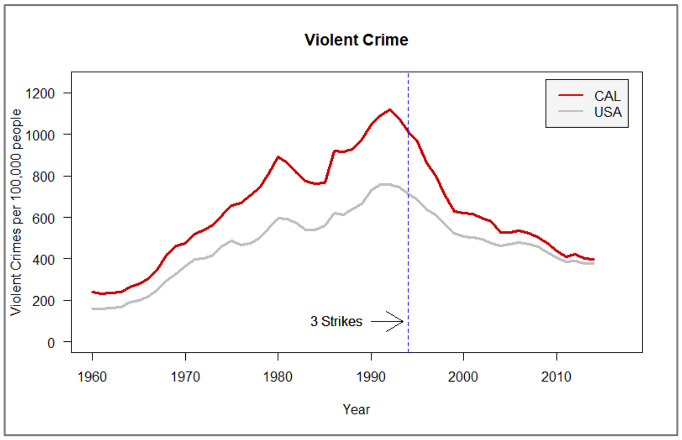
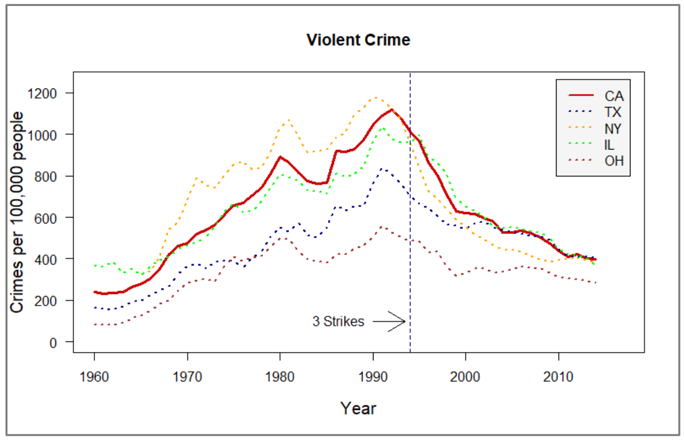
An Evaluation of California's Three Strikes Law
Gabelli School of Business: Observational Studies

Jonathan McWilliams
May 2019

Abstract

The stated goal of this research project was to test the effectiveness of California's Three Strikes law[1]. Enacted in 1994, the law was made as a reaction to three decades of increased crime in California and the US in general. The environment that led to the law's enaction was charged on both sides due to the nation's recent crack cocaine epidemic, police brutality witnessed during the Rodney King incident, and most importantly, a few high-profile cases of repeat offenders. The Three Strikes (3S) premise was simple; punish crime with increasingly harsh sentences for repeated offences. Crimes that would normally be reviewed on a case-by-case basis were





now pushed through courts with mandatory jail sentences for numerous crime types – not limited to just violent crime. A defendant facing a sentence for a third offense (notwithstanding whether it was violent or not) could face 25 years to life. The 3S enactment had a two-fold goal: (1) removing repeat offendors from the street and (2) discouraging would-be criminals from breaking the law. This project will attempt to answer if California's *violent crime rate* was affected by to the passage of California's 3S law.

Project Scope

Discussion of Analysis Method

The research method chosen for this study is Synthetic Control Method (SCM). This is a fairly new statistical method that is related to the Difference in Differences technique. SCM is unique in that it creates a fake (i.e. synthetic) control group that correlates to the treated sample *before* treatment and then measures the difference between the synthetic control and the treated group *after* treatment to assess the treatment's impact. This synthetic control is created by taking parts of individual control states and creating a weighted average of the synthetic group. The weights applied to the individual control states is dependent on the control group's predictive variable correlation to the treatment state's predictive variables.

For this analysis, the application of SCM was done at a state level. Our potential control states (or *"donor regions"* as they are more formally known)[4] were all US states (Washington D.C. included) that did *not* have similar crime legislation as California. If we mistakenly used a donor region with similar legislation as the treatment state, the impact of the treatment could not be attributed to the actual treatment itself – rendering our research useless. However, there is a very specific exception to this rule, If the donor states (i.e. controls) had similar treatment – in this case, harsh criminal sentencing - passed outside of the testing period, then they could be used as donor states. This exception is allowed because SCM cares about the *correlation* of the donor state to the treatment more than the absolute difference between the two. In our case, two specific exceptions were included because of this rule that would have been ruled out otherwise: Texas and New York. Both states had harsh legislation similar to 3S but both had that legislation passed significantly before our research timeframe – therefore we can conclude that the effects of these laws on Texas and New York were already included in our sample and we could use these states similar to any other donor region.

