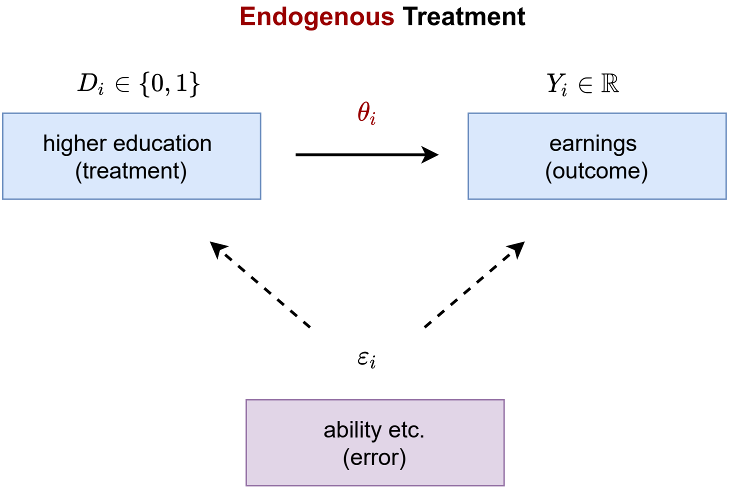
Adaptive Estimation of Partially Identified Treatment Effects

Maria Nareklishvili

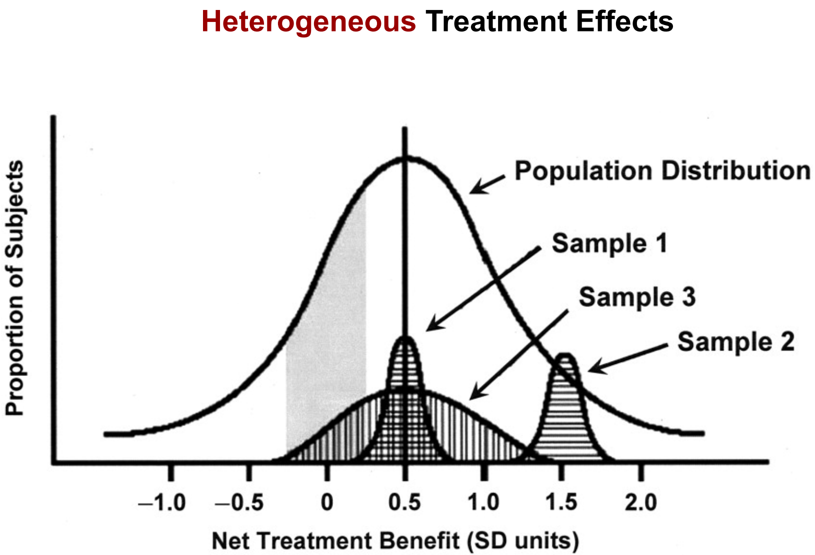
The University of Oslo
maria.nareklishvili@frisch.uio.no

December 11, 2022

Motivation: Endogeneity



Motivation: Heterogeneity



Research Questions

Impose causal assumptions to deal with endogeneity and use random forests to infer partially identified treatment effects.

- *Conditional treatment effect bounds* tighter and more informative?
- Do conditional bounds reflect *heterogeneous subgroups*?
- Statistically and economically *significant* heterogeneity?

Contribution

- **Introduce flexible direction** of the monotonicity assumptions to identify conditional treatment effect bounds [Manski, 1990; Balke and Pearl, 1997; Heckman and Vytlacil, 1999; Horowitz and Manski, 2000; Beresteanu and Manski, 2000; Chernozhukov et al., 2007; Manski and Pepper, 2009; Lee, 2009; Romano and Shaikh, 2010; Shaikh and Vytlacil, 2011; Beresteanu et al., 2011; Huber and Mellace, 2015; Mogstad and Torgovitsky, 2018; Torgovitsky, 2019; Semenova, 2020; Heiler, 2022].
- Introduce **multivariate random forests** and investigate large sample properties [Breiman, 2001, 2004; Lin and Jeon, 2006; Meinshausen and Ridgeway, 2006; Biau, 2012; Denil et al., 2014; Wager, 2014; Scornet et al., 2015; Athey and Imbens, 2016; Wager and Athey, 2018; Athey et al., 2019; Nekipelov et al., 2018; Chernozhukov et al., 2017, 2018; Li, 2020].

