Heterogeneous Treatment Effects

Drew Dimmery

Structure

- General Definitions
- What do we know about best practices?
  - The alphabet soup of learners
- What tools can you use to do this?
- How can you peel open the black box?

The HTE Problem

- For each unit, we observe Y(1) xor Y(0)
- We want to estimate the supervised regression problem of:
  - Y(1) - Y(0) ~ X
  - This is impossible! Holland (1986)
- Lots of approaches, some of them are very good!

Assumptions

- Consistency: $Y_i = Y_i(A_i)$
- No Unmeasured Confounding: $A \perp (Y(1), Y(0)) \mid X$
- Positivity: $0 < \epsilon \leq \pi_i \leq 1-\epsilon < 1$ with probability 1

Under these assumptions, $\tau(x) = E[Y(1) - Y(0) \mid X = x]$

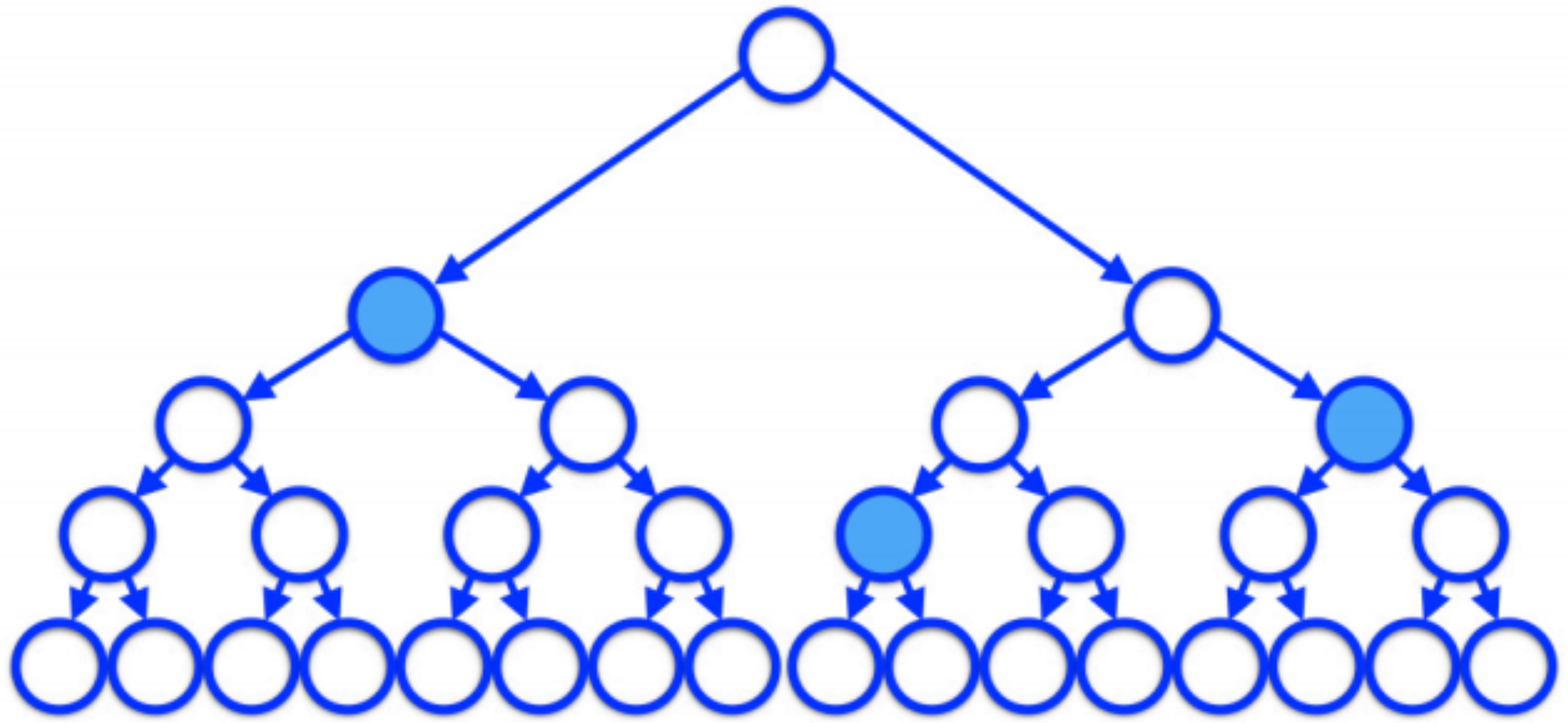
- Regularity conditions based on chosen model

Alphabet Soup

The Zoo of HTE Learners

S-learner

Single Regression model approach (e.g. Hill 2011)



- Easy to estimate!
- Treatment is just another feature
- Over-regularizes

Algorithm SI 2 S-learner

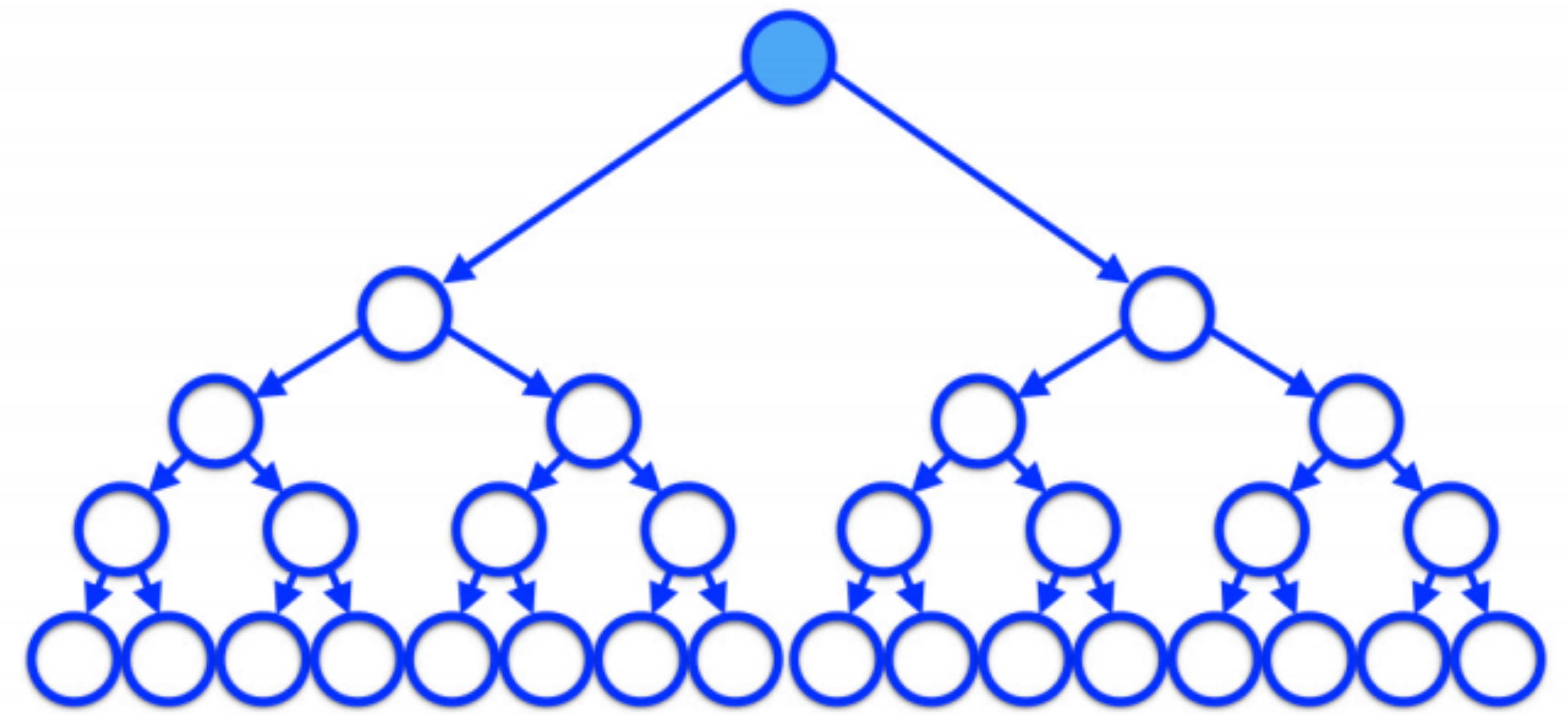
procedure S-LEARNER(X, Y, W)

