Synthetic controls with staggered adoption

Yunlin Zhou

2023-04-21

library(augsynth)  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.1 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

# Import the dataset

merge\_df <- read.csv("fc\_rucc\_pm25\_merge.csv")

# Create a new dataset for multisynth()

analysis\_df <- merge\_df %>%  
 mutate(year\_opened = ifelse(is.na(year\_opened),   
 Inf, year\_opened),  
 cbr = 1 \* (year >= year\_opened))%>%  
 select(cbr, id, year, PM2.5, RUCC\_2013) %>%  
 distinct(cbr, id, year, PM2.5, RUCC\_2013, .keep\_all = TRUE) %>%  
 filter(complete.cases(.))  
  
analysis\_df <- distinct(analysis\_df, id, year, .keep\_all = TRUE)  
  
analysis\_df$id <- as.factor(analysis\_df$id)

# Using Synthetic controls with staggered adoption (without RUCC)

ppool\_syn <- multisynth(form = PM2.5 ~ cbr,  
 unit = id,  
 time = year,  
 data = analysis\_df,  
 fixedeff = TRUE,  
 time\_cohort = TRUE,  
 n\_leads = 3)

print(ppool\_syn$nu)

## [1] 0.9199388

ppool\_syn\_summ <- summary(ppool\_syn)  
ppool\_syn\_summ

##   
## Call:  
## multisynth(form = PM2.5 ~ cbr, unit = id, time = year, data = analysis\_df,   
## n\_leads = 3, fixedeff = TRUE, time\_cohort = TRUE)  
##   
## Average ATT Estimate (Std. Error): 0.114 (0.144)  
##   
## Global L2 Imbalance: 0.000  
## Scaled Global L2 Imbalance: 0.000  
## Percent improvement from uniform global weights: 100  
##   
## Individual L2 Imbalance: 0.027  
## Scaled Individual L2 Imbalance: 0.002  
## Percent improvement from uniform individual weights: 99.8   
##   
## Time Since Treatment Level Estimate Std.Error lower\_bound upper\_bound  
## 0 Average 0.05919265 0.1320794 -0.2203274 0.3081469  
## 1 Average 0.11974715 0.1591569 -0.1855765 0.4264856  
## 2 Average 0.11632673 0.1874523 -0.2407812 0.4940995

plot(ppool\_syn\_summ)

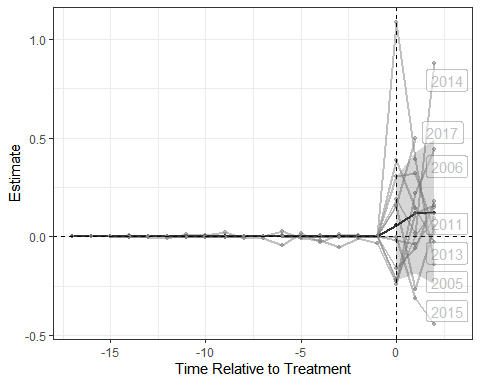
## Joining with `by = join\_by(Level)`

## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none" instead as  
## of ggplot2 3.3.4.  
## ℹ The deprecated feature was likely used in the augsynth package.  
## Please report the issue to the authors.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

## Warning: Removed 84 rows containing missing values (`geom\_line()`).

## Warning: Removed 84 rows containing missing values (`geom\_point()`).

## Warning: ggrepel: 6 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps

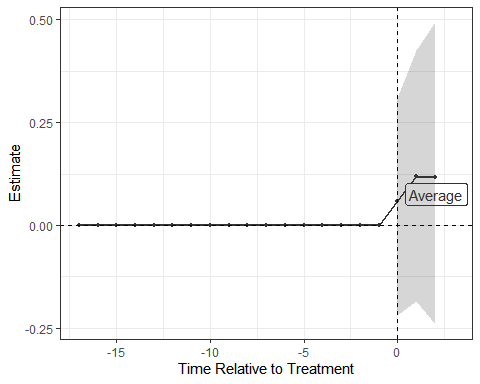


plot(ppool\_syn\_summ, levels = "Average")

## Joining with `by = join\_by(Level)`

## Warning: Removed 1 row containing missing values (`geom\_line()`).

## Warning: Removed 1 rows containing missing values (`geom\_point()`).



# Using Synthetic controls with staggered adoption (with RUCC)

ppool\_syn\_rucc <- multisynth(form = PM2.5 ~ cbr| RUCC\_2013,  
 unit = id,  
 time = year,  
 data = analysis\_df,  
 fixedeff = TRUE,  
 time\_cohort = TRUE,  
 n\_leads = 3)

print(ppool\_syn\_rucc$nu)

## [1] 0.681342

ppool\_syn\_rucc\_summ <- summary(ppool\_syn\_rucc)  
ppool\_syn\_rucc\_summ

##   
## Call:  
## multisynth(form = PM2.5 ~ cbr | RUCC\_2013, unit = id, time = year,   
## data = analysis\_df, n\_leads = 3, fixedeff = TRUE, time\_cohort = TRUE)  
##   
## Average ATT Estimate (Std. Error): 0.100 (0.150)  
##   
## Global L2 Imbalance: 0.001  
## Scaled Global L2 Imbalance: 0.000  
## Percent improvement from uniform global weights: 100  
##   
## Individual L2 Imbalance: 0.083  
## Scaled Individual L2 Imbalance: 0.007  
## Percent improvement from uniform individual weights: 99.3   
##   
## Time Since Treatment Level Estimate Std.Error lower\_bound upper\_bound  
## 0 Average 0.06089087 0.1445502 -0.2224020 0.3474358  
## 1 Average 0.10153993 0.1636978 -0.2273861 0.4111674  
## 2 Average 0.06556047 0.1818975 -0.2811778 0.4245036

