ARE 213 Problem Set 2b

Nick Depsky, Will Gorman, Peter Worley

November 9, 2018

Import Data

```
#setwd("~/Dropbox/Berkeley_tings/Fall 2018/ARE213/Problem Sets/SharedFiles/are213/PS1b")
#setwd("C:\\Users\\will-\\Desktop\\are213\\PS2b")
setwd("C:\\Users\\Will\\Desktop\\are213\\PS2b")

dat <- read.dta("traffic_safety2.dta")
```

1a - Aggregate treatment analysis

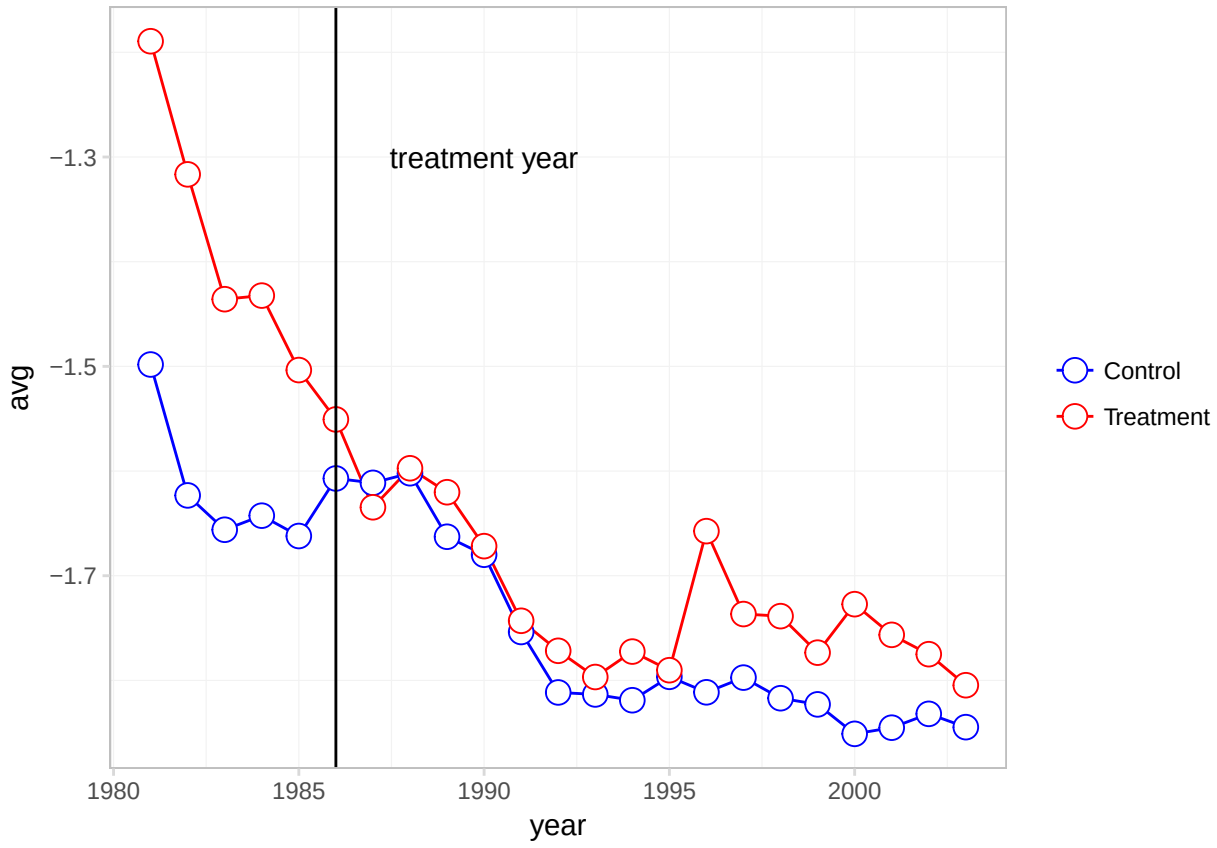
In the first tibble printed below, we see that the average pre-period dependent variable for the TU site is -1.38 while that for the control states is -1.71 suggesting that the treatment states have a higher per capita fatality rate than the control states.

```
dat$fatal_pc <- log(dat$fatalities/dat$population)
dat$control <- ifelse(dat$state %in% c(6,10,30,41),1,ifelse(dat$state == 99,2,0))

dat %>% group_by(control,primary) %>% summarize(avg = mean(fatal_pc))
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.4
```

```
## # A tibble: 6 x 3
## # Groups:   control [?]
##   control primary   avg
##     <dbl>   <int> <dbl>
## 1       0       0 -1.71
## 2       0       1 -1.92
## 3       1       0 -1.49
## 4       1       1 -1.76
## 5       2       0 -1.38
## 6       2       1 -1.72
```

```
## In the above table,  the pre-treatment average are when primary = 0
## The treatment aggregate is equal to 2 and
## the control average is equal to 0

summary_yr <- dat %>% group_by(control,year) %>% summarize(avg = mean(fatal_pc))

ggplot(summary_yr[summary_yr$control != 1,],
       aes(x=year, y=avg, group = factor(control), color = factor(control))) +
  geom_line() +geom_point( size=4, shape=21, fill="white") + theme_plot +
  scale_color_manual(labels = c("Control", "Treatment"), values = c("blue","red")) +
  geom_vline(xintercept = 1986) + annotate("text", x=1990, y=-1.3, label= "treatment year")
```