$\quad \hat{\mu} = M(Y \sim (X, W))$

$\quad \hat{\tau}(x) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0)$

T-learner

Two Regression model approach (e.g. Athey et al 2015)



Algorithm SI 1 T-learner

1: **procedure** T-LEARNER(X, Y, W)

2: $\hat{\mu}_0 = M_0(Y^0 \sim X^0)$

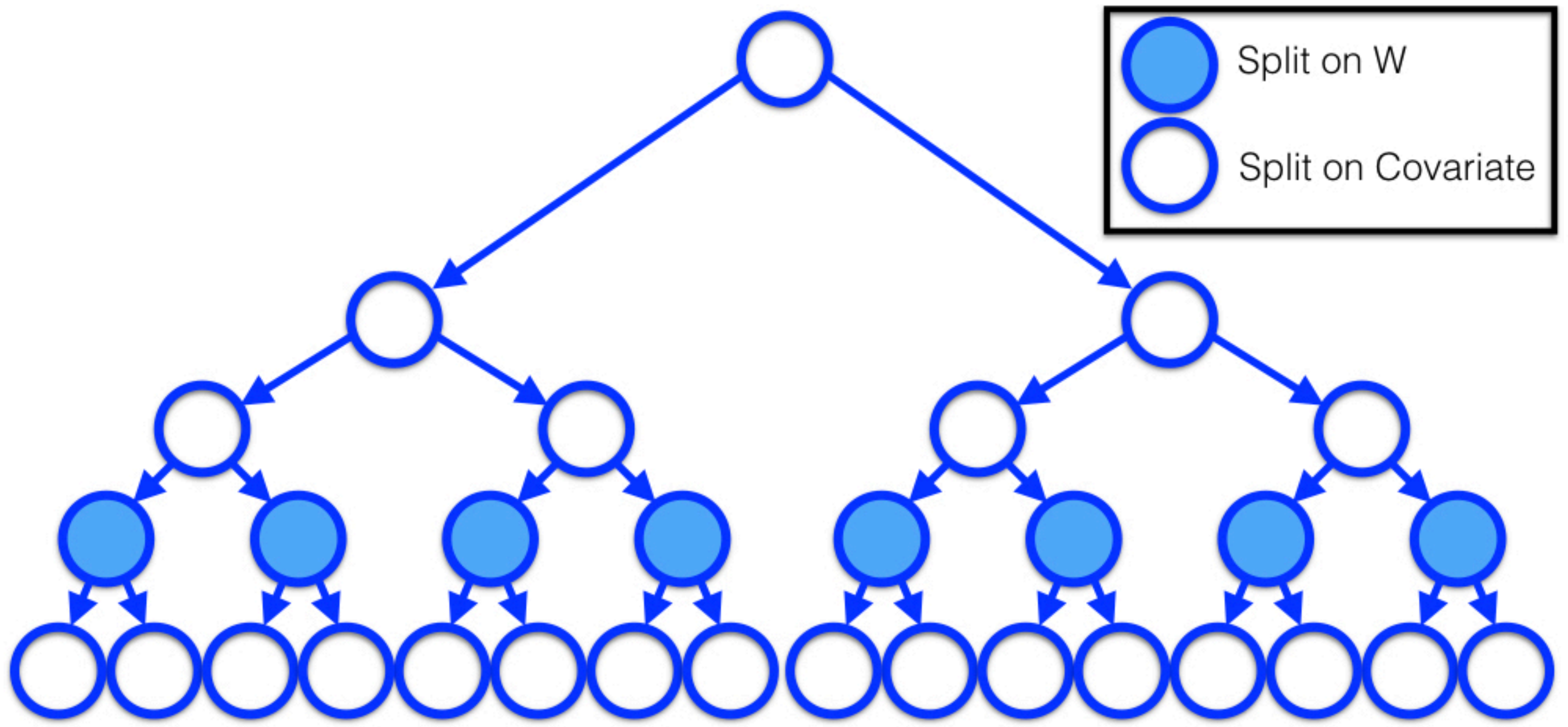
3: $\hat{\mu}_1 = M_1(Y^1 \sim X^1)$

4: $\hat{\tau}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x)$

- Easy to estimate!
- *Under*-regularizes

Causal Forests

Modify the splitting criteria of RFs (e.g. Wager and Athey 2017)



- There's software out there for you.
- You get a *very particular* form of unbiasedness!
- Important insight!
  - **HONESTY**

X-learner

T-learner then run some extra regressions (Künzel et al 2019)

Algorithm SI 3 X-learner

procedure X-LEARNER(X, Y, W, g)

$\hat{\mu}_0 = M_1(Y^0 \sim X^0)$ ▷ Estimate response function

$\hat{\mu}_1 = M_2(Y^1 \sim X^1)$

$\tilde{D}_i^1 = Y_i^1 - \hat{\mu}_0(X_i^1)$ ← **Estimates CATT** ▷ Compute imputed treatment effects

$\tilde{D}_i^0 = \hat{\mu}_1(X_i^0) - Y_i^0$ ← **Estimates CATC**

$\hat{\tau}_1 = M_3(\tilde{D}^1 \sim X^1)$ ▷ Estimate CATE in two ways

$\hat{\tau}_0 = M_4(\tilde{D}^0 \sim X^0)$

$\hat{\tau}(x) = g(x)\hat{\tau}_0(x) + (1 - g(x))\hat{\tau}_1(x)$ ▷ Average the estimates

- Now we're getting somewhere!
- More complicated and not doubly-robust
- Regularizes reasonably!
- Under unconfoundedness, CATT = CATC = CATE

The Approach

Kennedy (n.d.)

