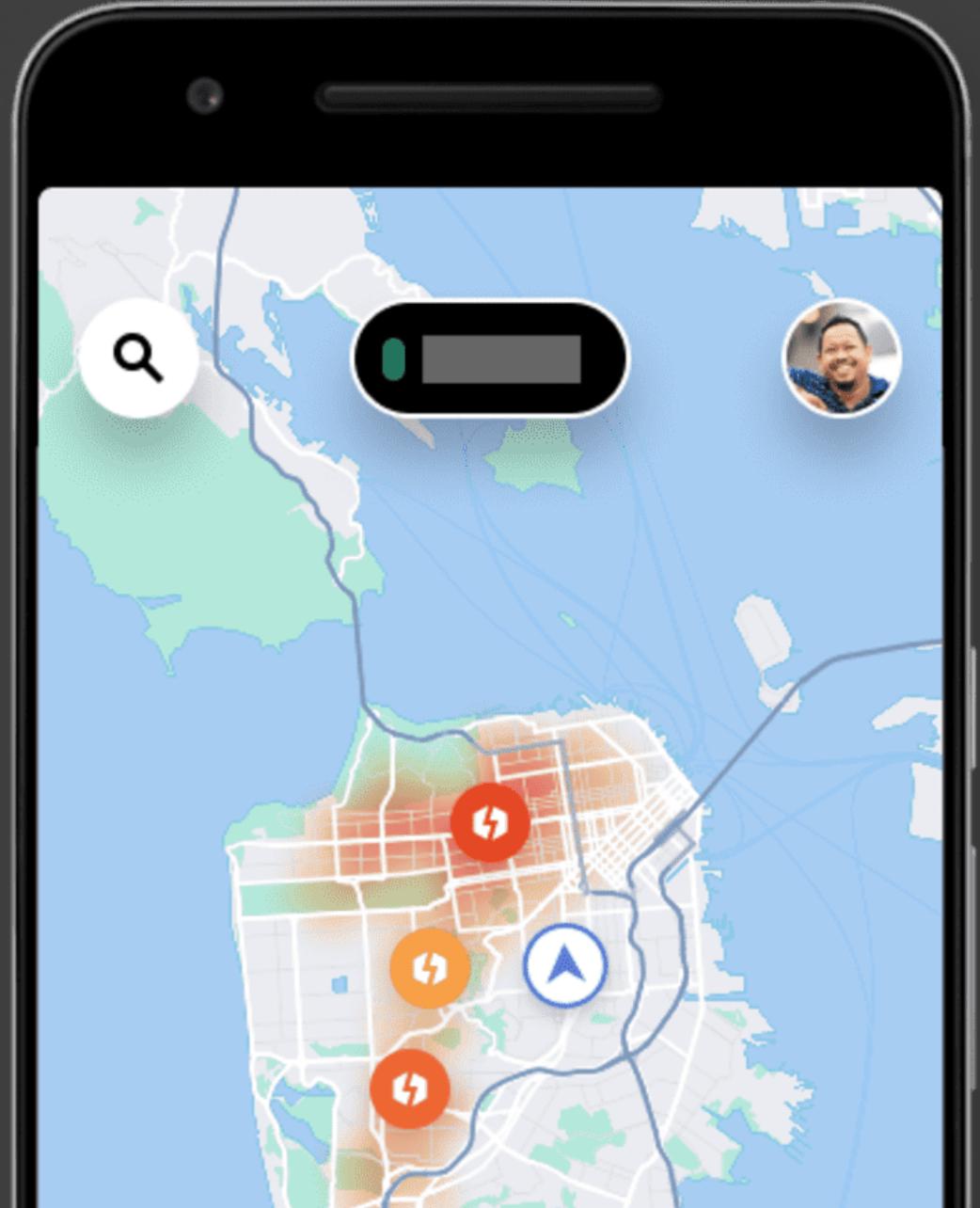
**Alpaca Therapeutic Socks**  
www.alpacasofmontana.com/ • Warm, Soft, Alpaca Diabetic Socks  
Non-restrictive promotes blood flow

**Women's SmartWool Socks**  
www.sierratradingpost.com/SmartWool • 4.6 ★★★★★ advertiser rating  
Great Selection of SmartWool Socks.  
Shop SmartWool Socks For Women Now.

