Machine Learning Estimation of Heterogeneous Treatment Effects with Instruments

Vasilis Syrgkanis
Microsoft Research
New England

Victor Lei
TripAdvisor

Miruna Oprescu
Microsoft Research
New England

Maggie Hei
Microsoft Research
New England

Keith Battocchi
Microsoft Research
New England

Greg Lewis
Microsoft Research
New England

Causal Inference and Machine Learning

- Machine learning infiltrating decision making
- Most decision making questions are causal/counterfactual
- At odds with ML power: prediction vs counterfactual prediction

- Many times we can estimate “causal” ML models, if we use auxiliary models for de-biasing (e.g. nuisance models)

Goal: a general framework of ML with nuisance models

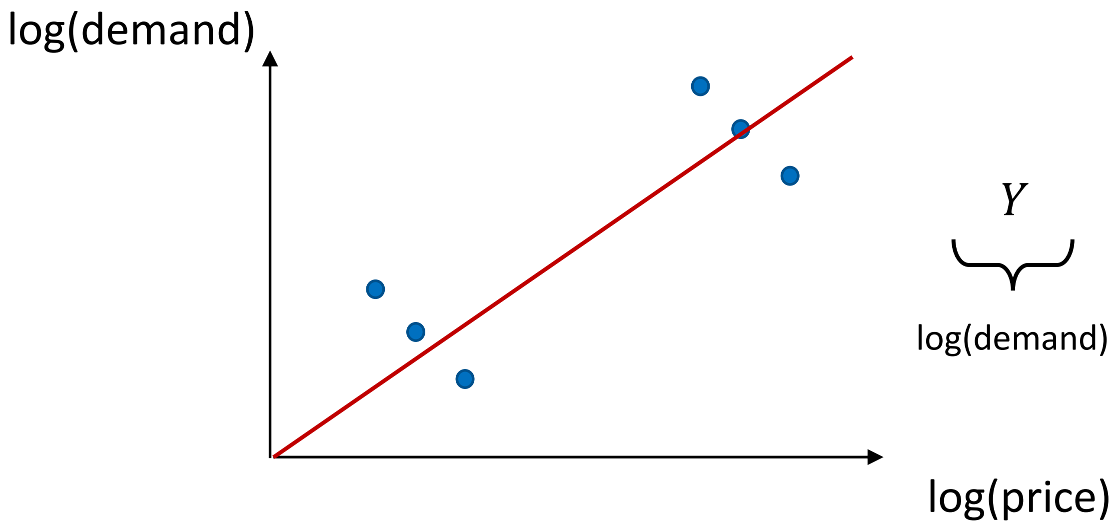
- Econometrics -> ML: use of the notion of Neyman orthogonality for robust generalization bounds
- ML -> Econometrics: focus on “counterfactual” generalization bounds can avoid many assumptions and allow more flexible target models

Walkthrough Example 1:

Estimating Heterogeneous Elasticity of Demand

Example: Estimating Price Elasticity of Demand

Goal: Estimate elasticity, the effect of a change in price on demand



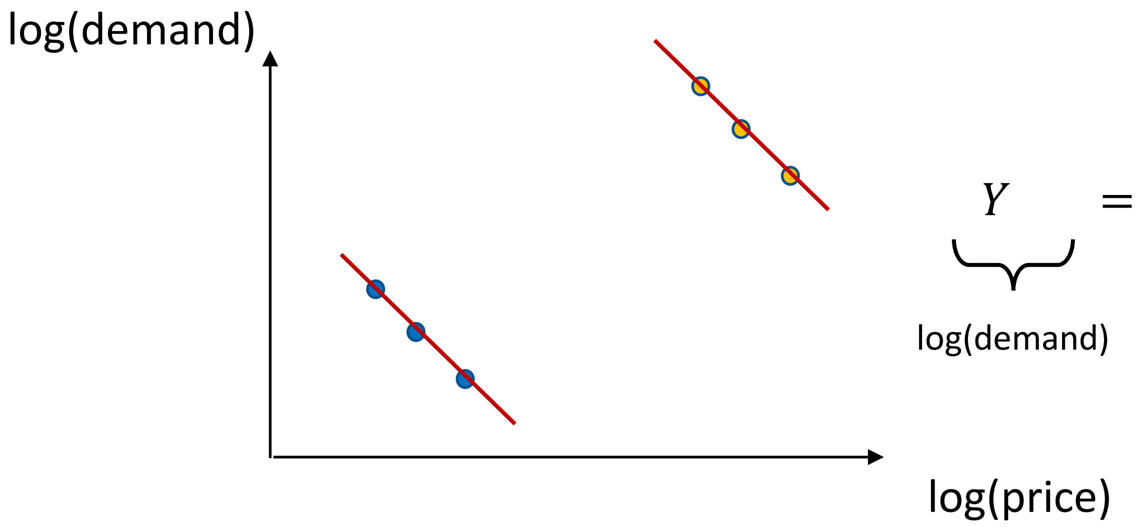
$$\underbrace{Y}_{\text{log(demand)}} = \underbrace{\theta_0}_{\text{elasticity}} \cdot \underbrace{T}_{\text{log(price)}} + \underbrace{\epsilon}_{\text{noise}}$$