The general steps to complete SCM are as follows:

1. Filter out all states that had similar legislation passed during the testing timeframe[3]
2. Of the states left from step 1, assess the correlation of their crime rate to California's
3. Using the most promising donor states from step 2, assess which predictor variables have the most significant correlation (or anti-correlation) to violent crime (the predicted variable) and use these in the model

Data Selection and Sources

Most data used was gotten from .GOV sites (BEA, BLS, BJS, etc). Data was gathered at the state and year level from 1960 to 2017. Since few of the variables had consistent data through the entire 1960 - 2017 timeframe, the analysis was condensed to 1978 – 2014, maximizing full data coverage. The data was pulled from its respective site, input into excel, manipulated using pivot tables, then combined into one holistic dataset. There were instances of complications when finding and gathering the data – the most common issues were gaps in annual data (i.e. proportion of population living in urban neighborhood) or consistency in a variable's measurement changing from year to year (i.e. % 18-24 YoA with HS degree not available before 2005, only 25+ YoA).

Discussion of Variables

1. ***Violent crime rate***[5] *(the predicted variable)* represents the violent occurrences per 100,000 people. This statistic is made up of 4 main crime types (as specified by the Bureau of Justice Statistics): *Aggravated Assault*, *Robbery*, *Rape*, and *Murder*. This variable was available for all years in the study and all 51 regions (50 states plus Washington D.C.).

2. State level ***population***[5] was used both as a lone predictor and as the denominator in another predictor variable (% of population incarcerated).

3. ***Real GDP per capita***[6] was assumed to be negatively correlated with violent crime. A state with a high GDP is assumed to have a lower relative risk of crime. An important step in using this variable was restating everything to base year equivalency using the Consumer Price Index.

4. ***Median family income***[7] was included as a *personal* measure of a state's wealth. While this could be looked at as a redundant measure to GDP, GDP includes value created by corporations along with individuals. MFI is a more personal measure of what one could hope to earn in a given state.

5. ***Proportion of population 25+ YoA with High School education***[9] was assumed to be a solid predictive measure of a state's education level. This variable was assumed to have a negative correlation to the violent rate.

6. ***Unemployment rate***[12] was assumed to put a state at higher risk for crime.

7. The USA's growth toward urban living seemed to coincide with a dramatic rise in crime over the same timeframe. Therefore, the ***proportion of population that is urban***[11] was included in the analysis.

8. ***Incarcerated population***, ***admissions***, and ***releases***[10] were included to create an average duration of prisoner stay. This calculation will act much the same way one would calculate inventory turnover for a business; the higher turnover points toward more lenient state sentencing, which in turn could be a slight incentive for increased crime.

9. ***Proportion of population incarcerated*** was included in the hopes that it could show a negative correlation to violent crime; as more criminals are incarcerated there are less people to commit crime. The calculation was *Incarcerated Population / Total Population.*

Adjustments Made to Data

Most data was taken at face-value but there was significant changes made to a few select variables.

- GDP pCap was downloaded in two separate blocks: 1963 to 1997, and 1997 to forward. The first block was chained to 1997 dollars while the second block was tied to 2012 dollars (this is how the BEA published it). The second block was converted to 1997 levels using CPI from both years multiplied by the GDP pCap: (1997 CPI of 160.5 ÷ 2012 CPI of 212.6) x GDP per Capita = restated GDP per Capita. This conversion was done so that all GDP could be compared without the effect of inflation.

- Proportion of adults 25+ with High School degree had spotty information. It was available for the 15 largest states but was frequently unavailable for the smaller states. Additionally, between 1977 and 1988, 6 years of data was missing (large states included) so a straight-line average was employed to fill in the missing spaces.

- Median Family Income was adjusted for inflation using CPI. All MFI was adjusted back to 1974 levels for consistency.

Exploratory Data Analysis

Eligible and Non-Eligible Donor States

There were 50 potential donor states before any exclusions (inclusive of D.C). For SCM, we cannot include treated states in our pool of donor states – the reasoning is explained in the *Discussion of Method* above. So all states with a Three Strikes law (or a similar version) that was passed during the analysis timeframe (1978 – 2014) would have to be excluded. This first step filtered out 21 of the initial 50, leaving 29 potential donor states.