Looking at the TU pre-period (red line before 1986), we see that there was a significantly different pre-period trend that we would worry about when running our econometric analysis. The graph also shows that the treatment states had a noticeably higher pre-treatment per capita fatality rate than the control states.

```
rm_treat <- dat[dat$control != 1,]

result <- rm_treat %>%
             group_by(control, primary, state) %>%
             filter(year == max(year))

result <- result[result$primary == 0,c(1,2,15,16)]
result <- result[result$year != 2003,]
kable(round(result[order(result$fatal_pc),],3))
```

state	year	fatal_pc	control
29	1999	-2.444	0
45	2001	-2.223	0
32	1983	-2.145	0
18	1997	-2.138	0
4	1992	-2.001	0
20	1999	-1.969	0
7	2002	-1.872	0
13	1997	-1.851	0
16	1995	-1.601	0

state	year	fatal_pc	control
35	1990	-1.597	0
9	1995	-1.594	0
99	1985	-1.504	2
34	1997	-1.393	0
1	1999	-1.359	0
25	1986	-1.345	0

```
rm_treat <- dat[dat$control != 1,]
```

Georgia (which is state number 9) appears to be the closest state to our treatment state (state number 99)

```
rm_comp <- dat[dat$state %in% c(9,99),]

store <- rm_comp %>% group_by(primary, secondary,state) %>%
             summarise_all(funs(mean))

kable(round(store,3))
```

primary	secondary	state	year	college	beer	population	unemploy	fatalities	totalvmt	precip	snow32
0	0	9	1984.5	0.197	1.171	5919.153	6.425	1436.750	53249.75	4.155	0.000
0	0	99	1983.0	0.224	1.620	11466.998	6.753	2903.171	95298.49	2.444	0.178
0	1	9	1992.0	0.215	1.197	6836.821	5.557	1458.143	76547.43	4.450	0.000
1	0	9	1999.5	0.233	1.206	8133.751	4.350	1563.500	101109.12	4.183	0.000
1	0	99	1994.5	0.236	1.476	14189.515	5.900	2540.081	137211.03	2.518	0.134

The independent variables do not appear very similar. The population, for instance, is much larger in the treated states, resulting in much larger vmt trips. Furthermore, it appears the weather is different with less rain in the treated states and marginally more snow.

