

AIPW for project

Diane Sarkis

5/25/2023

```
library(lmtest)

## Loading required package: zoo
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

library(sandwich)
library(grf)
library(glmnet)

## Loading required package: Matrix
## Loaded glmnet 4.1-7

library(splines)
library(ggplot2)
library(reshape2)
library(RColorBrewer)

library(haven)
data <- read_dta("Replication_Dataset.dta")

Y <- data$vote_lega_euro
W <- data$diesel_euro4_ass

covariates <- c("age", "factor(eco_mode)", "vote_lega_municipal", "EDU1", "EDU2", "EDU3",
  "INC1", "INC2", "INC3", "INC4", "INC5", "INC6", "INC7", "INC8", "INC9",
  ↪ "INC10", "INC11", "INC12",
  "INC13", "INC14", "INC15", "INC16", "everyweek", "female",
  ↪ "gov_firms_responsibility",
  "green_policies_positive", "climate_neutrality", "km_1k_to_5k",
  ↪ "km_5k_to_10k",
  "km_10k_to_20k", "km_20k_to_30k", "km_less_1k", "km_more_30k",
  ↪ "pay_eco_friendly",
  "factor(recycled_materials)", "taxes_eco_friendly", "use_month",
  ↪ "use_week", "use_year",
  "water_bottle")

X <- model.matrix(formula(paste0("~", paste0(covariates, collapse="+"))), data=data)
```

```
X <- X[, -1]
```

```
# Input covariates need to be numeric.
```

```
# Estimate a causal forest.
```

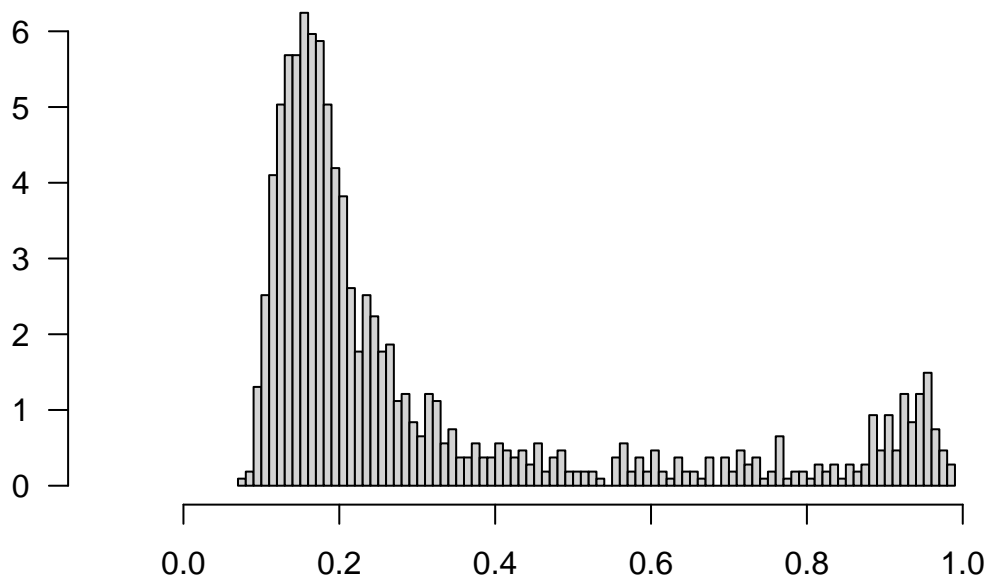
```
forest <- causal_forest(X=X, W=as.vector(W), Y=as.vector(Y), num.trees = 100)
forest.ate <- average_treatment_effect(forest)
```

```
## Warning in average_treatment_effect(forest): Estimated treatment propensities
## go as high as 0.99 which means that treatment effects for some treated units
## may not be well identified. In this case, using `target.sample=control` may be
## helpful.
```

```
print(forest.ate)
```

```
##      estimate      std.err
## 0.03977937 0.02173932
```

```
e.hat <- forest$W.hat
hist(e.hat, main="", breaks=100, freq=FALSE,
     xlab="", ylab="", xlim=c(-.1, 1.1), las=1)
```



```
PAST CODE # {r} # W_formula <- as.formula(paste("W ~", paste(covariates, collapse="+")))
# # Fit the logistic regression model # W_model <- glm(W_formula, family=binomial(link="logit"),
data=data) # # Obtain predicted probabilities of treatment # W_hat <- predict(W_model,
type="response") # # Define the outcome model # Y_formula <- as.formula(paste("Y ~ W
+", paste(covariates, collapse="+"))) # # Fit the linear regression model # Y_model <-
glm(Y_formula, data=data) # # Obtain predicted outcomes # Y_hat <- predict(Y_model) # #
Compute weights for AIPW # weights <- ifelse(W==1, 1/W_hat, 1/(1-W_hat)) # # # Compute
AIPW estimate of ATE # ATE_AIPW <- mean((W - W_hat)*(Y - Y_hat)*weights + W_hat*Y_hat -
(1-W_hat)*(1-Y_hat)*weights) # {r} # ATE_AIPW ““ Fit two models: a logistic regression to estimate
the probability of treatment, and a linear regression to estimate the potential outcome under treatment and
control. We then compute the AIPW estimate of the ATE.
```