Personalised Medicine

Methods to Estimate Personalised Treatment Effects in Observational Data

Daniela Buchwald

14 December 2018

Overview

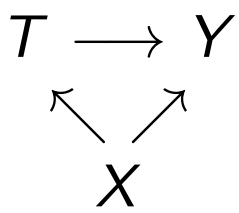
1. The Propensity Score for Observational Data
2. Estimating Personalised Treatment Effects

Part 1

The Propensity Score for Observational Data

Introduction

Goal: Estimation of the causal effect of a treatment T on the outcome Y given a confounder X:



T: Treatment (y/n)

Y: Outcome (y/n)

X: Confounder

$\Rightarrow$ Create a pseudo-randomized sample by eliminating the backdoor path from T to Y, i.e. erasing the arrow from X to T.

Propensity Score - A balancing Score

$$\text{Propensity Score} = \mathbb{P}(T = 1|\mathbf{X})$$

1. Estimate propensity score per person with Logistic Regression, Machine Learning Methods, ...
2. Estimate the treatment effect with propensity adjustment:

```
model <- glm(Y ~ T + X, *)
```

Method	**Model adjustment**
1) Inverse Probability of Treatment Weighting (IPTW)	$*$ = Weights
2) Matching	$*$ = Matched Dataset
3) Stratification	$*$ = Separate Models
4) Include as a Covariate	$*$ = Covariate "Propensity Score"

Motivating example

- $\mathbb{P}(T = 1|X = \mathit{Female}) = 0.1$:

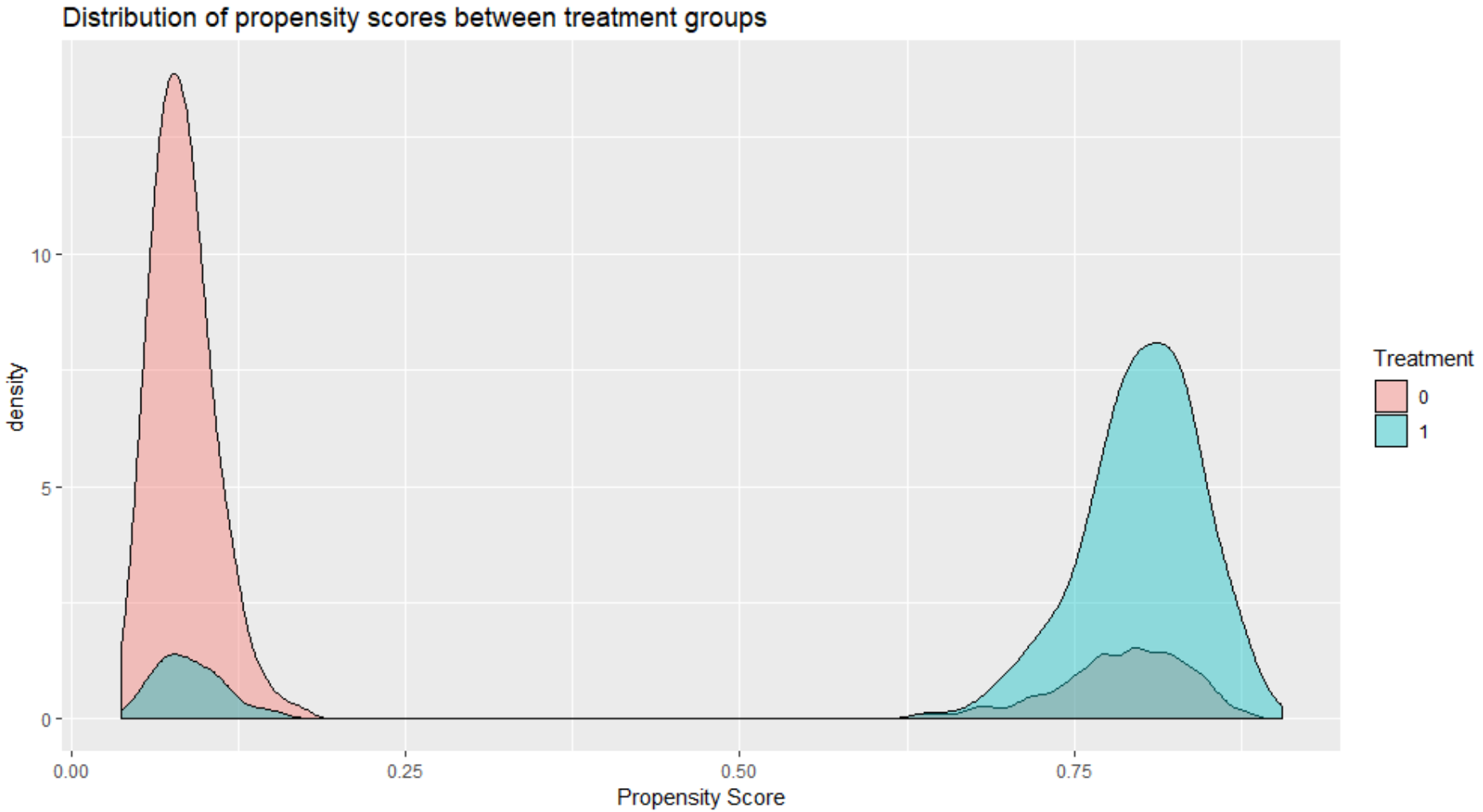
  Among women 10% receive the treatment

- $\mathbb{P}(T = 1|X = \mathit{Male}) = 0.8$:

  Among men 80% receive the treatment

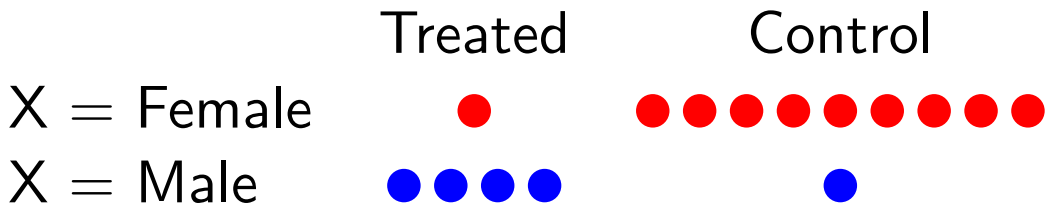
First Step: Estimating Propensity Score

Use of Logistic Regression:



Inverse Probability of Treatment Weighting (IPTW)

Original Data:



Assign a weight w to each patient:

$$w = \frac{T}{\mathbb{P}(T = 1|\mathbf{X})} + \frac{1 - T}{1 - \mathbb{P}(T = 1|\mathbf{X})}$$

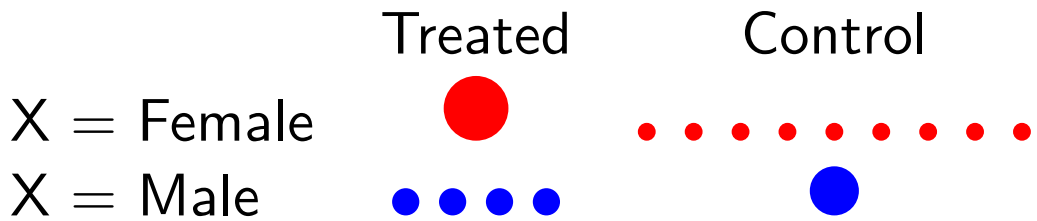
$\rightarrow$ Patients who are under-represented get more weight and vice versa

Inverse Probability of Treatment Weighting (IPTW)

Here:

	Treated	Control
X = Female	$\frac{1}{\mathbb{P}(T=1\|X=F)} = \frac{1}{0.1} = 10$	$\frac{1}{1-\mathbb{P}(T=1\|X=F)} = \frac{1}{0.9} = 1.11$
X = Male	$\frac{1}{\mathbb{P}(T=1\|X=M)} = \frac{1}{0.8} = 1.25$	$\frac{1}{1-\mathbb{P}(T=1\|X=M)} = \frac{1}{0.2} = 5$

Weighted Data:




```
model <- glm(Y ~ T + X, weights = w)
```

Inverse Probability of Treatment Weighting (IPTW)

Standardized Mean Differences:

$$\text{SMD} = \frac{\bar{x}_{treatm} - \bar{x}_{control}}{\sqrt{\frac{s^2_{treatm} - s^2_{control}}{2}}}$$