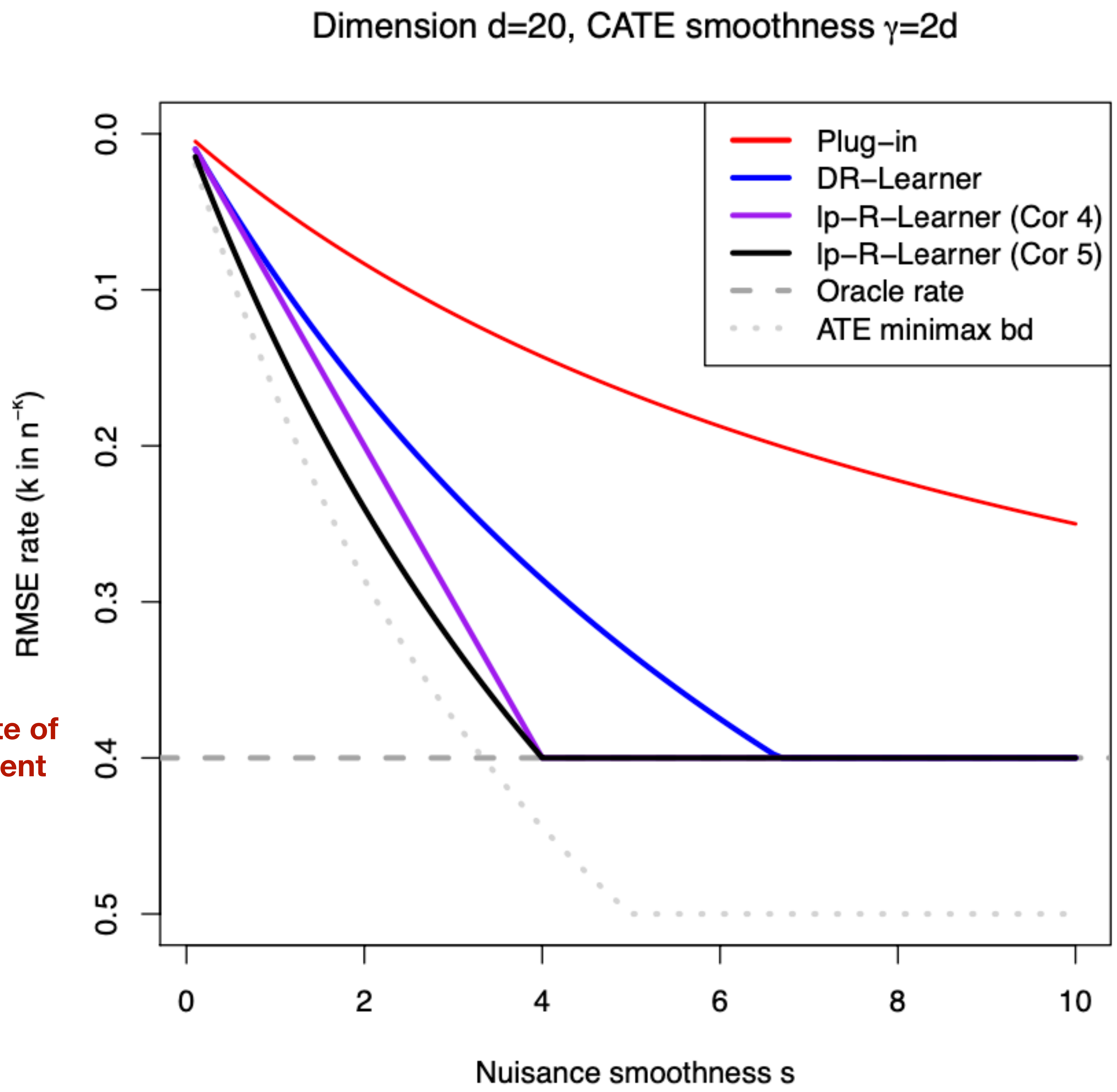
- Nuisance training
  - Propensity score (for us: known)
  - Regression functions (Y^0 / Y^1)
- Make an unbiased estimate of each unit's CATE (pseudo-outcome)

$$\widehat{\varphi}(Z) = \frac{A - \widehat{\pi}(X)}{\widehat{\pi}(X)\{1 - \widehat{\pi}(X)\}}\left\{Y - \widehat{\mu}_A(X)\right\} + \widehat{\mu}_1(X) - \widehat{\mu}_0(X)$$

Unbiased estimate of the unit's treatment effect

- Smooth pseudo-outcome **Smoothing**
- Training nuisance models and smoothing should be on separate subsamples (**cross-fitting**)

Optimality!