1b - Synthetic control method

The synthetic control method compares aggregated outcomes in a number of different control observations to a measured treatment effect. A synthetic control is formed from a group of potential control units, and is designed by evaluating covariates in a pre-treatment periods, developing weights for each potential control based on its distance from the treated group. Once the weights are formulated using pre-treatment data, The the post-treatment outcome are computed for the aggregated control unit to synthesize a counterfactual to the treated unit. Finally, the treatment effect is calculated by comparing the treated unit to the synthetic control unit in the post period

Benefits of the method: - Transparency - since a synthetic control is a "weighted average of the available control units" the contributions and similarities of each control unit is made transparent.

- In a quasi-experiment, synthetic controls allows for the empiricist to limit bias in selection of controls by chosing them based on how closely they resemble treated units rather than based on arbitrary characteristics. (i.e. its more rigorous and less ad-hoc)
- Extrapolation can be avoided, due to the fact the sum of the weights can be restricted to one.
- The large number of available control units allows for estimation of appropriate standard errors due to "placebo" estimates.

Drawbacks of the method:

- The method is effective in using and simulating aggragate data, so is not applicable to all empirical situations – especially with a smaller pool of potential controls.
- The method assumes outcomes post-treatment in the controls are not impacted by treatment in the treatment unit,or there is no spillover effects. This may not be a reasonable assumption, but Abadie talks through how to test the treatment effect assuming one of the control units had actually been treated, and contends, in his 2010 paper, that this proves the treatment effect is signifcant.

```
data <- dat[dat$control != 1,]
data$state <- as.numeric(data$state)

dataprep.out.mean<-
dataprep(
foo = data,
predictors = c("college", "beer","unemploy","totalvmt","precip","snow32"),
predictors.op = "mean",
dependent = "fatal_pc",
unit.variable = "state",
time.variable = "year",
treatment.identifier = 99,
controls.identifier = c(1:5,7:9,11:29,31:40,42:48),
time.predictors.prior = c(1981:1985),
time.optimize.ssr = c(1981:1985),
time.plot = 1981:2003
)

synth.out.mean <- synth(dataprep.out.mean)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 0.001287399
##
## solution.v:
##  0.02467012 0.08644213 0.5643016 0.001777663 0.06078132 0.2620272
##
## solution.w:
##  5.58e-08 6.67e-08 3.7024e-06 0.1651352 2.052e-07 1.258e-07 0.2400315 1.145e-07 7.93e-08 8.4e-08 6.73
```

This command is creating weights for each of the control observations we provide it based on using the mean of the observed covariates during the pretreatment period as well as a selection of outcome variables for both the treated and control units. The weights are selected such that they minimize the distance between the treated and control units.

Our preferred specification followed the above logic. We rely on the mean of the observed covariates, as this is the default. We also did not see any reason why other formulations would make sense given we have a large enough number of states. We also selected the time to optimize over to match the pre-treatment

period. Since the treatment would likely affect affect the outcome, we did not want to construct a synthetic control around this affected data. Finally, we had to select the predictors of interest. We figured most of our variables would likely have an effect on fatalities, so we included them all in our specifications.

1c - Graphical interpretation and treatment

```
dataprep.out.med<-
dataprep(
foo = data,
predictors = c("college", "beer","unemploy","totalvmt","precip","snow32"),
predictors.op = "median",
dependent = "fatal_pc",
unit.variable = "state",
time.variable = "year",
treatment.identifier = 99,
controls.identifier = c(1:5,7:9,11:29,31:40,42:48),
time.predictors.prior = c(1981:1985),
time.optimize.ssr = c(1981:1985),
time.plot = 1981:2003
)

synth.out.med <- synth(dataprep.out.med)
```

