Taking the High Way
-
The Short-Term Effects of Recreational Marijuana Legalization on Traffic Road Accidents in the State of California

Nikolas Anic, Yannik Haller
Final Project in Program Evaluation and Causal Inference
Department of Economics
University of Zurich

May 29, 2020

Abstract

Per January 2018, California became the eighth state that legalized the recreational cultivation and commercialization of marijuana. Being currently available to more than a quarter of all Americans, policy makers have continuously expressed concerns regarding the adverse effects that may follow the introduction of mind-altering substances, taking evidence from similar settings with alcohol, amphetamines or opioids. Although a substantial wealth of research exists on impairment-based driving with regard to cannabis, validity as well as transferability of tested causal effects remains difficult based on confounding individual characteristics. To deliver an improved substance for causality we analyze the effect of recreational marijuana legalization on traffic accidents in California with daily and monthly data from February 2016 to December 2019 using both a Synthetic Control as well as a weighted Difference-in-Difference approach. Using placebo tests, we find that the synthetic control groups saw similar trends following the legislation despite not having legalized the substance, finding that the mere legalization had no systematic effect on traffic accidents. This result can be justified when controlling for additional factors via the Difference-in-Difference method. However, when the focal treatment is changed to permitting the commercial distribution of recreational marijuana, the preferred Difference-in-Difference estimator suggests a positive treatment effect of 0.879 daily accidents per million population. Although varying the specifications of this model reveals slightly different magnitudes for this effect, its direction appears to be supported by each of the applied alternatives.

1 Introduction

The subject about societal and legal attitudes to be adopted towards the dealing with marijuana is a debate which is concerning American society for over 100 years. Tracing back to the 1860s, marijuana was considered a pharmaceutical, used to cure mental conditions and mitigate adverse consequences experienced by forms of stress. In the wake of extended research on the health-based consequences of commonly accepted pharmaceuticals, international politics adopted a more careful attitude towards the consumption and distribution of substances such as opioids, cocaine as well as marijuana, which was manifested in the International Opium Convention (UN, 1922). Although maintaining legal status until 1938, distribution by pharmaceutical motives and economic feasibility of marijuana cultivation gradually declined beforehand (Sanna, 2014). While general perception towards the substance was negative due to health-related concerns, attitudes substantially declined once the drug was politicized and used as a symbol of a social class embodying confounding views of a politically anticipated system (Baer, 1973). Political considerations consequently led to the Controlled Substances Act of 1970, which signified a major milestone in prohibition by reclassifying the drug to a class one substance, similar to the likes of heroin, crack, cocaine or methamphetamine, and enforced mandatory sentence laws conducted on each felony involving the drug. Although nation-wide legalization remained, political views towards marijuana have become more lenient, especially for health-related purposes. Starting in 1996, California was the first US state to enact marijuana legalization for medical use, with both Oregon and Washington following suit in 1998. Despite continued prohibition on a federal level, currently 27 US states employ laws that enable cultivation of marijuana for medical reasons in a controlled setting. Support for cultivation was able to gain another breakthrough by political legislation, as Colorado and Washington voted to legalize the production and distribution of the drug for recreational purposes in 2012. Since the respective decision, eight additional US states enacted similar forms of legislation (Hansen et al., 2017). Figure 1 displays current law-related marijuana practices each state employs.[1]

With more supportive political attitudes adopted towards the substance, security-related considerations became an important feature in analyzing civil consequences of the respective legislations, as was pointed out by Hansen et al. (2020). Provided that traffic accidents are constantly climbing and currently cost lives in excess of 38'000 victims in the US each year alone (ASIRT, 2020), considerable interest has been given to research focusing on factors potentially impairing driving abilities. While studies observing the effects of alcohol- and opioid-related driving impairment are well established (Fishbain et al., 2003; Zhao et al., 2014; Papalimperi et al., 2019), less research is devoted towards the implications of marijuana consumption on driving behavior. Notable work commonly focuses on roadside surveys or laboratory settings. As Hansen et al. (2020) rightly point out, such studies suffer from self-selection bias or heterogeneity in observational settings, which further set limits to assess, compare and generalize drug-related effects of driving behavior. A perhaps more nuanced form of research can be found in non-experimental data, which commonly analyzes

[1] The marijuana related laws of each state are obtained from the National Conference of State Legislatures (2020) and Ballotpedia (2020).

the effect of particular laws on fatalities via difference-in-difference approaches, which found confounding results of marijuana consumption on traffic-related incidents and driving behavior (Mark Anderson et al., 2013; Santaella-Tenorio et al., 2017; Sevigny, 2018). Such methods have the advantage that they solely rely on adjacent trend directions of the treatment group once set into comparison with the control states. Although more promising, they rely on background assumptions which are commonly difficult to ensure if the necessary data structure is not provided, a topic we focus on in the subsequent chapter. Most importantly, factors influencing state-level supply and demand characteristics, consumption preferences and potential spill-over effects make the detection of appropriate comparison groups difficult, if the required data cannot provide an adequate variation structure (Hansen et al., 2017).

We address this issues in the same manner as Hansen et al. (2020), and assess the effect of marijuana legalization on traffic accidents through a Synthetic Control Method. This approach allows us to create an artificially weighted counterfactual of states that did not pursue legalization practices to match the variation of predictor variables in the pre-treatment period (such as Vehicle Miles Travelled (VMT), rural and urban road mileage, Registered Motor Vehicles (RMV), population density and alcohol consumption). In a next step, we measure the discrepancies in post-treatment trends of our treatment group and its synthetic counterfactual to estimate the causal impact of our problem structure.

We find that legalization of marijuana does not systematically increase rates of traffic accidents relative to movements of the synthetic control. By using placebo studies, which assign treatment to the individual members of the control group and perform the identical matching strategy, we can further show that, once treatment is assigned at random, the probability of receiving similar post-treatment results is high, indicating that the observed variation is well within commonly expected intervals. In addition, we attempt to strengthen our results in two forms. For once, by applying a weighted Difference-in-Difference approach, which quantifies justifiably similar results to the ones obtained in the Synthetic Control. However, using the same approaches to analyze the effect of commercial distribution instead of legal enactment on traffic accidents appears to reveal similar results to the previous analysis for the Synthetic Control method, while the Difference in Difference estimation suggests a relative increase in the outcome variable for the treatment group.

Putting all together, we reject the hypothesis that the legalization of recreational marijuana leads to systematic increases of traffic-related incidents, while we fail to reject the according hypothesis from a commercial standpoint. Still, in the discussion we mention which factors prevent us from absolutely ruling out that the actual effects might differ from our findings.

The remaining part of the paper is constructed as follows. In Chapter 2, we give an overview of the data and methodology used for the analysis. Chapter 3 summarizes the results of the Synthetic Control Method and the Difference-in-Difference estimation for the legalization of recreational marijuana in a first step, before we then display the results for the commercial distribution in a second step. Lastly, we deliver a discussion of our results in Chapter 4.

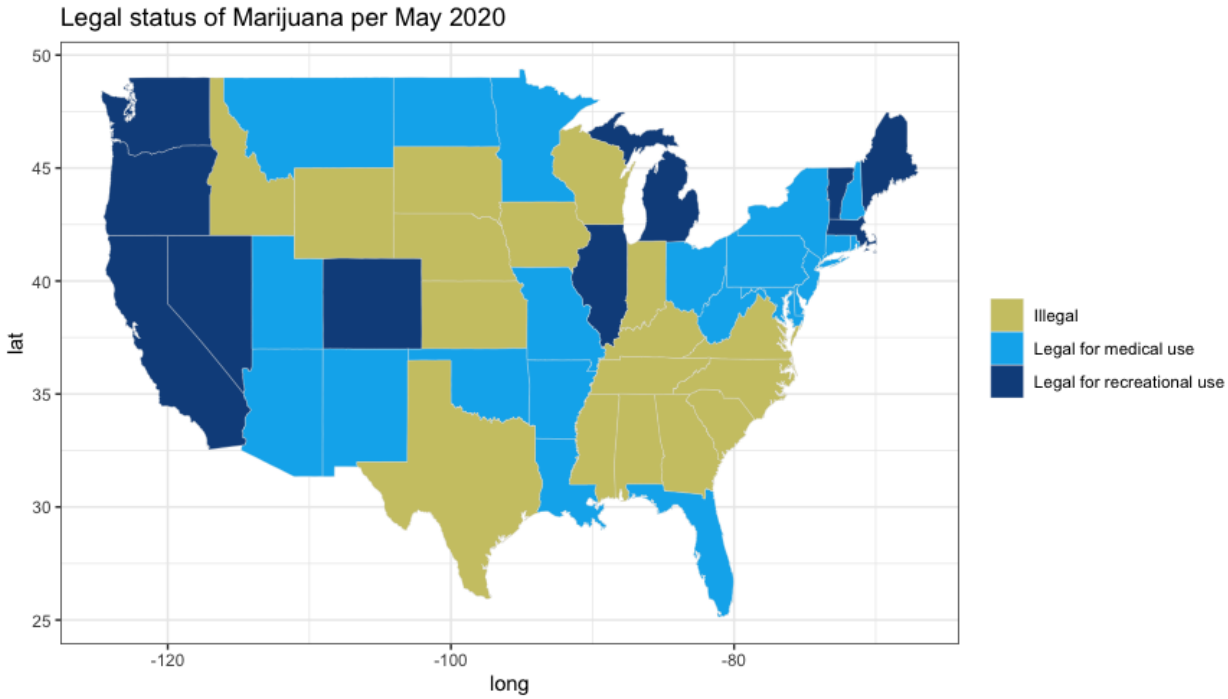

Figure 1: Legal status of Marijuana

2 Data and Methodology

2.1 Data on Traffic Accidents

In order to analyze the relation of marijuana legalization and traffic accidents, we obtain a rich data set of Moosavi et al. (2019a,b) from the University of Ohio. They collect real-time traffic data from "Mapquest Traffic" and "Microsoft Bing Traffic Data". Using continuous access to streamline services, both data providers broadcast traffic events through a variety of entities such as the US state department of transportation, local and federal law enforcement agencies as well as official traffic observation channels. The authors ensured non-disrupted data availability by refreshing traffic information every 90 seconds between 6 am and 11 pm and every 150 seconds from 11 pm to 6 am for each day throughout the measuring period. In total, they obtained 2.27 million traffic incidents between February 2016 and March 2019 for 49 states from both collection resources.[2] To address potential duplication issues, the authors followed a Haversine distance measurement approach. That is, they defined two events to be duplicates if their radius distance as well as incidence record time both fall within a pre-defined threshold, whereas they set bandwidths of 250 meters and ten minutes, respectively. Accounting for the according margins, they removed approximately 24'600 duplicates, which account for roughly 1 percent of all registered accidents. Each accident contains a number

[2]The states excluded are Alaska and Hawaii. However, they included the District of Columbia as individual state, totaling the overall number of states to 49 (Moosavi et al., 2019a,b). Also, after publishing the first version which ranges only until March 2019, they updated the data set per 31st of December 2019. However, the documentation files still contain the old numbers.

of observational features, such as geographic location indicators, meteorological information, road-based characteristics as well as day-and-time related records, whereas each account is obtained through official source. A full set of features is available online.[3] Using the latest version of the data and accumulating each data point, we obtain a data set of 49 feature variables for 2.97 million traffic accidents between February 2016 and December 2019.

In order to obtain a comparable accident structure, we need to construct a panel according to the implicitly required level-time structure. This requires two modifications. First, as we attempt to construct and compare traffic accident distributions based on state-wide legislation factors, we are required to aggregate the individual measures into state bins. Further, to get a reliable number of pre-treatment observations, we aggregate the accident timeline into daily intervals. In order to do so, we take the reported start time of each accident and assign the respective day to it. There are primarily two reasons for the daily aggregation. On the one hand, daily observations allow for week trends to be observed. Secondly, we cannot provide a balanced panel by looking at higher observation frequency, such as hours, since for several state-hour combinations we have zero observations. Finally, we summarize all measures according to their geographic and time-spatial coordinates. Unfortunately, the aggregation leads to the loss of a considerable amount of baseline characteristics for the individual measures as they cannot be reliably aggregated. For instance, meteorological factors, such as heavy fog or extreme winds, are mostly local phenomena. Also, road-based characteristics (indicators for traffic signals turning loops, junctions) are excluded from the analysis as they cannot be reasonably summarized on a state-level. A possible solution to this issue would be to analyze accident trends based on pre-defined features. However, practically no listed characteristic occurs in every state-day combination. This would require us to aggregate the data into a lower frequency, such as months, in order to get a balanced panel for specific accident features, which ultimately leads to fewer pre-treatment observations.

2.2 Parallel Trends and Measurement Error

As with all Difference-in-Difference estimation techniques, a crucial assumption of the panel data is that underlying trends between different observational units follow a parallel pattern in pre-treatment settings (Abadie et al., 2015). Existing literature often attempts to fulfill this requirement by aggregating control state data into higher levels and summarizing the individual observations. For instance, Firpo and Possebom (2018) do the same for Spain, as they analyze the economic consequences of the ETA terrorism activities mainly experienced in Basque country. This process is justifiable as it is the aim to take a weighted average of all control states that best possibly mimic the trend of the treated unit. However, any higher-level average implies that the authors assume no individual state of the aggregated data suffers from any type of measurement error, as incorrectly specified individual units may also threaten the accuracy of the aggregated unit.

Although the data gathering process of Moosavi et al. (2019a,b) was clearly carried out with

[3]Further information about the data gathering and structuring processes can be taken via the Kaggle Web page, from which the data set originated, or via the Official Notes from Moosavi et al. (2019a,b).

a high degree of precision and professionalism, we realize that the categorical observations partially show substantial differences in magnitude as well as frequency between both sources of real-time accident data. In order to understand the reason for this discrepancy, we analyze the density plots for accidents gathered over time from both source providers for each state and depict several suspicious mechanisms. First, as we can see in Figure 2, data gathering process commenced in Februrary 2016 for Microsoft Bing, but MapQuest only provided such data for a majority of states from the second half of 2017 onward. Moreover, approximately six states experience a considerable increase in accident density throughout 2019 within Microsoft Bing, whereas MapQuest appears to have a quite consistent observational density after it started providing data for all states on from July 2017. One argumentation could be that the discrepancies are based on differences in data access rights due to state-wide law enforcement. We assume that both real-time data providers cannot simply grant access rights on traffic accident observations to external sources due to data privacy issues. Consequently, any access right must be pre-approved by local law, which requires more time in some states than in others. Data access based on legal characteristics may also hold when considering an expansion of source provisions. For instance, it might be that some states only allow access from certain sources at the beginning and then gradually increase the options from which data can be gathered. Another reason could be technological development. If a state increases its means by which it can observe traffic accidents (such as cameras or online registration of accidents) it is likely that accident are more accurately registered and can be more easily accessed. Unfortunately, for neither assumption the authors provide any type of justification.

Although the exact reasons remain speculative, we strongly assume that discrepancies are not due to the absence of actual accidents for some states but are rather based on measurement error due to limited data availability. Consequently, even if the aggregated US data shows parallel trends, we cannot plausibly assume that these trends are indeed accurate as individual state-wide observations clearly don't show such trends. In order to account for this fact, we modify our data set. For the period of February 2016 to June 2017, we only include observations from Microsoft Bing, as MapQuest does not observe incidents for all states. For the period between July 2017 and May 2019 we gather data from both sources as we see a consistent observational density. Then, from June 2019 onward, we only use data from MapQuest in order to account for the large increases for several states throughout the source of Microsoft Bing. This process can be seen in Figure 3. Such modifications almost entirely account for discrepancies of parallel trends throughout our states. Lastly, we restrict the data to states for which the density plot does not depict a measurement error suggesting shape and exclude states for which the data contains accidents on less than 500 days throughout the 1423 days in our observational period, leaving out Arkansas, Idaho, Montana, North Dakota, Nevada, Oregon, Pennsylvania, South Dakota, Utah, Vermont and Wyoming.

All steps taken together, we obtain a data set consisting of 38 states over a time span of three years and nine months and summarising 54'074 observations. The density plots for each state of this final data set are shown in Figure 4.

2.3 Pre-Treatment Characteristics and Predictor Variables

As we mentioned in the previous paragraph, accidents are a heterogeneous phenomenon. As they depend on a considerable amount of factors varying for each particular situation, it is not surprising that we are barely able to find any observation that can be mirrored according to their aforementioned baseline features. Consequently, existing literature on panel-induced data analysis (Abadie et al., 2010, 2011) recommends to analyze indicators in the pre-treatment period that are known to have an impact on the variable of interest. Such pre-intervention characteristics can further increase robustness of our estimated effects as the conclusion does not merely rely on any individual characteristic but rather on a set of them. Notably, we expect that differences between states may also account for diverging features in traffic related accidents. As a consequence, we include trends for traveled vehicle miles (VMT), public urban and rural road mileages, number of registered motor vehicles (RMV), population density as well as alcohol sales per state into our analysis. We select these variables based on two considerations. First, existing research (Salomonsen-Sautel et al., 2014; Migoya, 2017; Hansen et al., 2020) assumes that these variables, although not the only ones, are likely to be a factor influencing traffic accident probability. Further, and more importantly, we observe that these variables don't follow similar trends for different states throughout the observation period. Consequently, to consistently observe that differences in traffic accident probabilities are based on our treatment indicator, we must account for trend differences of our predictor variables.

We obtain monthly data on VMT levels (in millions) from the U.S. Department of Transportation's Federal Highway Administration (2020) and divide the monthly total by the respective number of days for each month to receive a daily estimate. Further, we use yearly data on public rural and urban road mileage and number of registered vehicles, which we also obtain from the U.S. Department of Transportation's Federal Highway Administration (2015a, 2016a, 2017a, 2018a, 2019a, 2015b, 2016b, 2017b, 2018b, 2019b). Alcohol sales, measured by the total amount of ethanol contained in the sold beverages, are taken from the National Institute on Alcohol Abuse and Alcoholism (2019). Lastly, we download yearly population estimates as well as land area (measured in square miles) per state from the U.S. Census Bureau (2019) and the World Population Review (2020) to obtain population density. An important consideration is the limited comparability of absolute magnitudes for individual groups. In order to address this concern, we decide to take per million population values for our predictor variables whenever such a transformation appears to be reasonable. Lastly, we use a linear interpolation approach to transform the variables with lower frequency into daily observations.[4] Accordingly, we can ensure that we have a balanced panel.

[4]Whenever the acquired data was only available until 2018, we imposed the growth rates observed in the period from 2017 to 2018 on the period from 2018 to 2019.

