# Colorado and the Economic Effect of Cannabis: A Synthetic Control Approach

Afif Mazhar

5/13/2022

## I. Introduction

The rapid industrialization across the United States and a proactive research process on marijuana has explored the conversation to include recreational drugs as a sector in states' industrial makeup. There exists a plethora of reproducible research that detail the benefits of marijuana consumption; however, this paper will focus on a highly, imperative incentive of the legalization of cannabis: project economic growth. More specifically, determining the causality of the impacts from recreational dispensaries can be vital for further development on cannabis' economic ramifications. Although the relationship between the two is already logically sound, this research project focuses on the empirical hypothesis: Does decriminalization of cannabis increase economic growth? Colorado, the first state to introduce recreational dispensaries following the legalization of cannabis through a state-wide policy in 2014, will be the main entity of interest. A synthetic control model was employed with Colorado as the delineated aggregate unit for treatment versus a 'donor pool' of other states as the comparative case study. I find that the decriminalization of cannabis has lead to an increase of real gross domestic product.

I am interested in proving the causal effect between the 2014 legalization of cannabis in Colorado and the state's real gross domestic product.

## II. Background

In 2014, Colorado introduced cannabis into retail stores for consumers to purchase legally. From a logical standpoint, it is reasonable to make the assumption that distribution of marijuana-related consumables will promptly increase the real gross domestic product. Is it correlative to assume similar scenarios occur as most individuals consume marijuana illegally. The legalization of an ex-contraband would hypothetically increase the real gross domestic product. Papers that deal with cannabis, in reference to proving causality, deal with other issues not specific to GDP growth or using the synthetic control method. Some deal with the the relationship of cannabis to emergency visits, while others deal with its effects on the suicide rate. One paper, in particular, focuses on the relationship of cannabis with the unemployment rate and the labor market effect, but fails to use the synthetic control method for further analysis - instead, it deployed the contemporary difference-in-differences methods with the two way fixed effects design. To date, no research paper specifically proves causality between GDP growth and the decriminalization of cannabis using the synthetic control method. Using unique indicator variables that are not typically used for these two units, we will attempt to prove that the causal relationship does indeed exist.

## III. Data

The data set includes 176 total observations and 17 total columns. The first three columns indicate State number (used for the unit variable), State name, and year. The dependent variable to be estimated is the real gross domestic product, categorized by State. The next number of columns are the real gross domestic product of attributed industries that play a role on the real gross domestic product. *Mining* is labeled as the real gross domestic product of "Natural Resources and Mining" as a percentage of real GDP of the respective state. *Agriculture* is labeled as the real gross domestic product of "Agriculture, Forestry, Fishing and Hunting" as a percentage of real GDP of the respective state. *Construction* is labeled as the real gross domestic product of "Construction" as a percentage of real GDP of the respective state. *Manufacturing* is labeled as the real gross domestic product of "Manufacturing" as a percentage of real GDP of the respective state. These 4 variables are the highest percentage breakdown of Colorado's state GDP and were strategically placed in the methodology to provide the industrial makeup as accurately as possible for the model. *PCE* is listed as the "Personal Consumption Expenditure" of each respective state. *Personal Income* is the "Per Capita Personal Income" by state. *High School* and *Bachelor* are estimated percentages of the state population that have that specific degree or higher. *Poverty Rate* is the estimated percent of individuals that are below the poverty line, aggregated by state. All data was sourced from the FRED database (Federal Reserve Economic Data) from 2005 - 2020.

Below is a summary statistics of each of the key variables.

| Statistic | Mean | St. Dev. | Pctl(25) | Pctl(75) |
|---|---|---|---|---|
| Mining | 2.665 | 3.110 | 0.598 | 4.001 |
| Agriculture | 1.265 | 1.539 | 0.331 | 1.515 |
| Construction | 2.395 | 1.832 | 0.603 | 3.880 |
| Manufacturing | 7.538 | 8.023 | 0.887 | 14.113 |
| PCE | 179,942.400 | 238,379.100 | 32,714.600 | 199,385.000 |
| Personal Income | 23,714.680 | 18,205.960 | 6,169.1 | 40,024.7 |
| High School | 44.216 | 43.323 | 1.832 | 86.281 |
| Bachelor | 14.781 | 13.123 | 2.284 | 27.106 |
| Poverty Rate | 8.994 | 8.199 | 1.490 | 15.811 |

Note that the dataset is limited; real gross domestic product by sector has expanded incredibly with regards to the ever-growing complexity of the states' industrial makeup. Only 4 sectors were deployed in the model, and 5 predictor variables were used as covariates. To create a much more sophisticated analysis, expanding the sectors with most, if not all sectors displayed on the FRED database would duly increase the precision of the model. Furthermore, the timeframe for the model is solely estimated from 2005 as FRED has data for some variables (High School, Bachelor, Poverty Rate, etc.) cannot date back further than 2005. Even more, including other states that decriminalized marijuana such as South Carolina, South Dakota, Tennessee, Utah, West Virgina, Wisconsin, Wyoming, Oklahoma, and Iowa could potentially ramify the model to perform better. I chose not to use these states to keep the model relatively simple; an extension of this research design with the above limitations could produce different results although unlikely given that the synthetic control method is specific in accounting for additional observations.