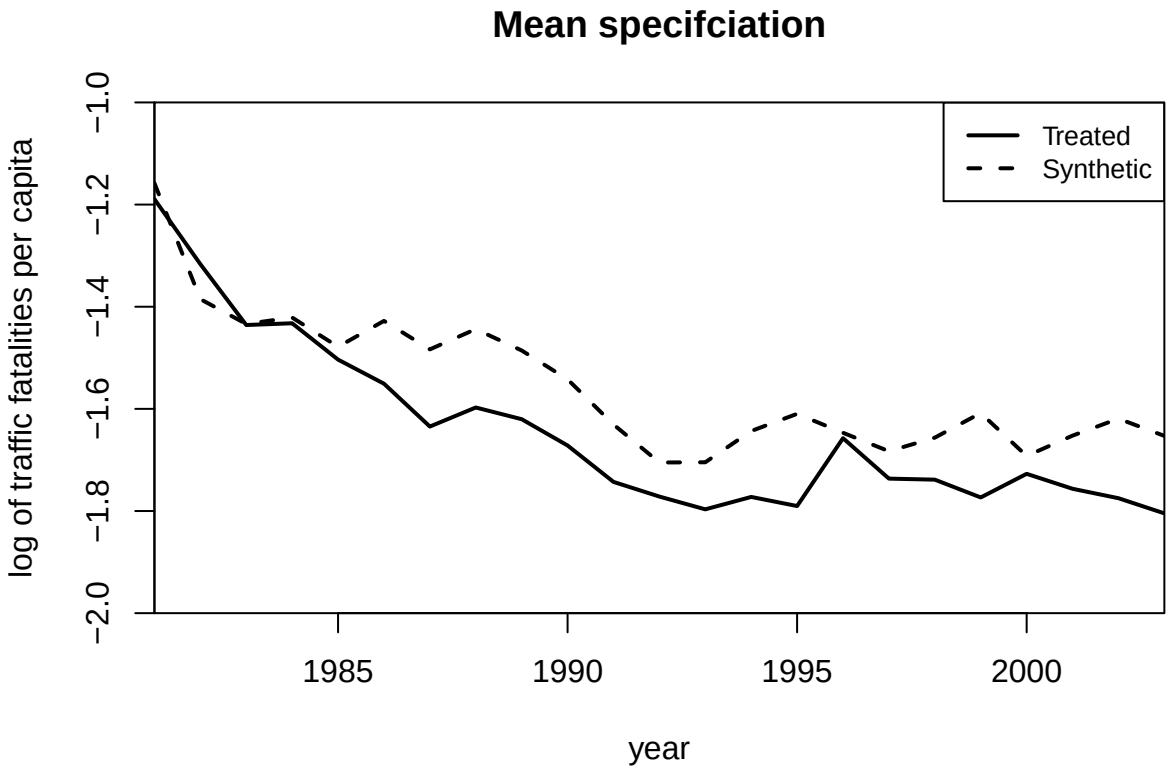
```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 0.001240674
##
## solution.v:
##  0.03139319 0.3250322 0.5003968 0.003401803 0.01302912 0.1267469
##
## solution.w:
##  2.53771e-05 3.1301e-05 1.895e-06 0.216871 0.003098196 5.83064e-05 0.2482756 4.76307e-05 3.55017e-05
```

```
dataprep.out.full<-
dataprep(
foo = data,
predictors = c("college", "beer","unemploy","totalvmt","precip","snow32"),
predictors.op = "mean",
dependent = "fatal_pc",
unit.variable = "state",
time.variable = "year",
treatment.identifier = 99,
controls.identifier = c(1:5,7:9,11:29,31:40,42:48),
```