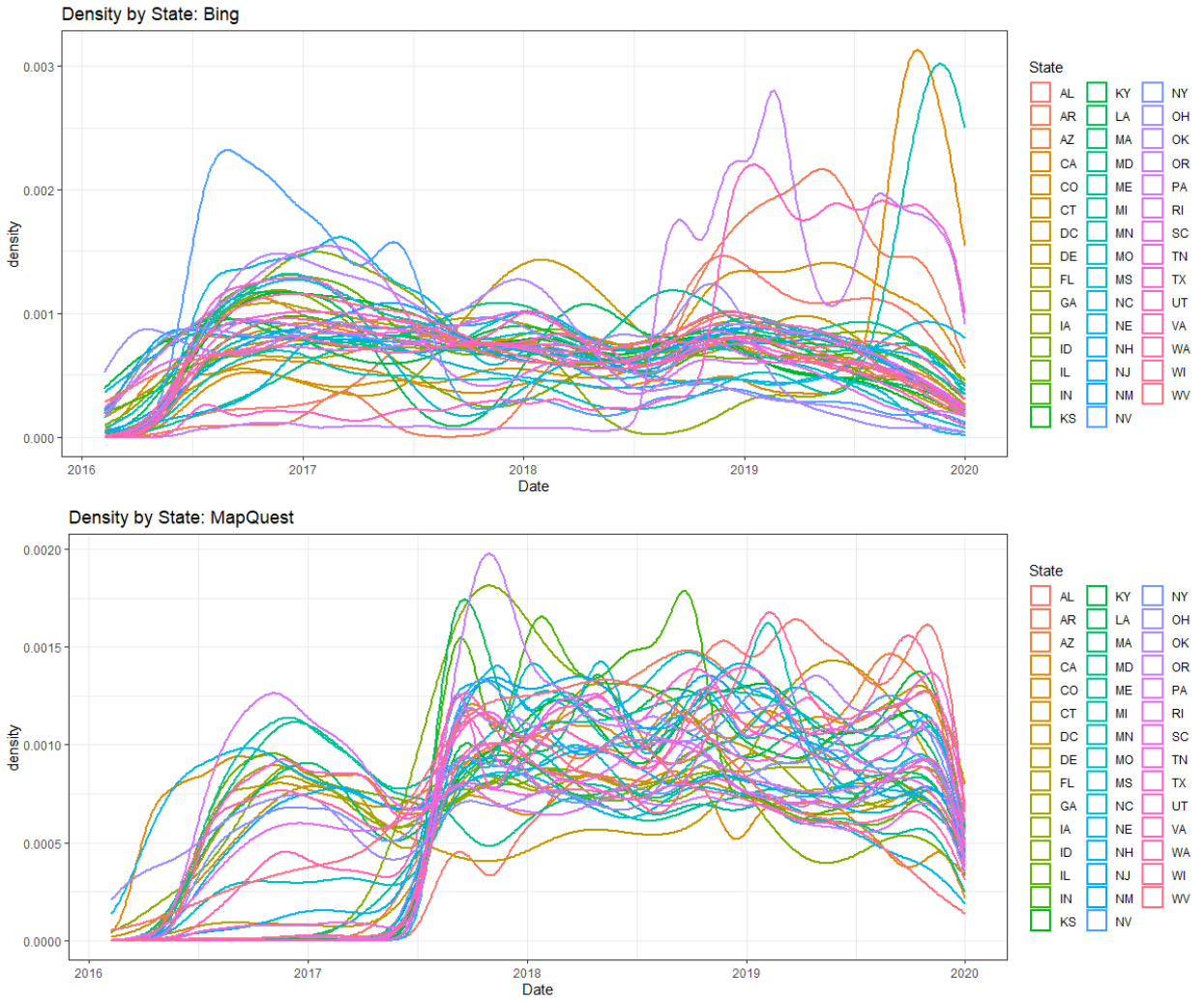

Figure 2: Accident occurrence density by State separated by data source (Bing and MapQuest)

We are aware that linear interpolation techniques may fail to explain the true within-year trends of our predictors. Essentially, we force the data to assume a linear trend throughout the year, which implies that we fail to account for seasonal trends. This approach is especially vulnerable for the number of registered vehicles as well as alcohol consumption levels, since we assume them to have quite considerable seasonal variation. However, as mentioned above, exclusion of the respective predictors is likely to result in an inconsistent estimation of our outcome variable. Nevertheless, based on the observations, we decide to include the predictor variables to control for potential heterogeneous effects in observed characteristics at the expense of not being able to accurately predict within-year trends for our sample.

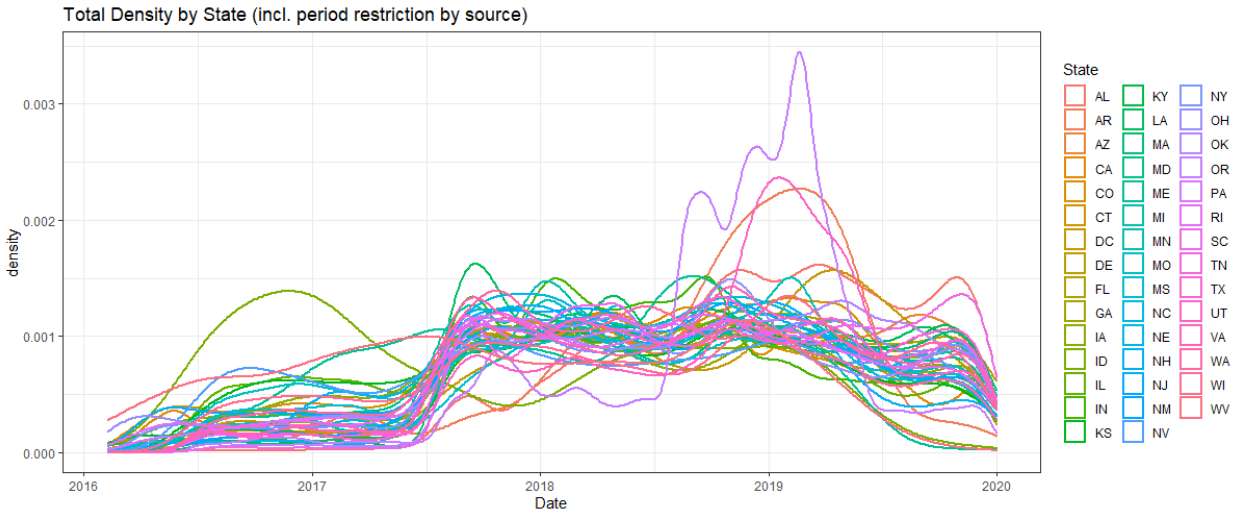

Figure 3: Accident occurrence density by State after applying the source specific time frame restriction

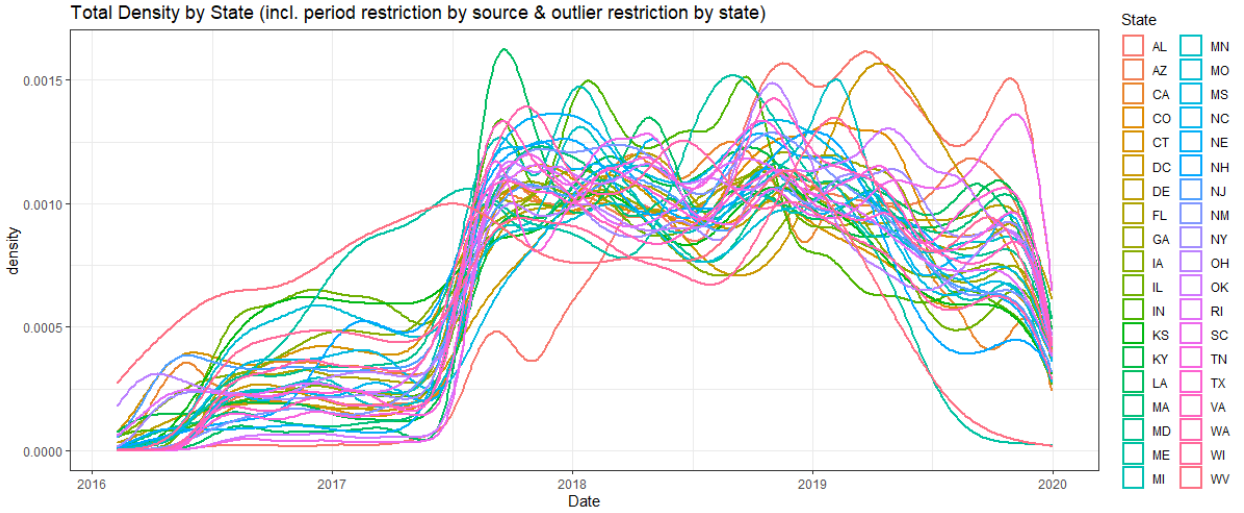

Figure 4: Accident occurrence density by State after applying the source specific time frame restriction and removing selected states

2.4 Synthetic Control Method

2.4.1 The Synthetic Control Estimator

It is our aim to measure the causal effect of marijuana legislation on traffic accidents. However, as with all evaluation practices, we cannot observe the effect of the true counterfactual. That is, we are unable to observe what results would have been obtained for our treatment group in absence of such treatment. Thus, we need to find a counterfactual representing the potential outcomes of our non-treated treatment group, which we assign the control status. Importantly, any such control group must satisfy two conditions. For once, any characteristics likely to influence the treatment outcome must be accounted for, either by

including them into the model or by ensuring similarity between treatment and control group prior to treatment. Secondly, our control group must be excluded from any type of similar treatment throughout the observation period. However, the variations observed in Figure 5 highlight the difficulty of compliance. Comparing California with US average trends, neither the variable of interest nor the predictors show similar trends throughout the observational period. The parallel trends assumption is unlikely to hold, making the US, as is, a susceptible comparison group. Still, if we were to narrow the control down to neighboring states, the exclusion of treatment restriction no longer fully holds, as all of the four neighboring states either legalized the recreational use of marijuana shortly before or after the enactment in California (Nevada, Oregon and Washington) or already had existing laws permitting the use of medical marijuana (Arizona).

In order to mitigate potential endogeneity arising from the factors discussed above, we decide to use the Synthetic Control Method (SCM) approach, first introduced by Abadie et al. (2010). This approach uses a synthetically constructed control group that closely mimics trends of the treatment group prior to treatment. The fundamental idea is that groups showing parallel trends in the pre-treatment period are expected to continue this trajectory if treatment had not taken place for one group. Consequently, differences in post-treatment trends are expected to solely rely on the treatment indicator, rendering the synthetically constructed control group as a precise counterfactual. To get an insight into the fundamental functions of the SCM, we describe shortly the logic behind its estimation.

The SCM follows the intuition that the pre-intervention characteristics of the treatment group can be more precisely approximated by a combination of untreated units than by any single untreated unit. Now suppose that we have a sample of J + 1 units, such as states in our case, where we indicate j = 1 to the treatment group and j = 2 to j = J + 1 as comparison units, or donor pool. Because our donor pool is meant to approximate the counterfactual of the treatment group in absence of treatment, it is important that this donor pool shows similar characteristics as the treatment group in predictor variables, that is we need to ensure that the outcome of the predictor variables are driven by the same structural processes as the treatment group and that the control units were not affected by any treatment of similar effect throughout the observational period. Further, they should not have been exposed to structural shocks. All summed up together, they must show a parallel trend to the treatment group in the pre-treatment period. Moreover, we must clarify that the sample is balanced, meaning that observations for our variables of interest are available for each time, t = 1,...,T, period and unit. Then, we define the Synthetic Control as a weighted average of the units in the donor pool, or control group. That means that we define a (Jx1) vector of weights $W = (w_2, ..., w_{J+1})'$ with $\sum_{j=2}^{J+1} w_j = 1$ for units in the control group. We choose a particular W such that the characteristics of the treatment unit are best described by the characteristics of the control unit. For that, existing literature defines X_1 as (kx1) vector containing the values of the pre-intervention characteristics for the treated unit that we aim to match as closely as possible and X_0 as (kxJ) matrix containing the values of the pre-intervention characteristics for the control group members. That is, both groups experience k time periods in the pre-treatment periods for their respective characteristics.

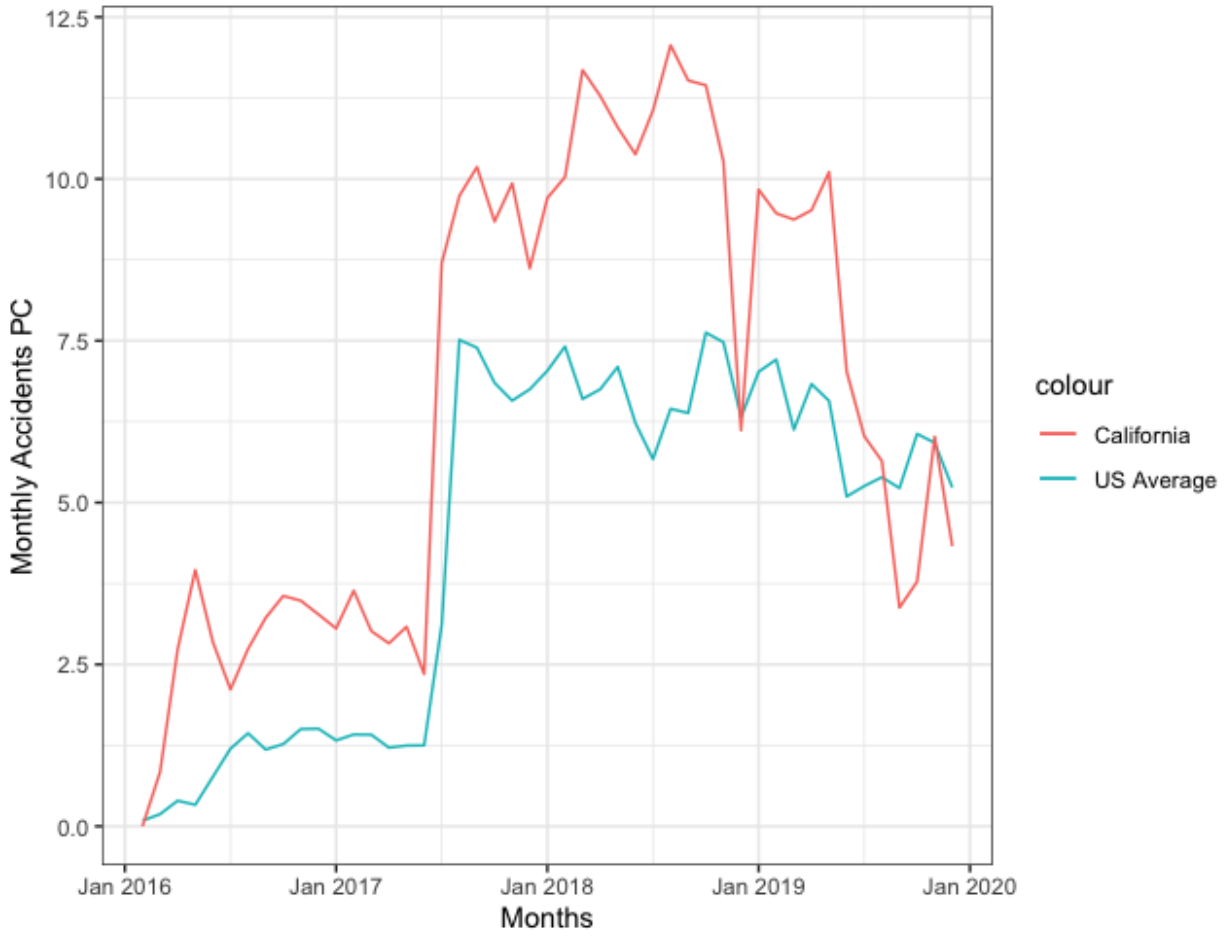

Figure 5: Distribution of accidents for California and the US Average based on monthly observations

The difference between the pre-intervention characteristics of the treated unit and a synthetic control is given by the vector $X_1 - W * X_0$. We then choose a W* such that the **size of this difference is minimized**. Then we can conceptualize this function by minimizing the sum of the differences for each of the m variables that we use to map pre-treatment characteristics, when further assigning a weight that reflects the relative importance that we assign to the m-th variable when we measure the discrepancy between $X_1 - W * X_0$.

$$\sum_{m=1}^{k} v_m (X_1 - X_0 W)^2 \tag{1}$$

To summarize this formula: what we do is that for each control variable of interest we subtract the matrix of control group units multiplied with the respective weights from the vector of our treatment values in each time period for the pre-treatment period. We then assign the resulting difference with a weight that reflects the relative importance that we assign to the m-th variable. Then we do this procedure for each of our m control variables and

take the sum. Importantly, we **take W* such that it minimizes this sum!** Recall that we take the difference between actual and synthetically weighted values for each predictor variable in each pre-treatment observation. When we have variables with large predictive power, the difference between actual and synthetically weighted values for all observations in the pre-treatment period for this predictor will be very close to zero. As we take the sum of these differences for all predictor variables, it is important that, in order to minimize the weighted average of these differences, we assign large weights, v_m, to the predictor variables that have a large predictive power and minimize this difference of actual and synthetically weighted values.

If we have sufficiently enough observations and a control group that is quite similar to the treatment group in its predictor variables, then the gap between actual and synthetically weighted values should be very small and, consequently, they should show a very similar curve. That is for the pre-treatment period. Now, we take the synthetically produced weights and look at the post-treatment period. For that we take the actual outcome variables where Y_1 is the (τ x 1) vector collecting the post-intervention values of the outcome for the treated unit, such that $Y_1 = (Y_{1,T_0+1}, ..., Y_{1,T_0+\tau})'$. Similarly, we define Y_0 to be a (τ x j) matrix that takes the outcome variable values for each control group member and for each post-treatment period. Therefore, for time $t > T_0$ of the post-intervention period , the synthetic control estimator for the effect of the treatment is given by the difference between the outcome for the treated unit and the outcome for the synthetic control unit at time t:

$$Y_{1,t} - \sum_{j=2}^{J+1} w_j * Y_{0,t,j} \tag{2}$$

2.4.2 Inference Procedure and Statistical Power

A way to test for statistical significance of these relationships is to compare the Root Mean-Squared Prediction Errors (RMSE) of pre- and post-treatment as proposed by Abadie et al. (2010) and Abadie et al. (2015). The RMSE measures the magnitude of the gap in the outcome variable of interest between each state and its synthetic counterpart. A large post-treatment RMSE does not tell us a large effect of the intervention if the synthetic control does not closely reproduce the outcome of interest prior to the treatment. That is, a large post-treatment value is not indicative of a large effect of the treatment if the pre-treatment RMSE is also large. Therefore, what we can do to further analyze the relationships is to divide post-treatment by pre-treatment RMSE to get the RMSE ratio (RMSER) and then compare the distribution for our state of interest in comparison to the placebo states. Formally, we receive:

$$RMSER_j = \sqrt{\frac{\frac{1}{\tau-1}\sum_{t=T_0+1}^{T_0+\tau}(Y_{it} - \hat{Y}_{it}^N)^2}{\frac{1}{T_0}\sum_{t=1}^{T_0}(Y_{it} - \hat{Y}_{it}^N)^2}} \tag{3}$$

Where both numerator and denominator include multiple time periods.

Then, the rank in RMSER of our treatment state relative to the others serves as permutation-based test to reveal the empirical p value:

$$p = \frac{\sum_{j=1}^{J+1} \mathbf{1}[RMSER_j \geq RMSER_1]}{J+1} \tag{4}$$

We chose this inference method based on evidence delivered by Firpo and Possebom (2018). They summarized the formal theory behind the synthetic control inference procedure, laid out the respective baseline hypotheses and argued that, in a panel data context, four assumptions must hold to to ensure the accuracy of the inference procedure proposed Abadie et al. (2010) and Abadie et al. (2015), namely:

1. Stable Unit Treatment Value Assumption (SUTVA)
2. The choice of which unit will serve as treatment group is practically random
3. The potential outcomes per region and time period are quasi-fixed, yet unknown values *a priori*
4. All potential outcomes for each state are known, regardless of treatment status assignment

Under the condition that all four assumptions hold, we can obtain Fisher's Exact p values and reject the 0-Hypotheses at the commonly known intercept levels.

