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

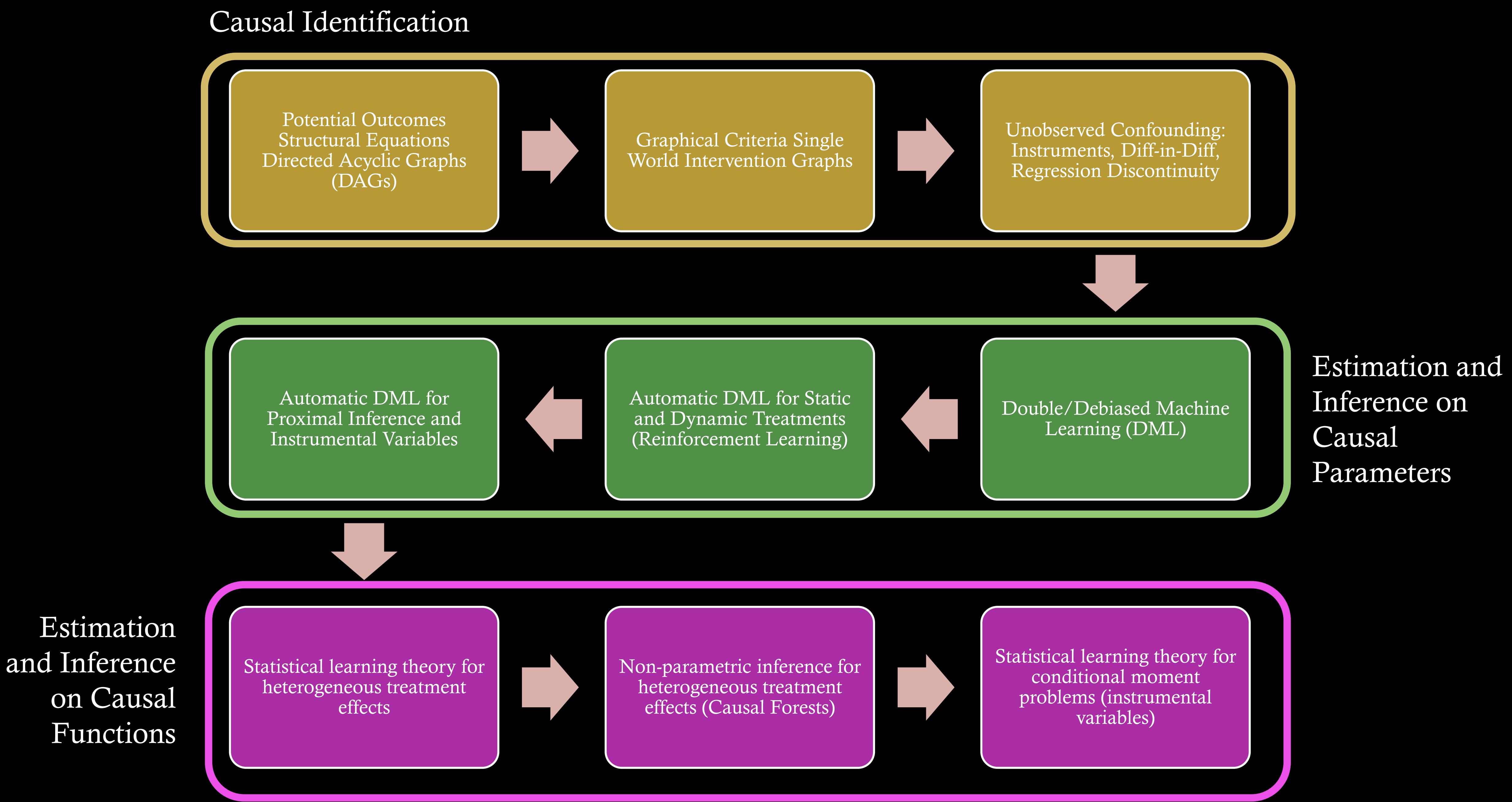
Logistics

- **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 1.30-4.20, 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
Potential Outcomes Structural Equations Directed Acyclic Graphs (DAGs)
Graphical Criteria Single World Intervention Graphs
Unobserved Confounding: Instruments, Diff-in-Diff, Regression Discontinuity
Estimation and Inference on Causal Parameters
Double/Debiased Machine Learning (DML)
Automatic DML for Static and Dynamic Treatments (Reinforcement Learning)
Automatic DML for Proximal Inference and Instrumental Variables
Estimation and Inference on Causal Functions
Statistical learning theory for heterogeneous treatment effects
Non-parametric inference for heterogeneous treatment effects (Causal Forests)
Statistical learning theory for conditional moment problems (instrumental variables)

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

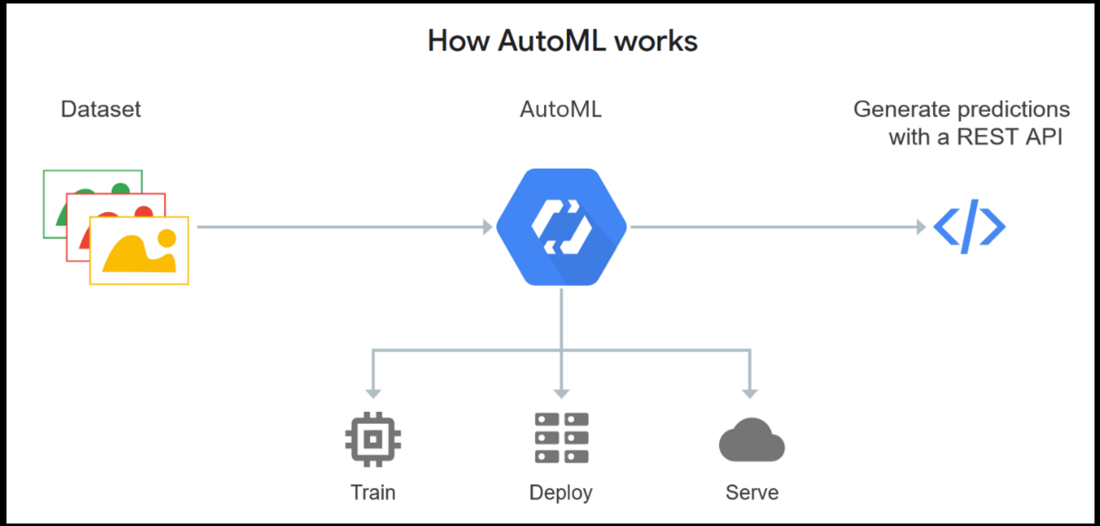
Introduction
Causal Machine Learning in Practice
Case Studies and Challenges

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

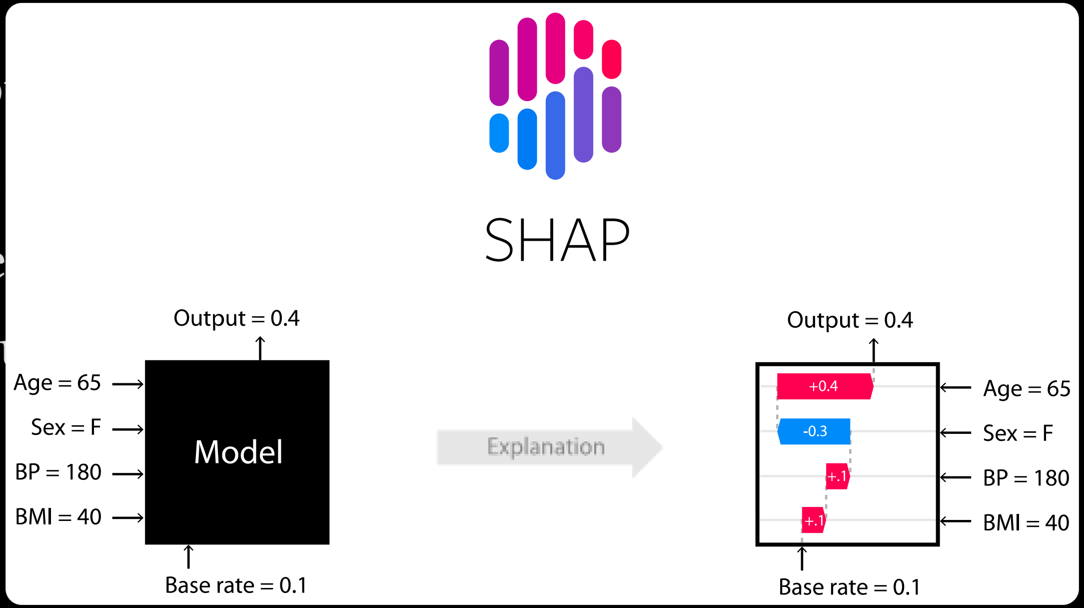


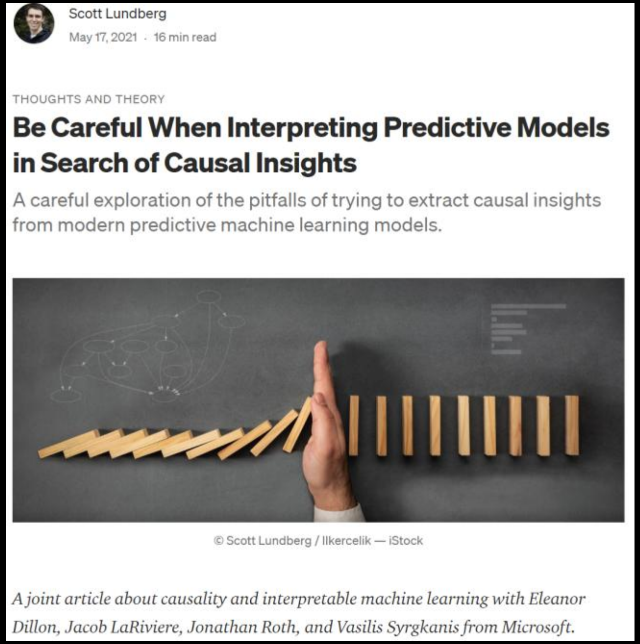
The Ease of (mis)Use of Machine Learning