Finding Optimal Predictor Variables

Not all of the predictor variables available are statistically significant to predict crime – so further analysis must be done to select the correct control ones. A two-step process was employed to find these optimal predictor variables

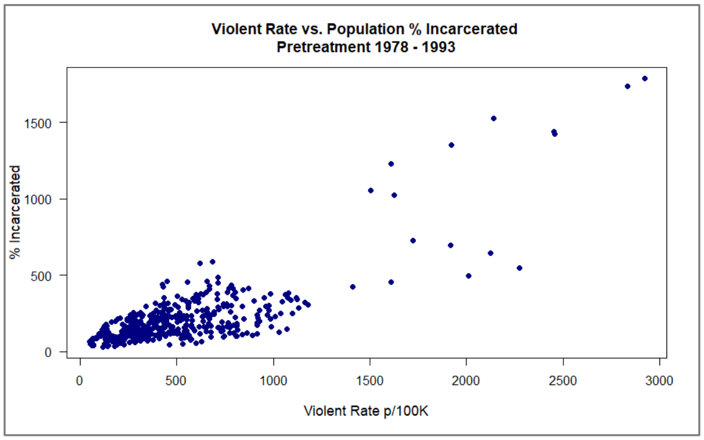
1. Find which potential donor states had crime rates correlated to California
2. Of these donors, identify which variables are correlated to the violent crime rate
3. Once these variables were identified, build the model with these selected predictive variables
4. Allow the SCM algorithm to pick the weights for donor states and variables

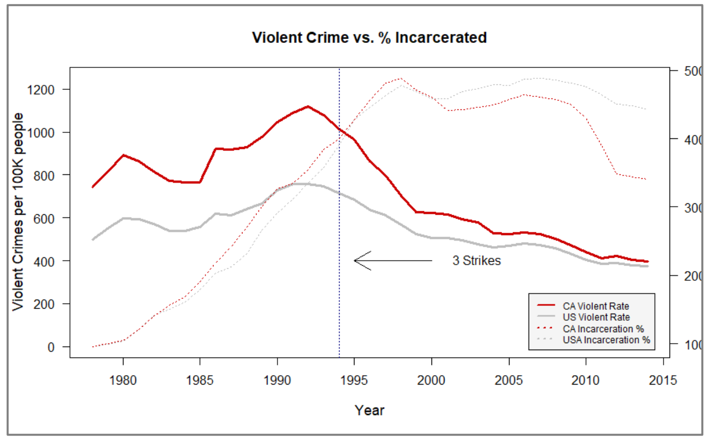
Of the 29 potential donor states, correlation between their crime rates and Californias' was analyzed and the top 14 states were chosen as close matches to California. These 14 states had correlation of 0.83 or better to California's crime rate. Once these states were identified, their predictive variables (along with California) were assessed for correlation to their respective crime rates. The outcome from this exercise would identify which predictive variables were high-quality. The results (below) shows poor correlation for *Jail Population Churn*, *Median HH Income*, and *Unemployment*. These variables were left out of the final model.

PREDICTOR (X)	CORRELATION TO Y	PREDICTOR (X)	CORRELATION TO Y
INCARCERATED POPULATION %	0.841	~~*JAIL POPULATION CHURN*~~	~~*0.284*~~
GDP	0.775	~~*MEDIAN HH INCOME*~~	~~*0.150*~~
EDUCATION	0.655	~~*UNEMPLOYMENT*~~	~~*<0.188>*~~
URBAN %	0.598		