Further, comparing twelve different test statistics to analyze the size and power of the inference method described above, they showed that, when only analyzing one treated unit, the approach proposed by Abadie et al. (2010) and Abadie et al. (2015) provides the most powerful test statistic for Synthetic Control settings.

2.5 Difference-in-Difference Method

In order to further evaluate the conclusions drawn by the SCM application, we additionally apply a Difference-in-Difference (DiD) approach, as manifested by Ashenfelter (1978) and Ashenfelter and Card (1978) and summarized by Imbens and Wooldridge (2009), to our data. To get credible results from a DiD estimation it is crucial to argue and verify that the parallel trend assumption is likely to hold. This means that, firstly, the pre-treatment trends of the treatment and control group have to be parallel to each other, and secondly, that the counterfactual post-treatment trend of the treated group is the same as the observed trend of the control group. In the present analysis this means that we need to find an accurate control group for California, for which we can reasonably justify that the above stated assumptions are likely to hold. As argued before, we can follow the procedure of the above described SCM to generate a plausibly accurate control group, which in the following we refer to as the synthetic control group (SCG). As soon as the SCG is generated, we use the weights assigned to each of the 31 states in our donor pool ($w_j, j \in 1, .., 31$) to calculate the values of

the outcome variable for our control group ($Y_{0,t}$) at each day t in the observational period. Formally, we receive:

$$Y_{0,t} = \sum_{j=1}^{31} w_j * Y_{0,t,j} \tag{5}$$

Assuming that the treatment effect remains constant in every period after the treatment occurred, we then are able to calculate a point estimate of the treatment effect. Theoretically, this can be done in two different ways. Either by just comparing the first differences (i.e. the difference between the post-treatment and pre-treatment average of the outcome variable) in daily accidents per million population between California and the SCG, or by running a regression of the following form:

$$Y_t = \beta_0 + \beta_1 I_{t \geq T_0} + \beta_2 D + \delta(D \times I_{t \geq T_0}) + X_t \gamma + \epsilon_t \tag{6}$$

where $I_{t \geq T_0}$ is a dummy variable indicating whether the observation is a post-treatment outcome, D is a dummy variable indicating the treatment group, $D \times I_{t \geq T_0}$ is an interaction term of the two prior dummies, X_t incorporates potential controls (e.g. dummies to capture fixed effects or variables influencing both, the outcome variable and treatment) and ϵ_t captures the error term. The resulting coefficient on the $D \times I_{t \geq T_0}$ variable, $\hat{\delta}$, then reveals the estimation of the treatment effect. Running such a regression has two crucial advantages compared to a calculation of the treatment effect by hand: it allows to control for additional variables and it is much more efficient. Considering this fact, we intend to apply this DiD method to estimate the treatment effect by means of appropriately specified regression models later on.

3 Results

The results section is divided as follows. First, we show the prevalence of the synthetic control method by showing the fit in related accidents of our treatment state and its synthetically created counterpart and provide information about the individual weights. Then we analyze the trend differences post treatment and attempt to provide reasonable point estimates for the individual intervals. Thereafter, reflecting the work of Firpo and Possebom (2018), we identify the precision of the obtained results through modification of Fisher's p values. Lastly, we compare the insights from the SCM to the resulting point estimates from the applied DiD models. Note that the considerable amount of data due to its daily availability made it quite difficult for us to provide an accurate image of the trends throughout the observational period. Therefore, any plots lack clear predictability when displayed in a daily manner. As a consequence, we ran the synthetic control with monthly aggregated data in order to provide a better image of what we would like to show. As was expected, the resulting trends and estimates for monthly data closely resemble the ones for daily data, when taken in relative terms. Therefore, we will display the results for both, daily and monthly, data.

3.1 Marijuana-Related Trends on Traffic Accidents prior to Legalization - Precision of the Estimated Model

In Figures 6 & 7 we can illustrate the relationship of accident trends over the observational period from February 2016 to December 2019 and observe which states were used to reach a synthetic counterpart. Figure 6 shows the development of California as well as its synthetic counterpart on a daily basis, whereas Figure 7 matches this relationship based on monthly aggregated data. In the period until enactment date, we can observe two distinctive features. In the beginning of said period it appears as if the trend as well as levels of the synthetic control group closely mirror the actual course of our treatment group. This changes after around three to four months as California shows increases in accidents while its synthetic counterpart remains constant. From that time on, the both variables solely show precise similarities in pre-treatment trends, although the levels of California constantly exceed the ones of its synthetic counterpart. As a consequence, we presume that the synthetic control provides a somewhat systematic solution to explain the counterfactual. In order to pin down the relation with greater precision, we show the pre-treatment predictor variables used for our synthetic counterpart in Table 1.

Table 1: Predictor Means for California, Synthetic counterpart and Average of Control States

	Treated	Synthetic	Sample Mean
Acc_PC	2.554	1.558	0.842
VMT_PC	25.680	23.919	29.922
Rural_Mileage_PC	5154.523	6268.076	26924.565
Urban_Mileage_PC	6362.035	8660.792	8896.392
RMV_PC	0.741	0.765	0.859
Pop_Density	251.366	696.692	249.380
Alc_Cons_PC	5243.542	5493.647	5402.149

Table 1 compares the actual values of California with the synthetically weighted ones as well as the average values of the 31 states in the donor pool. We can observe that for nearly all predictor variables the synthetically weighted form provides a more accurate image of the true values compared to the average. Notably, we see that accidents per million population prior to treatment are significantly lower throughout the comparison states. Also, VMT, Urban as well as Rural Road Mileage as well as RMV are substantially higher compared with the Californian trends. On the opposite, population density and alcohol consumption per million population appears to be better resembled by the average of the control states compared to our synthetically produced estimates. A reason for this might be taken from Abadie et al. (2010) analysis of cigarette consumption after Proposition 99 in California. As explained in section two, the weighted averages of each of the m predictor variables are assigned an individual weight, v_m, to minimize the mean squared prediction error of our variable of interest. The v_m for population density is close to zero (0.009), indicating that, compared to the other variables, this predictor has substantially less power in estimating the actual trends of accidents prior to the enactment of recreational marijuana legalization in California. Consequently, they argue that any failure in accurate predictability of a predictor is likely to account for the discrepancies between the actual and synthetically

produced values. Although the predictor variables are unable to exactly match the pre-treatment trends, they provide a quite considerable match when compared to the unweighted average of the control states. Also, the larger number of pre-treatment observations likely further strengthens the accuracy of the indicators. Still, both observations do not indicate that this fit is the best possible prediction overall, rather they show that, compared to traditional settings with average values, they do provide a better predictability for the specific setting. Taking all together, we conclude that the synthetic control provides a reasonably satisfying, although not significantly accurate, estimation of the accidents recorded prior to the enactment of recreational marijuana.

3.2 The Potential Effects of Marijuana Legalization on Traffic Accidents

The estimated effect of the enactment of recreational marijuana legalization on traffic accidents in California is the difference in accidents per million population between California and its synthetically produced counterpart after the enactment. These differences are shown in Figure 8 for daily and in Figure 9 for monthly data. After enactment, the trend appears to remain on a constant level compared to the rather poor pre-treatment fit, which indicates that right after the beginning no noticeable difference in accidents was observed based on the legalization. This trend then drastically changes in month 16 of our observational period as, after a short decline in accidents for the actual California, the discrepancy between both states first rises and then falls dramatically in a period of three months. This is expected. As proposed in the previous chapter, we account for discrepancies in parallel trends which we assume to be based on measurement error between the individual states by including observations from the MapQuest source only from July 2017 on.

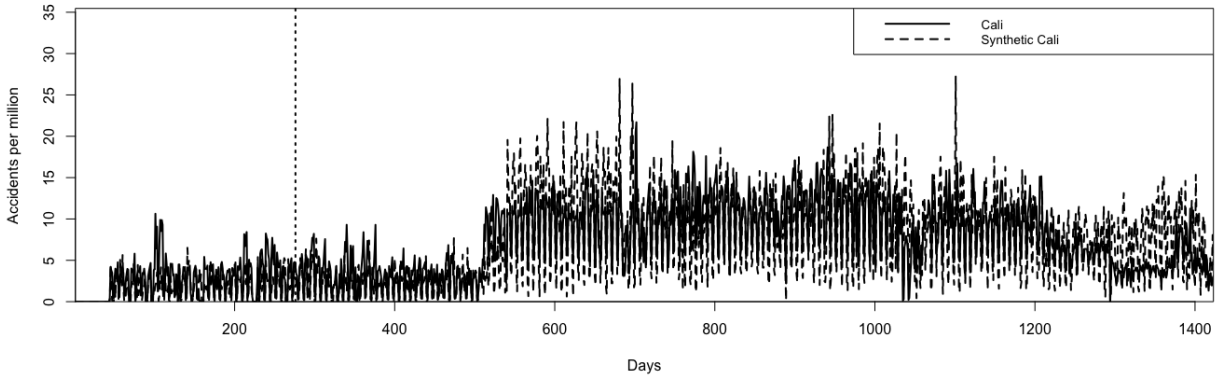

Figure 6: Daily trends of accidents for California and Synthetic counterpart

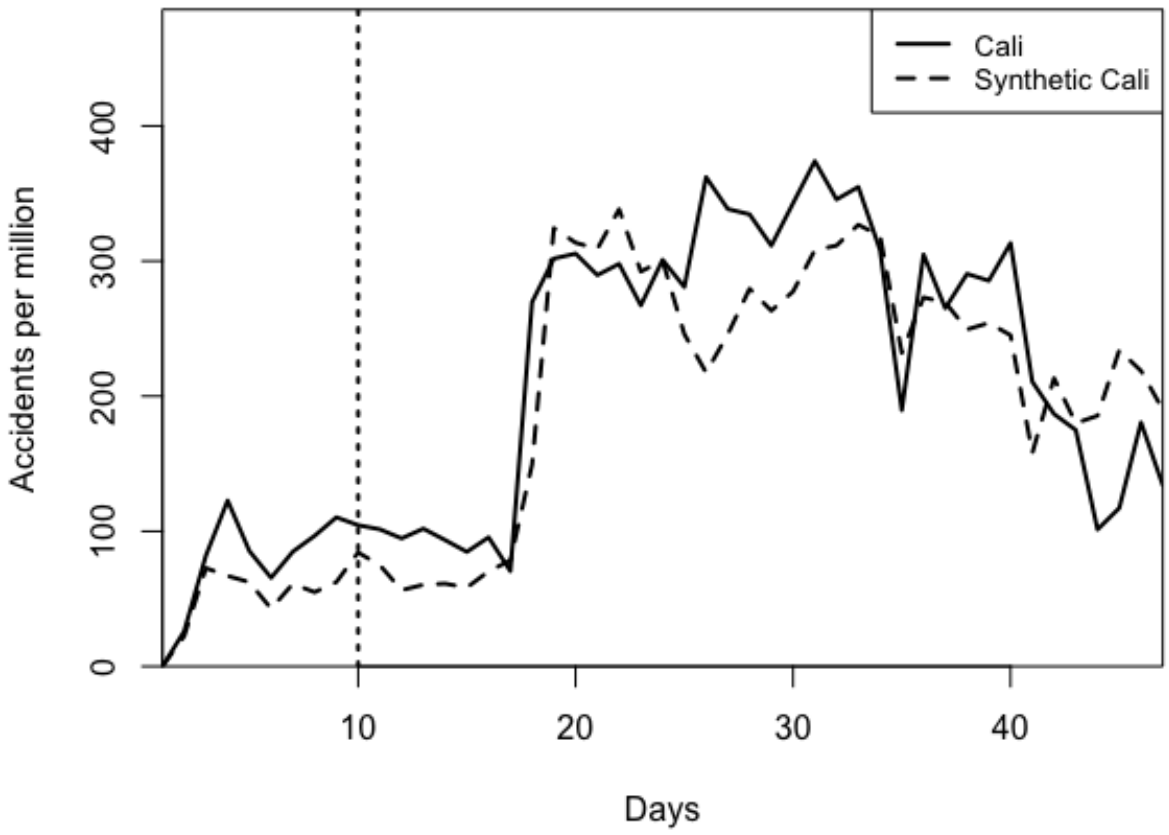

Figure 7: Monthly trends of accidents for California and Synthetic counterpart

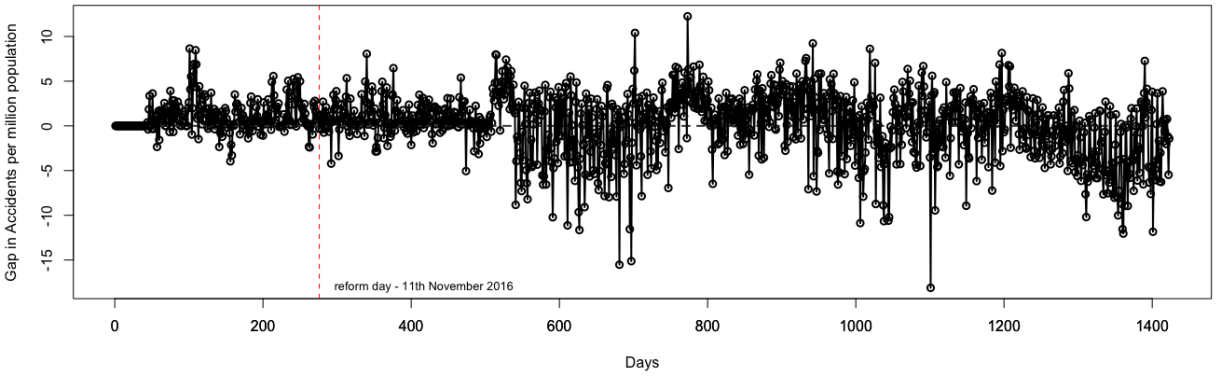

Figure 8: Gap in trends between California and its Synthetic counterpart on daily basis

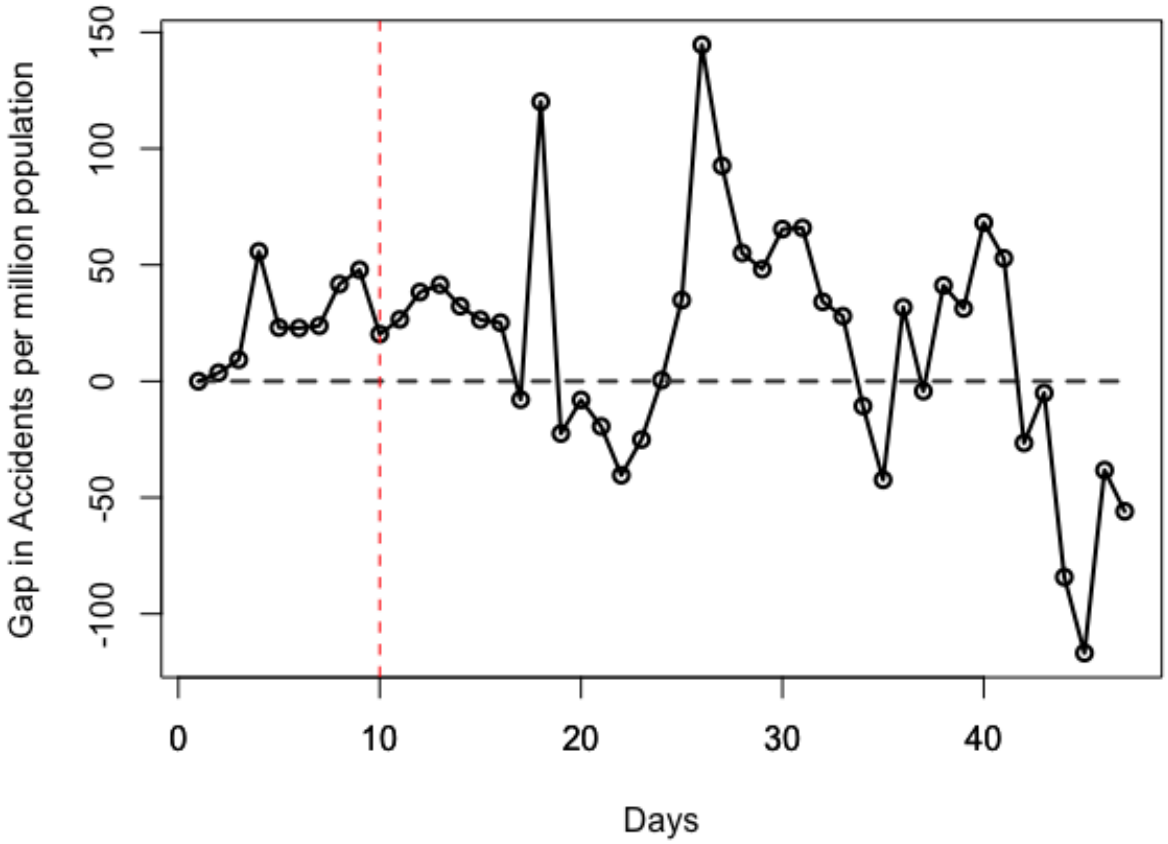

Figure 9: Gap in trends between California and its Synthetic counterpart on monthly basis

Coincidentally, California is one of the few states for which the respective source offers incident data throughout the observational period. For many control states, data from MapQuest is only provided on a gradual basis from July 2017 on, which likely explains a substantial part of said short-term variation. Similar trends are applicable when analyzing the time around month 40, which was when we excluded observations from the other source, Microsoft Bing, in order to account for the same reasons as described above. As a consequence, the subsequent, short-term variations in the gap of accidents per million population applicable in month 16 to 18 as well as in month 40 to 41 in Figure 9 can likely be traced back to the disproportional gain or loss of accident data of the treatment group compared to its synthetically weighted control group. Note that both data sources constantly provided stable observations of traffic accident data for the time period in which we included them.

Consequently, after accounting for the temporal effects for both changes of our data source structure, we can more clearly depict the proposed effect that the legalization had on traffic accidents. After month 20, we see that accidents gradually increase on a greater basis for California, which suggests that the legalization had a quite large negative effect in the short period after enactment. Following this half-year time span, excessive accidents gradually decline and follow a lightning-shaped distribution compared to the synthetically produced trend, which indicates that the drastic increase in accidents smooths out over the medium run and, if anything, potentially even decrease overall accidents per million population compared to the synthetic counterpart. Quantifying these assumptions, our results suggest that, after accounting for the data structure, the legalization of recreational marijuana lead to an average increase of roughly 5 daily (100 monthly) accidents per million population in the short run and a subsequent decline in the medium run, which amounts to an overall increase of roughly 2 daily (60 monthly) accidents per million population for our observational time horizon.

3.2.1 Statistical Significance and Inference of the Estimates

To evaluate the statistical significance of our estimates, we use placebo tests described in section three. The logic behind such placebo tests is that we assess to what probability we would have observed outcomes of similar magnitude to those of our treatment state if treatment were allocated at random to one of our control states. By iteratively applying the same synthetic control procedure to our donor states we obtain a distribution of estimated gaps for the non-intervention units. If the magnitudes of the placebo states indicate that the gap in accidents per million population is not disproportionally large for compared to the states that did not implement the legalization, then the analysis is unable to provide conclusive evidence that marijuana legalization had a positive effect on overall traffic-related incidents for California. Figures 10 and 11 show the distributions of our placebo studies as grey solid lines and California as black solid line for daily and monthly data.