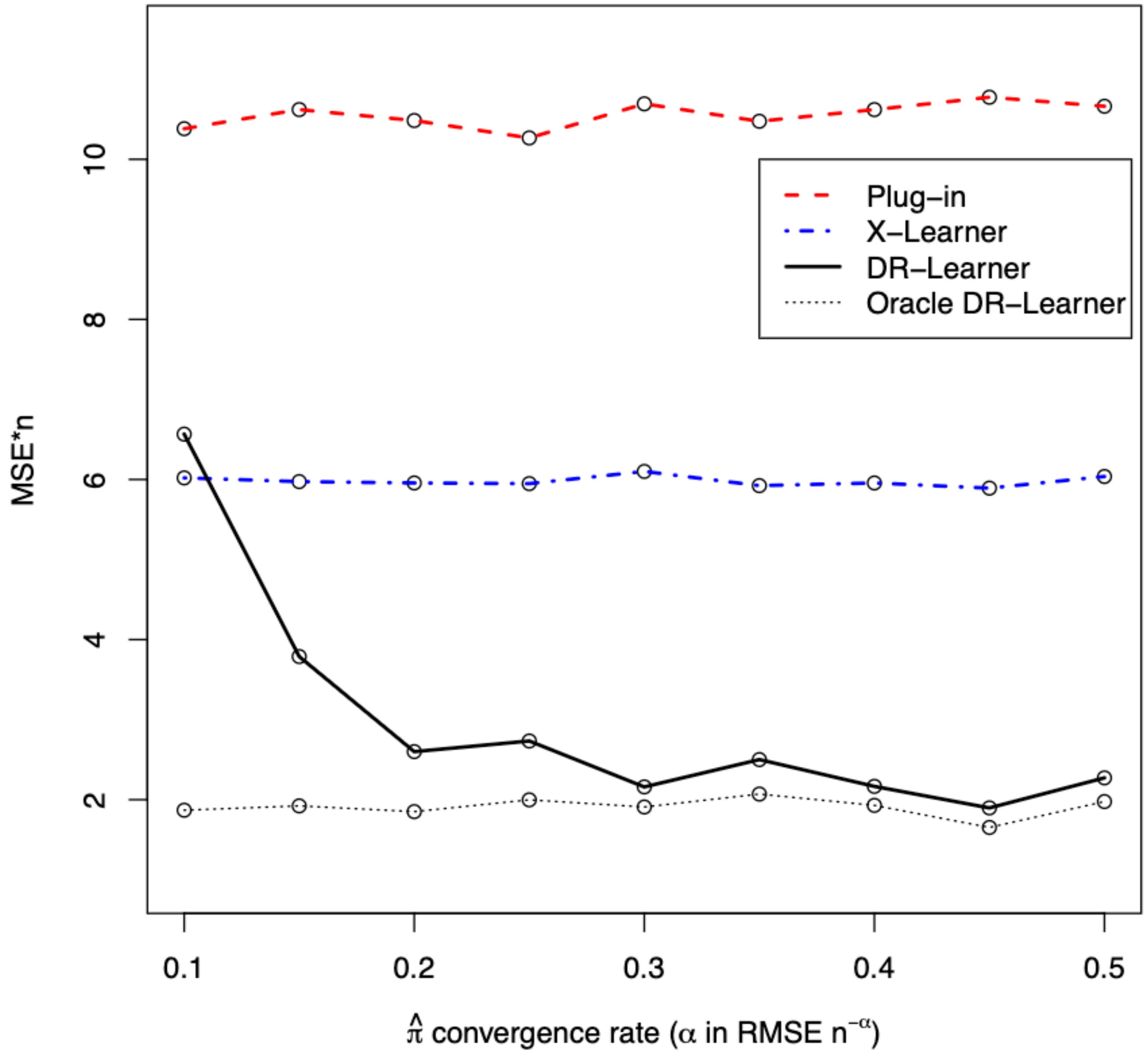


Why does this work?

- Pseudo-outcome is the uncentered *efficient influence function* of the ATE
- Estimation is harder than the ATE by a factor increasing in dimension of covariates and decreasing in smoothness of the true CATE function

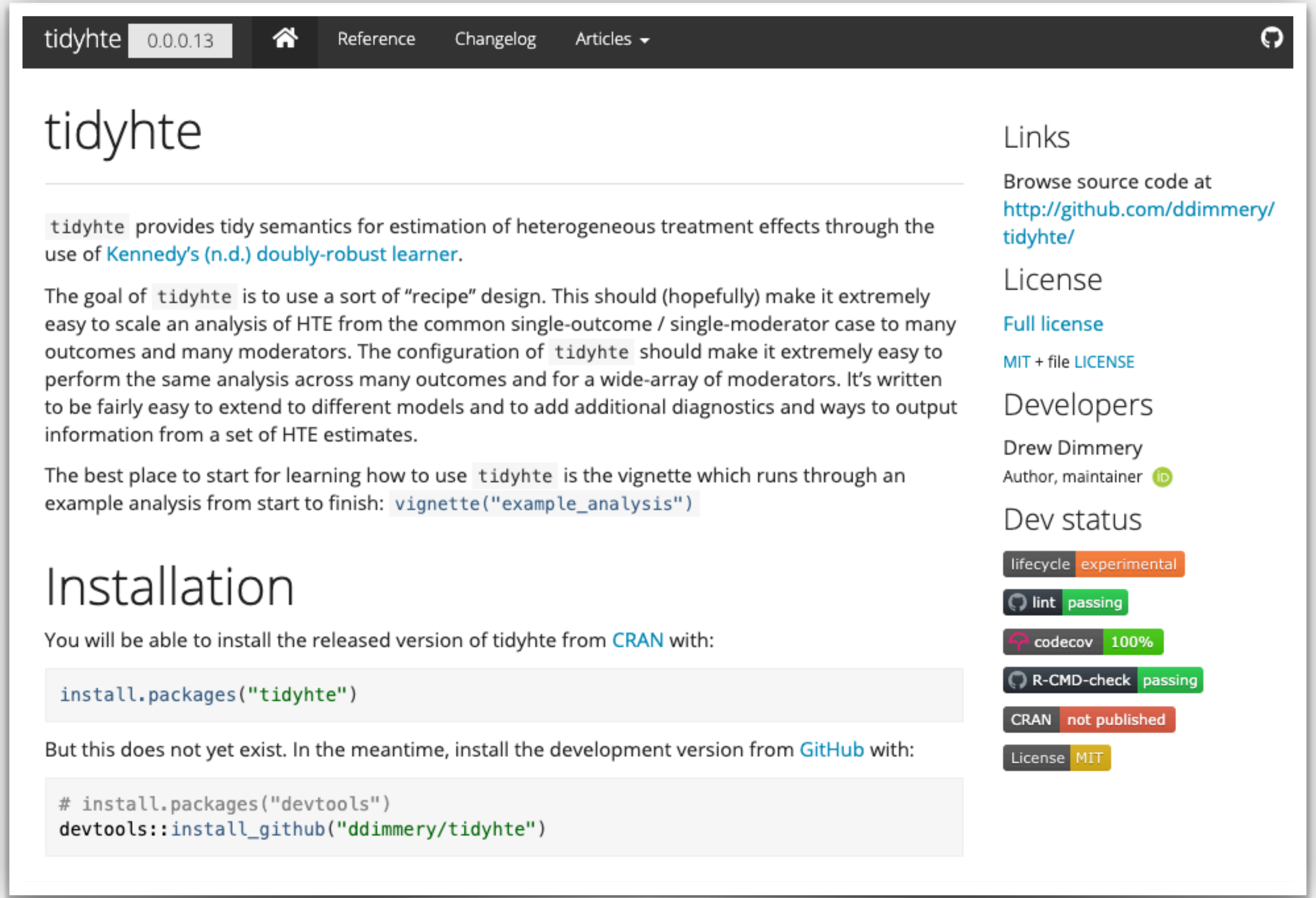
Why does it work better than X-learner?

- X-learner doesn't benefit from convergence in propensity score
- In these experiments we *know* the true propensity score!
- We're using our outcome models as a *control variate* to reduce variance.
    - For a stratified second stage model, *this is just AIPW*.



