

MS&E328/CS328

Foundations of Causal Machine Learning

Vasilis Syrgkanis

Stanford University

Assistant Professor

Management Science and Engineering

(by courtesy) Computer Science and Electrical Engineering

Institute for Computational and Mathematical Engineering

Logistics

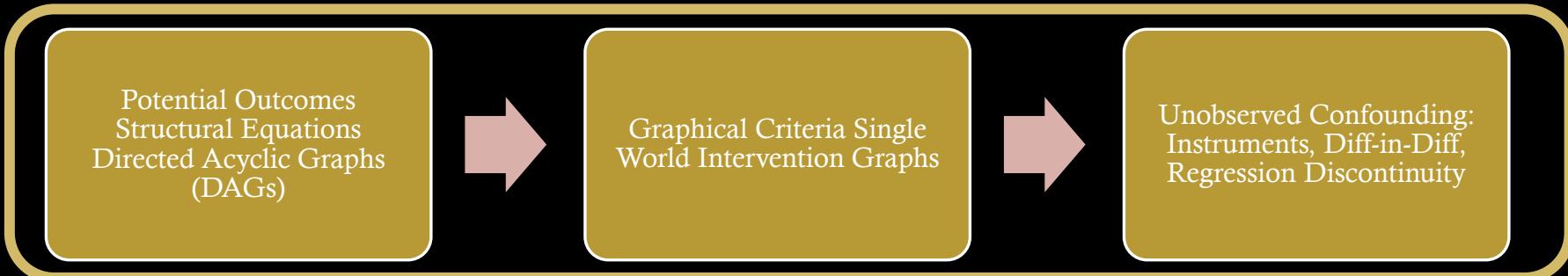
- ❖ **Instructor:** Vasilis Syrgkanis, Asst. Professor of Management Science and Engineering, (courtesy) Computer Science and Electrical Engineering
- ❖ **Office Hours:** Monday 5-6pm, Huang 252
- ❖ **Lectures:** Friday 1.30-2.50 (10min break) 3-4.20, Turing Auditorium
- ❖ **Grading:** Class Presentations (30%), Project Report (70%)
- ❖ **Webpage:** <https://stanford-msande328.github.io/winter26/>
- ❖ **Prior offering:** <https://stanford-msande328.github.io/fall24/>

- ❖ **Project:** Literature review on a particular topic related to class, covering at least 3 papers in depth. Replication of experiments in one paper or improvement on theoretical results (even if incremental). Proposal Due: Jan 30th. Lit Review Due: Feb 28th. Final Report Due: March 20th
- ❖ **Presentations:** 20min presentation of your research project

Class Goals

- ❖ Uncover the statistical and computational foundations of modern methods in causal effect identification and machine learning based estimation
- ❖ Be able to prove results on the identification of causal effects and understand novel arguments on causal effect identification
- ❖ Be able to prove results on the statistical properties of machine learning based estimation of causal parameters and causal functions, such as asymptotic normality and rates of convergence
- ❖ Be able to present understand and present papers on these topics

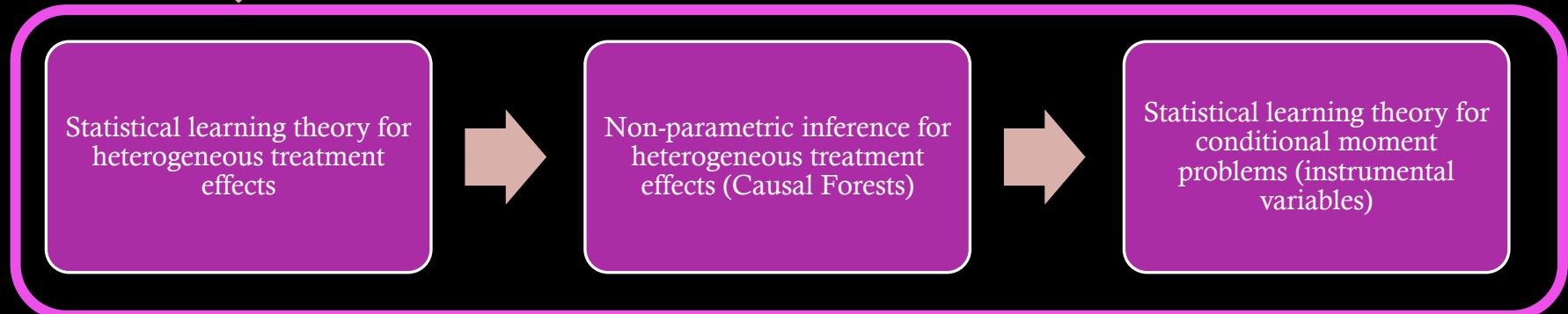
Causal Identification



Estimation and
Inference on
Causal
Parameters



Estimation and Inference on Causal Functions



Related Courses

- ❖ Applied Causal Inference with Machine Learning and AI
- ❖ <https://stanford-msande228.github.io/winter26/>

- ❖ ECON 272: Intermediate Econometrics III: Methods for Applied Econometrics (Imbens)
- ❖ MGTECON 634: Machine Learning and Causal Inference (Athey, Wager)
- ❖ MGTECON 614: Emerging Topics in Econometrics (Lihua Lei)
- ❖ MS&E 226: Fundamentals of Data Science: Prediction, Inference, Causality (Johari)
- ❖ STATS 209: Introduction to Causal Inference (Rothenhaeusler)
- ❖ STATS 361: Causal Inference (Rothenhaeusler, Wager)
<https://datascience.stanford.edu/causal/courses>

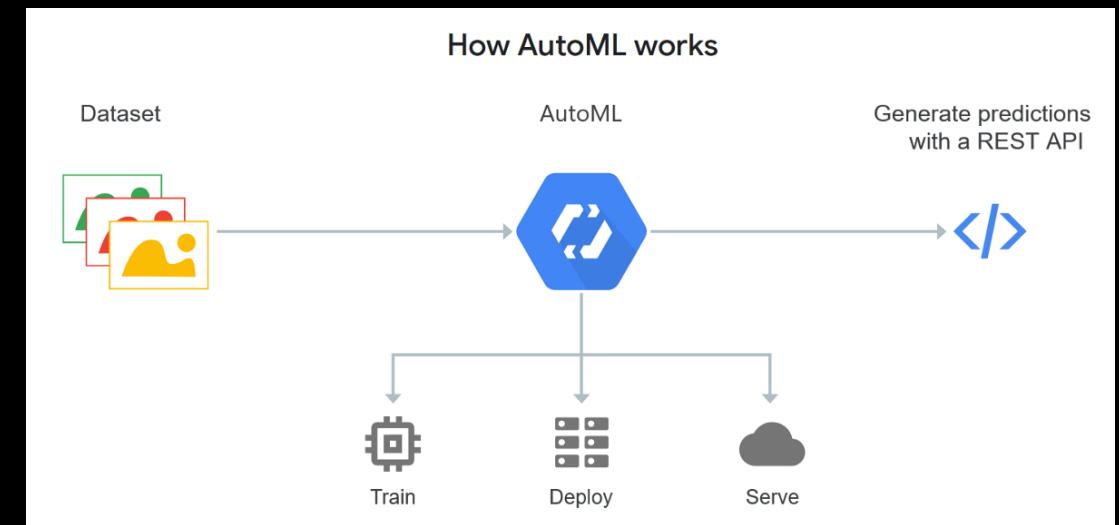
Introduction Causal Machine Learning in Practice Case Studies and Challenges

Vasilis Syrgkanis

Stanford University

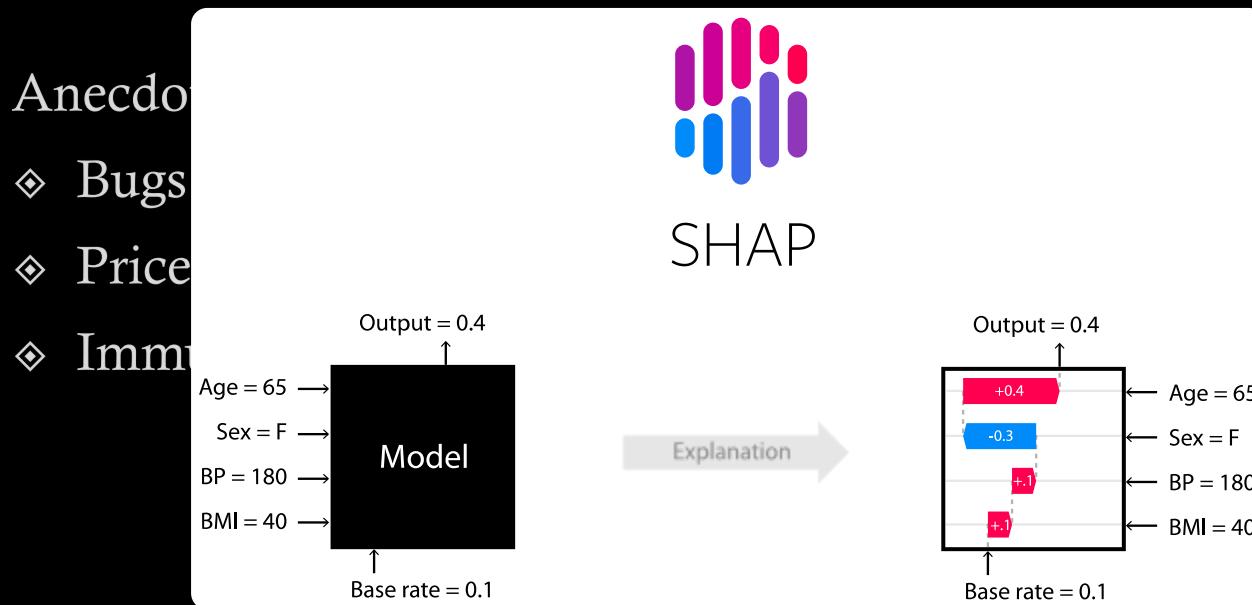
The Ease of Use of Machine Learning

- ❖ Almost everyone with access to a dataset and an outcome of interest can construct an ML predictive model



The Ease of (mis)Use of Machine Learning

- ❖ Almost everyone with access to a dataset and an outcome of interest can construct an ML predictive model
- ❖ Time and again people use insights from predictive models to make decisions
- ❖ AutoML + interpretability ≠ Causal inference



Scott Lundberg
May 17, 2021 · 16 min read

THOUGHTS AND THEORY
Be Careful When Interpreting Predictive Models in Search of Causal Insights