- 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

Anecdo

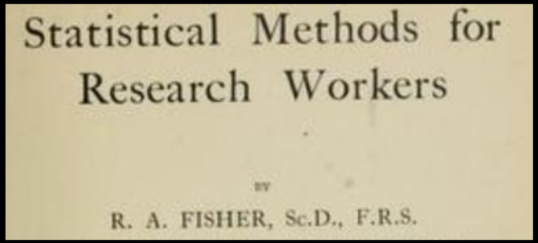
- Bugs
- Price
- Imm





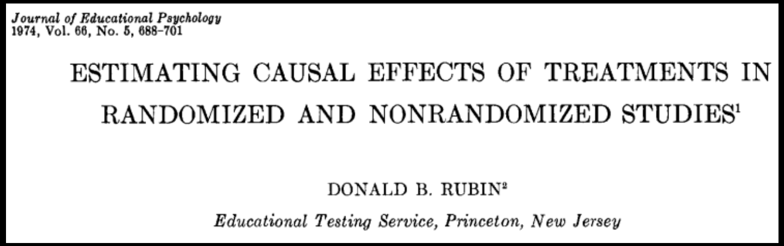
Causal Inference

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


Statistical Methods for Research Workers

BY

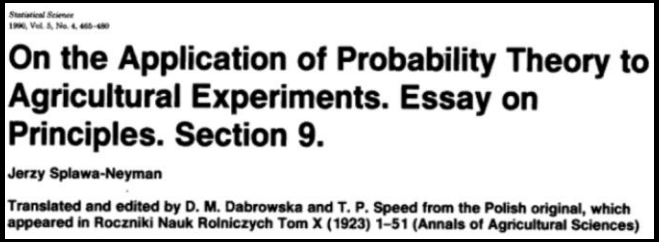
R. A. FISHER, Sc.D., F.R.S.


Journal of Educational Psychology
1974, Vol. 66, No. 5, 688–701

ESTIMATING CAUSAL EFFECTS OF TREATMENTS IN RANDOMIZED AND NONRANDOMIZED STUDIES[1]

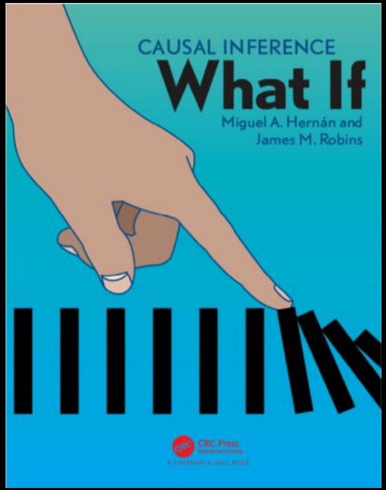
DONALD B. RUBIN[2]

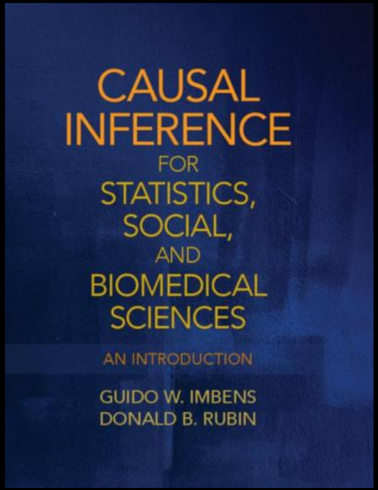
Educational Testing Service, Princeton, New Jersey

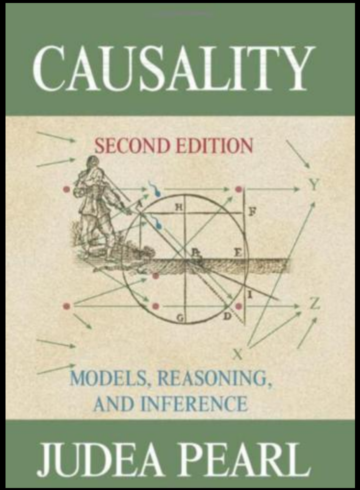

On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.

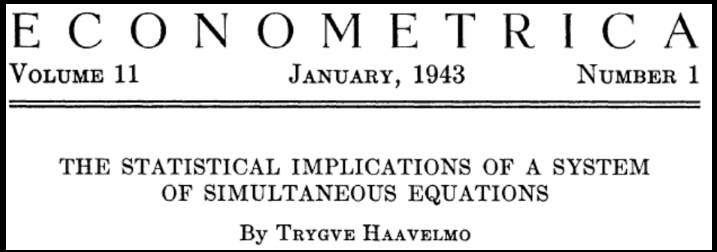
Jerzy Splawa-Neyman

Translated and edited by D. M. Dabrowska and T. P. Speed from the Polish original, which appeared in Roczniki Nauk Rolniczych Tom X (1923) 1–51 (Annals of Agricultural Sciences)








ECONOMETRICA

VOLUME 11 JANUARY, 1943 NUMBER 1

THE STATISTICAL IMPLICATIONS OF A SYSTEM OF SIMULTANEOUS EQUATIONS

BY TRYGVE HAAVELMO

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

(**A**utomated **L**earning and **I**ntelligence for **C**ausation and **E**conomics)

- 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

Microsoft Walmart Save money. Live better. Electrolux

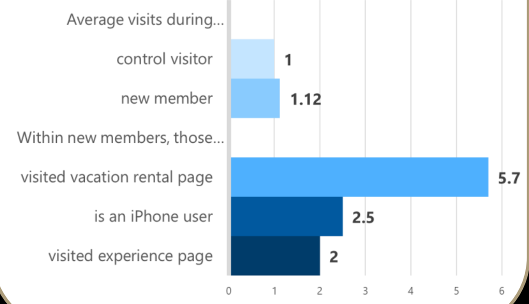
downloads 704k | Starred 1.9k | downloads/month 88k

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

Heterogeneous Effect of Membership



Using Non-Linear Models to Study Aerosol-Cloud Interactions in the Southeast Pacific

Environmental Science & Technology — Machine Learning-Aided Causal Inference Framework for Environmental Data Analysis: A COVID-19 Case Study

Yuya Sugano · Mar 22, 2020 · 13 min read — A World of Causal Inference with EconML by Microsoft Research

The Role of "Live" in Livestreaming Markets: Evidence Using Orthogonal Random Forest

nature › scientific reports › articles › article — Youth well-being predicts later academic success

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

EconML with R in the Tidyverse — Example Usage with Multiple Continuous Treatment, Multiple Outcome Observational Data — Itamar Caspi

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

Combinatorial Polycation Synthesis and Causal Machine Learning Reveal Divergent Polymer Design Rules for Effective pDNA and Ribonucleoprotein Delivery

Empirical Estimates on Hand Manipulation are Recoverable: A Step Towards Individualized and Explainable Robotic Support in Everyday Activities

Causal Inference and COVID-19 Nursing Home Patie… Identifying Factors That Reduced Mortality Risk

Zillow · Humana · Grab · Google[X] · ZS · Spotify — Data Scientist

AIM — 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

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

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

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

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

ML

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

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

Predict	unexplained variation in outcome	from	exogenous variation of treatment induced by instrument	using	linear model where the coefficient can vary with observable characteristics
	$\tilde{Y}$		$\Delta(Z,X)$		$\theta(X)\cdot\Delta(Z,X)$

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]

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

Debiasing Correction: uses quantities that ideally are mean zero

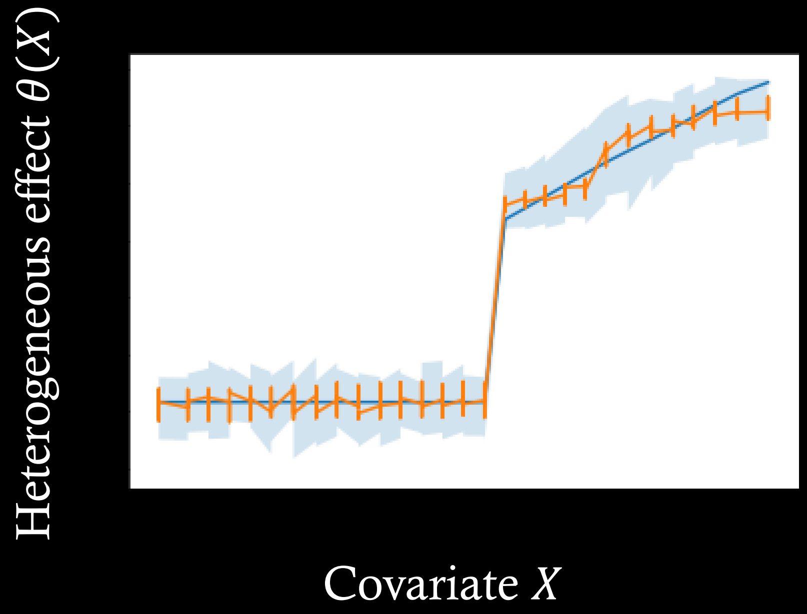
- **Robust heterogeneous treatment effect**

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

ML

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

Inference



- 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]

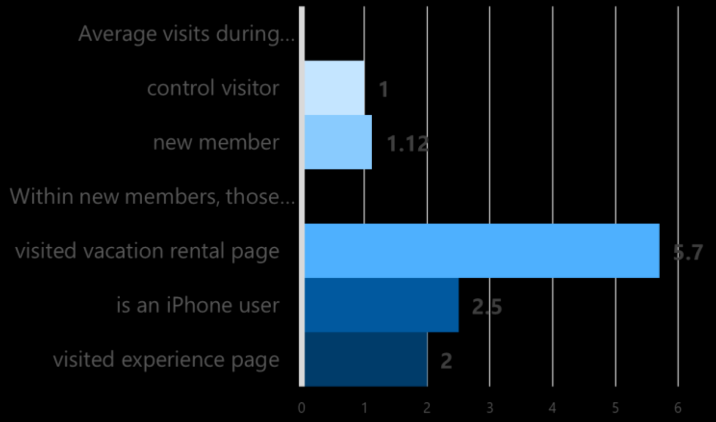
Some Caution

- 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

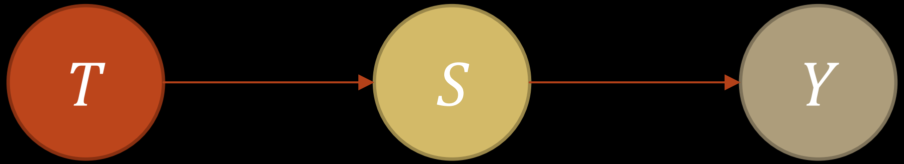
Through the lens of a Return-on-Investment (ROI) Case-Study at Microsoft

