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/PS2b")
#setwd("C:\\Users\\will-\\Desktop\\are213\\PS1b")
#setwd("C:\\Users\\Will\\Desktop\\are213\\PS2b")

dat <- read.dta("traffic_safety2.dta")
dat$vmtPC <- (1000000*dat$totalvmt)/(dat$population*1000)
var.labs <- attributes(dat)$var.labels
var.labs[15] <- "Vehicle Miles Traveled (VMT) per capita"

dat$lfatalPC <- log(dat$fatalities/(dat$population*1000))
var.labs[16] <- "log-fatalities per capita"

snames <- data.frame(state = c(attributes(dat)$label.table$state_number,99), slab = c(names(attributes(
dat <- merge(dat,snames)
rm(snames)
dat$slab <- as.character(dat$slab)

yrs <- unique(dat$year)
#Load the synthetic controls R package
library(Synth)
```

```
## ##
## ## Synth Package: Implements Synthetic Control Methods.

## ## See http://www.mit.edu/~jhainm/software.htm for additional information.
```

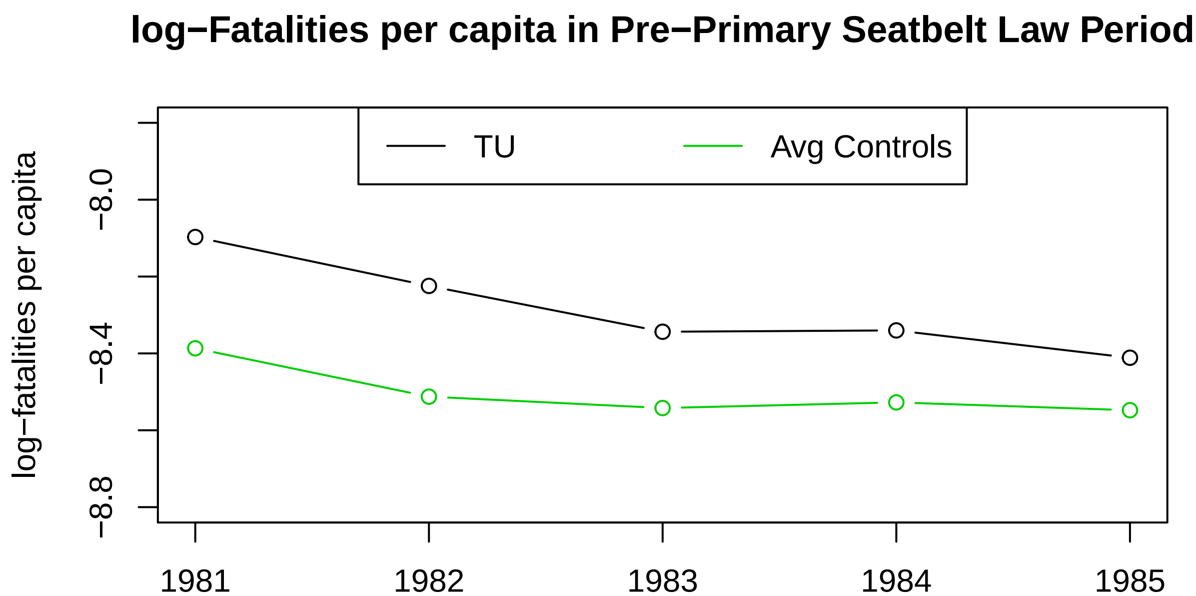
1a - Aggregate treatment analysis

```
TUpre <- dat %>% filter(state == 99 & primary == 0) %>% summarise(mean1 = mean(lfatalPC)) %>% as.numeri

#Exclude the 4 states (CT-6,IA-10,NM-30,TX-41) that comprise TU from controls and NJ (29) and NY (32) w
controls <- c(1:5,7:9,11:28,31,33:40,42:48)

ALLpre <- dat %>% filter(state %in% controls & primary == 0) %>% summarise(mean1 = mean(lfatalPC)) %>% a

kable(data.frame(Pre_TU = TUpre, Pre_ALLcontrols = ALLpre), caption = "Average log-fatalities for the T
```

Table 1: Average log-fatalities for the TU state and all the control states in the pre-primary seatbelt law period

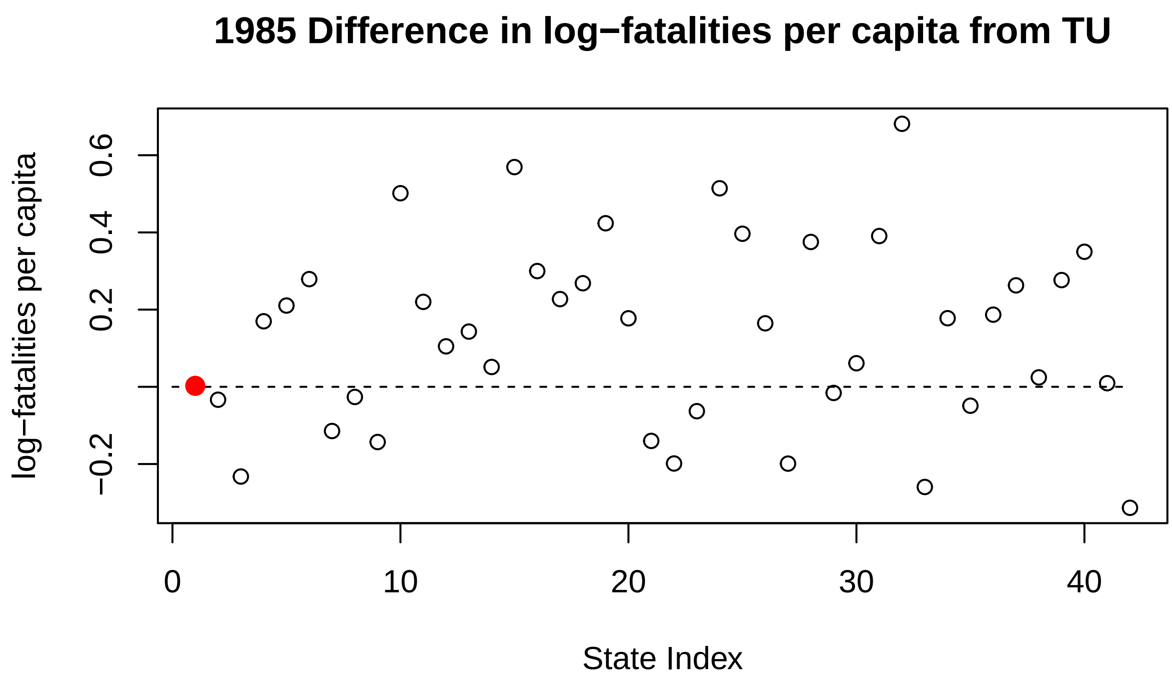
Pre_TU	Pre_ALLcontrols
-8.283	-8.606