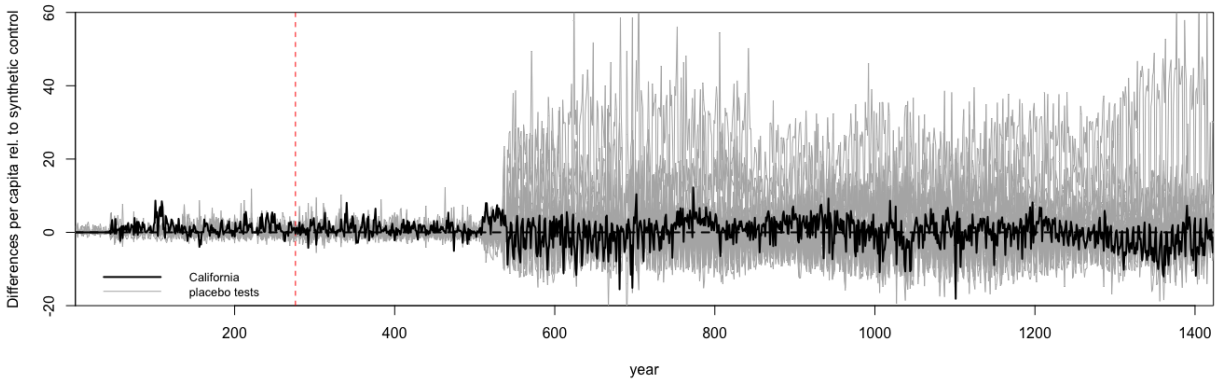

Figure 10: Placebo test runs for California (black) and Donor Pool (grey) on a daily basis

As is apparent, the gap for California is not unusually large relative to the distribution of the donor states. Prominently, the proposed increases based on the adjustments of the data set are apparent for both times at month 20, where we observe increases for all states, and 40, where we observe disproportional declines for California as compared to the donor pool. In addition, the short-term increase and subsequent medium run decline in accidents is mirrored and well within the frame of magnitudes we observe in the placebo distributions. Both observations further strengthen the assumption that the trend we observe for California following the legalization of recreational marijuana is not indicative of a systematic increase in traffic accidents. To further strengthen our argumentative basis, we calculate the distribution of the RMSER for our treatment state and the donor pool. Figure 12 exhibits the distribution of the RMSE ratios for all 31 states included in the synthetic control as well as for California on a daily basis. It can be observed that California is well below the average value of the distribution with a ratio of approximately 1.79. That said, post-treatment prediction error is approximately 1.79 times the size of pre-treatment prediction error. More precisely, if intervention were assigned on a random basis, one would receive a RMSER with the same magnitude of California with a probability of 1. This basically indicates that, in any event, the trend post treatment based on the trend prior to treatment would have been at least as extreme as in the case of California.

The distribution in gaps in combination with the given point estimates strengthen two assumptions. Initially, the trends based on observational discrepancies between the treatment and control states can be replicated through the placebo distributions. This indicates that the temporary movements at the data source changes indeed rely on disproportional variations between the member states. In other words, at least some form of variation is based on heavier increases or declines for some member states due to changes in the data source (highlighting potential measurement errors) and this variation is accounted for when replicating the study, providing less biased observational estimates for our data. Furthermore, these short- and medium run movements are not exceptional. This is important as we can expect similar lightning-shaped trends to occur when we apply the method to our control states and that these trends are within ordinary magnitudes. Nevertheless, we cannot automatically assume the accuracy of the treatment inference by solely basing our assessment on said ratio. Remember that, we base our assessment on the relation between matching before and after treatment occurred. If pre-treatment observations lack precision in predictability compared to the synthetic counterparts, then precision in post-treatment settings is also likely to be inferior. This results in two issues. For once, post-treatment observations can have a substantially larger discrepancy and still not be considered unusual. Further, even if post-treatment observations obtain reasonable variation, we cannot be certain that this variation is indeed accurate based on the matching ability beforehand. Putting all together, we expect the data to display a conditionally accurate relation on traffic accidents for the observational period and still reject the Null Hypothesis that the effect of marijuana legalization has a systematic effect on traffic accidents in the three years following the legislation. However, in order to address the concerns postulated during the assessment, we apply a different form of estimation technique to our data.

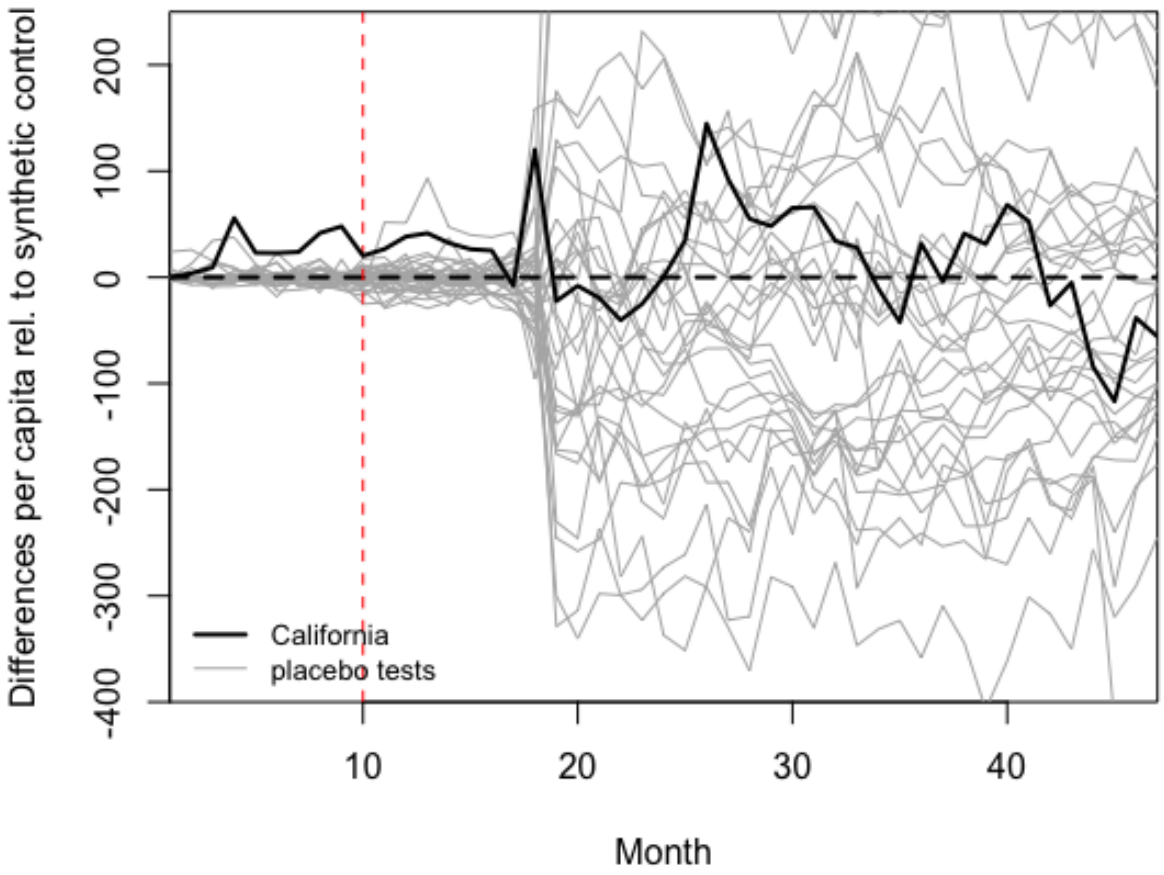

Figure 11: Placebo test runs for California (black) and donor pool (grey) on a monthly basis

3.2.2 Evaluation using Difference-in-Difference Estimation Models

As the applied analyses from the SCM approach deliver rather qualitative results, we now want to evaluate whether these results can be supported from a quantitative perspective. That is, we want to get a quantitative point estimate for the effect of legalizing recreational marijuana in California. We therefore use a SCG generated on a daily basis within the pre-treatment period from February 8^{th}, 2016, to the November 8^{th}, 2016, to estimate several different regression models.[5] The models we apply are based on differing versions of equation

[5]The weight assigned to each state in the donor pool is shown in Table 13 in the appendix. To gain interpretability while illustrating the trends of California and the SCG we decide to accumulate the daily

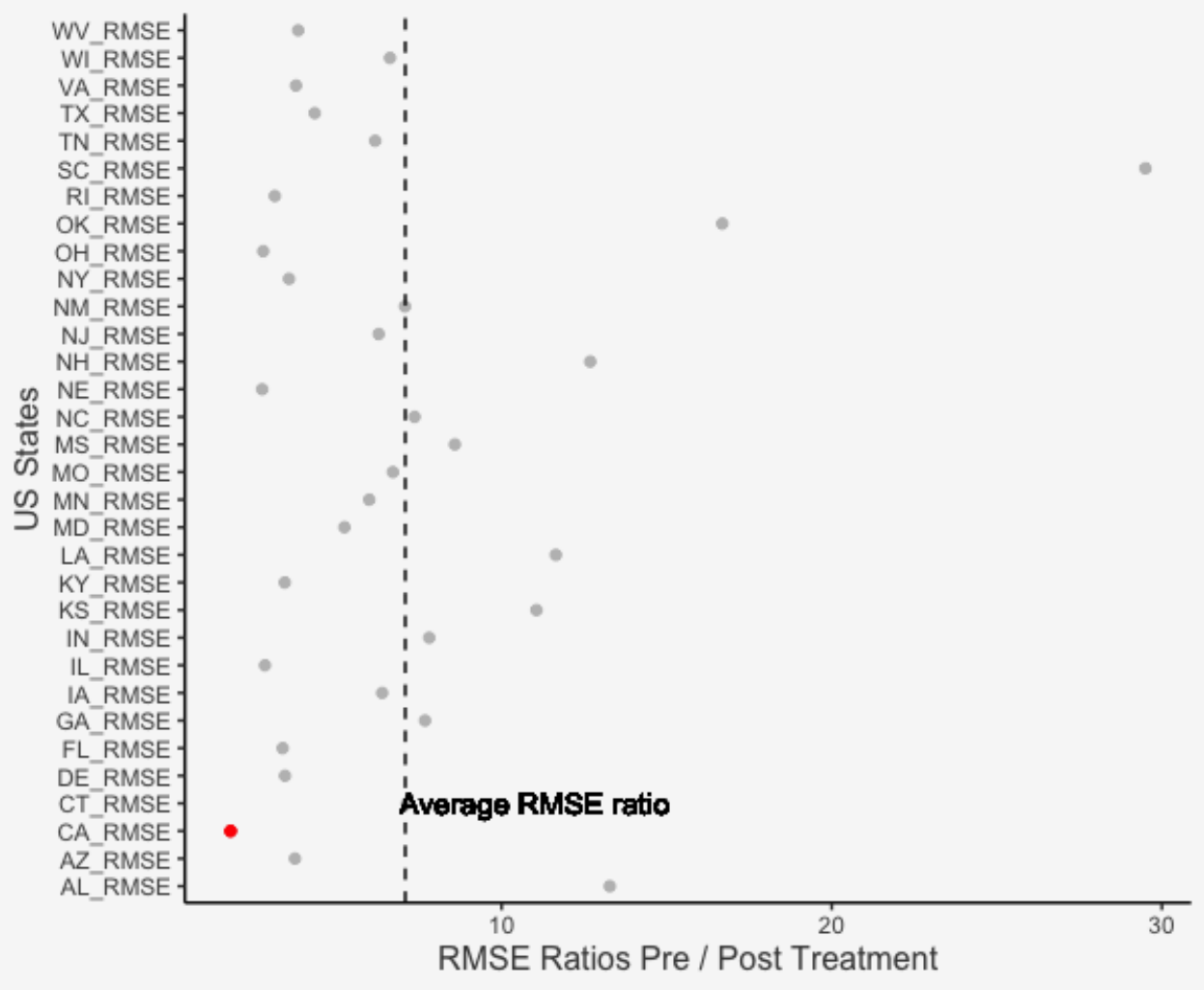

Figure 12: RMSE ratios for California and Synthetic Controls

(6). When estimating a model of this form, we implicitly assume that the treatment effect is constant in every period after the treatment occurred. For the treatment we are analyzing one could plausibly argue that this assumption is likely to hold. However, the data set we use underlies a substantial problem: it contains observations from two different data sources, which are considered for two different but overlapping periods. More precisely, data from Bing is considered from February 16 to May 19, while data from MapQuest is considered from July 17 to December 19. Although the constant treatment effect assumption is likely to hold in reality, this idiosyncrasy of our data set leads to discontinuities in the outcome variable at the cutoffs where a data source is added or removed (i.e. on 1^{st} July 17 and on 1^{st} June 19). Due to this reason the constant treatment effect assumption is probably violated in our data set, and we therefore need to account for this problem while estimating the treatment effect.

entries of the outcome variable for both, treatment and control group, over each month and plot the accidents per million population on a monthly basis. This plot, together with the indication of treatment enactment (i.e. red dotted vertical line), is shown in the appendix in Figure 14.

In a first step we run two different regression models. [6] The first model (i.e. Model (1)) is just a naive regression of equation (6) without any controls and considering the whole available time frame (i.e. February 16 - December 19). This model serves as the baseline model, which we now extended sequentially. In the second model (i.e. Model (2)) we avoid the inherent problem of our data set by restricting the time frame to last only from February 16 to June 17 (i.e. the period where only observations from Bing are considered). As this approach removes a substantial part of our data, it clearly leads to a less robust result. However, as only data from one source is considered, it resolves the key issue of our data set and therefore is likely to deliver a much less biased treatment effect. The results of these two first models can be seen in Table 7 in the appendix.[7] We observe that Model (1) reveals a negative and significant treatment effect of -0.847 daily accidents per million population, while Model (2) results in a much weaker and insignificant effect of -0.182 daily accidents per million population. Since we suspect that the treatment effect unveiled by Model (1) underlies substantial bias, we prefer the specification of Model (2).

Since increasing the data base with reliable observations always enhances the robustness of a model, we then apply further attempts to account for the inherent problem in our data set, while being able to keep all available observations. To do so, we add two dummy variables to our model, indicating each days in the observation period where data from Bing or MapQuest, respectively, is considered. We refer to this model as Model (3). However, since the effect of adding or removing a source appears to be different for the treatment and the control group, we additionally add two interaction terms in Model (4). These interaction terms are simply the product of the newly added source dummies and the treatment group indicator dummy.[8] The results of Models (3) and (4) can be seen in Table 8 in the appendix. We observe that Model (3) reveals a negative and significant treatment effect of -0.847 daily accidents per million population, while Model (4) unveils a much weaker and insignificant effect of -0.182 daily accidents per million population. Furthermore, we notice that Model (3) delivers the exact same estimate of the treatment effect as Model (1). This intuitively makes sense, as we are estimating the difference between the first differences of two curves, which both are shifted by the same extent by introducing the source dummies. However, Model (4) has the advantage that group specific effects of adding or removing a data source are accounted for. Hence, Model (4) is the preferred specification thus far.

In the last step we apply a final attempt to further enhance the robustness of our model. To do so we extend the specification of Model (4) by adding fixed effects (FE) for calendar years, months and weekdays. We refer to this final model as Model (5). An overview of the results of our preferred models (i.e. Models (2), (4) and (5)) is given by Table 2. We observe that all models reveal a negative but insignificant treatment effect of -0.182 daily

[6] We use the output format given by stargazer, which we cite here: Hlavac (2018).

[7] The point estimate of the treatment effect is represented by the coefficient on the *Post_Treat_X_Treat_Grp* variable (i.e. $\hat{\delta}$ from equation (6)).

[8] In Figure 15 (in the appendix) the residuals from the estimations of Models (1), (3) and (4) are visualized. As one can easily observe, the discrepancies in July 17 and June 19 could indeed be accounted for by adding the described dummy variables.

accidents per million population.[9] Hence, these results seemingly provide further support to the conclusions drawn by the SCM in the previous section. Namely, the resulting point estimates for the treatment effect from the DiD analyses suggest that we have to reject the Null Hypothesis that the effect of marijuana legalization has a systematic effect on traffic accidents in the three years following the legislation.

3.3 Sales Start

The current analyses prove little evidence that the legalization of recreational marijuana indeed had adverse effects on traffic-related incidents. Nevertheless, an important practical implication is missing. So far, we only analysed the impact of legislation, thereby assuming that the political decision-making process prompted individuals to adopt attitudes towards marijuana consumption that could then have negative implications for traffic behavior. However, two confounding considerations remain. For once, legalization did, in the case of California, not imply that access to the substance was improved immediately. While the bill passed in November 2016 it was only from the January 1^{st}, 2018, that commercial sale was allowed. While, potentially, the changes in legal consequences following the consumption and distribution lead to an increase in general accessibility, it is likely that the actual sales start significantly improved access to marijuana compared to pre-sales levels. Especially, sales start could have lead to the adoption of a more welcoming attitude towards the substance, in that commercial distribution enabled the promotion of the drug as lifestyle product, comparable to the likes of alcohol, tobacco or coffee. On the other hand, the general accessibility of the substance based on the already lenient management of medical marijuana practices as well as the long-lasting promotion of marijuana-related subjects throughout one of California's strongest and economically most viable entertainment industries, namely hip hop music, may be factors mitigating expected increases in traffic-related incidents based on the effects of marijuana sales compared to its mere legalization. Consequently, we decided to perform an identical Synthetic Control Method where we set the reform date to January 1^{st}, 2018. This approach allows for two advantages. On the one hand, we receive a larger pre-treatment horizon, as the enactment data switched from day 276 to day 694, which may give us a more informative estimate of pre-treatment levels. Moreover, we can include the differences in source data, which occur around July 2017, into our pre-treatment observational period, which allows us to account for said differences even before treatment commences.

3.3.1 Evaluation using the Synthetic Control Method

Table 9 and Table 15 in the appendix again pin down the predictor means as well as the composition of the (on a daily basis) synthetically constructed weights for the donor pool. As expected, the values in predictor means differ for accidents per million population, which resembles the increase in accident observations based on the additional source we included

[9]Note that adding time FE still does not change the estimated treatment effect. This intuitively makes sense, since the introduced time dummies shift the curves from the treatment and control group to the same extent. However, controlling for time FE reduces the variance of the estimated coefficients, which allows us to evaluate the results more precisely.