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

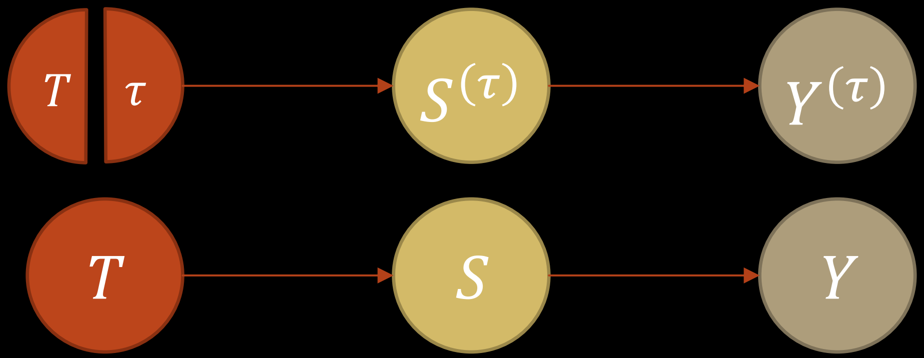
Causal Inference with Surrogates 101

(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

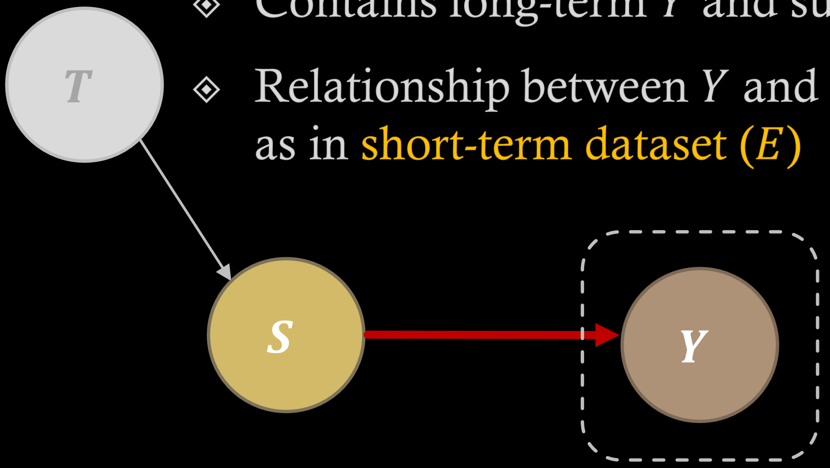
$$\underbrace{E\left[Y^{(\tau)}\right]}_{\substack{\text{Average reward}\\ \text{if intervene and}\\ \text{set investment}=\,\tau}} = \underbrace{E\left[Y^{(\tau)} \middle| T=\tau\right]}_{\substack{\text{Independence in}\\ \text{counterfactual}\\ \text{graph}}} = \underbrace{E[Y|T=\tau]}_{\substack{\text{Average reward of}\\ \text{samples that received}\\ \text{investment}=\tau\text{ in data}}} = \underbrace{E[E[Y|T=\tau,S]|T=\tau]}_{\substack{\text{Tower Law of}\\ \text{Expectations}}} = E[\underbrace{E[Y|S]}_{\substack{\text{Forecasted}\\ \text{reward from}\\ \text{surrogates}}}|T=\tau]$$



Causal Inference with Surrogates 101

historical/long-term (O)

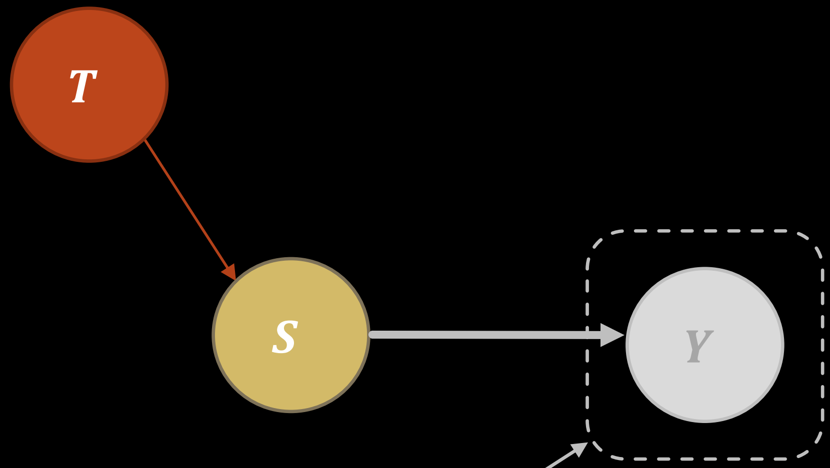
- Contains long-term Y and surrogates S
- Relationship between Y and S the same as in short-term dataset (E)



$g(S) = E_o[Y|S]$

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

recent/short-term (E)



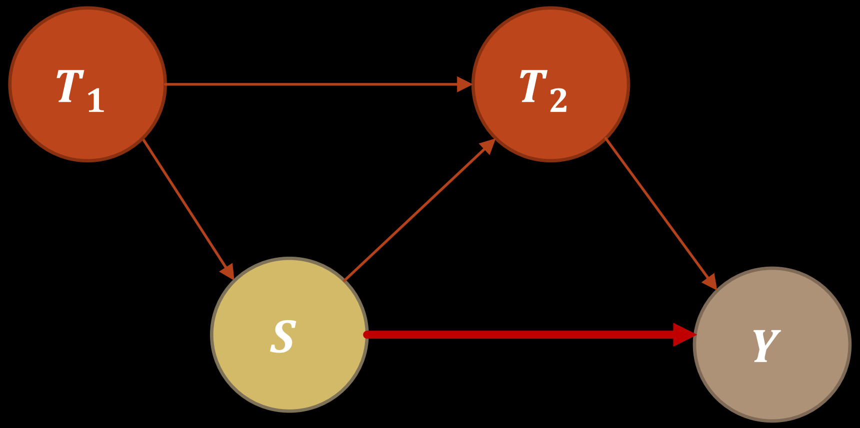
2. Impute expected long-term outcomes in (E)
3. Regress $g(S) \sim T$ to estimate effect of T on Y from (E)

Key Assumptions

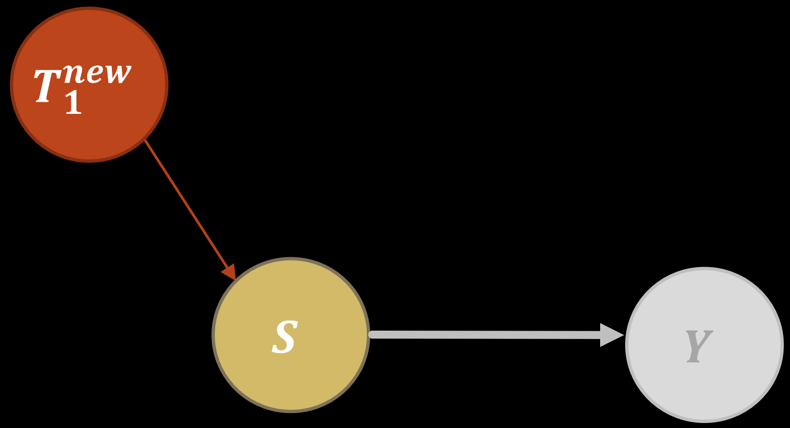
- Long-term effect only goes through surrogates
- Expected relationship between surrogates and long-term reward is the same long-term setting (O) and in short-term setting (E)

Key Assumptions can be Easily Violated
Investment policies are dynamic and change

historical/long-term (O)

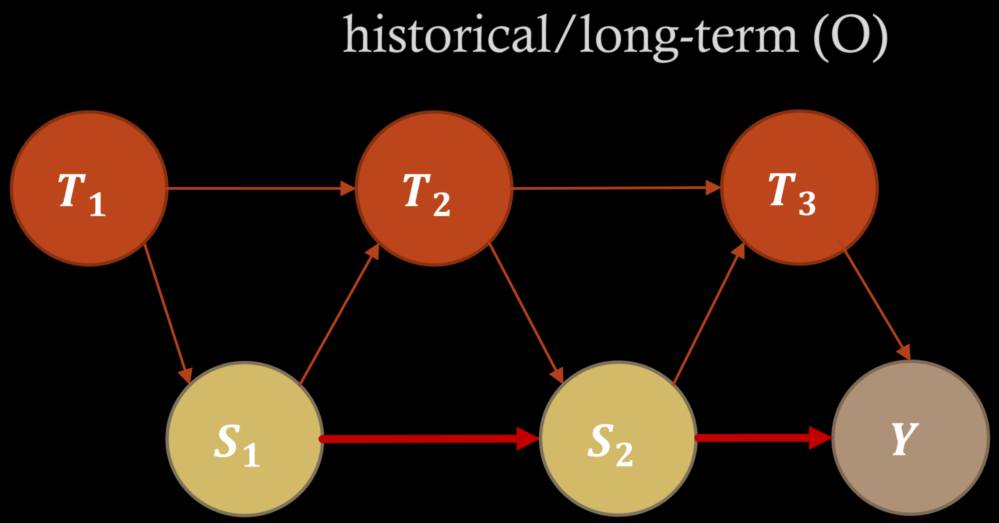


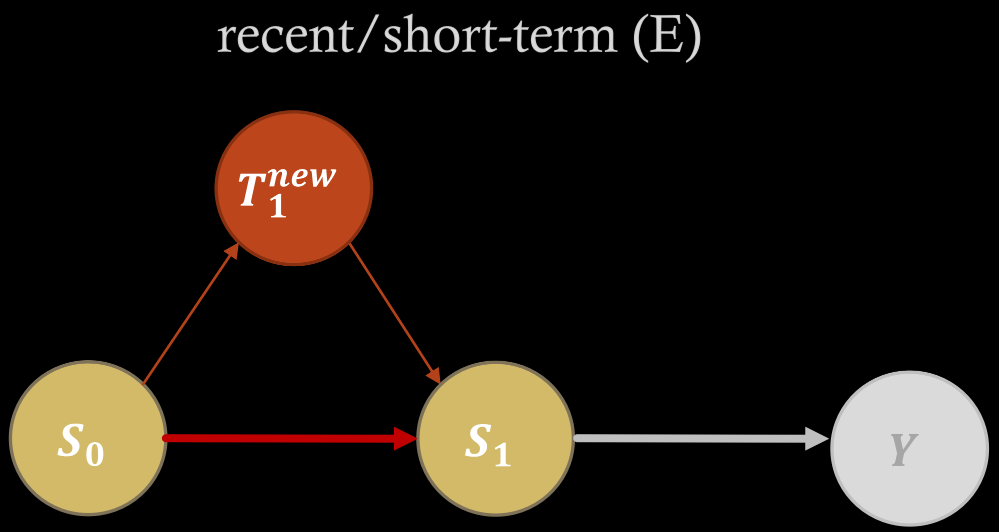
recent/short-term (E)