Data	Variable	n (%) (t = 0)	n (%) (t = 1)	SMD
Original	Gender (M)	204 (18.3)	801 (90.5)	2.105
Weighted	Gender (M)	996.3 (50.0)	1004.2 (49.6)	0.009

Inverse Probability of Treatment Weighting (IPTW)

Pros:

- Easy to implement
- Retains the whole data set

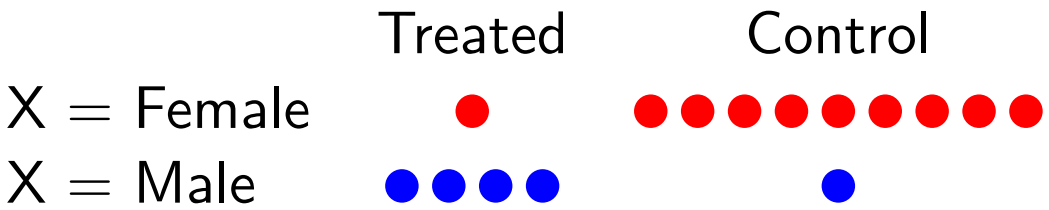
Cons:

- Can lead to high variance in estimated treatment effects with propensity scores close to 0 or 1 ($\rightarrow$ very large weights)

$\rightarrow$ Possibility to trim weights

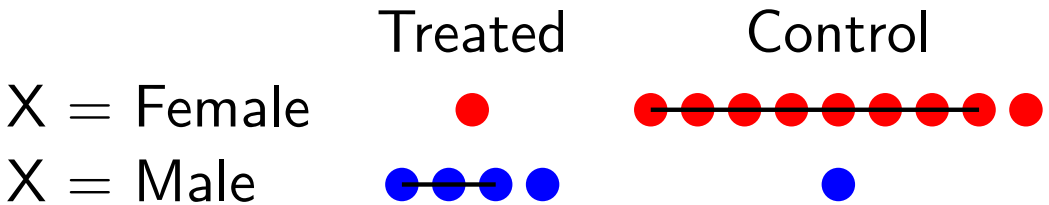
Matching

Original Data:



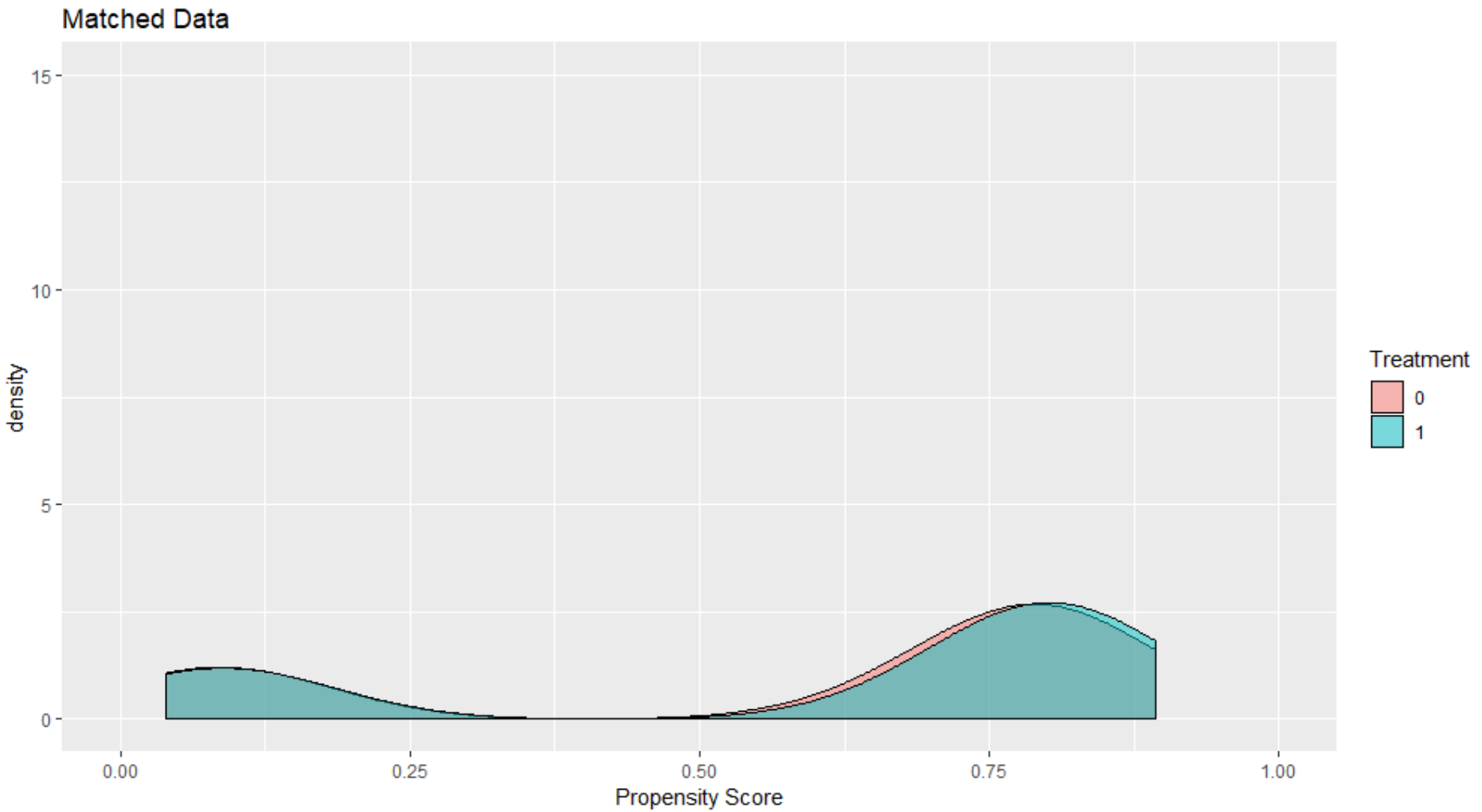
Choose pairs of patients between treatment and control group with similar ("1:1 nearest neighbor") propensity scores (on a logit scale) and discard the unmatched patients.

Matched data:




```
model <- glm(Y ~ T + X, data = matcheddata)
```

Matching


Matched Data
15
10
5
0
density
0.00
0.25
0.50
0.75
1.00
Propensity Score
Treatment
0
1

Matching

Pros:

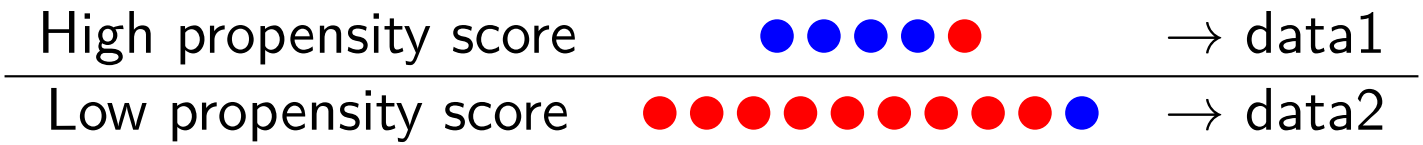
- Good balance
- Simple to analyse

Cons:

- Discard individuals
- Choose the right matching method ("Nearest neighbor", "Optimal", "Exact", "Full", combination with Mahalanobis metric matching, ...)

