Applied Causal Inference Powered by ML and AI

Vasilis Syrgkanis

MS&E, Stanford

Instructors

Vasilis Syrgkanis

vsyrgk stanford edu

Course Assistants

Jikai Jin

jkjin stanford edu

Shiangyi Lin

shiangyi stanford edu

A Data Science Tail

Credit: [joint blogpost with Scott Lundberg, Eleanor Dillon, Jacob LaRiviere and Jonathan Roth]

Somewhere in the world right now…

- (M)anager: "Build a model that predicts whether a customer will renew their product subscription"
- (D)ata (S)cientist: "I'll collect many factors from our database that I believe are predictive of renewal"
- $X = \{$*customer discount, ad spending, customer's monthly usage, last upgrade, bugs reported by a customer, interactions with a customer, sales calls with a customer, and macroeconomic activity*$\}$
- DS: "Using last year's data, I fitted a state-of-the-art ML model (xgboost; gradient boosted forest) to predict $y = \{$ renewal $\}$ from X!"

```
[2]: X, y = user_retention_dataset()
     model = fit_xgboost(X, y)
```

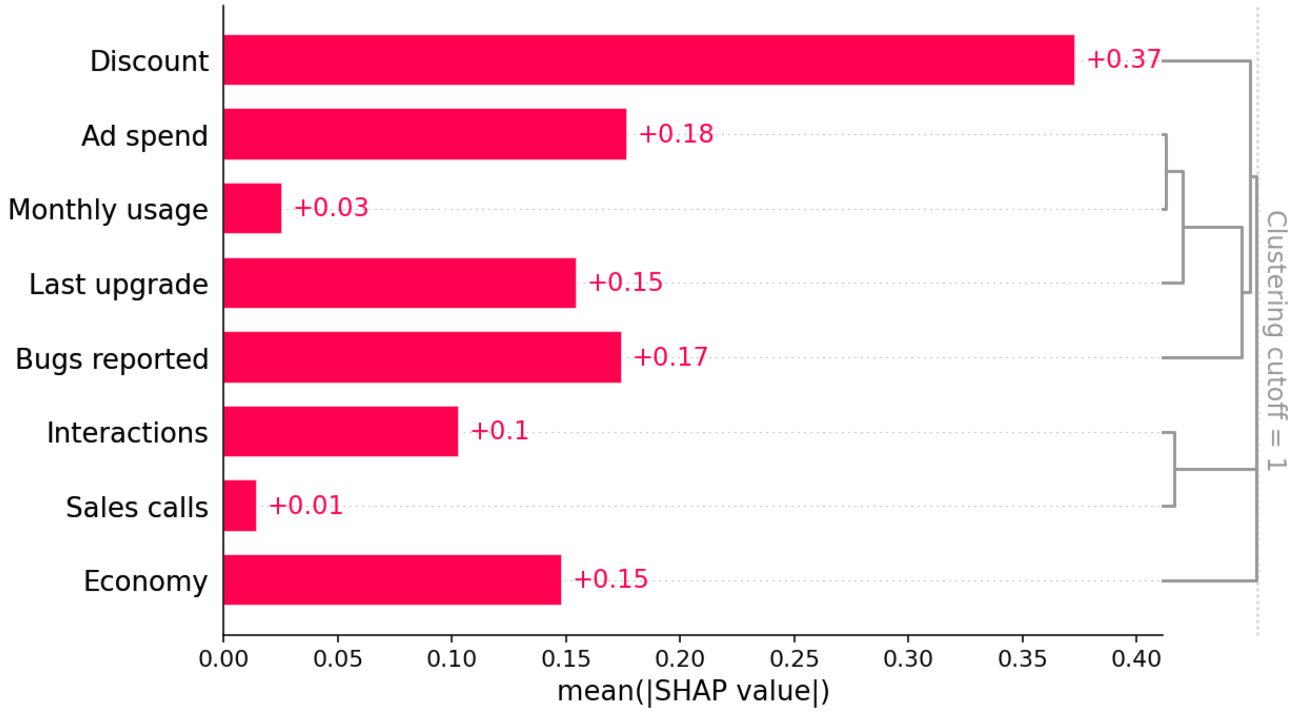
So what…

- DS: “It learned a function $f: X \to y$ that represents the relationship between the variables X and the outcome y!”
- M: “Fantastic! How accurate does it predict when given new data it hasn’t seen?”
- DS: “It gives the correct answer 99% of the time!”
- M: “Fantastic! It’s a great model! We can use it to project next year’s revenue!”
- M: “Oh; Maybe we can also see what it learned and try to prevent churn proactively!”
- DS: “Yeah! I know an amazing new explainable machine learning tool called SHAP; you give it any model and it returns what variables were important and in which direction they influence the outcome.”
- M: “Let’s see it!”