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

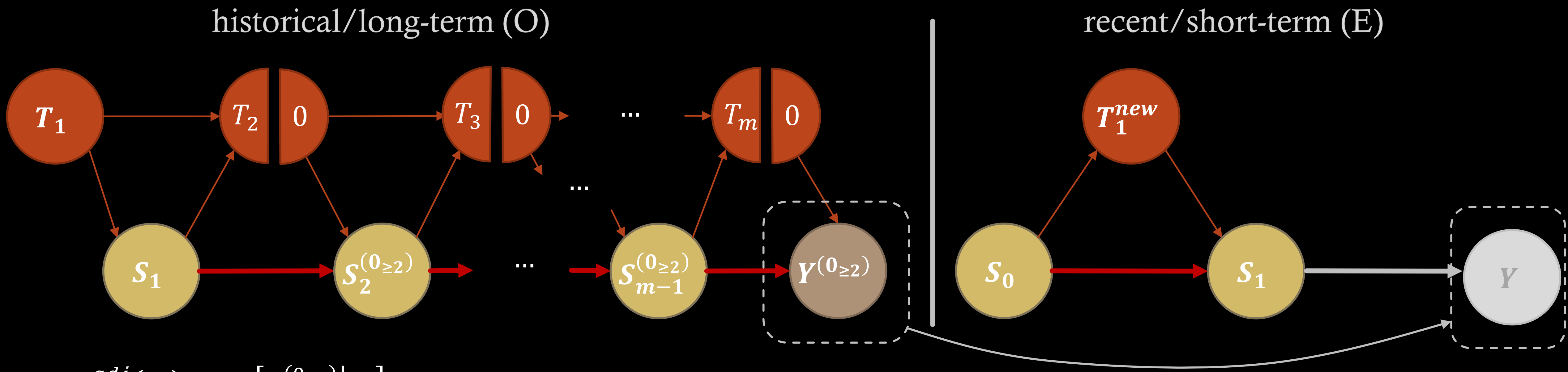
Beyond Two Periods





- 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



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

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

"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

Data Collection → Estimation (training) → Validation (testing)

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

Data Collection → Estimation (training) → Validation (testing) → Interpretation

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

Data Collection → Domain Assumption Elicitation → Identification Strategy → Estimation (training) → Validation (testing) → Interpretation Policy Recommendations

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

Data Collection → Domain Assumption Elicitation → Identification Strategy → Estimation (training) → Validation (testing) → Interpretation Policy Recommendations → Sensitivity analysis

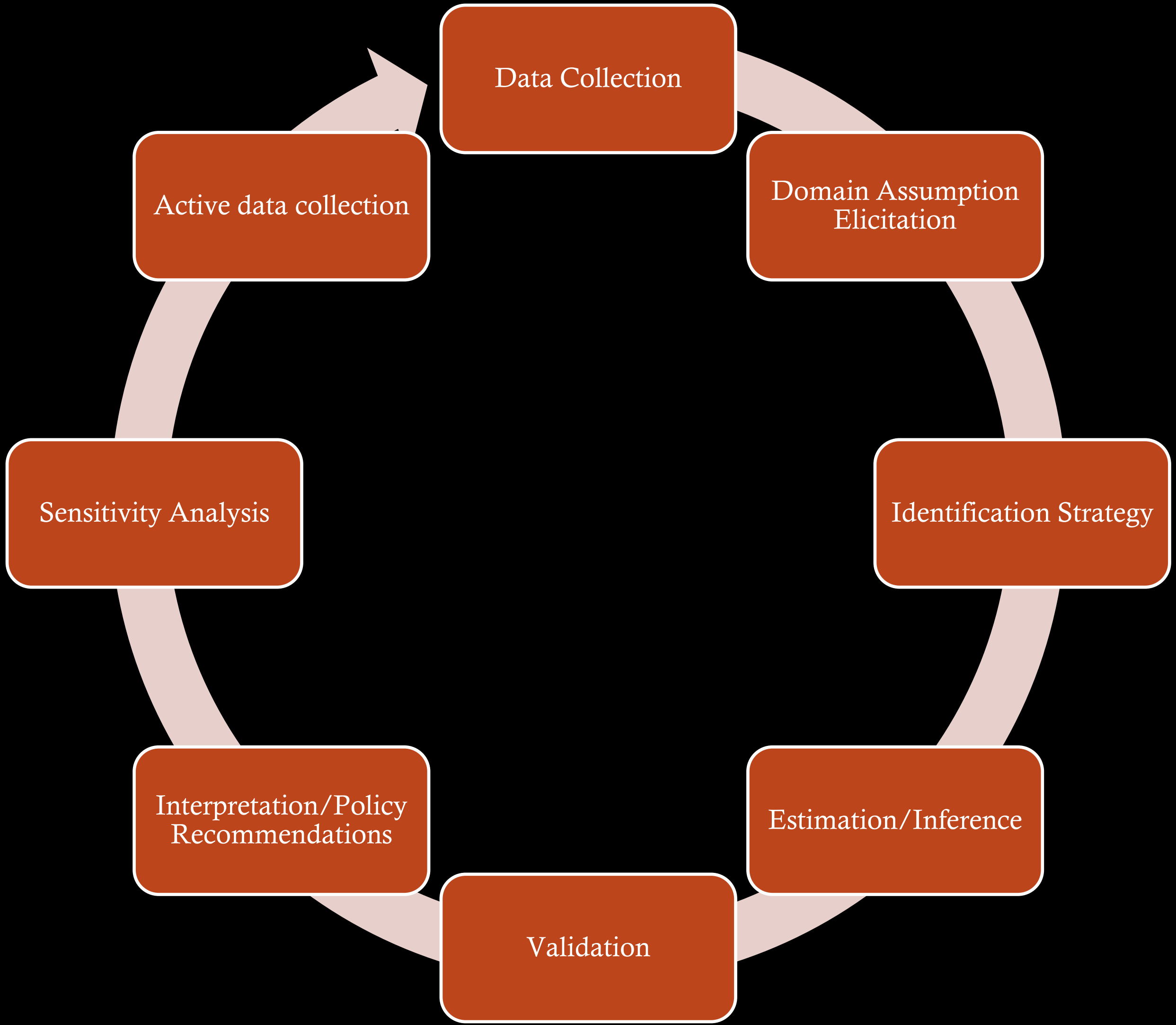
Making CausalML Accessible to Every Decision-Maker

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

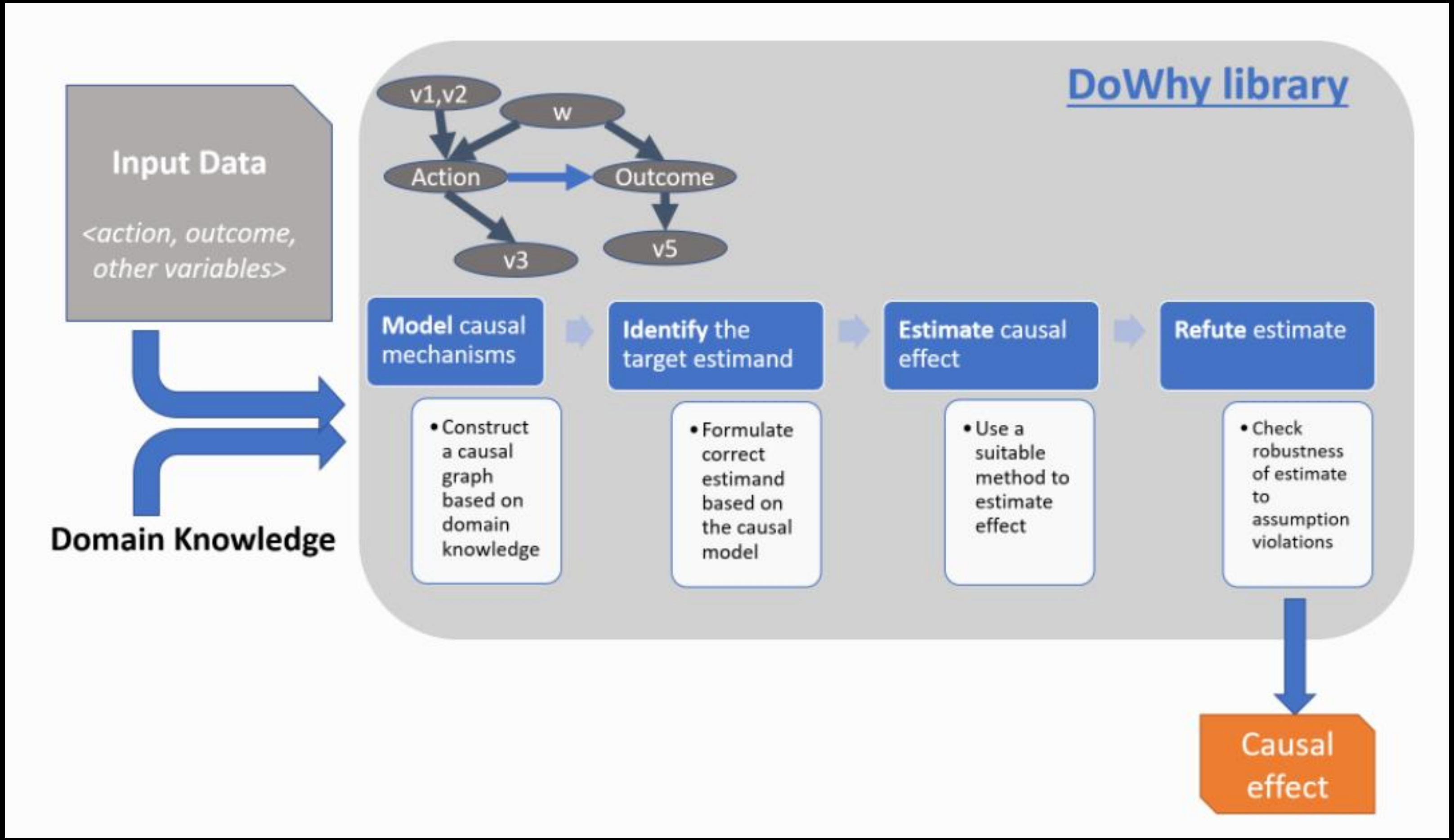
Causal modelling

Data Collection → Domain Assumption Elicitation → Identification Strategy → Estimation Inference → Validation → Interpretation Policy Recommendations → Sensitivity analysis