Illustrative Example

National Longitudinal Survey of Youth, 1979 (De Haan and Leuven, 2020). Treatment: the preschool program; outcome: years of schooling.

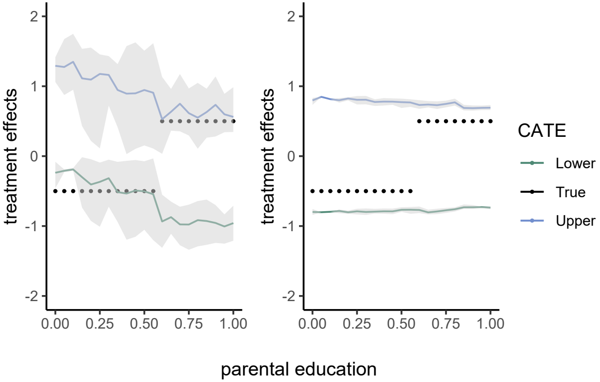


Figure 1: negative treatment selection, subgroups recovered by causal forests **(left)**, flexible treatment selection, subgroups recovered by MRF **(right)**.

Table of Contents

Identification: Monotonic Treatment Selection

1. Bounded support of the outcome

2. MTS

A treatment is a preschool program, designed for children with a disadvantaged family background

Observed

$$\mathbb{E}(Y(1)|D_i = 1, X_i)$$

positive selection

$\geq$

or

negative selection

$\leq$

Unobserved

$$\mathbb{E}(Y(1)|D_i = 0, X_i)$$

Identification: Bounds of CATE

Proposition 1.1

Under a negative treatment selection (Manski, 1990; Manski and Pepper, 2000):

$$\theta^L(x) = \mathbb{E}\big(Y_i(1)|D_i = 1, X_i = x\big) - \mathbb{E}\big(Y_i(0)|D_i = 0, X_i = x\big)$$
$$\theta^U(x) = \mathbb{E}\big(Y_i(1)|D_i = 1, X_i = x\big) \cdot P(D_i = 1|X_i = x) + Y^U(x) \cdot P(D_i = 0, X_i = x) - \mathbb{E}\big(Y_i(0)|D_i = 0, X_i = x\big) \cdot P(D_i = 0|X_i = x) - Y^L(x) \cdot P(D_i = 1|X_i = x).$$

Proof

$Y^B(x)-$ user-specified quantiles for $B \in \{L, U\}$.

Assumptions

Assumption 2.1 (Honesty)

$\mathcal{F}(Y_{im}|X_i, s) = \mathcal{F}(Y_{im}|X_i)$. $s = (j, c))$ *denotes the splitting coordinates and the splitting indices.*

Assumptions

Assumption 2.1 (Honesty)

$\mathcal{F}(Y_{im}|X_i, s) = \mathcal{F}(Y_{im}|X_i)$. $s = (j, c))$ *denotes the splitting coordinates and the splitting indices.*

Assumption 2.2 (Random Split Trees)

The probability that the next split occurs at j-th covariate is bounded below by π/p for $\pi \in (0, 1]$, for all $j = 1, \ldots, p$.

Assumptions

Assumption 2.1 (Honesty)

$\mathcal{F}(Y_{im}|X_i, s) = \mathcal{F}(Y_{im}|X_i)$. $s = (j, c))$ *denotes the splitting coordinates and the splitting indices.*

Assumption 2.2 (Random Split Trees)

The probability that the next split occurs at j-th covariate is bounded below by π/p for $\pi \in (0, 1]$, for all $j = 1, \dots, p$.

Assumption 2.3 (The Splitting Algorithm is (α, k)-regular)

Each leaf contains at least $\alpha > 0$ fraction of observations. Moreover, the splitting ceases when $N_\ell \leq k$ for some $k \in \mathbb{N}$.