The important factors

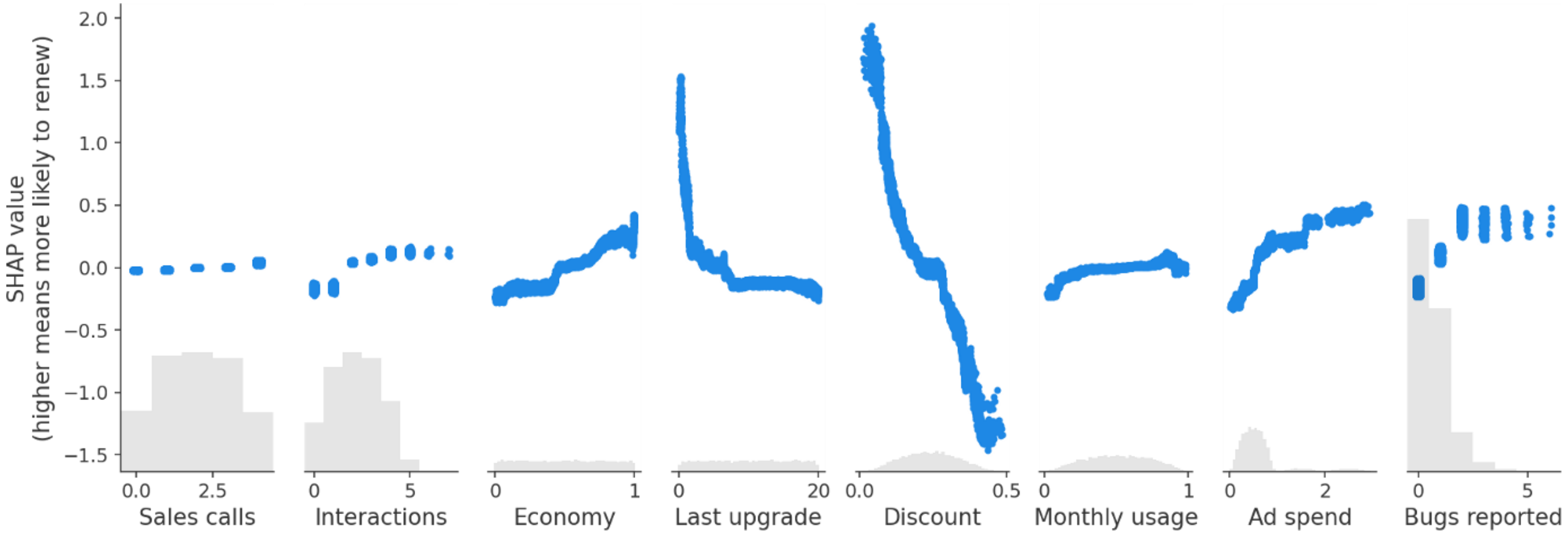
```python
import shap

explainer = shap.Explainer(model)
shap_values = explainer(X)

clust = shap.utils.hclust(X, y, linkage="single")
shap.plots.bar(shap_values, clustering=clust, clustering_cutoff=1)
```

- DS: “It seems that discounts and ad spend are important! Also bugs!”
- M: “Great let’s see how much each one affects the outcome?”



The awkwardness

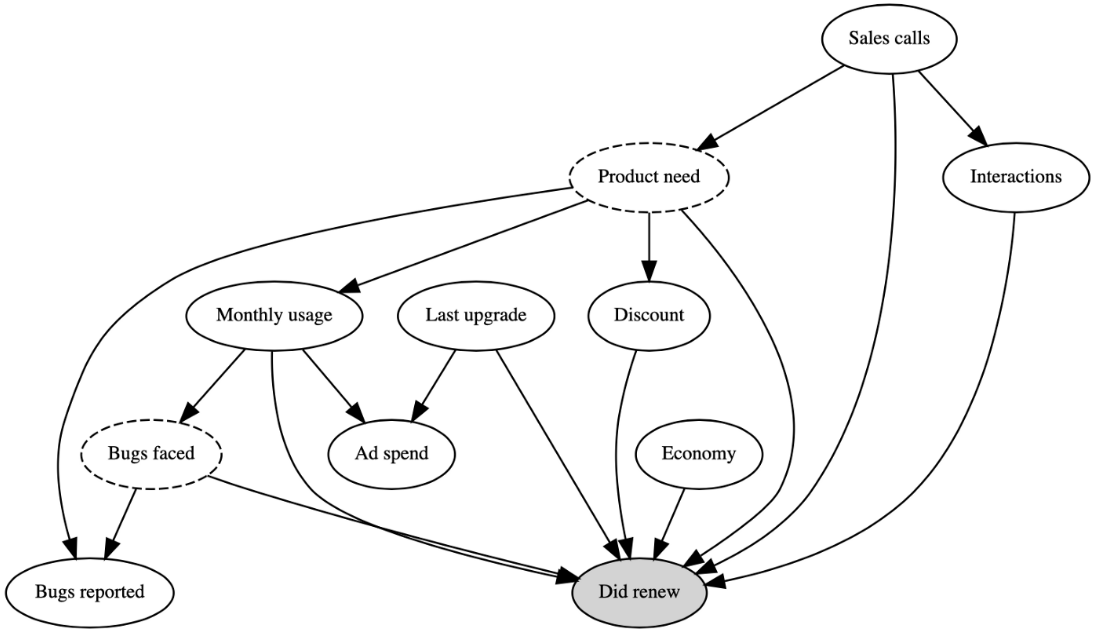
```
shap.plots.scatter(shap_values)
```



- DS: “So larger discounts reduces renewal! Also, more bugs lead to renewal! Oh, and ads are very important!”
- M: “Great let’s increase prices, add more bugs and spam everyone!”

What happened?

- Business expert:
  - “Users with high usage who value the product are more likely to report bugs and to renew their subscriptions.”
  - “The sales force tends to give high discounts to customers they think are less likely to be interested in the product, and these customers have higher churn.”