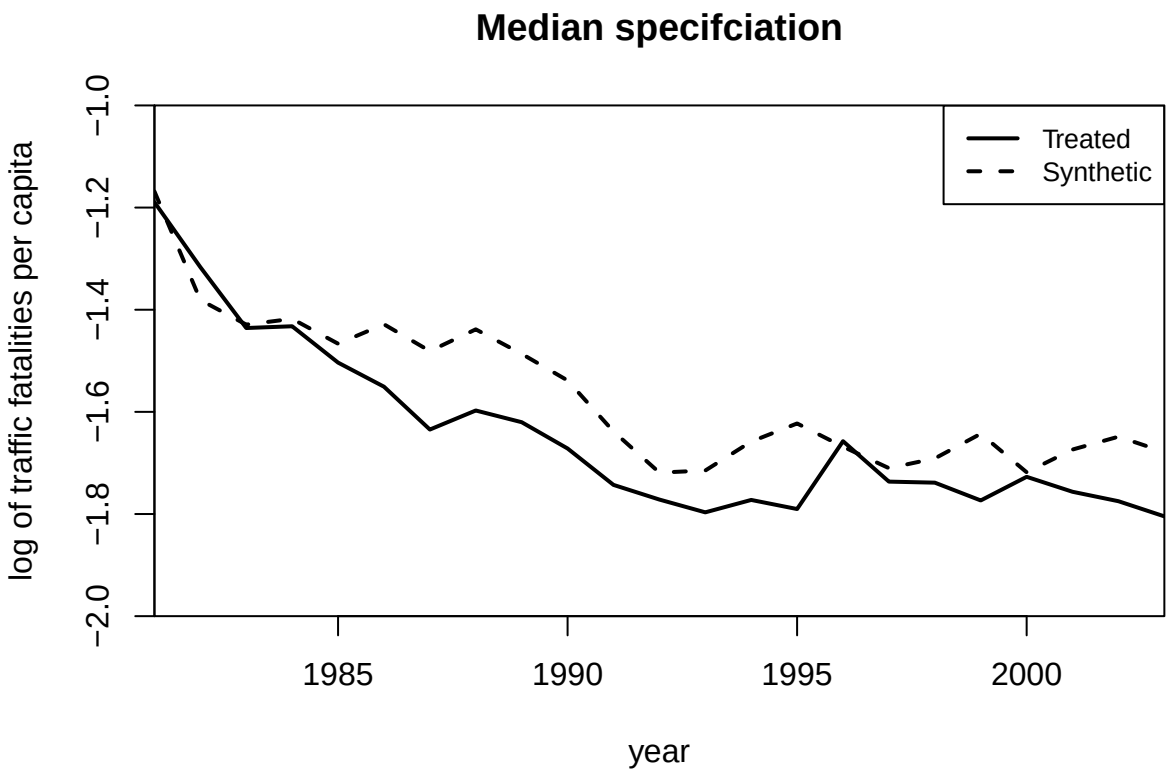
```
time.predictors.prior = c(1981:1985),
time.optimize.ssr = c(1981:2003),
time.plot = 1981:2003
)

synth.out.full <- synth(dataprep.out.full)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 0.00430354
##
## solution.v:
##  0.2528337 0.2482077 0.2508223 0.005938748 0.06230082 0.1798967
##
## solution.w:
##  3.42557e-05 4.44871e-05 3.3672e-06 0.2331292 0.001250623 6.30296e-05 0.1937911 4.96284e-05 4.22918e
```