Assumptions

Assumption 2.4 (Distributional Assumptions on the Data Generating Process)

$X_i \in [0,1]^p$, $\mathbb{E}(Y_{im}|X_i = x)$, $\mathbb{E}\big((Y_{im})^2|X_i = x\big)$ *are Lipschitz-continuous. Furthermore,* $Var(Y_{im}|X_i = x)$ *is bounded away from 0 (i.e.,* $inf_{x\in\mathcal{X}} Var(Y_{im}|X_i = x) > 0$*).*

Assumptions

Assumption 2.4 (Distributional Assumptions on the Data Generating Process)

$X_i \in [0,1]^p$, $\mathbb{E}(Y_{im}|X_i = x)$, $\mathbb{E}\big(\big(Y_{im}\big)^2|X_i = x\big)$ *are Lipschitz-continuous. Furthermore,* $Var(Y_{im}|X_i = x)$ *is bounded away from 0 (i.e.,* $inf_{x\in\mathcal{X}} Var(Y_{im}|X_i = x) > 0$*).*

Assumption 2.5 (Overlap)

We assume that for all $x \in [0,1]^p$*:*

$$0 < \mathbb{P}(D_i = 1|X_i = x) < 1.$$

Inference: Höeffding Decomposition I

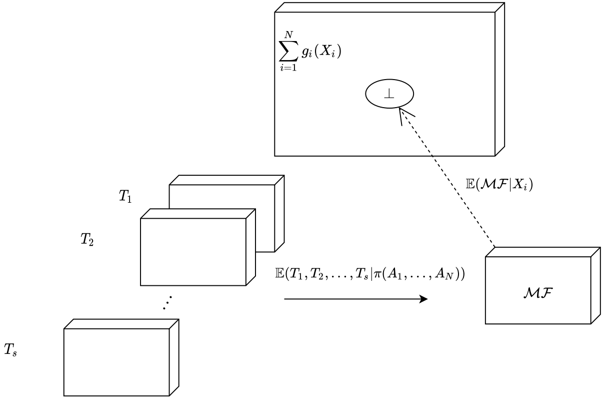


Figure 2: Hajek Projection (Hájek, 1968; Athey and Wager, 2019) of a (vector-valued) U-statistic.

Inference: Höeffding Decomposition II

More formally,

$$\underbrace{\Sigma^{-1/2}\big(\mathcal{M}\mathcal{F}-\mu\big)}_{Y}=\underbrace{\Sigma^{-1/2}\big(\dot{\mathcal{M}}\mathcal{F}-\mu\big)}_{\mu(X_i)}+\underbrace{\Sigma^{-1/2}\big(\mathcal{M}\mathcal{F}-\dot{\mathcal{M}}\mathcal{F}\big)}_{\varepsilon},$$

The objective:

$$\Sigma^{-1/2}\big(\mathcal{M}\mathcal{F}(x,A_1,\ldots,A_N)-\dot{\mathcal{M}}\mathcal{F}(x,A_1,\ldots,A_N)\big)\xrightarrow{p}0.$$

Inference: Asymptotic Theory

Lemma 1

The mean squared difference of $\mathcal{MF}$ *and* $\dot{\mathcal{M}}\mathcal{F}$ *has the upper bound:*
$\mathbb{E}\big(\mathcal{MF} - \dot{\mathcal{M}}\mathcal{F}\big)^T \Sigma^{-1} \big(\mathcal{MF} - \dot{\mathcal{M}}\mathcal{F}\big) \leq \frac{s}{N} tr\Big(\big(\mathbb{V}(\dot{\widetilde{T}})\big)^{-1} \mathbb{V}(\widetilde{T})\Big)$, *where* $\mathbb{V}(\dot{\widetilde{T}})$ *and* $\mathbb{V}(\widetilde{T})$ *denote the variance of the tree projection and a tree, respectively.* *Sketch of Proof*