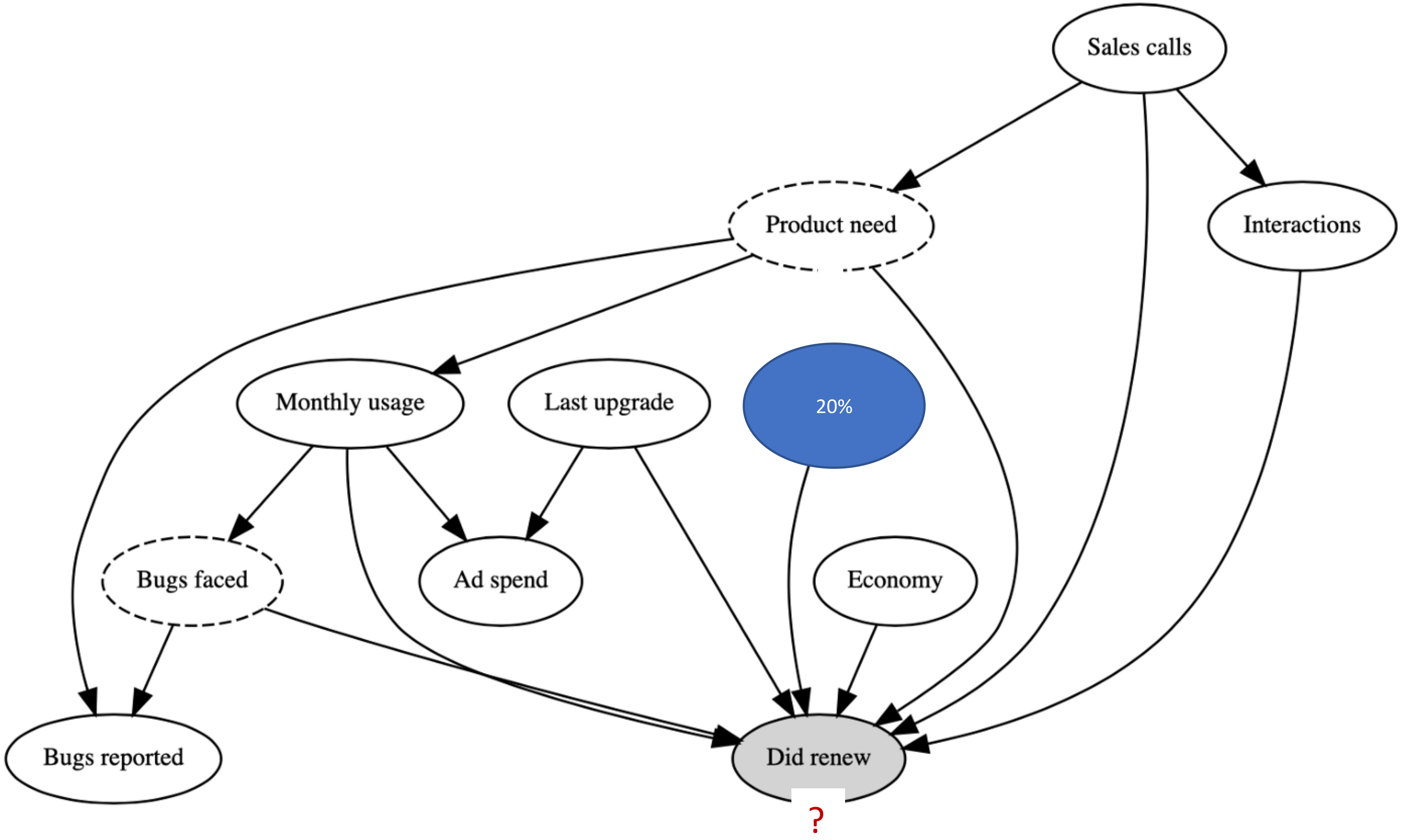
Are the counter-intuitive relationships that the model learned problematic?

It depends

- If our goal was to simply project next year's revenue (without any intervention), then these relationships are not problematic
- Such tasks that ask for "projecting" some outcome variable in the absence of any intervention are "predictive tasks"

- If our goal is to understand what would happen if we intervene in one of the variables to increase retention, then these relationships are problematic
- Such tasks that ask for "what-if" or "counterfactual" values of an outcome under some intervention are "causal tasks"

Causal/Interventional Question



How do we answer causal questions?

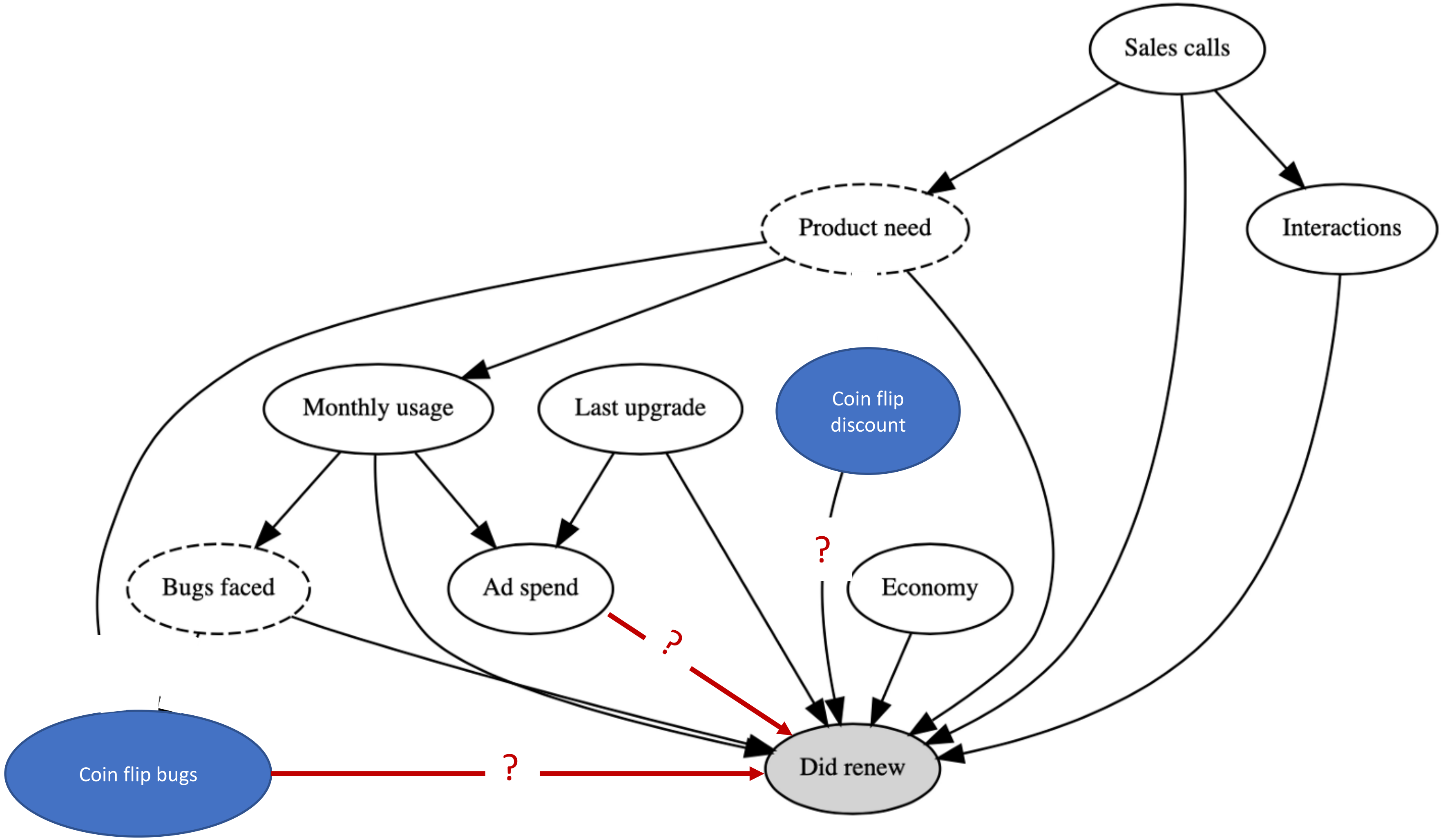
Experiments:
The ideal solution

Why the ideal

- By randomizing the treatment have two populations that are statistically indistinguishable other than that they differ in the assigned treatment