Conclusion: Increasing price increases demand!

Problem: Demand increases in winter and price anticipates demand

Example: Estimating Price Elasticity of Demand

Goal: Estimate elasticity, the effect of a change in price on demand

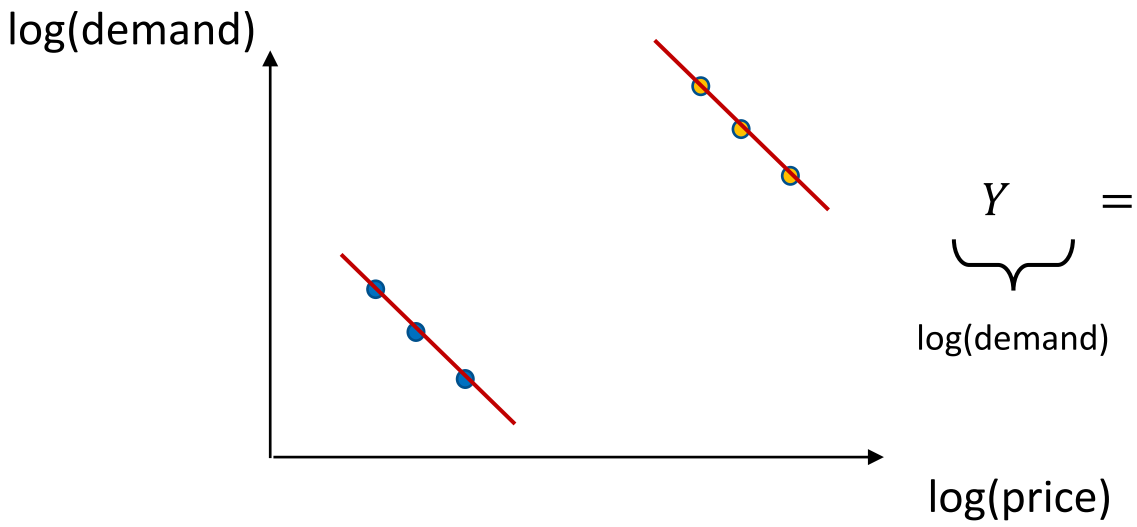


$$\underbrace{Y}_{\text{log(demand)}} = \underbrace{\theta_0}_{\text{elasticity}} \cdot \underbrace{T}_{\text{log(price)}} + \beta_0 \cdot \underbrace{X}_{\text{season indicator}} + \underbrace{\epsilon}_{\text{noise}}$$

Idea: Introduce confounder (the season) into regression

Example: Estimating Price Elasticity of Demand

Goal: Estimate elasticity, the effect of a change in price on demand



$$\underbrace{Y}_{\text{log(demand)}} = \underbrace{\theta_0}_{\text{elasticity}} \cdot \underbrace{T}_{\text{log(price)}} + \beta_0 \cdot \underbrace{X}_{\text{season indicator}} + \underbrace{\epsilon}_{\text{noise}}$$

Problem: What if there are 100s or 1000s of potential confounders?

