Statistical Report on Qualified Immunity

## Setup

## Introduction and Background

Qualified immunity is a court-established doctrine that shields government officials from personal liability for constitutional violations unless the officials violated clearly established laws (<https://www.lawfareblog.com/what-qualified-immunity-and-what-does-it-have-do-police-reform>). After the killing of George Floyd sparked movements against police violence around the country, many activists directed their attention towards qualified immunity as a subject of reform. Activists argued that qualified immunity prevents police officers from being held accountable for excessive uses of force within civil court. Supporters of qualified immunity argued that efforts to limit qualified immunity would prevent police officers from effectively performing their job out of fear of frivolous lawsuits. In this statistical report, we aim to provide preliminary data-driven insights on the effects of recently passed qualified immunity reform on violent and property crime rates in major urban jurisdictions.

Four states have passed measures to limit qualified immunity: Colorado, Connecticut, New Mexico, and New York. Of those states, Colorado passed its reform the earliest, with its measure taking effect June 19, 2020. As a result, we decided to analyze Colorado crime data to determine the effects of qualified immunity on crime rates. In particular, our research question was as follows: Was the passage of qualified immunity reform in Colorado in June of 2020 correlated with significantly larger proportional increases in average daily violent and property crime incidents compared to increases in control jurisdictions? Because statewide data for 2020 and 2021 YTD from Colorado was not publicly available, we further narrowed the scope of our analysis to Denver, Colorado Springs, and Aurora, the three largest jurisdictions within Colorado.

Although we attempt to establish some level of causation in this study through the use of a synthetic control method, we lack the volume of observational data needed to successfully establish causation. In particular, we are missing observations on several key lurking variables, including the effects of COVID on poverty rates in each jurisdiction, 2020 and 2021 census data, community attitudes towards policing as a result of the George Floyd protests, amidst several other control variables. Much of this data will only be released a few years from now, severely limiting the contours of the present analysis. However, due to the prescience of the qualified immunity question and the need for data within the debate, we decided to produce this preliminary report to at least illustrate the plausible effects of qualified immunity on crime rates in Colorado. None of the findings in this report should be interpreted as demonstrating a conclusive causal relationship between qualified immunity reform and crime.

## Overview of Available Data

Many of our Methodology choices can only be understood in the context of the available data and policy context at our disposal.

We obtained three different types of data from three sources, almost entirely through data released from official sources (with the exception of land area data, which was obtained from a website reporting census data but may have errors). First, we received socioeconomic indicator data from the 2011-2019 American Community Survey estimates as found through the census data website. The predictor data we utilized was organized “by Place,” meaning the data was largely aggregated in terms of local unit boundaries (towns, cities, CDPS, etc). This data was collected with the intention of serving as crime predictor data, a use that will be described further in the “Methodology” section. Unfortunately, at the time of the creation of this report, neither 2020 nor 2021 census data had been publicly released. Second, we collected crime rate data by state by city through the UCR Crime in the United States Reports released by the FBI. This data was complete from 2011-2020, and the first 3 quarters of both 2020 and 2021 had been released in the Quarterly UCR Report from roughly 155 large agencies (limited by the number of agencies that reported their crime rates). This data included statistics on jurisdiction population, violent crime numbers, property crime numbers, and numbers for individual crimes (such as forcible rape, nonnegligent homicide/murder, larceny, etc). The quarterly data was unfortunately not disaggregated by quarter, which presented problems later down the line in Methodology A, as the 2020 data incorporated 2 quarters without qualified immunity legislation effect in Denver and 1 quarter after the passage of the qualified immunity legislation. Third, we received incident level crime data by downloading the data from various agency websites and submitting FOIA requests for agencies that had not released their data (such as Champaign Police Department). Several times, these FOIA requests returned data that was unfeasible to use (such as PDF reports of individual crimes), were too costly (totaling greater than $100 for smaller agencies), or were flatly denied on the basis that data was not kept or that state FOIA laws only permitted in-state residents to make FOIA requests (in the case of Clarksville). As a result, the usefulness of FOIA requested data was incredibly limited; however, we incorporated the data that we could obtain using this method into the analysis. The incident-level crime data gave us daily incident crime numbers for various offenses. We categorized incidents into violent and property crimes based on the UCR definitions (Footnote).

In sum, the data that we could obtain was extremely limited in scope, largely due to the combination of limited fiscal/temporal resources, difficulty in obtaining data from police departments, and late releases of census and UCR data. These data limitations then significantly harmed the soundness of the analysis and forced the methodology to deviate significantly from its ideal form.

## Methodology

In this study, we employed two different methodologies, one to increase the accuracy of the treatment date and the other to increase the total completeness of the data. We have primarily incorporated findings from Methodology A, but findings from Methodology B (performed months before Methodology A was performed) are included in the appendices of this report if they are found to be useful.

In both methodologies, we employed, to varying extents, a synthetic control methodology as described by Abadie in his article “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects.” In a synthetic control method, researchers create a weighted average of jurisdictions and data points with the goal of minimizing the distance between the weighted average and the true jurisdiction’s pre-treatment predictor and response values. The synthetic control method has the advantage of systematically identifying the strongest control jurisdictions based on predictor numbers over time, as opposed to manually identifying a geographically close jurisdiction and claiming the existence of sufficient connection to isolate the effects of the policy intervention, thus removing the effects of researcher bias from the analysis. Generally, the synthetic control method is then followed by a series of placebo synthetic control constructions to determine if the post-treatment to pre-treatment MSPE (root mean squared prediction error) ratio for the treated jurisdiction is extreme compared to those of placebo synthetic controls (<https://mixtape.scunning.com/synthetic-control.html>). If employed successfully, the synthetic control method can successfully prove the existence of a causal connection between different variables.

### Sampling Methodology

Using data provided by the FBI UCR reports, we created a dataset of about 100 jurisdictions and tracked their violent and property crime numbers from 2011 to the first 3 quarters of 2021. Jurisdictions were chosen on the basis of two criteria: First, the jurisdictions’ violent and property crime numbers must have been published every year (from 2011 to 2021) by the UCR, including in the UCR’s 2021 quarterly crime report. Because the UCR’s 2021 quarterly crime report only received numbers from less than 50% of agencies and only published figures from jurisdictions greater than 100,000 in population, the sample of jurisdictions is influenced by some self-selection sampling bias; those jurisdictions that chose not to report figures for one of the years would be automatically excluded. Second, jurisdictions must have been greater than 85K in population according to UCR estimates for every year from 2011 to 2020 (no population data for 2021). This measure is meant to exclude excessively small jurisdictions at the beginning time period that experienced extreme population growth. The number “85,000” was chosen, in part, as an arbitrary value for a jurisdiction that could experience average population growth from 2011 to reach at least 100,000 by 2020. The second selection criterion suffers from some arbitrariness; however, the criterion only excludes 4 jurisdictions from the analysis, each of which were likely too small to provide significant insight.

After compiling the database of crime numbers, we then compiled a set of yearly socioeconomic indicators from each jurisdiction to serve as predictor values for the ‘Synth’ package to average when creating a synthetic control. In particular, we chose indicators on the basis of sociological evidence that such indicators could serve as moderately strong predictors of metropolitan violent or property crime rates. Based on the results found in Wells and Weishelt’s “Explaining Crime in Metropolitan and Non-Metropolitan Counties,” we chose to record jurisdictions’ single female-led household percentage, high school education percentage, residential stability (or percent of people living in the same property that they lived in 1 year ago), percentage of population over 18, percentage of population who is white, percentage of population who is self-employed, unemployment rate, median income, child poverty rate, and population density (or population divided by land area). Certain variables recorded in the Wells and Weishelt study were excluded from our analysis because they were either found to be largely non-significant for metropolitan counties in the study (such as South vs. non-South or owner-occupied housing) or were difficult to collect (such as % voting in last election). These indicators were collected by jurisdiction for 2011 to 2019 from the census tables. If a jurisdiction was missing data from any of those years on any of the collected variables, the jurisdiction was excluded from the analysis.

Certain errors occurred when combining census data with UCR data, particularly around the naming schemes of the cities. While the UCR names cities by their given names, the census data often adds addendum names such as “CDP,” “city,” “town,” and others. We attempted to correct for these errors by erasing addendum words from census names (for instance, removing " City" from names as in the case of “Boise City, Idaho” or “Houston City, Texas”). For large jurisdictions (usually above 100000 in population), we further went back and individually corrected names to match. We believe we caught most of these errors, but inevitably some errors slipped through the cracks, leading to randomly lost data. Regardless, we find it unlikely that these random errors significantly hindered our analysis.