```r
pre <- dat[dat$year %in% dat$year[which(dat$state==99 & dat$primary==0)],]
cavgpre <- pre %>% filter(state %in% controls) %>% group_by(year) %>% summarise(meanf = mean(lfatalPC))

plot(pre$year[pre$state==99],pre$lfatalPC[pre$state==99], ylab = "log-fatalities per capita", xlab = ""
lines(cavgpre$year,cavgpre$meanf, type = 'b', col = 3)
legend('top', legend = c("TU","Avg Controls"), lty = 1, col = c(1,3), horiz = T)
```



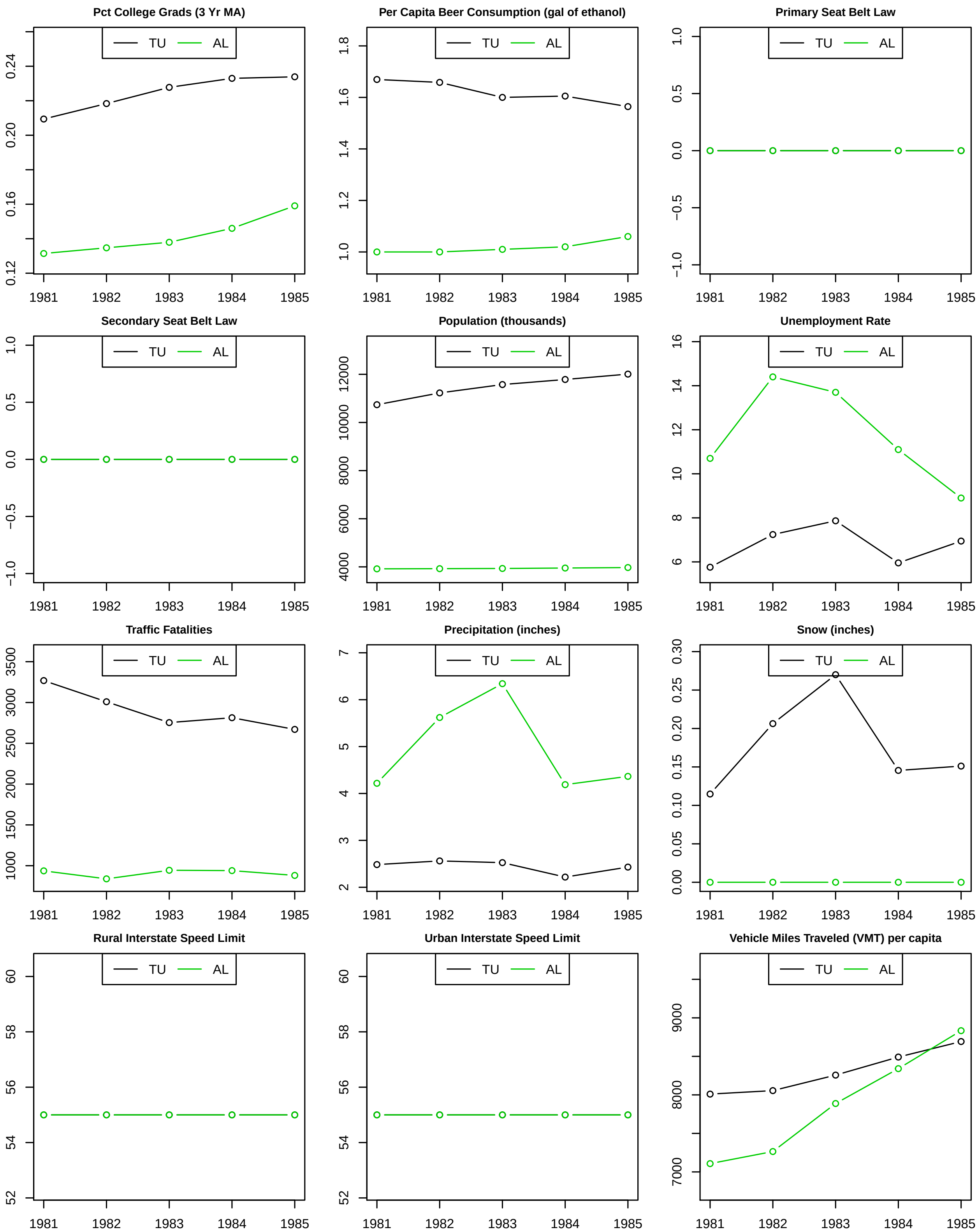
As seen in the graph above, we can see that there were on average fewer traffic fatalities in the control states during the pre-intervention period for the aggregate "treatment" state TU from 1981 - 1985. Averged over these 5 years, the treatment state TU had roughly 0.32 more log-traffic fatalities per capita and 7.2e-05 more (nonlog)-traffic fatalities per capita compared to the rest of the control states. This suggests that the states the received the primary seatbelt laws (the ones that make up TU) in 1985 were good candidates for such a law given their above-average rate of traffic fatalities.

```r
pre1 <- pre[pre$year==1985,]
anom85 <- pre1$lfatalPC[pre1$state==99]-pre1$lfatalPC[pre1$state %in% controls]
```



We can see that in 1985 - the year before the intervention in TU - the state most similar to TU in terms of log-fatalities per capita is state number 1 (AL-Alabama).

Comparison of covariates from Alabama to TU during the pre-intervention period.



Generally speaking, the covariates for Alabama during the 1981-1985 pre-treatment period for TU are similar in that Alabama also has no primary or secondary seatbelt law and in their rural and interstate speed limits of 55mph. The number of vehicle miles traveled per capita is fairly similar between TU and AL as well, especially in 1984-1985 just before the treatment begins in TU. However, the majority of other covariates are fairly different. For example, TU has a much higher college graduate rate (21-23% vs. 13-16%) during this period, more per capita beer consumption (0.5-0.6 more gallons of ethanol per capita), roughly half the unemployment rate of Alabama, much less average precipitation though much higher average snowfall. Other gross metrics, such as population, total traffic fatalities, and total vehicle miles traveled are more difficult to compare betwee TU and Alabama because they are not scaled by population, though their totals are much different all the same.

1b - Synthetic control method

The method of using synthetic controls, as introduced by Abadie et al. in 2010, is advantageous because it addresses two large issues facing diff-in-diffs, other differencing methods, and matching. These are the fact that oftentimes choosing control units, as in many diff-in-diffs studies, can be seemingly arbitrary, with a cursory analysis of similarities between treatment and control units yielding a subset of controls that may have been entirely different if a different researcher had made the decisions on controls. The choice of these controls, however, does still bear significant influence on the results, and therefore the arbitrary decision-making process potentially plays a significant role in conclusions being made from the study. Synthetic controls on the other hand, takes advantage of all the available control units, using them to create a weighted single "synthetic" control unit to compare to the treatment based on the similarities of all available control units' covariate values to the treatment unit. Additionally, the synthetic controls method rectifies the issue of dubious standard errors facing other methods given the fact that panel data can oftentimes have only a few number of observations for each unit over time, meaning the effective n of each unit is too small to calculate a reasonable variance or standard errors and there is no good representation of unit-by-time specific shocks. It does so by utilizing the multitude of available control units by constructing "placebo" estimates for untreated units, maing the estimation of estimator variance more robust.

However, potential drawbacks of the synthetic controls procedure include the fact that it relies on available data for a number of possible control units, uniformly available over the panel. Furthermore, even if there exist such data for a number of controls, it's possible that no combination of them as a synthetic control provide a reasonably good fit to the treated unit characterstics, meaning the researcher much assess as to whether the synthetic control is reasonable for the given treated unit. Similarly, if the control units being used are very different from one another in a geographic or regional sense, it may be difficult to avoid interpolation bias introduced from interpolating across regions with very different traits (Abadie et al. 2010).