Tools

Introducing `tidyhte`



tidyhte 0.0.0.13 Reference Changelog Articles

tidyhte

`tidyhte` provides tidy semantics for estimation of heterogeneous treatment effects through the use of Kennedy's (n.d.) doubly-robust learner.

The goal of `tidyhte` is to use a sort of "recipe" design. This should (hopefully) make it extremely easy to scale an analysis of HTE from the common single-outcome / single-moderator case to many outcomes and many moderators. The configuration of `tidyhte` should make it extremely easy to perform the same analysis across many outcomes and for a wide-array of moderators. It's written to be fairly easy to extend to different models and to add additional diagnostics and ways to output information from a set of HTE estimates.

The best place to start for learning how to use `tidyhte` is the vignette which runs through an example analysis from start to finish: `vignette("example_analysis")`

Installation

You will be able to install the released version of tidyhte from CRAN with:

```
install.packages("tidyhte")
```

But this does not yet exist. In the meantime, install the development version from GitHub with:

```
# install.packages("devtools")
devtools::install_github("ddimmery/tidyhte")
```

Links

Browse source code at http://github.com/ddimmery/tidyhte/

License

Full license

MIT + file LICENSE

Developers

Drew Dimmery

Author, maintainer

Dev status

lifecycle experimental

lint passing

codecov 100%

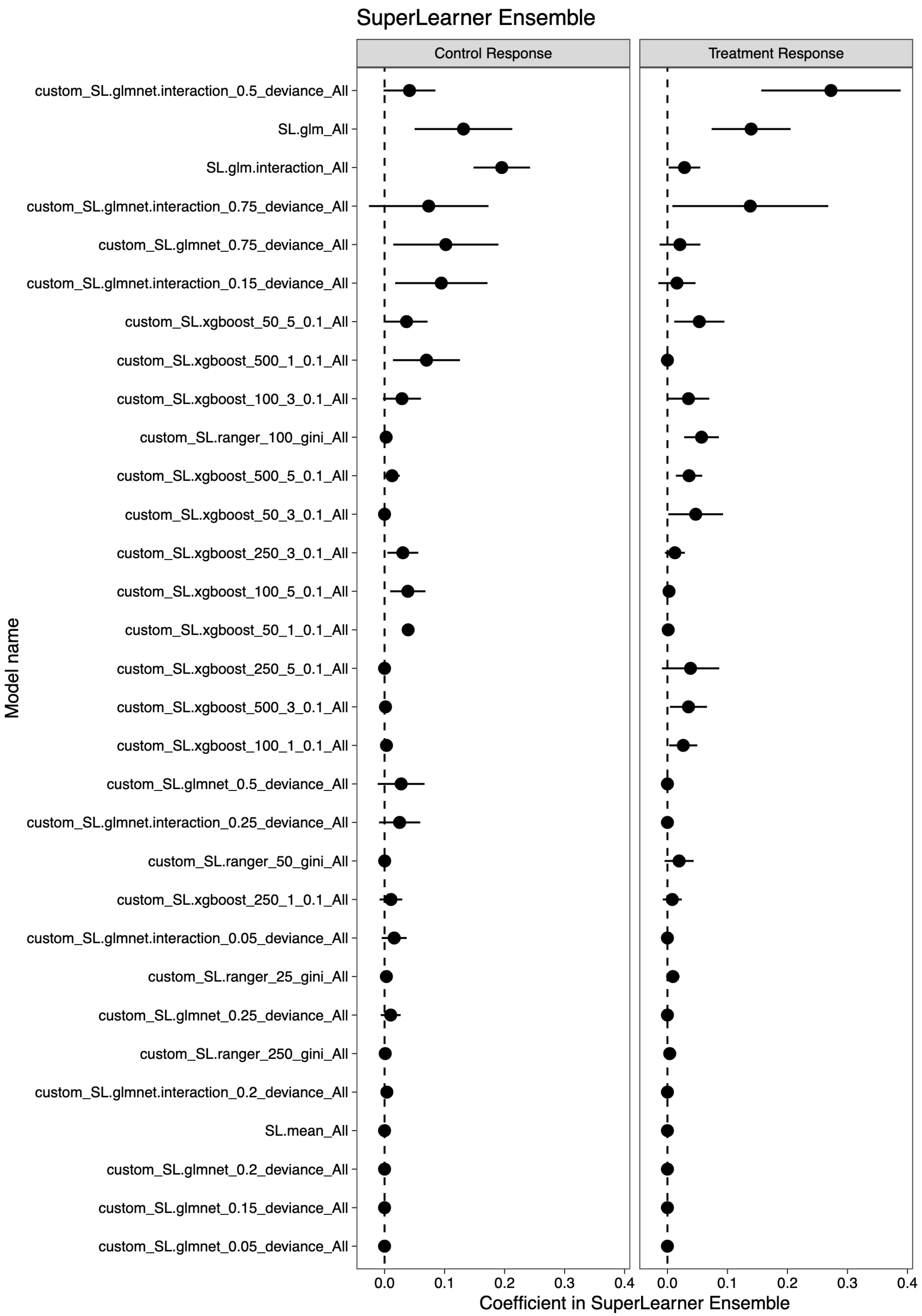
R-CMD-check passing

CRAN not published

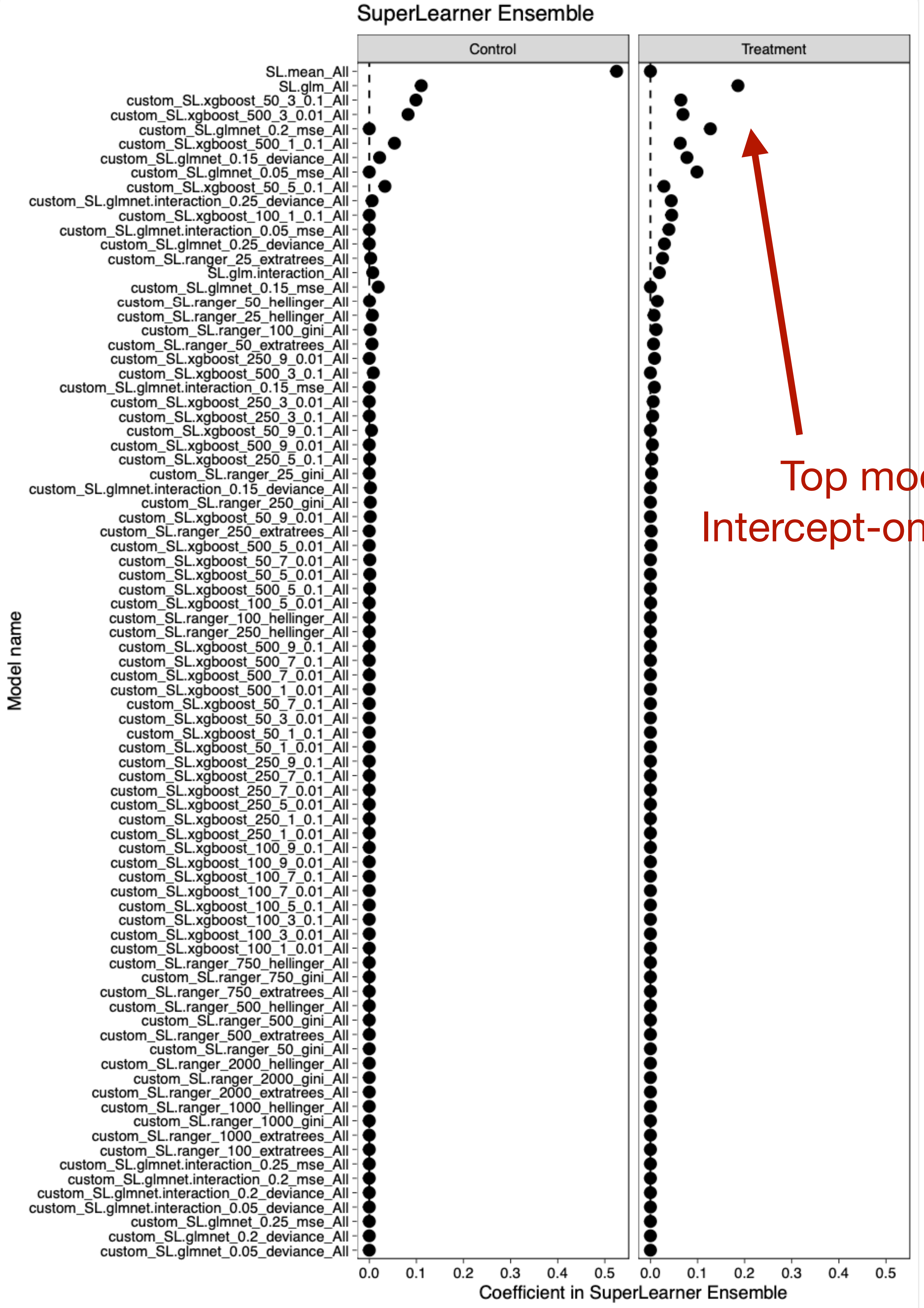
License MIT

Nuisance Models

- SuperLearner to learn an ensemble of machine learning models.
- Component models:
  - Intercept-only
  - OLS
  - OLS + 2-way interactions
  - Elastic Net
  - Elastic Net + 2-way interactions
  - Random forests (up to 2000 trees)
  - GBDTs (up to 500 iterations)

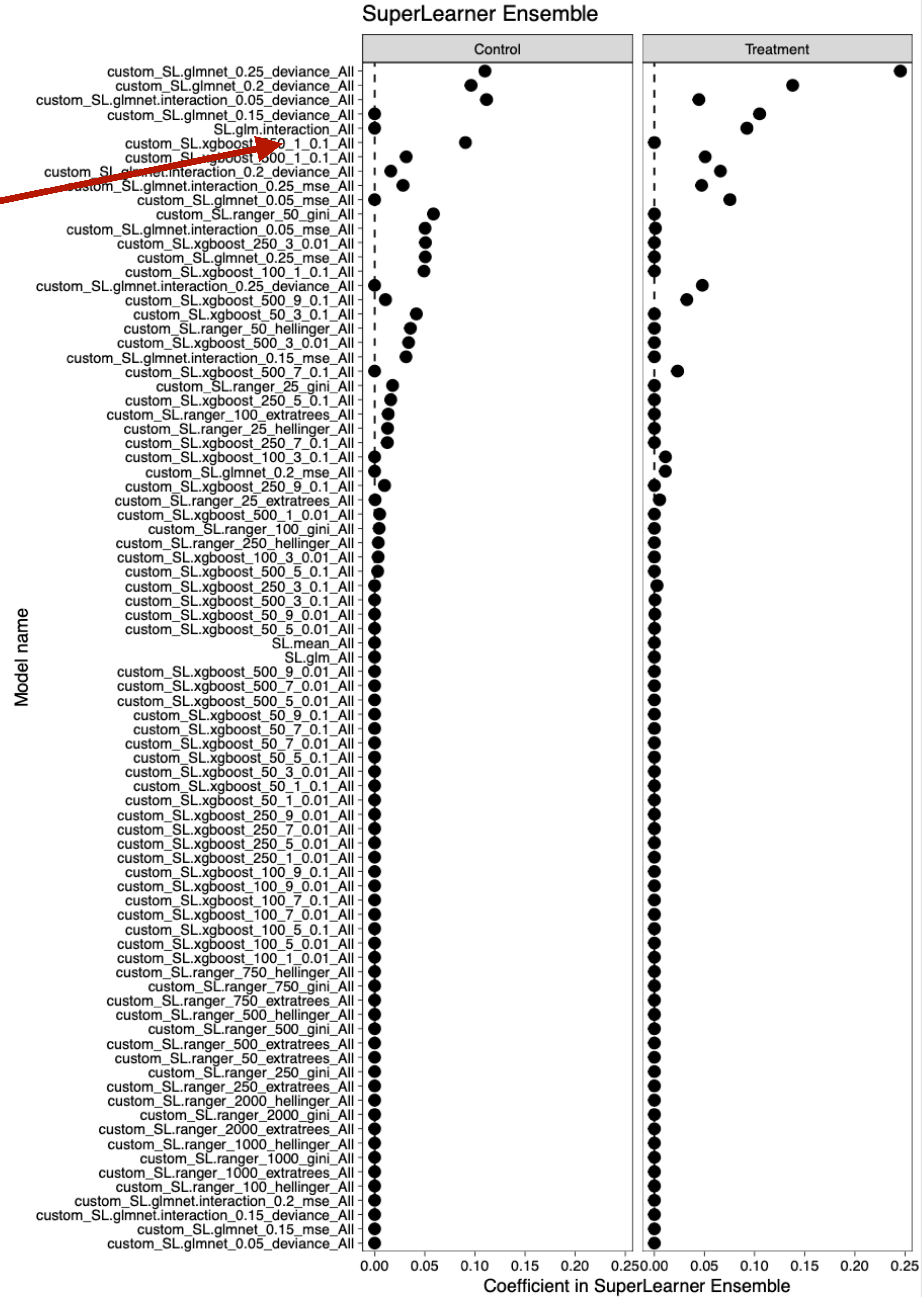


Low heterogeneity



Top models are
Intercept-only and OLS.