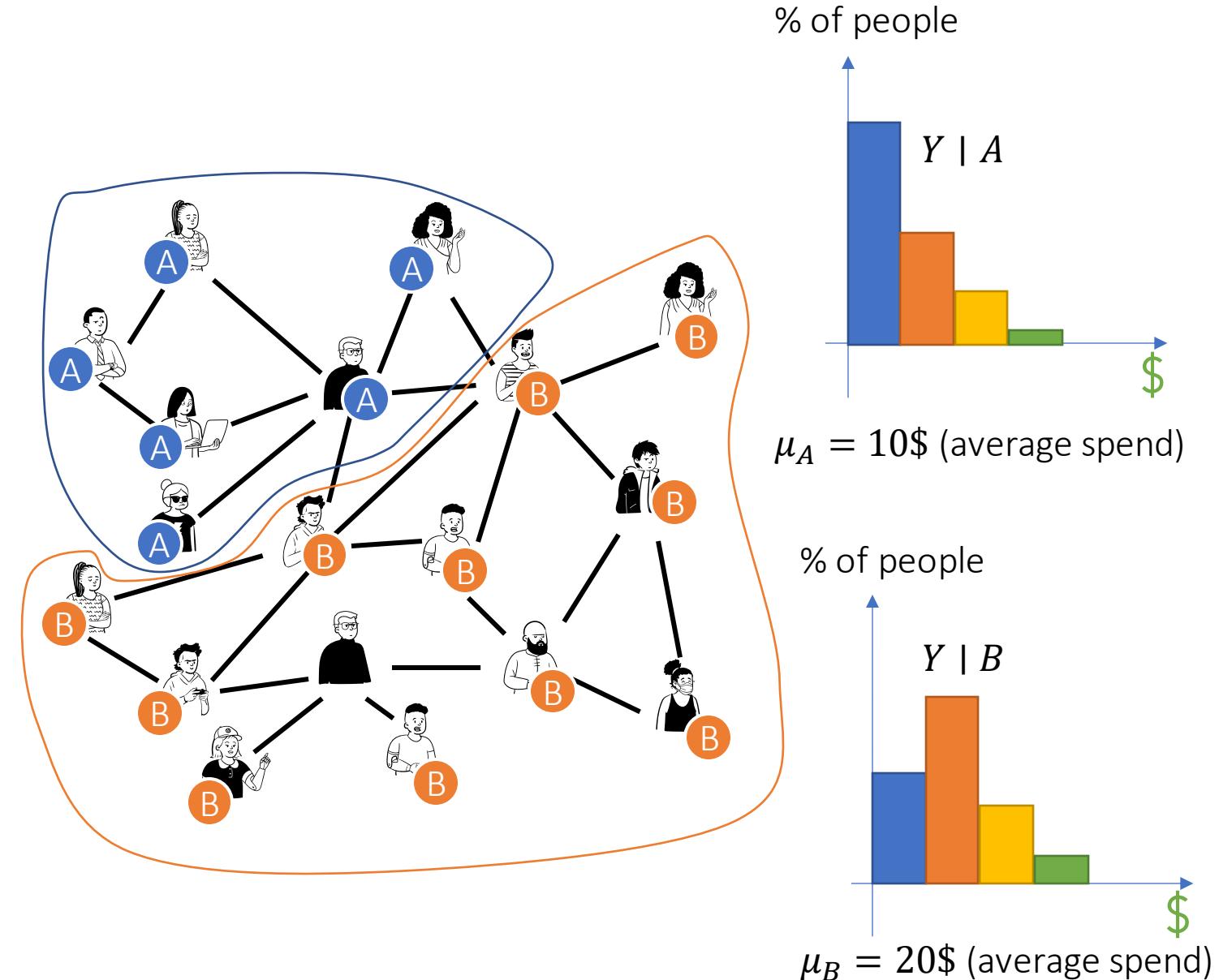
Search Network Ads

Shopping Ads

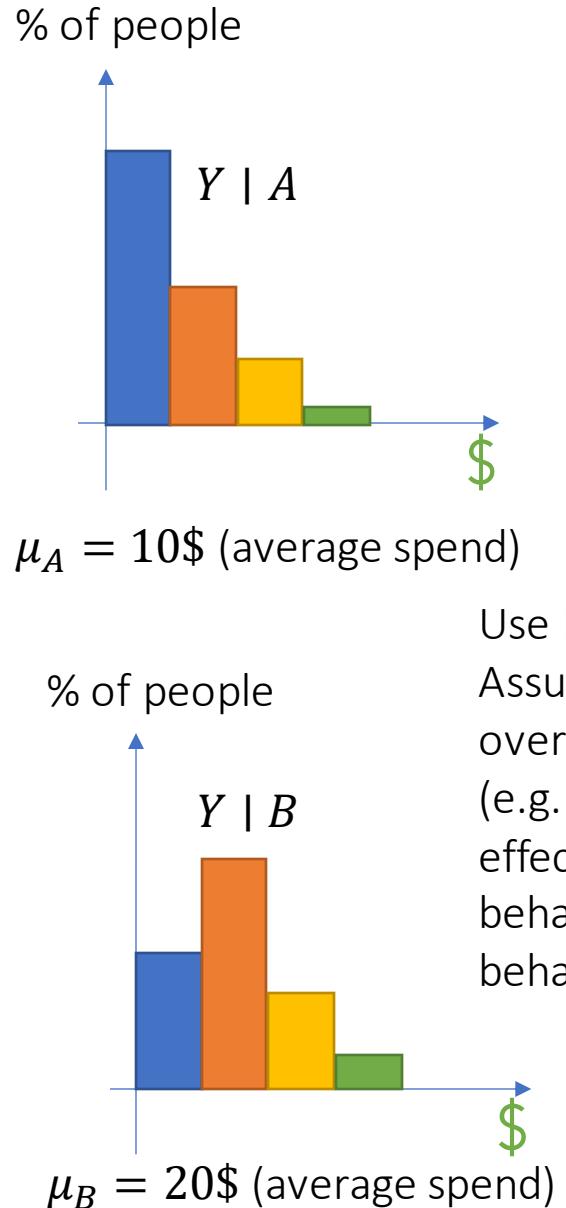
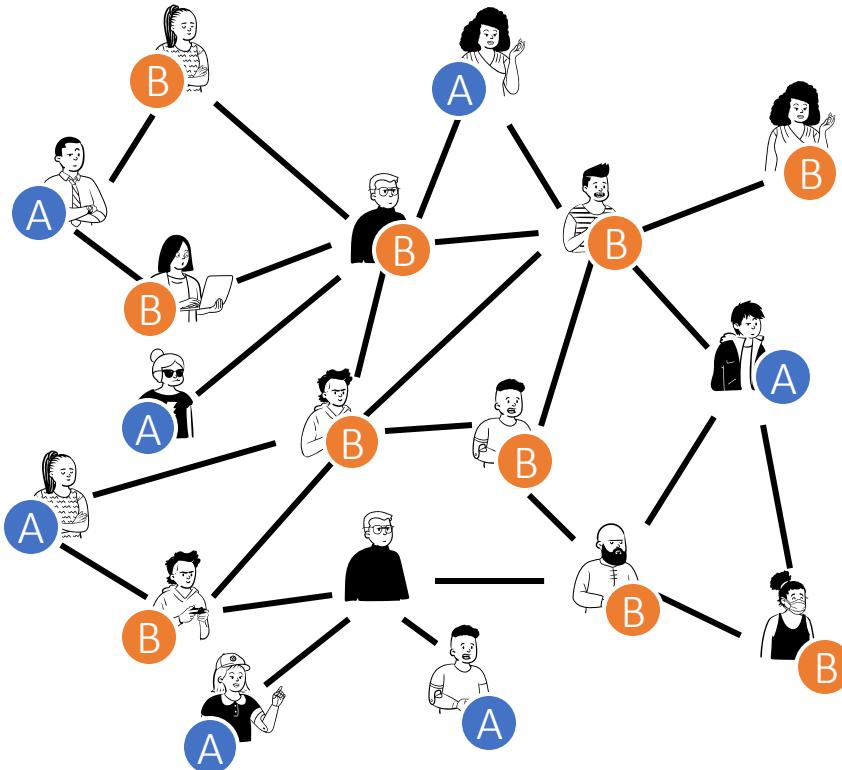
Image source: <https://googleadsstrategy.com/google-adwords-search-network-vs-display-network/>



# Approach: Clustering



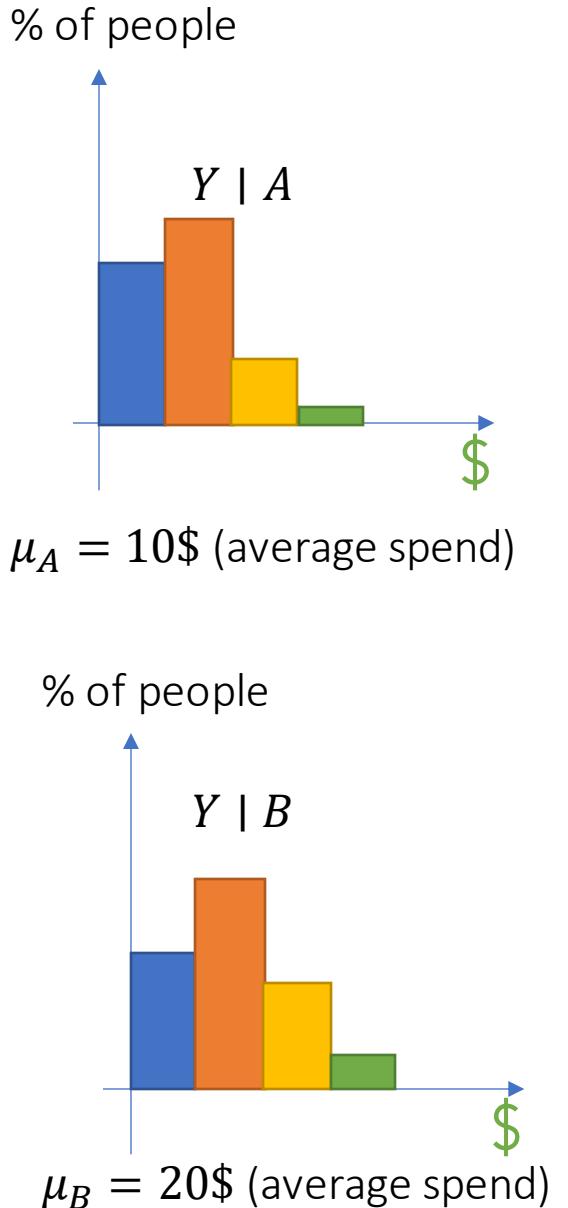
# Approach: Structural Bias Correction



Correct  
Spill-Over  
Bias



Use Network Information +  
Assumptions on how spill-  
overs change outcome  
(e.g. additive homophily  
effects, market equilibrium  
behavior, Nash equilibrium  
behavior)



**Yifan Shen** 2d

Can we briefly talk about some real-life examples of survival analysis in longitudinal studies and how can apply what we have learnt in class to it?

 2 Reply Edit Delete ...

# Survival Data

- If your outcome is survival then you need to adjust for censoring
- Easiest way to combine with what we've learned
  - Focus on un-censored samples
  - De-bias the "selection" by dividing the observed outcomes by the probability that censoring doesn't happen by the time of the observation  $\Pr(C > t | X)$
  - Can be estimated via reduction to many classification problems or via CDF estimators (Nelson-Aalen estimator; Survival Forests)
  - Then apply all that we learned on these "transformed outcomes"
- More tailored:
  - [Estimation of Heterogeneous Effects with Survival Outcomes](#)
  - [Meta Learners for Survival Data R-package](#)

# Survival Analysis and CATE

- Xu et al. "The Systolic Blood Pressure Intervention Trial (SPRINT) is a multicenter RCT that evaluated the effectiveness of an intensive blood pressure (BP) treatment goal (systolic BP < 120 mm Hg) as compared to the standard BP goal (< 140 mm Hg) on reducing risks of cardiovascular disease (CVD). SPRINT recruited 9,361 participants and found that the intensive treatment reduced the risk of fatal and nonfatal CVD and all-cause mortality for patients at high risk of CVD events [Wright et al., 2015]."