Unobserved Confounders and Instrumental Variables

Unobserved Confoundedness

$$\underbrace{Y}_{\text{outcome}} = \underbrace{\theta_0(X)}_{\text{heterogeneous effect}} \cdot \underbrace{T}_{\text{treatment}} + \cancel{\underbrace{f(W)}_{\text{effect of potential confounders}}} + \underbrace{\epsilon}_{\text{noise}}$$

W is not observed

- One solution: Instrumental variables
  - Variables Z that affect T but does not directly affect Y

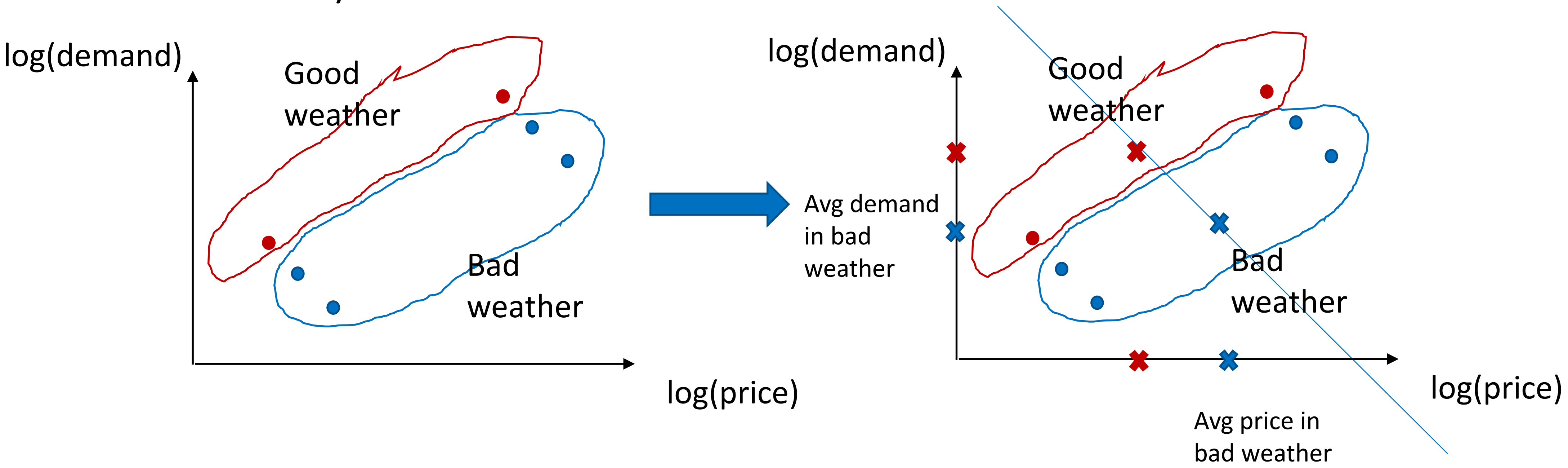
$$Y = \theta_0(X) \cdot T + \epsilon$$

$$T = g(Z) + \eta$$

Example: Estimating Price Elasticity of Demand

Goal: Estimate elasticity, the effect of a change in price on demand

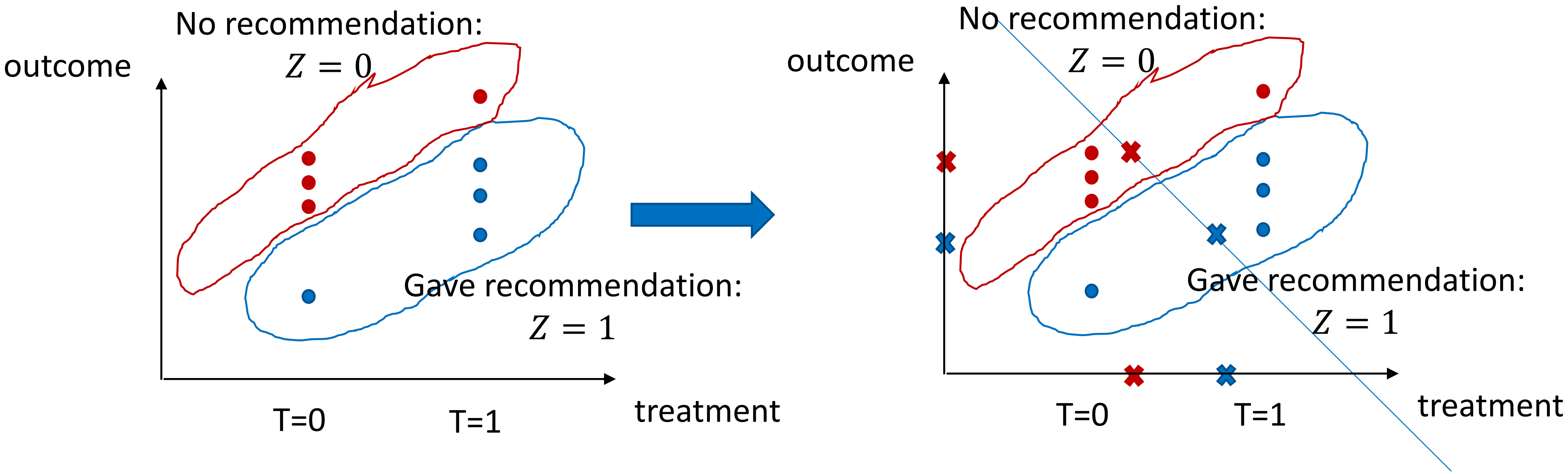
- Instrument: weather in brasil affects production cost of coffee and hence price of coffee but does not directly affect the demand in US