Table 2: DiD estimation - Models (2), (4) and (5) for cannabis legalization analysis

	Dependent variable:		
	Acc_PC		
	Model (2)	Model (4)	Model (5)
Constant	1.558*** (0.120)	−0.946*** (0.354)	0.788* (0.434)
Post_Treat	0.698*** (0.178)	0.698** (0.319)	1.298*** (0.355)
Treat_Grp	0.995*** (0.170)	−1.383*** (0.501)	−1.383*** (0.361)
Post_Treat_X_Treat_Grp	−0.182 (0.251)	−0.182 (0.452)	−0.182 (0.325)
SRC_Bing		2.504*** (0.280)	2.297*** (0.279)
SRC_MQ		7.257*** (0.271)	6.988*** (0.274)
SRC_MQ_X_Treat_Grp		−0.278 (0.383)	−0.278 (0.276)
SRC_Bing_X_Treat_Grp		2.378*** (0.397)	2.378*** (0.286)
Day_of_WeekDo.			−0.409** (0.181)
Day_of_WeekFr.			−0.042 (0.181)
Day_of_WeekMi.			−0.147 (0.181)
Day_of_WeekMo.			−0.739*** (0.181)
Day_of_WeekSa.			−5.468*** (0.181)
Day_of_WeekSo.			−5.773*** (0.181)
Month2			−0.155 (0.262)
Month3			−0.356 (0.254)
Month4			0.332 (0.255)
Month5			0.307 (0.254)
Month6			0.271 (0.266)
Month7			−0.390 (0.277)
Month8			0.665** (0.277)
Month9			0.588** (0.279)
Month10			0.813*** (0.277)
Month11			0.856*** (0.287)
Month12			−0.744** (0.291)
Year2017			−0.498 (0.341)
Year2018			0.016 (0.411)
Year2019			−0.535 (0.471)
Observations	1,018	2,846	2,846
R^2	0.070	0.479	0.731
Adjusted R^2	0.068	0.477	0.728
Residual Std. Error	1.997 (df = 1014)	3.590 (df = 2838)	2.588 (df = 2818)

from July 2017 on. Overall, the synthetic control improves predictability of the treated unit in almost all predictor variables relative to the sample mean of our donor states once we set the treatment date to sales start, which is also true for the rather imprecise variable of vehicle miles travelled. This likely resembles improved predictability based on a larger prediction horizon. Further, we observe that the new sample substituted the weight of Rhode Island for Florida as largest weight while the two previous weights for New York as well as North Carolina remained in the sample. Improvements in predictability are also shown in Figure 18 and Figure 19 (also in appendix), which portray the gap in accidents per million population for California as well as its synthetic counterpart on daily and monthly basis, respectively. Although losing some precision at the very beginning of the observational period, especially daily trends show improvements in pre-treatment gaps compared to the previous setting. However, this cannot be observed for monthly trends. Overall, we assume that, if anything, the change of treatment date lead to a slightly improved replication of California's pre-treatment trends by its synthetic counterpart.

Looking at post treatment, we are unable to pin down clear differences in accident trends between California and its synthetic counterpart when comparing it to the trends of analysing the effect of legalization. Although the increase in gap in the immediate aftermath of the treatment is less severe for the analysis of the sales start, subsequent movements resemble quite accurately the trends observed for the legalization setting. Consequently, we observe that the change of pre-treatment period does not substantially influence post-treatment discrepancies of accidents based on the sales start.

Lastly, Figure 20 and Figure 21 in the appendix show the distribution of the placebo test runs if treatment were substituted to any of our donor states. Interestingly, both daily as well as monthly trends appear to have far lower discrepancies in gap values for all of our control states, which at least partly indicates that the predictor variables more reliably explain movements in accidents, after accounting for the artificial increase in accident observations. As for the placebo runs conducted in the previous setting, we can observe that the synthetic control can only partially replicate actual accident trends with sufficient precision before sales start and that the movements for California are not unusually large compared to the donor states in the post-treatment period. Further, looking at the RMSER distribution depicted by Figure 22 and Figure 23 in the appendix, we can see that, although not as clearly as before, California's ratio lies very much within the usual range. More precisely, if one were to assign intervention to a state at random, the probability of getting a RMSER with at least the same magnitude as California would be $\frac{8}{32}$ or 0.25, which is still well above commonly acceptable thresholds for statistical significance.

Overall, all results strengthen two assumptions. First, we can strengthen our assumption that the effect of recreational marijuana legalization had no systematic effect on accident developments in California for the observational time period. Both the pre-treatment matching, post-treatment magnitudes as well as statistical inference closely resemble the results obtained in the first specification, indicating that the effect of sales start potentially did not have additional effects on traffic-related incidents that mere legalization was unable to capture. However, the problem that the synthetic counterpart is only able to replicate

pre-treatment movements of our variable of interest cannot be mitigated by applying the synthetic control to our new setting. As a consequence, the accuracy of our obtained estimates is still likely to be limited due to a potential lack of fit. Hence, in order to assess our estimated effects, we again use a weighted Difference-in-Difference approach.

3.3.2 Evaluation using Difference-in-Difference Estimation Models

Alike in the previous DiD evaluation, we again intend to assess whether the qualitative results from the SCM approach can be supported from a quantitative perspective. To do so, we calculate point estimates for the treatment effect of issuing licences to commercially sell recreational marijuana by means of several different versions of the DiD regression model specified in equation (6), similar to the ones we estimated in the DiD evaluation above. However, as argued before our data set suffers the substantial problem that the constant treatment effect assumption probably is violated. The big difference between the two treatments we are analyzing is that the treatment of sales start occurs only after observations of MapQuest are already added to the data set. To account for this problem we estimate the applied regression models with three different SCGs.[10] The big difference between these SCGs is that they are estimated on differing pre-treatment periods. The first one is estimated within the time frame between February 16 and December 17, which includes the cutoff in July 17 where observations from MapQuest are added to the data set. The second and third SCG avoid including this cutoff on the cost of a shorter pre-treatment period. Namely, the second SCG is generated using July 17 to December 17 as the pre-treatment period, while the third SCG considers the time between August 17 and December 17 as the pre-treatment period on which it is generated.[11] However, all of these three SCGs are estimated on a daily basis within the corresponding pre-treatment period. Worth mentioning is also that the regression models using the first SCG are estimated on the whole available time frame (i.e. February 16 to December 19), while the models using the second or third SCG only rely on corresponding subsets of the data, in which observations from both sources are available (i.e. data between July 17 and May 19).

In the following we discuss the results from the regression models we estimated using the first SCG as described above. The first model we apply here, which we refer to as Model (6), relies on the exact same specifications as Model (1) from above. Obviously, the important bit that is different now is the treatment, which has changed from legalizing recreational marijuana to issuing licences to commercially sell recreational marijuana. The result of this model can be seen in Table 11. We observe that a negative and insignificant treatment effect of -0.088

[10] The weight assigned to each state in the donor pool is shown in the appendix in Table 15 for the first SCG, in Table 18 for the second SCG and in Table 17 for the third SCG. Furthermore, the monthly trends in accidents per million population of California and the corresponding SCGs are shown in Figure 24 for the first SCG, in Figure 25 for the second SCG and in Figure 26 for the third SCG (also in the appendix). The treatment enactment is indicated by the red dotted vertical line in each of these figures.

[11] The reason for estimating a third SCG which uses an even shorter pre-treatment period is that we suspect the second SCG to be inaccurately weighted due to the huge discrepancy between the treatment and the control group's outcome values in July 17. In Figure 25 can be seen that California depicts a much higher number of monthly accidents per million population in the initial month than its synthetic counterpart. Furthermore, when looking at Figure 26 one can easily see that such a huge discrepancy no longer occurs.

daily accidents per million population is revealed. However, as argued before we strongly suppose the resulting treatment effect from this basic model to be biased. We therefore again extend the basic model by introducing the previously suggested source dummies to get Model (7). Furthermore, by additionally including the interaction term between these source dummies and the treatment group indicator dummy we get Model (8). Alike Models (3) and (4) from above, Models (7) and (8) are able to account for the discrepancies occurring when a data source is added or removed from the data set. Moreover, Model (8) has the advantage that it additionally can account for group specific effects of adding or removing a data source. The results of these two extended models are shown in Table 12 in the appendix. We observe that Model (7) still delivers the exact same treatment effect as Model (6). This clearly makes sense, since the treatment effect measures the difference in the first difference between the treatment and the control group's curves, which both are shifted to exact same amount by introducing the source dummies. Thus, the result from Model (8), which depicts a positive and significant treatment effect of 0.879 daily accidents per million population, is probably more reliable. In a last step we extend Model (8) by adding FE for calendar years, months and weekdays, which results in Model (9). The results of the preferred models we estimated using the first SCG (i.e. Models (8) and (9)) are shown in Table 3. As expected, we observe that controlling for time FE does not change the magnitude of the estimated treatment effect, but further improves its precision by reducing the noise of the estimated coefficient.

In the next step we discuss the results from the regression models we apply using the second and third SCGs as described above. Since these models are exclusively estimated on periods for which the data provides observations from both sources, there is no longer a need to add source dummies. Thus, the pool of potentially interesting regression models we can apply to estimate the treatment effect reduces to two remaining specifications. The first one is just the basic model specification as proposed in equation (6), without using any control variables. Estimating this basic model with the second and third SCG, delivers Models (10) and (12), respectively. The second interesting model specification is the one where controls for time FE are added to the basic model. Using this extended specification with the second and third SCG as described above, delivers Models (11) and (13), respectively. The results of Models (10) and (11) (i.e. the models which rely on the second SCG) are summarized in Table 4. We observe that both models deliver the same positive treatment effect of 0.737 daily accidents per million population, whereas only in Model (11) the resulting estimate is declared to be statistically significant. Furthermore, when looking at the results of Models (12) and (13) (i.e. the models which rely on the third SCG), which are shown in Table 5, we observe that an even stronger treatment effect of 1.449 daily accidents per million population is revealed. Moreover, the estimate is highly statistically significant in both models. However, since the pre-treatment periods to generate the second and third SCG are substantially shorter than the one used for the first SCG, we think that Model (9) probably delivers the most confidential results. Nevertheless, the results of Models (11) and (13) seemingly support the conclusions we can draw from the results of Model (9). Taking all together we think that we can reasonably argue that the sales start of recreational marijuana indeed had a positive effect on traffic-related incidents in California. Namely, our preferred DiD model proposes a positive and significant treatment effect of 0.879 daily accidents per million population. Hence, we fail to reject the Null Hypothesis that the effect of issuing licences to commercially

Table 3: Did estimation - Models (8) and (9) for sales start analysis

	Dependent variable:	
	Acc_PC	
	Model (8)	Model (9)
Constant	0.160 (0.286)	1.178*** (0.345)
Post_Treat	−0.014 (0.265)	0.648** (0.303)
Treat_Grp	−2.480*** (0.404)	−2.480*** (0.273)
Post_Treat_X_Treat_Grp	0.879** (0.375)	0.879*** (0.253)
SRC_Bing	1.472*** (0.251)	1.433*** (0.225)
SRC_MQ	6.485*** (0.265)	5.842*** (0.226)
SRC_MQ_X_Treat_Grp	0.136 (0.375)	0.136 (0.253)
SRC_Bing_X_Treat_Grp	3.638*** (0.355)	3.638*** (0.239)
Day_of_WeekDo.		−0.150 (0.146)
Day_of_WeekFr.		−0.091 (0.146)
Day_of_WeekMi.		−0.014 (0.146)
Day_of_WeekMo.		−0.517*** (0.146)
Day_of_WeekSa.		−4.970*** (0.146)
Day_of_WeekSo.		−5.179*** (0.146)
Month2		−0.145 (0.210)
Month3		−0.088 (0.204)
Month4		0.389* (0.205)
Month5		0.366* (0.204)
Month6		0.418* (0.214)
Month7		−0.024 (0.223)
Month8		0.985*** (0.223)
Month9		0.524** (0.224)
Month10		0.953*** (0.223)
Month11		1.064*** (0.224)
Month12		−0.109 (0.223)
Year2017		0.698*** (0.150)
Year2018		0.365** (0.156)
Year2019		
Observations	2,846	2,846
R^2	0.525	0.785
Adjusted R^2	0.523	0.783
Residual Std. Error	3.085 (df = 2838)	2.081 (df = 2819)
F Statistic	447.459*** (df = 7; 2838)	396.144*** (df = 26; 2819)

sell recreational marijuana in California has a systematic effect on traffic accidents in the two years following the enactment. This result differs from the conclusions drawn from the above applied SCM (i.e. that sales start has no systematic effect on traffic accidents). However, since the conclusions from the DiD method can be supported by numerous models we applied, we think that these quantitative results have a higher validity than the qualitative conclusions from the SCM.

Table 4: Did estimation - Models (10) and (11) for sales start analysis

	Dependent variable:	
	Acc_PC	
	Model (10)	Model (11)
Constant	8.046^{***} (0.279)	10.124^{***} (0.296)
Post_Treat	0.128 (0.325)	-0.561^{**} (0.251)
Treat_Grp	1.365^{***} (0.395)	1.365^{***} (0.208)
Post_Treat_X_Treat_Grp	0.737 (0.460)	0.737^{***} (0.243)
Day_of_WeekDo.		−0.176 (0.200)
Day_of_WeekFr.		−0.052 (0.200)
Day_of_WeekMi.		0.062 (0.200)
Day_of_WeekMo.		-0.695^{***} (0.200)
Day_of_WeekSa.		-6.954^{***} (0.200)
Day_of_WeekSo.		-7.299^{***} (0.200)
Month2		0.275 (0.260)
Month3		0.609^{**} (0.254)
Month4		0.578^{**} (0.256)
Month5		0.469^{*} (0.254)
Month6		−0.038 (0.325)
Month7		-0.805^{***} (0.276)
Month8		0.752^{***} (0.276)
Month9		0.722^{***} (0.278)
Month10		0.686^{**} (0.276)
Month11		0.431 (0.278)
Month12		-0.947^{***} (0.277)
Year2018		0.676^{***} (0.163)
Year2019		
Observations	1,400	1,400
R^2	0.064	0.743
Adjusted R^2	0.062	0.739
Residual Std. Error	3.785 (df = 1396)	1.997 (df = 1378)
F Statistic	32.078^{***} (df = 3; 1396)	189.544^{***} (df = 21; 1378)

Table 5: Did estimation - Models (12) and (13) for sales start analysis

	Dependent variable:	
	Acc_PC	
	Model (12)	Model (13)
Constant	8.678*** (0.305)	10.584*** (0.312)
Post_Treat	−0.728** (0.347)	−1.085*** (0.267)
Treat_Grp	0.878** (0.431)	0.878*** (0.232)
Post_Treat_X_Treat_Grp	1.449*** (0.491)	1.449*** (0.264)
Day_of_WeekDo.		−0.073 (0.207)
Day_of_WeekFr.		0.102 (0.207)
Day_of_WeekMi.		0.167 (0.207)
Day_of_WeekMo.		−0.512** (0.208)
Day_of_WeekSa.		−6.846*** (0.208)
Day_of_WeekSo.		−7.134*** (0.208)
Month2		0.231 (0.264)
Month3		0.460* (0.258)
Month4		0.571** (0.260)
Month5		0.567** (0.258)
Month6		0.244 (0.330)
Month7		0.024 (0.326)
Month8		0.365 (0.283)
Month9		0.686** (0.285)
Month10		0.574** (0.283)
Month11		0.352 (0.285)
Month12		−1.223*** (0.283)
Year2018		0.282* (0.165)
Year2019		
Observations	1,338	1,338
R^2	0.071	0.735
Adjusted R^2	0.069	0.731
Residual Std. Error	3.769 (df = 1334)	2.027 (df = 1316)
F Statistic	34.149*** (df = 3; 1334)	173.767*** (df = 21; 1316)

4 Discussion and Implications

All obtained results only provide little support to the hypothesis that the legalization of recreational marijuana has adverse effects on traffic-related incidents. However, several limitations arise which should be addressed before making a concluding remark to the results obtained.

Initially, we provided the reader with important assumptions on which we base our statistical inference. Particularly, we assume that the choice of which unit will serve as treatment group is practically random. Although our placebo-based strategy allows us to assign treatment status to all members of the donor pool and compare individual post-treatment trends which show that California is well within the range of expected outcomes, we cannot rule out that certain unobserved characteristics prevailed which potentially influence both the treatment decision as well as the treatment outcome. That is, one could assume that several components influencing road safety, risk-behaviour or even attitudes towards impaired driving fundamentally differ between the treatment group and our donor pool, which even made the step to legalization possible in the first place. Such factors clearly could alter the observed relationships or even substantially mitigate comparability. As a consequence, the used methods to obtain statistical inference cannot exclusively be regarded as superior method and, hence, may lack credibility.

Another factor in need of consideration is the relatively imprecise matching ability of pre-treatment variations through our predictor variables which we mentioned in previous chapters. Although the sample mean of our synthetic counterpart outperforms average control levels, we still lack a clear predictability of pre-treatment trends. Such lack of fit is then likely to repeat after treatment started. As a consequence, the movements of the counterfactual are likely to suffer from imprecision, which would at least partly account for differences in RMSER. Unfortunately, the respective bias cannot be assigned a clear direction. Although a more precise replication of pre-treatment trends would increase the RMSER, we cannot depict if the increase in precision would lead to a con- or divergence of actual and expected movements post-treatment, leading to further uncertainty about the statistical inference procedure.

Most notably, however, we regard the modifications of the data set as key factors which mitigate our ability to draw concise conclusions from our statistical analysis. On the one hand, we assume that the generating process with which the data was provided itself suffers from measurement error. This is given by the large discrepancies of accident data both between states and throughout time, that we address in chapter two. Although we attempt to account for this source of bias by modifying the data such that no clear dissimilarities both within states and time exist, we still suffer from two uncertainties. First, when we assume differences in accident movements are at least partly based on an improved observation practice, then excluding observations which bear patterns of unusually large increases could further bias the data as we artificially exclude movements which are potentially closer to the truth (assuming that increases in observations are indeed correct). Secondly, even if we adjust source availability throughout time, we cannot guarantee that the trends obtained

are indeed an accurate depiction of the present.

This also counts for our predictor variables. None of the data which predicts the pre-treatment movements, besides accident trends, is based on daily levels. Although our approach of linear interpolation allowed us to create a balanced panel, we must make the assumption that linearity is indeed true for our variables. While it may be true that trends of some variables, such as road construction or population density, underlie patterns which can more accurately be depicted by linear trends, others are likely to fall short. Based on the analysis of daily data, we especially have to consider that alcohol consumption or VMT is likely to follow inter-weekly trends (such as an increase throughout weekends). Continuing with monthly data, it is probable that RMV, VMT as well as alcohol consumption follow monthly trends which one should account for (such as increases over holiday seasons). However, the underlying data structure, which is commonly based on yearly levels, does not allow for such a nuanced differentiation.

To sum up the last consideration: all we can provide with these methods is to partly ensure that measurement error is somewhat consistent throughout the individual treatment and control group members. However, consistency does not allow us to draw causal conclusions of the estimation procedure as long as we assume that the underlying data structure is unable to provide an adequate display of the reality. In addition, the first two considerations point out the difficulty in assessing and inferring variation even if we assume the data to display reality in an accurate manner. All things considered, although the Synthetic Control Method provides a powerful tool in estimating causal inference, it still is based on a range of assumptions which heavily depend on the underlying structure. Although we attempt to balance the trade-off between data availability and observational accuracy, we must be careful in assigning a causal relation of the analyzed variables. Nevertheless, we assume that the SCM, as used for this data set, is at least able to provide us with the notion that the legal enactment of recreational marijuana does not lead to systematic increases in traffic accidents for the observational time period.