Roberto Lobato Lopez 2d

I would love to see some examples of RCT gone wrong. Like when researchers just assumed that because they could leverage enough datapoints, naive A/B testing was enough.

♡ 2 Reply Edit Delete ...

# Modern Challenges of A/B tests

- Interference
- Long range objectives
- Multiple exposures over time
- Non-compliance

Roberto Lobato Lopez 2d

I have seen some examples of synthetic controls (mainly in finance) that are constructed by weighting "similar" subjects. I'm a little bit skeptic on them. Any comments?

♥ Reply Edit Delete ...

## Synthetic A/B Testing using Synthetic Interventions

Anish Agarwal, Devavrat Shah, and Dennis Shen\*

### Abstract

Suppose there are  $N$  units and  $D$  interventions. We aim to learn the average potential outcome associated with every unit-intervention pair, i.e.,  $N \times D$  causal parameters. While running  $N \times D$  experiments is conceivable, it can be expensive or infeasible. This work introduces an experiment design, *synthetic A/B testing*, and the *synthetic interventions* (SI) estimator to recover all  $N \times D$  causal parameters while observing each unit under at most two interventions, independent of  $D$ . Under a novel tensor factor model for potential outcomes across units, measurements, and interventions, we establish the identification of each parameter. Further, we show the SI estimator is finite-sample consistent and asymptotically normal. Collectively, these also lead to novel results for panel data settings, particularly for synthetic controls. We empirically validate our experiment design using real e-commerce data from a large-scale A/B test.

Riley Juenemann 2d

Are there application domains where causal inference is not (often) used that you think could really benefit from incorporating some of these techniques?

 Reply Edit Delete ...

Operations Management and business decision making (non-digital)

Alex Desronvil 2d

Thank you -- two questions: (1) I was wondering if it would be possible to talk about good rules-of-thumb in selecting/pursuing a causal inference strategy and (2) I was wondering if you could offer some good directions to start with diff-in-diff, please.

Heart Reply Edit Delete ...

Diff-in-Diff requires observing two periods from each unit Y<sub>1</sub>, Y<sub>2</sub>

Y<sub>1</sub> pre-treatment Y<sub>2</sub> post-treatment

Assumes conditional exogeneity on time-difference in potential outcomes (aka parallel trends)

Y<sub>1(0)</sub> - Y<sub>2(0)</sub> independent of D given X => We can estimate E[Y<sub>1(0)</sub> - Y<sub>2(0)</sub> | D=1] (identification by conditioning)

We then observe

$$E[Y_2(1) - Y_2(0) | D=1] = E[Y_2(1) - Y_1(0) + Y_1(0) - Y_2(0) | D=1] = E[Y_2 - Y_1 | D=1] - E[Y_2(0) - Y_1(0) | D=1]$$

Equivalent to Average Treatment effect on the treated (ATT)! Can deploy doubly robust estimate for ATT.

## What's Trending in Difference-in-Differences?

### A Synthesis of the Recent Econometrics Literature\*

Jonathan Roth<sup>†</sup>

Pedro H. C. Sant'Anna<sup>‡</sup>

Alyssa Bilinski<sup>§</sup>

John Poe<sup>¶</sup>

December 26, 2022

### Structural Nested Mean Models Under Parallel Trends Assumptions

Zach Shahn<sup>1,2</sup>, Oliver Dukes<sup>3</sup>, David Richardson<sup>4</sup>, Eric Tchetgen Tchetgen<sup>3</sup>, and James Robins<sup>5</sup>

<sup>1</sup>CUNY School of Public Health, New York, NY, USA

<sup>2</sup>IBM Research, Yorktown Heights, NY, USA

<sup>3</sup>University of Pennsylvania, Philadelphia, PA, USA

<sup>4</sup>University of California Irvine, Irvine, CA, USA

<sup>5</sup>Harvard TH Chan School of Public Health, Boston, MA, USA

John Kohler 2d

Sometimes in econometrics, we transform our outcome variables in order to give special meaning to our parameter estimates (for example, taking the log of Y and X to calculate elasticities). Is machine learning capable of flexibly capturing these sorts of transformations in the output space? (we have only discussed input space transformation so far) Is there any connection to class material or is this a completely separate question of interpretation rather than estimation?

♥ Reply Edit Delete \*\*\*

Log-like? Identified ATEs defined with zero-valued outcomes are  
(arbitrarily) scale-dependent\*

Jiafeng Chen  
Harvard Business School  
Department of Economics, Harvard University

Jonathan Roth  
Department of Economics, Brown University

February 1, 2023

#### Abstract

Economists frequently estimate average treatment effects (ATEs) for transformations of the outcome that are well-defined at zero but behave like  $\log(y)$  when  $y$  is large (e.g.,  $\log(1 + y)$ ,  $\text{arcsinh}(y)$ ). We show that these ATEs depend arbitrarily on the units of the outcome, and thus cannot be interpreted as percentage effects. Moreover, we prove that when the outcome can equal zero, there is no parameter of the form  $E_P[g(Y(1), Y(0))]$  that is point-identified and unit-invariant. We discuss sensible alternative target parameters for settings with zero-valued outcomes that relax at least one of these requirements.

John Kohler 2d

Tell us about your career so far! Would love to hear about your experience at Microsoft Research, your decision to come to Stanford, and any career advice you have for causal inference in industry

♡ 2 Reply Edit Delete ...

EECS (04-09) (Athens,  
Greece)

Cornell CS PhD (09-14):  
Theoretical Computer  
Science; Algorithmic Game  
Theory

Microsoft Research NYC  
Postdoc (14-16): Machine  
Learning theory, Online  
Learning Theory

Microsoft Research New  
England (16-22)  
(Researcher; Principal  
Researcher): Causal Machine  
Learning, Econometrics; The  
ALICE Project and EconML

Stanford MS&E (22-)

# The ALICE project

## (Automated Learning and Intelligence for Causation and Economics)

- ◊ Research + Industry Problems + Software Development
- ◊ New application domains for causal inference, lead to novel CausalML methodologies
  - ◊ Recommendation A/B tests at TripAdvisor  
⇒ ML heterogeneous effects with instruments
  - ◊ Long-term Return-on-Investment (ROI) at Microsoft ⇒ Dynamic effects + surrogates in high dimensions
  - ◊ Personalized pricing and heterogeneous demand ⇒ Statistical learning for heterogeneous causal effects

### Major research challenges addressed by my work

- ◊ High-dimensionality of data and valid inference [ICML'18a, COLT'20, NeurIPS'21, CLeaR'22, Arxiv'21a,b,c]
- ◊ Unobserved confounding [NeurIPS'19, NeurIPS'20]
- ◊ Dynamic aspects of decision making [NeurIPS'15 (best paper), ICML'16, NeurIPS'16, FOCS'17 & JACM'20, EC'15 & OR'20, ICML'18b,c, ICML'21, NeurIPS'21a,b]
- ◊ Personalized effects and policies [ICML'19, NeurIPS'19, COLT'19 (best paper) & R&R Annals of Stats, CLeaR'22]

Contributions to statistical learning theory, high-dimensional statistics, semi-parametric and non-parametric inference theory, optimization theory, online learning theory

downloads 704k  
Starred 1.9k  
downloads/month 88k

**EconML**

"Developing a deep understanding of our travelers so we can create truly relevant experiences is at the core of what we do at TripAdvisor. Our partnership with Microsoft Research has allowed us to unlock critical insights that inform how to improve those experiences in a 1:1 manner."