The Causal Inference Pipeline



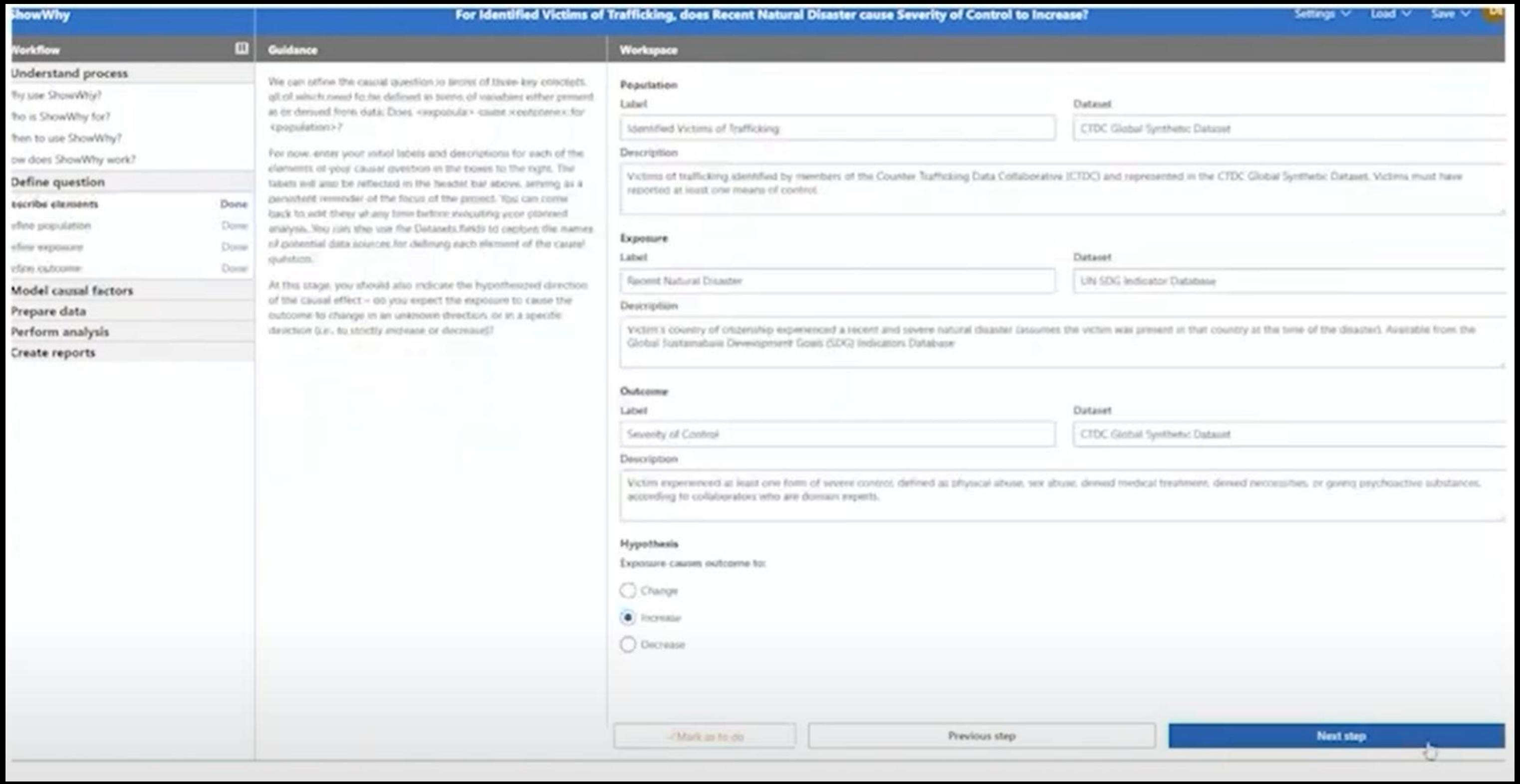
DoWhy Python Package
Guiding the Data Scientists through the Causal Pipeline



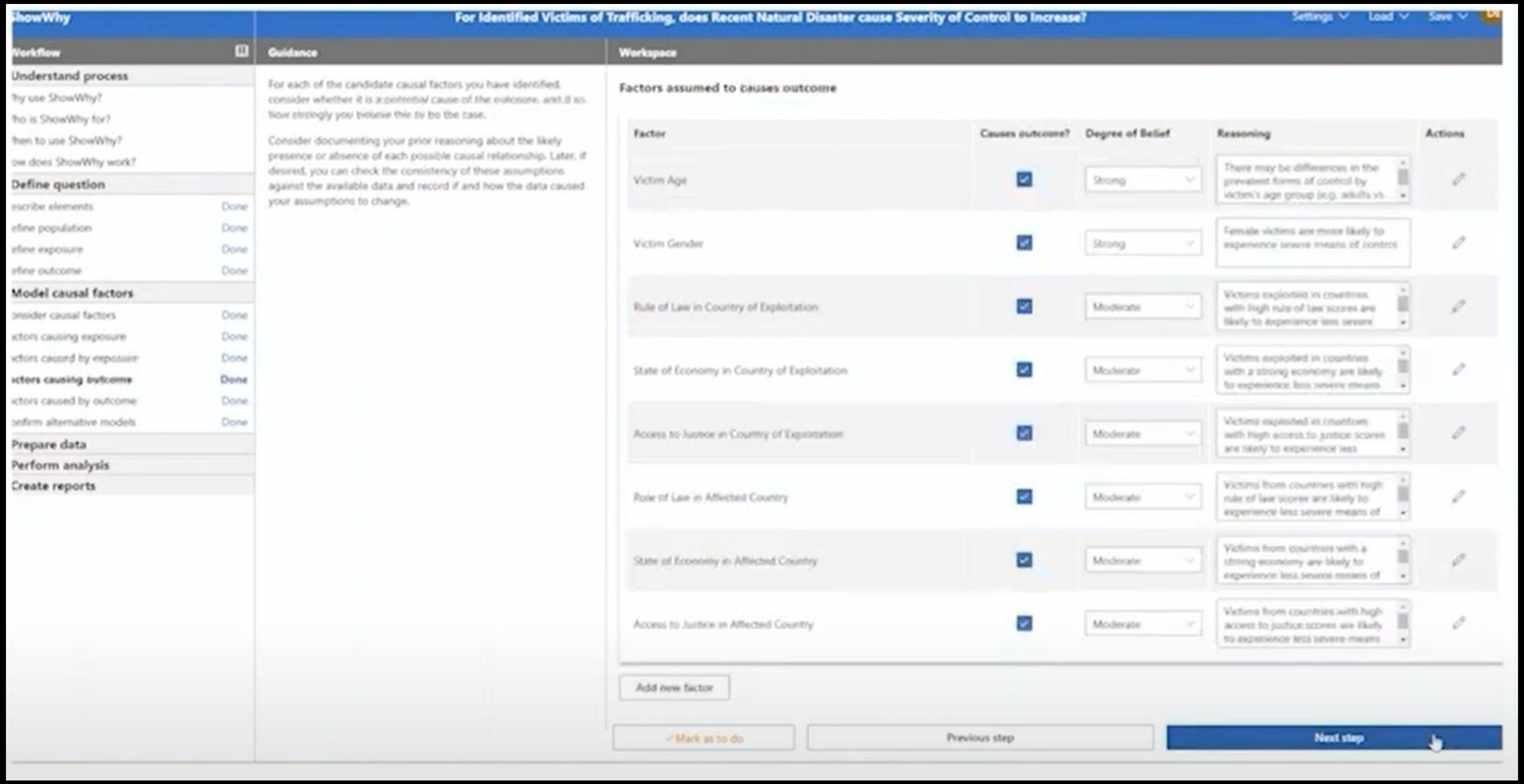
[2108.13518] DoWhy: Addressing Challenges in Expressing and Validating Causal Assumptions (arxiv.org)