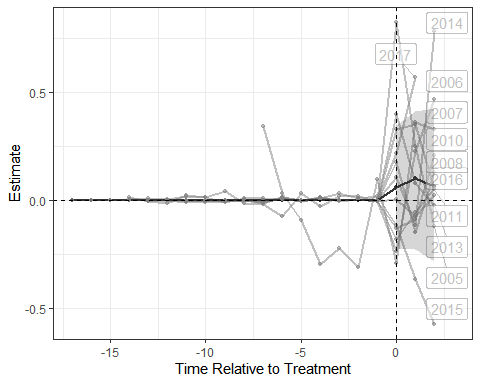
plot(ppool\_syn\_rucc\_summ)

## Joining with `by = join\_by(Level)`

## Warning: Removed 84 rows containing missing values (`geom\_line()`).

## Warning: Removed 84 rows containing missing values (`geom\_point()`).

## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps

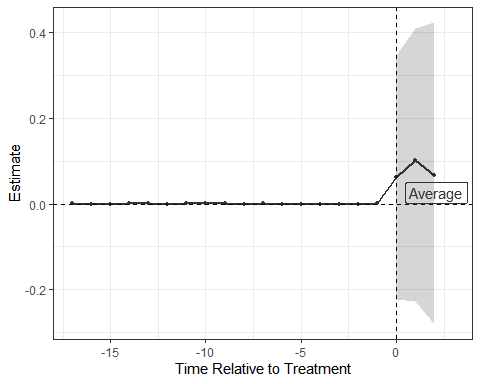


plot(ppool\_syn\_rucc\_summ, levels = "Average")

## Joining with `by = join\_by(Level)`

## Warning: Removed 1 row containing missing values (`geom\_line()`).

## Warning: Removed 1 rows containing missing values (`geom\_point()`).



# Using Synthetic controls with staggered adoption (with RUCC, nu=.1, lambda=.000135)

ppool\_syn\_rucc\_nu <- multisynth(form = PM2.5 ~ cbr| RUCC\_2013,  
 unit = id,  
 time = year,  
 data = analysis\_df,  
 fixedeff = TRUE,  
 time\_cohort = TRUE,  
 nu=.1,  
 lambda=.000135,  
 n\_leads = 3)

ppool\_syn\_rucc\_nu\_summ <- summary(ppool\_syn\_rucc\_nu)  
ppool\_syn\_rucc\_nu\_summ

##   
## Call:  
## multisynth(form = PM2.5 ~ cbr | RUCC\_2013, unit = id, time = year,   
## data = analysis\_df, n\_leads = 3, nu = 0.1, lambda = 0.000135,   
## fixedeff = TRUE, time\_cohort = TRUE)  
##   
## Average ATT Estimate (Std. Error): 0.098 (0.156)  
##   
## Global L2 Imbalance: 0.006  
## Scaled Global L2 Imbalance: 0.001  
## Percent improvement from uniform global weights: 99.9  
##   
## Individual L2 Imbalance: 0.048  
## Scaled Individual L2 Imbalance: 0.004  
## Percent improvement from uniform individual weights: 99.6   
##   
## Time Since Treatment Level Estimate Std.Error lower\_bound upper\_bound  
## 0 Average 0.06368450 0.1531511 -0.2354696 0.3809860  
## 1 Average 0.09622663 0.1690647 -0.2341437 0.4165696  
## 2 Average 0.06545656 0.1824701 -0.3077763 0.4465734

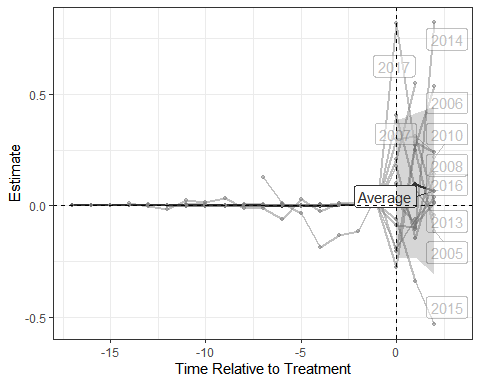
plot(ppool\_syn\_rucc\_nu\_summ)

## Joining with `by = join\_by(Level)`

## Warning: Removed 84 rows containing missing values (`geom\_line()`).

## Warning: Removed 84 rows containing missing values (`geom\_point()`).

## Warning: ggrepel: 2 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps



plot(ppool\_syn\_rucc\_nu\_summ, levels = "Average")

## Joining with `by = join\_by(Level)`

## Warning: Removed 1 row containing missing values (`geom\_line()`).

## Warning: Removed 1 rows containing missing values (`geom\_point()`).