Example: A/B testing with non-compliance

Goal: Estimate effect of treatment without ability to enforce treatment

- Run an A/B test in the form of recommendation:
  - Recommend a user to take an action/treatment with some probability
  - User decides to take the recommended action/treatment
  - Estimate the effect of the treatment
- Instrument: **the recommendation** (assuming that recommendation increases the prob of treatment)



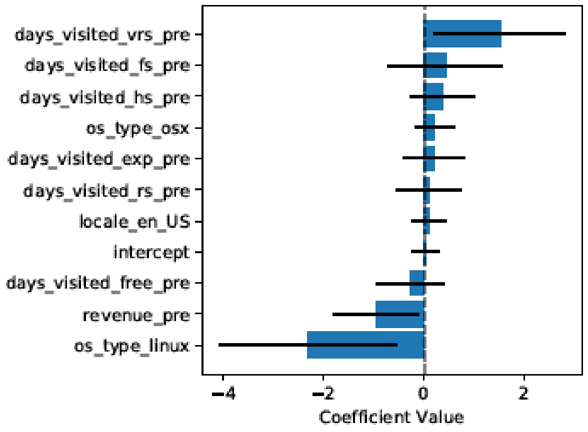
Effect of Membership on TripAdvisor

[S., Lei, Oprescu, Hei, Battocchi, Lewis, ’19]

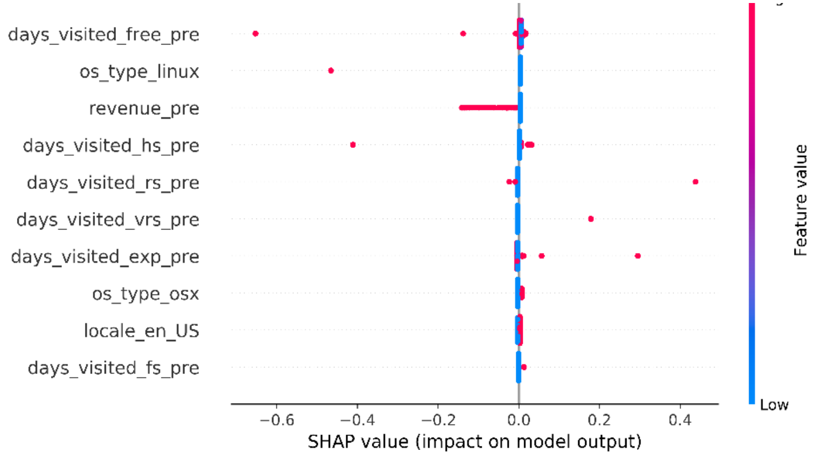
A/B Test: For random half of 4million users, easier sign-up flow was enabled

- Easier sign-up incentivizes membership
- A/B test can be used as an instrument for measuring effect of membership
- Outcome: number of visits in the next 14 days