When interpreting the results from the DiD analyses, one should always keep in mind that for both treatments we analyze we have rather short pre-treatment periods on which we generate the SCG, leading to the problem that the synthesised control group might not be accurate enough. Especially for the analysis of the sales start treatment, where the pre-treatment period includes the cutoff where MapQuest is added as a second data source, we need to be aware that this discontinuity could cause additional problems in the generation process of the SCG. Although we tried to avoid this problem by choosing a shorter pre-treatment period to generate the SCG, we still have the problem that this pre-treatment period might be too short to synthesise a robust control group. Further, as mentioned above we potentially have the problem of measurement error in our data set, which probably causes attenuation bias in the estimated treatment effects. Another potential source for bias in our estimates could spring from omitted variables, which tend to influence both, treatment status and accident frequency of a state. For instance, one such variable could be the unobserved portion of car drivers driving under the influence of marijuana. A high portion of drug impaired drivers is clearly associated with a higher number of daily traffic accidents per person and probably

also indicates that the resident population has a rather drug favouring attitude and therefore is more likely to accept the legalization of marijuana. Hence, as the portion of drug impaired drivers is not observable for us, this probably leads to positively biased estimates. However, one could also argue that this bias rather drags the estimated coefficients into the opposite direction by reasoning that a high portion of drug influenced drivers causes voters to rather vote against a legalization of marijuana. Since it is unclear which of these mechanisms apply in reality, it is hard to draw a unique conclusion about the direction of this bias. However, the results of our DiD analyses appear to conform with our expectations. We therefore suppose that latent biases are probably not strong enough to significantly change the treatment effects we revealed (i.e. no significant effect of cannabis legalization on traffic accidents per population, but a positive effect of cannabis sales start on accident frequency per capita). Furthermore, our Model (9) appears to deliver a rather precise estimate for the treatment effect of cannabis sales start, which suggests evidence for our conclusions.

The external validity of our results for other countries than the US is probably rather limited, because every country and population has so many idiosyncrasies, that in most cases it is rather unreasonable to directly compare two cultures or countries to each other. However, we argue that there is at least some external validity for other states in the US, as we are using data from the whole US to synthetically construct our control groups. Moreover, we think that demographics and people's attitudes across different states in the US are probably similar enough to draw analog conclusions. Nevertheless would it be interesting to investigate the effect of legalizing recreation marijuana on traffic accidents in other states or countries which recently accepted an according law. However, if the required data for such approaches is not yet available or rich enough, there are also several possibilities to extend the analysis proposed within this work. One attempt could be to conduct the proposed DiD analyses with more restrictive donor pools for the SCG. For instance, one could only consider states where even medical marijuana is illegal or exclude all states which share a boarder with an area where marijuana can be purchased legally. Such attempts could be used to address the concern of potential spillover effects. Another way to extend our approach could be to analyse how the treatment effects change, when only accidents of a certain severity level are considered.[12] One could also try to enhance the accuracy of our estimates by searching additional or alternative data sets, which enable a longer pre-treatment period to generate the SCG. However, we leave this open to further research.

[12]The data set from Moosavi et al. (2019a) and Moosavi et al. (2019b) provides an indicator for accident severity. This indicator categorizes each accident incident into one of four severity categories according to its impact on traffic.

References

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic Control Methods for comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American statistical Association*, 105(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2011). Synth: An R package for Synthetic Control Methods in Comparative Case Studies. *Journal of Statistical Software*, 42(13).

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59(2):495–510.

Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings. *The Review of Economics and Statistics*, 60:47–57.

Ashenfelter, O. and Card, D. (1978). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. *The Review of Economics and Statistics*, 67(4):648–660.

ASIRT (2020). *ASIRT Road Safety Statistics*. Retrieved May 25, 2020, from https://www.asirt.org/safe-travel/road-safety-facts/.

Baer, D. J. (1973). Attitudes about Marijuana and political Views. *Psychological reports*, 32(3_suppl):1051–1054.

Ballotpedia (2020). *Marijuana on the Ballot*. Retrieved May 28, 2020, from https://ballotpedia.org/Marijuana_on_the_ballot.

Firpo, S. and Possebom, V. (2018). Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets. *Journal of Causal Inference*, 6(2).

Fishbain, D. A., Cutler, R. B., Rosomoff, H. L., and Rosomoff, R. S. (2003). Are Opioid-dependent/tolerant Patients impaired in driving-related Skills? A structured Evidence-based Review. *Journal of pain and symptom management*, 25(6):559–577.

Hansen, B., Miller, K., and Weber, C. (2017). The Grass is Greener on the Other Side: How Extensive is the Interstate Trafficking of Recreational Marijuana? Technical report, National Bureau of Economic Research.

Hansen, B., Miller, K., and Weber, C. (2020). Early Evidence on Recreational Marijuana Legalization and Traffic Fatalities. *Economic Inquiry*, 58(2):547–568.

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Central European Labour Studies Institute (CELSI), Bratislava, Slovakia. R package version 5.2.2.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of economic literature*, 47(1):5–86.

Mark Anderson, D., Hansen, B., and Rees, D. I. (2013). Medical Marijuana Laws, Traffic Fatalities, and Alcohol Consumption. *The Journal of Law and Economics*, 56(2):333–369.

Migoya, D. (2017). Exclusive: Traffic Fatalities linked to Marijuana are up sharply in Colorado. Is legalization to blame. *The Denver Post*, 27.

Moosavi, S., Samavatian, M. H., Parthasarathy, S., and Ramnath, R. (2019a). A Country-wide Traffic Accident Dataset. *arXiv preprint arXiv:1906.05409*.

Moosavi, S., Samavatian, M. H., Parthasarathy, S., Teodorescu, R., and Ramnath, R. (2019b). Accident Risk Prediction based on Heterogeneous Sparse Data: New Dataset and Insights.

National Conference of State Legislatures (2020). *State Medical Marijuana Laws*. Retrieved May 28, 2020, from https://www.ncsl.org/research/health/state-medical-marijuana-laws.aspx.

National Institute on Alcohol Abuse and Alcoholism (2019). *Surveillance Report 115: Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends, 1977–2018*. Retrieved May 13, 2020, from https://pubs.niaaa.nih.gov/publications/surveillance.htm.

Papalimperi, A. H., Athanaselis, S. A., Mina, A. D., Papoutsis, I. I., Spiliopoulou, C. A., and Papadodima, S. A. (2019). Incidence of Fatalities of Road Traffic Accidents associated with Alcohol Consumption and the Use of Psychoactive Drugs: A 7-year Survey (2011-2017). *Experimental and therapeutic medicine*, 18(3):2299–2306.

Salomonsen-Sautel, S., Min, S.-J., Sakai, J. T., Thurstone, C., and Hopfer, C. (2014). Trends in Fatal Motor Vehicle Crashes before and after Marijuana Commercialization in Colorado. *Drug and alcohol dependence*, 140:137–144.

Sanna, E. (2014). *Marijuana: Mind-Altering Weed*. Simon and Schuster.

Santaella-Tenorio, J., Mauro, C. M., Wall, M. M., Kim, J. H., Cerdá, M., Keyes, K. M., Hasin, D. S., Galea, S., and Martins, S. S. (2017). US Traffic Fatalities, 1985–2014, and their Relationship to Medical Marijuana Laws. *American journal of public health*, 107(2):336–342.

Sevigny, E. L. (2018). The Effects of Medical Marijuana Laws on Cannabis-involved Driving. *Accident Analysis & Prevention*, 118:57–65.

UN (1922). *UNTC Opium Convention*. Retrieved May 25, 2020, from https://treaties.un.org/Pages/ViewDetailsIV.aspx?src=TREATYmtdsg_no=VI-2chapter=6Temp=mtdsg4clang=_en.

U.S. Census Bureau (2019). *State Population Totals and Components of Change: 2010-2019*. Retrieved May 11, 2020, from https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.htmlpar_textimage_1574439295.

U.S. Department of Transportation's Federal Highway Administration (2015a). *Functional System Lane-Length - 2014*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2014/hm60.cfm.

U.S. Department of Transportation's Federal Highway Administration (2015b). *State Motor-Vehicle Registrations - 2014*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2014/mv1.cfm.

U.S. Department of Transportation's Federal Highway Administration (2016a). *Functional System Lane-Length - 2015*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2015/hm60.cfm.

U.S. Department of Transportation's Federal Highway Administration (2016b). *State Motor-Vehicle Registrations - 2015*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2015/mv1.cfm.

U.S. Department of Transportation's Federal Highway Administration (2017a). *Functional System Lane-Length - 2016*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2016/hm60.cfm.

U.S. Department of Transportation's Federal Highway Administration (2017b). *State Motor-Vehicle Registrations - 2016*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2016/mv1.cfm.

U.S. Department of Transportation's Federal Highway Administration (2018a). *Functional System Lane-Length - 2017*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2017/hm60.cfm.

U.S. Department of Transportation's Federal Highway Administration (2018b). *State Motor-Vehicle Registrations - 2017*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2017/mv1.cfm.

U.S. Department of Transportation's Federal Highway Administration (2019a). *Functional System Lane-Length - 2018*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2018/hm60.cfm.

U.S. Department of Transportation's Federal Highway Administration (2019b). *State Motor-Vehicle Registrations - 2018*. Retrieved May 13, 2020, from https://www.fhwa.dot.gov/policyinformation/statistics/2018/mv1.cfm.

U.S. Department of Transportation's Federal Highway Administration (2020). *Traffic Volume Trends*. Retrieved May 12, 2020, from https://www.fhwa.dot.gov/policyinformation/travel_monitoring/tvt.cfm.

World Population Review (2020). *United States by Area 2020*. Retrieved May 11, 2020, from https://worldpopulationreview.com/states/states-by-area/.

Zhao, X., Zhang, X., and Rong, J. (2014). Study of the Effects of Alcohol on Drivers and Driving Performance on straight Road. *Mathematical Problems in Engineering*, 2014.

Appendix

Table 6: Predictor means for cannabis legalization on monthly basis

	Treated	Synthetic	Sample Mean
Acc_PC	74.833	49.502	24.363
VMT_PC	758.710	675.285	888.808
Rural_Mileage_PC	5181.852	4006.069	26927.148
Urban_Mileage_PC	6357.953	9373.658	8895.002
RMV_PC	0.741	0.817	0.859
Pop_Density	251.335	918.991	249.361
Alc_Cons_PC	155567.438	172412.520	160271.772

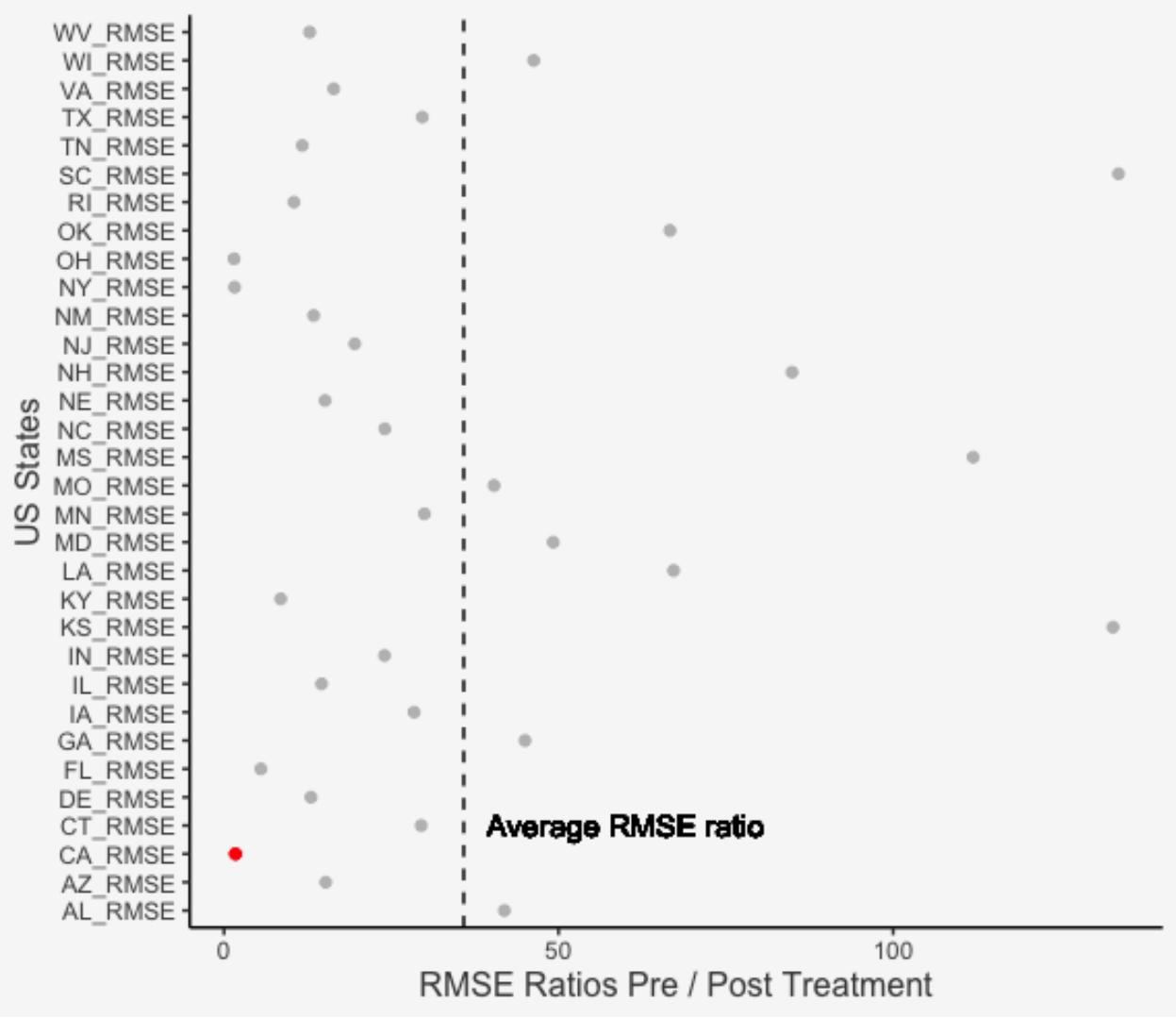

Figure 13: RMSE ratios for California and Synthetic Controls on monthly basis

Table 7: DiD estimation - Models (1) and (2) for cannabis legalization analysis

	Dependent variable:	
	Acc_PC	
	Model (1)	Model (2)
Constant	1.558^{***} (0.268)	1.558^{***} (0.120)
Post_Treat	6.009^{***} (0.298)	0.698^{***} (0.178)
Treat_Grp	0.995^{***} (0.379)	0.995^{***} (0.170)
Post_Treat_X_Treat_Grp	-0.847^{**} (0.422)	-0.182 (0.251)
Observations	2,846	1,018
R^2	0.199	0.070
Adjusted R^2	0.199	0.068
Residual Std. Error	4.445 (df = 2842)	1.997 (df = 1014)
F Statistic	236.061^{***} (df = 3; 2842)	25.620^{***} (df = 3; 1014)

Table 8: DiD estimation - Models (3) and (4) for cannabis legalization analysis

	Dependent variable:	
	Acc_PC	
	Model (3)	Model (4)
Constant	-2.135^{***} (0.296)	-0.946^{***} (0.354)
Post_Treat	1.030^{***} (0.285)	0.698^{**} (0.319)
Treat_Grp	0.995^{***} (0.308)	-1.383^{***} (0.501)
Post_Treat_X_Treat_Grp	-0.847^{**} (0.343)	-0.182 (0.452)
SRC_Bing	3.694^{***} (0.200)	2.504^{***} (0.280)
SRC_MQ	7.119^{***} (0.193)	7.257^{***} (0.271)
SRC_MQ_X_Treat_Grp		-0.278 (0.383)
SRC_Bing_X_Treat_Grp		2.378^{***} (0.397)
Observations	2,846	2,846
R^2	0.471	0.479
Adjusted R^2	0.470	0.477
Residual Std. Error	3.615 (df = 2840)	3.590 (df = 2838)
F Statistic	505.669^{***} (df = 5; 2840)	372.009^{***} (df = 7; 2838)

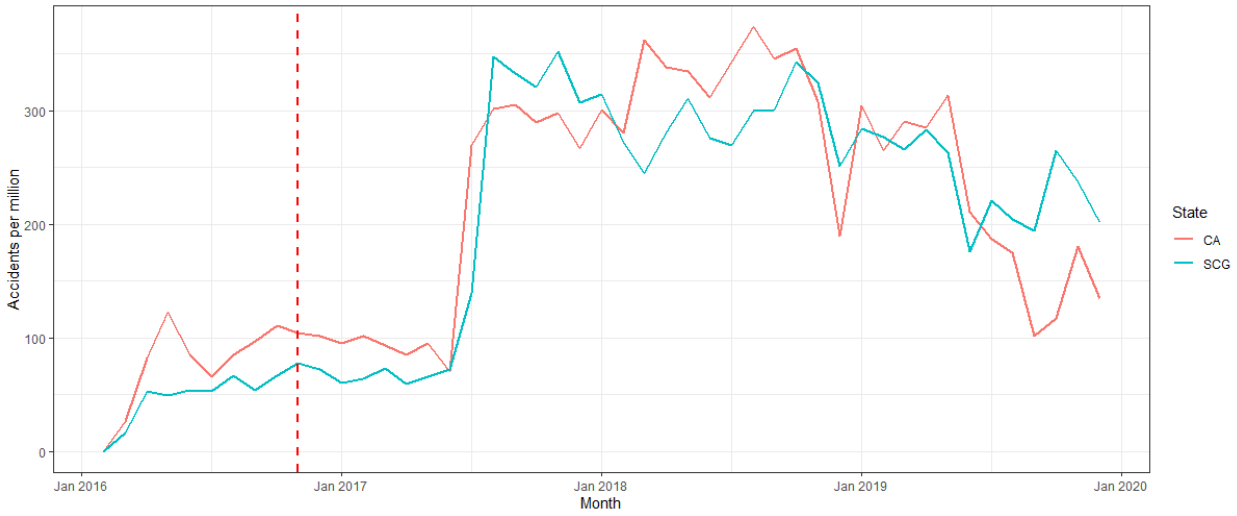

Figure 14: Monthly trends between California and its Synthetic counterpart on daily basis with cannabis legalization 09.11.16 (pre-treatment period Feb. 16 - Nov. 16)