Towards No-Code Interfaces

ShowWhy

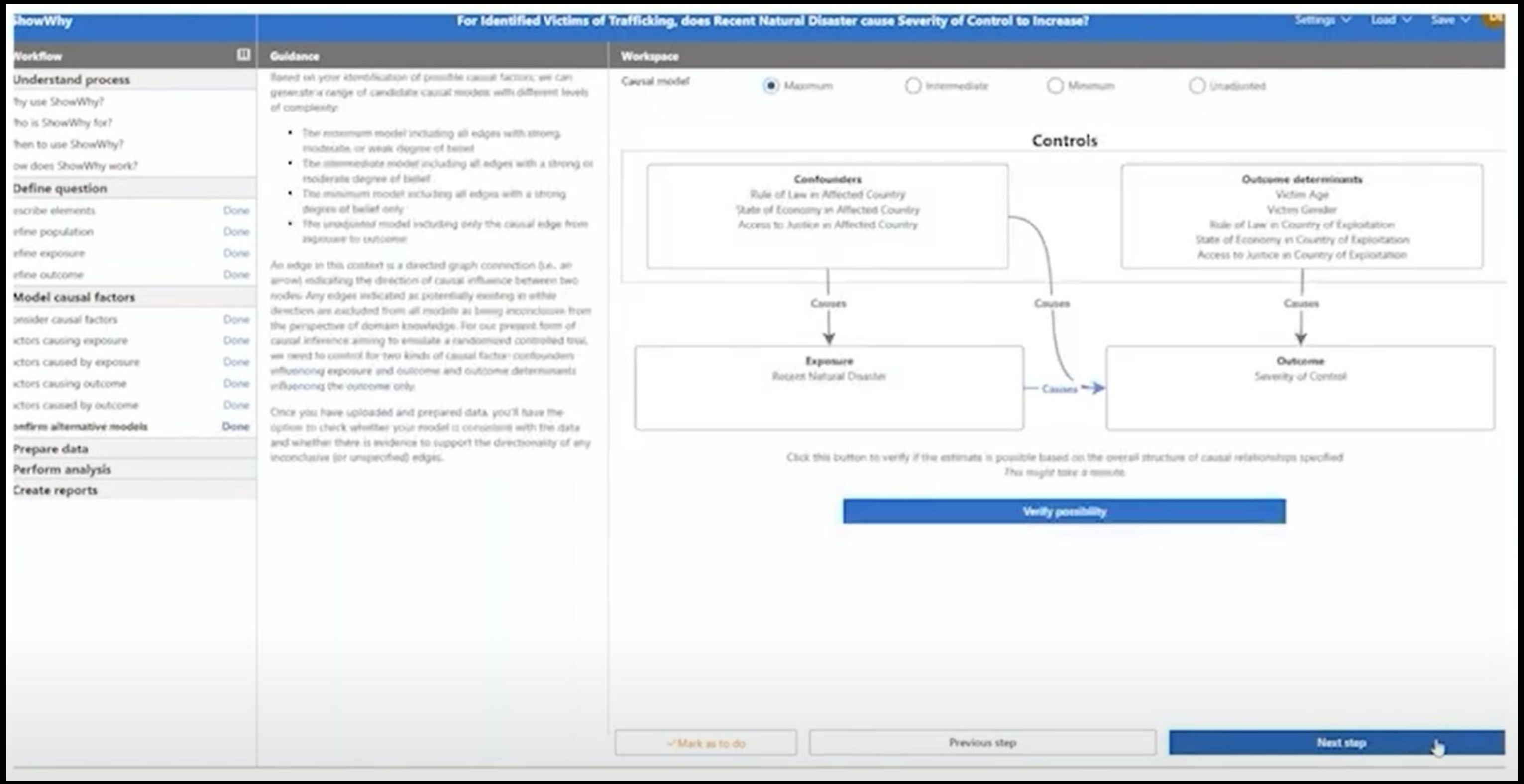

ShowWhy
For Identified Victims of Trafficking, does Recent Natural Disaster cause Severity of Control to Increase?
Workflow
Guidance
Workspace
Understand process
Define question
Model causal factors
Prepare data
Perform analysis
Create reports
Population
Label
Identified Victims of Trafficking
Dataset
CTDC Global Synthetic Dataset
Description
Exposure
Recent Natural Disaster
UN SDG Indicator Database
Outcome
Severity of Control
Hypothesis
Change
Increase
Decrease
Previous step
Next step

ShowWhy



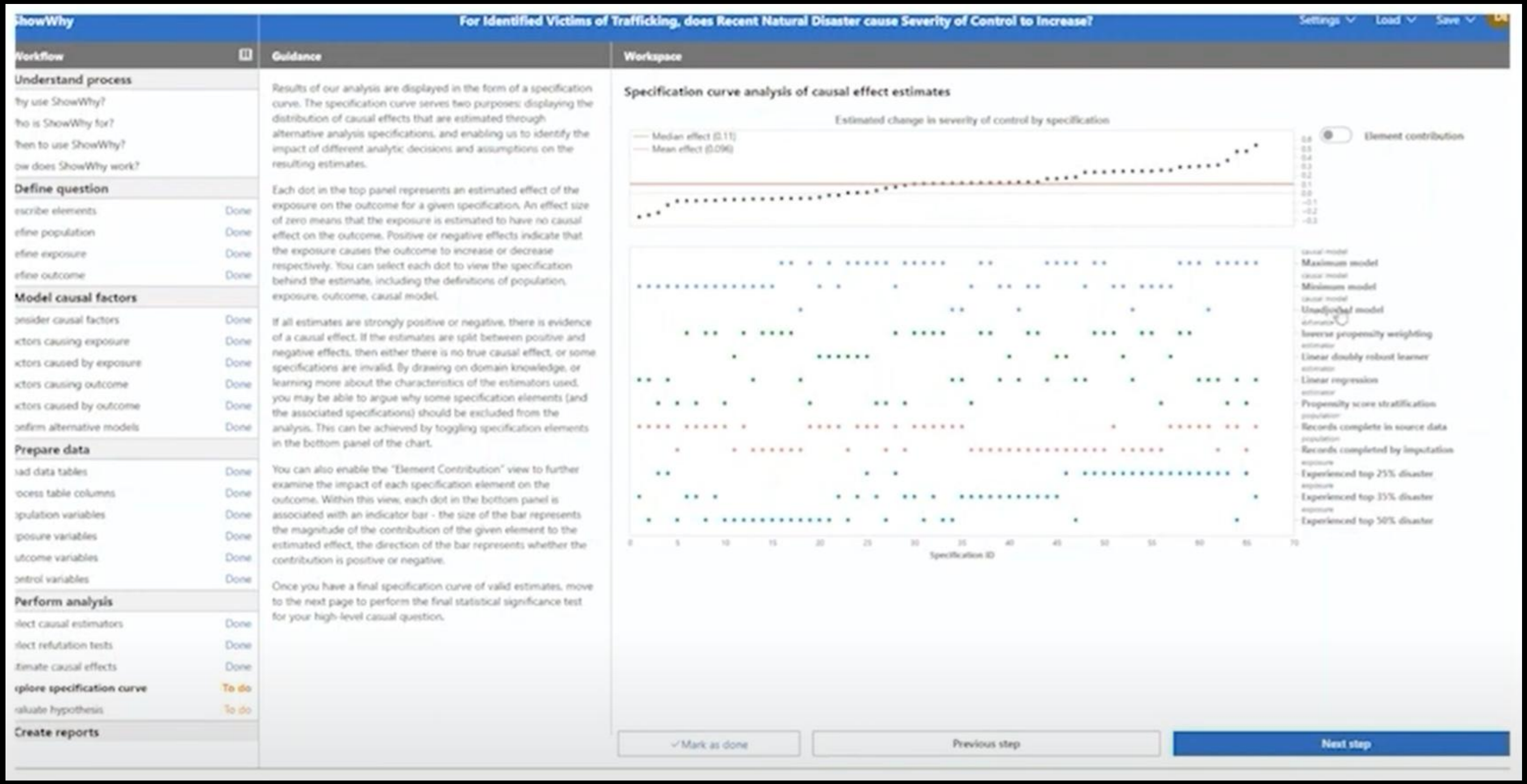
For Identified Victims of Trafficking, does Recent Natural Disaster cause Severity of Control to Increase?
Settings
Load
Save
Workflow
Understand process
Define question
Model causal factors
Prepare data
Perform analysis
Create reports
Guidance
Workspace
Factors assumed to causes outcome
Factor
Causes outcome?
Degree of Belief
Reasoning
Actions
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
Rule of Law in Affected Country
State of Economy in Affected Country
Access to Justice in Affected Country
Add new factor
Previous step
Next step

ShowWhy



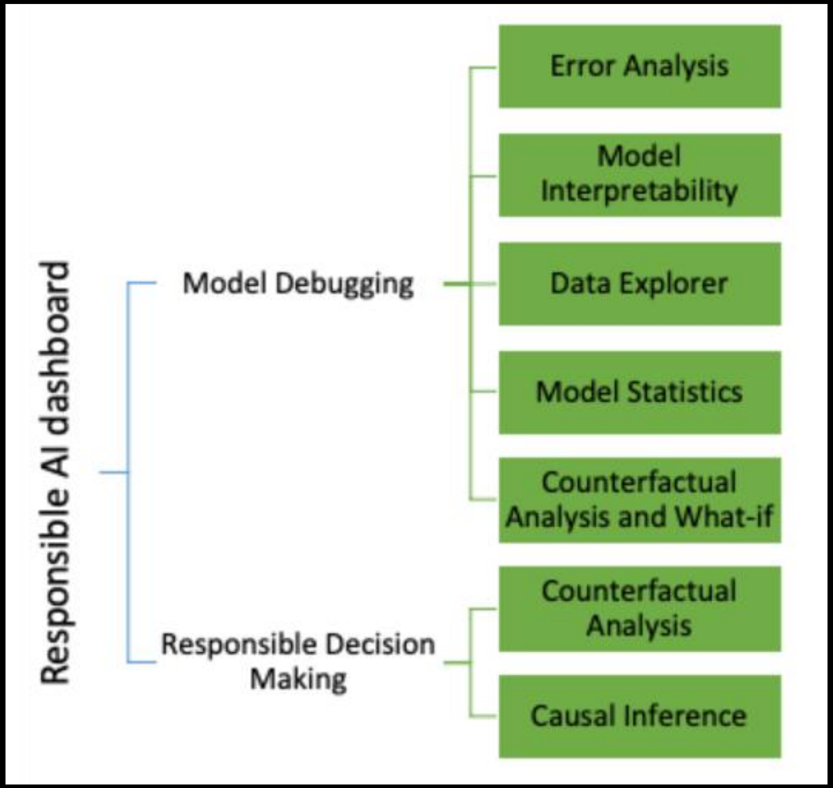
For Identified Victims of Trafficking, does Recent Natural Disaster cause Severity of Control to Increase?
Workflow
Guidance
Workspace
Understand process
Define question
Model causal factors
Prepare data
Perform analysis
Create reports
Controls
Confounders
Outcome determinants
Causes
Exposure
Recent Natural Disaster
Outcome
Severity of Control
Verify possibility
Previous step
Next step

ShowWhy



[microsoft/showwhy (github.com)](https://github.com/microsoft/showwhy)

AzureML Responsible AI Dashboard



AzureML Responsible AI Dashboard

Model Debugging via Responsible AI dashboard

Identify

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

Diagnose

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

Mitigate

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

Exploratory-Data-Analysis
Understand dataset characteristics.