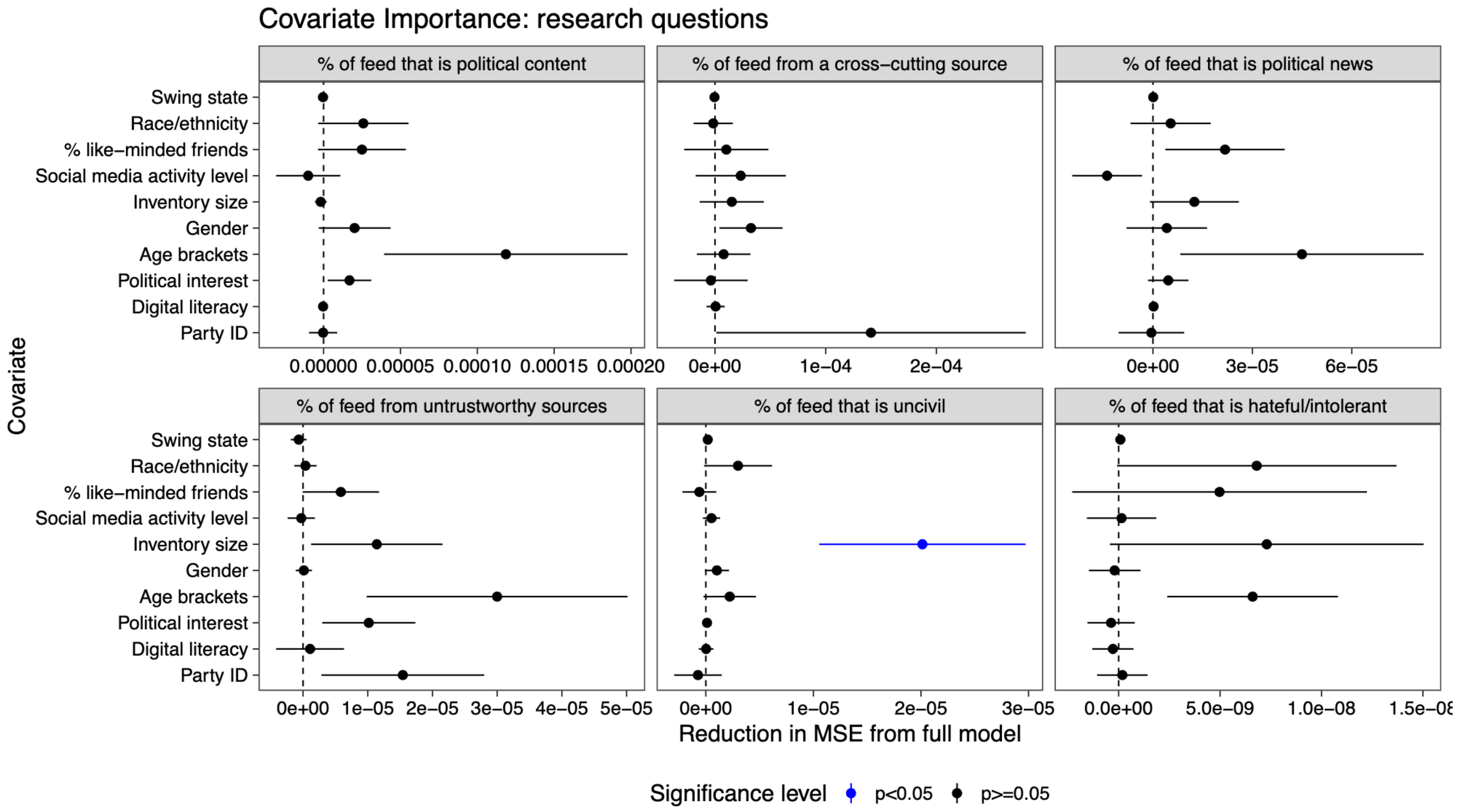
High heterogeneity



Top models are
interaction-heavy.