```r
# Specify column number of dependent variable
dcol <- which(names(dat) == "lfatalPC")
# Identify pre-intervention periods
pre_period <- dat$year[dat$state==99 & dat$primary==0]
MPSE <- rep(NA,4)
```

The covariate specifications were experimented with to construct the synthetic controls during the pre-intervention period:

Specification 1:

```r
# Specify column numbers of predictors to use
pcols <- c(3:4,7:12,15)
noquote(c("Covariates used to construct synthetic control:",var.labs[pcols]))
```

```
## [1] Covariates used to construct synthetic control:
## [2] Pct College Grads (3 Yr MA)
## [3] Per Capita Beer Consumption (gal of ethanol)
## [4] Population (thousands)
## [5] Unemployment Rate
```

```
##  [6] Traffic Fatalities
##  [7] Vehicle Miles Traveled (VMT)
##  [8] Precipitation (inches)
##  [9] Snow (inches)
## [10] Vehicle Miles Traveled (VMT) per capita
```

```
# Prepare dataframe for synth function
synthdat1 <- dataprep(foo = dat,
                      predictors = pcols,
                      dependent = dcol,
                      unit.variable = "state",
                      time.variable = "year",
                      treatment.identifier = 99,
                      controls.identifier = controls,
                      time.predictors.prior = pre_period,
                      time.optimize.ssr = pre_period,
                      time.plot = yrs,
                      unit.names.variable = "slab")
synth.out1 <- synth(synthdat1)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 0.001266968
##
## solution.v:
##  0.01045675 0.1022152 0.2832795 0.002760243 0.002304067 0.1796989 0.227689 0.003823182 0.1877732
##
## solution.w:
##  5.66249e-05 5.42585e-05 0.0001038873 0.362404 0.0001001451 9.46631e-05 0.21796 0.0001072291 8.96564
```

```
MPSE[1] <- synth.out1$loss.v
```

Specification 2:

```
pcols <- c(3:4,8:12,15)
noquote(c("Covariates used to construct synthetic control:",var.labs[pcols]))
```

```
## [1] Covariates used to construct synthetic control:
## [2] Pct College Grads (3 Yr MA)
## [3] Per Capita Beer Consumption (gal of ethanol)
## [4] Unemployment Rate
## [5] Traffic Fatalities
## [6] Vehicle Miles Traveled (VMT)
## [7] Precipitation (inches)
## [8] Snow (inches)
## [9] Vehicle Miles Traveled (VMT) per capita
```

```r
# Prepare dataframe for synth function
synthdat2 <- dataprep(foo = dat,
                      predictors = pcols,
                      dependent = dcol,
                      unit.variable = "state",
                      time.variable = "year",
                      treatment.identifier = 99,
                      controls.identifier = controls,
                      time.predictors.prior = pre_period,
                      time.optimize.ssr = pre_period,
                      time.plot = yrs,
                      unit.names.variable = "slab")
synth.out2 <- synth(synthdat2)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 0.001317934
##
## solution.v:
##  0.0106391 0.08721108 0.0003481696 0.02065657 0.2495813 0.3495101 0.00103241 0.2810212
##
## solution.w:
##  4.46894e-05 4.43772e-05 0.0001021641 0.401198 8.08893e-05 7.80667e-05 0.2194026 5.52027e-05 7.58034
```

```r
MPSE[2] <- synth.out2$loss.v
```

Specification 3:

```r
pcols <- c(3:4,8,11:12,15)
noquote(c("Covariates used to construct synthetic control:",var.labs[pcols]))
```

```
## [1] Covariates used to construct synthetic control:
## [2] Pct College Grads (3 Yr MA)
## [3] Per Capita Beer Consumption (gal of ethanol)
## [4] Unemployment Rate
## [5] Precipitation (inches)
## [6] Snow (inches)
## [7] Vehicle Miles Traveled (VMT) per capita
```

```r
# Prepare dataframe for synth function
synthdat3 <- dataprep(foo = dat,
                      predictors = pcols,
                      dependent = dcol,
                      unit.variable = "state",
                      time.variable = "year",
                      treatment.identifier = 99,
```

```
                     controls.identifier = controls,
                     time.predictors.prior = pre_period,
                     time.optimize.ssr = pre_period,
                     time.plot = yrs,
                     unit.names.variable = "slab")
synth.out3 <- synth(synthdat3)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 0.0009380342
##
## solution.v:
##  0.04987788 0.3263691 0.1404927 0.2481937 0.2341473 0.0009193423
##
## solution.w:
##  0.0007622605 0.0009771509 0.0004160553 0.002366852 0.09811248 0.004966034 0.2249363 0.002297663 0.0
```

```
MPSE[3] <- synth.out3$loss.v
```

Specification 4:

```
pcols <- c(3:4)
noquote(c("Covariates used to construct synthetic control:",var.labs[pcols]))
```

```
## [1] Covariates used to construct synthetic control:
## [2] Pct College Grads (3 Yr MA)
## [3] Per Capita Beer Consumption (gal of ethanol)
```

```
# Prepare dataframe for synth function
synthdat4 <- dataprep(foo = dat,
                     predictors = pcols,
                     dependent = dcol,
                     unit.variable = "state",
                     time.variable = "year",
                     treatment.identifier = 99,
                     controls.identifier = controls,
                     time.predictors.prior = pre_period,
                     time.optimize.ssr = pre_period,
                     time.plot = yrs,
                     unit.names.variable = "slab")
synth.out4 <- synth(synthdat4)
```

```
##
## X1, X0, Z1, Z0 all come directly from dataprep object.
##
##
```

```
## ****************
##  searching for synthetic control unit
##
##
## ****************
## ****************
## ****************
##
## MSPE (LOSS V): 0.01060304
##
## solution.v:
##  0.01846698 0.981533
##
## solution.w:
##  0.007739791 0.00791832 0.01684898 0.01153733 0.218101 0.01461501 0.01485703 0.009418557 0.01361989 (
```