- Any statistical differences in the outcome between the two populations can then safely be attributed to the treatment

What would an A/B test do?



Limitations

- Ethical
- Practical
- Generalizability

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]

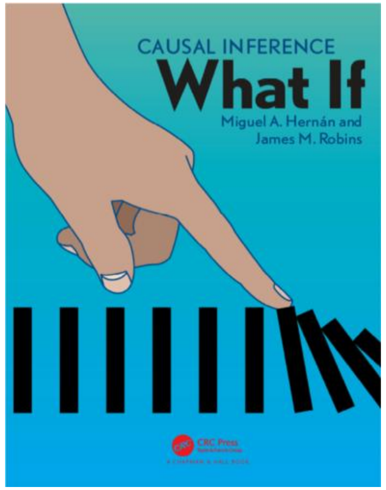
DONALD B. RUBIN[2]
Educational Testing Service, Princeton, New Jersey

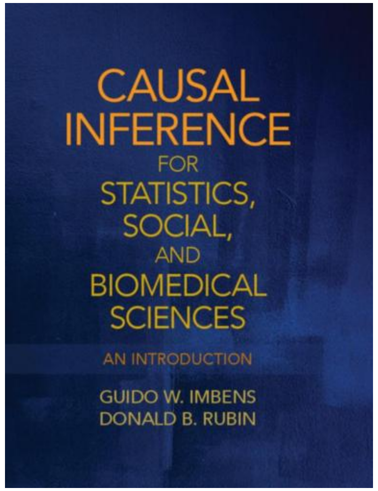
Statistical Science
1990, Vol. 5, No. 4, 465–480

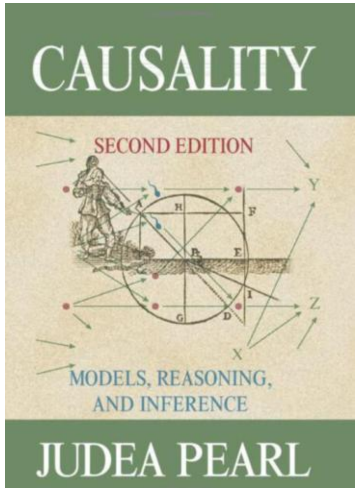
On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.

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)







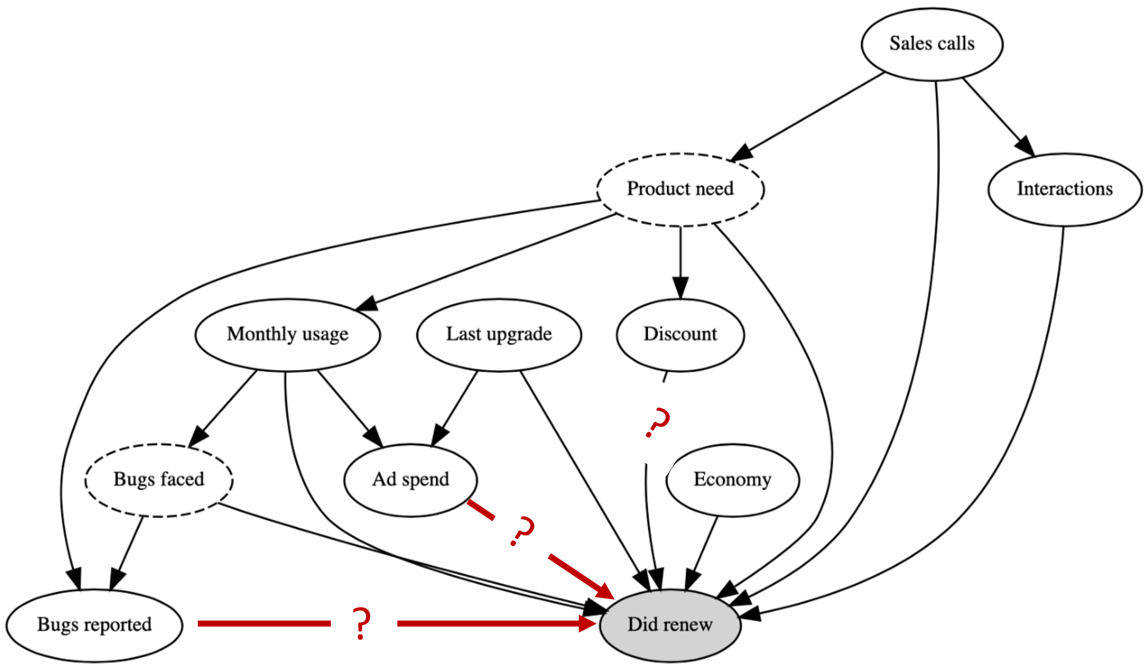
ECONOMETRICA
VOLUME 11 JANUARY, 1943 NUMBER 1