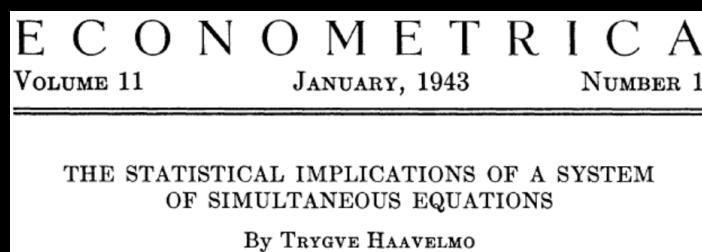
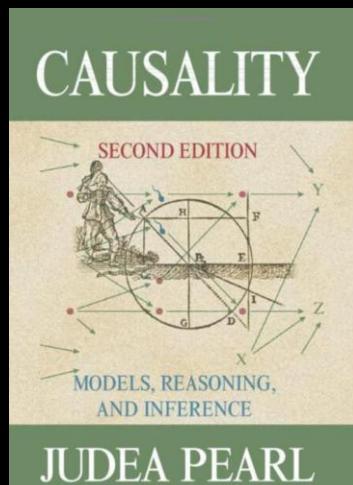
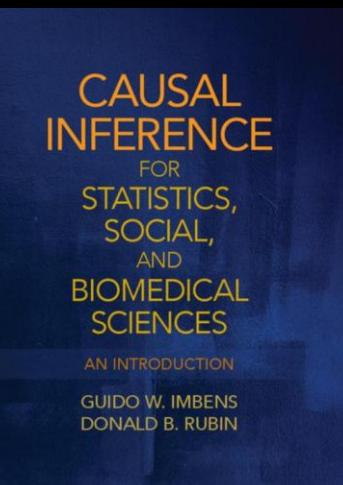
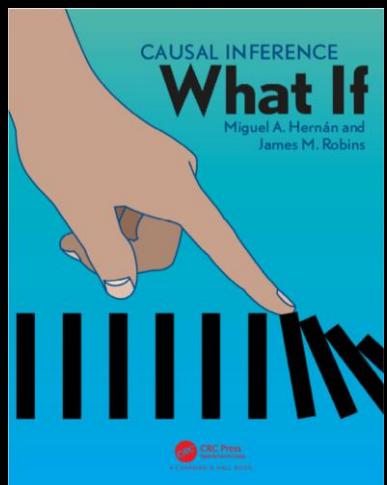
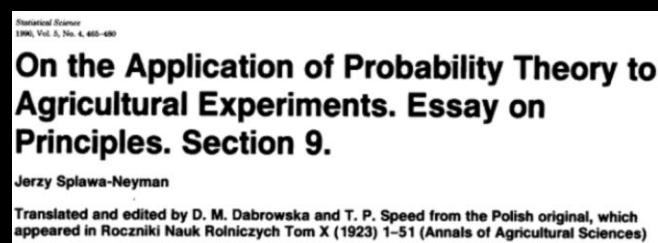
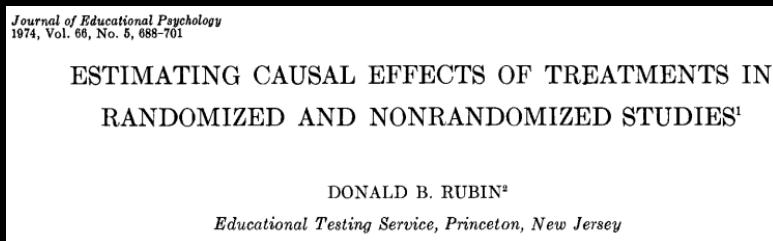
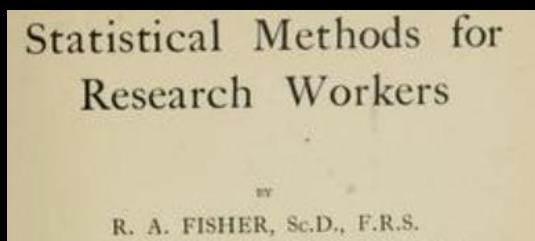
A careful exploration of the pitfalls of trying to extract causal insights from modern predictive machine learning models.

 © Scott Lundberg / llkcelik — iStock

A joint article about causality and interpretable machine learning with Eleanor Dillon, Jacob LaRiviere, Jonathan Roth, and Vasilis Syrgkanis from Microsoft.

Causal Inference

- ❖ Addresses interventional (what-if) statistical questions and the identification of causal relationships from data



Causal inference in new data modalities

- ❖ Richer datasets (10s or 100s of variables)
- ❖ Un-structured data (text, images, genomics)
- ❖ Personalization (heterogeneous effect models)
- ❖ Automatic model selection/model robustness/researcher bias robustness

Exactly the topics where predictive machine learning techniques really shine!

ML helps bypass the curse of dimensionality

- ❖ ML methods automatically adapt to unknown complexity of the truth
 - ❖ Adapt to number of relevant features (Lasso, Random Forests)
 - ❖ Adapt to low dimensionality of covariate distribution (Nearest Neighbor regression)
 - ❖ Adapt to the “complexity/smoothness” of true model (Kernel Ridge)
 - ❖ Build data-driven low dimensional feature representations (Neural Networks)

Causal Machine Learning

Re-directing the ability of machine learning estimators to bypass the curse of dimensionality, from the current focus of solving prediction problems to solving statistical problems that arise in causal inference.

Many industrial and scientific use cases

- ❖ Return-on-investment, pricing, customer segmentation and personalization
- ❖ Experimentation in tech industry
- ❖ Personalized medicine
- ❖ Heterogeneity of effect in social science studies

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

- ◊ 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, ICML'16, NeurIPS'16, FOCS'17, EC'15, ICML'18b,c, ICML'21, NeurIPS'21a,b]
- ◊ Personalized effects and policies [ICML'19, NeurIPS'19, COLT'19, CLeaR'22]

New problems in statistical learning theory, high-dimensional statistics, semi-parametric and non-parametric inference theory, optimization theory, online learning theory

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 8, 2021 IN NEWS

Microsoft Introduces New Resources & Tools To Help Implement AI Responsibly

Microsoft has launched new tools and guidelines to enable product leaders build AI responsibly from research to practice

EconML/CausalML KDD 2021 Tutorial

Causal Inference and Machine Learning in Practice with EconML and CausalML: Industrial Use Cases at Microsoft, TripAdvisor, Uber

Starred 1.9k

downloads 704k

downloads/month 88k

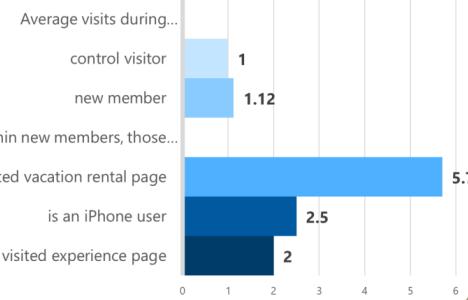


"Developing a deep understanding of our travelers so we can create truly relevant experiences is at the core of what we do at TripAdvisor. Our partnership with Microsoft Research has allowed us to unlock critical insights that inform how to improve those experiences in a 1:1 manner."



—Matthew Dacey, Vice President Membership and Growth at TripAdvisor

Heterogeneous Effect of Membership



Using Non-Linear Models to Study Aerosol-Cloud Interactions in the Southeast Pacific
Andrew Jessian, Peter Marshavinen, Alyson Douglas, Duncan Watson-Parris, Yarin Gal, Philip Stier
University of Oxford

Machine Learning-Aided Causal Inference Framework for Environmental Data Analysis: A COVID-19 Case Study
Qiao Kang,³ Xing Song,³ Xiaying Xin,³ Bing Chen,⁴ Yuanzhu Chen, Xudong Ye, and Baiyu Zhang
Cite This: Environ. Sci. Technol. 2021, 55, 13406–13410
Read Online

The Role of "Live" in Livestreaming Markets: Evidence Using Orthogonal Forest
Ziwei CONG
HUST Business School
Jia Liu
HUST Business School
Punet Manchanda
University of Michigan, Stephen M. Ross School of Business

Youth well-being predicts later academic success
Diana Cárdenas, Finnian Lattimore, Daniel Steinberg & Katherine J. Reynolds

When a Voice Assistant Asks for Feedback: An Empirical Study on Customer Experience with A/B Testing and Causal Inference Methods
Yuqi Deng
Amazon
Sudeeksha Murari
Amazon

Empirical Estimates on Hand Manipulation are Recoverable: A Step Towards Individualized and Explainable Robotic Support in Everyday Activities
Alexander Wichi
Institute for Computer Intelligence, University of Bremen, Bremen, Germany
Holger Schultheis
Institute for Computer Intelligence, University of Bremen, Bremen, Germany
Michael Beetz
Institute for Computer Intelligence, University of Bremen, Bremen, Germany
Amina Ahmed, Robert Goldberg, Joseph Swiderski, Zachary A.P. Wintrob, Naveen Ramya Kumar, Ngoc Le, Felipe Ormeño, Mary E. Brown, and Theresa M. Rieser
Cite This: JACS-Accts. 2022, XXXX, XXX, XXX-XXXX
Publication Date: February 7, 2022
Article Views | Alerts | Citation Metrics | Share | Learn About These Metrics

Combinatorial Polyacrylation Synthesis and Causal Machine Learning Reveal Divergent Polymer Design Rules for Effective pDNA and Ribonucleoprotein Delivery
Ramyaa Kumar, Ngoc Le, Felipe Ormeño, Mary E. Brown, and Theresa M. Rieser
Cite This: JACS-Accts. 2022, XXXX, XXX, XXX-XXXX
Publication Date: February 7, 2022
Article Views | Alerts | Citation Metrics | Share | Learn About These Metrics

Causal Inference and COVID-19 Nursing Home Patients Identifying Factors That Reduced Mortality Risk
Amina Ahmed, Robert Goldberg, Joseph Swiderski, Zachary A.P. Wintrob, Naveen Ramya Kumar, Ngoc Le, Felipe Ormeño, Mary E. Brown, and Theresa M. Rieser
doi: https://doi.org/10.1101/2021.11.18.21266489

Yuya Sugano
Mar 22, 2020 · 13 min read

A World of Causal Inference with EconML by Microsoft Research

微软EconML简介：基于机器学习的Heterogeneous Treatment Effects估计
会飞的船

EconML with R in the Tidyverse

Example Usage with Multiple Continuous Treatment, Multiple Outcome Observational Data
Itamar Caspi

Zillow Humana Grab

Combining data analytics with a personalized approach for better health outcomes.

MOBE Impact where it matters. Spotify

ZS ZS Impact where it matters. Management Consulting - Evanston, IL - 331,160 followers

Software Engineer Intern Systems Oncology - Part-time Sep 2020 - Jun 2021 - 10 mos Phoenix, Arizona, United States

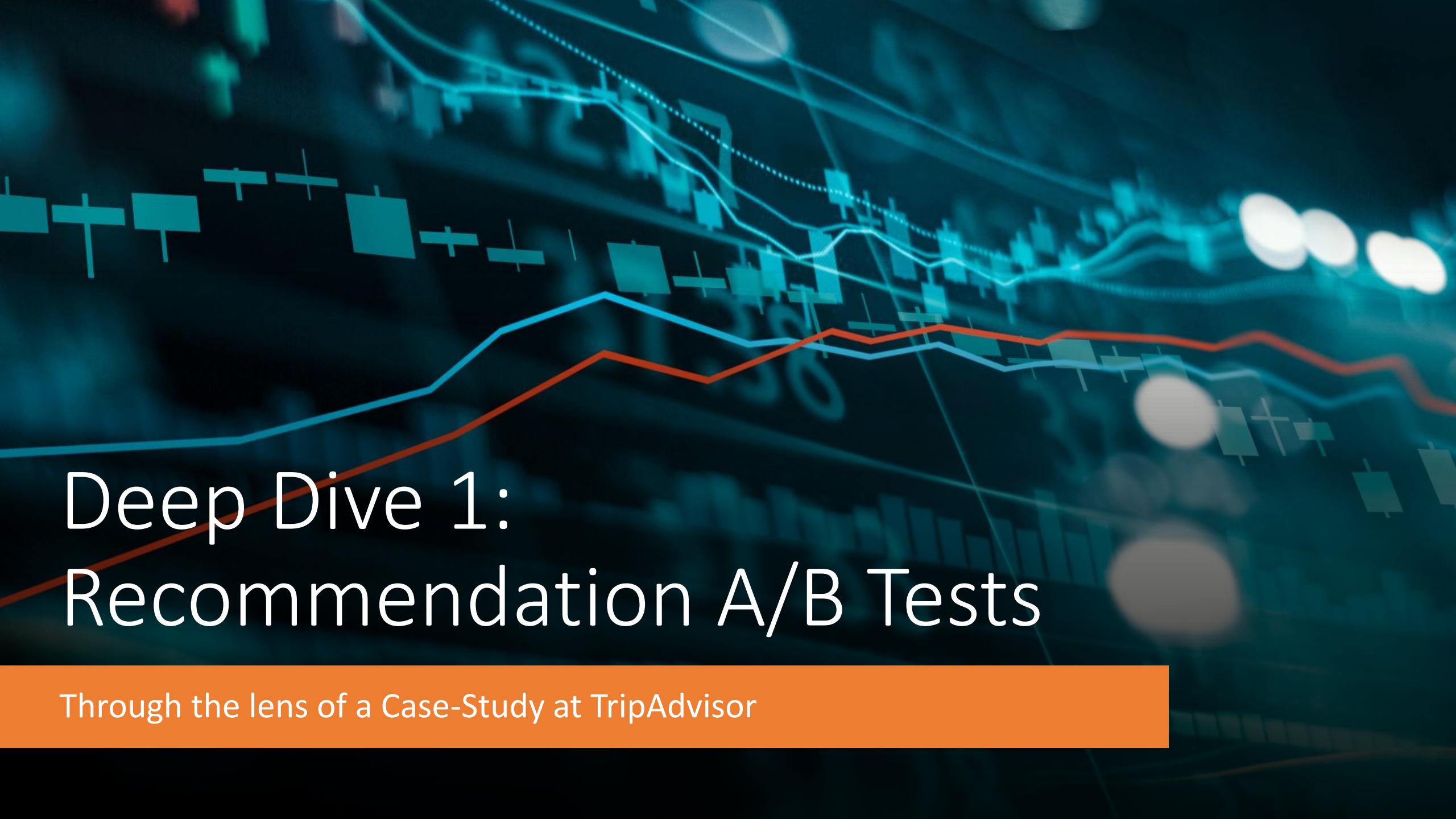
• Implemented and adapted Causal Inference libraries DoWhy and EconML