Because we lacked predictor data for 2020 and 2021, we extrapolated predictor data from 2019 to 2020 and 2021. In other words, 2020 and 2021 predictor data (outside of population and population density) were equivalent to 2019 data. Additionally, 2021 population and population density were extrapolated from 2020 population figures. We do not argue that this extrapolation is a fair representation of reality; of course, with COVID and the George Floyd protests of 2020, socioeconomic indicators in 2020 will be far different from those in 2019. Extrapolating will almost certainly skew our pre-treatment predictor averages to some extent; however, we do not think it will invalidate our results entirely. We discuss more in the “Methodology A” section what the implications of this choice will be on our results.

In the end, we had a dataset of 96 jurisdictions, with crime data from 2011 to 2021 and predictor data from 2011 to 2019. In total, our dataset had 1056 observations and 24 variables. In Table 1, we display the first 20 rows of our dataset.

## state city population  
## 1 Texas Abilene 119526  
## 2 Texas Abilene 119886  
## 3 Texas Abilene 119401  
## 4 Texas Abilene 120686  
## 5 Texas Abilene 121764  
## 6 Texas Abilene 122523  
## 7 Texas Abilene 122480  
## 8 Texas Abilene 122480  
## 9 Texas Abilene 123665  
## 10 Texas Abilene 124061  
## 11 Texas Abilene 124061  
## 12 Virginia Alexandria 141638  
## 13 Virginia Alexandria 145892  
## 14 Virginia Alexandria 148519  
## 15 Virginia Alexandria 151065  
## 16 Virginia Alexandria 152710  
## 17 Virginia Alexandria 155319  
## 18 Virginia Alexandria 162588  
## 19 Virginia Alexandria 162588  
## 20 Virginia Alexandria 162258

We decided to split our analysis into violent and property crime analysis for similar reasons as Wells and Weishelt did in their study. There is no evidence that violent and property crime trends are parallel, and ordinarily, property crime numbers would constitute 90% of the total crime rate. Additionally, because the UCR primarily reports property and violent crime numbers, property and violent crime numbers would already be standardized before we began our wrangling process.

### Methodology A: Approximating the Ideal Synthetic Control Methodology

In our first methodology, we analyzed 3 treated jurisdictions (Denver, Colorado Springs, and Aurora) and included roughly 91 control jurisdictions. The treated jurisdictions were chosen on the basis that they were the 3 largest cities within Colorado. We excluded cities in Connecticut, as they passed their own version of the qualified immunity bill around the same time as Colorado, and their law took effect in June of 2021, skewing the 2021 results. We did not need to further exclude New Mexico and New York City, as such jurisdictions were missing data and did not appear in our final dataset, regardless.

Using the Synth package, we created synthetic controls of each of the 3 treated jurisdictions for both violent and property crime rates. We tested a series of synthetic controls to determine the helpfulness of particular predictor variables in the analysis, but we ended up keeping all predictor variables that we collected data on to preserve methodological standardization. Regardless, if our results from removing predictors to improve the synthetic controls are of interest, we have included the results in Appendix 1.

When creating synthetic controls, we included all pretreatment time periods but optimized over 2012 to 2019, allowing the “Synth” function to automatically calculate the pre-treatment MSPE over those 8 years. We specified the pretreatment time period (time.predictors.prior) to be 2011 to 2020. Although 2020 was the year that the qualified immunity law was passed in Colorado, the function we used to calculate MSPE ratios was the “generate.placebos” function from the SCtools package, which included the final pretreatment year and the posttreatment years in calculating the post-treatment MSPE. Thus, although the pretreatment time period was specified to be 2011 to 2020, for functional purposes, 2012 to 2019 were the years relevant to the pre-treatment MSPE calculation, and 2020-2021 were the years relevant to the post-treatment MSPE calculation. Additionally, we specified for the package to employ every available optimization method and choose the best-performing method. We ended by creating 6 different synthetic controls, two for each Colorado jurisdiction, and within each jurisdiction, one for violent crime rates and one for property crime rates.

To determine the significance of our findings, we calculated the “MSPE (mean squared prediction error) ratio” for each of the synthetic controls. In other words, we averaged the squared amount that the synthetic violent or property crime rates differed from the observed violent or property crime rates over the optimized pretreatment time period, given by the Synth function as the loss.v value. We then averaged the squared amount that the synthetic violent or property crime rates differed from the observed violent or property crime rates over the post-treatment time period. To control for jurisdictions where the synthetic model was not a great fit to begin with, we then divided the post-treatment MSPE by the pre-treatment MSPE, creating an MSPE ratio. Theoretically, if the intervention had a significant effect on the property or violent crime rates in the treated jurisdictions, we should see significant increases in crime rates in 2020 and 2021 exceeding those of the synthetic control, and thus, the MSPE ratio of those jurisdictions should be high. However, because there is no objective metric for what a “high enough” MSPE ratio is, we further created placebo synthetic controls for every control jurisdiction in the dataset and calculated MSPE ratios for each placebo synthetic control. (FOOTNOTE) If the MSPE ratio of the treated jurisdiction was greater than 95% of placebo MSPE ratios, we concluded that the MSPE ratio of the treated jurisdiction was high enough to be statistically significant. The interpretations of such a conclusion will be further discussed in the “Results” section.

The synthetic control methodology that we employed suffered from several limitations. First, as noted in the “Sampling Methodology” section, we extrapolated predictor values from 2019 to 2020 and 2021. This biased the averages used when constructing the synthetic control. Since 2020 was included in the pretreatment time section, we functionally doubled the role of 2019 in calculating predictor averages for the treated jurisdiction for the synthetic jurisdiction to emulate. We do not believe this should, alone, invalidate our analysis however. Since the 2020 predictors data is only used in calculating an overall average of the predictors that the synthetic jurisdiction should approximate - and not to serve as predictors that should be held constant from year to year to isolate the effects of the intervention - the extrapolated 2020 data would only cause the synthetic control methodology to create weighted averages that matched treated jurisdictions’ 2019 data above other older years. For example, if researchers attempted to control the 2020 and 2021 MSPE for the predictor variables using the 2019 data, such an effort would clearly be invalid, as 2019 unemployment and child poverty rates clearly cannot be used to adjust for 2020 and 2021 data. However, because we employ a synthetic control methodology and do not calculate MSPEs differently based on predictor values, we do not suffer from such limitations. The methodology merely averages the predictor values of the treated jurisdiction over the pretreatment time period for the synthetic control to match, but does not attempt to hold such predictors constant from year to year or control for yearly shifts in those predictors. Thus, the skew created by such a flaw is minimal.

Second, the time of the treatment is not effectively accounted for by the synthetic control methodology. The passage of the qualified immunity bill in Colorado occurred in the middle of 2020; however, we do not have quarterly data by which we could isolate the 2 quarters of 2020 without the influence of the policy intervention from the 2 quarters influenced by the intervention. Instead, we simply sort 2020 and 2021 as broadly falling under the post-treatment time frame, operating on the assumption that if the qualified immunity legislation affected violent and property crime rates in the Colorado jurisdictions, the increase in violent and property crime rates for the whole of 2020 would be greater than those of treated jurisdictions. Unfortunately, such an assumption is not necessarily true, as 2020 introduced a series of different factors, ranging from COVID to the George Floyd protests, each of which influenced jurisdictions’ crime rates in unknown ways. As a result, we are hesitant to derive a causal conclusion from any of our analysis; our findings could very well be interpreted in several different lights, including that COVID impacted Colorado jurisdictions’ crime rates harder than other jurisdictions, or that police agencies in Colorado passed 2020 policies different from those of other jurisdictions. We attempt to solve this problem in Methodology B at the cost of other significant methodological limitations.

Third, in an ideal synthetic control, we would have a wealth of years both before and after the treatment to evaluate. Unfortunately, due to the recency of the legislation and the inability to divide years into quarterly data, we only had a total of 2 post-treatment time periods to evaluate. This may limit our insights, as a single year of increased property or violent crime rates in one of the treated jurisdictions would skew the mean post-treatment MSPE substantially, even if such a year occurred merely from chance. Placebo testing should be able to diminish the influence of chance in the analysis, but having more post-treatment time periods to calculate the MSPE would allow the analysis to be more reliable.

### Methodology B: Synthetic Control as Comparative Case Study Selection

Our second methodology was employed before the release of 2020 and 2021 data by the UCR and was meant to serve as a workaround to normal synthetic analysis. The second methodology was also performed in August of 2021 with only a few updates since then, limiting the scope of the data. As a result, the second methodology suffers from severe limitations, many of which could invalidate the analysis entirely. Regardless, we include the results from Methodology B in the appendices in case they are found to be useful in their treatment date precision and high jurisdiction inclusion that Methodology A lacks.