—Matthew Dacey, Vice President Membership and Growth at TripAdvisor

Using Non-Linear Models to Study Aerosol-Cloud Interactions in the Southeast Pacific  
Machine Learning-Aided Causal Inference Framework for Environmental Data Analysis: A COVID-19 Case Study  
A World of Causal Inference with EconML by Microsoft Research  
EconML with R in the Tidyverse  
When a Voice Assistant Asks for Feedback: An Empirical Study on Customer Experience with A/B Testing and Causal Inference Methods  
Empirical Estimates on Hand Manipulations are Recoverable: A Step Towards Individualized and Explainable Robotic Support in Everyday Activities  
Causal Inference and COVID-19 Nursing Home Patients Identifying Factors That Reduced Mortality Risk  
Combining data analysis with machine learning for better health outcomes  
Software Engineer Internships at Microsoft Research  
Data Scientist Business  
Impact where it matters

Microsoft Walmart Electrolux

Microsoft introduces new resources & tools to help implement AI responsibly

EconML/CausalML KDD 2021 Tutorial  
Causal Inference and Machine Learning in Practice with EconML and CausalML: Industrial Use Cases at Microsoft, Tel Aviv University

# The ALICE project

- ◊ Research + Industry Problems + Software Development
- ◊ New application domains for causal inference, lead to novel CausalML methodologies
  - ◊ Recommendation A/B tests at TripAdvisor  
ML heterogeneous effects with instruments
  - ◊ Long-term Return-on-Investment (ROI) at Microsoft ⇒ Dynamic effects + surrogates in high dimensions
  - ◊ Personalized pricing and heterogeneous demand ⇒ Statistical learning for heterogeneous causal effects

- ◊ Making CausalML methods accessible with the EconML library



PUBLISHED ON DECEMBER 18, 2020 IN NEWS  
**Microsoft introduces New Resources & Tools To Help Implement AI Responsibly**

## The MSR ALICE Project

Automated Learning and Intelligence for Causation and Economics

- **Research.** Advance methodological research in econometrics and ML
- **Impact.** Apply Econ + ML methods to industry and societal problems
- **Software.** Develop software tools that reduce barriers to entry



**EconML** is a Python package that applies the power of machine learning techniques to estimate individualized causal responses from observational or experimental data.



**Flexible** model forms avoid strong assumptions and can estimate personalized responses to treatment



**Unified API** brings together all the latest advances in causal machine learning and econometrics



**Familiar Interface** built on standard Python packages make causal analysis quicker and easier for a broad set of users

**Yuwei Wu** 1d

Can we go over more examples of policy optimization based on heterogeneous treatment effects?

 Reply Edit Delete ...

**Amod Hegde** 20h

Can you tell us how the Causal + ML combo has impacted the field of causal inference? Has this led to an increase in the adoption of causal inference in the industry? Does this open up new interesting application domains for causal inference which weren't accessible before?

 Reply Edit Delete ...

<https://causal-machine-learning.github.io/kdd2021-tutorial/>

<https://drive.google.com/file/d/1QJJUCo4LH5kGQP3kaJIG1RdhjhaJWp-5/view>

[https://docs.google.com/presentation/d/1FvRtis2fm4c2R7XmRKWM\\_TtZaZjUObW1fGxpNmapmjKI/edit#slide=id.ge6ef5be800\\_0\\_478](https://docs.google.com/presentation/d/1FvRtis2fm4c2R7XmRKWM_TtZaZjUObW1fGxpNmapmjKI/edit#slide=id.ge6ef5be800_0_478)

[https://drive.google.com/file/d/1yyIu\\_3epIVXbwzJj658lv4vxHGjtPh8n/view](https://drive.google.com/file/d/1yyIu_3epIVXbwzJj658lv4vxHGjtPh8n/view)

<https://drive.google.com/file/d/1FEKXFHHATntHjsEymXnEw6GAiUGMm8sG/view>

[Search docs](#)

## EconML User Guide

### Overview

Machine Learning Based  
Estimation of Heterogeneous  
Treatment Effects

#### Motivating Examples

Introduction to Causal Inference

#### Problem Setup and API Design

Library Flow Chart

Detailed estimator comparison

#### Estimation Methods under Unconfoundedness

# Overview

EconML is a Python package that applies the power of machine learning techniques to estimate individualized causal responses from observational or experimental data. The suite of estimation methods provided in EconML represents the latest advances in causal machine learning. By incorporating individual machine learning steps into interpretable causal models, these methods improve the reliability of what-if predictions and make causal analysis quicker and easier for a broad set of users.

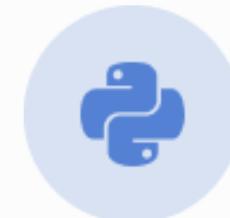
EconML is open source software developed by the [ALICE](#) team at Microsoft Research.



Flexible

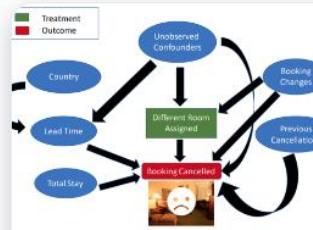


Unified



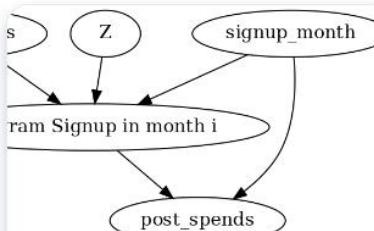
Familiar Interface

## Case Studies



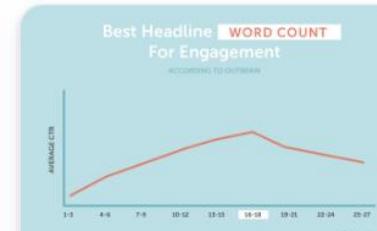
### Hotel Booking Cancellations

Beyond predictive models: The causal story behind hotel booking cancellations.



### Effect of Customer Loyalty Programs

Estimating the effect of a member rewards program.



### Optimizing Article Headlines

Introducing the do-sampler for causal inference.



### Effects of Home Visits on Infant Health (IHDP)

Understanding the question of why.

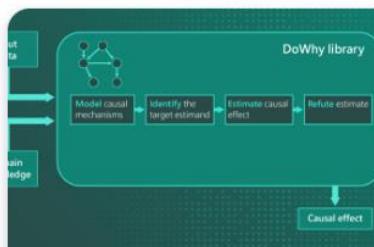
[See more case studies →](#)

## News



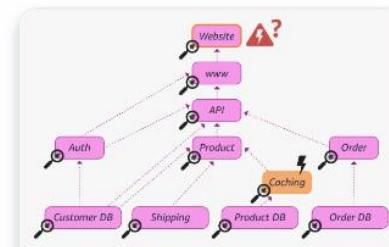
### National Association for Business Economics Conference

Causal inference at scale presented at NABE.



### DoWhy evolves to independent PyWhy model to help causal inference grow

Identifying causal effects is an integral part of scientific inquiry. It helps us understand everything from



### AWS contributes novel causal machine learning algorithms to DoWhy Python library

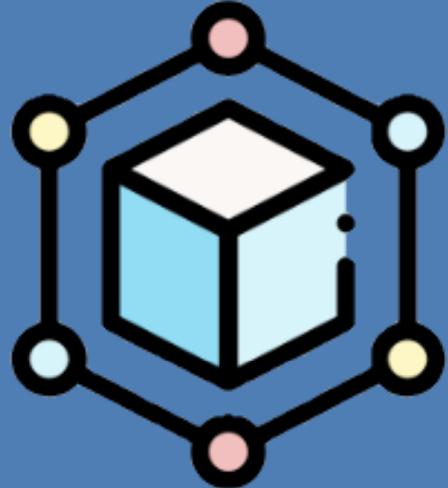
New features go beyond conventional effect estimation by attributing events to individual components of complex



Discover DoWhy  
A software library for causal inference

### Announcing DoWhy, a software library for causal inference

For decades, causal inference methods have found wide applicability in the social and biomedical sciences. As computing systems start



latest

Search docs

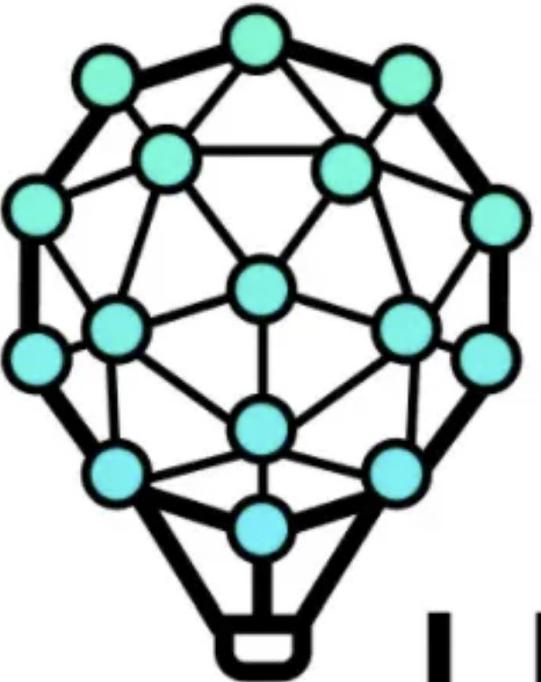
[Home](#) / About Causal ML

[Edit on GitHub](#)

## About Causal ML

[Causal ML](#) is a Python package that provides a suite of uplift modeling and causal inference methods using machine learning algorithms based on recent research. It provides a standard interface that allows user to estimate the **Conditional Average Treatment Effect (CATE)** or **Individual Treatment Effect (ITE)** from experimental or observational data. Essentially, it estimates the causal impact of intervention **T** on outcome **Y** for users with observed features **X**, without strong assumptions on the model form.