Stratification

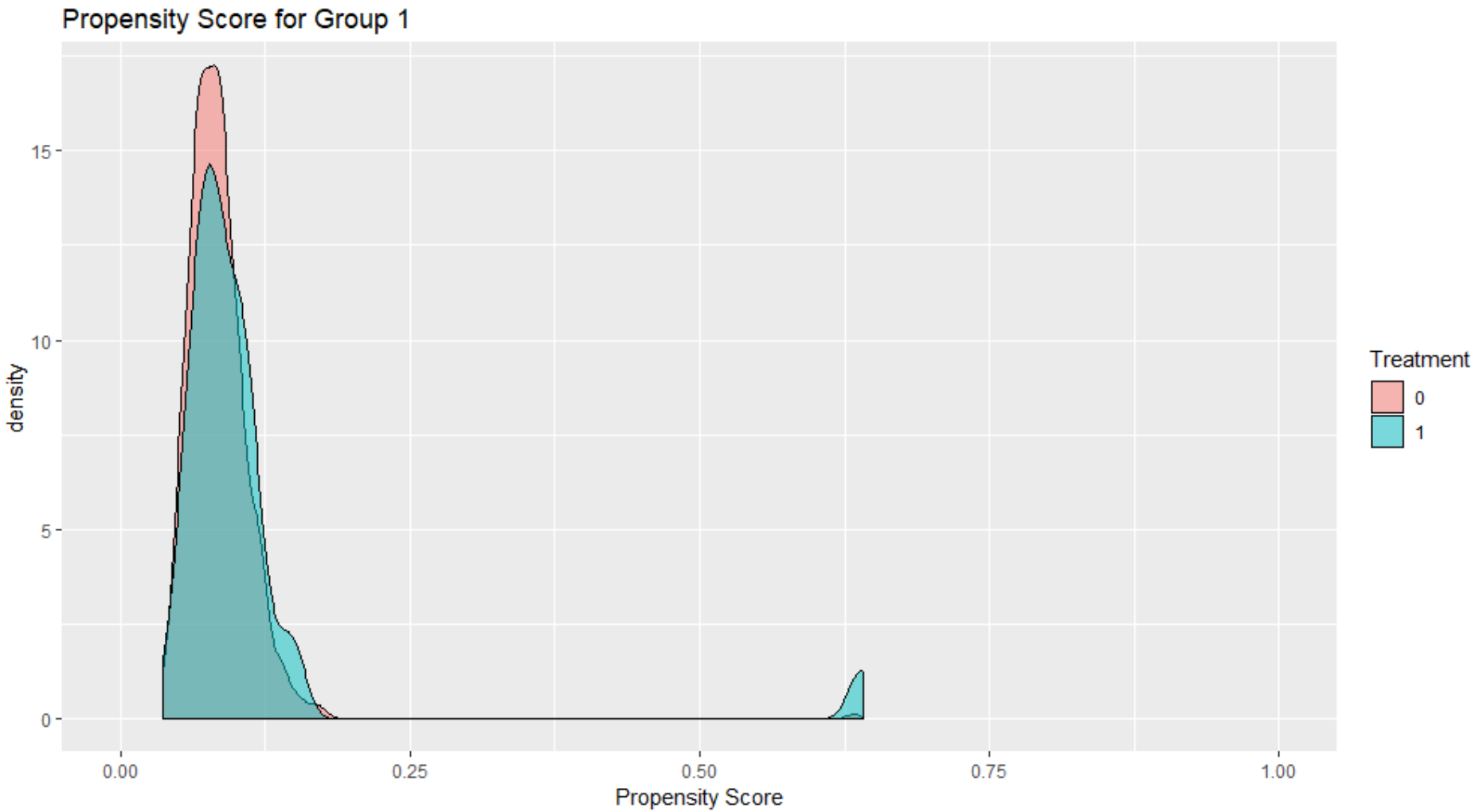
Group patients into bins of similar propensity score and fit model within each bin



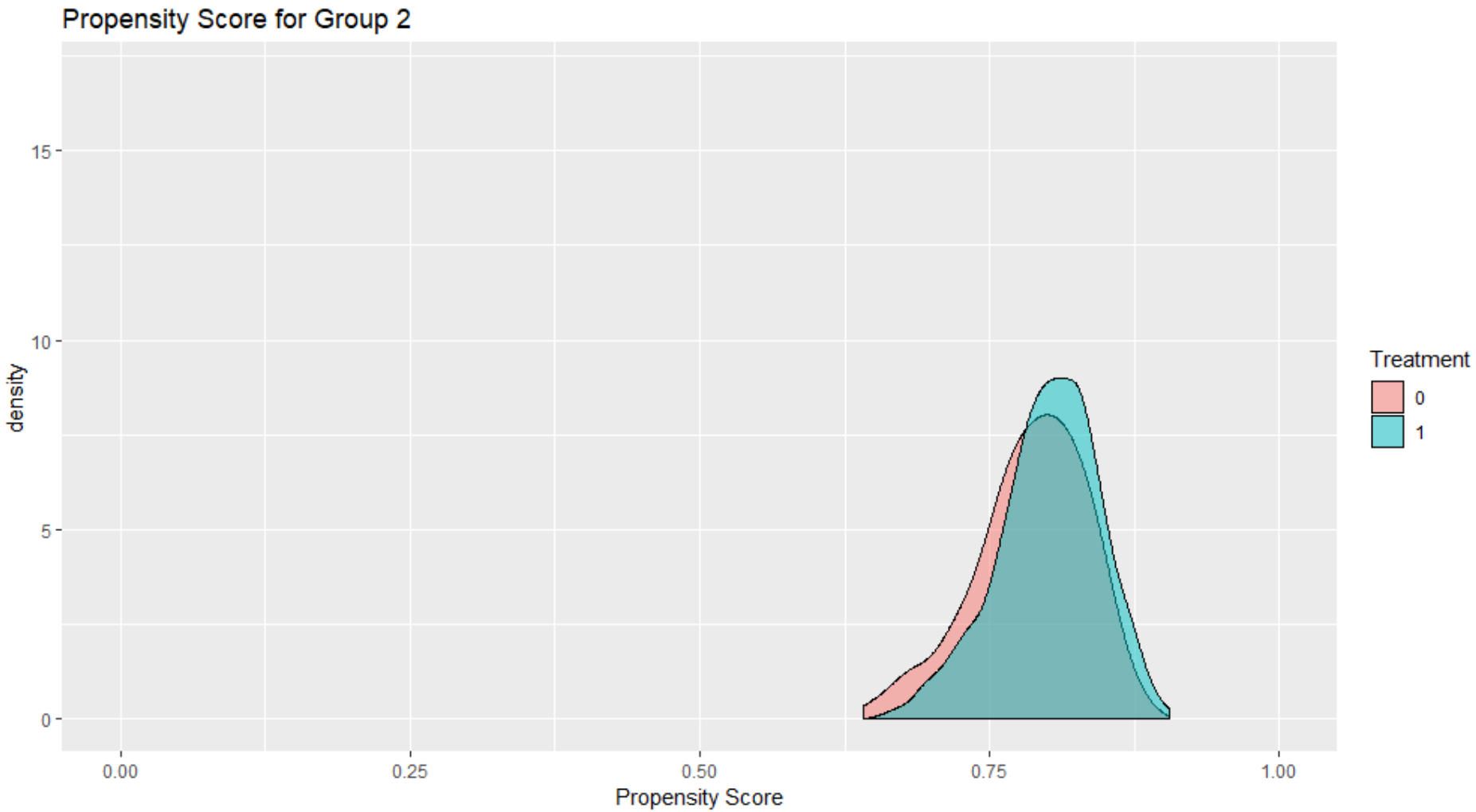

```
model1 <- glm(Y ~ T + X, data = data1)

model2 <- glm(Y ~ T + X, data = data2)
```

Stratification



Stratification



Stratification

Pros:

- Works well for not too unbalanced datasets
- Provides effect estimates for every stratum

Cons:

- Choice of number and size of strata (Trade-off between control of confounding and sufficient number of observations in each strata)

Include Propensity Score as a Covariate

- Until now: Balance covariates between treatment and control group
- Now: Covariate adjustment in the prediction model which aims to control for covariate effects (confounding)

```
model <- glm(Y ~ T + X + PS)
```

Include Propensity Score as a Covariate

Cons:

- Like traditional covariate adjustment

Pros:

- But over-parameterizing is no problem

Note

Causal Inference using propensity score requires certain assumptions:

- **Unconfoundedness** (Exchangeability, Ignorability): There are no unmeasured confounders

$$\{Y(0), Y(1)\} \perp\!\!\!\perp T|X$$

- **Consistency**: Subject's potential outcome under treatment received = Observed Outcome
- **Positivity**: $0 < \mathbb{P}(T = 1) < 1$
- **No misspecification** of the propensity score model

Part 2

Estimating Personalised Treatment Effects

Methods for Personalised Treatment Effect Estimation

1. Weighting
   - **Model-based Trees (MOB Trees)**
   - **Causal Trees**
   - Causal Forests
   - Pollinated Transformed Outcome (PTO) Forests
2. Stratification
   - Causal Mulitivariate Adaptive Regression Splines (Causal MARS)
   - Causal Boosting
3. Include Propensity Score as Covariate
   - **Bayesian Additive Regression Trees (BART)**

MOB Trees

Model-based recursive partitioning for automated detection of patient subgroups

Algorithm:

1. Fit parametric model $\mathcal{M}((Y, \mathbf{X}), \theta)$ to a dataset with all observations with:

$$\theta = \begin{pmatrix} \alpha \\ \tau \end{pmatrix} \begin{matrix} \text{intercept} \\ \text{treatment effect} \end{matrix}$$