Inference: Asymptotic Theory

Lemma 1

The mean squared difference of $\mathcal{MF}$ *and* $\dot{\mathcal{M}}\mathcal{F}$ *has the upper bound:* $\mathbb{E}\big(\mathcal{MF} - \dot{\mathcal{M}}\mathcal{F}\big)^T \Sigma^{-1} \big(\mathcal{MF} - \dot{\mathcal{M}}\mathcal{F}\big) \leq \frac{s}{N} tr\Bigg(\big(\mathbb{V}(\dot{\tilde{T}})\big)^{-1} \mathbb{V}(\tilde{T})\Bigg)$, *where* $\mathbb{V}(\dot{\tilde{T}})$ *and* $\mathbb{V}(\tilde{T})$ *denote the variance of the tree projection and a tree, respectively.* Sketch of Proof

Theorem 2

The entries of $\mathbb{V}(\tilde{T})$ *are bounded and its diagonal elements are bounded away from zero. Moreover, the lower bound of the off-diagonal terms of* $\mathbb{V}(\dot{\tilde{T}})$ *are on the order of* $o\big(\frac{1}{\log^p(s)}\big)$. *The upper bound in Lemma 1 converges to zero in the limit:* $\frac{s}{N} tr\Bigg(\big(\mathbb{V}(\dot{\tilde{T}})\big)^{-1} \mathbb{V}(\tilde{T})\Bigg) \rightarrow 0$. Sketch of Proof

Estimation: Method of Moments

Proposition 2.1

Deviation between the personalized and group level parameters:

$$\mathbb{E}\left[\left(\theta(X_i) - \widetilde{\theta}(X_i, S^{est}, \Pi)\right)^T \Sigma^{-1} \left(\theta(X_i) - \widetilde{\theta}(X_i, S^{est}, \Pi)\right)\right].$$

Quantile loss:

$$\sum_b \mathbb{E}(L_{\alpha_b}(Y_i, q)|X_i).$$

Sketch of Proof

where

$$L_{\alpha_b}(y, q) = \begin{cases} \alpha_b \times |y - q| \text{ if } y > q \\ (1 - \alpha_b) \times |y - q| \text{ if } y \leq q \end{cases}.$$

Simulation: NLSY 1979

Y_i is the years of schooling. D_i - Head Start program.

$$\begin{aligned} Y_i = 13.240 &- 0.5 \cdot D_i + \\ &\mathbb{1}\{\text{Parental Education}_i \geq 0.6\} \cdot D_i + X_i^R \cdot 0 + \varepsilon_i, \\ &\varepsilon_i \sim \mathcal{N}(0, 1 + D_i). \end{aligned}$$

Simulate 100 experiments with the upper quantiles of the outcome within $[1, 0.7]$, and the lower quantiles within $[0, 0.3]$.

Simulations: A Tree

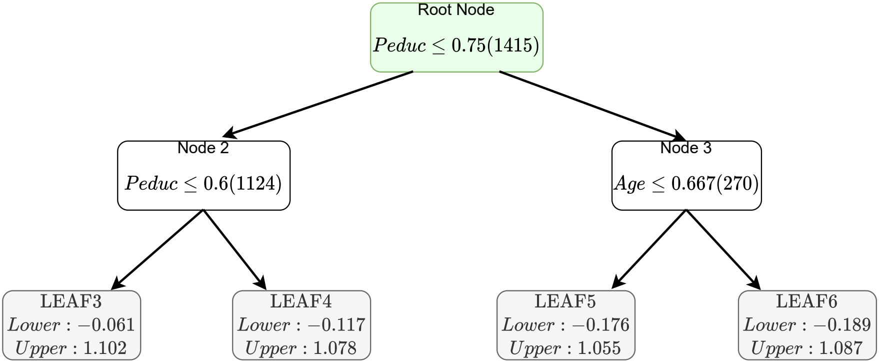
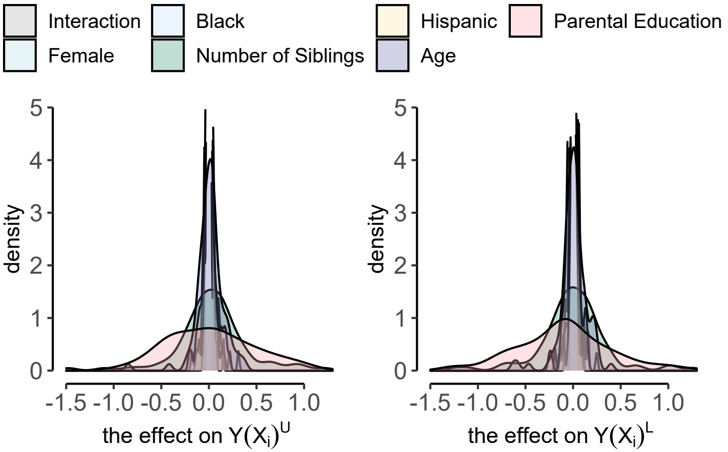