Uber



# UpliftML



Booking.com

UpliftML is a Python package for scalable unconstrained and constrained uplift modeling from experimental data. To accommodate working with big data, the package uses PySpark and H2O models as base learners for the uplift models. Evaluation functions expect a PySpark dataframe as input.



Azure

Explore ▾ Products ▾ Solutions ▾ Pricing ▾ Partners ▾ Resources ▾

Search



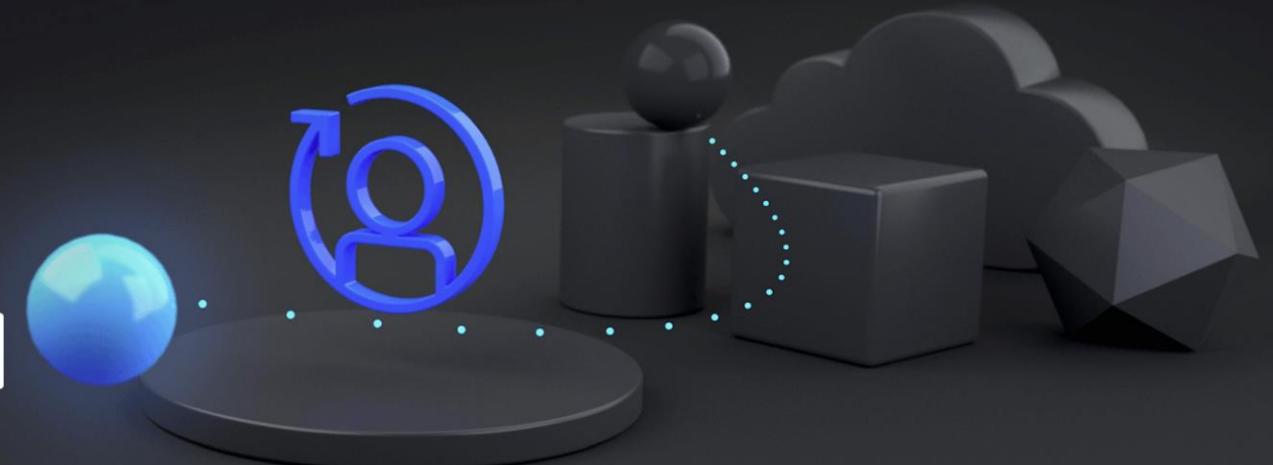
Learn Support Contact Sales

Free account

Sign in

## Personalizer

Deliver personalized, relevant experiences for each of your users.

[Try Personalizer free](#)[Create a pay-as-you-go account](#)[Overview](#) [Features](#) [Security](#) [Pricing](#) [Get started](#) [Resources](#) [FAQ](#)

### Create optimized user experiences

Boost conversion and engagement, and add real-time relevance to product recommendations, with reinforcement learning-based capabilities available only through Azure. Select hero content, optimize layouts, and personalize offers with two API calls. Use Personalizer, part of Azure Cognitive Services, as a standalone personalization solution or to complement existing ranking engines—with no machine learning expertise required.

[Chat with sales](#)



Contact Us Support▼ My Account▼ Sign In

re:Invent Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace Customer Enablement Event

Blog Home

Category ▾

Edition ▾

AWS Machine Learning Blog

# Power contextual bandits using continual learning with Amazon SageMaker RL

by Saurabh Gupta, Anna Luo, Bharathan Balaji, Siddhartha Agarwal, Vineet Khare, and Yijie Zhuang | on 29 AUG 2019 | in  
Amazon SageMaker, Artificial Intelligence | [Permalink](#) | [Comments](#) | [Share](#)

Resources

[Getting Started](#)  
[What's New](#)

# Online shopping gets more personal with Recommendations AI

July 22, 2020



**IKEA Retail (Ingka Group) has increased global average order value for eCommerce by 2% with Recommendations AI**



**Yasu Yoshida** 1d

I would like to know more about the frontiers of current research and the challenges you are trying to solve with your current research.

 Reply Edit Delete ...

**Cecile Bourbonnais** 13h

Are there specific emerging areas of research in causal inference right now that are most exciting to you, especially given recent technological advances? Or fields outside of medicine/econ that historically have not relied on causal inference as much, but that you feel should or are starting to?

 Reply Edit Delete ...

# Many Research Challenges

Long-term outcomes  
Safety constraints  
Fairness and ethical constraints  
Experimentation constraints (ability to intervene on subsets of variables)

Computationally efficient sensitivity analysis

Evidence based interpretations for policy makers

Causal attribution

Good losses for causal model selection

Active data collection  
[EC'15/OR'20, EC'26/OR'21,  
ICML'16, NeurIPS'16,  
FOCS'17, ICML'18a,b,  
ICML'21]

Data Collection  
[NeurIPS'17, NeurIPS'21  
MLEcon workshop]

Domain Assumption Elicitation  
[Clear'22]

Sensitivity Analysis  
[Arxiv'21]

Identification Strategy  
[NeurIPS'21a,b]

Interpretation/Policy Recommendations  
[NeurIPS'19,  
ICLR'21, Clear'22]

Estimation/Inference  
[COLT'19 R&R AoS,  
COLT'20, NeurIPS'20,  
NeurIPS'21]

Validation  
[COLT'19]

Robustness to data corruption/adversarial attacks

Interactive assumption elicitation (query complexity)

Causal Representation Learning for unstructured data

Incorporate non-graphical restrictions in automated graph-based identification algorithms

Automated de-biasing in general (e.g. dynamic regime, mediation effects)

Unobserved confounding

Post adaptive data collection inference  
Post causal discovery inference

❖ <https://github.com/causaltext/causal-text-papers>

---

**Adapting Text Embeddings for Causal Inference**

---

Victor Veitch\*      Dhanya Sridhar\*      David M. Blei

Department of Statistics and Department of Computer Science  
Columbia University

embedding methods. We illustrate the methods by answering the two motivating questions—the effect of a theorem on paper acceptance and the effect of a gender label on post popularity. Code and data available at [github.com/vveitch/causal-text-embeddings-tf2](https://github.com/vveitch/causal-text-embeddings-tf2).

## Table of Contents

1. Datasets and Simulations
2. Learning resources and blog posts
3. Causal Inference with Text Variables
  - i. Text as treatment
  - ii. Text as mediator
  - iii. Text as outcome
  - iv. Text as confounder
4. Causality to improve NLP
  - i. Causal interpretations and explanations
  - ii. Sensitivity and robustness
5. Applications in the Social Sciences
  - i. Linguistics
  - ii. Marketing
  - iii. Persuasion & Argumentation
  - iv. Mental health
  - v. Psychology
  - vi. Economics
  - vii. Bias and Fairness
  - viii. Social Media
  - ix. Law
  - x. Online Hate Speech

Max Schuessler 23h

Have you seen any of those methods being applied to -omics analyses in biology? I heard that in biostatistics, researchers traditionally use modelling to understand causal effects of a specific biomarker. What are the differences between modelling and the approaches we discussed in class in terms of their application?