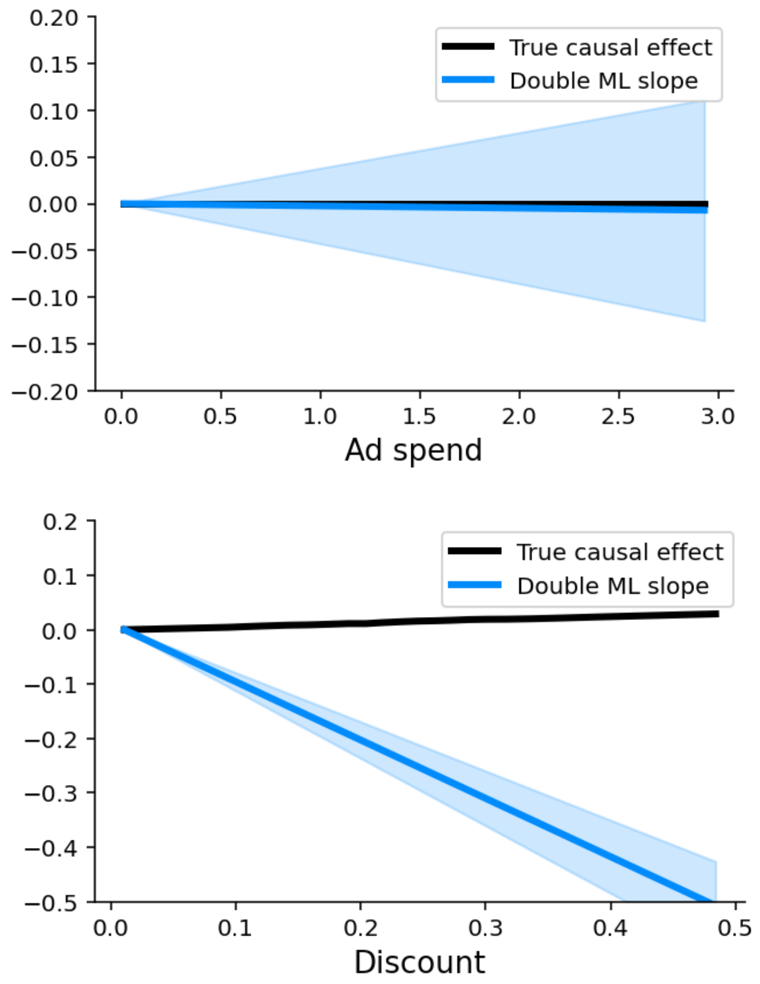
THE STATISTICAL IMPLICATIONS OF A SYSTEM OF SIMULTANEOUS EQUATIONS

By TRYGVE HAAVELMO

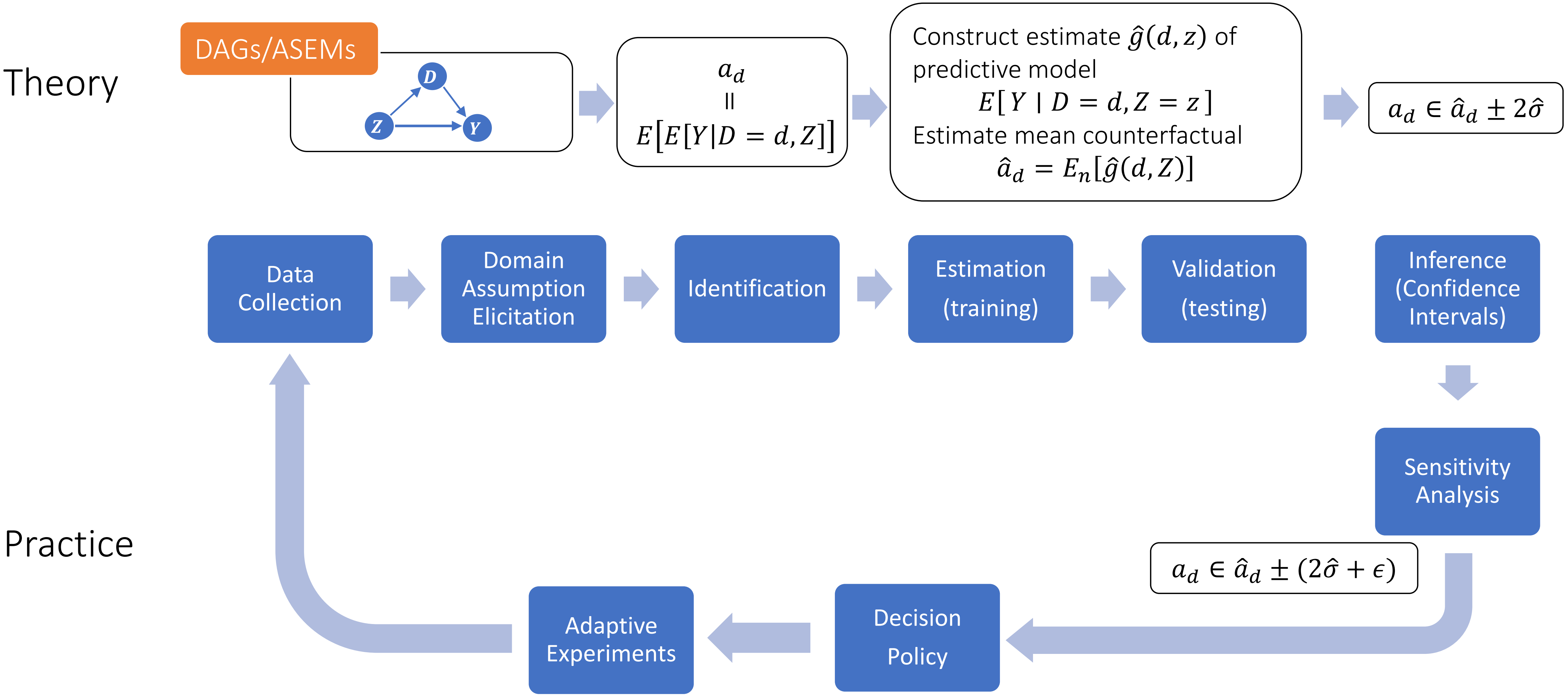
Require domain knowledge

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





Causal Inference Pipeline



The immense improvement in observational techniques

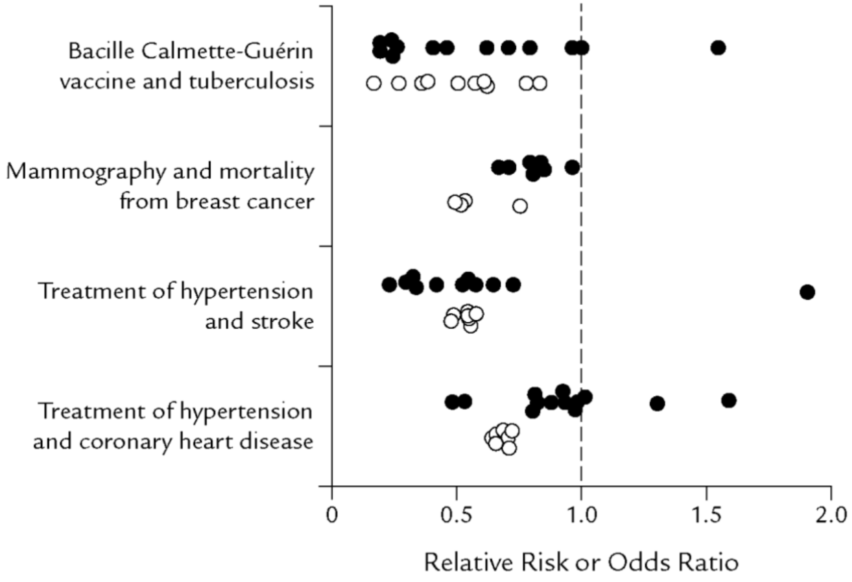


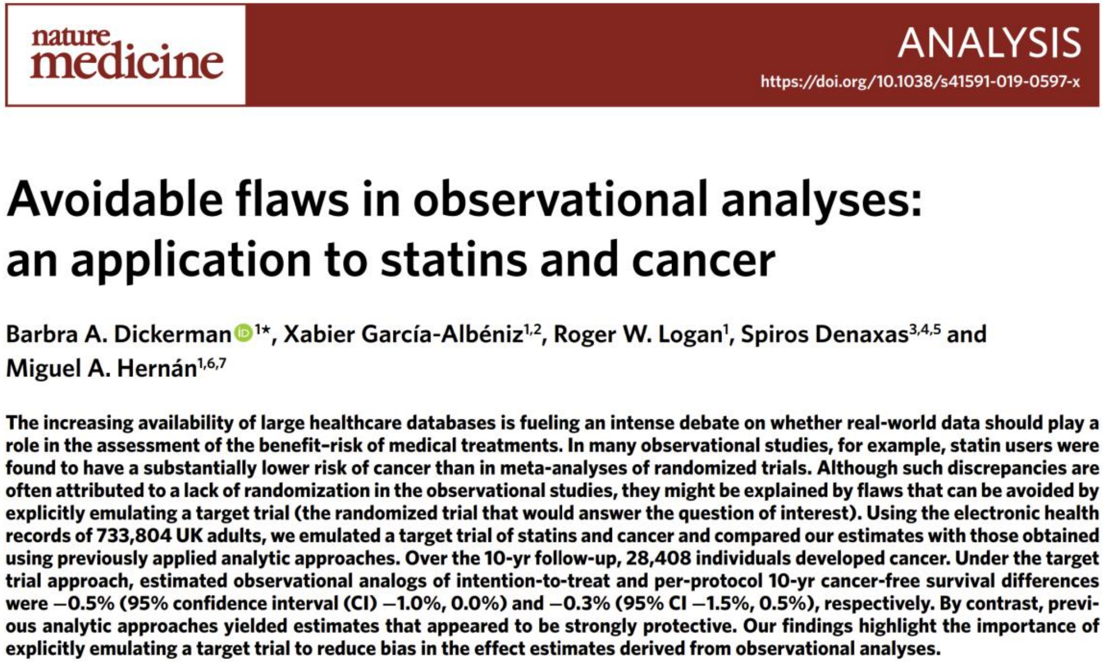
Figure 2. Comparison of odds ratio ranges for observational studies (○) and randomized controlled trials (●). Adapted with permission.[13]

Importance of Observational Studies in Clinical Practice

Robert J. Ligthelm, MD[1]; Vito Borzì, PhD[2]; Janusz Gumprecht, MD, PhD[3]; Ryuzo Kawamori, MD[4]; Yang Wenying, MD[5]; and Paul Valensi, MD[6]

The pitfalls

- Conclusions hinge on validity of assumptions
- Data quality is much more questionable in observational studies


nature medicine
ANALYSIS
https://doi.org/10.1038/s41591-019-0597-x

Avoidable flaws in observational analyses: an application to statins and cancer

Barbra A. Dickerman[1*], Xabier García-Albéniz[1,2], Roger W. Logan[1], Spiros Denaxas[3,4,5] and Miguel A. Hernán[1,6,7]

The increasing availability of large healthcare databases is fueling an intense debate on whether real-world data should play a role in the assessment of the benefit-risk of medical treatments. In many observational studies, for example, statin users were found to have a substantially lower risk of cancer than in meta-analyses of randomized trials. Although such discrepancies are often attributed to a lack of randomization in the observational studies, they might be explained by flaws that can be avoided by explicitly emulating a target trial (the randomized trial that would answer the question of interest). Using the electronic health records of 733,804 UK adults, we emulated a target trial of statins and cancer and compared our estimates with those obtained using previously applied analytic approaches. Over the 10-yr follow-up, 28,408 individuals developed cancer. Under the target trial approach, estimated observational analogs of intention-to-treat and per-protocol 10-yr cancer-free survival differences were −0.5% (95% confidence interval (CI) −1.0%, 0.0%) and −0.3% (95% CI −1.5%, 0.5%), respectively. By contrast, previous analytic approaches yielded estimates that appeared to be strongly protective. Our findings highlight the importance of explicitly emulating a target trial to reduce bias in the effect estimates derived from observational analyses.