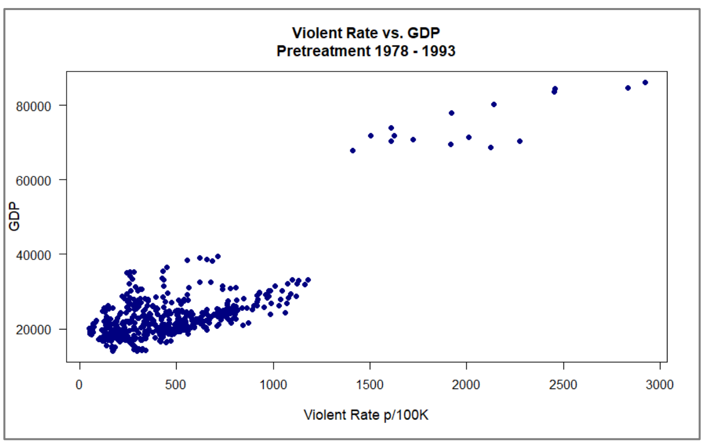
Visual checks where then completed on the 4 residual predictive variables to confirm a trend pre-treatment. The graphs below show relationships between the predictive variables and crime rates for the regions listed. The scatter plots represent only the pre-treatment years while the line graphs look at the entire time frame of 1978 – 2014.

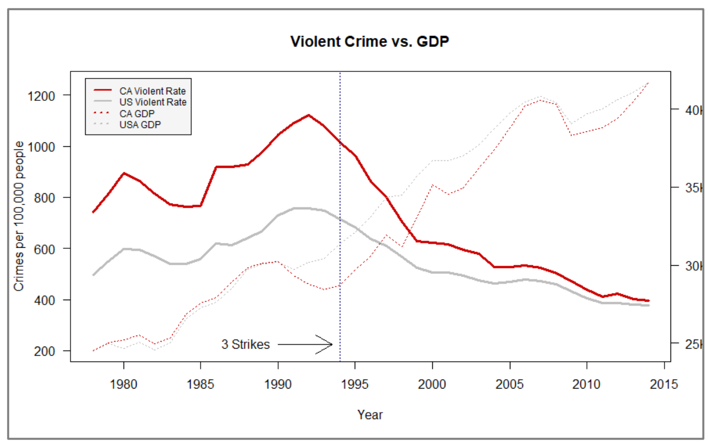
The proportion of population that is incarcerated has a remarkable correlation to the violence crime rate.



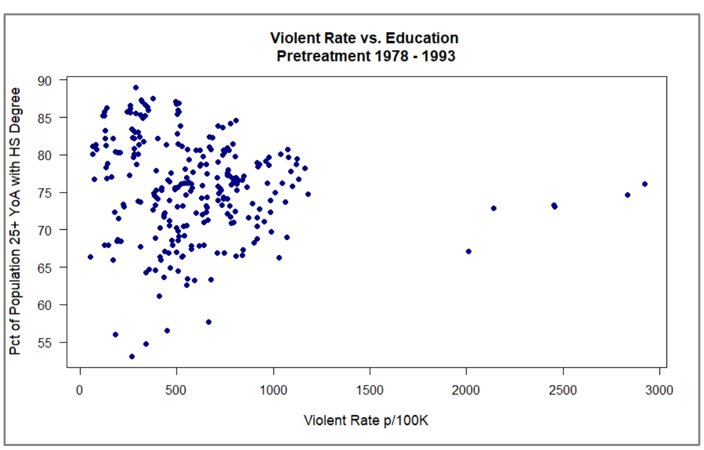


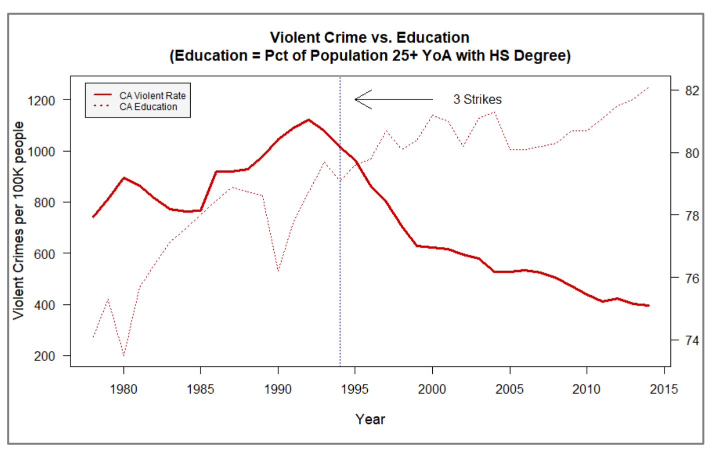
GDP had a surprising *positive* correlation to violent crime.