♥ Reply Edit Delete \*\*\*

---

<https://insitro.com/>

---

## A FLEXIBLE APPROACH FOR PREDICTIVE BIOMARKER DISCOVERY

---

PREPRINT

**Philippe Boileau**

Graduate Group in Biostatistics and  
Center for Computational Biology,  
University of California, Berkeley, and  
[philippe\\_boileau@berkeley.edu](mailto:philippe_boileau@berkeley.edu)

**Nina Ting Qi**

Genentech Inc.  
[qit3@gene.com](mailto:qit3@gene.com)

**Mark van der Laan**

Division of Biostatistics,  
Department of Statistics, and  
Center for Computational Biology,  
University of California, Berkeley  
[laan@berkeley.edu](mailto:laan@berkeley.edu)

**Sandrine Dudoit**

Department of Statistics,  
Division of Biostatistics, and  
Center for Computational Biology,  
University of California, Berkeley  
[sandrine@stat.berkeley.edu](mailto:sandrine@stat.berkeley.edu)

**Ning Leng**

Genentech Inc.  
[lengn@gene.com](mailto:lengn@gene.com)

Daniel Jenson 22h

I've been going through some of the earlier proofs and I'm still not sure what we mean when we say 'partial out X and D from 1' so that you get  $\tilde{1}$ . In my mind,  $X = (1, W)$ , so if you have  $1 \sim D + X$ , this is equivalent to  $1 \sim D + 1 + W$ , in which case it will assign the coefficient 1 to the bias term in X and ignore D and W?

Heart Reply Edit Delete ...

$$V_{11} = \frac{E\epsilon^2 \tilde{D}^2}{(E\tilde{D}^2)^2}.$$

Applying the same theory for  $\beta_1$  (the intercept coefficient), yields<sup>20</sup>

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \approx \sqrt{n} \frac{E_n \epsilon \tilde{1}}{E_n \tilde{1}^2} \stackrel{a}{\sim} N(0, V_{22}),$$

where  $\tilde{1} := (1 - D)$  is the residual after partialling out D and X from 1 and

20: To explain the derivation, note that by partialling out D and W (recall that  $X = (1, W)$ ) from 1 and Y, we obtain

$$\tilde{Y} = \beta_1 \tilde{1} + \epsilon; \quad \tilde{1} := (1 - D).$$

The projection of 1 on D and W is given by D since D is binary and we've assumed EW = 0.

Typo! Should have been W

$$V_{22} = \frac{E\epsilon^2 \tilde{1}^2}{(E\tilde{1}^2)^2}.$$

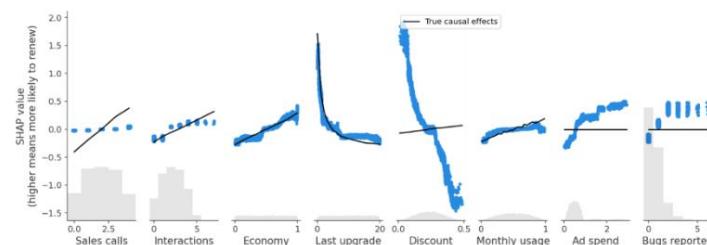
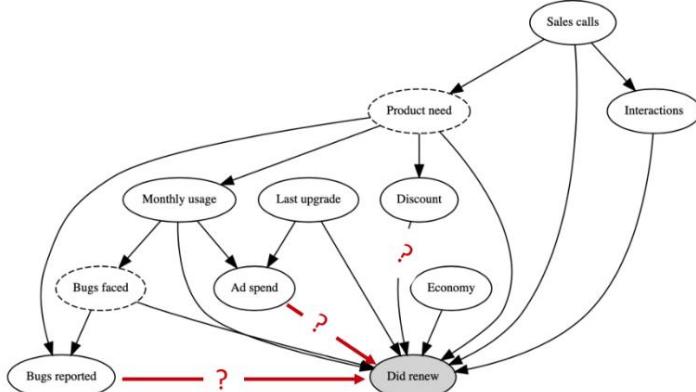
Louis Gautier 21h

Could you please go back over the motivation slides from the very first lecture (the example on building a model that predicts whether a customer will renew their product subscription) and explain how you got the true causal effect graphs on slide 17? It would be very useful to see how to use the tools of the course to tackle this problem.

Reply Edit Delete ...

## Require domain knowledge

of the high-level mechanisms that underlie the data collection process



<https://towardsdatascience.com/be-careful-when-interpreting-predictive-models-in-search-of-causal-insights-e68626e664b6>

Partialling Out DoubleML

Mike Van Ness 20h

Even though causal inference seems to be a large area of research and practice in Economics, I struggle to understand how causal inference is used in an economic setting, likely because of my lack of any knowledge of Economics. From my viewpoint, it seems that economics in the real world is so complex that it would be impossible to ever successfully do identification because there would always very likely be unobserved confounding. Can you explain a few simple problems in economics where causal inference is useful?

♡ 1 Reply Edit Delete \*\*\*

14 October 2019

The Royal Swedish Academy of Sciences has decided to award the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2019 to

**Abhijit Banerjee**

Massachusetts Institute of Technology, Cambridge, USA

**Esther Duflo**

Massachusetts Institute of Technology, Cambridge, USA

**Michael Kremer**

Harvard University, Cambridge, USA

*"for their experimental approach to alleviating global poverty"*

## Their research is helping us fight poverty

The research conducted by this year's Laureates has considerably improved our ability to fight global poverty. In just two decades, their new experiment-based approach has transformed development economics, which is now a flourishing field of research.

11 October 2021

The Royal Swedish Academy of Sciences has decided to award the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021

with one half to

**David Card**

University of California, Berkeley, USA

*"for his empirical contributions to labour economics"*

and the other half jointly to

**Joshua D. Angrist**

Massachusetts Institute of Technology, Cambridge, USA

**Guido W. Imbens**

Stanford University, USA

*"for their methodological contributions to the analysis of causal relationships"*

## Natural experiments help answer important questions for society

This year's Laureates – David Card, Joshua Angrist and Guido Imbens – have provided us with new insights about the labour market and shown what conclusions about cause and effect can be drawn from natural experiments. Their approach has spread to other fields and revolutionised empirical research.

# GENERIC MACHINE LEARNING INFERENCE ON HETEROGENOUS TREATMENT EFFECTS IN RANDOMIZED EXPERIMENTS, WITH AN APPLICATION TO IMMUNIZATION IN INDIA

VICTOR CHERNOZHUKOV, MERT DEMIRER, ESTHER DUFLO, AND IVÁN FERNÁNDEZ-VAL

**ABSTRACT.** We propose strategies to estimate and make inference on key features of heterogeneous effects in randomized experiments. These key features include *best linear predictors of the effects* using machine learning proxies, *average effects sorted by impact groups*, and *average characteristics of most and least impacted units*. The approach is valid in high dimensional settings, where the effects are proxied (but not necessarily consistently estimated) by predictive and causal machine learning methods. We post-process these proxies into estimates of the key features. Our approach is generic, it can be used in conjunction with penalized methods, neural networks, random forests, boosted trees, and ensemble methods, both predictive and causal. Estimation and inference are based on repeated data splitting to avoid overfitting and achieve validity. We use quantile aggregation of the results across many potential splits, in particular taking medians of p-values and medians and other quantiles of confidence intervals. We show that quantile aggregation lowers estimation risks over a single split procedure, and establish its principal inferential properties. Finally, our analysis reveals ways to build provably better machine learning proxies through causal learning: we can use the objective functions that we develop to construct the best linear predictors of the effects, to obtain better machine learning proxies in the initial step. We illustrate the use of both inferential tools and causal learners with a randomized field experiment that evaluates a combination of nudges to stimulate demand for immunization in India.

# Causal Inference in Economics

- Beyond RCTs, methods that handle unobserved confounding are much more credible and more frequently used
- Instrumental Variables (demand, labor, education)
- Difference in Difference
- Regression Discontinuity
- Synthetic Controls (country level program evaluation)

# Journal of Political Economy

**Lead Editor:** Magne Mogstad

**Editors:** John Asker, Andrew Atkeson, Mark Bils, Leonardo Bursztyn, Melissa Dell, Rachel Griffith, Emir Kamenica, Greg Kaplan, John List, Lance Lochner, Esteban Rossi-Hansberg, Azeem Shaikh

- Consumption and Income Inequality in the United States since the 1960s

Bruce D. Meyer and James X. Sullivan

pp. 247–284

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

FREE

- Getting Dynamic Implementation to Work

Yi-Chun Chen, Richard Holden, Takashi Kunimoto, Yifei Sun, and Tom Wilkering

pp. 285–387

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

- The Performance of School Assignment Mechanisms in Practice

Monique De Haan, Pieter A. Gautier, Hessel Oosterbeek, and Bas van der Klaauw

pp. 388–455

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

- Sources of Wage Growth

Jérôme Adda and Christian Dustmann

pp. 456–503

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

- Costless Information and Costly Verification: A Case for Transparency

Deniz Kattwinkel and Jan Knoepfle

pp. 504–548

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

- Large-Scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India

Gaurav Khanna

pp. 549–591

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

As education levels rise, we expect earnings and therefore the returns to be affected in a few ways. First, more educated workers are more productive and earn higher wages. Second, educated workers may reside in regions where there are fewer educated workers, making them relatively more valuable in the labor market. But if large numbers of people receive additional education, there is also a GE effect in the labor market: an increase in the abundance of high-skill labor puts downward pressure on the earnings skill premium. Yet as more skilled workers join the labor force, skill-biased capital may be adopted by firms in these regions, raising the premium. Indeed, as workers switch to more productive skill groups, overall output may increase, to the benefit of all workers. I estimate all of these components of the GE effects to better quantify the distributional impacts and the changes in labor market benefits.