What is new

- Richer datasets (small data) and high-dimensionality
- Larger datasets (big data) and the desire for personalization

- Advancement of modern predictive machine learning and large-scale computation
- Desire for more robust and flexible analysis even in classical datasets

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
- Digital experimentation, online ad targeting
- Personalized medicine
- Heterogeneity of effect in social science studies

Modern case studies

Sneak Peek 1: Digital 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

Sneak Peek 2: Long-term effects of new treatments

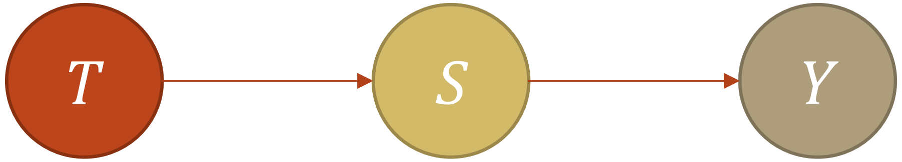
Through the lens of a Return-on-Investment (ROI) Case-Study at Microsoft

Estimating Long-Term Returns on Investment

- Companies frequently deploy 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

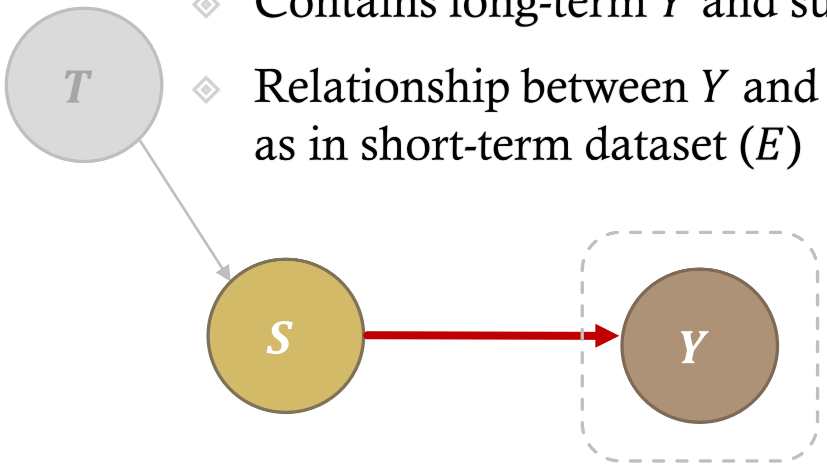
- Suppose that there are many short-term signals S that are indicative of a customer's long-term reward Y (e.g. the next 6-month purchase patterns of a customer could be indicative of their long-term spend)
- Suppose that investment program T affects long-term rewards if and only if it affects these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



- We will call these short-term signals S surrogates

Causal Inference with Surrogates 101

historical/long-term (O)

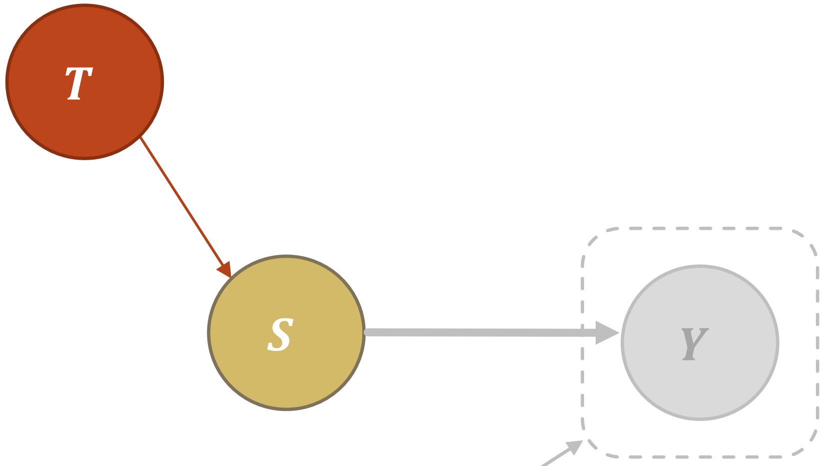


- Contains long-term Y and surrogates S
- Relationship between Y and S the same as in short-term dataset (E)

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

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

recent/short-term (E)



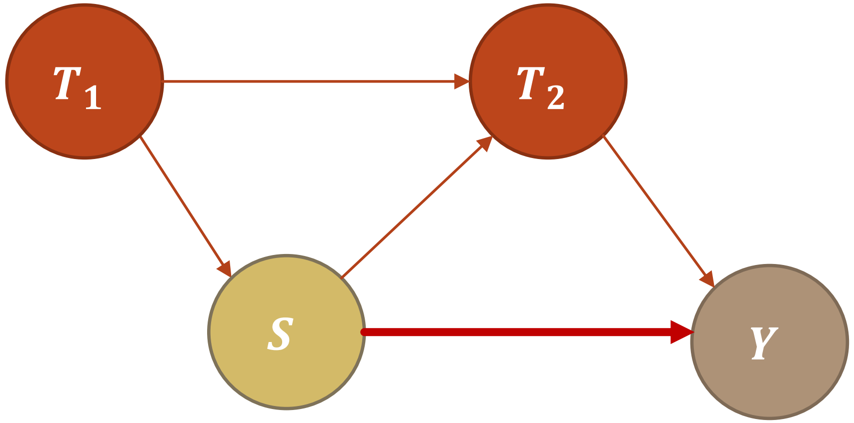
2. Impute expected long-term outcomes in (E)
3. Regress $g(S) \sim T$ to estimate effect of T on Y from (E)

Key Assumptions

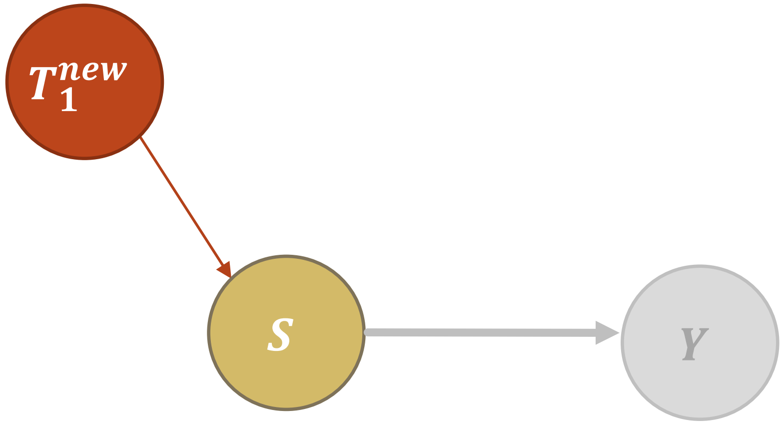
- Long-term effect only goes through surrogates
- Expected relationship between surrogates and long-term reward is the same long-term setting (O) and in short-term setting (E)