Inform Actions

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

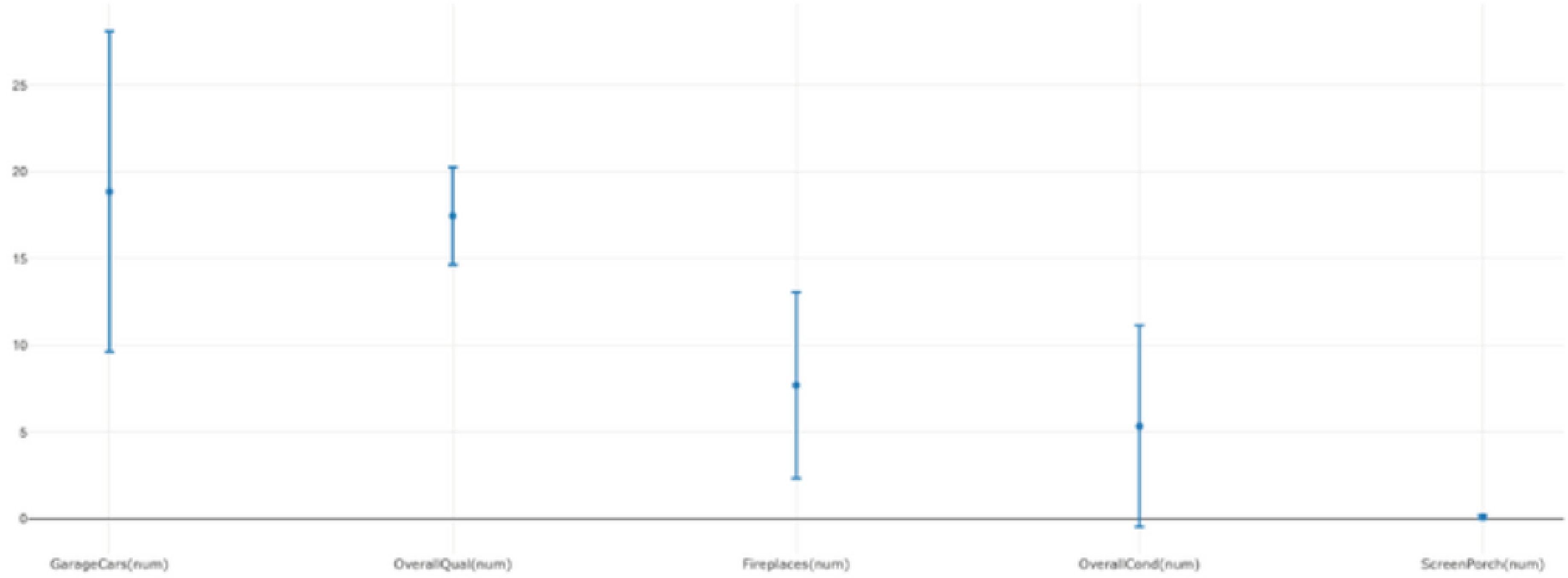
Counterfactual Analysis
Generate diverse counterfactual explanations for informing end users.

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

ⓘ 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)	8.797e-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

Aggregate causal effects | Individual causal what-if | Treatment policy

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

OverallCond <= 6.5	EnclosedPorch <= 15	EnclosedPorch > 15
	n = 508 Recommended treatment = increase	n = 73 Recommended treatment = decrease
OverallCond > 6.5	OpenPorchSF <= 151	OpenPorchSF > 151
	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 whet the individual will respond to the selected treatment. The table also specifies number datapoints in the current data sample assigned to each segment.

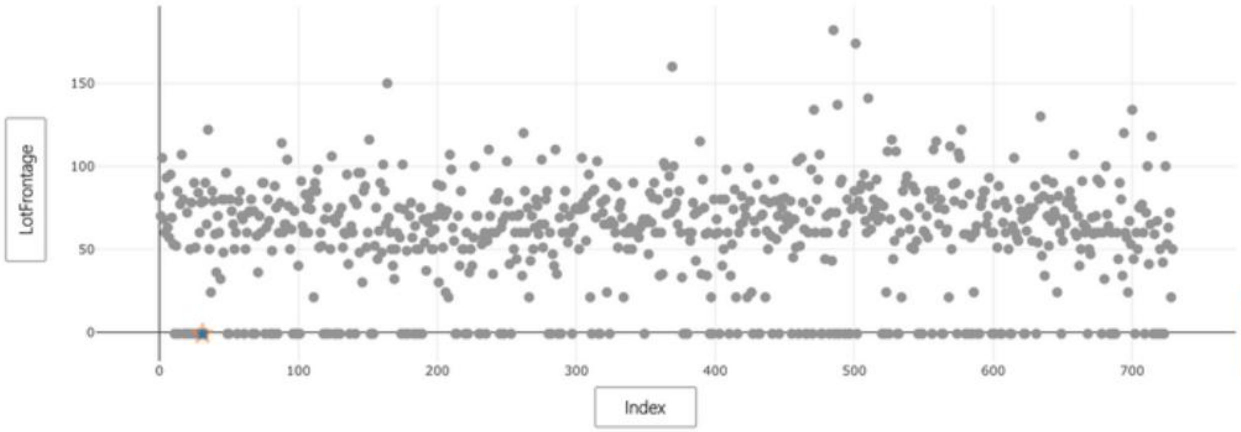
The table can be read by taking a row and then taking a column of that specific row.

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.



Datapoint index

Index 31

Select treatment

OverallQual

Current treatment value: 6

Set new treatment value

6

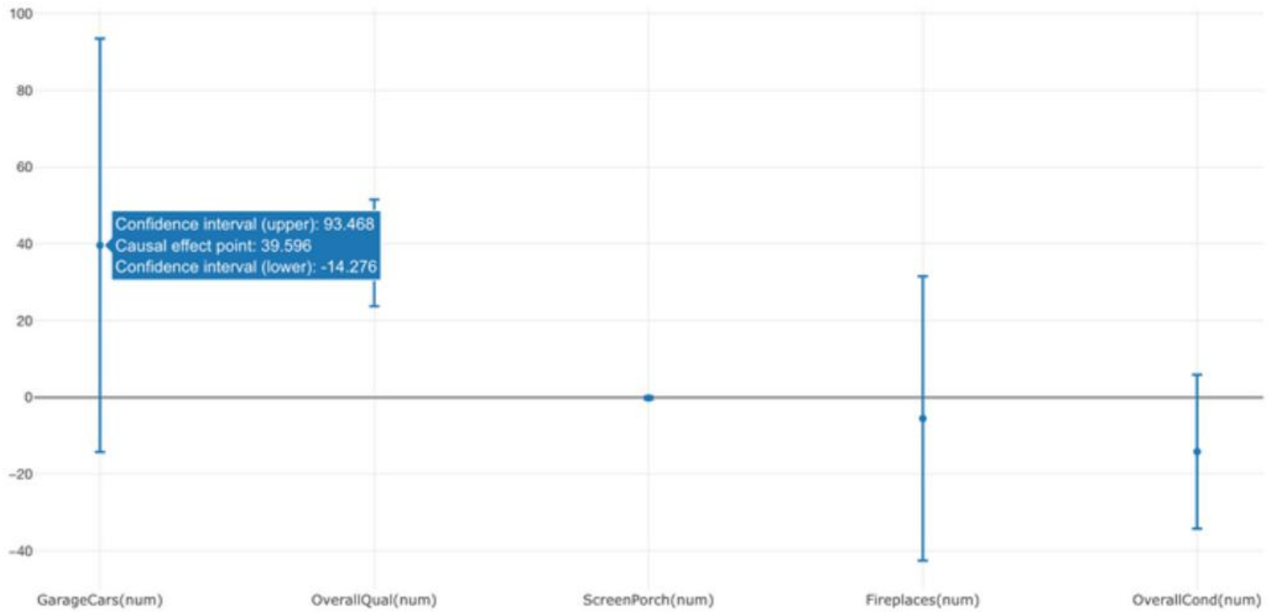
Treatment value min: 5.4 (-10%) , max: 6.6000000000000005 (+10%)

Current outcome

196

New outcome

196 (196, 196)

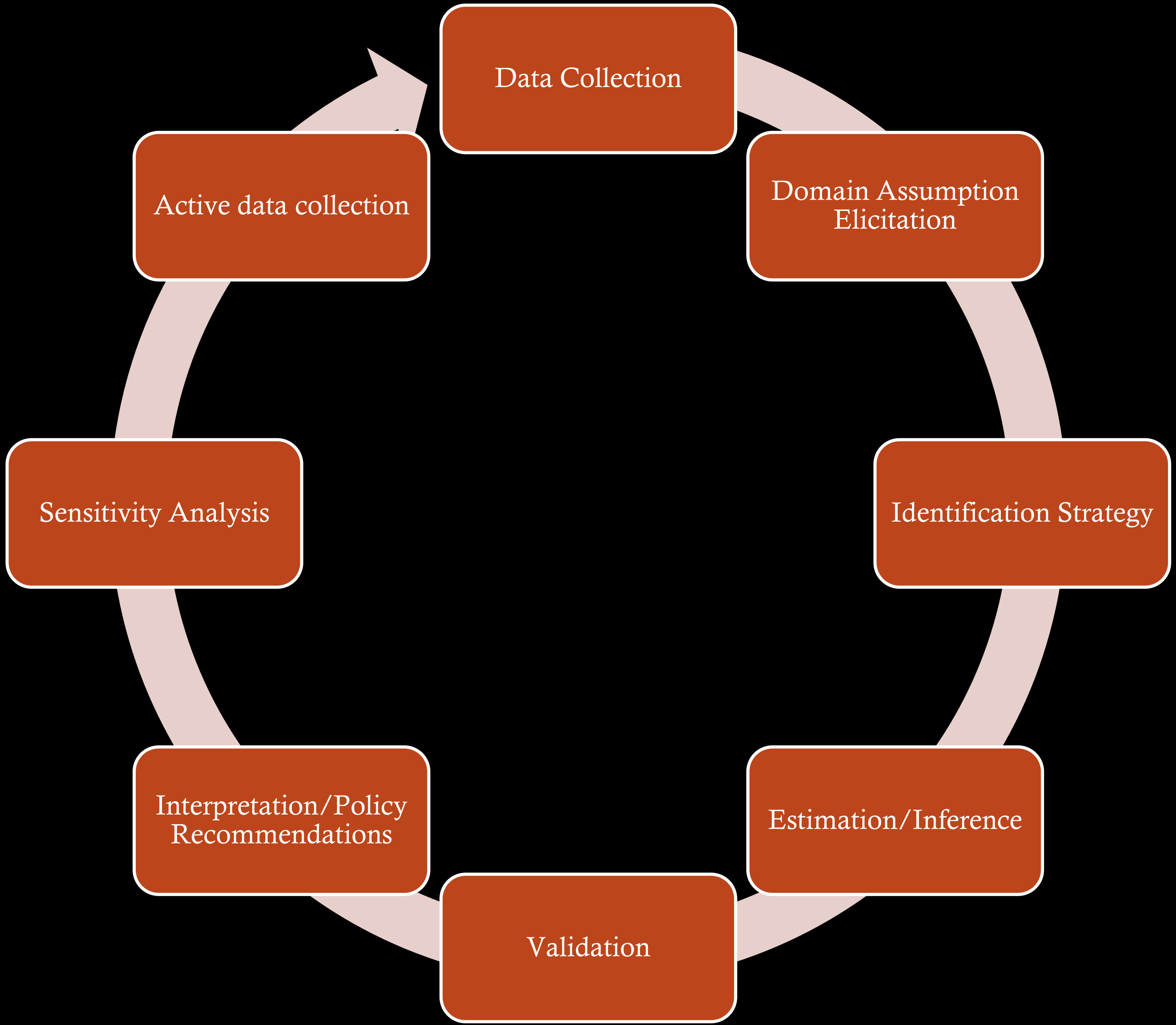


Continuous treatments: On average in this sample, increasing this feature by 1 unit will cause the probability of class/label 1 to increase by X units.