## IV. Model Design

Before applying the synthetic control method (SCM) to this particular research, it is best to have an understanding of how the SCM is derived. The synthetic control method is defined as an research design method used to evaluate the effect of an intervention in comparative case studies. It involves creation of a weighted combination of groups (donor pool) used as controls which is compared to the treatment group. Essentially, the comparison estimates the treatment group if it had not received the treatment. We utilize the combination of multiple untreated units and label it the synthetic control. This method is unique to that of difference-in-differences, where confounders, variables that may have a spurious relationship amongst others, are accounted for as the synthetic control group is weighted to match the treatment group before the treatment.

Colorado in 2014 will be assigned as the treatment group with the synthetic control group as a weighted estimate of 10 states that have not decriminalized cannabis: Arkansas, Alabama, Florida, Georgia, Idaho, Indiana, Kansas, Kentucky, Pennsylvania, and Texas. Before running the research design, it is necessary to make do with natural assumptions that are pre-designed to counter any impracticalities that may appear from the experiment.

*Assumption 1: Parallel trends are not necessarily need as long as the treatment group (Colorado) and the synthetic control group have similar pre-treatment trends.*

*Assumption 2: Only the treated subject undergoes the treatment, hence the distinguishment of independent and dependent variables. Colorado, in the experiment is the only state that decriminalized cannabis.*

*Assumption 3: The identifying assumption is that the exclusion restriction is met based on the pre-treatment outcomes.*

Let us use these assumptions and apply them to the mathematical representation and the causal equation of this research design. Suppose that we have $J+1$ units and assume that unit 1 is the unit affected by some policy intervention. In this case, it is the decriminalization of cannabis in Colorado, or the official introduction of dispensaries. Units $j$=2,..., $J+1$ is a collection of untreated units, otherwise commonly referred to as the "donor pool." This is the agglomeration of Arkansas, Alabama, Florida, Georgia, Idaho, Indiana, Kansas, Kentucky, Pennsylvania, and Texas. Also assume that the data we have span T time periods, with $\mathsf{T}_\theta$ periods before the intervention. For each unit $j$ and each time $t$, we observe the outcome $\mathsf{jtT_{jt}}$.

$$Y^I_{j,t} : The\, potential\, outcome\, with\, intervention$$
$$Y^N_{j,t} : The\, potential\, outcome\, without\, intervention$$

Thus, the effect of the treated unit $j = 1$ at time $t$, is listed below. This is also referred to as the average treatment effect