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

A Growing Software Tool EcoSystem

causaLens decisionOS Solutions

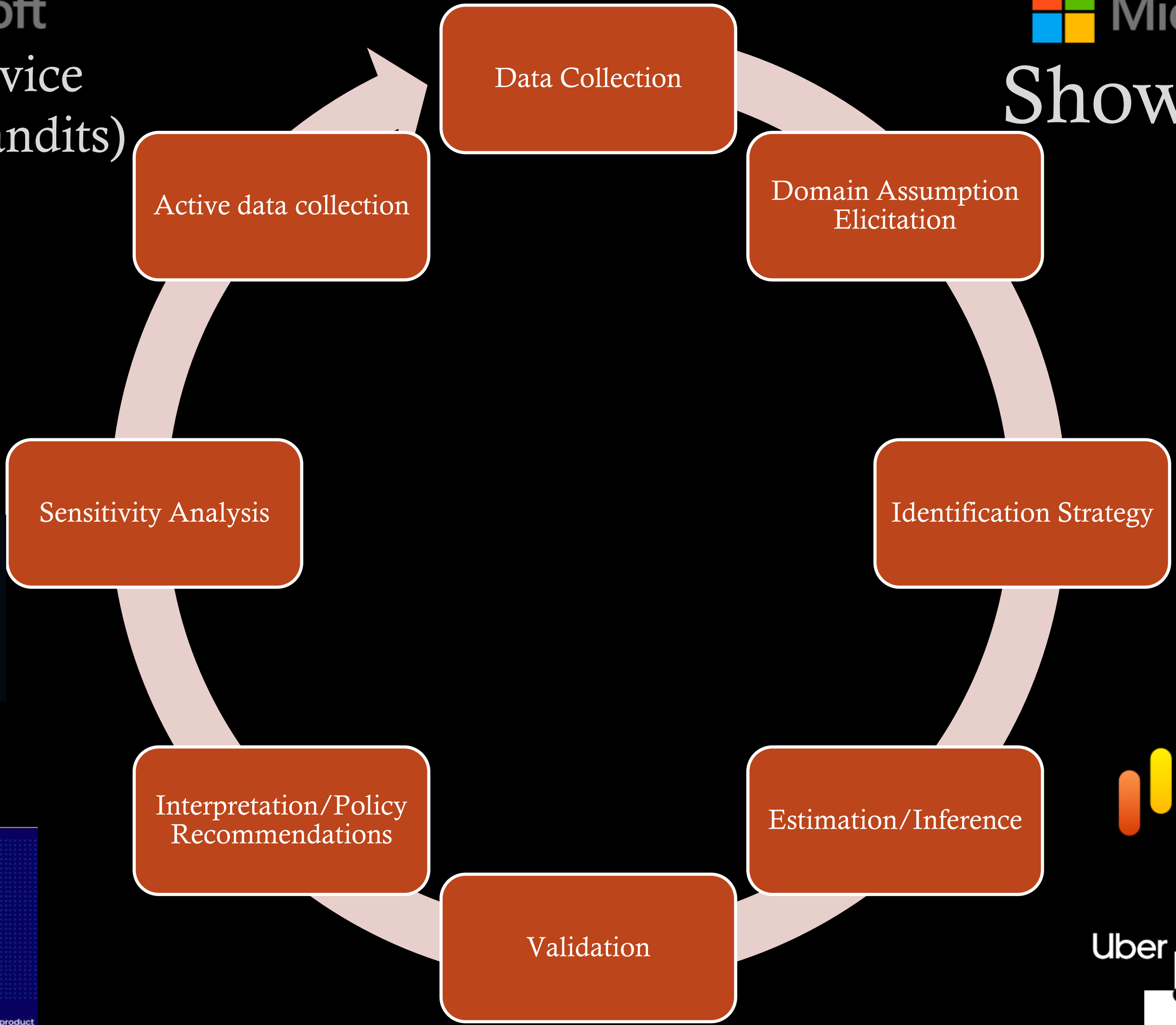
Causal AI: The next generation of Enterprise AI

Deploy Explore Learn Log

Microsoft Decision Service (Contextual Bandits)

Microsoft ShowWhy

GEMINOS



Microsoft DoWhy

Microsoft DoWhy

motif

Unlock the Power of Event Sequences to Answer the Why

tlverse

EconML

grf-labs

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

Uber CausalML

Booking.com UpliftML

WISE Auto-Causality

What we hope you’ll take away

Goals of the class

- Learn if and how you can identify causal effects from a dataset
- Learn how to properly use Machine Learning in causal estimation
- Practical experience implementing Causal ML methods in Python
- Practical experience applying Causal ML methods in real world datasets from social sciences, healthcare, tech

Structure and rough outline

Class Outline

- Section 1: Causal Effect Identification and Potential Outcomes
- Section 2: Estimation and inference with modern ML methods
- Section 3: Structural Equation Models and Directed Acyclic Graphs
- Section 4: Unobserved confounding: sensitivity analysis, instruments, proxy controls
- Section 5: Heterogeneous Treatment Effects
- Section 6: Topics:
  - Difference-in-Differences
  - Dynamic Treatment Effects
  - Long-Term Effects via Surrogates
  - Regression Discontinuity Designs

Logistics

- Class info: https://stanford-msande228.github.io/winter26/
- Approximately 7, roughly weekly homework assignments (90%)
- Class participation (more than half of in-class polls) (10%)
- Main text-book: https://causalml-book.org
- Discussion: *Ed Discussion*, Submissions: *Gradescope*

Office Hours: (Starting Week 2)

	Time	Location
Vasilis Syrgkanis	Tue 4.30-6pm, Fri 4.30-6pm	Huang 252
Shiangyi Lin	Mon 1.30-3pm, Thu 11-12.30	TBD
Jikai Jin	Mon 9-10.30, Fri 9-10.30	TBD