In our second methodology, we examined violent and property crime rates in two treated jurisdictions: Denver and Colorado Springs. In selecting possible control jurisdictions, we waived the requirements for 2020 and 2021 crime data, as such data was not relevant for the analysis. We additionally only filtered for jurisdictions greater than 50000 in population, as we only had access to a small number of time periods but an enormous sample of jurisdictions within the donor pool. We removed population density from the analysis and relied on population alone to serve as the “population” level statistic. This led to many nonsimilar jurisdictions being included in the analysis, significantly increasing the potential for bias. In total, we had roughly 530 jurisdictions in the donor pool when constructing the synthetic control.

To account for the lack of 2020 and 2021 date and to increase the precision of the treatment dates, we used the synthetic control methodology to identify similar jurisdictions to Denver or Colorado Springs and to provide weights for some of those jurisdictions. We optimized the synthetic controls for 2016 to 2019 to obtain a synthetic control that could follow the most recent trends in both Denver and colorado Springs. We then identified the top 5 jurisdictions with the highest weights and reran the synthetic control model with only those jurisdictions to recalculate the weights, relying on the premise that the synthetic of the top 4 or 5 jurisdictions that comprise a majority weight in the full synthetic control would be similar enough to the treated jurisdictions to analyze. With those identified control jurisdictions, we submitted requests for incident-level crime data from those departments for 2019-2021. When those requests were either unanswered or denied (as in the case of Ann Arbor Police Department), we removed the city from the synthetic control model and reran the model until we obtained at least three police departments with accessible incident level crime data.

We subdivided all 2019-2021 incident-level crime data into property and violent crimes based on UCR definitions. In particular, murder and nonnegligent homicide, aggravated assault, robbery, and forcible rape (including sexual assault with an object, fondling, and forcible sodomy) were identified as violent crimes. We categorized larceny charges, burglary, damage/destruction of property, arson, shoplifting, pocket-picking, and motor vehicle theft charges as property crimes. We calculated the daily numbers of violent and property offenses for June of 2019 to June of 2020 (before qualified immunity reform) and June of 2020 to June of 2021 (after qualified immunity reform) in control and treated jurisdictions. We then subtracted the daily numbers of violent and property offenses in the 2019-20 time period from the 2020-21 time period and divided by the total number of violent or property offenses in the 2019-20 time period to make the daily numbers of violent and property crimes proportionate to each jurisdiction’s respective crime numbers. Finally, we created a bootstrapped null distribution assuming no true difference between the daily increases of the synthetic jurisdiction compared to Denver or Colorado Springs and calculated a p-value based on the probability of observing the real difference or greater between Denver/Colorado Springs and the respective synthetic control difference based on the null distribution.

This methodology had a few critical limitations. First, because of the lack of 2020 and 2021 census data, we could only identify jurisdictions similar to either Denver or Colorado Springs until 2019 which is no guarantee that those similarities in control predictors continued until 2020 and 2021. In addition, since we did not have access to UCR data for 2020 and 2021, we had to use 2011-2019 weights in 2020 and 2021 calculations, which extrapolate beyond the capabilities of the synthetic control. Second, because of the several denied requests (especially from Ann Arbor and Clarksville Police Departments), we were forced to rely partially on convenience sampling in order to successfully carry out the study. Despite this, the plots comparing the synthetic control property/violent crime rates with the Denver property/violent crime rates look strong enough to indicate that the synthetic controls at least partially match the crime trends of the treated jurisdictions even when certain jurisdictions were removed due to lack of data. Third, because we needed to determine if the increases between the 2019-20 time period and 2020-21 time period were significantly greater than increases in control jurisdictions, we were forced to employ a test where we subtracted daily crimes in one time period from daily crimes in another time period. This method may have substantially exaggerated the standard deviation of violent and property crimes, since daily fluctuations in crime do not remain constant over the course of a year. The test may have been more successful on a monthly level, but we did not have enough monthly difference data to successfully arrive at statistical conclusions through simulation. Fourth, because we had to standardize the daily crime numbers by dividing crime numbers from some relative figure for each jurisdiction (in this case, the total number of offenses in the 2019-20 time period), smaller jurisdictions disproportionately influenced the variance of the synthetic control, since daily fluctuations of 1-2 offenses were much greater when standardized compared to larger jurisdictions.

The results of this methodology, in their entirety, will be described in Appendix 2.

## Exploratory Data Analysis

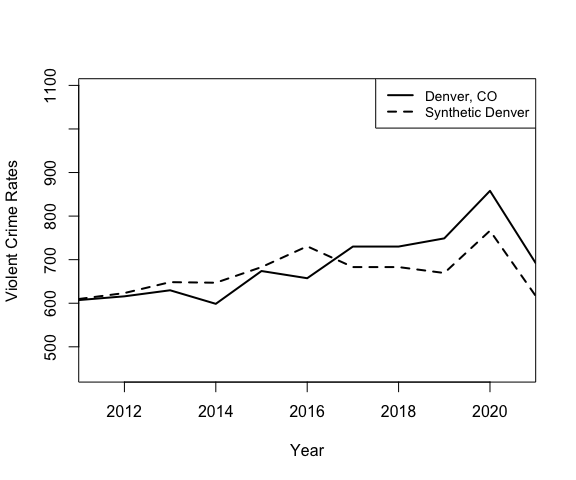
## Testing and Results

In this section, we will discuss the synthetic control diagnostics and MSPE test results for each of the 6 synthetic controls. Note that both population and population density were included as predictors, as both of them were found to be important by model selection.

### Synthetic Controls for Violent Crime

#### Denver:

We began with Denver, the largest jurisdiction in Colorado. Below, we included the graph comparing the violent crime rates of the synthetic jurisdiction and Denver itself.



The synthetic control is relatively strong with a pre-treatment MSPE value of synth\_denver\_violent$loss.v, as computed by the Synth package over the years of 2012-2019. As expected, synthetic Denver roughly follows the trends of Denver up until about 2016. In 2017, observed Denver’s violent crime rate increases as the synthetic’s decreases, causing observed Denver’s violent crime rates to be greater than those of Synthetic Denver for the next few years up until 2021. The trends on the graph are not perfectly parallel prior to 2016. Past 2016, however, Denver’s observed violent crime rate is parallel to the synthetic control’s violent crime rate.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Treated | Synthetic | Sample Mean |
| population | 684330.80 | 683871.385 | 307773.863 |
| pop\_density | 4472.75 | 4491.511 | 3771.049 |
| median\_income | 56967.30 | 59164.227 | 55965.052 |
| hs | 86.31 | 86.429 | 86.508 |
| res\_stability | 78.24 | 78.249 | 80.471 |
| over\_18 | 79.27 | 79.183 | 76.208 |
| white\_percent | 53.20 | 53.886 | 53.477 |
| self\_employed | 5.64 | 5.645 | 5.272 |
| unemployment | 6.12 | 6.151 | 7.475 |
| female\_household | 9.01 | 9.212 | 12.600 |
| owner\_occupied | 50.03 | 46.609 | 53.963 |
| child\_poverty | 24.34 | 22.035 | 23.527 |

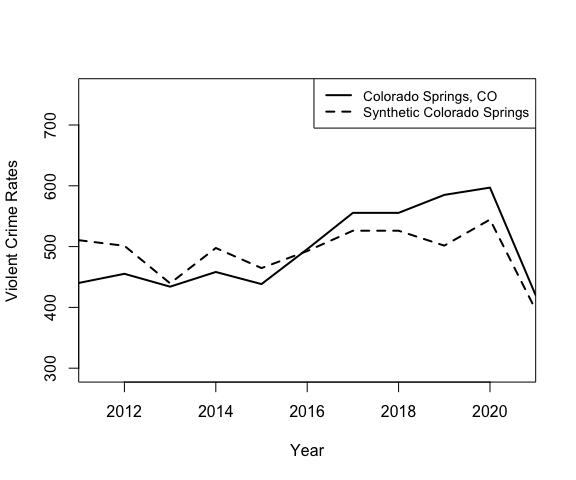
|  |  |  |
| --- | --- | --- |
|  | w.weights | unit.names |
| 79 | 0.334 | Salt Lake City, Utah |
| 83 | 0.310 | Seattle, Washington |
| 23 | 0.162 | Dallas, Texas |
| 40 | 0.073 | Houston, Texas |
| 42 | 0.047 | Irving, Texas |
| 3 | 0.020 | Ann Arbor, Michigan |
| 8 | 0.017 | Bellevue, Washington |
| 72 | 0.008 | Pearland, Texas |
| 49 | 0.005 | Laredo, Texas |
| 30 | 0.003 | Escondido, California |
| 6 | 0.002 | Austin, Texas |
| 71 | 0.002 | Pasadena, Texas |
| 11 | 0.001 | Brownsville, Texas |
| 36 | 0.001 | Garland, Texas |
| 59 | 0.001 | McAllen, Texas |
| 67 | 0.001 | Oceanside, California |
| 69 | 0.001 | Oklahoma City, Oklahoma |
| 81 | 0.001 | San Diego, California |
| 87 | 0.001 | Sterling Heights, Michigan |
| 1 | 0.000 | Abilene, Texas |

Table 2 indicates that the synthetic control matches the predictor values of observed Denver almost perfectly, even in terms of population and population density despite the sample means significantly deviating from those of Denver. When examining the weights (as found in Table 3), we can see that most weights are around 0, a sign of a strong synthetic control fit. We also see that Seattle, Salt Lake City, and Dallas comprise almost 80% of the model alone.