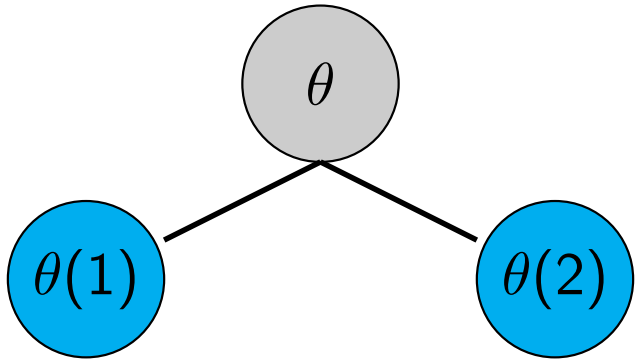
Here: Model = GLM with logit-Link

2. Test for parameter instability over a set of partitioning variables

MOB Trees

3. If there is some overall parameter instability, select the variable associated with the highest parameter instablity (i.e. the lowest p-value), otherwise stop
4. Repeat the procedure in each of the resulting subsamples

$\Rightarrow$ Fit separate models $\mathcal{M}((X, Y), \theta_b)$ in partitions




```
glmt <- glmtree(Y ~ T | X, data, family = binomial)
```

MOB Trees and IPTW - Data Example

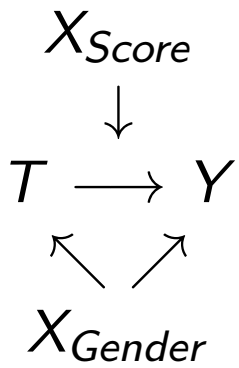
- n = 2000
- p = 20 (10 binary, 10 numeric)
- Confounder: X_{Gender}

$$\mathbb{P}(T = 1|\mathbf{X}) = \begin{cases} 0.8 & X_{Gender} = \text{Male} \\ 0.1 & X_{Gender} = \text{Female} \end{cases}$$

- Treatment Effect: X_{Score}

$$\tau(x) = \begin{cases} 5 & X_{Score} < 0 \\ 0 & X_{Score} > 0 \end{cases}$$

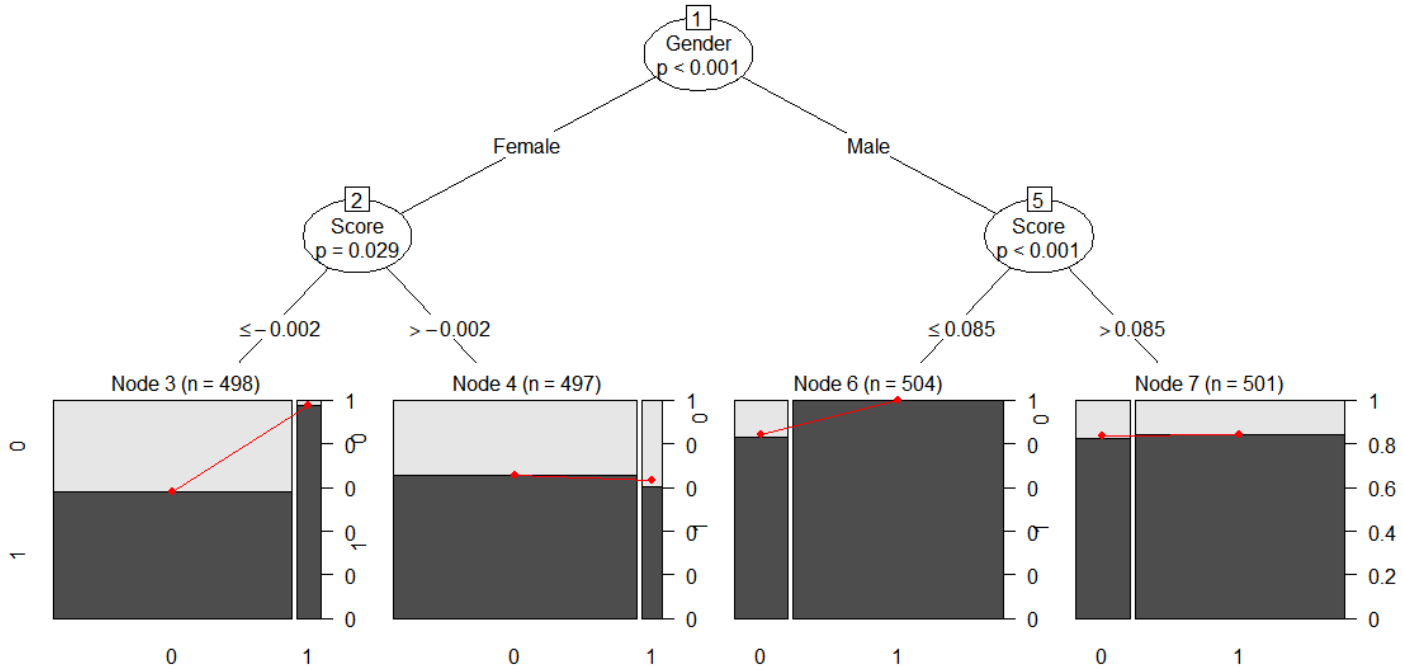
MOB Trees and IPTW - Possible results



$$\text{logit}(\mathbb{P}(Y = 1|T, X)) = \underbrace{0.5 + 1.2 \cdot I_{X_{\text{Male}}}}_{\text{Mean Effect (Intercept)}} + \underbrace{\tau(x) \cdot I_{t=1}}_{\text{Treatment Effect}}$$

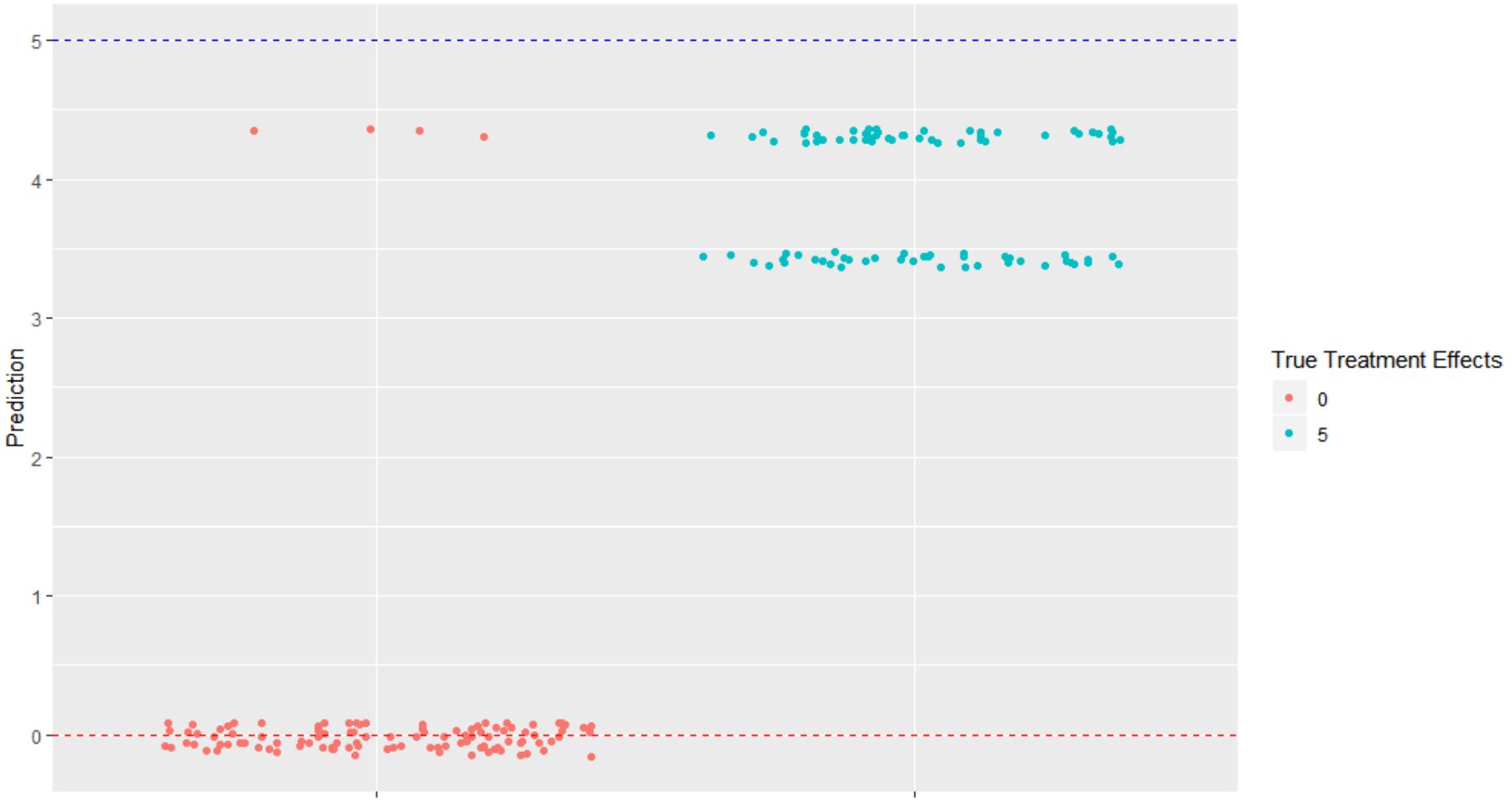
Case		Intercept	Treatment Effect
Male	Score < 0	1.7	5
Male	Score > 0	1.7	0
Female	Score < 0	0.5	5
Female	Score > 0	0.5	0