```
MPSE[4] <- synth.out4$loss.v
```

```
kable(data.frame(Specification = 1:4, MPSE = MPSE))
```

Specification	MPSE
1	0.0012670
2	0.0013179
3	0.0009380
4	0.0106030

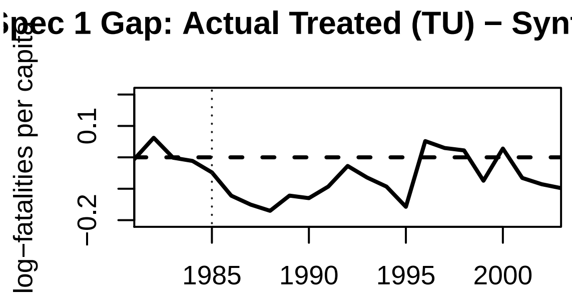
Ultimately, the specification 3 set of covariates was used for the final synthetic control construction due to its having the smallest MPSE during the pre-intervention period compared to other sets of covariates specifications. Specification 3 excludes population, since it was not scaled by area, and the gross population numbers do not necessarily convey any useful information for selecting control units. Also excluded were gross number of fatalities and totalvmt, which makes sense because they are also gross values and are not scaled to a per capita basis. The rural and urban speed limits are also excluded since they both are unvarying at 55mph across the country for the entire pre-intervention period. We can see that specification 4, which only used college education and beer consumption as its covariates had a much higher MPSE than the others, as expected.

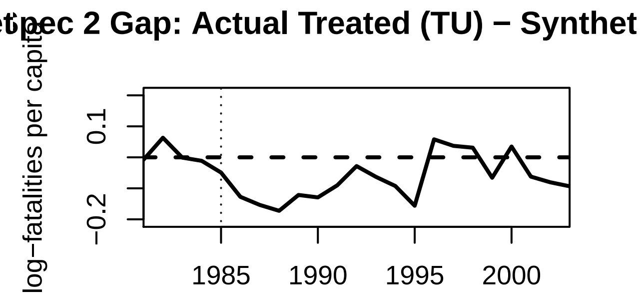
Generally speaking, the algorithm finds a synthetic control treatment and has reported the mean squared prediction error (MPSE) during the pre-intervention period as well as the weights applied to each covariate (solution.v) and to each control unit state (solution.w). Essentially, the command is averaging the pre-intervention period covariate values by each state and then comparing them to those pre-intervention averages of the treatment unit to determine the similarities of each state to each covariate. Then, it applies an optimization algorithm to test different weighting schemes for both the covariate and control unit-specific weights in order to construct a synthetic control unit that is most similar to the treated unit during the pre-intervention period (1981-1985), as defined by minimized MSPE during that period.

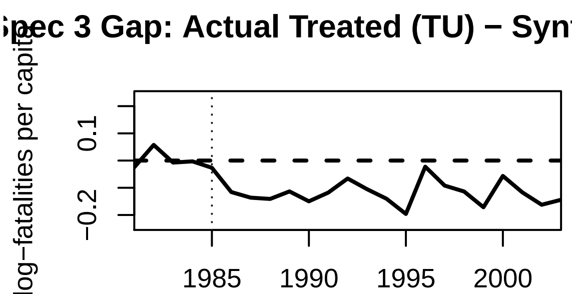
1c - Graphical interpretation and treatment

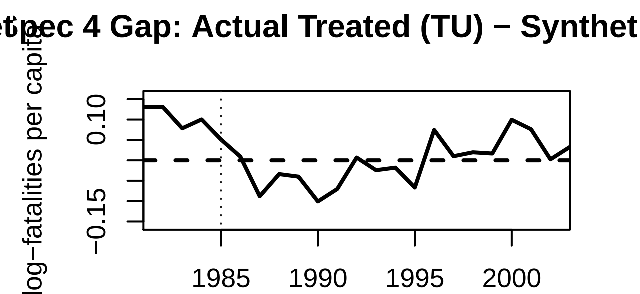
```
par(mfrow = c(2,2))
gaps.plot(synth.res = synth.out1, dataprep.res = synthdat1, Xlab = "", Ylab = "log-fatalities per capita
lines(c(1985,1985),c(-0.5,0.5), lty = 'dotted', col = 'grey10')
```

```r
gaps.plot(synth.res = synth.out2, dataprep.res = synthdat2, Xlab = "", Ylab = "log-fatalities per capit
lines(c(1985,1985),c(-0.5,0.5), lty = 'dotted', col = 'grey10')

gaps.plot(synth.res = synth.out3, dataprep.res = synthdat3, Xlab = "", Ylab = "log-fatalities per capit
lines(c(1985,1985),c(-0.5,0.5), lty = 'dotted', col = 'grey10')

gaps.plot(synth.res = synth.out4, dataprep.res = synthdat4, Xlab = "", Ylab = "log-fatalities per capit
lines(c(1985,1985),c(-0.5,0.5), lty = 'dotted', col = 'grey10')
```









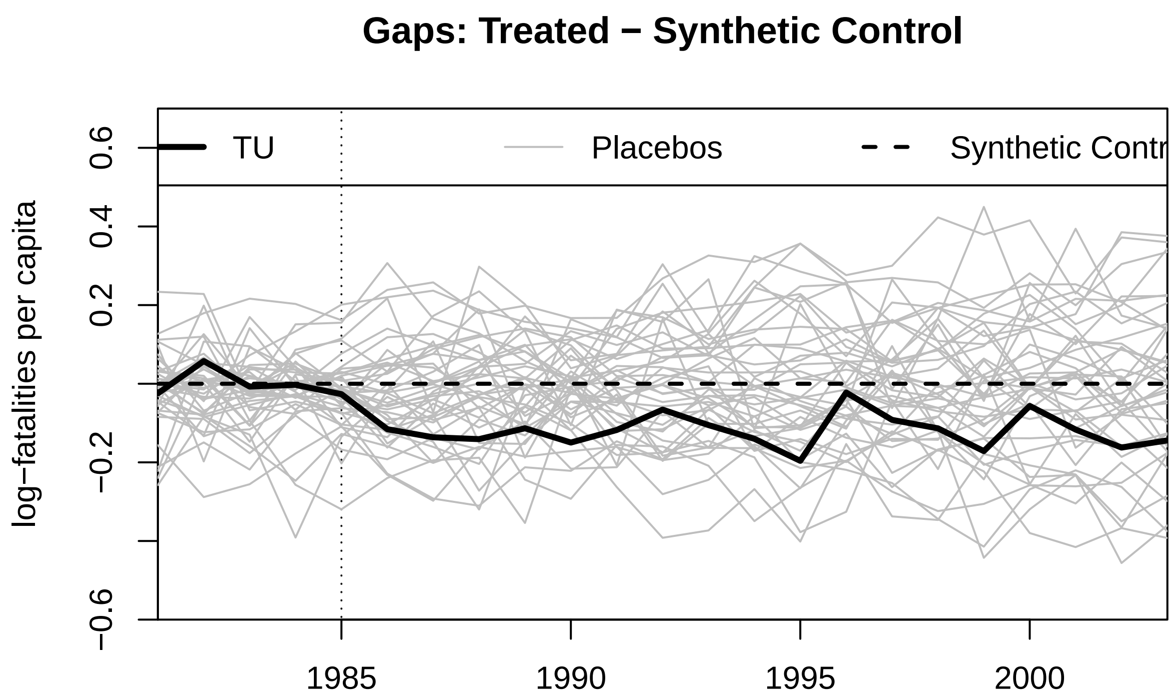
We can see in the plot of gaps between the treated TU and its synthetic control that during this 1981-1985 period the differences are very small in specification 3, which gives us confidence in the synthetic controls goodness of fit.

"Placebo" state treatments - Kept all years analysis (didn't worry about years in which primary or secondary laws were put into effect)

```r
# Run synthetic controls using the chosen specified covariates for all states (placebos)
controls <- controls[controls!=27]
gaps <- as.data.frame(matrix(NA,nrow = length(yrs),ncol = length(controls))) %>% set_colnames(controls)
pcols <- c(3:4,8,11:12,15)
for(c in controls){
  newcontrols <- c(controls[-which(controls==c)],99)
  synthdat <- dataprep(foo = dat,
                       predictors = pcols,
                       dependent = dcol,
                       unit.variable = "state",
                       time.variable = "year",
```