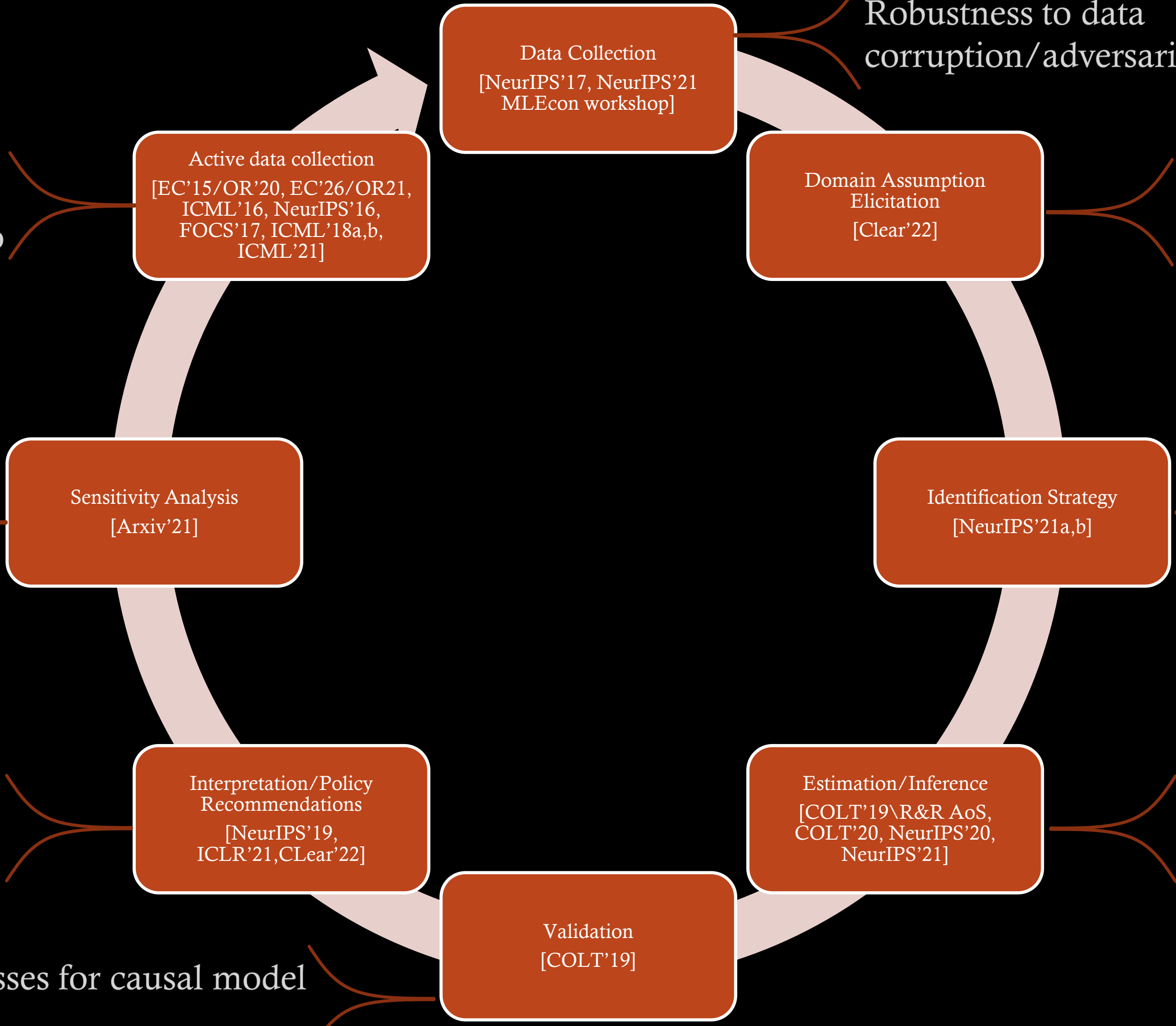
Binary treatments: On average in this sample, turning on this feature will cause the probability of class/label 1 to increase by X units.

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

The Causal Inference Pipeline



Many Research Challenges



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

Good losses for causal model selection

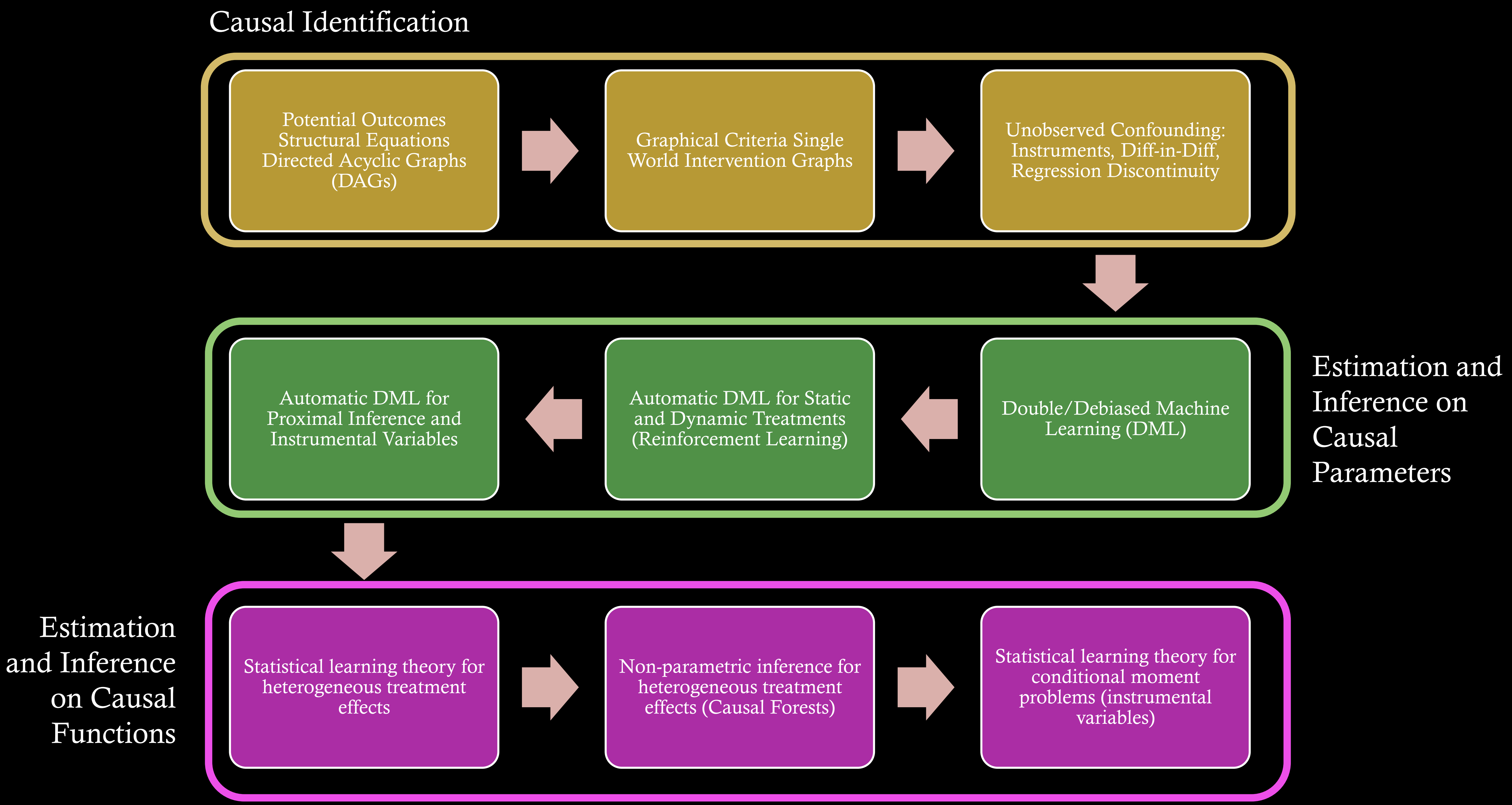
Evidence based interpretations for policy makers

Causal attribution

Computationally efficient sensitivity analysis

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

[2108.13518] DoWhy: Addressing Challenges in Expressing and Validating Causal Assumptions (arxiv.org)


Causal Identification
Potential Outcomes Structural Equations Directed Acyclic Graphs (DAGs)
Graphical Criteria Single World Intervention Graphs
Unobserved Confounding: Instruments, Diff-in-Diff, Regression Discontinuity
Estimation and Inference on Causal Parameters
Double/Debiased Machine Learning (DML)
Automatic DML for Static and Dynamic Treatments (Reinforcement Learning)
Automatic DML for Proximal Inference and Instrumental Variables
Estimation and Inference on Causal Functions
Statistical learning theory for heterogeneous treatment effects
Non-parametric inference for heterogeneous treatment effects (Causal Forests)
Statistical learning theory for conditional moment problems (instrumental variables)


Thank you!