MOB Trees and IPTW - Result

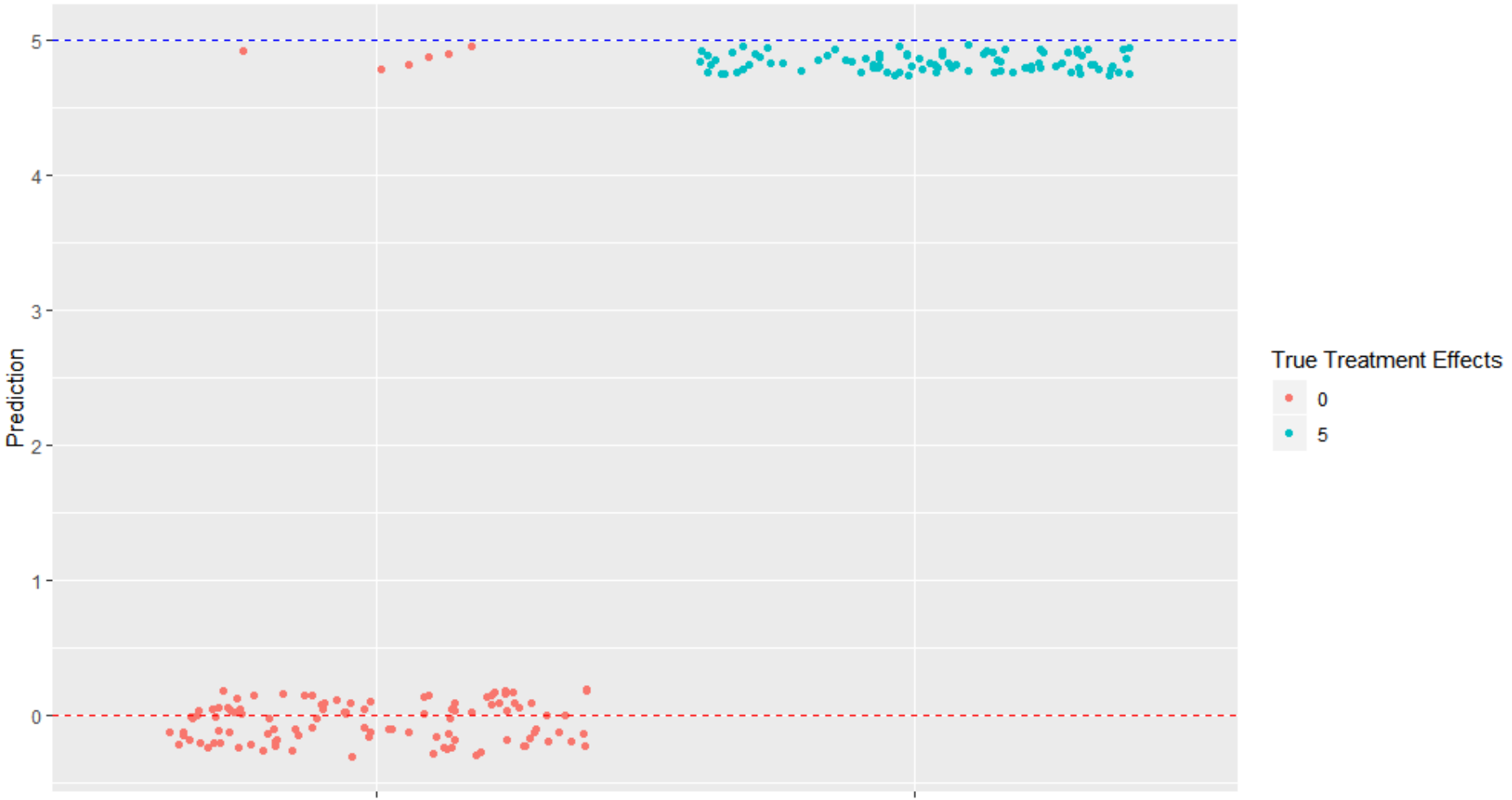


	Estimated		True	
Node	Intercept	Treatment Effect	Intercept	Treatment Effect
3	0.33	3.42	0.5	5
4	0.64	-0.10	0.5	0
6	1.67	4.32	1.7	5
7	1.64	0.03	1.7	0

MOB Trees and IPTW - Result with weights



MOB Trees and IPTW - Result without weights



Causal Trees

- Like a "rpart" (CART) object, but estimates treatment effects instead of outcomes

$\Rightarrow$ Data-driven approach to partition the data in to subpopulations

- "Honest" version: Two different samples for tree builing an treatment effect estimation $\rightarrow$ unbiased estimation
- Confounding: Include PS weights
- Tree splits are chosen by minimizing -$\widehat{EMSE}$

Causal Trees

True Treatment Effect:

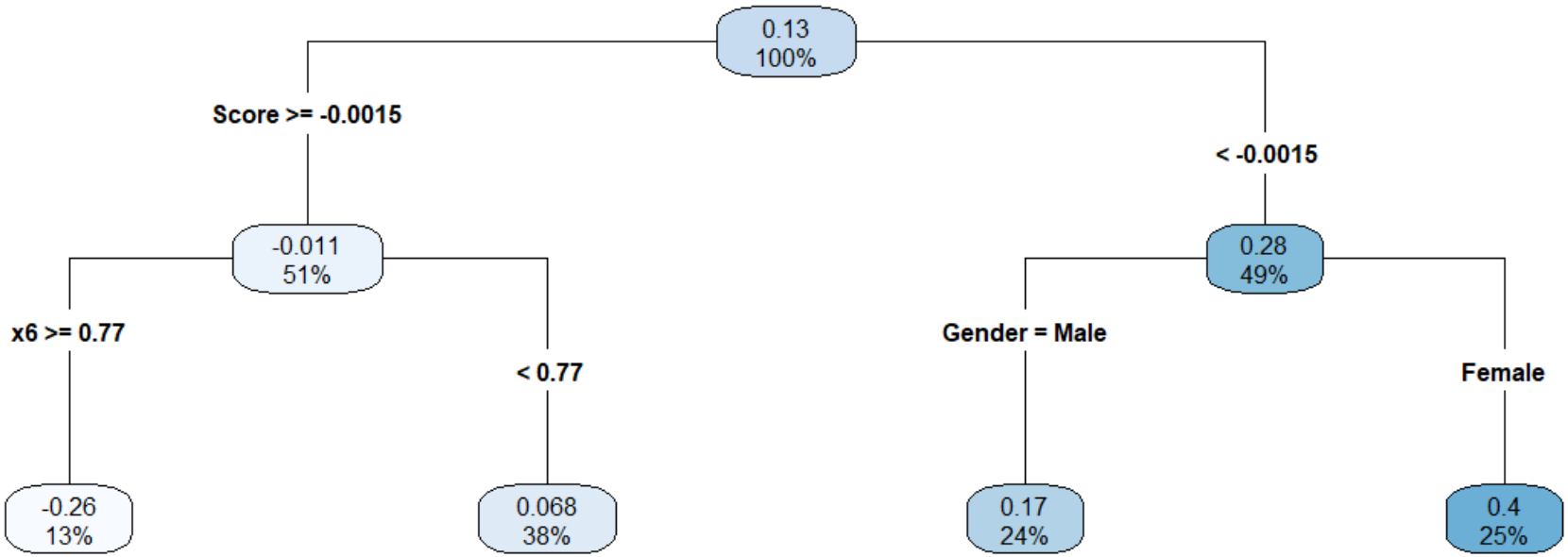
$$\tau(x) = \mathbb{P}(Y = 1 | T = 1, X) - \mathbb{P}(Y = 1 | T = 0, X)$$