## [1] 0.4594595

After calculating over 90 different placebo synthetic controls, we found that Denver’s MSPE ratio was nonsignificant at the 1%, 5%, or 10% level. In particular, we found that when excluding jurisdictions with pre-treatment MSPEs more than 5x greater than the pre-treatment MSPE of Denver, more than 45% of the placebo synthetic controls had MSPE ratios greater than that of Denver. Thus, the data does not provide sufficient evidence to indicate that Denver’s violent crime rate in 2020 or 2021 could not have resulted purely from chance.

#### Colorado Springs



|  |  |  |  |
| --- | --- | --- | --- |
|  | Treated | Synthetic | Sample Mean |
| population | 456052.600 | 451903.638 | 307773.863 |
| pop\_density | 2338.731 | 2376.101 | 3771.049 |
| median\_income | 57594.800 | 57686.668 | 55965.052 |
| hs | 93.160 | 90.809 | 86.508 |
| res\_stability | 76.380 | 76.468 | 80.471 |
| over\_18 | 75.880 | 75.878 | 76.208 |
| white\_percent | 69.540 | 58.832 | 53.477 |
| self\_employed | 5.550 | 5.472 | 5.272 |
| unemployment | 7.790 | 7.708 | 7.475 |
| female\_household | 10.310 | 11.606 | 12.600 |
| owner\_occupied | 59.040 | 55.500 | 53.963 |
| child\_poverty | 17.520 | 19.227 | 23.527 |

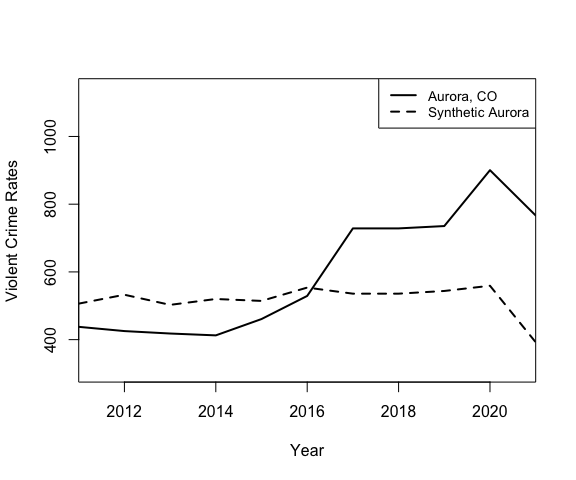
|  |  |  |
| --- | --- | --- |
|  | w.weights | unit.names |
| 18 | 0.416 | Clarksville, Tennessee |
| 73 | 0.174 | Peoria, Arizona |
| 3 | 0.165 | Ann Arbor, Michigan |
| 40 | 0.139 | Houston, Texas |
| 14 | 0.063 | Carlsbad, California |
| 35 | 0.007 | Frisco, Texas |
| 39 | 0.005 | Henderson, Nevada |
| 9 | 0.004 | Boise, Idaho |
| 50 | 0.004 | Las Vegas, Nevada |
| 62 | 0.002 | Mesa, Arizona |
| 19 | 0.001 | College Station, Texas |
| 36 | 0.001 | Garland, Texas |
| 52 | 0.001 | Lee’s Summit, Missouri |
| 56 | 0.001 | Lubbock, Texas |
| 65 | 0.001 | Murfreesboro, Tennessee |
| 67 | 0.001 | Oceanside, California |
| 69 | 0.001 | Oklahoma City, Oklahoma |
| 70 | 0.001 | Olathe, Kansas |
| 77 | 0.001 | Round Rock, Texas |
| 80 | 0.001 | San Antonio, Texas |

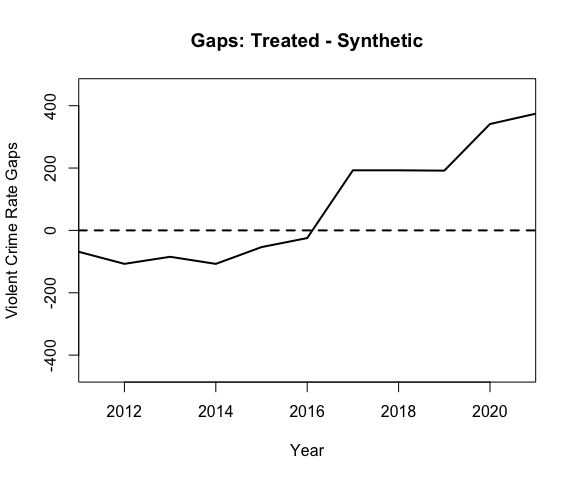
The synthetic control for Colorado Springs is very strong, with a pre-treatment MSPE of 1636.7726543, lower than that of Denver. Similar to Denver, per table x, Colorado Springs’ predictor values match exceptionally well with those of the synthetic jurisdiction, with the sole exception of racial homogeneity which is about 10% off. Nevertheless, given that the synthetic control tracks Colorado Springs relatively thoroughly, we find it appropriate to proceed with the given model.

## [1] 0.7042254

Using the same placebos generated for Denver, we found that Colorado Springs’ MSPE ratio of 1.019 was smaller than 70% of placebo jurisdictions after removing placebo jurisdictions with pre-treatment MSPEs 5x greater than Colorado Springs’. The data does not provide sufficient evidence to indicate that Colorado Springs’ violent crime rates in 2020 and 2021 were significantly different from those of control jurisdictions.

#### Aurora





|  |  |  |  |
| --- | --- | --- | --- |
|  | Treated | Synthetic | Sample Mean |
| population | 360080.600 | 359932.872 | 307773.863 |
| pop\_density | 2323.101 | 2419.512 | 3771.049 |
| median\_income | 56417.600 | 58072.485 | 55965.052 |
| hs | 86.620 | 86.739 | 86.508 |
| res\_stability | 79.030 | 79.063 | 80.471 |
| over\_18 | 73.610 | 73.614 | 76.208 |
| white\_percent | 46.090 | 46.184 | 53.477 |
| self\_employed | 5.000 | 5.003 | 5.272 |
| unemployment | 7.560 | 7.495 | 7.475 |
| female\_household | 12.920 | 13.025 | 12.600 |
| owner\_occupied | 58.640 | 58.412 | 53.963 |
| child\_poverty | 20.510 | 20.520 | 23.527 |

|  |  |  |
| --- | --- | --- |
|  | w.weights | unit.names |
| 18 | 0.253 | Clarksville, Tennessee |
| 72 | 0.169 | Pearland, Texas |
| 80 | 0.143 | San Antonio, Texas |
| 71 | 0.112 | Pasadena, Texas |
| 19 | 0.066 | College Station, Texas |
| 56 | 0.056 | Lubbock, Texas |
| 68 | 0.048 | Odessa, Texas |
| 15 | 0.043 | Chesapeake, Virginia |
| 26 | 0.025 | Detroit, Michigan |
| 77 | 0.010 | Round Rock, Texas |
| 1 | 0.008 | Abilene, Texas |
| 3 | 0.007 | Ann Arbor, Michigan |
| 73 | 0.007 | Peoria, Arizona |
| 28 | 0.006 | El Paso, Texas |
| 24 | 0.003 | Dayton, Ohio |
| 35 | 0.003 | Frisco, Texas |
| 21 | 0.002 | Columbia, Missouri |
| 49 | 0.002 | Laredo, Texas |
| 51 | 0.002 | League City, Texas |
| 69 | 0.002 | Oklahoma City, Oklahoma |

Despite the fact that Aurora’s synthetic control matches its predictor values extremely well, the graph of Aurora’s synthetic control makes it clear that Aurora’s synthetic control is a weak fit for the data. While Aurora experiences a large uptick in violent crime in 2016 and 2017, the synthetic control experiences no such uptick. Aurora’s pretreatment MSPE is also extremely high with a value of 1.8092381^{4}, almost 20 times that of Colorado Springs. We find it unlikely that insights derived from this synthetic control will be helpful, but we run the significance test regardless.