Locations Being Here Diversity, Equity & Impact Students How We Hire

Experience with causal inference methodologies and hands-on experience with packages like CausalML, EconML, etc. is a strong plus

Where you'll be:

Data Scientist Business Insights-



Deep Dive 1: Recommendation A/B Tests

Through the lens of a Case-Study at TripAdvisor

TripAdvisor Membership Problem

- ❖ What is the causal effect of becoming a member on TripAdvisor on downstream activity on the webpage?
- ❖ How does that effect vary with observable characteristics of the user?
- ❖ Useful for understanding the quality of membership offering/improvements/targeting

TripAdvisor Membership Problem

- ❖ What is the causal effect of becoming a member on TripAdvisor on downstream activity on the webpage?
- ❖ How does that effect vary with observable characteristics of the user?
- ❖ Useful for understanding the quality of membership offering/improvements/targeting

Standard approach: Let's run an A/B test!

Not applicable: We cannot enforce the treatment!

- ❖ We cannot take a random half of the users and make them members
- ❖ Membership is an action that requires user engagement!

Recommendation A/B Tests

- ❖ In optimizing a service we want to understand the causal effects of actions that involve user engagement (e.g. becoming a member)

Recommendation A/B Tests

- ❖ In optimizing a service we want to understand the causal effects of actions that involve user engagement (e.g. becoming a member)
- ❖ We can run a **recommendation A/B test**:
 - ❖ “recommend/create extra incentives” to half the users to take the action/treatment
- ❖ *Example at TripAdvisor*: enable an easier sign-up flow process for a random half of users

Recommendation A/B Tests

- ❖ In optimizing a service we want to understand the causal effects of actions that involve user engagement (e.g. becoming a member)
- ❖ We can run a **recommendation A/B test**:
 - ❖ “recommend/create extra incentives” to half the users to take the action/treatment
- ❖ *Example at TripAdvisor*: enable an easier sign-up flow process for a random half of users
- ❖ **Non-Compliance**: ``user’s choice to comply or not`` can lead to biased estimates

Instrumental Variables (IV)

- ❖ **Instrumental Variable:** any random variable Z that affects the treatment assignment T but does not affect the outcome Y other than through the treatment [Wright'28, Bowden-Turkington'90, Angrist-Krueger'91, Imbens-Angrist'94]
- ❖ Cohort assignment in recommendation A/B test is an instrument
- ❖ We can apply IV methods to estimate average treatment effect θ

TripAdvisor Experiment

For random half of 4 million users, easier sign-up flow was enabled

- ❖ Easier sign-up incentivizes membership

For each user we observe

- ❖ Instrument Z : whether the easier sign-up flow was enabled
- ❖ Variables X : observed characteristics of each user: e.g. prior history on platform, location
- ❖ Treatment T : whether the user became a member
- ❖ Outcome Y : number of visits in the next 14 days

Effect and Compliance Heterogeneity

- ❖ Typical IV methods do not account for complex effect or compliance heterogeneity
- ❖ Not accounting for heterogeneity in compliance and effect can lead to **biased average effect**
- ❖ Personalization requires estimates of **heterogeneous effect** $\theta(X)$

Given observed characteristics X , what is the probability of compliance?

Given observed characteristics X , what is the effect of the treatment?

Can we decompose IV estimation
into (ML) *prediction* problems?

De-composing into Prediction Problems

- ❖ **Compliance score** (Abadie'03): variation of probability of treatment induced by instrument

$$\Delta(Z, X) = \mathbb{P}(T = 1|Z, X) - \mathbb{P}(T = 1|X)$$

- ❖ **Residual outcome:** variation in outcome that is unpredictable by observed characteristics

$$\tilde{Y} = Y - \mathbb{E}[Y|X]$$

- ❖ **Heterogeneous effect** is the minimizer of the prediction loss:

$$\hat{\theta} = \operatorname{argmin}_{\theta(\cdot)} \mathbb{E} \left[(\tilde{Y} - \theta(X) \cdot \Delta(Z, X))^2 \right]$$

Predict

unexplained
variation in
outcome

\tilde{Y}

from

exogenous variation
of treatment induced
by instrument

$\Delta(Z, X)$

using

linear model where
the coefficient can
vary with observable
characteristics

$\theta(X) \cdot \Delta(Z, X)$

ML

Some Caution

- ❖ Only requirement from ML models should be a guarantee on expected prediction error on data from same distribution as the training data
- ❖ ML methods balance bias and variance; leads to biased estimates and wrong p-values
- ❖ Many times we need to solve much more complex prediction tasks than the quantity we care about; can we not depend heavily on the estimation error of these tasks



Robustness via Debiasing

- ❖ **De-biased** preliminary $\hat{\theta}(X)$ [Okui et al'12, Chernozhukov et al'18, Athey-Wager'19]

$$\hat{\theta}_{debiased}(X) = \hat{\theta}(X) + \frac{\tilde{Y} - \hat{\theta}(X) \cdot (T - \mathbb{P}(T = 1|X))}{\Delta(Z, X)}$$

Debiasing
Correction:
uses quantities
that ideally
are mean zero

- ❖ **Robust heterogeneous treatment effect**

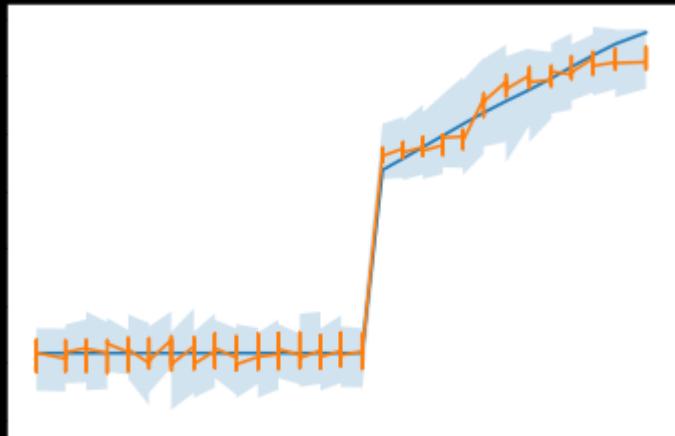
$$\min_{\theta(\cdot)} \mathbb{E} \left[\left(\hat{\theta}_{debiased}(X) - \theta(X) \right)^2 \right]$$

ML

- ❖ Mean-Squared-Error of final $\theta(X)$ robust to errors in first stage models [Chernozhukov, Nekipelov, Semenova, Syrgkanis'18, Foster – Syrgkanis'19]

Inference

Heterogeneous effect $\theta(X)$



Covariate X

- ❖ When final ML method supports CI construction, debiased final estimate typically **preserves the asymptotic validity of the intervals**
 - ❖ Inference on best linear projection of heterogeneous effect via OLS [Chernozhukov et al'16]
 - ❖ Inference on high-dimensional linear projections via Debiased Lasso [Zhang-Zhang'14, Javanmard-Montanari'14, van der Geer et al'14, Chernozhukov et al'17]
 - ❖ Non-Parametric inference via Honest Regression Forests [Athey-Wager'15, Athey-Wager-Tibshirani'16, Mentch-Hooker'16, Oprescu et al'19, Syrgkanis-Zampetakis'20]

Some Caution

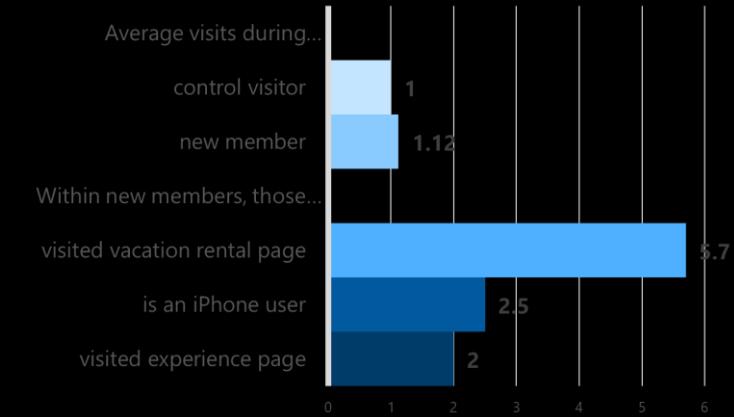
- ❖ Only requirement from ML models should be a guarantee on expected prediction error on data from same distribution as the training data
- ❖ ML methods balance bias and variance; leads to biased estimates and wrong p-values
- ❖ Many times we need to solve much more complex prediction tasks than the quantity we care about; can we not depend heavily on the estimation error of these tasks



TripAdvisor Experiment

For random half of 4 million users, easier sign-up flow was enabled

- ❖ Easier sign-up incentivizes membership
- ❖ Outcome: number of visits in the next 14 days



High Level Empirical Findings

- ❖ Large effect heterogeneity based on which pages were recently visited
- ❖ Large effect heterogeneity based on platform of access (e.g. iPhone, Linux etc.)
- ❖ Results enable **better targeting** of right user population and **improvements of membership offering for user segments** with small/almost zero effects

EconML in Action

EconML python library for ML Estimation of Heterogeneous Treatment Effects

- ❖ <https://github.com/microsoft/EconML>
- ❖ `pip install econml`

```
dr_cate = IntentToTreatDRIV(model_y_x=RandomForestRegressor(),
                             model_t_xz=RandomForestClassifier(),
                             prel_model_effect=RandomForestRegressor(),
                             final_model_effect=LinearRegression())
dr_cate.fit(y, T, X, Z)
dr_cate.effect(X)
```

ALICE (Automated Learning and Intelligence for Causation and Economics) project:

- ❖ <https://www.microsoft.com/en-us/research/project/alice/>



Deep Dive 2: Long-term effects of new treatments

Through the lens of a Return-on-Investment (ROI) Case-Study at Microsoft

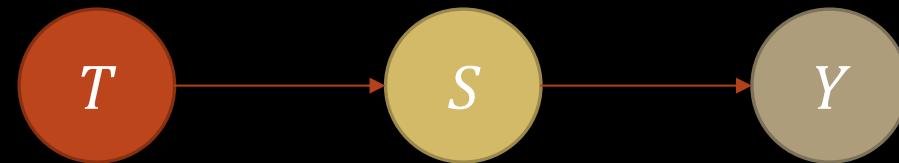
Estimating Long-Term Returns on Investment

- ❖ Companies frequently deploys new discount or customer support programs
- ❖ Which of these programs (“investments”) are more successful than others?
- ❖ Success is a **long-term** objective: what is the effect of the program on the two-year customer journey (e.g., effect on two-year revenue)
- ❖ We cannot wait two years to evaluate a program
- ❖ **Main Question.** Can we construct estimates of the values of these programs with **short-term** data, e.g. after 6 months?