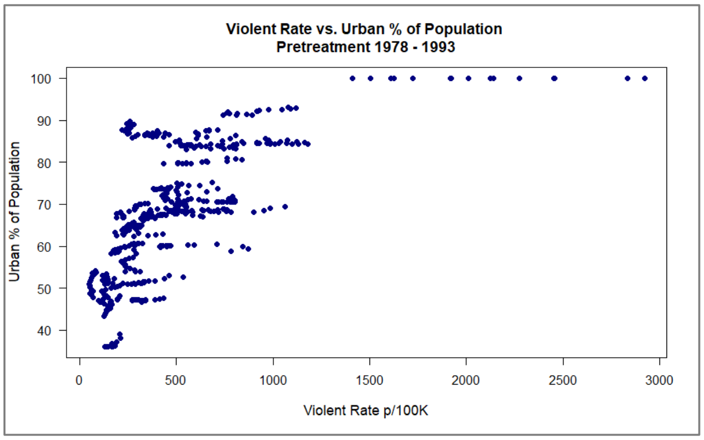


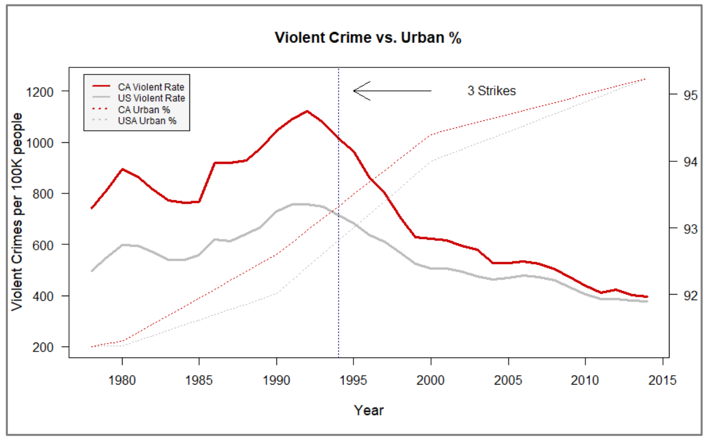
Education showed a positive correlation as well.





The proportion of the population that lives in an urban environment also has positive correlation to violent crime during the pre-treatment years – which is the signal that we expected.