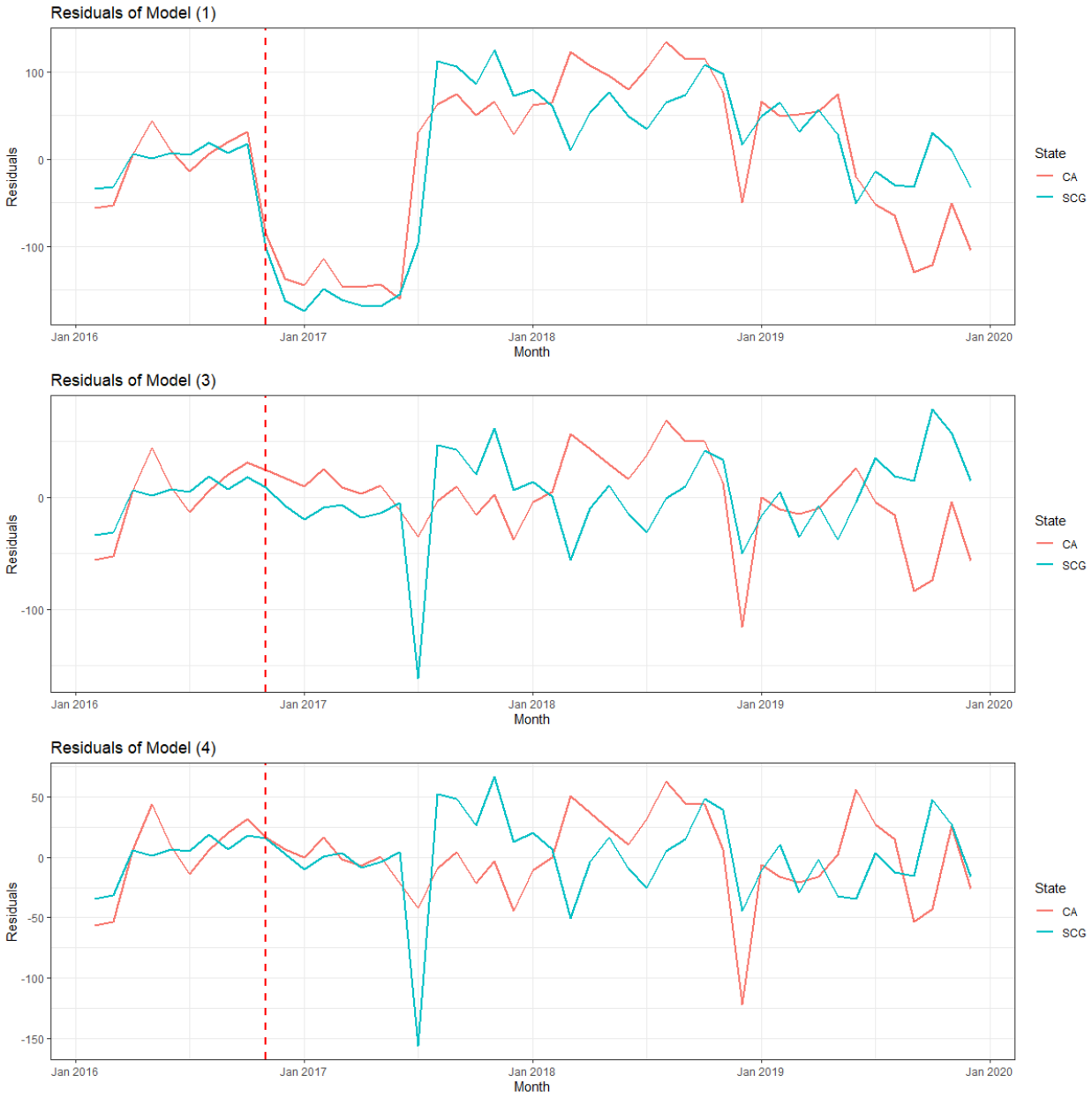

Figure 15: Residuals of Models (1), (3) and (4) for cannabis legalization analysis

Table 9: Predictor means for cannabis sales start on daily basis (pre-treatment period: Feb. 16 - Dec. 17)

	Treated	Synthetic	Sample Mean
Acc_PC	4.548	3.350	2.573
VMT_PC	25.252	25.636	29.482
Rural_Mileage_PC	4597.982	6008.148	26880.542
Urban_Mileage_PC	6328.586	8067.391	8932.743
RMV_PC	0.750	0.701	0.862
Pop_Density	252.055	373.211	249.858
Alc_Cons_PC	5225.150	5599.948	5397.208

Table 10: Predictor means for cannabis sales start on monthly basis (pre-treatment period: Feb. 16 - Dec. 17

	Treated	Synthetic	Sample Mean
Acc_PC	137.046	94.879	77.524
VMT_PC	760.848	734.753	888.312
Rural_Mileage_PC	4609.577	8201.436	26881.541
Urban_Mileage_PC	6328.391	6758.608	8932.044
RMV_PC	0.750	0.658	0.862
Pop_Density	252.040	476.511	249.849
Alc_Cons_PC	157436.043	149996.899	162620.216

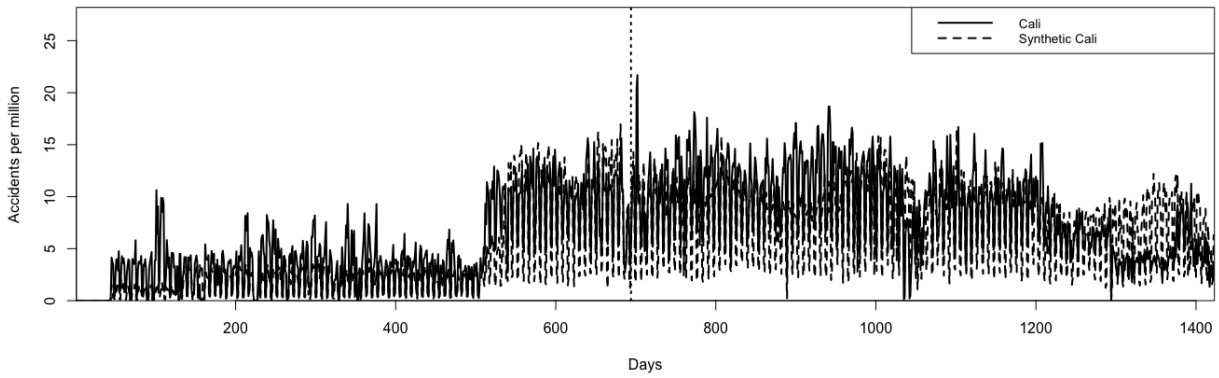

Figure 16: Daily trends between California and its Synthetic counterpart on daily basis with sales start 01.01.18

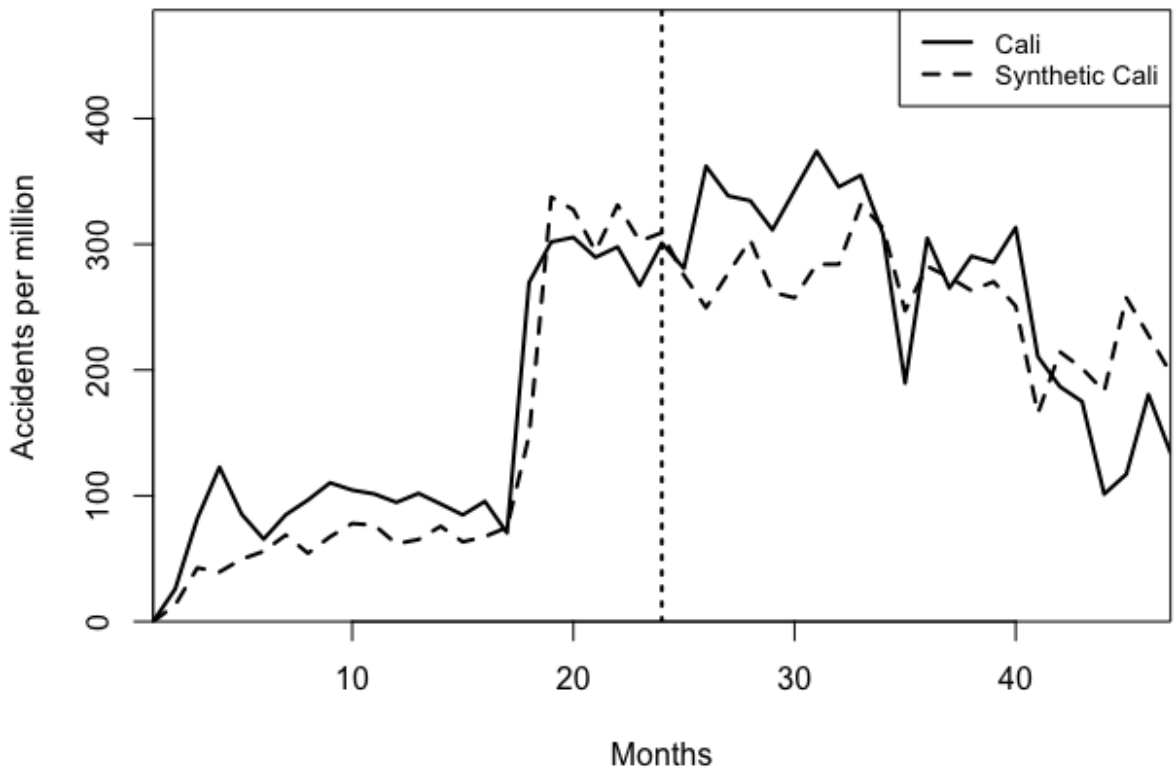

Figure 17: Monthly trends between California and its Synthetic counterpart on monthly basis with sales start 01.01.18

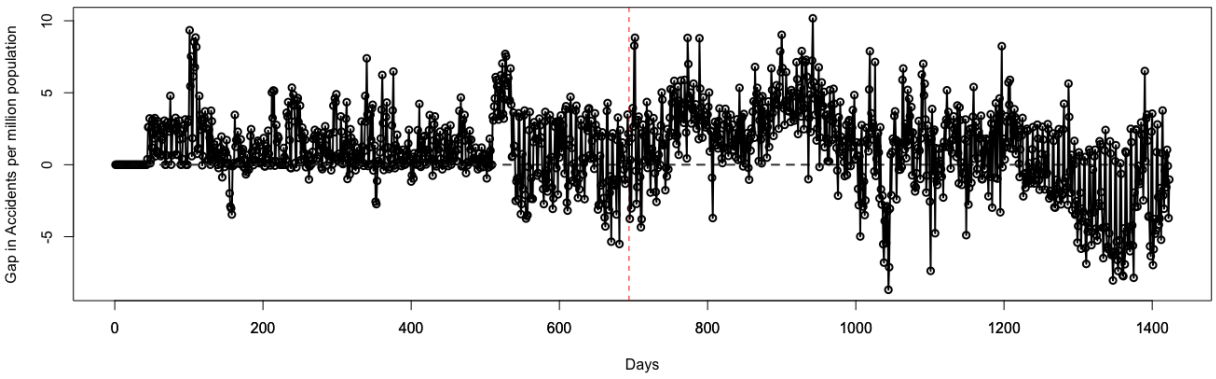

Figure 18: Gap in trends between California and its Synthetic counterpart on daily basis with sales start 01.01.18

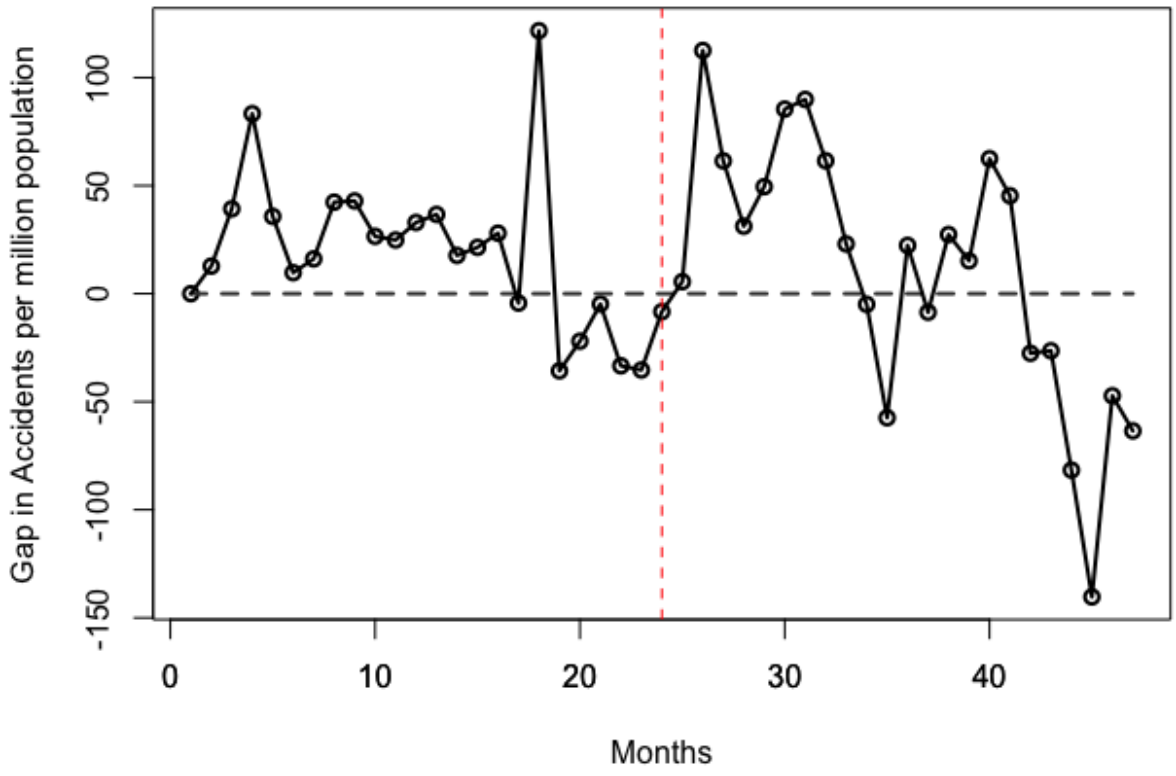

Figure 19: Gap in trends between California and its Synthetic counterpart on monthly basis with sales start 01.01.18

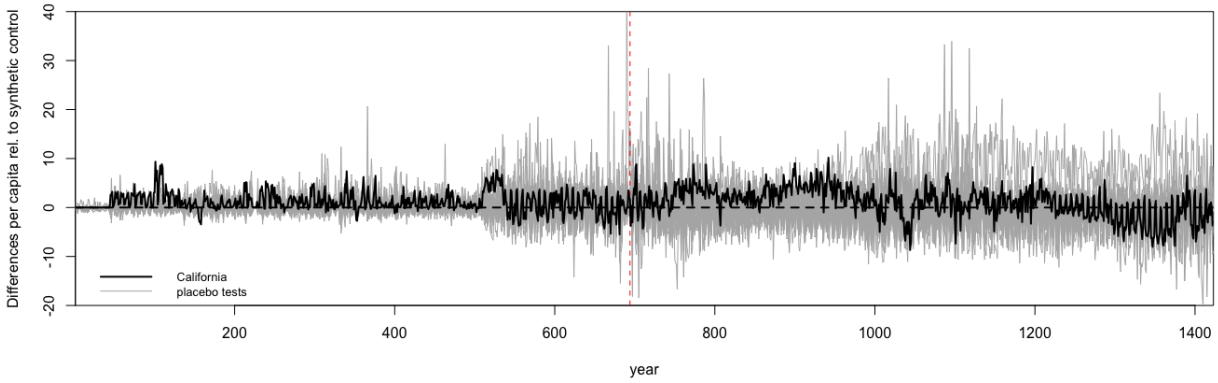

Figure 20: Placebo test runs for California (black) and Donor Pool (grey) on a daily basis with sales start 01.01.18

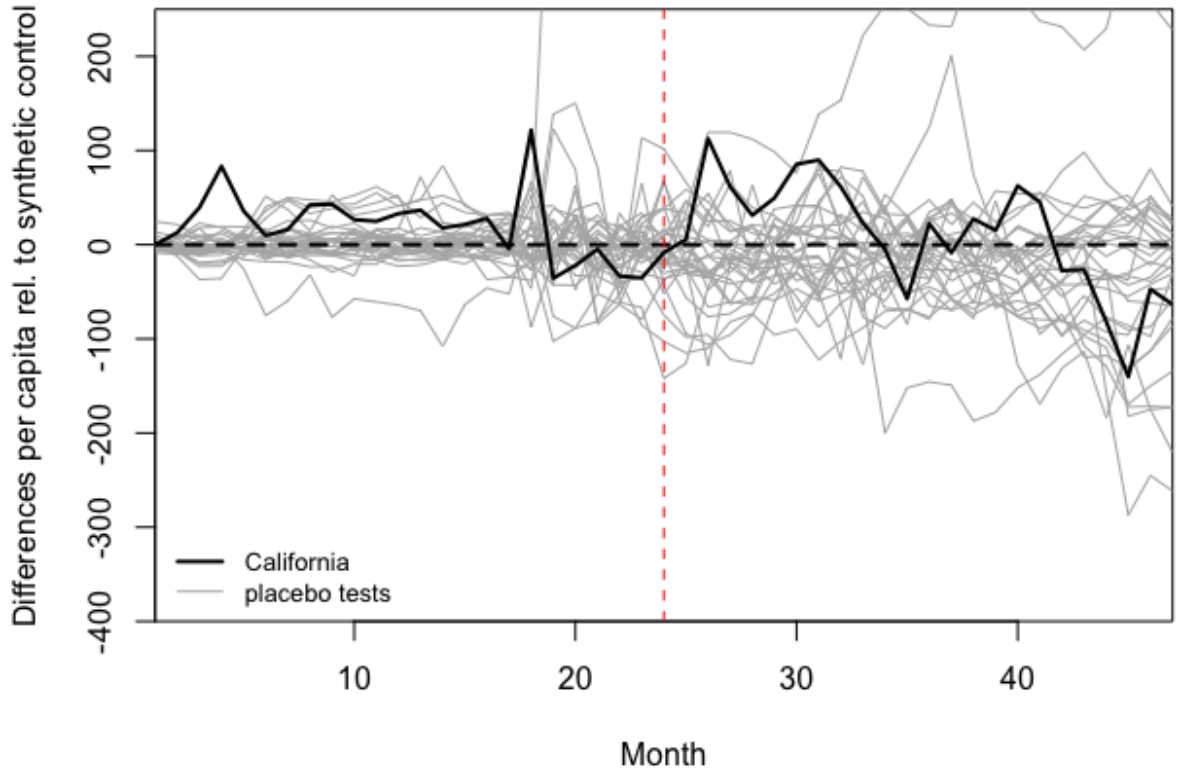

Figure 21: Placebo test runs for California (black) and Donor Pool (grey) on a monthly basis with sales start 01.01.18