Long-Term Effects from Short-Term Surrogates

- ❖ Suppose that there are many short-term signals S that are indicative of a customer's long-term reward Y (e.g. the next 6-month purchase patterns of a customer could be indicative of their long-term spend)
- ❖ Suppose that investment program T affects long-term rewards if and only if it affects these short-term signals



- ❖ We will call these short-term signals S surrogates

Causal Inference with Surrogates 101

(Prentice, 1989; Begg & Leung, 2000; Frangakis & Rubin, 2002; Freedman et al., 1992; Athey et al., 2020)

- ❖ Since long-term effect goes only through surrogates:

expected effect on long-term reward = effect on projected long-term reward based on surrogates

$$E[Y^{(\tau)}] = E[Y^{(\tau)}|T = \tau] = E[Y|T = \tau] = E[E[Y|T = \tau, S]|T = \tau] = E[E[Y|S]|T = \tau]$$

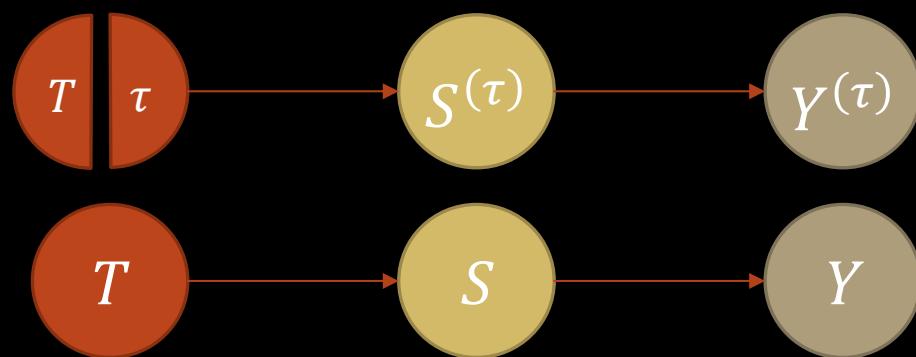
Average reward if intervene and set investment= τ

Independence in counterfactual graph

Average reward of samples that received investment= τ in data

Tower Law of Expectations

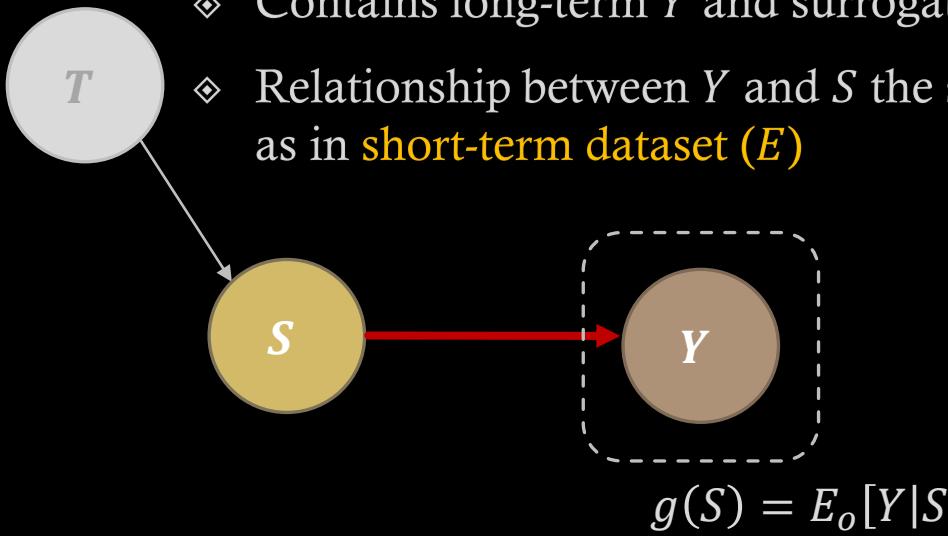
Forecasted reward from surrogates



Causal Inference with Surrogates 101

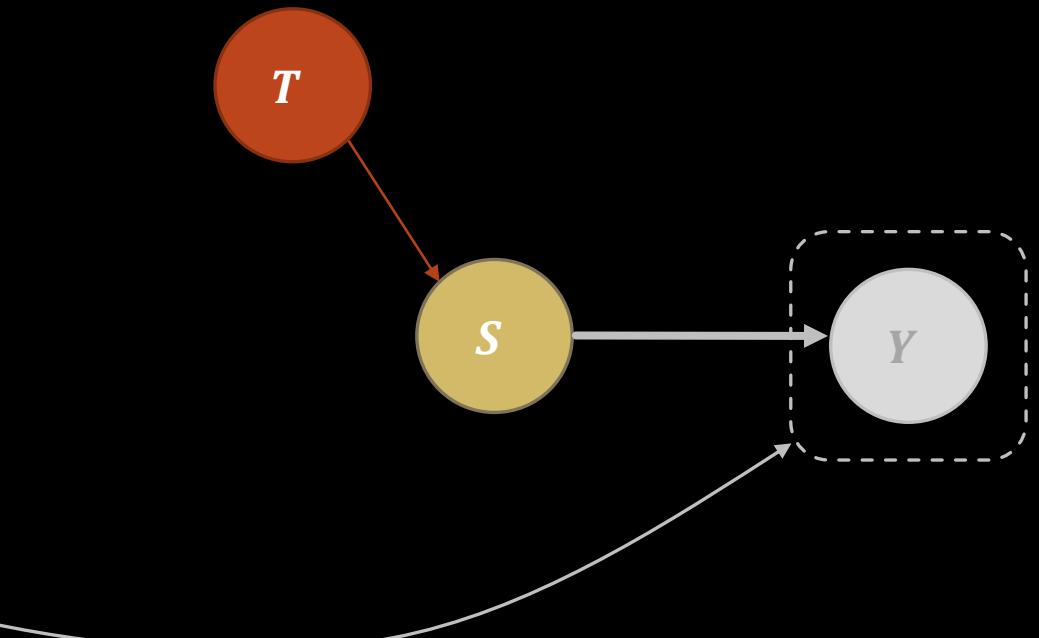
historical/long-term (O)

- ❖ Contains long-term Y and surrogates S
- ❖ Relationship between Y and S the same as in **short-term dataset (E)**



1. Estimate $g(S) := E[Y|S]$ (**surrogate index**) from (O) by regressing $Y \sim S$

recent/short-term (E)



2. Impute expected long-term outcomes in (E)
3. Regress $g(S) \sim T$ to estimate effect of T on Y from (E)

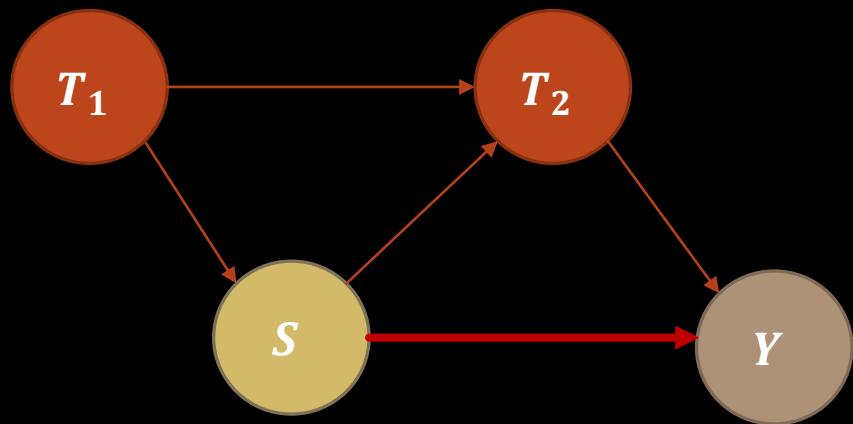
Key Assumptions

- ❖ Long-term effect only goes through surrogates
- ❖ Expected relationship between surrogates and long-term reward is the same long-term setting (O) and in short-term setting (E)

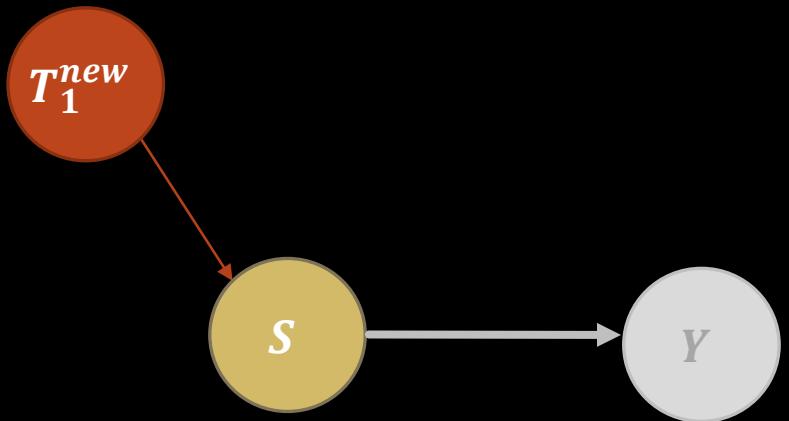
Key Assumptions can be Easily Violated

Investment policies are dynamic and change

historical/long-term (O)



recent/short-term (E)



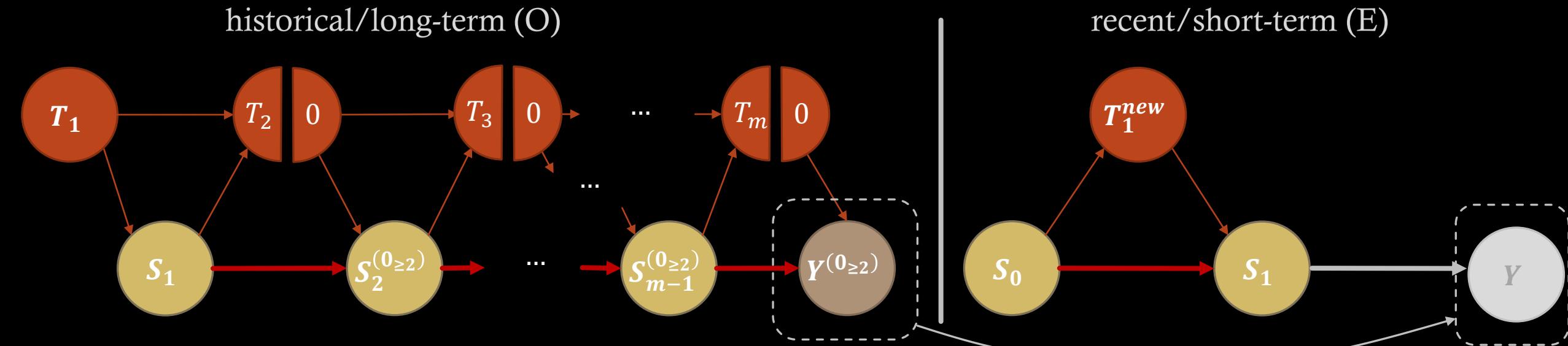
- ◊ We deployed older/deprecated investments
- ◊ In a potentially long-term highly auto-correlated manner
- ◊ Investments are potentially adaptive
- ◊ Investment policies change

Beyond Two Periods



- ❖ Treatments are offered in an adaptive manner, in response to previous period surrogate/state
- ❖ The surrogate – treatment feedback precludes viewing this as a one-shot treatment problem
- ❖ Setting is known as the **dynamic treatment regime** [Robins'94, '04, Chakraborty-Murphy'14]

General Setting



$$g^{adj}(S_1) = E_o[Y^{(0_{\geq 2})} | S_1]$$

$$= E_o \left[Y - \sum_{t \geq 2} \underbrace{\gamma_t(T_t, S_{t-1})}_{\text{Blip effect}} \middle| S_1 \right]$$

“blip” effect of treatment at period t given prior surrogate S_{t-1} and continuing with “zero” treatment subsequently

Semi-parametric restriction (Linear Structural Nested Mean Models, J. Robins '94, '04).