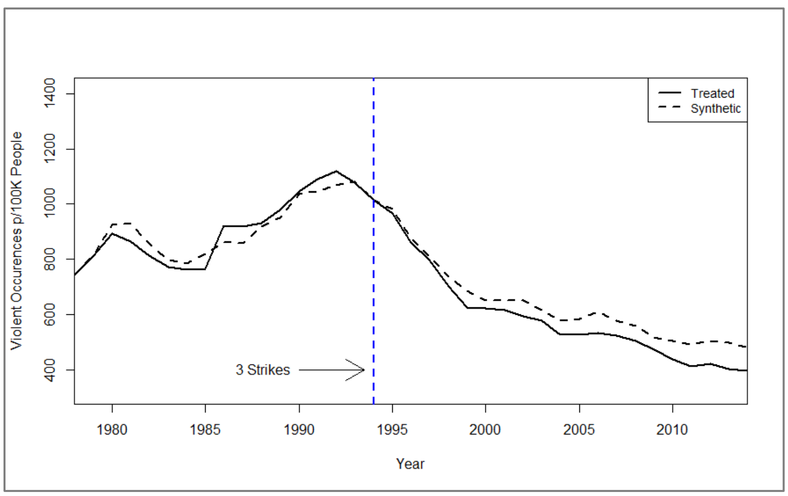


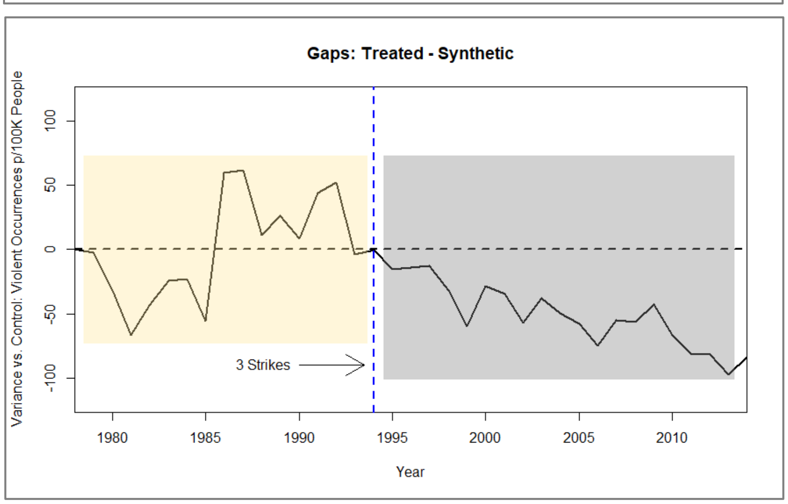


Modeling

Model Build

As stated above, the model was built with 4 predictor variables and given the option to choose from 29 donor regions. Here are the results of the model when given the full pool of 29 potential donors. The first graph shows the difference between real California (solid line) and it's synthetic control (dashed line). The synthetic was made from Arizona 48%, New York 32%, D.C. of 11%, Illinois of 8%, and Massachusetts/Rhode Island making up the residual 1%. It's worth noting that only New York was found in the highly correlated donor list discussed in the *EDA* section of this study.

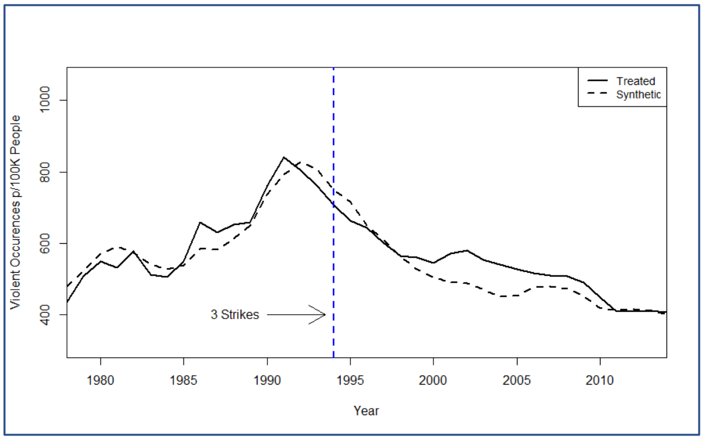


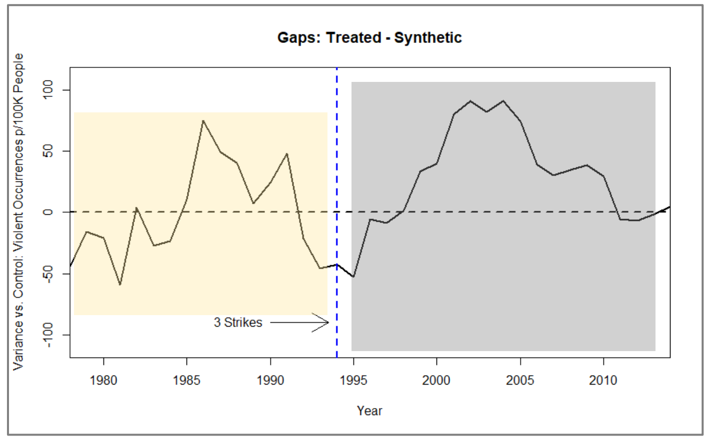


The second graph uses the same data as the first, but the synthetic control is laid down flat as the x-axis. We can clearly see that the pre-treatment fit leaves much to be desired in the way of a fit. The pretreatment MSE (yellow box) is 32 – meaning the model differs from real California by 32 violent crimes per 100K annually. The post-treatment MSE (grey box) differs from the synthetic by 49 violent crimes per 100K annually. The extra variance in the post-treatment timeframe is somewhat of a signal that Three Strikes had an impact on crime (albeit a weak signal). We can check our results by setting up a model with a different treatment state or moving the date of the treatment in the original model.