```r
                        treatment.identifier = c,
                        controls.identifier = newcontrols,
                        time.predictors.prior = pre_period,
                        time.optimize.ssr = pre_period,
                        time.plot = dat$year[dat$state==c],
                        unit.names.variable = "slab")
  synth.out <- synth(synthdat)
  gaps[1:length(dat$year[dat$state==c]),which(controls==c)] <- synthdat$Y1plot - (synthdat$Y0plot %*% sy
}
```

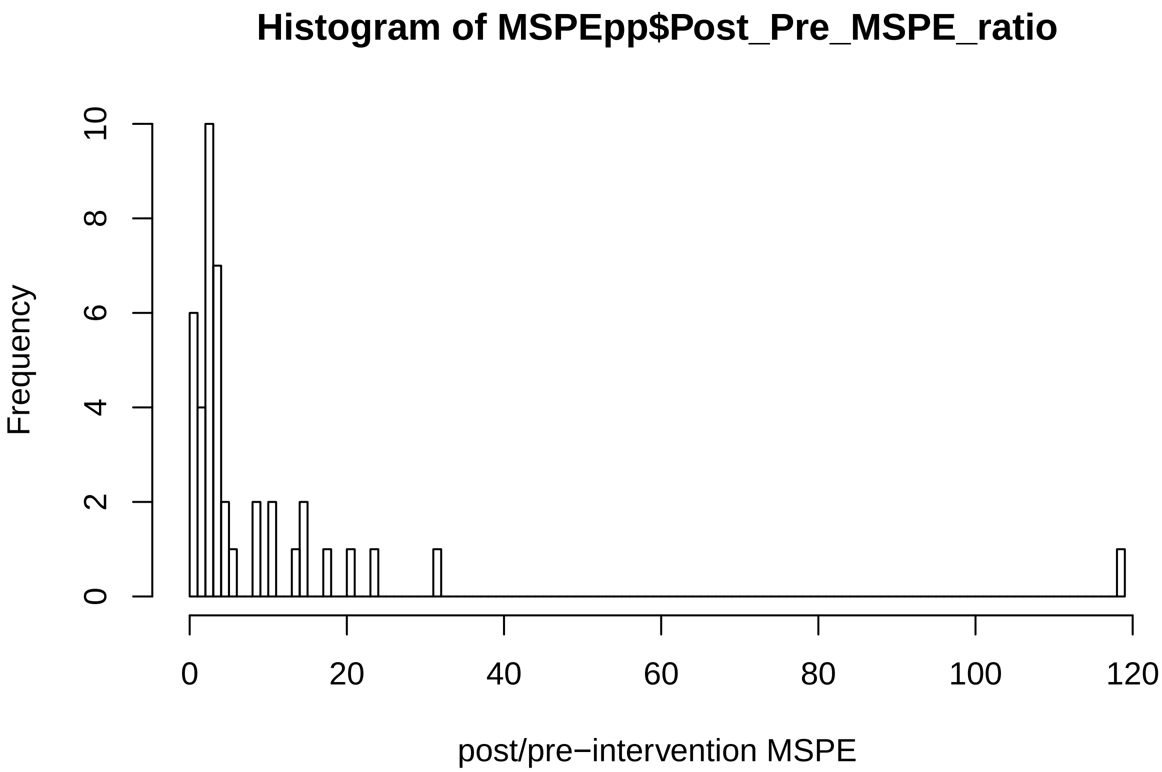
```r
gaps.plot(synth.res = synth.out3, dataprep.res = synthdat3, Xlab = "", Ylab = "log-fatalities per capita
lines(c(1985,1985),c(-5,5), lty = 'dotted', col = 'grey10')

for(c in 1:ncol(gaps)){
  lines(yrs,gaps[,c], col = 'grey')
}
lines(yrs,rep(0,length(unique(dat$year))),lty = 'dashed', lwd = 2)
TUgap <- synthdat3$Y1plot - (synthdat3$Y0plot %*% synth.out3$solution.w)
lines(yrs,TUgap, lwd = 3)
legend('top', legend=c("TU","Placebos","Synthetic Control"), horiz = T,col = c(1,"grey",1), lty = c(1,1
```



We can see that the TU is not well-outsidde the variance of other "placebo" control states relative to their respective synthetic controls. Therefore, it is hard from this plot to condlude that the treatment was significant. However, the control states in this case are not exactly great placebo controls because many of them experienced primary or secondary seatbelt laws at one point or another between 1985-2003, making the "placebo" concept not as substantive since in fact they may actually have been subject to a similar seatbelt law treatment.

```r
MSPEpp <- data.frame(Post_Pre_MSPE_ratio = rep(NA,length(controls)+1)) %>% set_rownames(unique(dat$slab
for(i in 1:(nrow(MSPEpp)-1)){
  # Only calculate ratio if there are at least 5 post-intervention years
  if(sum(!is.na(gaps[6:23,i])) >= 5){
    MSPEpp[i,1] <- mean(gaps[6:23,i]^2, na.rm = T)/mean(gaps[1:5,i]^2, na.rm = T)
```