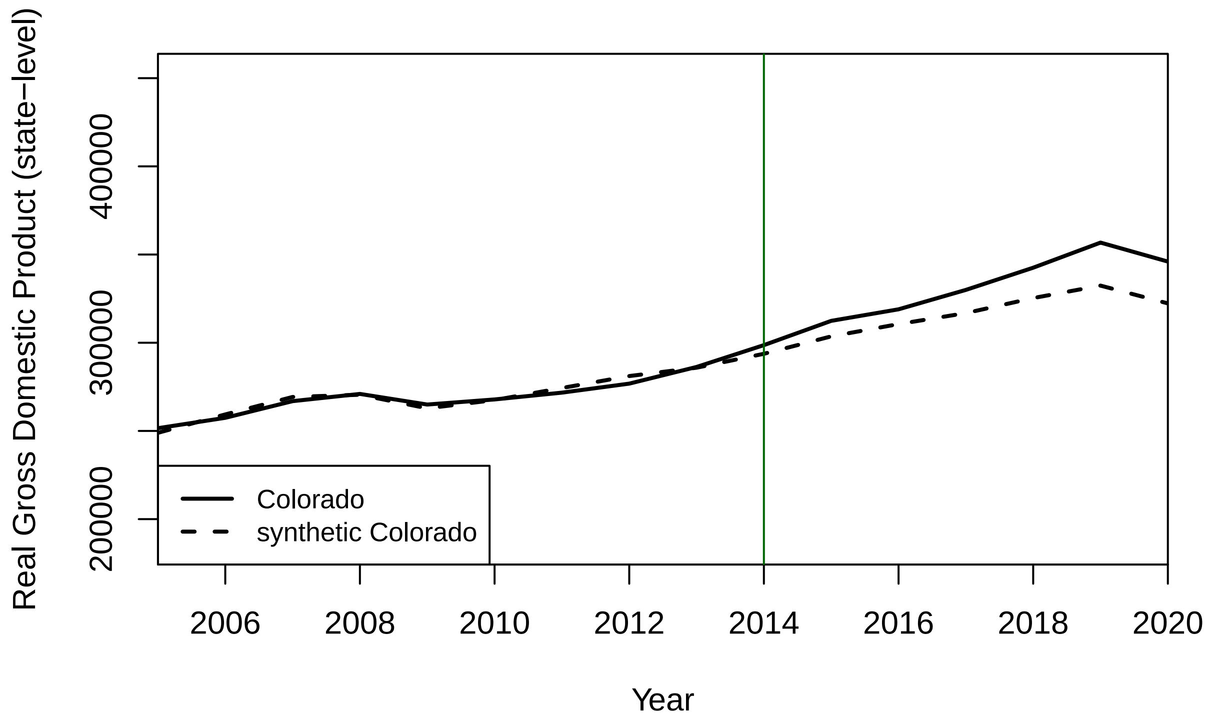
$$\tau_{1,t} = Y^I_{j,t} - Y^N_{j,t}$$

Note that $Y^I_{j,t}$ is Colorado and the interpretation after 2014 while $Y^N_{j,t}$ is defined as the synthetic control group of states that did not decriminalize cannabis in 2014. Our goal is to estimate $Y^N_{j,t}$. and to use a weighted average of the units in the donor pool to produce the counterfactual. In the post-intervention period, we have to use the synthetic control estimator $w^*_j$, which is the combination of the 10 other states and the predictor variables, to measure the causal effect of the following equation:

$$Y_{1,t} - \sum_{j=2}^{J+1} w^*_j, Y_{j,t}$$

$w^*_j$, however, is subject to weight constraints to minimize $||X_1 - X_0 W||$. Therefore, it is imperative to weight the estimator properly with the specified data given. Fortunately, the *Synth* package in R is able to reproduce the mathematical modeling and produce a proper synthetic control estimator.

## V. Results




Graphed above is a visual representation of the effects of decriminalizing cannabis in Colorado. The green, vertical line represents the treatment timing (2014, year of decriminalization). Note that the Colorado sample and the synthetic Colorado both follow the same general trend pre-treatment, but differ post-treatment. In the post-treatment phase (right of the green line) Colorado shifts higher, and presumably, produces a higher economic output in terms of real gross domestic product in comparison to the synthetic control group. Both groups falter off near 2020, seemingly due to the correlative nature of the COVID-19 pandemic which may have spuriously effected a decrease in economic output.

To have a greater understanding of how the weights were interpolated, below lists which weights were used in the model.