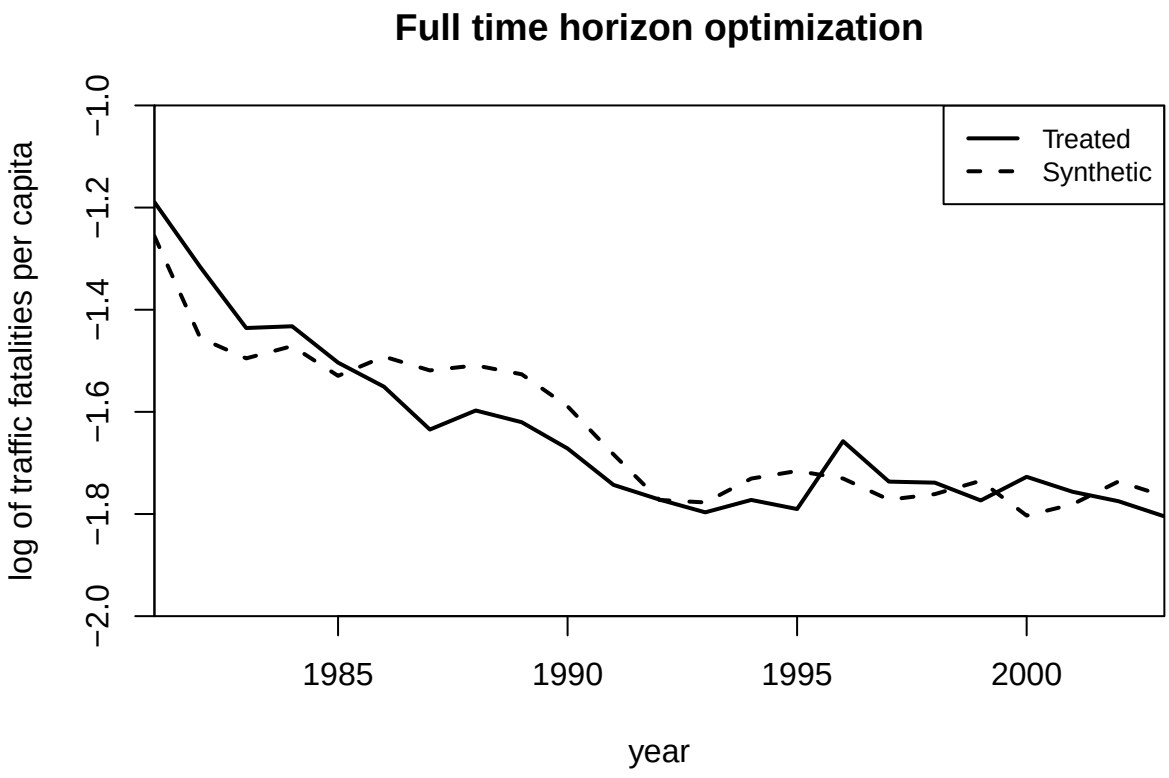
```
## preferred specification
path.plot(dataprep.res = dataprep.out.mean,synth.res = synth.out.mean,Ylab = c("log of traffic fatalitie
Xlab = c("year"),
Ylim = c(-2,-1),
Main = "Mean specifciation")
```




```
## specification #2
path.plot(dataprep.res = dataprep.out.med,synth.res = synth.out.med,Ylab = c("log of traffic fatalities
Xlab = c("year"),
Ylim = c(-2,-1),
Main = "Median specifciation")
```