```
  } else {
    MSPEpp[i,1] <- NA
  }
}
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")
```



From this plot we can see that indeed there is a state in which the post MSPE is much greater (~120x) than the pre-intervention MSPE. However, this state is in fact Florida, and not TU. The TU ratio was only roughly 17 times greater, which is still significant. It turns out that Florida passed a secondary seatbelt law in 1987 - right after the pre-treatment period for TU - which is likely contributing to some of this big difference in pre and post-intervention period MSPEs. The ratios for all states are shown below.

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

	Post_Pre_MSPE_ratio
AL	4.72
AR	3.06
AZ	2.14
CA	31.50
CO	2.96
DE	0.62
FL	118.28
GA	3.34
ID	10.08
IL	8.07
IN	2.41
KS	1.02
KY	2.34

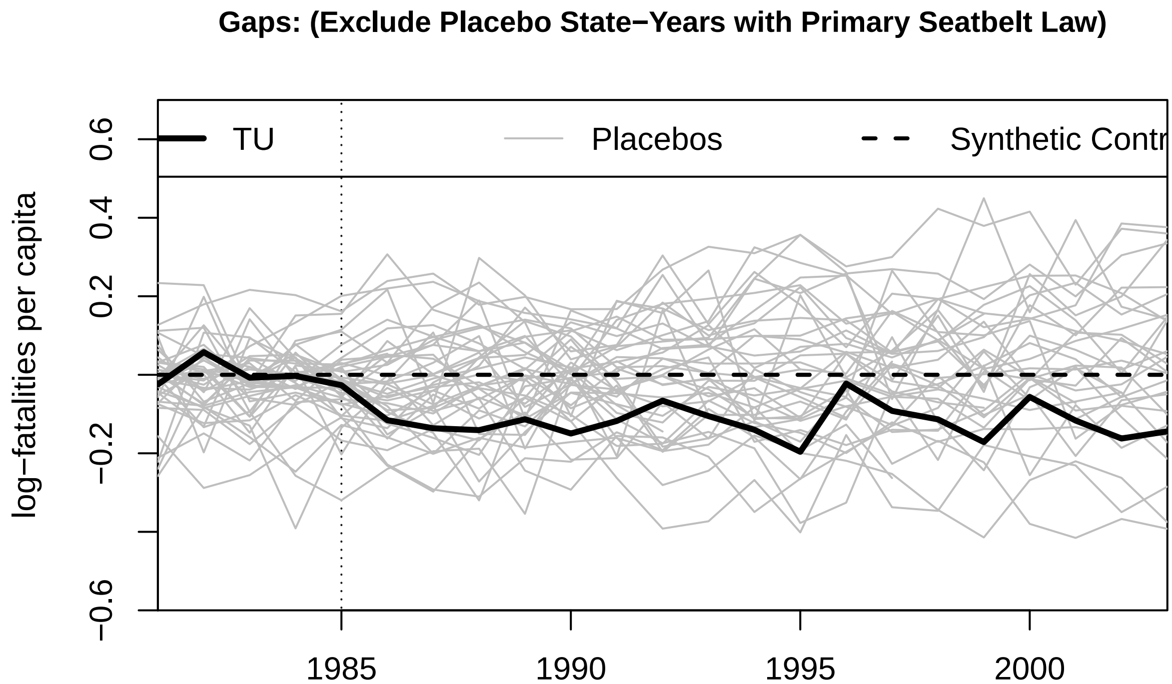
	Post_Pre_MSPE_ratio
LA	1.50
MA	8.08
MD	23.84
ME	3.40
MI	0.26
MN	0.50
MO	13.40
MS	2.79
MT	2.22
NC	14.09
ND	1.06
NH	5.29
NV	14.16
OH	1.44
OK	4.28
OR	20.27
PA	3.44
RI	0.58
SC	2.21
SD	3.60
TN	2.59
UT	10.43
VA	0.51
VT	3.40
WA	3.03
WI	0.82
WV	2.92
WY	2.03
TU	17.15

"Placebo" state treatments - Excluded years in which primary seatbelt laws were passed in other states from analysis (didn't worry about secondary seatbelt laws)

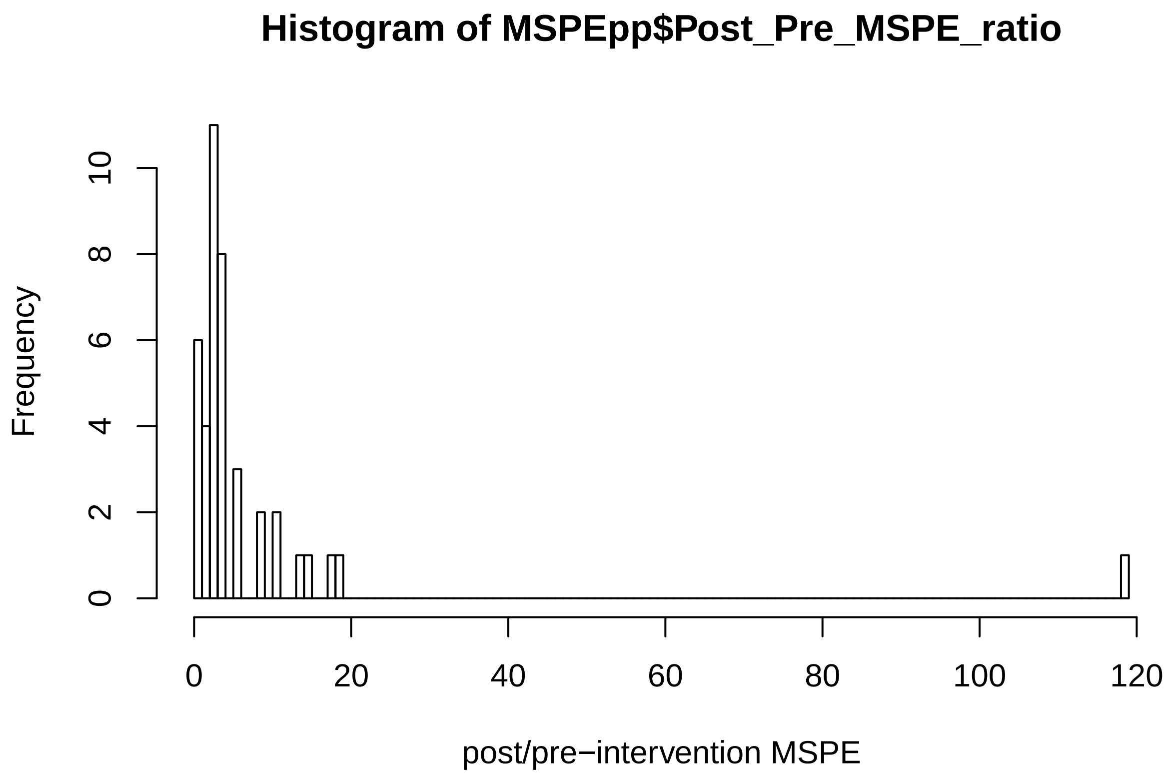
Therefore, to account for this 'contaminated' control data, which in fact has primary and seatbelt law treatment occurring throughout the study period, we can instead only use the years for each "placebo" control which do not have a primary and/or secondary treatment law in effect, to make them true "placebos". The following is the same process as above, but with state-years with primary seatbelt laws in effect removed.

```
par(cex.main = 0.9)
gaps.plot(synth.res = synth.out3, dataprep.res = synthdat3, Xlab = "", Ylab = "log-fatalities per capita
lines(c(1985,1985),c(-5,5), lty = 'dotted', col = 'grey10')

gaps_p <- as.data.frame((matrix(NA,nrow = nrow(gaps), ncol = ncol(gaps)))) %>% set_colnames(colnames(gap

for(c in 1:ncol(gaps)){
  gaps_p[which(yrs %in% dat$year[dat$state==controls[c] & dat$primary==0]),c] <- gaps[which(yrs %in% dat
  lines(yrs,gaps_p[,c], col = 'grey')
}
lines(yrs,rep(0,length(unique(dat$year))),lty = 'dashed', lwd = 2)
```

```
TUgap <- synthdat3$Y1plot - (synthdat3$Y0plot %*% synth.out3$solution.w)
lines(yrs,TUgap, lwd = 3)
legend('top', legend=c("TU","Placebos","Synthetic Control"), horiz = T,col = c(1,"grey",1), lty = c(1,1
```