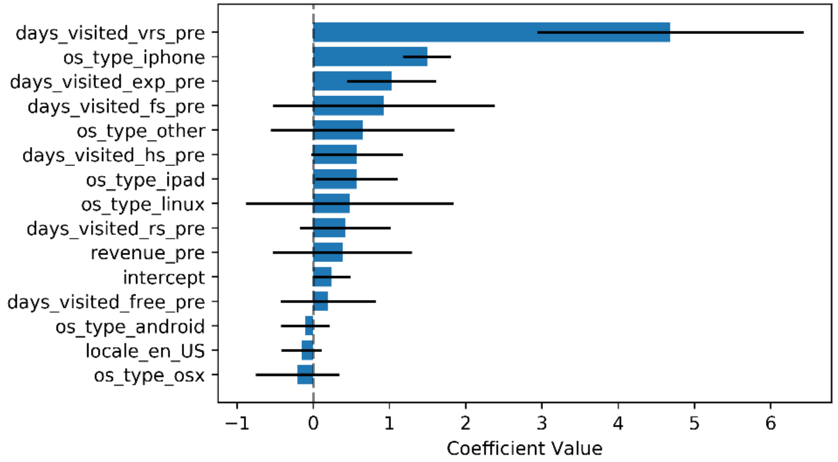
Linear Effect Model



SHAP Interpretation of Shallow RF Model

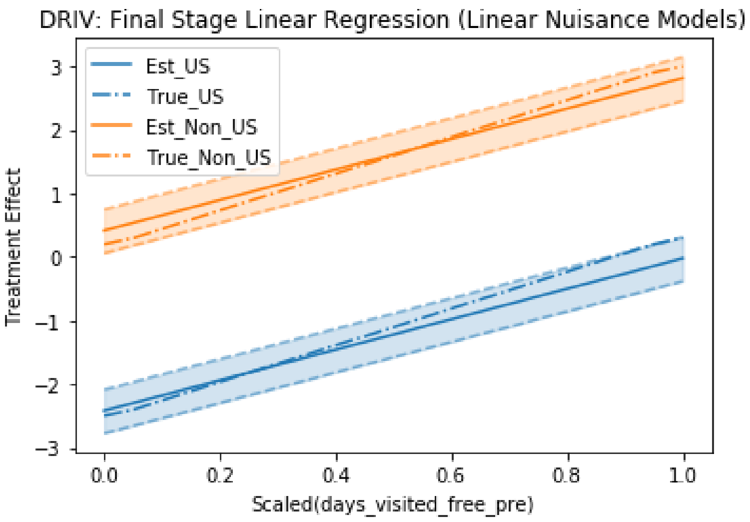


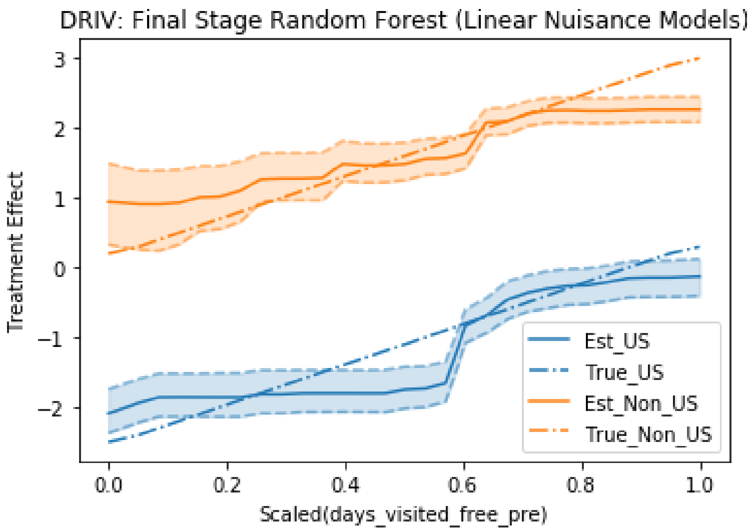