## [1] 0.1744186

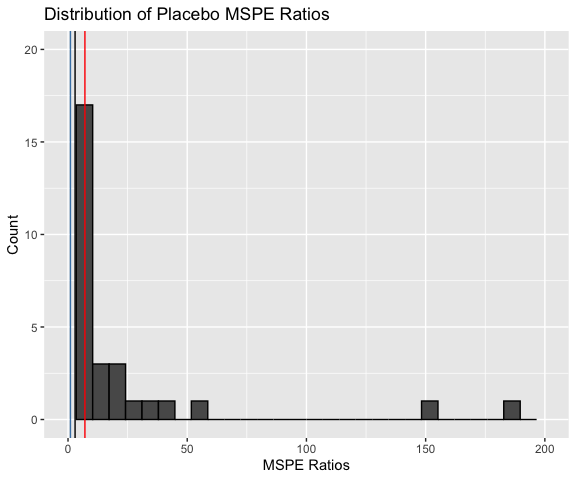
We find that when calculating the MSPE ratio of Aurora and comparing with placebos, the MSPE ratio is not significant at the 1%, 5%, or 10% level. When excluding placebos with pre-treatment MSPEs over 5x greater than that of Aurora, about 17% of the placebo synthetic controls have MSPE ratios greater than that of Aurora. It is plausible that if we further examined whether each placebo synthetic control’s large MSPE ratio was from higher than expected violent crime rates or lower than expected violent crime rates (functionally turning the test into a one-sided test), Aurora’s increase may become significant. However, given that Aurora’s synthetic control is already so weak, we do not feel it would be valuable to conduct such an analysis.

#### Placebos for Violent Crimes

As we noted earlier, we generated over 90 different placebo synthetic controls with the same settings as the original synthetic control and compiled all the MSPE ratios from each placebo synthetic control into a single dataset for significance testing. We would like to take a moment to comment on these placebos.

Below, we visualized these placebos’ MSPE ratios and noted their summary statistics.

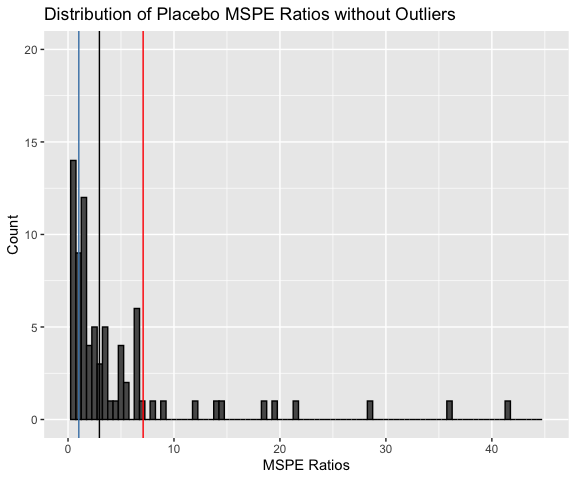
|  |  |  |
| --- | --- | --- |
| mean | sd | median |
| 8.563 | 25.842 | 1.779 |



As the histogram displays, most placebo MSPE ratios are centered at 0-10 with the exception of two placebo MSPE ratios above 150. Those MSPE ratios represent Evansville and Manchester and likely occurred from an exceptionally good fit with the data in pretreatment years with some declining fit in 2020. The outlier will be further explored in Appendix 2; however, we do not see a good reason to remove the outlier, as removing the outlier would not change the significance of any of our results.

Re-visualizing without outliers and reducing the binwidth to further detail the smaller MSPE ratios in the spectrum, we arrive at the second graph below.

|  |  |  |
| --- | --- | --- |
| mean | sd | median |
| 4.986 | 9.068 | 1.746 |



As we can see, Colorado Springs and Denver are both squarely within the center of the distribution. Aurora’s MSPE ratio is larger than most of the MSPE ratios within the dataset, but the ratio is still smaller than enough placebos to not constitute statistically significant evidence.

We display these placebos to give the reader an idea of what synthetic controls we ended up creating, any outliers or flaws within the synthetic controls, as well as where the treated jurisdictions lie on the distribution.

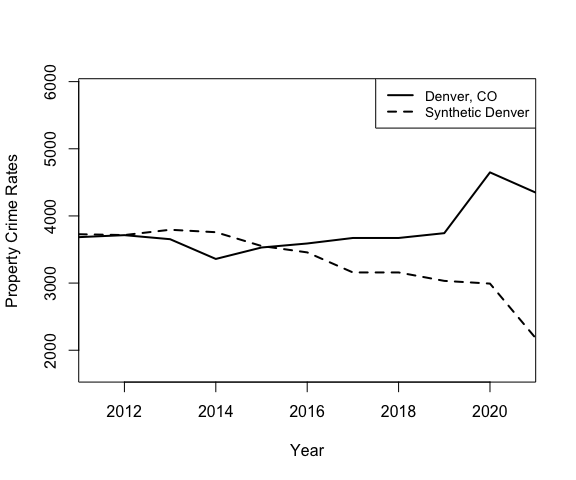
#### Overall Results for Violent Crime

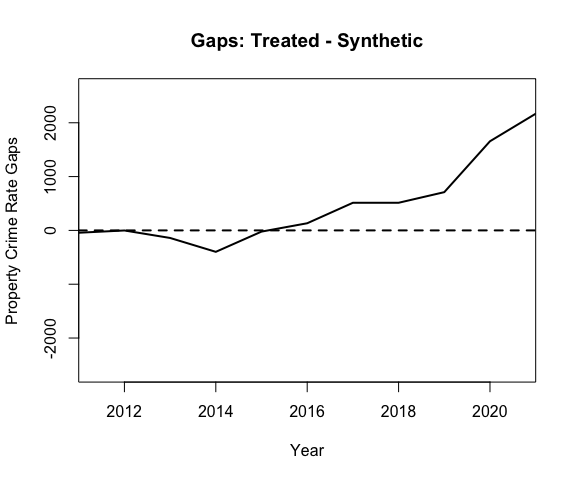
In terms of violent crime, we created three different synthetic controls, one for each of the three treated jurisdictions in Colorado. After generating over 90 placebo MSPE ratios, we found that the MSPE ratios of Denver, Colorado Springs, and Aurora were not large enough to constitute statistically significant evidence that violent crime rates in those areas were significantly different from violent crime rates in control jurisdictions post-treatment. The synthetic controls were strong fits for Denver and Colorado Springs, but the synthetic control was extremely weak in the context of Aurora, which may partially explain why Aurora’s data was the closest to being significant.

### Property Crime

Our property crime results varied significantly from our violent crime results in terms of the significance of our findings. Just like in the case of violent crime, we created three synthetic controls, one for each of three treated jurisdictions. All 3 synthetic controls were at least moderate fits with the data, with significant variation in how well the synthetic controls fit based on treated jurisdiction.

#### Denver





|  |  |  |  |
| --- | --- | --- | --- |
|  | Treated | Synthetic | Sample Mean |
| population | 684330.80 | 647536.384 | 307773.863 |
| pop\_density | 4472.75 | 4773.902 | 3771.049 |
| median\_income | 56967.30 | 67673.174 | 55965.052 |
| hs | 86.31 | 89.969 | 86.508 |
| res\_stability | 78.24 | 77.797 | 80.471 |
| over\_18 | 79.27 | 79.274 | 76.208 |
| white\_percent | 53.20 | 54.716 | 53.477 |
| self\_employed | 5.64 | 5.641 | 5.272 |
| unemployment | 6.12 | 6.128 | 7.475 |
| female\_household | 9.01 | 9.018 | 12.600 |
| owner\_occupied | 50.03 | 51.700 | 53.963 |
| child\_poverty | 24.34 | 16.904 | 23.527 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | w.weights | unit.names | unit.numbers |
| 83 | 0.257 | Seattle, Washington | 83 |
| 3 | 0.170 | Ann Arbor, Michigan | 3 |
| 40 | 0.121 | Houston, Texas | 40 |
| 72 | 0.083 | Pearland, Texas | 72 |
| 81 | 0.048 | San Diego, California | 81 |
| 35 | 0.035 | Frisco, Texas | 35 |
| 42 | 0.020 | Irving, Texas | 42 |
| 6 | 0.017 | Austin, Texas | 6 |
| 82 | 0.010 | Sandy Springs, Georgia | 82 |
| 8 | 0.009 | Bellevue, Washington | 8 |
| 79 | 0.009 | Salt Lake City, Utah | 79 |
| 51 | 0.008 | League City, Texas | 51 |
| 74 | 0.008 | Plano, Texas | 74 |
| 23 | 0.007 | Dallas, Texas | 23 |
| 70 | 0.006 | Olathe, Kansas | 70 |
| 77 | 0.006 | Round Rock, Texas | 77 |
| 30 | 0.005 | Escondido, California | 30 |
| 67 | 0.005 | Oceanside, California | 67 |
| 69 | 0.005 | Oklahoma City, Oklahoma | 69 |
| 2 | 0.004 | Alexandria, Virginia | 2 |

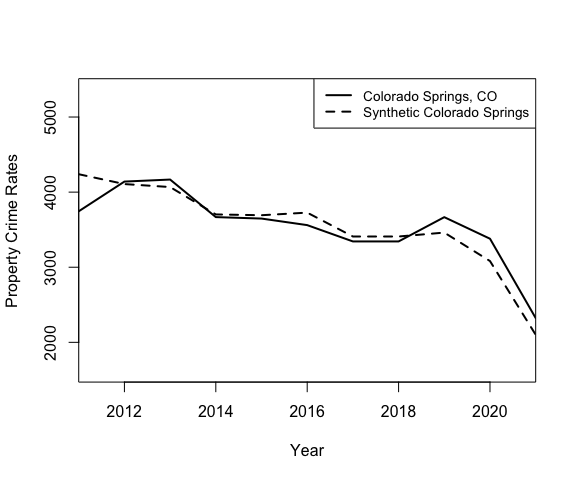
As seen in Figures \_\_\_\_ (path and gaps plots), the synthetic control is not an exceptional fit for the Denver data. In particular, while the synthetic control’s property crime rate steadily decreases, Denver’s property crime rate has the opposite trend from 2014. Notably, however, the gaps between Denver and the synthetic control’s property crime rates are not especially large until the time periods post-treatment (2020 and 2021), where Denver has an immense increase in property crime rates that is not at all followed by the synthetic control.