Model Verification Check One – changing the treatment state

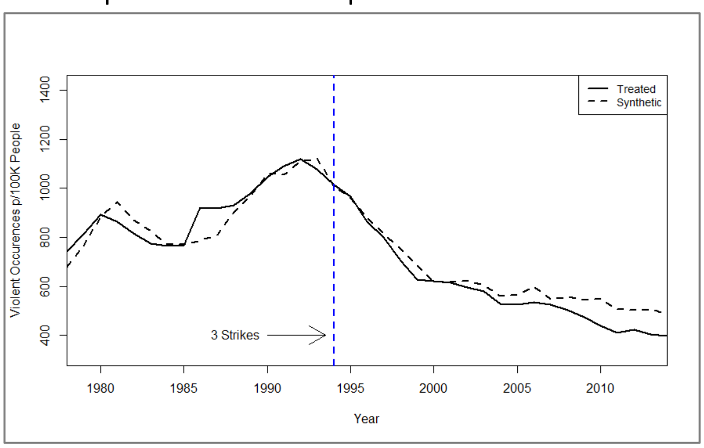
Applying treatment to a control state will be done by telling the model that *Texas* is the treated state and California goes into the pool of 29 donor states. The results show the opposite of the real model above. The pre-treatment MSE (yellow) is 32, very close to the real model, but the post-treatment MSE is only 38, showing little significance of the treatment impact. Additionally, the *real* Texas actually paces *above* it's synthetic control.

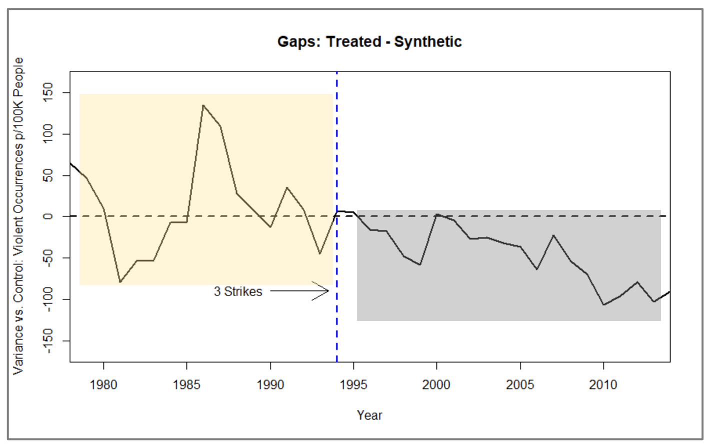




Model Verification Check Two – changing the treatment date

We now revert back to the original (i.e. real) model but will change the treatment date to 2000 instead of the actual 1994. The pre and post treatment MSE are 44 and 46 respectively, barely any difference between them. We can also see that the pre-treatment fit is quite bad.





Model Conclusion

While the our original model leaves a lot to be desired for a better pre-treatment fit, it *does* perform better than the two fake models we use to check it. Based on the data we have and the output we have seen it seems to signal that Three Strikes ***was*** part of the reason that California's crime decreased through the 1990s and 2000s. Our conclusion should be the model shows significance behind the 1993 treatment to California.

References

Discussion & Analysis behind CA's Three Strikes

1. Description of Three Strikes Law: https://www.aerlawgroup.com/overview-of-californias-three-strikes-sentencing-law/ and https://www.courts.ca.gov/20142.htm
2. State rankings of criminal leniency: http://www.unz.com/anepigone/strictness-index-which-states-are/
3. State list of Three Strikes implementation: http://www.inquiriesjournal.com/articles/696/assessing-the-impact-of-three-strikes-laws-on-crime-rates-and-prison-populations-in-california-and-washington

Synthetic Control Method description
4. https://www.urban.org/sites/default/files/publication/89246/the_synthetic_control_method_as_a_tool_0.pdf

Predicted Variable
5. Crime Statistics: Department of Justice: https://www.ucrdatatool.gov/Search/Crime/Crime.cfm

Predictor Variables
6. GDP: https://apps.bea.gov/itable/iTable.cfm?ReqID=70&step=1#reqid=70&step=1&isuri=1
7. Median Family Income (1974 – 2005): www.census.gov/data/tables/time-series/demo/income-poverty/4-person.html
(2006 – 2009): https://nces.ed.gov/programs/digest/d10/tables/dt10_025.asp
8. CPI & Annual Inflation: https://www.minneapolisfed.org/community/financial-and-economic-education/cpi-calculator-information/consumer-price-index-and-inflation-rates-1913
9. Education Statistics: 18-24 with HS: https://www.census.gov/topics/education/educational-attainment.html
https://nces.ed.gov/pubs98/98018.pdf
https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_1YR_S1501&prodType=table
10. Incarceration Statistics: Bureau of Justice Statistics: https://www.bjs.gov/index.cfm?ty=nps
11. Urban Population Statistics: US Census Bureau: https://www.census.gov/data/tables.html
12. Unemployment Statistics: Bureau of Labor Statistics: https://www.bls.gov/data/#unemployment