Blip effect has linear parametric form

$$\gamma_m(T_m, S_{m-1}) = \alpha'_m \phi(T_m, S_{m-1})$$

e.g. $\gamma_m(T_m, S_{m-1}) = \alpha_m T_m$ in illustrative example

Machine Learning Estimation

- ❖ In practice, we might have many surrogate variables, or features of them
- ❖ We also want to control for many customer characteristics to ensure un-confoundedness
- ❖ Can easily lead to high-dimensional settings, i.e. S_t is high-dimensional
- ❖ Estimation of several regression models in the above approach cannot be handled

- ❖ We might also want to use ML for effect heterogeneity and automated population partition

Machine Learning and Bias

- ❖ Can we simply replace linear regression with high-dimensional linear regression analogues in these models (e.g. lasso)
- ❖ Can lead to biased effect estimates and non-asymptotically normal estimates
- ❖ Inability to construct confidence intervals for our estimates

- ❖ Developed a novel debiased machine learning inference procedure for this setting

EconML in Action



EconML: Adjusted Surrogate Index



Historical Long-Term Data

Company	Year	Features	Controls/Surrogates	Investment	AdjRevenue
1	A	2018	...	\$1,000	\$9,000
2	A	2019	...	\$2,000	\$7,000
3	A	2020	...	\$3,000	\$10,000
4	B	2018	...	\$0	\$5,000
5	B	2019	...	\$100	\$9,500
6	B	2020	...	\$1,200	\$5,000
7	C	2018	...	\$1,000	\$19,000
8	C	2019	...	\$1,500	\$20,000
9	C	2020	...	\$500	\$10,000

```
from econml.dml import DynamicDML, LinearDML

est = DynamicDML(model_y=LassoCV(), # Any ML model for E[Y_t | S_{tau}]
                  model_t=LassoCV()) # Any ML model for E[T_t | S_{tau}]

# on historical data construct adjusted outcomes
for t in np.arange(1, period): # for each target period 1...m
    est.fit(long(y[:, 1:t]), long(T[:, 1:t]), W=long(W[:, 1:t]), groups=groups)
    yadj[:, t] = y[:, t] - est.effect(T0=0, T1=wide(T[:, 1:t]))

# fit surrogate model using W[:, 1] as surrogates
adj_surr_model = LassoCV().fit(W[:, 1], np.sum(yadj, axis=1))
```

Short-Term Data with New Investment

Company	Year	Features	Controls/Surrogates	New Investment	AdjSurrIndex
1	A	2021	...	\$1,000	\$20,000
2	B	2021	...	\$0	\$15,000
3	C	2021	...	\$1,000	\$40,000

```
# on recent data impute surrogate index
adj_surr_index = adj_surr_model.predict(Wnew[:, 1])

# estimate causal effect on adj_surr_index controlling for Wnew[:, 0]
new_est = LinearDML(model_y=LassoCV(), model_t=LassoCV())
new_est.fit(adj_surr_index, Tnew, W=Wnew[:, 0]).summary()
```

	point_estimate	stderr	zstat	pvalue	ci_lower	ci_upper
cate_intercept	1.289	0.154	8.374	0.0	1.036	1.542

Practical Challenges

- ❖ Unobserved confounding
- ❖ Outlier robustness
- ❖ Asymptotic nature of many inference procedures
- ❖ Interpretability of assumptions and conclusions
- ❖ Measures of “success”
- ❖ Scalability to ultra large data-sets (many features or many samples)

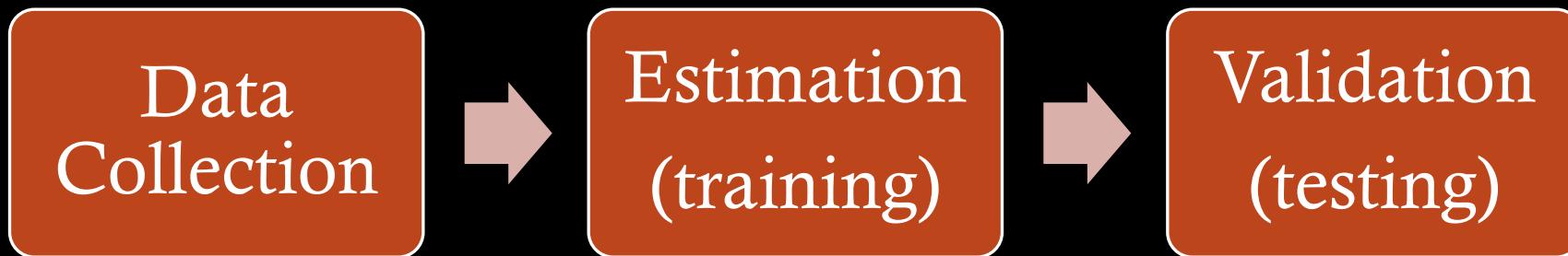


Deep Dive 3: Further tools for reducing barriers to entry

Making CausalML Accessible to Every Decision-Maker

- ❖ Need to make it as accessible as modern AutoML systems
- ❖ Causal inference is inherently harder than prediction

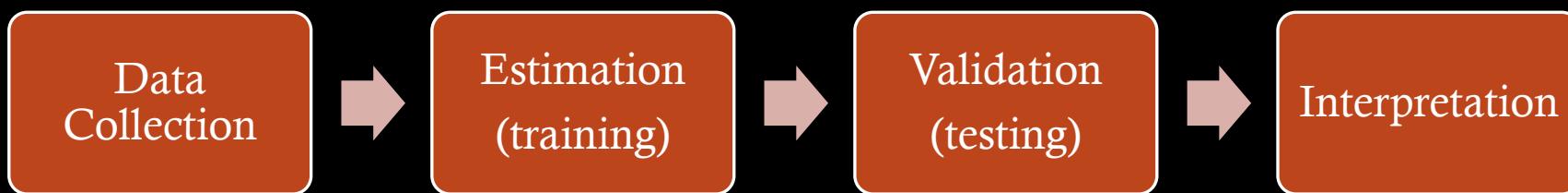
Predictive modelling



Making CausalML Accessible to Every Decision-Maker

- ❖ Need to make it as accessible as modern AutoML systems
- ❖ Causal inference is inherently harder than prediction

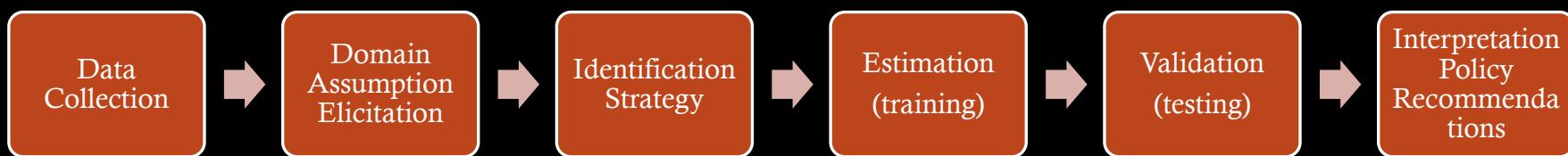
Predictive modelling



Making CausalML Accessible to Every Decision-Maker

- ❖ Need to make it as accessible as modern AutoML systems
- ❖ Causal inference is inherently harder than prediction

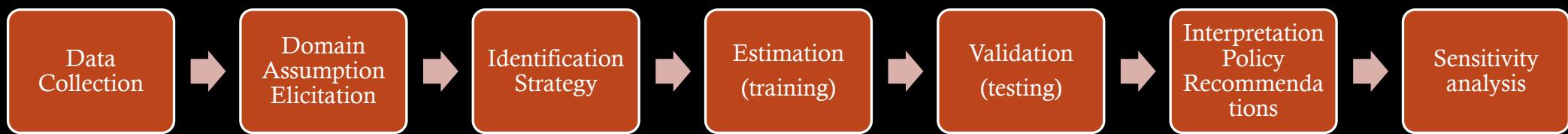
Causal modelling



Making CausalML Accessible to Every Decision-Maker

- ❖ Need to make it as accessible as modern AutoML systems
- ❖ Causal inference is inherently harder than prediction