Estimated Treatment Effect in each leaf:

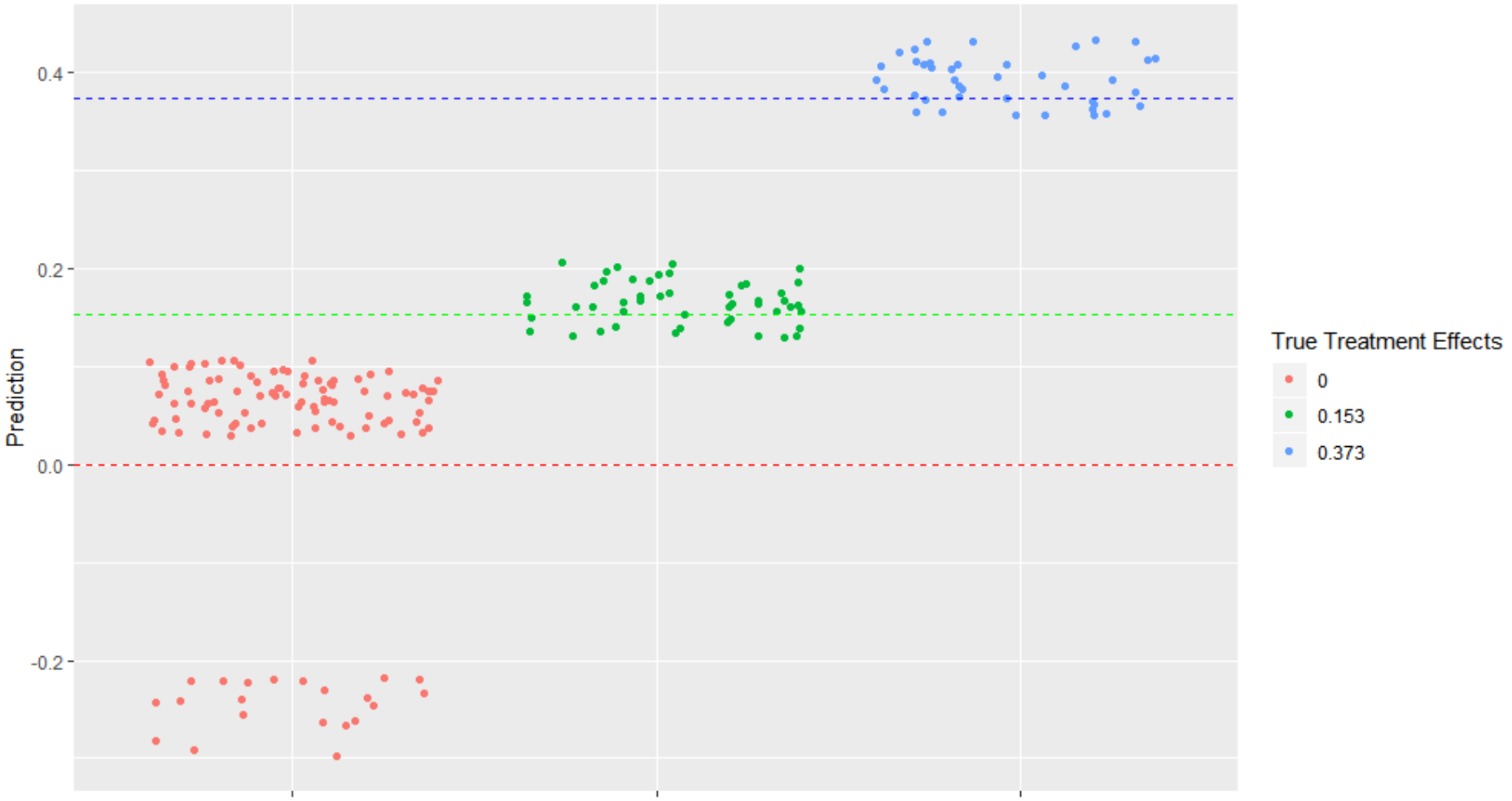
$$\hat{\tau}(x) = \hat{\mathbb{P}}(Y = 1 | T = 1, X) - \hat{\mathbb{P}}(Y = 1 | T = 0, X)$$

```
ct <- causalTree(Y ~ X, treatment = T, weights = weight)
```

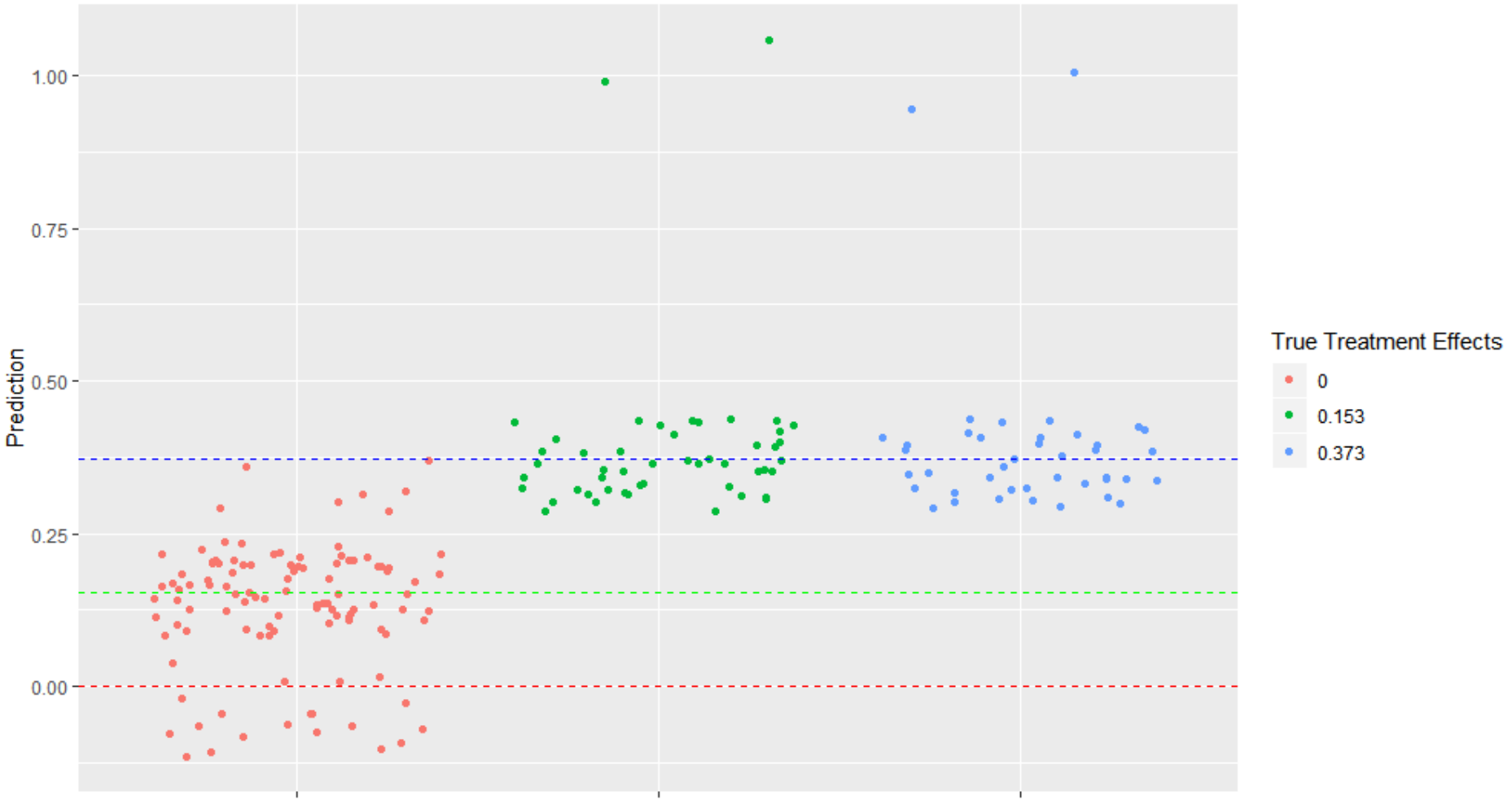
Estimated Tree



Causal Trees - Result with weights



Causal Trees - Result without weights



Bayesian Additive Regression Trees

Estimation of potential outcomes:

$$\tau(x) = f(x, 1) - f(x, 0)$$

with $f(x, t)$ = the estimated conditional mean functions

$f(x)$ = "sum of regression trees $g_l(x)$"

$$\hat{Y} = f(x) = \sum_{l=1}^{L} g_l(x)$$

with $l = 1, \ldots, L$ = Number of trees

Bayesian Additive Regression Trees

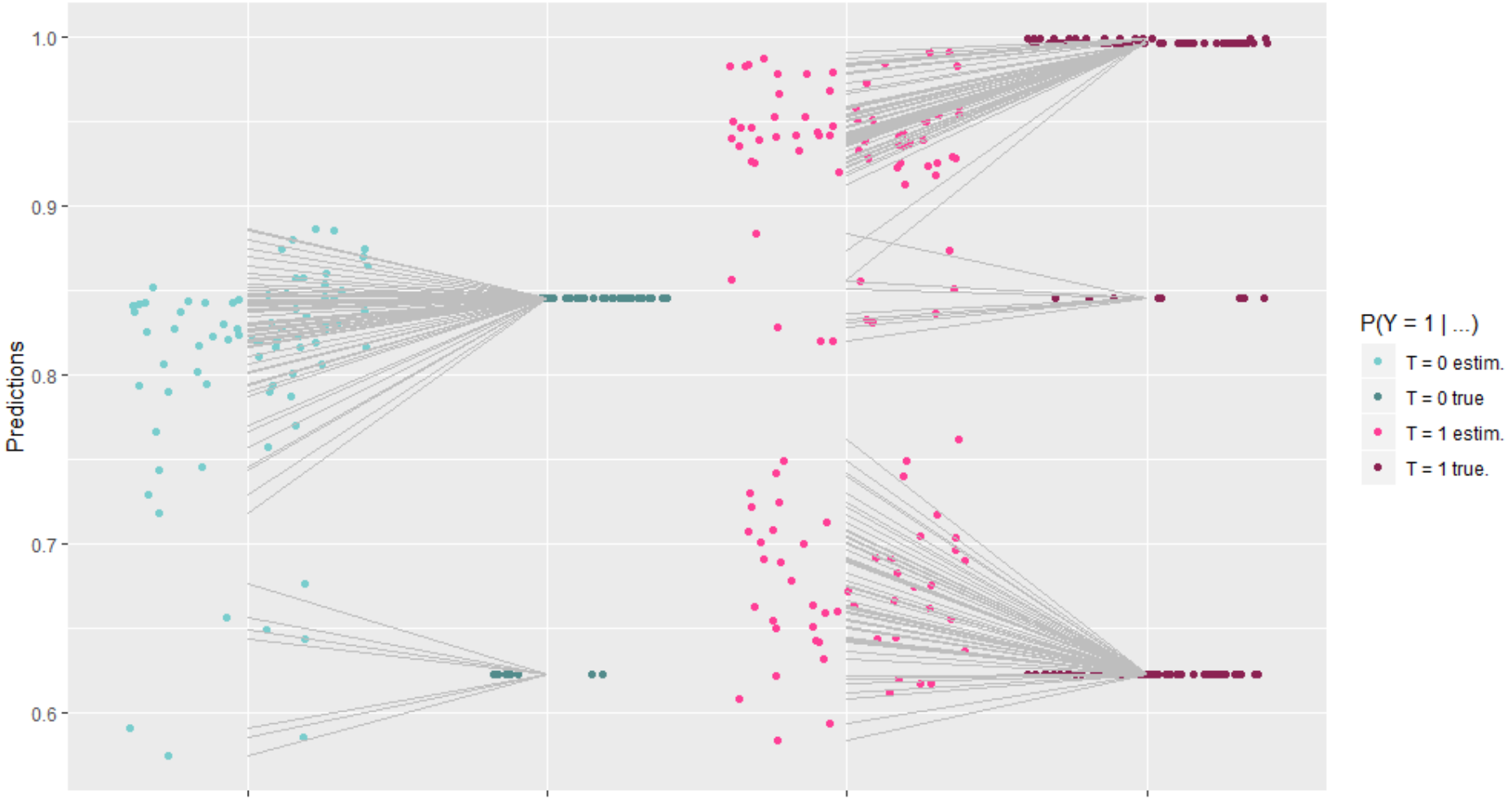
- Refits tree residuals $\rightarrow$ Bayesian regularized tree boosting procedures
- Choose priori for f: keep trees small, s.t. they are "weak learners" and avoid overfitting

$\Rightarrow$ BART consists of:

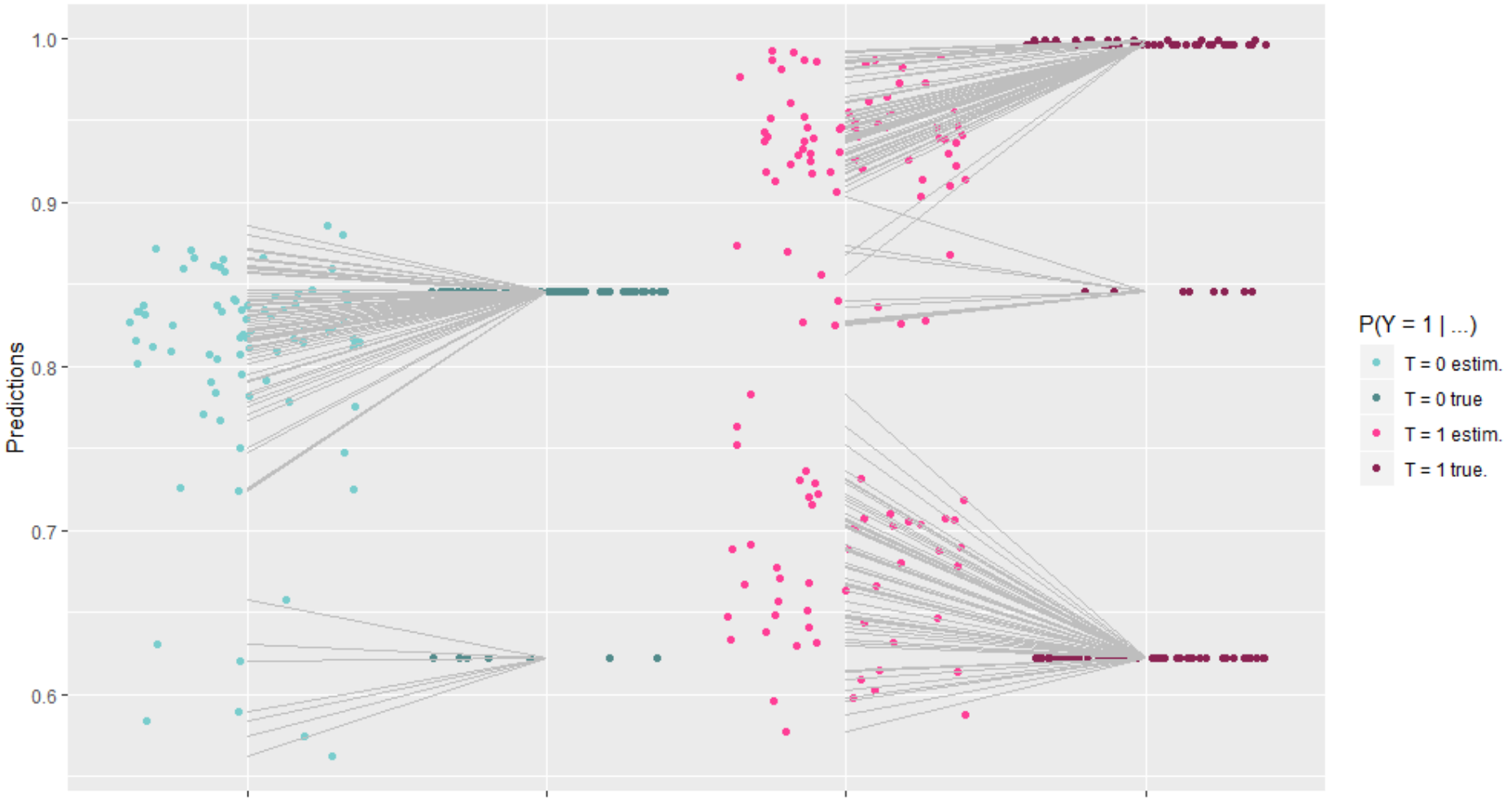
1. Sum-of-tree models
2. Regularization prior