Linear Effect Model: 2019 Experiment

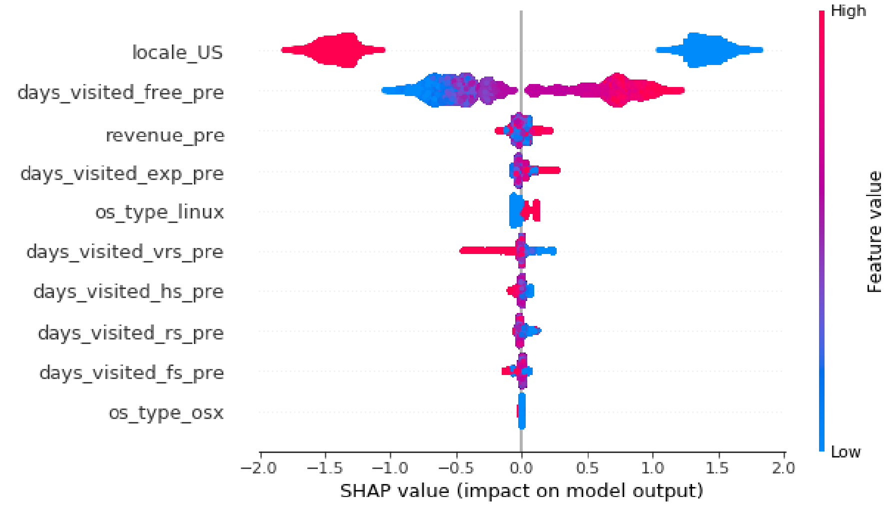


Semi-Synthetic Data

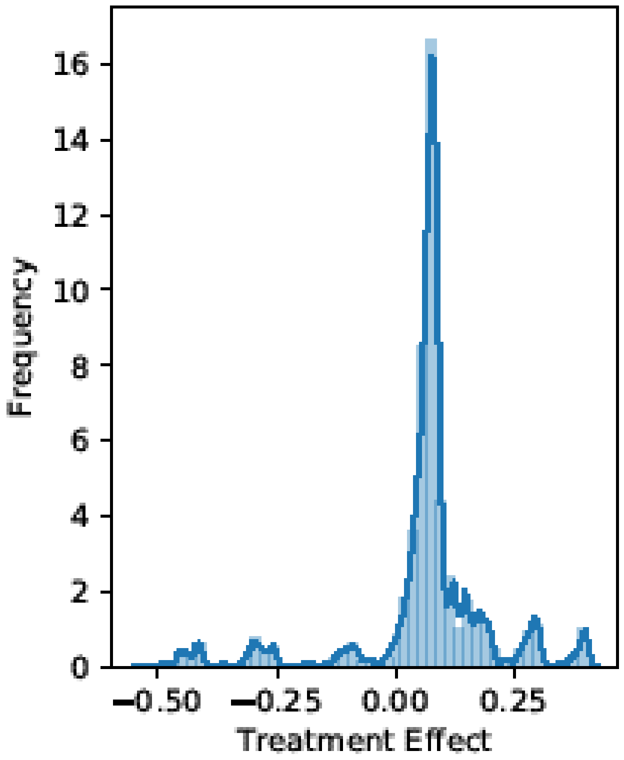
Simulated Data with similar marginals and known ground truth