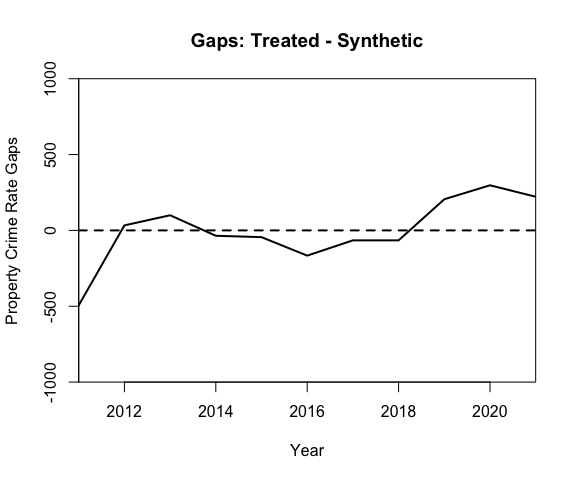
In terms of predictors, Table \_\_\_ displays that the syntheitc control does a good job of creating a synthetic jurisdictions with predicotrs that match the observed Denver well for every predictor except child poverty. In the case of child poverty, despite the predictor being one of the more significant predictors of property crime rates, the synthetic control deviates from observed Denver’s child poverty rate by almost 8%. Regardless, we believe that the synthetic control fits well enough to run a statistical significance analysis with meaningful insights. It is important to note, however, that the conclusions reached by the significance analysis are, of course, limited by the fact that the synthetic control and Denver had opposite crime trends even prior to treatment, even if such crime trends did not reach the magnitude of difference post-treatment.

## [1] 0.0125

After generating over 90 placebos, we found the MSPE ratio of Denver to be statistically significant. In particular, after discarding jurisdictions with pre-treatment MSPEs greater than 5x that of Denver’s, we found that Denver’s MSPE ratio of 24.234 was extremely high, only exceeded by the outlier jurisdiction of Madison, Wisconsin. Although not significant at the 1% level, such a finding is significant at the 5% and 10% levels. The data does provide sufficient evidence to indicate that Denver’s property crime rate in 2020 and 2021 is significantly greater than those of control jurisdictions and makes it unlikely that Denver’s heightened property crime rate merely resulted from chance.

#### Colorado Springs





|  |  |  |  |
| --- | --- | --- | --- |
|  | Treated | Synthetic | Sample Mean |
| population | 456052.600 | 446271.616 | 307773.863 |
| pop\_density | 2338.731 | 2647.140 | 3771.049 |
| median\_income | 57594.800 | 57295.690 | 55965.052 |
| hs | 93.160 | 90.656 | 86.508 |
| res\_stability | 76.380 | 77.000 | 80.471 |
| over\_18 | 75.880 | 76.050 | 76.208 |
| white\_percent | 69.540 | 65.919 | 53.477 |
| self\_employed | 5.550 | 5.575 | 5.272 |
| unemployment | 7.790 | 7.354 | 7.475 |
| female\_household | 10.310 | 10.765 | 12.600 |
| owner\_occupied | 59.040 | 54.581 | 53.963 |
| child\_poverty | 17.520 | 20.780 | 23.527 |

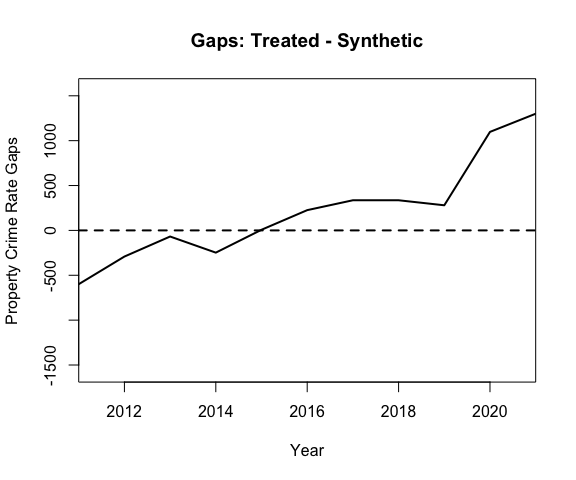
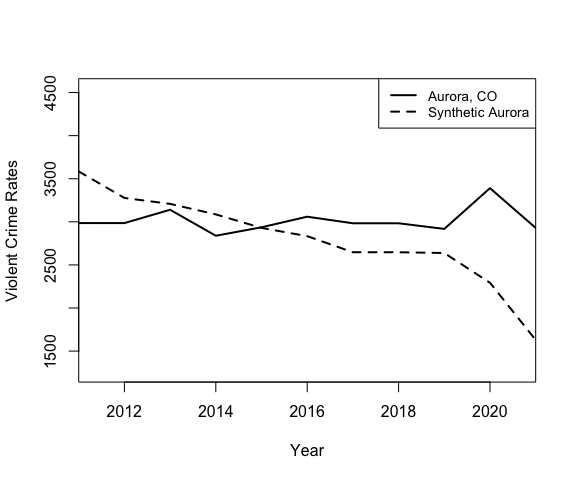
|  |  |  |  |
| --- | --- | --- | --- |
|  | w.weights | unit.names | unit.numbers |
| 18 | 0.249 | Clarksville, Tennessee | 18 |
| 85 | 0.216 | Springfield, Missouri | 85 |
| 9 | 0.160 | Boise, Idaho | 9 |
| 35 | 0.143 | Frisco, Texas | 35 |
| 40 | 0.085 | Houston, Texas | 40 |
| 81 | 0.048 | San Diego, California | 81 |
| 19 | 0.045 | College Station, Texas | 19 |
| 39 | 0.027 | Henderson, Nevada | 39 |
| 50 | 0.026 | Las Vegas, Nevada | 50 |
| 1 | 0.000 | Abilene, Texas | 1 |
| 2 | 0.000 | Alexandria, Virginia | 2 |
| 3 | 0.000 | Ann Arbor, Michigan | 3 |
| 4 | 0.000 | Arlington, Texas | 4 |
| 6 | 0.000 | Austin, Texas | 6 |
| 7 | 0.000 | Beaumont, Texas | 7 |
| 8 | 0.000 | Bellevue, Washington | 8 |
| 11 | 0.000 | Brownsville, Texas | 11 |
| 12 | 0.000 | Buffalo, New York | 12 |
| 13 | 0.000 | Cambridge, Massachusetts | 13 |
| 14 | 0.000 | Carlsbad, California | 14 |

Unlike in the case of Denver, Colorado Springs’ synthetic control is an exceptionally good fit to the data. With a pretreatment MSPE of only 1.1586726^{4} (note that because we are using property crime rates, numbers are expected to be much higher than in the case of violent crime rates), the synthetic control follows Colorado Springs’ trend exceptionally well until about 2019. From 2019-2021, Colorado Springs had a slightly higher property crime rate than its synthetic control.

## [1] 0.212766

Using significance testing, we found that Colorado Springs’ MSPE ratio of about 7.7, although higher than the average, was not statistically significant at any of the 3 levels after removing jurisdictions with pretreatment MSPEs more than 5x greater than that of Colorado Springs. Roughly 21% of placebo jurisdictions had MSPE ratios greater than the one in Colorado Springs. Thus, the data does not provide sufficient evidence to indicate that Colorado Springs’ different property crime rates from control jurisdictions could not have resulted simply from chance.