- Binary outcome: $Y = P(Y = 1|X) = F(f(x))$, with $F(x) =$ probit link
- Confounding: Include PS in the model

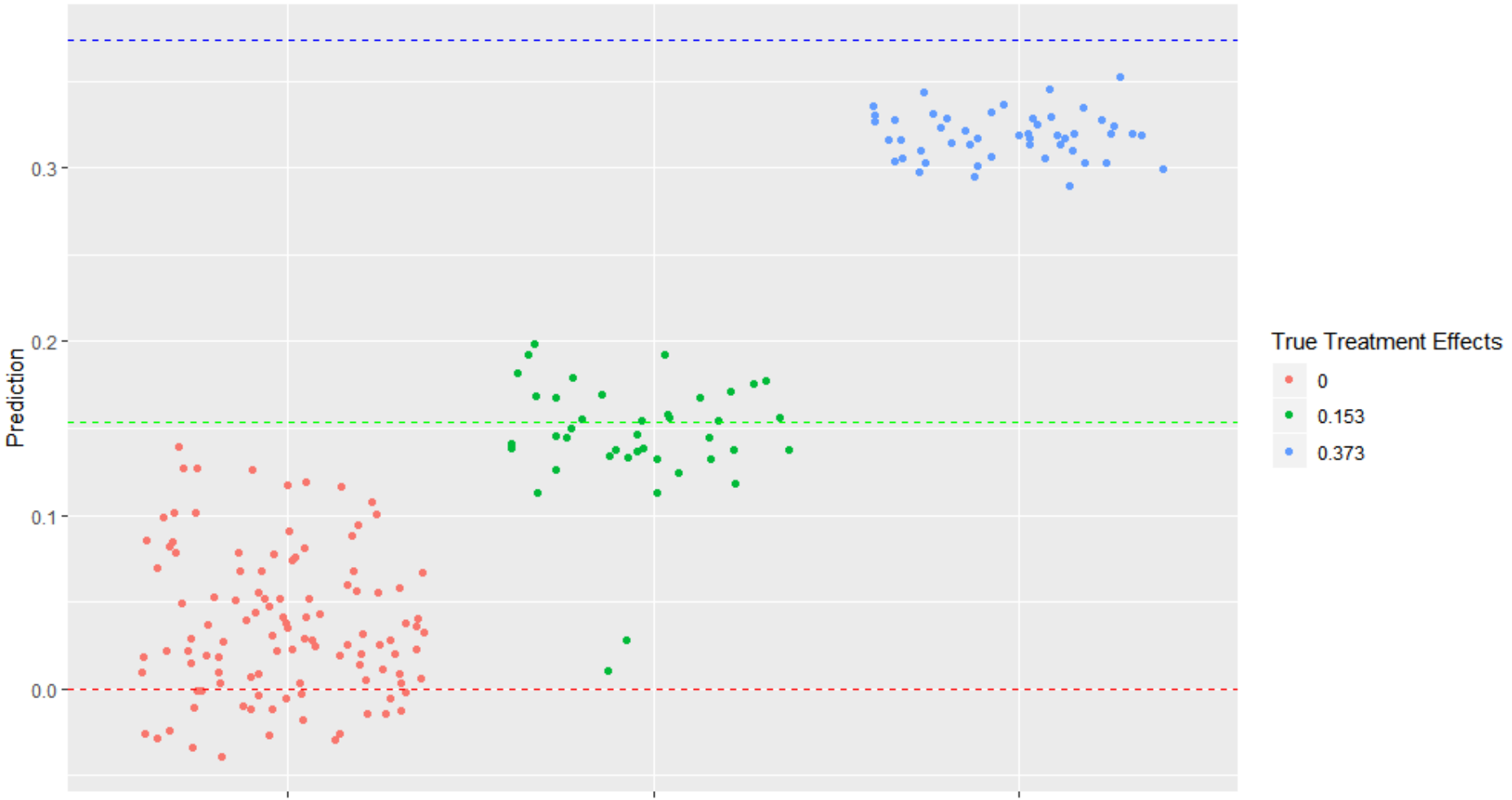
BART - Result with PS as Covariate



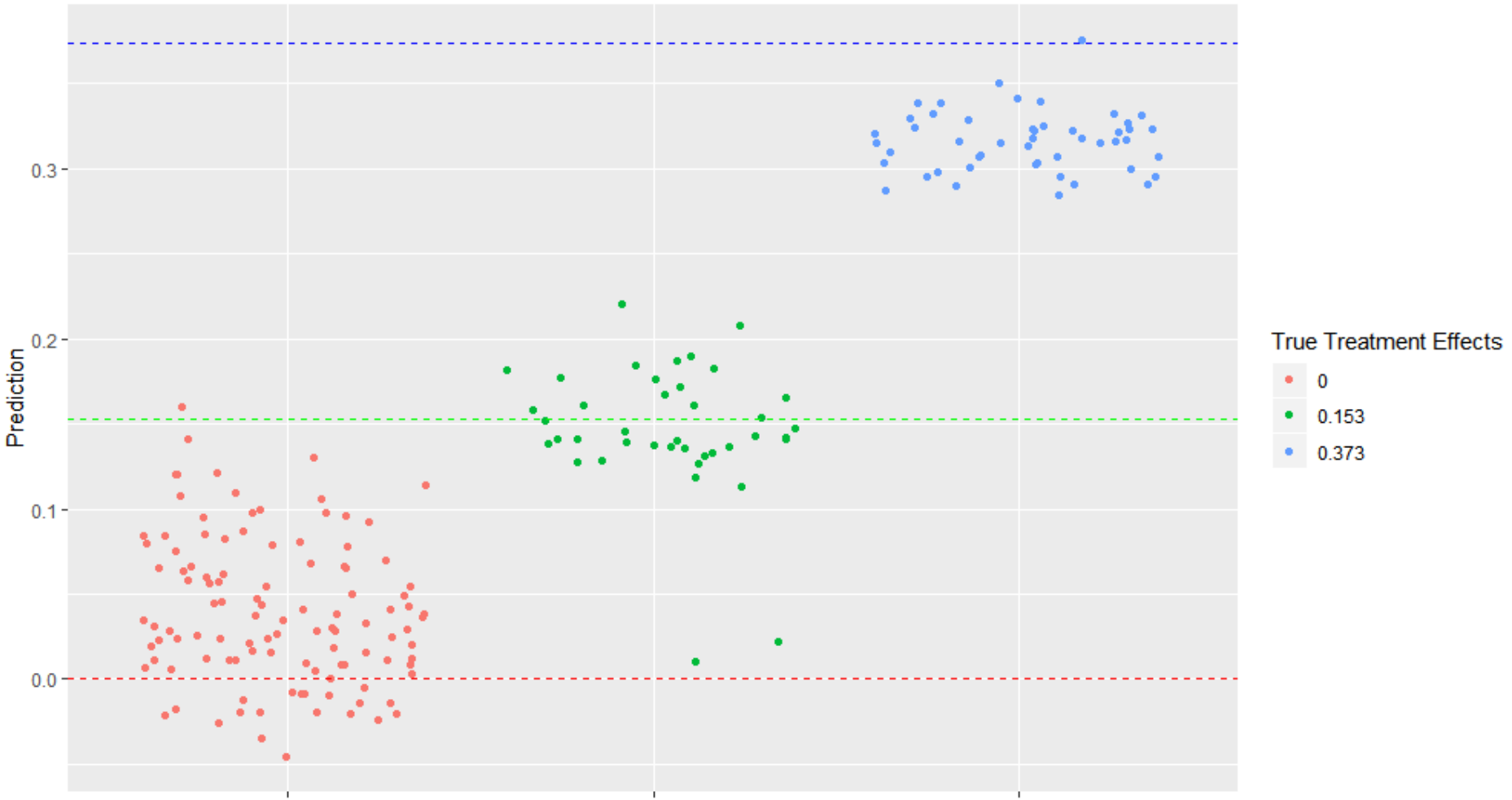
BART - Result without PS as Covariate



BART - Result with PS as Covariate



BART - Result without PS as Covariate



Discussion

- GLMTrees:
  - Little change in dataset results in big change in tree
  - Weights worsen the fit
- Causal Tree:
  - Trees get too big, pruning via complexity parameter not possible
    $\rightarrow$ Use of maxdepth?