The policy I study was India's flagship education scheme in the 1990s and early 2000s, the District Primary Education Program (DPEP), which expanded public schooling in half the country by targeting low-literacy regions. At that time, it was the largest program for primary education in the

Under the allocation rule, districts that had a female literacy rate below the national average were more likely to receive the program. I compare regions on either side of the cutoff to estimate causal impacts. The RD allows me to tackle biases that arise when estimating the individual returns to

# Journal of Political Economy

**Lead Editor:** Magne Mogstad

**Editors:** John Asker, Andrew Atkeson, Mark Bils, Leonardo Bursztyn, Melissa Dell, Rachel Griffith, Emir Kamenica, Greg Kaplan, John List, Lance Lochner, Esteban Rossi-Hansberg, Azeem Shaikh

Morgan Kelly, Joel Mokyr, and Cormac Ó Gráda

pp. 59–94

Abstract

Full Text

PDF

Supplemental Material

## Optimal Cooperative Taxation in the Global Economy

V. V. Chari, Juan Pablo Nicolini, and Pedro Teles

pp. 95–130

Abstract

Full Text

PDF

Supplemental Material

## Personalized Pricing and Consumer Welfare

Jean-Pierre Dubé and Sanjog Misra

pp. 131–189

Abstract

Full Text

PDF

Supplemental Material

## A Semistructural Methodology for Policy Counterfactuals

Martin Beraja

pp. 190–201

Abstract

Full Text

PDF

Supplemental Material

## Attraction versus Persuasion: Information Provision in Search Markets

Pak Hung Au and Mark Whitmeyer

pp. 202–245

Abstract

Full Text

PDF

Supplemental Material

## Personalized Pricing and Consumer Welfare

Jean-Pierre Dubé and Sanjog Misra

PDF

PDF PLUS

Abstract

Full Text

Supplemental Material



### Abstract

We study the welfare implications of personalized pricing implemented with machine learning. We use data from a randomized controlled pricing field experiment to construct personalized prices and validate these in the field. We find that unexercised market power increases profit by 55%. Personalization improves expected profits by an additional 19% and by 86% relative to the nonoptimized price. While total consumer surplus declines under personalized pricing, over 60% of consumers benefit from personalization. Under some inequity-averse welfare functions, consumer welfare may even increase. Simulations reveal a nonmonotonic relationship between the granularity of data and consumer surplus under personalization.

# Journal of Political Economy

**Lead Editor:** Magne Mogstad

**Editors:** John Asker, Andrew Atkeson, Mark Bils, Leonardo Bursztyn, Melissa Dell, Rachel Griffith, Emir Kamenica, Greg Kaplan, John List, Lance Lochner, Esteban Rossi-Hansberg, Azeem Shaikh

## The Comparative Advantage of Firms

Johannes Boehm, Swati Dhingra, and John Morrow

pp. 3025–3100

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

## Exit versus Voice

Eleonora Broccardo, Oliver Hart, and Luigi Zingales

pp. 3101–3145

[Abstract](#) [Full Text](#) [PDF](#)

FREE

## Asset Management Contracts and Equilibrium Prices

Andrea M. Buffa, Dimitri Vayanos, and Paul Woolley

pp. 3146–3201

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

## Vertical Contracts with Endogenous Product Selection: An Empirical Analysis of Vendor Allowance Contracts

Sylvia Hristakeva

pp. 3202–3252

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

## Breaking the Cycle? Intergenerational Effects of an Antipoverty Program in Early Childhood

Andrew Barr and Chloe R. Gibbs

pp. 3253–3285

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

## Aggregate Implications of Changing Sectoral Trends

Andrew T. Foerster, Andreas Hornstein, Pierre-Daniel G. Sarte, and Mark W. Watson

pp. 3286–3333

[Abstract](#) [Full Text](#) [PDF](#) [Supplemental Material](#)

# Vertical Contracts with Endogenous Product Selection: An Empirical Analysis of Vendor Allowance Contracts

Sylvia Hristakeva

[PDF](#) [PDF PLUS](#) [Abstract](#) [Full Text](#) [Supplemental Material](#)

## Abstract

Producers frequently provide retailers with financial incentives to secure product distribution. These payments often take the form of vendor allowances: lump-sum transfers to retailers that do not directly depend on quantity sold. I introduce an estimation strategy that uses observed product selections to inform unobserved allowances. I use retailers' replacement threats, which may allow them to capture both vendor transfers and lower wholesale prices. A counterfactual restricts firms to contract on only wholesale prices. Results show that vendor allowances may have not only (negative) product distortion effects but also (potentially positive) pricing effects.

The estimation methodology addresses retail price endogeneity by employing cost shifters as instrumental variables. I create a distance measure to capture transportation costs from each producer's manufacturing facility to each market. I locate yogurt plants in the United States that were used during the sample period. Appendix C summarizes the collected geographic distance information. To calculate a proxy for transportation costs between plants and each market, I combine these geographic distances with gas prices obtained from the US Energy Information Administration.

Hailong Chen 19h

What are the most common mistakes people do in causal inference practice?

1 Reply Edit Delete ...

- Believing that if you blindly control for a gazillion of features you have addressed unobserved confounding
- Trying to extract heterogeneity from very few samples
- Blindly using recent causal ml techniques as if it was a predictive problem without understanding anything about the data
- Believing that if you use machine learning to control, you fixed your unobserved confounding problem

Megan Li 19h

Are there any causal inference techniques that are widely used but you believe should be deprecated? If so, why?

♡ 1 Reply Edit Delete ...

# Why Propensity Scores Should Not Be Used for Matching\*

Gary King<sup>†</sup>      Richard Nielsen<sup>‡</sup>

November 10, 2018

## Abstract

We show that propensity score matching (PSM), an enormously popular method of preprocessing data for causal inference, often accomplishes the opposite of its intended goal — thus increasing imbalance, inefficiency, model dependence, and bias. The weakness of PSM comes from its attempts to approximate a completely randomized experiment, rather than, as with other matching methods, a more efficient fully blocked randomized experiment. PSM is thus uniquely blind to the often large portion of imbalance that can be eliminated by approximating full blocking with other matching methods. Moreover, in data balanced enough to approximate complete randomization, either to begin with or after pruning some observations, PSM approximates random matching which, we show, increases imbalance even relative to the original data. Although these results suggest researchers replace PSM with one of the other available matching methods, propensity scores have other productive uses.

Zhuoyang Liu 14h

In SEM, suppose we have

$$Y = \alpha D + \beta^T W + \epsilon$$

Can we write down the following:

$$Y(1) = \alpha + \beta^T W + \epsilon_1$$

and

$$Y(0) = \beta^T W + \epsilon_0$$

If yes, are

$$\epsilon_1$$

and

$$\epsilon_0$$

the same? (so that

$$Y(1) - Y(0) = \alpha \text{ a.s.}$$

).

 Reply Edit Delete \*\*\*

Yes, if you assume such a linear SEM this is the implicit assumption on the potential outcomes! (quite restrictive indeed)

Can be relaxed if we only assume that the epsilon is un-correlated with all the other variables, not just independent. Then that implicitly allows for epsilon to depend on treatment d, but should always be mean zero

Junting Duan 14h

I'd love to know some popular topics in current research in causal inference.

I also have two technical questions: 1. In Lecture 5, we learned that single Lasso is not Neyman orthogonal and it might omit some strong predictors for the treatment. Could you intuitively explain why omit strong predictors for the treatment is a problem for estimating the treatment effect? 2. For the second plot on P29 Lecture 14, how do we derive the confidence bounds for ATE?