#### Aurora



|  |  |  |  |
| --- | --- | --- | --- |
|  | Treated | Synthetic | Sample Mean |
| population | 360080.600 | 349718.242 | 307773.863 |
| pop\_density | 2323.101 | 2364.647 | 3771.049 |
| median\_income | 56417.600 | 56716.680 | 55965.052 |
| hs | 86.620 | 87.554 | 86.508 |
| res\_stability | 79.030 | 79.036 | 80.471 |
| over\_18 | 73.610 | 73.614 | 76.208 |
| white\_percent | 46.090 | 48.233 | 53.477 |
| self\_employed | 5.000 | 5.126 | 5.272 |
| unemployment | 7.560 | 7.553 | 7.475 |
| female\_household | 12.920 | 12.987 | 12.600 |
| owner\_occupied | 58.640 | 57.123 | 53.963 |
| child\_poverty | 20.510 | 20.524 | 23.527 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | w.weights | unit.names | unit.numbers |
| 18 | 0.322 | Clarksville, Tennessee | 18 |
| 37 | 0.126 | Grand Prairie, Texas | 37 |
| 80 | 0.112 | San Antonio, Texas | 80 |
| 73 | 0.063 | Peoria, Arizona | 73 |
| 19 | 0.061 | College Station, Texas | 19 |
| 72 | 0.053 | Pearland, Texas | 72 |
| 35 | 0.034 | Frisco, Texas | 35 |
| 71 | 0.025 | Pasadena, Texas | 71 |
| 28 | 0.016 | El Paso, Texas | 28 |
| 42 | 0.014 | Irving, Texas | 42 |
| 77 | 0.013 | Round Rock, Texas | 77 |
| 68 | 0.009 | Odessa, Texas | 68 |
| 40 | 0.008 | Houston, Texas | 40 |
| 16 | 0.007 | Chula Vista, California | 16 |
| 75 | 0.005 | Reno, Nevada | 75 |
| 21 | 0.004 | Columbia, Missouri | 21 |
| 22 | 0.004 | Corpus Christi, Texas | 22 |
| 56 | 0.004 | Lubbock, Texas | 56 |
| 69 | 0.004 | Oklahoma City, Oklahoma | 69 |
| 1 | 0.003 | Abilene, Texas | 1 |

Unlike in the case of the Aurora violent crime synthetic control, the Aurora property crime synthetic control is a moderate fit to the data. Per figure \_\_\_\_, the gaps between Aurora’s property crime rate and synthetic Aurora’s property crime rates are less than 500 until the treatment year. Similar to Denver’s property crime synthetic control, Aurora’s proeprty crime synthetic control suffers from opposite trends; while the synthetic control’s property crime rates are steadily decreasing every year from 2011, Aurora’s property crime rates remain steady until its increase in 2020. However, Aurora’s predictors are well-matched by the synthetic control, and Aurora’s pretreatment MSPE of 6.3259631^{4} is substantially lower than Denver’s property crime pretreatment MSPE. Overall, the fit with the data is strong enough to derive some meaningful insights.

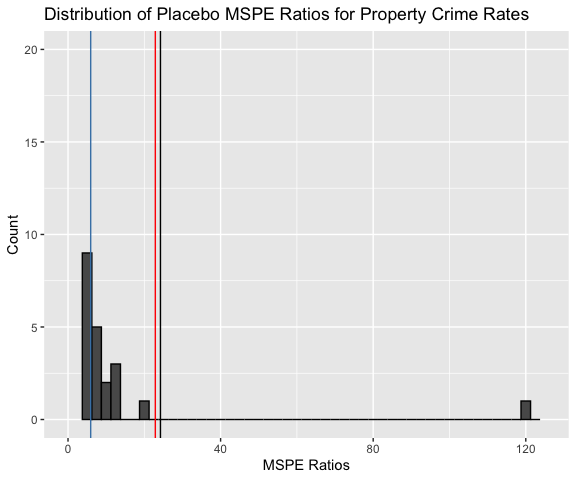
## [1] 0.01315789

After performing placebo test analysis, we find that the MSPE ratio of Aurora is statistically significant at both the 10% and 5% levels after removing jurisdictions with pretreatment MSPEs more than 5x greater than that of Aurora. In particular, Aurora’s MSPE ratio of 22.887 is surpassed by only one jurisdiction, the extreme outlier of Madison, Wisconsin. When removing the outlier, the data would be significant even at the 1% level. The data provides sufficient evidence to indicate that Aurora’s 2020 and 2021 property crime rates were significantly greater than those of similar jurisdictions, although the significance of the evidence is debatable based on the alpha level employed.

#### Visualizing the Placebos for Property Crime

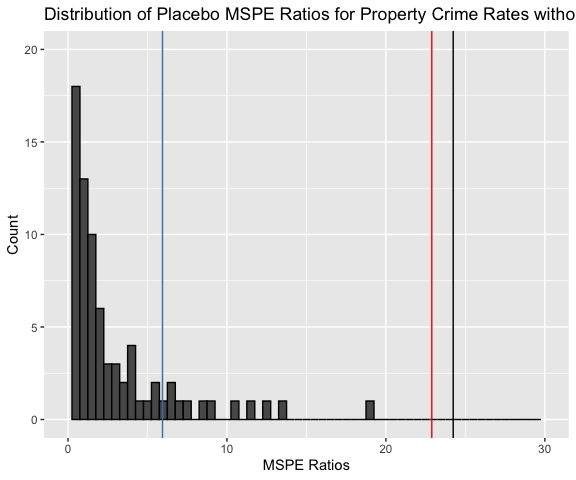
Below, we created a histogram to visualize the placebos for property crime.

|  |  |  |
| --- | --- | --- |
| mean | sd | median |
| 4.045 | 12.96 | 1.253 |



Similar to the violent crime MSPE ratios, we had one outlier at around 120, likely due to an incredibly low pretreatment MSPE coupled with some increase or decrease in property crime. In Appendix 2, we discuss this outlier more. For now, we can re-visualize the histogram without the outlier.

|  |  |  |
| --- | --- | --- |
| mean | sd | median |
| 2.763 | 4.104 | 1.24 |



As we can see on the histogram, both Denver and Aurora’s MSPE ratios are extreme compared to the placebo synthetic controls, implying that the property crime increases in both jurisdictions likely did not result purely from chance. Additionally, Colorado Springs’ property crime MSPE ratio occurs on the right end of the graph, implying that Colorado Springs’ property crime MSPE ratio is close to extreme but does not quite reach the level of statistical significance.

#### Overall Property Crime Results

After constructing over 90 different placebos and 3 synthetic controls for each of the treated jurisdictions, we found that Denver and Aurora both experienced property crime increases significantly greater than those of similar jurisdictions and that such increases likely did not result purely from chance (given that few placebos, outside of the one outlier, experienced similar increases). Such a finding will be discussed in further detail in the Discussion section. We additionally found that Colorado Springs’ MSPE ratio, although not significant, was also somewhat extreme, only surpassed by around 20% of placebo MSPE ratios. Each of the synthetic controls fit their jurisdictions at least moderately well, with significant variation in how well the controls performed. Colorado Springs possessed the strongest synthetic control by far, while both Denver and Aurora’s synthetic controls suffered from nonparallel trends. This may indicate that Colorado Springs’ data provided the most reliable conclusions, a subject that will be expounded upon in further detail in the “Discussion” section.

### Summary Table

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pre-Treatment MSPE | MSPE Ratio | P-value |
| Denver | 2369.431 | 2.9630 | 0.4595 |
| Colorado Springs | 1636.773 | 1.0191 | 0.7042 |
| Aurora | 18092.381 | 7.0896 | 0.1744 |
| Denver | 153592.680 | 24.2335 | 0.0125 |
| Colorado Springs | 11586.726 | 5.9443 | 0.2128 |
| Aurora | 63259.631 | 22.8868 | 0.0132 |

## Discussion

We found no statistically significant evidence in favor of the conclusion that Denver, Colorado Springs, or Aurora experienced unusually high violent crime rates after the passage of the police accountability bill in 2020 compared to control jurisdictions. On the other hand, we did find statistically significant evidence in favor of the conclusion that Denver and Aurora experienced unusually high property crime rates in 2020 and 2021 compared to placebo jurisdictions. In particular, Denver and Aurora’s property crime rates increased in 2020 and 2021 to be far above the synthetic control, and their calculated MSPE ratios were unusualy even in the context of placebo synthetic controls, decreasing the ilkelihood that chance was the explanation for the property crime increase. Additionally, Colorado Springs’ property crime rate was somewhat extreme, even if it was not significant in the placebo test. We now detail the implications of this analysis.

Several factors may cast doubt on the property crime findings that we have made here today. First, because both the Denver and Aurora synthetic controls were trending the opposite direction from the observed property crime rates, such synthetic controls are only expected to continue decreasing in 2020 and 2021. The fact that Denver and Aurora experienced large MSPE ratios may simply represent a flaw in the synthetic control itself, not a representation that Denver and Aurora possessed higher property crime rates than expected. Additionally, Colorado Springs, the jurisdiction with the strongest synthetic control, did not experience a statistically significant rise in property crime.