Figure 3: A single (randomly chosen) tree out of 200 trees.

Simulation: Splitting Accuracy

The correlation of parental education with age, the number of siblings, black, Hispanic and female have is -39%, -43%, -5%, 3%, and 2%, respectively.



Application: Treatment Effect Bounds

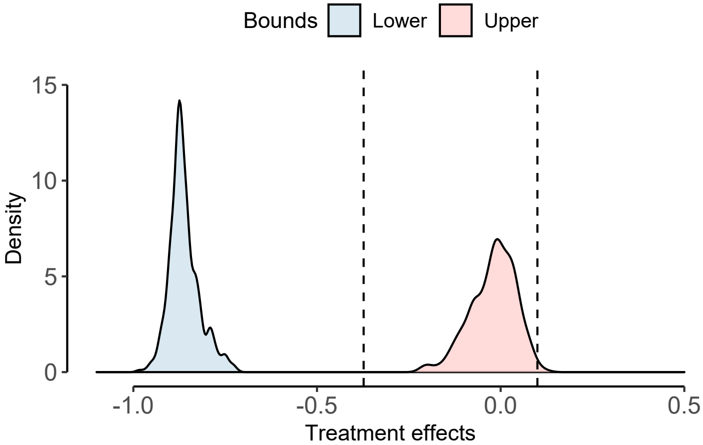
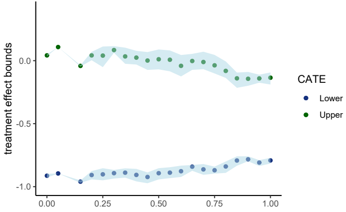
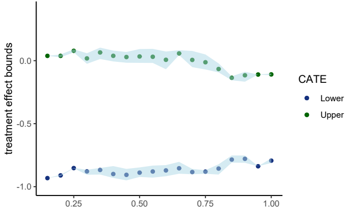


Figure 4: The density of the bounds for the effect of Head Start participation on years of schooling. We use 60 and 30-th percentiles of the outcome. Dashed lines represent unconditional bounds.

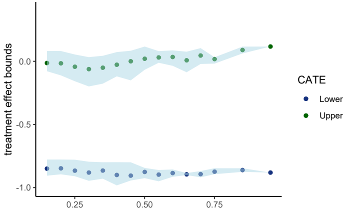
Application: Heterogeneity



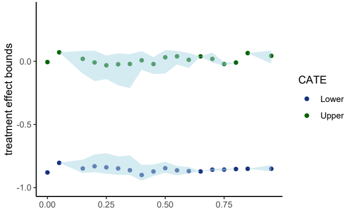
(a) P.education for female



(b) P.education for male



(c) N.Sibs for female



(d) N.Sibs for male

Future Work

This paper introduced

- Multivariate forests for partially identified CATE with an endogenous treatment,
- A test for heterogeneity - See
- Extension to anomalous data,
- Extension to network data.

Future Work

This paper introduced

- Multivariate forests for partially identified CATE with an endogenous treatment,
- A test for heterogeneity - See
- Extension to anomalous data,
- Extension to network data.

Empirical extensions:

- Investigate personalized effects of the education on earnings induced by the compulsory schooling reform in Norway (1959-1974).
- Investigate personalized effects of various cancer treatment methods in Norway.

Thank You!

E-mail: maria.nareklishvili@frisch.uio.no
Twitter: https://twitter.com/mnareklishvili