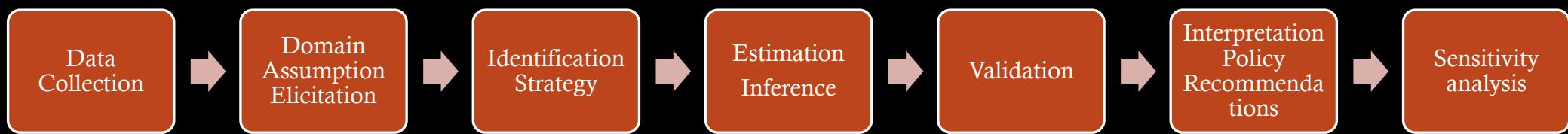
Causal modelling



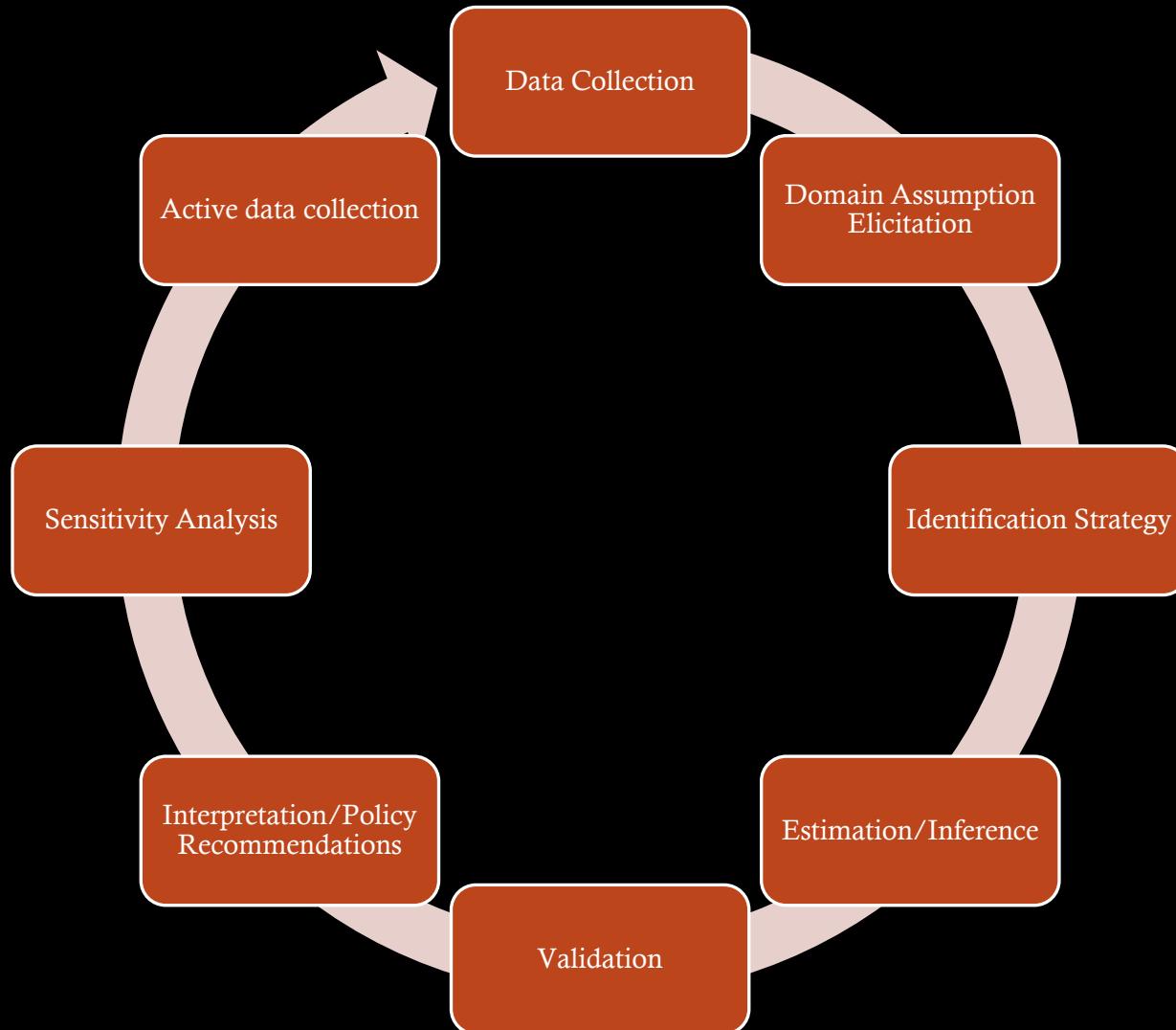
Making CausalML Accessible to Every Decision-Maker

- ❖ Need to make it as accessible as modern AutoML systems
- ❖ Causal inference is inherently harder than prediction

Causal modelling

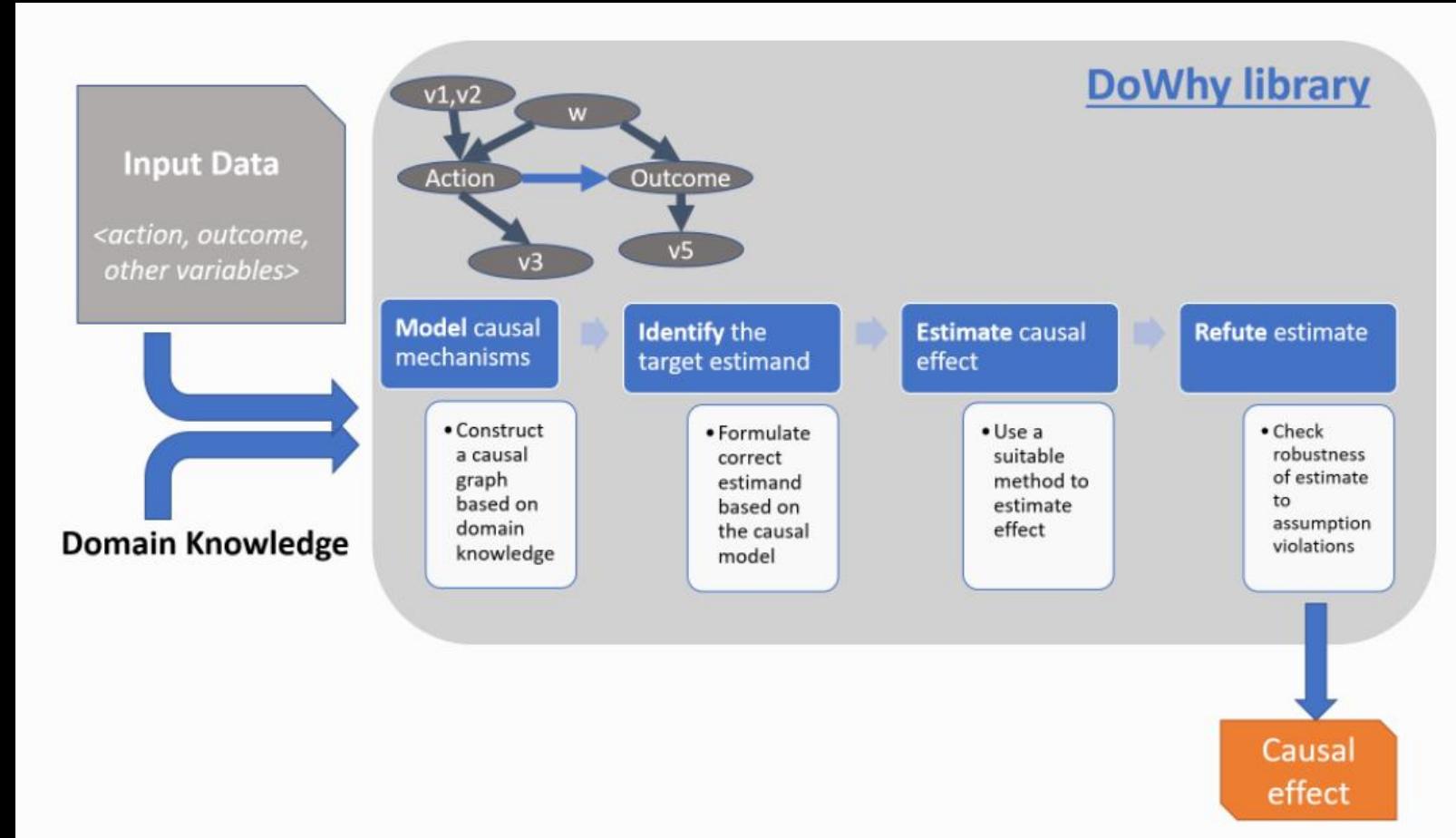


The Causal Inference Pipeline



DoWhy Python Package

Guiding the Data Scientists through the Causal Pipeline



Towards No-Code Interfaces

ShowWhy

For Identified Victims of Trafficking, does Recent Natural Disaster cause Severity of Control to Increase?

Settings ▾ Load ▾ Save ▾ DR

Workflow	Guidance	Workspace
Understand process Why use ShowWhy? This is ShowWhy? Then to use ShowWhy? How does ShowWhy work?	We can refine the causal question to focus on three key concepts, all of which need to be defined in terms of variables either present as or derived from data: Does, <impose>, victim <experience> for <population>?	Population Label: Identified Victims of Trafficking Dataset: CTDC Global Synthetic Dataset Description Victims of trafficking identified by members of the Counter Trafficking Data Collaborative (CTDC) and represented in the CTDC Global Synthetic Dataset. Victims must have reported at least one means of control.
Define question scribe elements define population define exposure define outcome	Done Done Done Done	Exposure Label: Recent Natural Disaster Dataset: UN SDG Indicator Database Description Victim's country of citizenship experienced a recent and severe natural disaster (assumes the victim was present in that country at the time of the disaster). Available from the Global Justifications Development Goals (GJDG) Indicators Database.
Model causal factors Prepare data Perform analysis Create reports		Outcome Label: Severity of Control Dataset: CTDC Global Synthetic Dataset Description Victim experienced at least one form of severe control, defined as physical abuse, sex abuse, denied medical treatment, denied necessities, or giving psychoactive substances, according to collaborators who are domain experts.
		Hypothesis Exposure causes outcome to: <input type="radio"/> Change <input checked="" type="radio"/> Increase <input type="radio"/> Decrease

[Mark as to do](#) [Previous step](#) [Next step](#)

ShowWhy

Workflow

Understand process

- Why use ShowWhy?
- Who is ShowWhy for?
- When to use ShowWhy?
- How does ShowWhy work?

Define question

- Describe elements Done
- Define population Done
- Define exposure Done
- Define outcome Done

Model causal factors

- Consider causal factors Done
- Actors causing exposure Done
- Actors caused by exposure Done
- Actors causing outcome Done
- Actors caused by outcome Done
- Define alternative models Done

Prepare data

Perform analysis

Create reports

Guidance

For Identified Victims of Trafficking, does Recent Natural Disaster cause Severity of Control to Increase?

Workspace

Factors assumed to cause outcome

Factor	Causes outcome?	Degree of Belief	Reasoning	Actions
Victim Age	<input checked="" type="checkbox"/>	Strong	There may be differences in the prevalent forms of control by victim's age group (e.g. adults vs. children).	Edit
Victim Gender	<input checked="" type="checkbox"/>	Strong	Female victims are more likely to experience severe means of control.	Edit
Rule of Law in Country of Exploitation	<input checked="" type="checkbox"/>	Moderate	Victims exploited in countries with high rule of law scores are likely to experience less severe means.	Edit
State of Economy in Country of Exploitation	<input checked="" type="checkbox"/>	Moderate	Victims exploited in countries with a strong economy are likely to experience less severe means.	Edit
Access to Justice in Country of Exploitation	<input checked="" type="checkbox"/>	Moderate	Victims exploited in countries with high access to justice scores are likely to experience less severe means.	Edit
Role of Law in Affected Country	<input checked="" type="checkbox"/>	Moderate	Victims from countries with high rule of law scores are likely to experience less severe means.	Edit
State of Economy in Affected Country	<input checked="" type="checkbox"/>	Moderate	Victims from countries with a strong economy are likely to experience less severe means.	Edit
Access to Justice in Affected Country	<input checked="" type="checkbox"/>	Moderate	Victims from countries with high access to justice scores are likely to experience less severe means.	Edit

Add new factor

Mark as to do

Previous step

Next step

ShowWhy

For Identified Victims of Trafficking, does Recent Natural Disaster cause Severity of Control to Increase?

Settings ▾ Load ▾ Save ▾

Workflow

Understand process

- Why use ShowWhy?
- Who is ShowWhy for?
- Then to use ShowWhy?
- Our does ShowWhy work?

Define question

- Describe elements Done
- Define population Done
- Define exposure Done
- Define outcome Done

Model causal factors

- Consider causal factors Done
- Actors causing exposure Done
- Actors caused by exposure Done
- Actors causing outcome Done
- Actors caused by outcome Done
- Address alternative models Done

Prepare data

Perform analysis

Create reports

Guidance

Based on your identification of possible causal factors, we can generate a range of candidate causal models with different levels of complexity:

- The maximum model including all edges with strong, moderate or weak degree of belief
- The intermediate model including all edges with a strong or moderate degree of belief
- The minimum model including all edges with a strong degree of belief only
- The unadjusted model including only the causal edge from exposure to outcome

An edge in this context is a directed graph connection (i.e., an arrow) indicating the direction of causal influence between two nodes. Any edges indicated as potentially existing in either direction are excluded from all models as being inconclusive from the perspective of domain knowledge. For our present form of causal inference aiming to emulate a randomized controlled trial, we need to control for two kinds of causal factors: confounders influencing exposure and outcome and outcome determinants influencing the outcome only.

Once you have uploaded and prepared data, you'll have the option to check whether your model is consistent with the data and whether there is evidence to support the directionality of any inconclusive (or unspecified) edges.

Workspace

Causal model: Maximum Intermediate Minimum Unadjusted

Controls

```
graph TD; Confounders[Confounding  
Rule of Law in Affected Country  
State of Economy in Affected Country  
Access to Justice in Affected Country] -- Causes --> Exposure[Exposure  
Recent Natural Disaster]; Confounders -- Causes --> Outcome[Outcome  
Severity of Control]; Outcome determinants[Outcome determinants  
Victim Age  
Victim Gender  
Rule of Law in Country of Exploitation  
State of Economy in Country of Exploitation  
Access to Justice in Country of Exploitation] -- Causes --> Outcome;
```

Click this button to verify if the estimate is possible based on the overall structure of causal relationships specified.
This might take a minute.