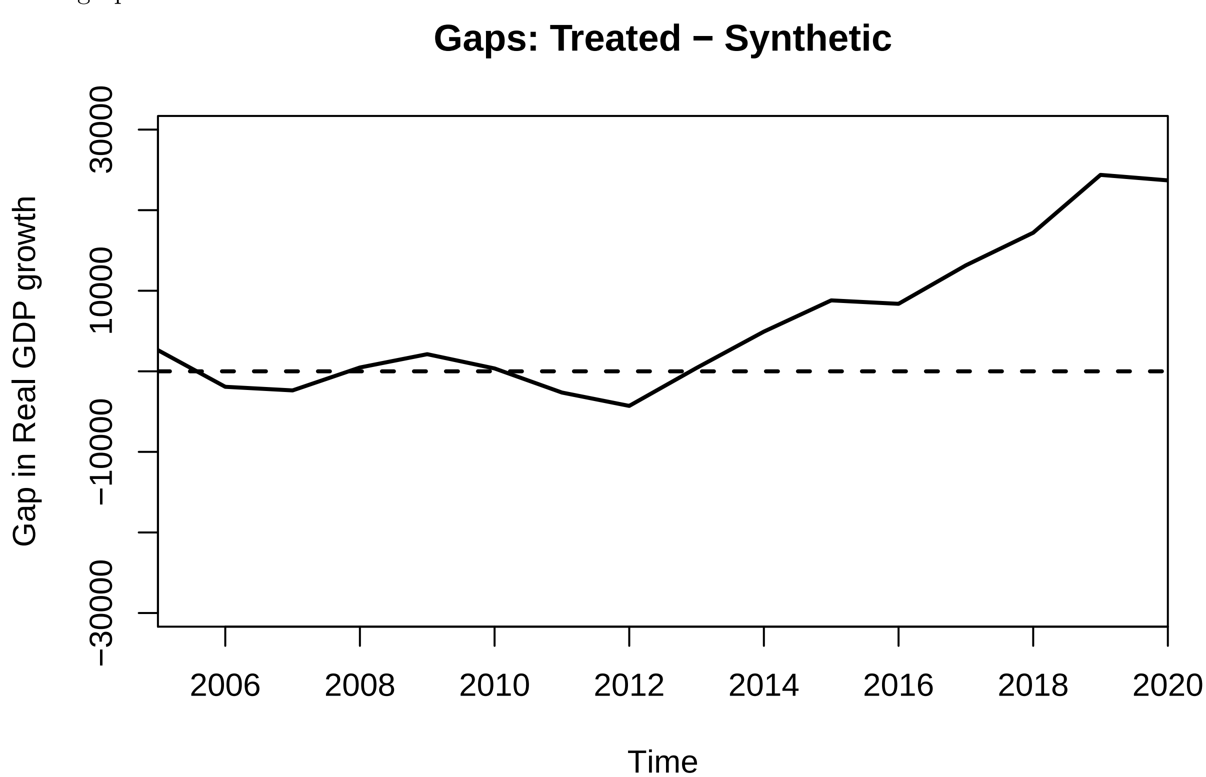
| | w.weights | unit.names | unit.numbers |
|---|---|---|---|
| 2 | 0.000 | Arkansas | 2 |
| 3 | 0.000 | Alabama | 3 |
| 4 | 0.001 | Florida | 4 |
| 5 | 0.118 | Georgia | 5 |
| 6 | 0.000 | Idaho | 6 |
| 7 | 0.000 | Indiana | 7 |
| 8 | 0.760 | Kansas | 8 |
| 9 | 0.000 | Kentucky | 9 |
| 10 | 0.067 | Pennsylvania | 10 |
| 11 | 0.054 | Texas | 11 |

Kansas was 76% of the weight distribution, Georgia was 11.8%, Pennsylvania 6.7%, and Texas was 5.4%. All other control cases make no contribution to the synthetic control estimator. Recall that the dataset is limited in observations; if the additional states mentioned in the **Data** section were used in the model, the weight estimator may shift to include those other units.

Listed below is the comparison of the pre-treatment predictor values for the treated unit, the synthetic control, and all units in the sample.

| | Treated | Synthetic | Sample Mean |
|---|---|---|---|
| Poverty Rate | 12.322 | 14.337 | 16.996 |
| PCE | 173507.433 | 171355.712 | 277492.131 |
| Personal Income | 42281.778 | 39680.041 | 36163.522 |
| High School | 89.233 | 88.111 | 84.718 |
| Bachelor | 35.911 | 28.918 | 24.441 |
| special.GDP.2005.2013 | 268273.944 | 268854.338 | 414048.317 |
| special.Mining.2005.2013 | 5.598 | 5.076 | 4.377 |
| special.Agriculture.2005.2013 | 0.865 | 3.357 | 2.230 |
| special.Construction.2005.2013 | 4.724 | 3.761 | 4.353 |
| special.Manufacturing.2005.2013 | 7.533 | 15.180 | 15.052 |

There is a massive similarity between most of the variables, excluding Manufacturing and Agriculture which satisfies the parallel trends assumption on pre-treatment values given that the model was able to properly synthesize the dependent variables and generate similar values to use for post-treatment. Furthermore, to picture the average treatment effect, I display the differences between the two estimators and display it below in the graph.



## VI. Conclusion

Using a synthetic control approach to identifying causality from Colorado's decriminalization in cannabis provides value to understanding how the model operates with regards to the weight estimators and producing counterfactuals. Not only did we not use all the variables provided in the methodical analysis, but the few that were used such as Kansas and Georgia were heavily weighted in determining the optimal estimator coefficient. It is interesting to note that not all variables were weighted in the estimator given that there were 10 units for comparison. The initial fear was that there would be multi-causal effects from vast discrepancies that may have occur beyond the treatment year, but the synthetic control method is able to account for those discrepancies and still deliver a sound counterfactual used for post-treatment analysis. Ultimately, there is proven, positive causality between the 2014 decriminalization of cannabis and the real GDP state growth. Given the results of the study, it is presumable that GDP growth on a state-level can arguably be used as an economic incentive in adoption policies for future recreational drugs.