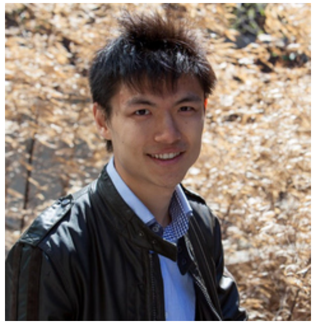
Orthogonal Random Forests for Causal Inference

Steven Wu
University of Minnesota

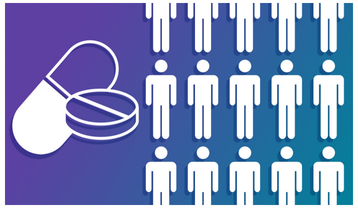
Joint work with: Miruna Oprescu and Vasilis Syrgkanis
Microsoft Research—New England



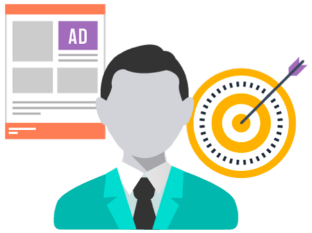


Motivating examples

Dynamic pricing



Clinical trials



Targeted advertising

- Conditional average treatment estimation (CATE) from observational data
  - Outcome Y_i (demand)
  - Treatment T_i (pricing)
  - Feature X_i that captures heterogeneity (income level)
  - Confounders W_i(other observed variables)

$$Y = \overbrace{\mu(X,W)}^{\text{Treatment effect}} \cdot T + f_0(X,W) + \epsilon$$
$$T = g_0(X,W) + \eta$$

Our Goal: CATE estimation

$$\theta_0(x) = E[\mu_0(X,W) |\, X = x]$$

More generally…

- In the language of econometrics:

Given a target feature x, find a solution $\theta_0(x)$ to

$$E[\,\psi(Z;\theta, h_0(X,W)) \mid X = x] = 0$$

with score function ψ and nuisance function h_0

- Other examples: non-parametric regression, instrumental variable regression, local maximum likelihood estimation, etc.

Orthogonal Random Forest (ORF)

Orthogonality (or double ML)

[Neyman1979; Chernozhukov et al. 2017]

+

Generalized Random Forest (GRF)

[Wager & Athey 2018; Athey et al. 2019]

Method:
- Perform two-stage estimation: first estimate nuisance, then estimate target θ_0

Pros:
- Robust to high-dimensional confounders

Cons:
- Assumes parametric form θ_0

Method:
- Non-parametric random forest-based estimation

Pros:
- Allows more general functions θ_0

Cons:
- Does not directly handle high-dimensional nuisance functions

Main theoretical results for ORF

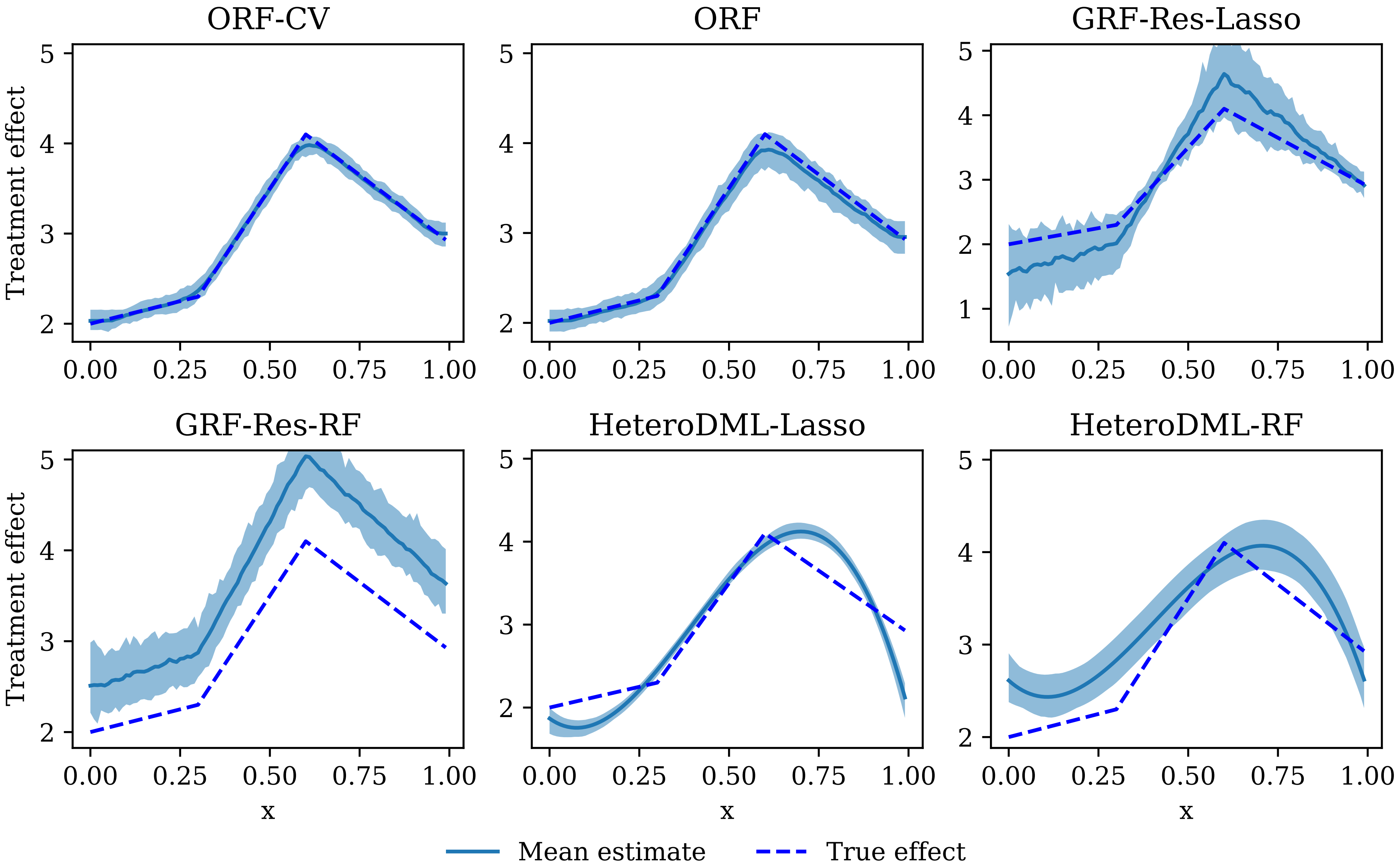
Accuracy for ORF estimate $\hat{\theta}$

- Consistency error rate
- Asymptotic normality

Nuisance estimation procedure

- Forest Lasso method that leverages locally sparse structure

Empirical Evaluation



Orthogonal Random Forests for Causal Inference

Poster: Wed Jun 12th @ Pacific Ballroom #195