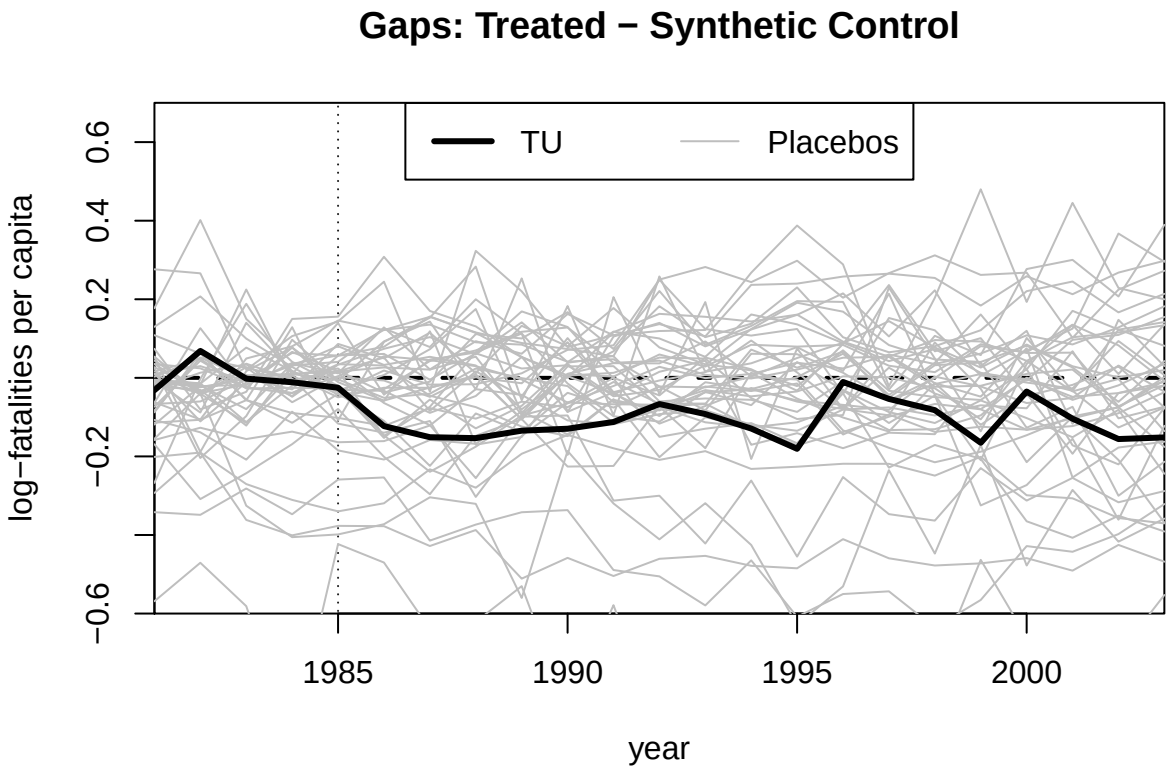

```
## specification #3
path.plot(dataprep.res = dataprep.out.full,synth.res = synth.out.full,Ylab = c("log of traffic fatalitie
Xlab = c("year"),
Ylim = c(-2,-1),
Main = "Full time horizon optimization")
```




```
controls <- c(2:5,7:9,11:29,31:40,42:48)
gaps <- as.data.frame(matrix(NA,nrow = length(1981:2003),ncol = length(controls))) %>% set_colnames(con
for(c in controls){
  newcontrols <- c(controls[-which(controls==c)],99)
  synthdat <- dataprep(foo = data,
                       predictors = c("college", "beer","unemploy","totalvmt","precip","snow32"),
                       predictors.op = "mean",
                       dependent = "fatal_pc",
                       unit.variable = "state",
                       time.variable = "year",
                       treatment.identifier = c,
                       controls.identifier = newcontrols,
                       time.predictors.prior = c(1981:1985),
                       time.optimize.ssr = c(1981:1985),
                       time.plot = 1981:2003)
  synth.out <- invisible(synth(synthdat))
  gaps[1:length(1981:2003),which(controls==c)] <- synthdat$Y1plot - (synthdat$Y0plot %*% synth.out$solu
}
```

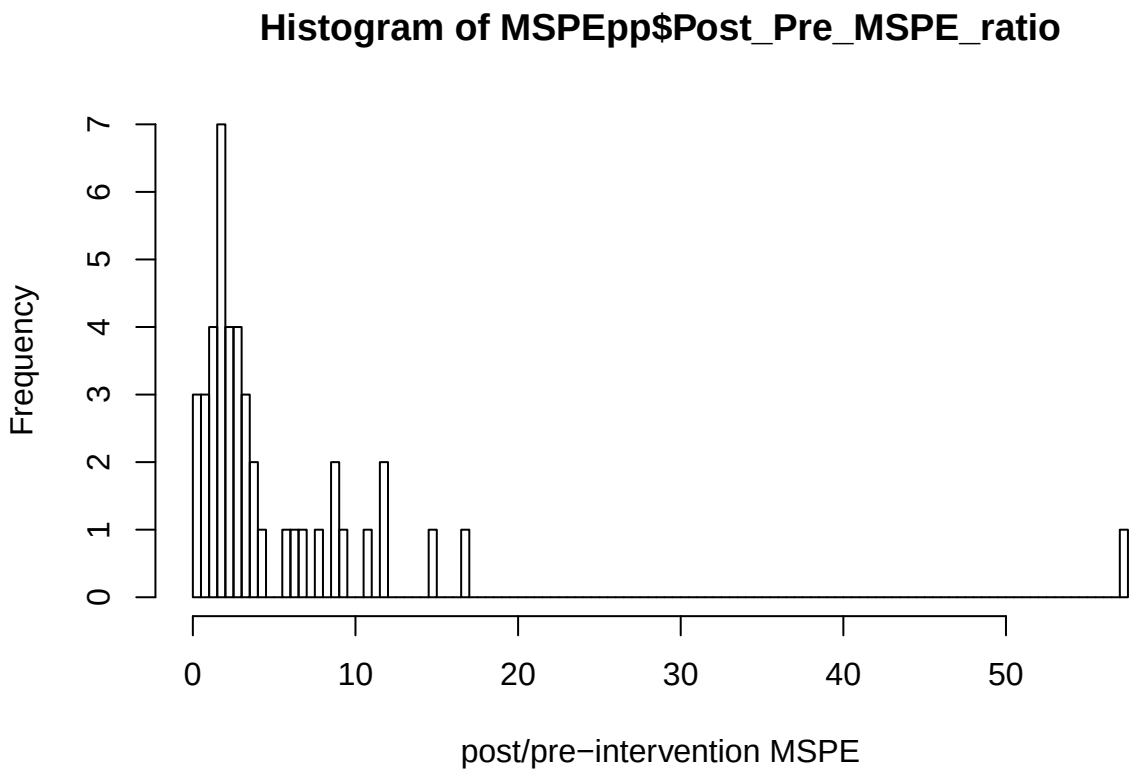
```
gaps.plot(synth.res = synth.out.mean, dataprep.res = dataprep.out.mean, Xlab = "year", Ylab = "log-fata
lines(c(1985,1985),c(-5,5), lty = 'dotted', col = 'grey10')

for(c in 1:ncol(gaps)){
  lines(1981:2003,gaps[,c], col = 'grey')
}
TUgap <- dataprep.out.mean$Y1plot - (dataprep.out.mean$Y0plot %*% synth.out.mean$solution.w)
```

```
lines(1981:2003,TUgap, lwd = 3)
legend('top', legend=c("TU","Placebos"), horiz = T,col = c(1,"grey"), lty = c(1,1), lwd = c(3,1))
```