```
MSPEpp <- data.frame(Post_Pre_MSPE_ratio = rep(NA,length(controls)+1)) %>% set_rownames(unique(dat$slab
for(i in 1:(nrow(MSPEpp)-1)){
  # Only calculate ratio if there are at least 5 post-intervention years
  if(sum(!is.na(gaps_p[6:23,i])) >= 5){
    MSPEpp[i,1] <- mean(gaps_p[6:23,i]^2, na.rm = T)/mean(gaps_p[1:5,i]^2, na.rm = T)
  } else {
    MSPEpp[i,1] <- NA
  }
}
MSPEpp[nrow(MSPEpp),1] <- mean(TUgap[6:23]^2, na.rm = T)/mean(TUgap[1:5]^2, na.rm = T)
hist(MSPEpp$Post_Pre_MSPE_ratio, breaks = 100, xlab = "post/pre-intervention MSPE")
```




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

	Post_Pre_MSPE_ratio
AL	5.65
AR	3.06
AZ	2.14
CA	3.84
CO	2.96
DE	0.64
FL	118.28
GA	5.15
ID	10.08
IL	8.07
IN	1.75
KS	1.02
KY	2.34
LA	2.14
MA	8.08
MD	18.56
ME	3.40
MI	0.17
MN	0.50
MO	13.40
MS	2.79
MT	2.22
NC	NA
ND	1.06
NH	5.29
NV	14.16

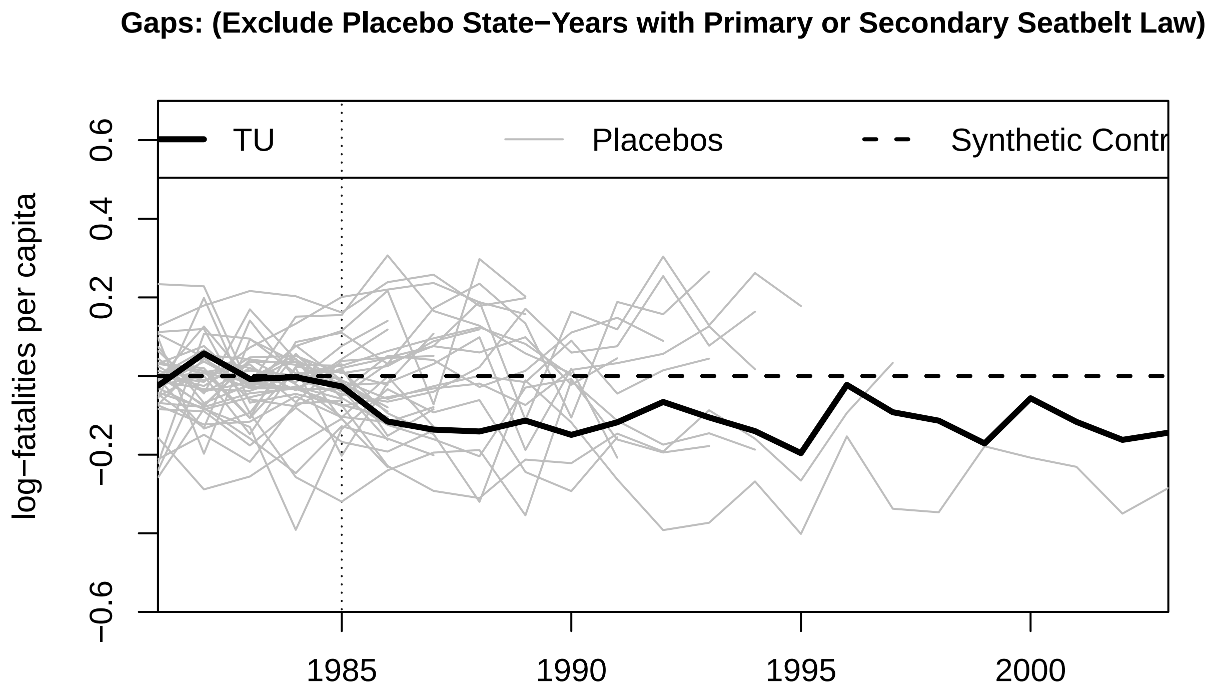
	Post_Pre_MSPE_ratio
OH	1.44
OK	3.80
OR	3.79
PA	3.44
RI	0.58
SC	2.21
SD	3.60
TN	2.59
UT	10.43
VA	0.51
VT	3.40
WA	2.57
WI	0.82
WV	2.92
WY	2.03
TU	17.15

We can see that even with excluding state-years with primary seatbelt laws in effect, not much has change. While TU is still among the states with the highest post/pre MSPE ratio, Florida is much higher and Maryland is slightly higher, making it difficult to concretely conclude that the TU primary seatbelt law was solely responsible for its decreased traffic fatalities.

"Placebo" state treatments - Excluded years in which primary or secondary seatbelt laws were passed in other states from analysis

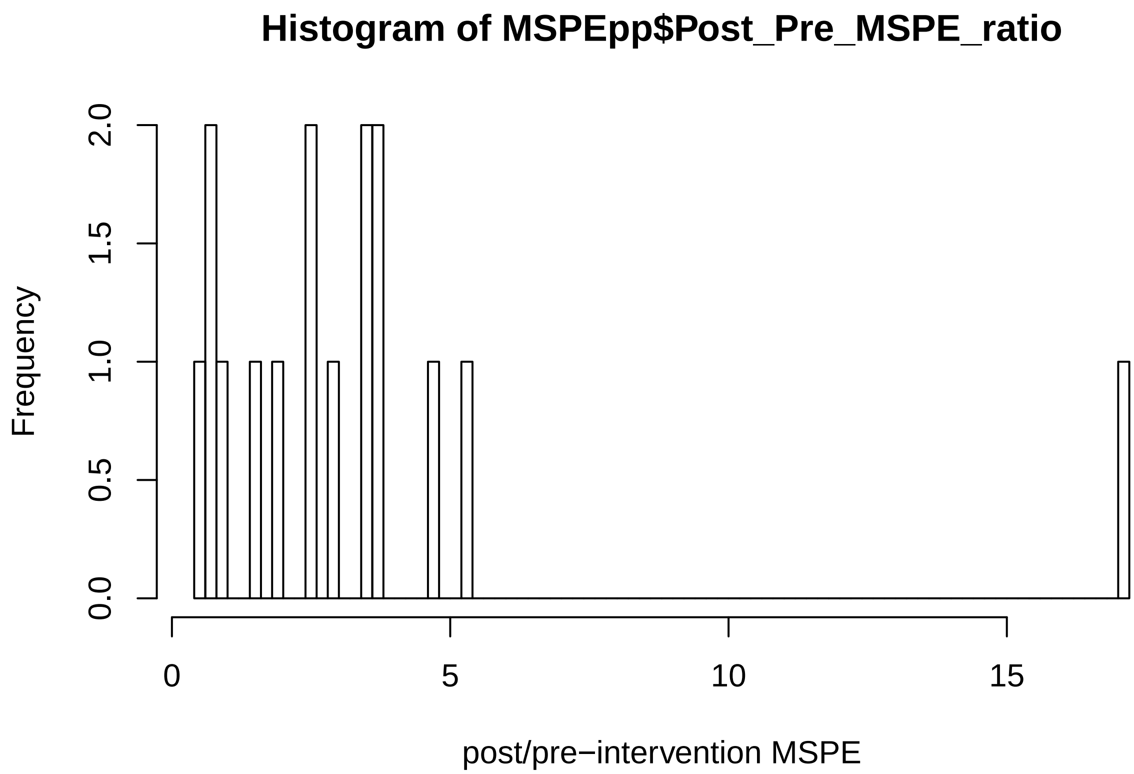
The final experiment worth running is to exclude all state-years that have either a primary or secondary seatbelt law in effect, therefore limiting the influence of similar treatments to the supposed "placebo" control states, which are not supposed to actually receive any treatment.