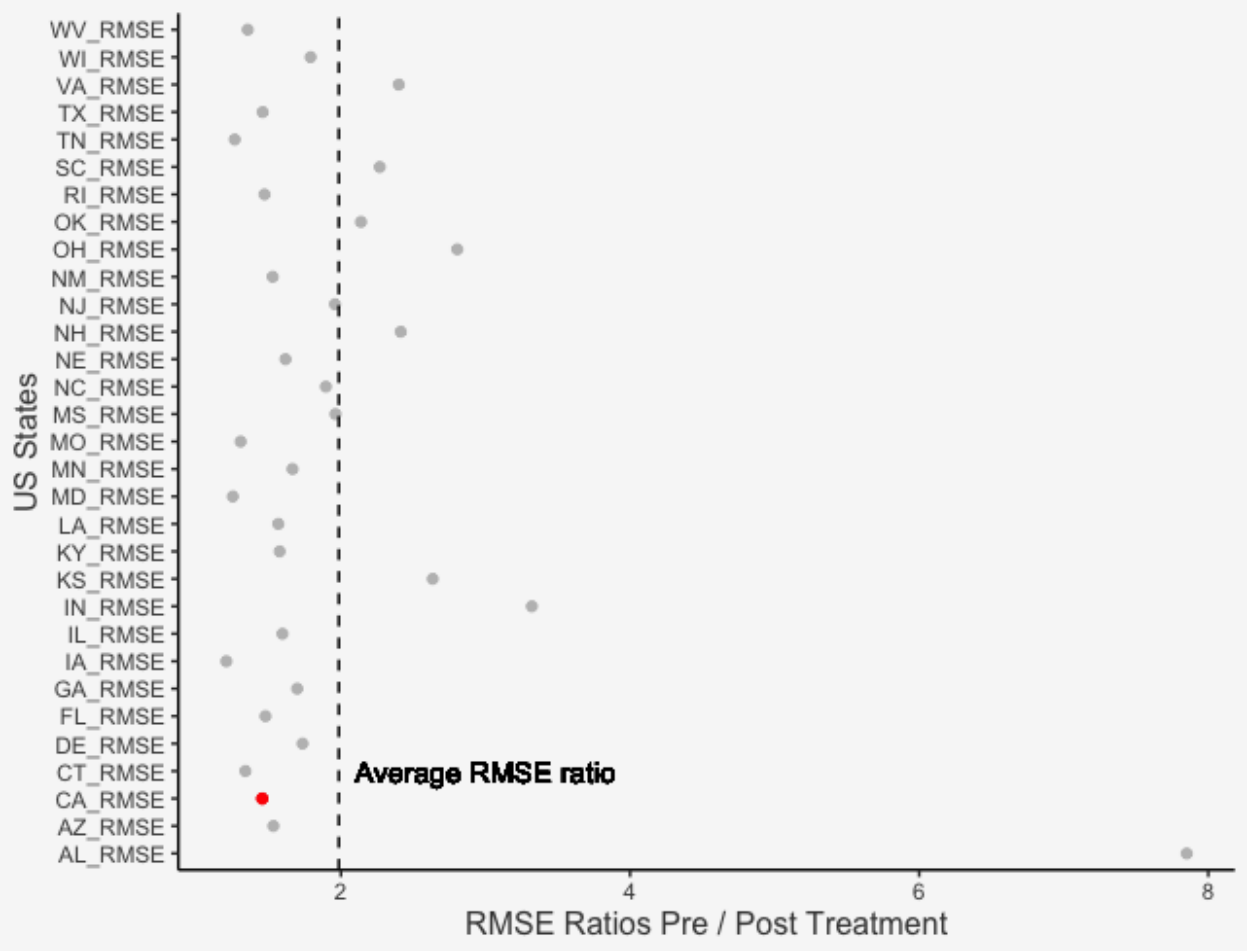

Figure 22: RMSE ratios for California and Synthetic Controls with sales start 01.01.18 and on daily basis

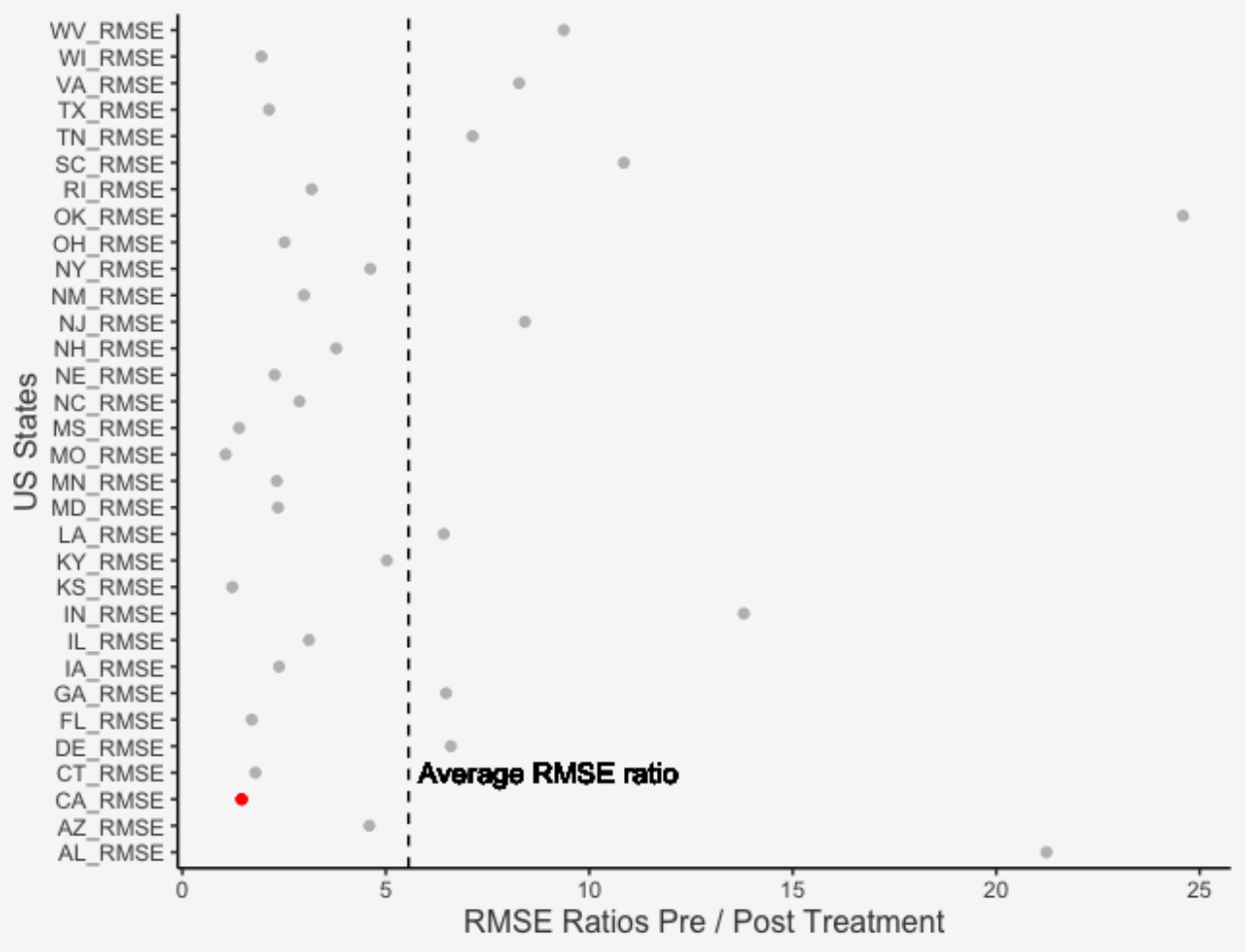

Figure 23: RMSE ratios for California and Synthetic Controls with sales start 01.01.18 and on monthly basis

Table 11: Did estimation - Model (6) for sales start analysis

	Dependent variable:
	Acc_PC
	Model (6)
Constant	3.354^{***} (0.148)
Post_Treat	4.317^{***} (0.206)
Treat_Grp	1.195^{***} (0.209)
Post_Treat_X_Treat_Grp	-0.088 (0.291)
Observations	2,846
R^2	0.245
Adjusted R^2	0.244
Residual Std. Error	3.885 (df = 2842)
F Statistic	307.575^{***} (df = 3; 2842)

Table 12: Did estimation - Models (7) and (8) for sales start analysis

	Dependent variable:	
	Acc_PC	
	Model (7)	Model (8)
Constant	-1.677^{***} (0.222)	0.160 (0.286)
Post_Treat	0.469^{**} (0.224)	-0.014 (0.265)
Treat_Grp	1.195^{***} (0.169)	-2.480^{***} (0.404)
Post_Treat_X_Treat_Grp	-0.088 (0.236)	0.879^{**} (0.375)
SRC_Bing	3.291^{***} (0.181)	1.472^{***} (0.251)
SRC_MQ	6.553^{***} (0.191)	6.485^{***} (0.265)
SRC_MQ_X_Treat_Grp		0.136 (0.375)
SRC_Bing_X_Treat_Grp		3.638^{***} (0.355)
Observations	2,846	2,846
R^2	0.507	0.525
Adjusted R^2	0.506	0.523
Residual Std. Error	3.141 (df = 2840)	3.085 (df = 2838)
F Statistic	584.123^{***} (df = 5; 2840)	447.459^{***} (df = 7; 2838)

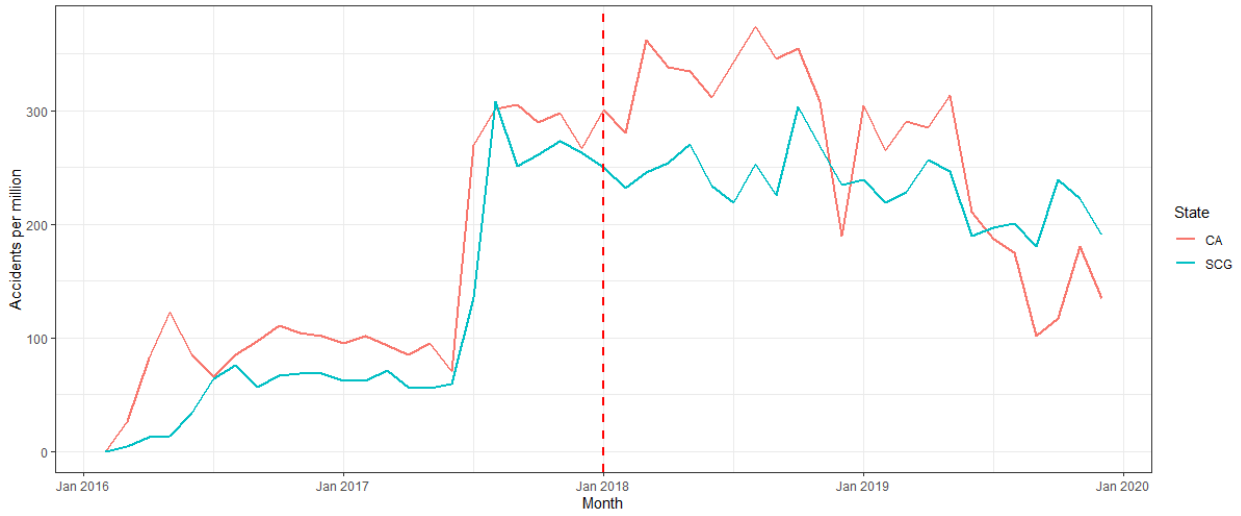

Figure 24: Monthly trends between California and its first Synthetic counterpart on daily basis with cannabis sales start 01.01.18 (pre-treatment period Feb. 16 - Dec. 17)

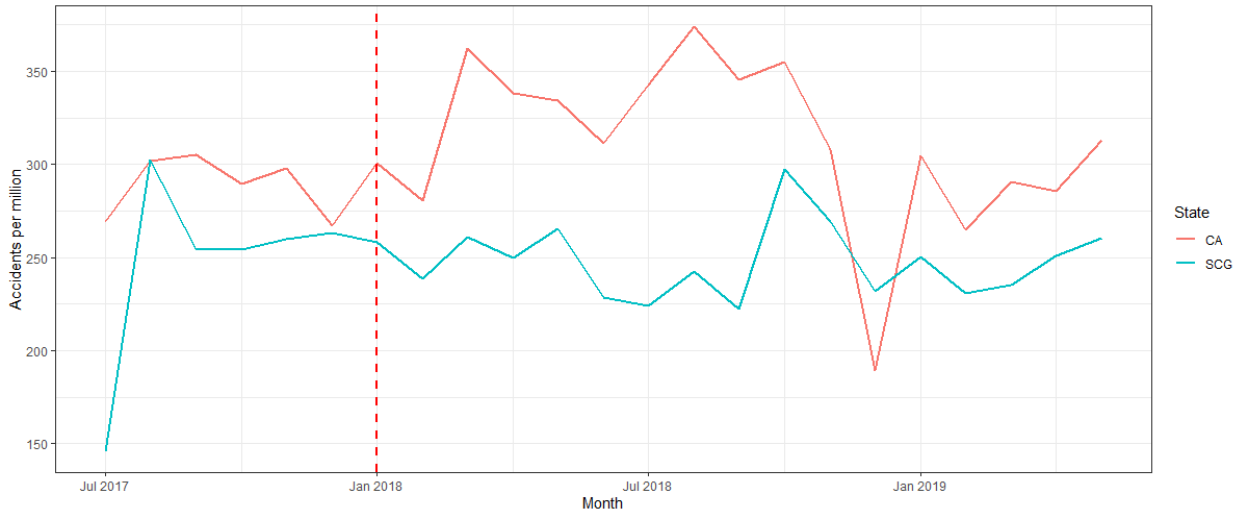

Figure 25: Monthly trends between California and its second Synthetic counterpart on daily basis with cannabis sales start 01.01.18 (pre-treatment period Jul. 17 - Dec. 17)

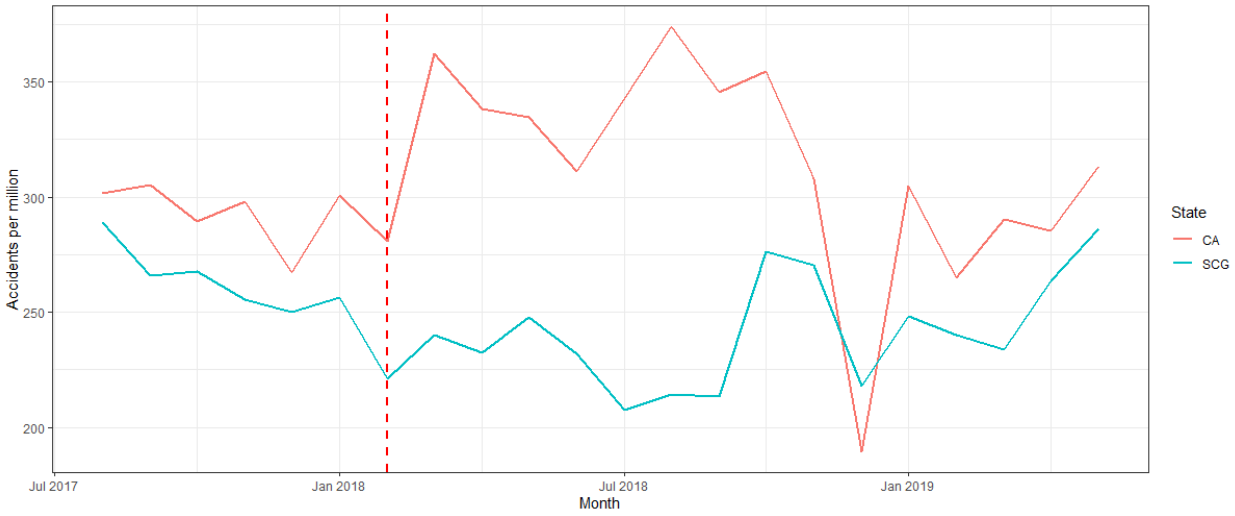

Figure 26: Monthly trends between California and its third Synthetic counterpart on daily basis with cannabis sales start 01.01.18 (pre-treatment period Aug. 17 - Dec. 17)

Table 13: Weights of the Synthetic Control for cannabis legalization analysis on daily basis

Weights	State	Unit Numbers
0.557	RI	26
0.275	NC	18
0.168	NY	23
0.000	AL	1
0.000	AZ	2
0.000	CT	4
0.000	DE	5
0.000	FL	6
0.000	GA	7
0.000	IA	8
0.000	IL	9
0.000	IN	10
0.000	KS	11
0.000	KY	12
0.000	LA	13
0.000	MD	14
0.000	MN	15
0.000	MO	16
0.000	MS	17
0.000	NE	19
0.000	NH	20
0.000	NJ	21
0.000	NM	22
0.000	OH	24
0.000	OK	25
0.000	SC	27
0.000	TN	28
0.000	TX	29
0.000	VA	30
0.000	WI	31
0.000	WV	32

Table 14: Weights of the Synthetic Control for cannabis legalization analysis on monthly basis

Weights	State	Unit Numbers
0.873	RI	26
0.127	NC	18
0.000	AL	1
0.000	AZ	2
0.000	CT	4
0.000	DE	5
0.000	FL	6
0.000	GA	7
0.000	IA	8
0.000	IL	9
0.000	IN	10
0.000	KS	11
0.000	KY	12
0.000	LA	13
0.000	MD	14
0.000	MN	15
0.000	MO	16
0.000	MS	17
0.000	NE	19
0.000	NH	20
0.000	NJ	21
0.000	NM	22
0.000	NY	23
0.000	OH	24
0.000	OK	25
0.000	SC	27
0.000	TN	28
0.000	TX	29
0.000	VA	30
0.000	WI	31
0.000	WV	32

Table 15: Weights of the first Synthetic Control for cannabis sales start analysis on daily basis (pre-treatment period: Feb. 16 - Dec. 17)

Weights	State	Unit Numbers
0.530	FL	6
0.335	NY	23
0.136	NC	18
0.000	AL	1
0.000	AZ	2
0.000	CT	4
0.000	DE	5
0.000	GA	7
0.000	IA	8
0.000	IL	9
0.000	IN	10
0.000	KS	11
0.000	KY	12
0.000	LA	13
0.000	MD	14
0.000	MN	15
0.000	MO	16
0.000	MS	17
0.000	NE	19
0.000	NH	20
0.000	NJ	21
0.000	NM	22
0.000	OH	24
0.000	OK	25
0.000	RI	26
0.000	SC	27
0.000	TN	28
0.000	TX	29
0.000	VA	30
0.000	WI	31
0.000	WV	32

Table 16: Weights of the first Synthetic Control for cannabis sales start analysis on monthly basis (pre-treatment period: Feb. 16 - Dec. 17)

Weights	State	Unit Numbers
0.479	MD	14
0.372	NY	23
0.148	SC	27
0.000	AL	1
0.000	AZ	2
0.000	CT	4
0.000	DE	5
0.000	FL	6
0.000	GA	7
0.000	IA	8
0.000	IL	9
0.000	IN	10
0.000	KS	11
0.000	KY	12
0.000	LA	13
0.000	MN	15
0.000	MO	16
0.000	MS	17
0.000	NC	18
0.000	NE	19
0.000	NH	20
0.000	NJ	21
0.000	NM	22
0.000	OH	24
0.000	OK	25
0.000	RI	26
0.000	TN	28
0.000	TX	29
0.000	VA	30
0.000	WI	31
0.000	WV	32

Table 17: Weights of the third Synthetic Control for cannabis sales start analysis on daily basis (pre-treatment period: Aug. 17 - Dec. 17)

Weights	State	Unit Numbers
0.796	NY	23
0.102	AZ	2
0.086	SC	27
0.014	NH	20
0.001	NC	18
0.000	AL	1
0.000	CT	4
0.000	DE	5
0.000	FL	6
0.000	GA	7
0.000	IA	8
0.000	IL	9
0.000	IN	10
0.000	KS	11
0.000	KY	12
0.000	LA	13
0.000	MD	14
0.000	MN	15
0.000	MO	16
0.000	MS	17
0.000	NE	19
0.000	NJ	21
0.000	NM	22
0.000	OH	24
0.000	OK	25
0.000	RI	26
0.000	TN	28
0.000	TX	29
0.000	VA	30
0.000	WI	31
0.000	WV	32

Table 18: Weights of the second Synthetic Control for cannabis sales start analysis on daily basis (pre-treatment period: Jul. 17 - Dec. 17)

Weights	State	Unit Numbers
0.392	FL	6
0.246	NY	23
0.211	MD	14
0.091	RI	26
0.060	SC	27
0.000	AL	1
0.000	AZ	2
0.000	CT	4
0.000	DE	5
0.000	GA	7
0.000	IA	8
0.000	IL	9
0.000	IN	10
0.000	KS	11
0.000	KY	12
0.000	LA	13
0.000	MN	15
0.000	MO	16
0.000	MS	17
0.000	NC	18
0.000	NE	19
0.000	NH	20
0.000	NJ	21
0.000	NM	22
0.000	OH	24
0.000	OK	25
0.000	TN	28
0.000	TX	29
0.000	VA	30
0.000	WI	31
0.000	WV	32