I do not conclude that the treatment was significant as it does not appear that it is significantly different from other placebo tests.

```
MSPEpp <- data.frame(Post_Pre_MSPE_ratio = rep(NA,length(controls)+1)) %>%
  set_rownames(c(controls,99))
for(i in 1:(nrow(MSPEpp)-1)){
    MSPEpp[i,1] <- mean(gaps[6:23,i]^2, na.rm = T)/mean(gaps[1:5,i]^2, na.rm = T)
}

MSPEpp[nrow(MSPEpp),1] <- mean(TUgap[6:23]^2, na.rm = T)/mean(TUgap[1:5]^2, na.rm = T)
hist(MSPEpp$Post_Pre_MSPE_ratio, breaks = 100, xlab = "post/pre-intervention MSPE")
```




```
kable(round(MSPEpp,2))
```

	Post_Pre_MSPE_ratio
2	7.76
3	2.24
4	4.03
5	3.25
7	2.33
8	11.51
9	1.81
11	2.73
12	0.25
13	5.78
14	1.27
15	6.82
16	1.57
17	2.19
18	8.71
19	2.65
20	2.86
21	0.49
22	57.40
23	3.05
24	2.08
25	1.57
26	1.72

	Post_Pre_MSPE_ratio
27	1.53
28	3.76
29	1.88
31	10.67
32	0.89
33	1.20
34	0.28
35	8.60
36	6.38
37	1.34
38	0.77
39	3.15
40	16.74
42	9.46
43	1.20
44	3.61
45	14.82
46	0.63
47	2.91
48	1.81
99	11.55

1d - Comparison to fixed effects

```
conv <- exp(dataprep.out.mean$Y1plot)
mean(conv)
```

```
## [1] 0.1961412
```

```
conv2 <- exp(dataprep.out.mean$Y0plot %*% synth.out.mean$solution.w)
mean(conv2)
```

```
## [1] 0.2131594
```

In the fixed effects analysis, we found an average treatment effect of -9% as a result of the primary law. On the other hand, the approach relying on synthetic controls found an everage treatment effect of -8%. We calculate this by first taking the average gap in the post period and find that the law reduced fatalities by 0.017 per capita. Dividing that by the synthetic control average, that is a -8% reduction in fatalities.