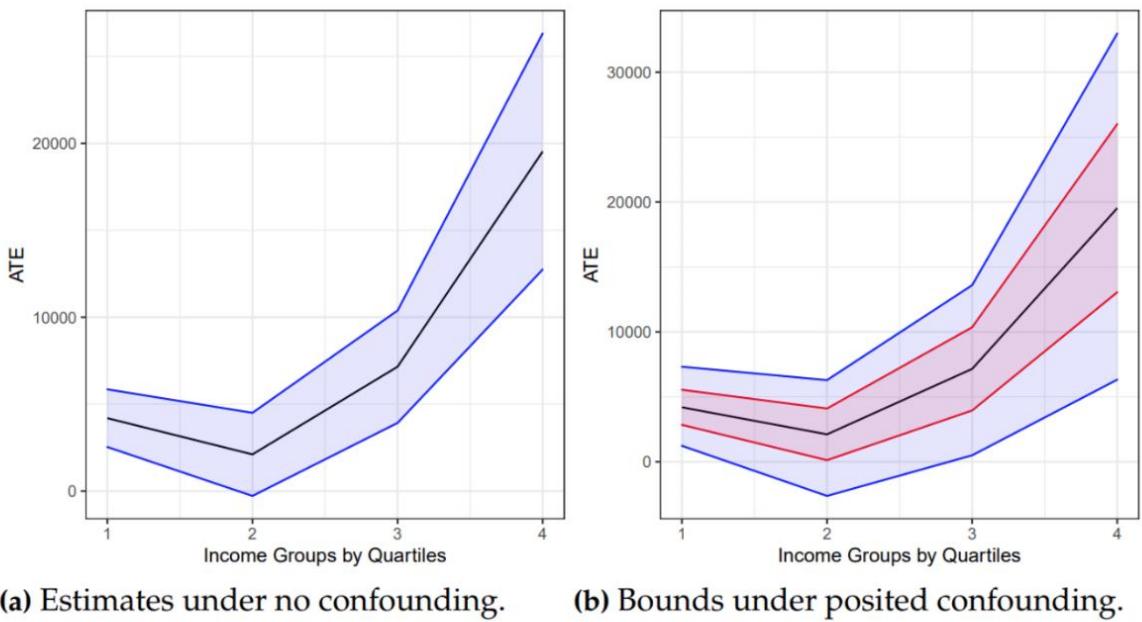
Thanks a lot!

Reply Edit Delete ...

- If you have a strong predictor of the treatment that was omitted by the single lasso due to regularization (as other factors were much stronger for the outcome), it is as if you did not control for that variable. This yields strong confounding bias.

## LONG STORY SHORT: OMITTED VARIABLE BIAS IN CAUSAL MACHINE LEARNING

V. CHERNOZHUKOV, C. CINELLI, W.K. NEWHEY, A. SHARMA, V. SYRGKANIS



**Note:** Estimate (black), bounds (red), and confidence bounds (blue) for the ATE. Confounding scenario:  $\rho^2 = 1$ ;  $C_Y^2 \approx 0.04$ ;  $C_D^2 \approx 0.031$ . Significance level of 5%.

$$\theta_s = E[m(W^s, g_s) + (Y - g_s)\alpha_s],$$

018d, 2021a). This representation is Neyman causal which is a key property required for DML. An

$$E(Y - g_s)^2 =: \sigma_s^2,$$

gonal with respect to  $g_s$ . The final component wing formulation:

$$E\alpha_s^2 = 2Em(W^s, \alpha_s) - E\alpha_s^2 =: v_s^2,$$

estimators are defined as

$$\hat{\theta}_s := \text{DML}(\psi_\theta); \quad \hat{\sigma}_s^2 := \text{DML}(\psi_{\sigma^2}); \quad \hat{v}_s^2 := \text{DML}(\psi_{v^2});$$

$$\begin{aligned}\psi_\theta(Z; \theta, g, \alpha) &:= m(W^s, g) + (Y - g(W^s))\alpha(W^s) - \theta; \\ \psi_{\sigma^2}(Z; \sigma^2, g) &:= (Y - g(W^s))^2 - \sigma^2; \\ \psi_{v^2}(Z; v^2, \alpha) &:= (2m(W^s, \alpha) - \alpha^2) - v^2.\end{aligned}$$

$$\varphi_\pm^o(Z) = \psi_\theta^o(Z) \pm \frac{|\rho|}{2} \frac{C_g C_\alpha}{S} (\sigma_s^2 \psi_{v^2}^o(Z) + v_s^2 \psi_{\sigma^2}^o(Z)).$$

: confidence bound

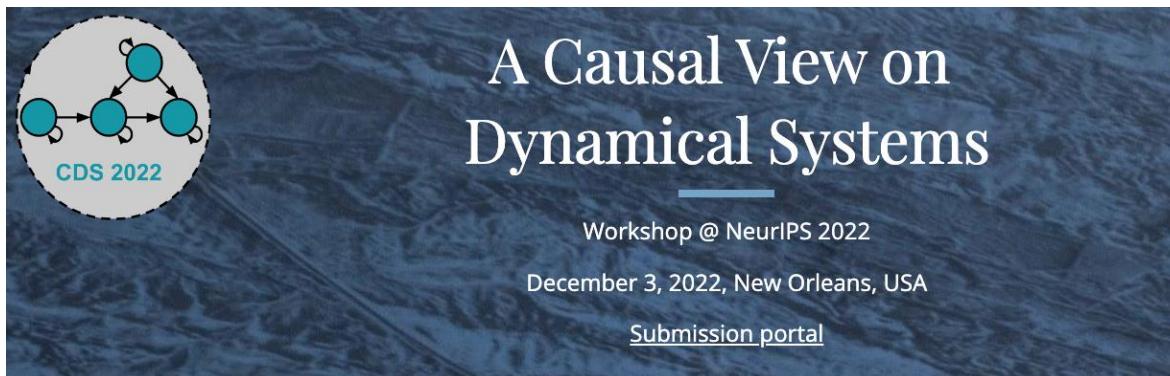
$$[\ell, u] = \left[ \hat{\theta}_- - \Phi^{-1}(1-a) \sqrt{\frac{E\varphi_-^o}{n}}, \hat{\theta}_+ + \Phi^{-1}(1-a) \sqrt{\frac{E\varphi_+^o}{n}} \right]$$

Soumya Koppaka 11h

How can we improve our ability to identify and estimate causal effects in complex and dynamic systems?

Because in many real world settings, there exist multiple correlated covariates, and feedback loops, some of which may be dynamic and difficult to measure.

♥ Reply Edit Delete ...



<https://sites.google.com/view/caudyn2022>

<http://networks.ece.mcgill.ca/sites/default/files/A%20Tutorial%20on%20Causal%20Inference%20in%20Dynamical%20Systems.pdf>

Shannon Marie Meyer 4h

Sorry this one is late - do you have any favorite exploratory data analysis methods for the setting of high-D causal inference? It can be easy to get lost in the details when starting on a project... (though maybe our best tools are the current literature and DAGs)

♥ Reply Edit Delete ...

- Double lasso for ATE (for continuous treatments); Doubly Robust (for categorical)
- Double lasso with feature interactions for quickly seeing if there is heterogeneity
- Generic meta learners with out of sample doubly robust validation for more flexible heterogeneity

David Troxell 1h

Since there is no way to validate the causal estimates we obtain given observational data (as discussed in lecture), are there any other ways besides confidence intervals to quantify the uncertainty of our estimate? For example, can full/split conformal inference provide some insight as to the stability of our modeling process (as opposed to the estimate itself)?

Another question I have is: after obtaining a DAG and identifying a valid adjustment set, is there a go-to process you usually follow in what to do next? Since we have discussed so many models/options I'm a bit confused on a good practical first step to do after identifying a valid adjustment set.

♡ Reply Edit Delete ...

- Stability and calibration
- Using a small experimental sample to validate
- Using multiple observational datasets and separately estimating, with agreement
- Conformal inference cannot target expected effects, only individual (random effects) and typically will yield very large intervals

GoTo:

- Doubly robust ATE estimate for binary and Partialling out estimate for categorical (first with just Lasso models on expanded features and then with generic AutoML)
- Meta learners for CATE with out-of-sample doubly robust based evaluation (calibration, hypothesis tests)