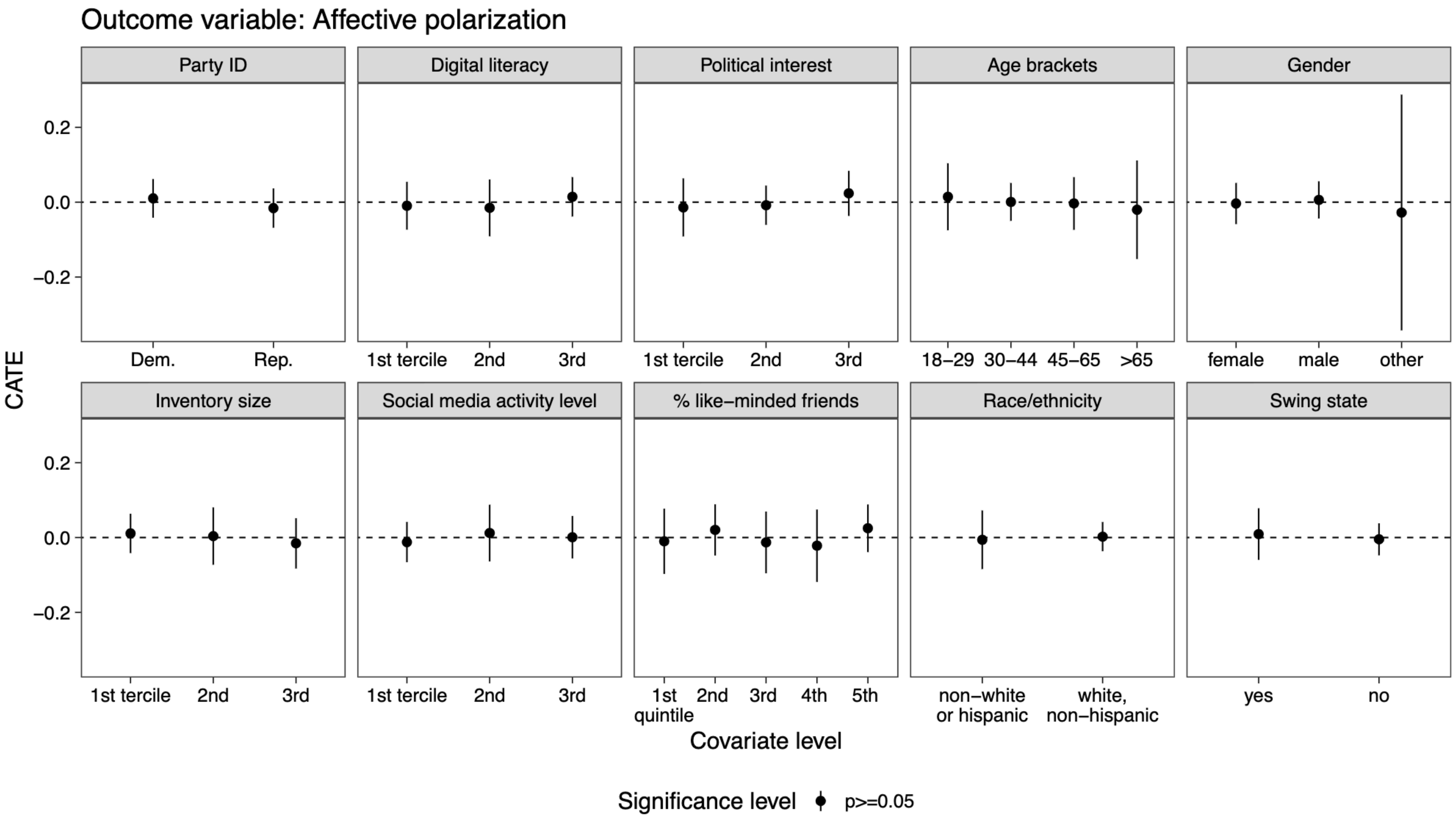
Variable Importance

- R^2 - based feature importance on the *pseudo-outcome*.
- Fully semiparametric as in nuisance function estimation.
- Williamson, Gilbert, Carone and Simon (2020) "Nonparametric variable importance assessment using machine learning techniques" *Biometrics*
- Shows the reduction in R^2 from removing a given covariate from a joint model of HTE.
- A well-defined quantity, but not quite as causal as you'd probably like.

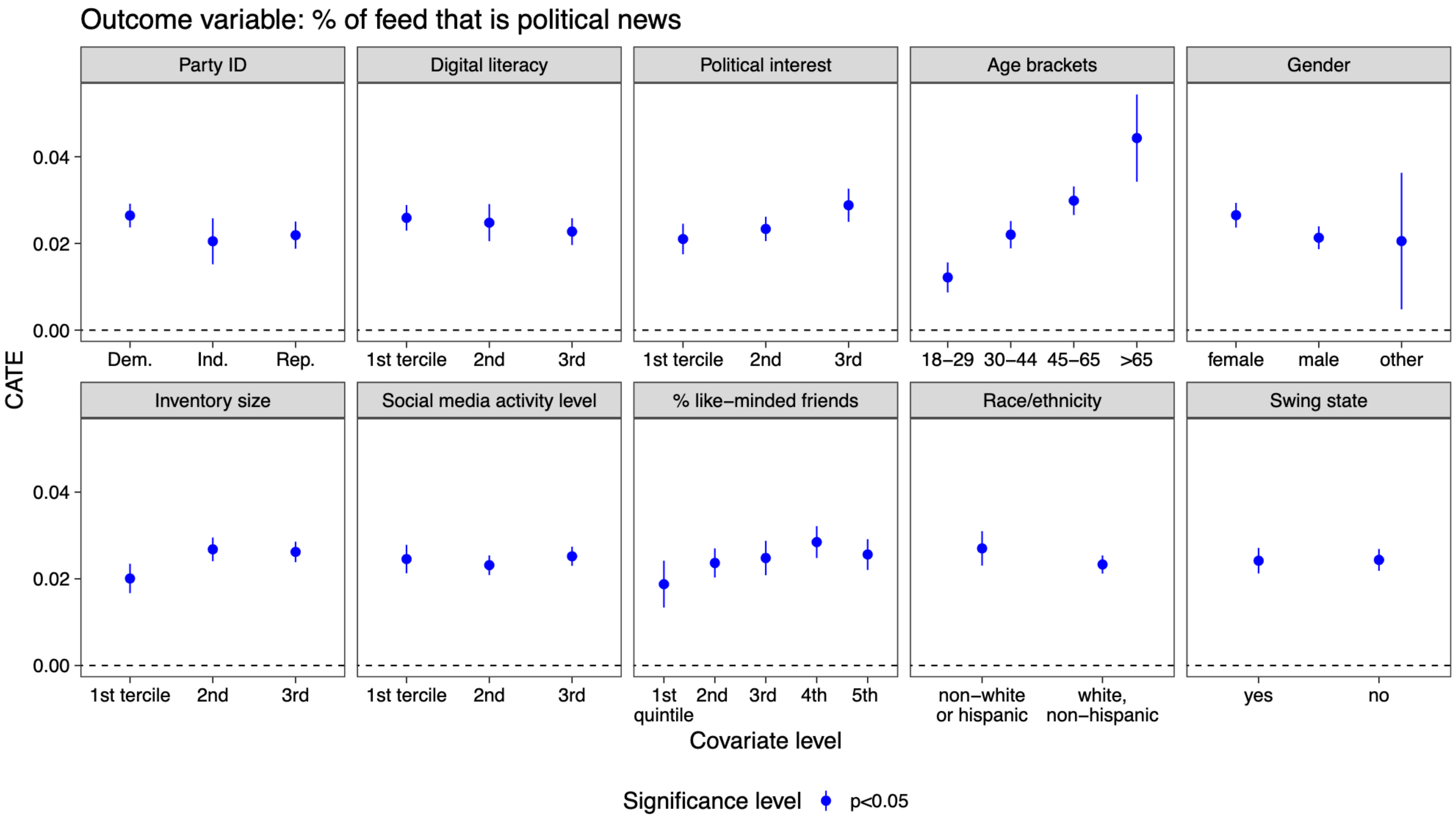


HTEs

- About to show "Marginal" CATEs:
  - One-dimensional slices of the HTEs.
  - This is simply the *average treatment effect at the given covariate level*.
  - This does not disentangle which covariates are ✌️driving✌️ heterogeneity.



Outcome variable: Affective polarization
Party ID
Digital literacy
Political interest
Age brackets
Gender
Inventory size
Social media activity level
% like−minded friends
Race/ethnicity
Swing state
CATE
0.2
0.0
−0.2
Dem.
Rep.
1st tercile
2nd
3rd
18−29
30−44
45−65
>65
female
male
other
1st quintile
4th
5th
non−white or hispanic
white, non−hispanic
yes
no
Covariate level
Significance level
p>=0.05


Outcome variable: % of feed that is political news
Party ID
Digital literacy
Political interest
Age brackets
Gender
CATE
0.04
0.02
0.00
Dem.
Ind.
Rep.
1st tercile
2nd
3rd
1st tercile
2nd
3rd
18–29
30–44
45–65
>65
female
male
other
Inventory size
Social media activity level
% like–minded friends
Race/ethnicity
Swing state
1st tercile
2nd
3rd
1st tercile
2nd
3rd
1st quintile
2nd
3rd
4th
5th
non–white or hispanic
white, non–hispanic
yes
no
Covariate level
Significance level
p<0.05

Vignettes

ddimmery.github.io/tidyhte

- `devtools::install_github("ddimmery/tidyhte")`
- `vignette("experimental_analysis")`