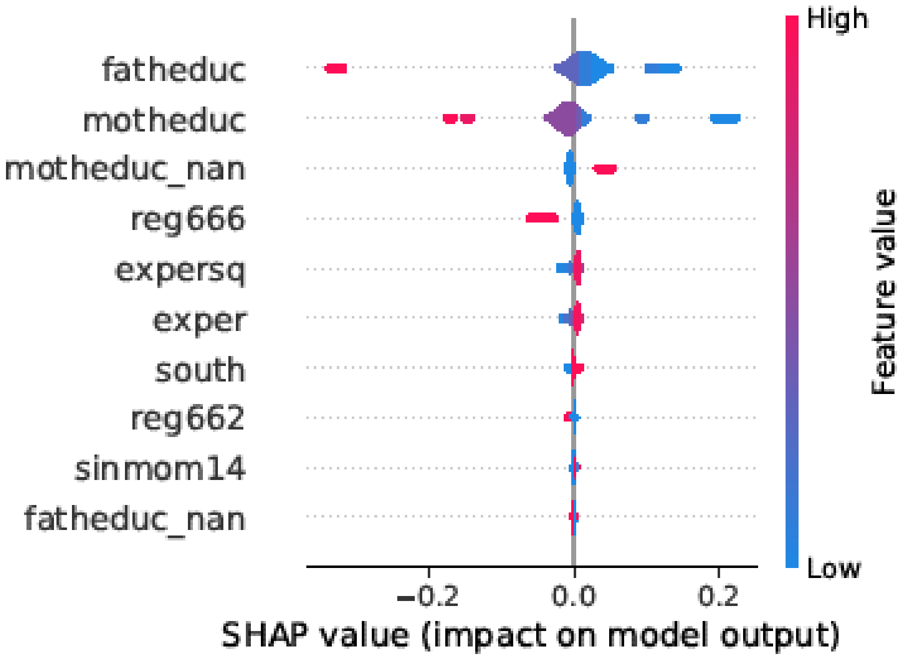


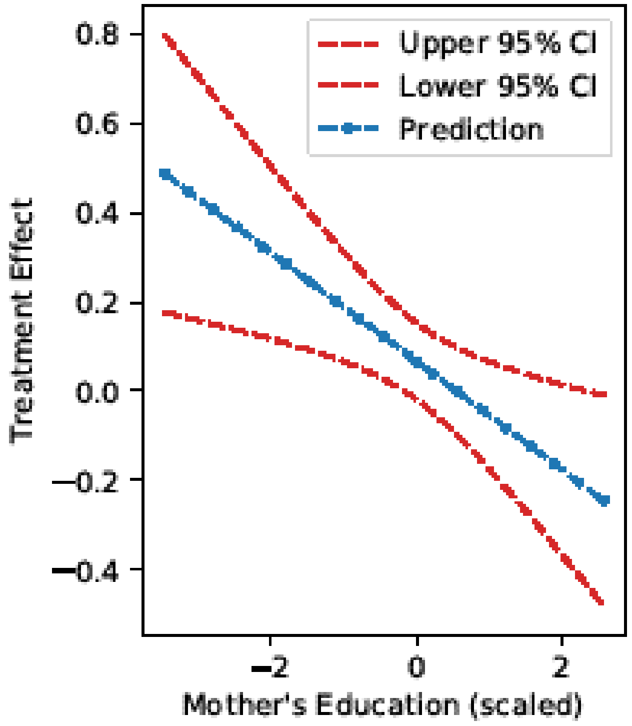




Returns of Schooling to Wages

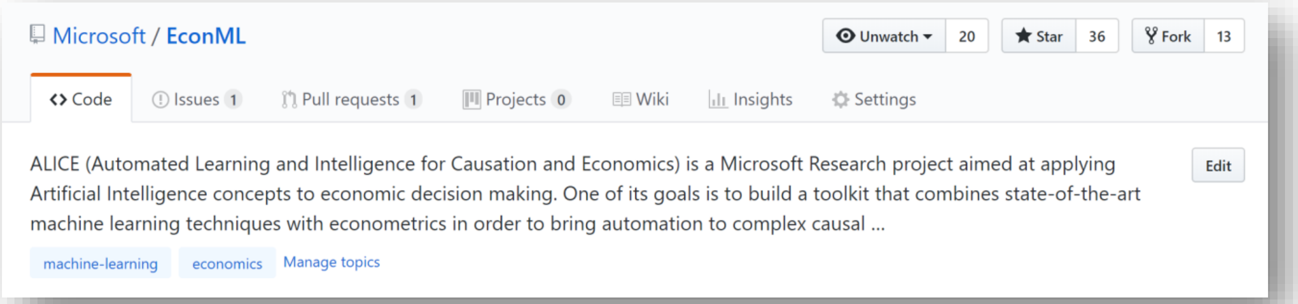




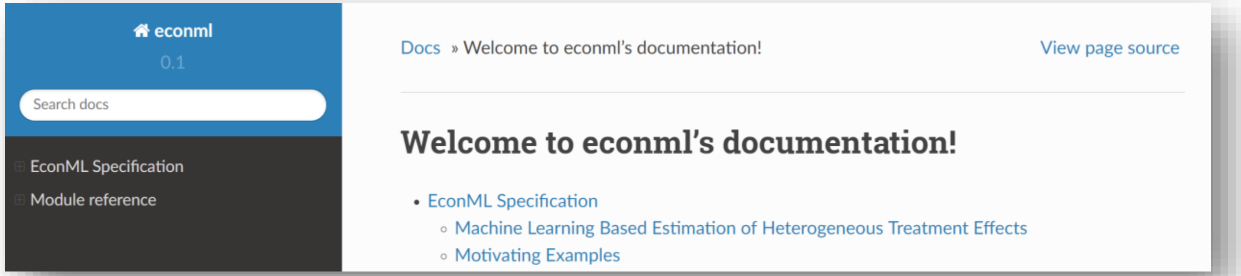


A Python Library

- Go to our GitHub repo: https://github.com/microsoft/econml



- Check out our documentation: https://econml.azurewebsites.net/



- Install EconML: "pip install econml"