```
par(cex.main = 0.9)
gaps.plot(synth.res = synth.out3, dataprep.res = synthdat3, Xlab = "", Ylab = "log-fatalities per capita
lines(c(1985,1985),c(-5,5), lty = 'dotted', col = 'grey10')

gaps_ps <- as.data.frame((matrix(NA,nrow = nrow(gaps), ncol = ncol(gaps)))) %>% set_colnames(colnames(ga

for(c in 1:ncol(gaps)){
  gaps_ps[which(yrs %in% dat$year[dat$state==controls[c] & dat$primary==0 & dat$secondary==0]),c] <- gap
  lines(yrs,gaps_ps[,c], col = 'grey')
}
lines(yrs,rep(0,length(unique(dat$year))),lty = 'dashed', lwd = 2)
TUgap <- synthdat3$Y1plot - (synthdat3$Y0plot %*% synth.out3$solution.w)
lines(yrs,TUgap, lwd = 3)
legend('top', legend=c("TU","Placebos","Synthetic Control"), horiz = T,col = c(1,"grey",1), lty = c(1,1
```



MSPE post/pre

```
MSPEpp <- data.frame(Post_Pre_MSPE_ratio = rep(NA,length(controls)+1)) %>% set_rownames(unique(dat$slab
for(i in 1:(nrow(MSPEpp)-1)){
  # Only calculate ratio if there are at least 5 post-intervention years
  if(sum(!is.na(gaps_ps[6:23,i])) >= 5){
    MSPEpp[i,1] <- mean(gaps_ps[6:23,i]^2, na.rm = T)/mean(gaps_ps[1:5,i]^2, na.rm = T)
  } else {
    MSPEpp[i,1] <- NA
  }
}
MSPEpp[nrow(MSPEpp),1] <- mean(TUgap[6:23]^2, na.rm = T)/mean(TUgap[1:5]^2, na.rm = T)
hist(MSPEpp$Post_Pre_MSPE_ratio, breaks = 100, xlab = "post/pre-intervention MSPE")
```




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

	Post_Pre_MSPE_ratio
AL	3.49
AR	2.89
AZ	2.45
CA	NA
CO	NA
DE	0.68
FL	NA
GA	NA
ID	NA
IL	NA
IN	NA
KS	NA
KY	1.45
LA	NA
MA	3.40
MD	NA
ME	2.50
MI	NA
MN	NA
MO	NA
MS	NA
MT	NA
NC	NA
ND	0.97
NH	5.29
NV	NA

	Post_Pre_MSPE_ratio
OH	NA
OK	3.80
OR	3.79
PA	NA
RI	0.58
SC	NA
SD	1.90
TN	NA
UT	NA
VA	NA
VT	4.77
WA	NA
WI	NA
WV	0.74
WY	NA
TU	17.15

Now, with both primary and secondary seatbelt law state-years removed from this analysis we do indeed see that TU has the highest post/pre MSPE ratio. However, states that had fewer years in the post-intervention period under these restrictions than in the pre-intervention period did not receive a ratio calculation due to the sample size being too small. In other words, only states that had no primary or secondary seatbelt laws from at least 1986-1990 or longer were subject to the ratio calculation, which excluded Florida and Maryland, making TU the state with the highest ratio value. However, with so many restrictions in place the number of placebo states and data points decline dramatically, making the analysis less robust.

1d - Comparison to fixed effects

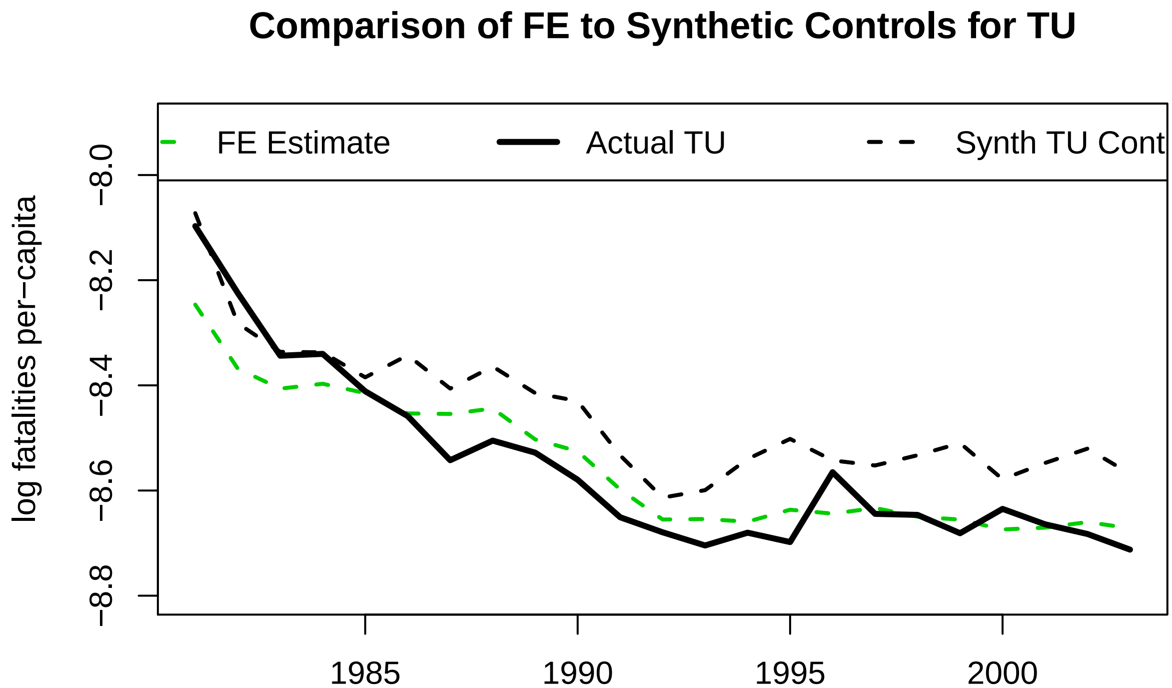
```
dat$time <- as.numeric(as.character(dat$year))-1981
dat$time_sq <- (dat$time)^2
dat$state <- as.factor(dat$state)
dat$year <- as.factor(dat$year)
```

```
library(lfe)
formula <- 'lfatalPC ~ primary + time_sq'
fixed_eff <- felm(as.formula(paste(formula, '+ year | state | 0 | 0' )), dat)
```

```
## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```

```
fe_fit <- fixed_eff$fitted.values[dat$state==99]
```

```
plot(1981:2003,fe_fit, type = 'l', ylim = c(-8.8,-7.9), col = 3, lty = 'dashed', main = "Comparison of F
lines(1981:2003,synthdat3$Y1plot, lwd = 3)
lines(1981:2003,synthdat3$Y0plot %*% synth.out3$solution.w, lwd = 2, lty = 'dashed')
legend('top',legend = c("FE Estimate", "Actual TU", "Synth TU Control "), col = c(3,1,1), lty = c(2,1,2
```



The FE and synthetic controls estimators look very similar in fact with no systematic differences between them. The FE estimates during the control period are slightly lower than the synthetic control estimates during that time, which we know are very close to the actual values for TU during 1981-1985. This suggestes that the FE estimator is slightly biased low for those 5 years, which could be due to the use of the quadratic time term, which has a negative coefficient, which may be absorbing some of the effects of secondary seatbelt laws coming into effect over time in many states over the panel, and therefore downweighting estimates in states merely due to passing time rather than due to the enacting of those laws in the FE model.