Verify possibility

[Mark as to do](#) [Previous step](#) [Next step](#)

ShowWhy

For Identified Victims of Trafficking, does Recent Natural Disaster cause Severity of Control to Increase?

Specification curve analysis of causal effect estimates

Estimated change in severity of control by specification

Median effect (0.11)
Mean effect (0.09)

Element contribution

Specification ID

Legend:

- Maximum model
- Minimum model
- Median model
- Unmodified model
- Estimator
- Invert propensity weighting estimator
- Linear doubly robust learner estimator
- Linear regression estimator
- Propensity score stratification
- Records complete in source data population
- Records completed by imputation exposure
- Experienced top 25% disaster
- Experienced top 35% disaster exposure
- Experienced top 50% disaster

Mark as done Previous step Next step

Workflow

Understand process

- Why use ShowWhy? Done
- What is ShowWhy? Done
- Then to use ShowWhy? Done
- How does ShowWhy work? Done

Define question

- Describe elements Done
- Define population Done
- Define exposure Done
- Define outcome Done

Model causal factors

- Consider causal factors Done
- Actors causing exposure Done
- Actors caused by exposure Done
- Actors causing outcome Done
- Actors caused by outcome Done
- Define alternative models Done

Prepare data

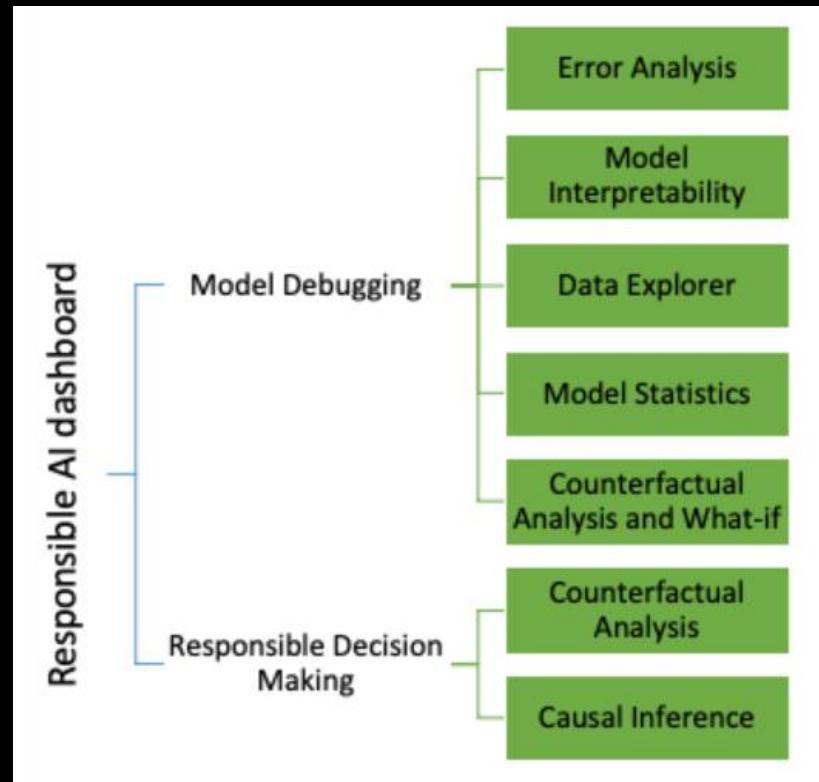
- Add data tables Done
- Process table columns Done
- Specification variables Done
- Exposure variables Done
- Outcome variables Done
- Control variables Done

Perform analysis

- Select causal estimators Done
- Select refutation tests Done
- Estimate causal effects Done
- Explore specification curve To do
- Evaluate hypothesis To do

Create reports

AzureML Responsible AI Dashboard



AzureML Responsible AI Dashboard

Model Debugging via Responsible AI dashboard

Identify

Diagnose

Mitigate



Error Analysis

Identify cohorts with high error rate versus benchmark and visualize how the error rate distributes



Model Statistics

Aggregate a variety of model Assessment metrics, showing model prediction distributions



Model Interpretability

Interpret and debug model.



Counterfactual Analysis and What If

Generate diverse counterfactual explanations for debugging. Perform feature perturbations



Exploratory Data Analysis

Understand dataset characteristics



Unfairness Mitigation

Mitigate fairness issues (via Fairlearn.org)



Data Enhancements

Enhance your dataset and retrain model

AzureML Responsible AI Dashboard

Decision Making via Responsible AI dashboard

Understand data



Inform Actions

 Exploratory-Data-Analysis

Understand dataset characteristics.



Causal Inference

Understand the causal impact of your features on real-world outcomes



Counterfactual Analysis

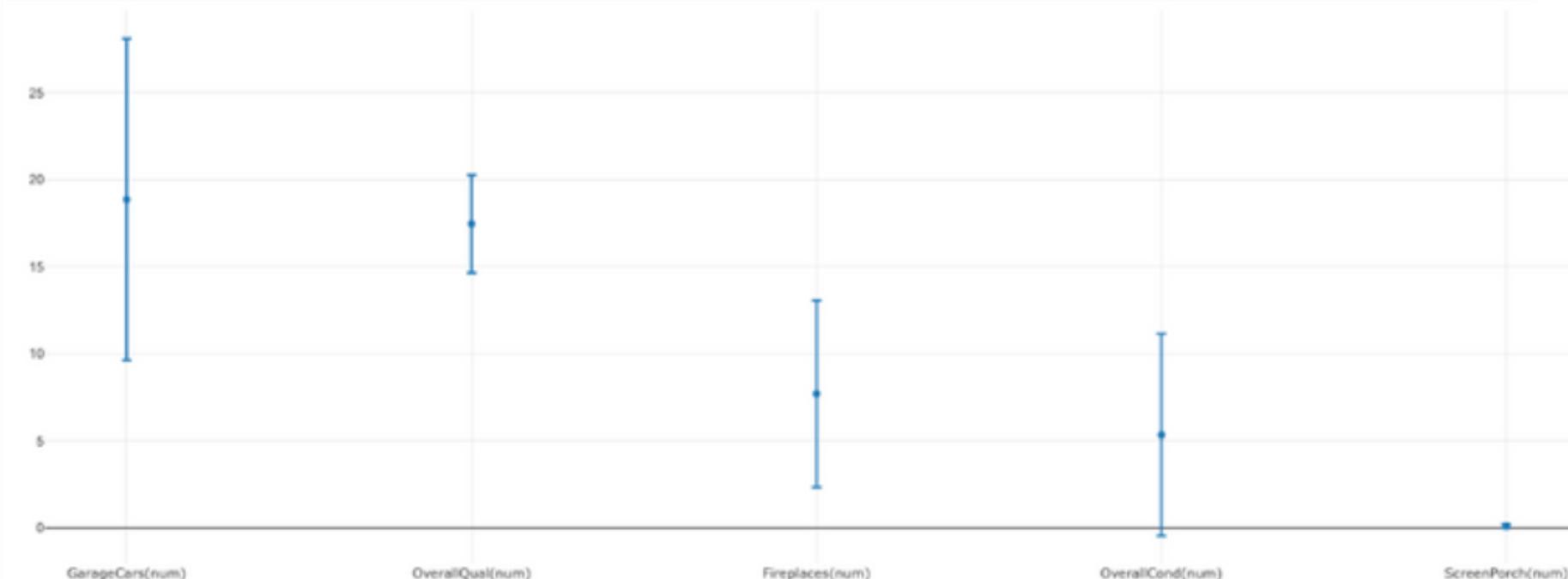
Generate diverse counterfactual explanations for informing end users.

Causal analysis answers "what if?" questions about how real world outcomes would have changed under different policy choices, such as a different pricing strategy for a product or an alternative treatment for a patient. Unlike model predictions that identify important correlation patterns, these tools help you identify the most important causal features that directly affect your outcome of interest. These models identify the causal effect of one feature (typically referred to as a "treatment"), holding other confounding features constant. For best results, make sure that the full dataset contains all available features that may correlate with the outcome as confounders.

Direct aggregate causal effect of each treatment with 95% confidence interval

① Why is it important to include confounding features?

Feature	Effect estimate	Standard error	Z-score	P-value	Confidence interval ...	Confidence interval (upper)
GarageCars(num)	1.885e+1	4.712e+0	4.001e+0	6.306e-5	9.618e+0	2.809e+1
OverallQual(num)	1.745e+1	1.432e+0	1.219e+1	3.564e-34	1.464e+1	2.025e+1
Fireplaces(num)	7.686e+0	2.734e+0	2.812e+0	4.927e-3	2.328e+0	1.304e+1
OverallCond(num)	5.329e+0	2.960e+0	1.800e+0	7.180e-2	-4.725e-1	1.113e+1
ScreenPorch(num)	8.297e-2	5.433e-2	1.619e+0	1.054e-1	-1.851e-2	1.944e-1



Continuous treatments: On average in this sample, increasing this feature by 1 unit will cause the probability of class/label 1 to increase by X units.

Binary treatments: On average in this sample, turning on this feature will cause the probability of class/label 1 to increase by X units.

A lasso (or logistic regression if y is binary) was fit to predict y from $X[-i]$, and a lasso (or logistic regression if $X[i]$ is categorical) was fit to predict $X[i]$ from $X[-i]$. The causal effect can be viewed as the average correlation of the residuals/surprise variation of the two prediction tasks. Learn more about Double Machine Learning [here](#).

These tools help build policies for future interventions. You can identify what parts of your sample experience the largest responses to changes in causal features, or treatments, and construct rules to define which future populations should be targeted for particular interventions.

Set treatment feature

ScreenPorch ▼

Interpretable recommended global treatment policy for sample size (n) = 730

	EnclosedPorch <= 15	EnclosedPorch > 15
OverallCond <= 6.5	n = 508 Recommended treatment = increase	n = 73 Recommended treatment = decrease
	OpenPorchSF <= 151	OpenPorchSF > 151
OverallCond > 6.5	n = 138 Recommended treatment = decrease	n = 11 Recommended treatment = increase

This table shows a recommended treatment policy that can be applied to the current data sample or other populations. The table provides a simple rule to segment observations into data cohorts based on the features with the largest impact on whether the individual will respond to the selected treatment. The table also specifies number datapoints in the current data sample assigned to each segment.

The table can be read by taking a row and then taking a column of that specific row.