We believe that this concern, although valid and important to consider, should not alone be enough to invalidate our Denver and Aurora findings. While synthetic Denver and Aurora did trend opposite from the observed cities, they still matched the predictors for both cities extremely well, and if we threw out every synthetic control merely because of non-ideal trending, many of the placebo synthetic controls likely would be removed as well. Standardizing MSPE ratios by dividing by pre-treatment MSPE to account for models that aren’t well-fit should be able to compensate for some of the error. Additionally, the gaps between the synthetic jurisdictions and the observed jurisdictions were relatively moderate before 2020 and only expanded dramatically after 2020 and 2021. The synthetic controls, although weak, still imply that some causal factor is driving up property crime rates in Aurora and Denver that is not necessarily influencing other jurisdictions (or only influencing other jurisdictions to a lesser extent).

Second, the predictors we utilized were imperfect. As noted in Appendix 1, the predictors for violent crime only had an R-squared value of \_\_\_\_, while the predictors for proeprty crime only had an R-squared value of \_\_\_\_. These predictors, thus, were largely imperfect in explaining jurisdiction to jurisdiction variations in property and violent crime rates, and as a result, synthetic jurisdictions created on the basis of these predictors will be imperfect as well. Once again, this concern is valid but should not be enough to discredit the analysis. Especially for phenomenon that is as variable as crime rates, we must accept some imperfection in choosing the predictors to explain jurisdictional variations. Although data on certain predictors may improve the analysis (such as data on trust in police), the predictors that we have ensure that the synthetic controls will mimic the real jurisdictions in enough key socioeconomic indicators for the two to possess at least somewhat similar crime dynamics.

We are confident in our ability to generate valid, albeit flawed, insights from the methodology that we used. However, we do not believe definitive causal conclusions can be generated from our report.

The data does suggest that some causal factor or combination of causal factors is uniquely affecting large urban jurisdictions in Colorado in a way that other control jurisdictions are not experiencing. The data also suggests that such a causal factor likely became prominent in 2020, at least in the cases of Denver and Aurora (Colorado Springs experienced an increase in 2019 that simply perpetuated into 2020 and 2021). However, because of the limitations of our synthetic control methodology, we cannot pinpoint the causal factors that explain such an increase. The problem is especially exacerbated given the random variations and unknowns that COVID-19 and the George Floyd protests presented. For instance, it is plausible that the Colorado state government’s policy response to the COVID-19 pandemic was weaker than surrounding states, leading to higher unemployment and more property crimes. It is also plausible that the state simply experienced more property destructing during the protests compared to other states, leading to more reported property crimes. Even if we eventually gather the data needed to remove these “unknown” lurking variables, other confounding variables still hinder our ability to make an effective causal judgment. It is fully possible that the factors leading to the passage of the police accountability measure also led to increased property crime rates. For instance, citizens’ attitudes towards police officers have been shown to significantly impact crime rates. When citizens distrust police officers, more violent and property crimes tend to occur due to a decrease in proactive policing. At the same time, citizen distrust could have also generated the political momentum to pass the qualified immunity legislation in the first place. With the presence of all these different plausible causal chains, using these insights to create a definitive claim on what causal factor caused increased property crime in large urban jurisdictions in Colorado would be both improper and invalid. For causal inference to be valid, we need more than just statistics; we require cotemporal factors to be controlled for and social scientific evidence that a causal chain is plausible. Unfortunately, this report only provides the statistics and is unable to control for cotemporal factors.

Additionally, another key limitation on the ability for this report to contribute causal conclusions on qualified immunity reform is the bill itself. Although we centered this report on qualified immunity reform, in reality, Colorado’s policing legislation passed in June of 2020 included a wealth of other reform measures, ranging from police transparency to general reform of police practices. We had no way of isolating qualified immunity reform from these other measures, and as a result, our conclusions only apply to the police accountability legislation. We hoped that the effects of the other parts of the legislation would be partially accounted for by the fact that most state jurisdictions passed some level of policing reform in the summer of 2020, likely including measures present in Colorado’s bill except for qualified immunity reform. Thus, this effect may be coincidentally controlled for by the cotemporal passage of other police accountability legislation, but there is no guarantee that this will be the case, and we do not add a variable to check if police accountability legislation was passed in a particular jurisdiction or not. THus, our ability to attribute effects to qualified immunity reform, as opposed to other parts of the bill, is severely limited.

With these limitations in mind, what are our insights useful for? We believe that our insights can inform the public debate on qualified immunity and police accountability in two ways:

First, although the debate on police accountability usually centers on the effects of such bills on violent crime rates (such as murders and aggravated assaults), our report indicates that there is no significant evidence in favor of the idea that the qualified immunity legislation in Colorado coincided with a greater-than-expected increase in violent crime. In fact, especially in the cases of Denver and Colorado Springs, both jurisdictions experienced changes in violent crime that were relatively middle of the road compared to placebo jurisdictions. Although the lack of significance does *not* entail that there truly is no relationship between the two variables, we find it unlikely based on Denver and Colorado Springs’ MSPE ratios that the Colorado police accountability measure increased violent crime rates in large jurisdictions. In terms of police accountability, our insights suggest that the true evidentiary debate should center on property crime rates.

Second, we can conclude that some unique causal factor increased property crime rates in Denver and Aurora. Our analysis rules out the idea that there is no surge in property crime rate in either jurisdiction; it also rules out chance as the explanation behind the increase. Whether the trends generalize to other cities in Colorado is debatable at best; we did not see the same level of significance in Colorado Springs even if we saw some extremity. However, at least in those two cities, we believe that our analysis reveals future directions for statistical and social scientific research in determining why those cities experienced such increases.

## Appendix 1: Results from Removing Predictors

Using a combination of linear model selection procedures (such as forward selection and backward selection) and haphazard experimentation, we removed certain predictors from the analysis to improve the pre-treatment MSPE values. In particular, we removed median income and owner-occupied housing from violent crime analysis and removed owner-occupied housing and single female-led household precentage from property crime analysis. Below, we depicted a table comparing synthetic control diagnostics and results from both strategies. There was no real change in the significance of any of our results from removing these predictors.

Table 6: Comparing Results With and Without Predictors

|  |  |  |  |
| --- | --- | --- | --- |
| Model Name | Pre-Treatment MSPE | MSPE Ratio | P-Value |
| Denver Violent Crime | 2369.431 | 2.9630 | 0.4595 |
| Colorado Springs Violent Crime | 1636.773 | 1.0191 | 0.7042 |
| Aurora Violent Crime | 18092.381 | 7.0896 | 0.1744 |
| Denver Property Crime | 153592.680 | 24.2335 | 0.0125 |
| Colorado Springs Property Crime | 11586.726 | 5.9443 | 0.2128 |
| Aurora Property Crime | 63259.631 | 22.8868 | 0.0132 |
| Denver Violent Crime | 2138.568 | 3.1475 | 0.4079 |
| Colorado Springs Violent Crime | 852.016 | 0.4761 | 0.9032 |
| Aurora Violent Crime | 16656.917 | 6.9069 | 0.1860 |
| Denver Property Crime | 164062.726 | 18.6066 | 0.0247 |
| Colorado Springs Property Crime | 8536.855 | 7.6995 | 0.1500 |
| Aurora Property Crime | 42629.831 | 27.9284 | 0.0141 |

On average, in terms of pre-treatment MSPE, the models with excluded predictors tend to perform much better, the sole exception being Denver’s Property Crime synthetic controls. However, we recognize that such an effect is inevitable given the nature of the synthetic control method. As the number of predictors deccreases, the model will generally have stronger and stronger pre-treatment MSPE values, simply because the model no longer has to choose jurisdictions with matching predictors to the observed jurisdiction and can focus purely on optimizing the pre-treatment fit. Thus, the decrease in pre-treatment MSPE, alone, cannot be justification for removing the predictors from the model. Additionally, the inclusion of predictors ensures that the synthetic model creates jurisdictions that match treated jurisdictions in meaningful predictors instead of simply choosing jurisdictions that, by chance, match the crime rates of treated jurisdictions prior to treatment. Although our analysis certainly could improve if we removed some of the predictors from our analysis (such as single female-led household percentage for property crimes or percentage of population who are over 18, which were generally found to be nonsignificant in our linear modeling), we decide to keep all predictors to ensure that our methodology remains standardized.

Regardless, we include the above table to demonstrate that conclusions do not vary dramatically with the exclusion of certain unhelpful predictors, at least in the example provided.

## Appendix 2: Methodology B Results

Methodology B has the advantage of being precise about the treatment dates, allowing us to possibly isolate the qualified immunity bill as a factor, instead of other lurking variables such as COVID-19. However, we do not include the results in our main analysis due to critical methodological limitations that likely invalidate our results.