Causal analysis

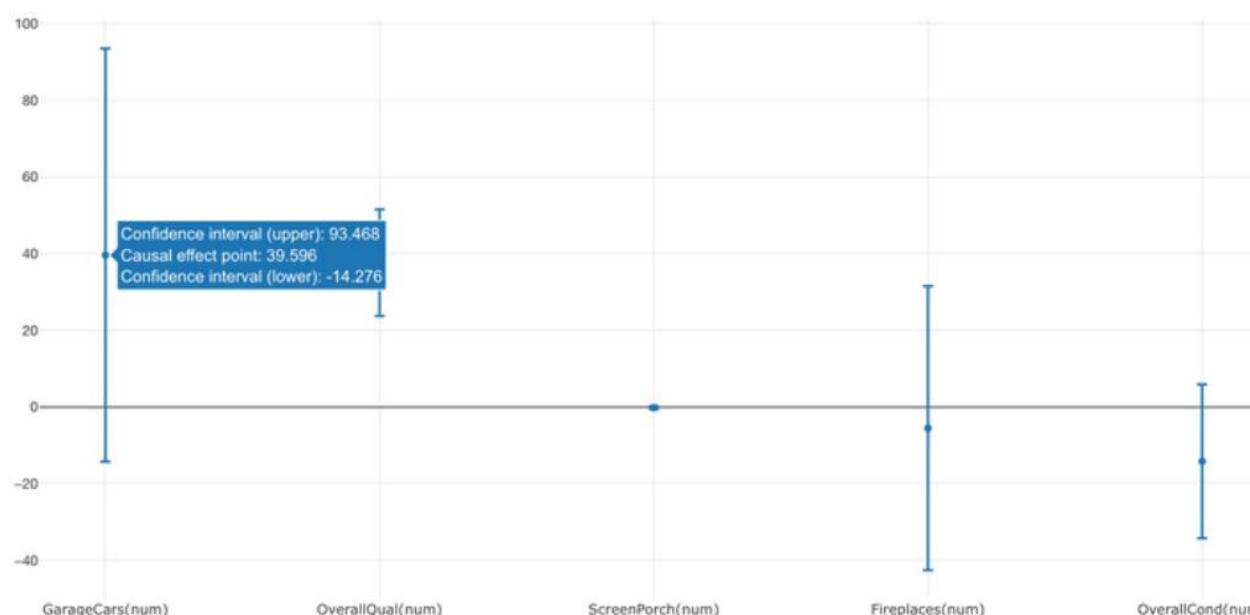
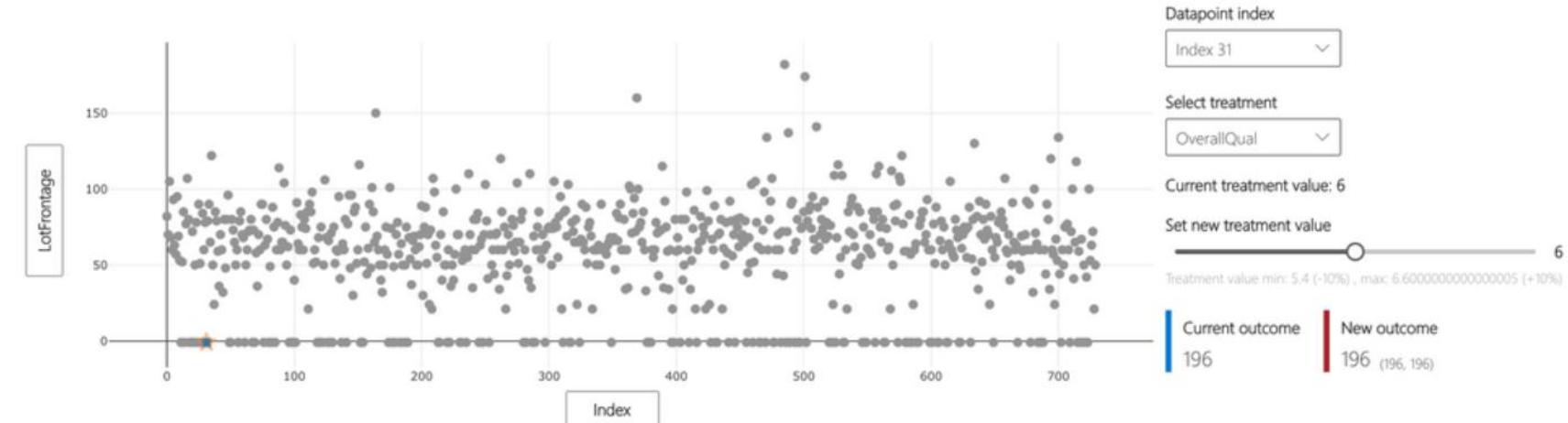
ⓘ Global cohorts are not currently supported for causal analysis. All causal analyses will be shown for all data.

Aggregate causal effects

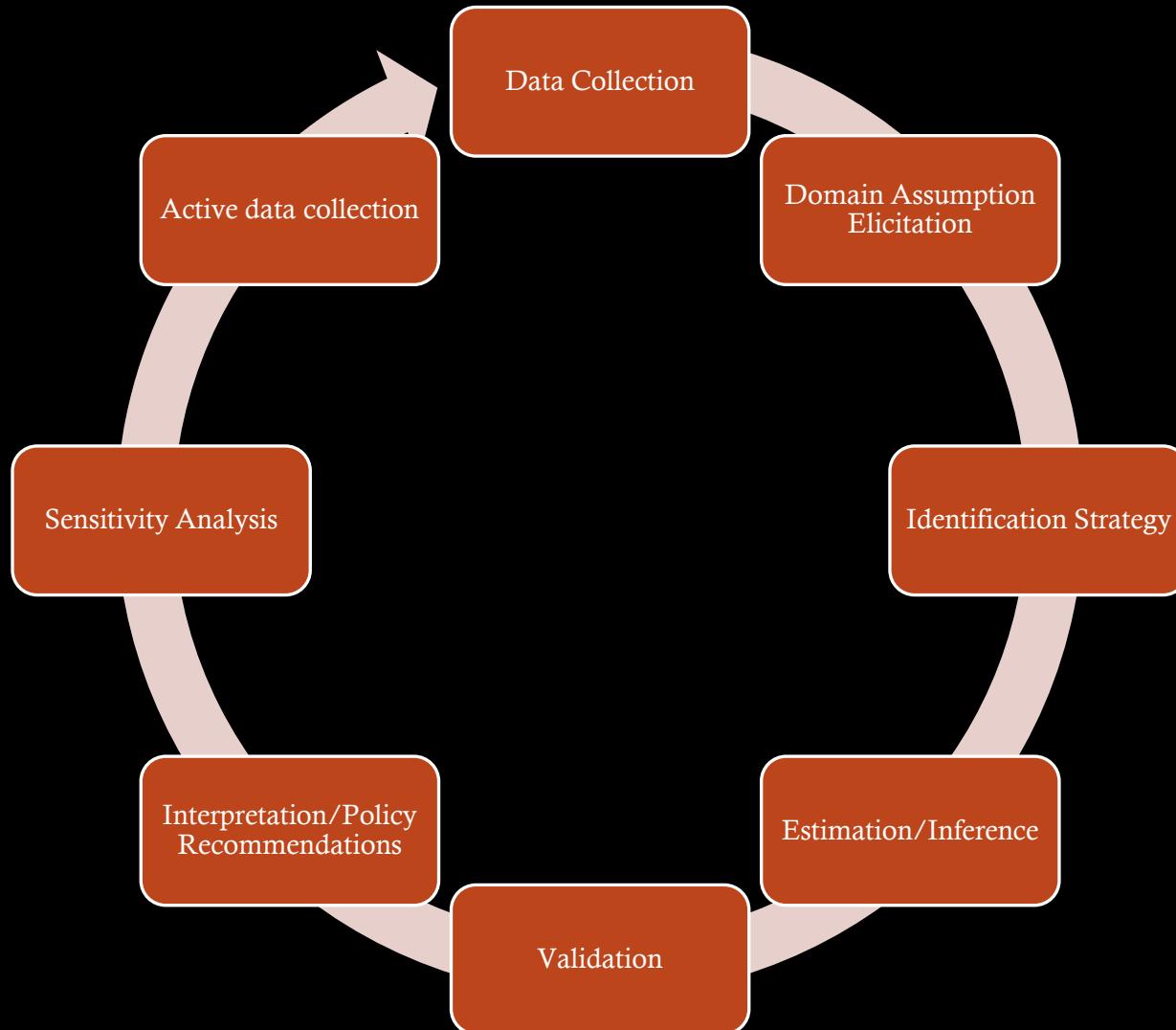
Individual causal what-if

Treatment policy

Individual causal effects can inform personalized interventions, such as a targeted promotion to customers or an individualized treatment plan. How would an individual with a particular set of features respond to a change in a causal feature, or treatment? The causal what-if tool calculates marginal changes in real-world outcomes for a particular individual if you change their level of a treatment. This analysis enables you to understand how real-world outcomes would have changed under different policy choices, such as a different pricing strategy for a product or an alternative treatment for a patient. Specify the treatment of interest and observe how the real-world outcome would change.



The Causal Inference Pipeline



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

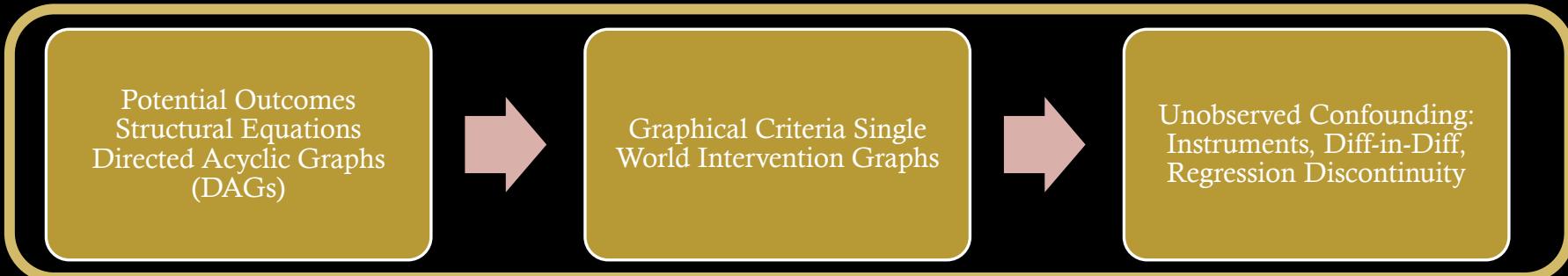
Incorporate non-graphical restrictions in automated graph-based identification algorithms

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

Unobserved confounding

Post adaptive data collection inference
Post causal discovery inference

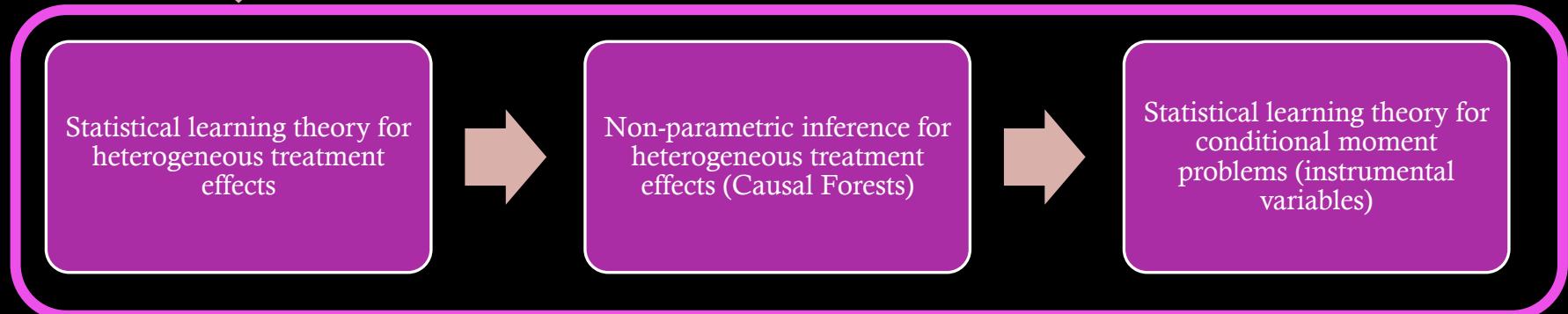
Causal Identification



Estimation and
Inference on
Causal
Parameters



Estimation and Inference on Causal Functions



An aerial photograph of a long bridge spanning across a body of water with a vibrant turquoise hue. The bridge features a multi-lane highway with several vehicles, including trucks and cars, traveling in both directions. The water has a fine, wavy texture.

Thank you!
