AN UPDATE ON THE SYNTHETIC CONTROL METHOD AS A TOOL TO UNDERSTAND STATE POLICY

Robert McClelland and Livia Mucciolo

May 2022

INTRODUCTION

In 2017, the Tax Policy Center (TPC) published "The Synthetic Control Method as a Tool to Understand State Policy," a guide for using the synthetic control method (SCM) as a quantitative adjunct to case studies (McClelland and Gault 2017). In this new report, we update the guide to account for more recent research. The method, developed in Abadie and Gardeazabal (2003) and popularized in Abadie, Diamond, and Hainmueller (2010), has been used to study a wide variety of topics.[1] Abadie, Diamond, and Hainmueller (2010) alone have been cited more than four thousand times (Google Scholar 2021).

The initial popularity of the method stems from its ability to evaluate treatments in which there is only one treated unit, typically a city, county, state, or other region, and no readily available control unit. Instead of an actual control unit, the method creates a synthetic control so that the treatment can be evaluated by comparing the outcome of the treated unit with the outcome of the synthetic control unit. The synthetic control is created by selecting and then weighting together a small number of control units drawn from a pool of potential donors. The weights are selected based on a set of predictors, with the objective of minimizing the distance between the outcome of the treated unit prior to treatment with the outcome of the synthetic control during that time. The effectiveness of the treatment is then measured as the differences between the synthetic control and the treated unit after treatment.

The SCM is also remarkably transparent and accessible. The small number of donors lets analysts determine if the makeup of the synthetic control is sensible. Sensitivity analyses can be conducted by selectively removing suspect donors and reestimating the synthetic control or by systematically removing each donor and reestimating. Weights for the predictors can be similarly reviewed by analysts. The availability of the software in Matlab, Stata, and R means this approach is easily accessible, not simply an interesting, but unused, empirical tool.

Since McClelland and Gault's (2017) publication, many articles have proposed improvements or extensions to the SCM. One improvement allows the SCM to better handle cases in which there are a very large pool of donors. Abadie, Diamond, and Hainmueller (2010) use the SCM to analyze the effects of California cigarette taxes on smoking, with the pool of donors as the 49 remaining states and the District of Columbia. The pool is then limited by excluding states that had enacted similar cigarette taxes, leaving a pool of 37 potential donors. But other analyses might have a pool with hundreds or even thousands of donors. Although for many estimation methods more information is better than less, in the SCM, a large number of donors can lead to low precision through overfitting and, in certain circumstances, eliminate a condition necessary for the synthetic control to be constructed from a small number of donors. Some new methods, such as singular value decomposition and thresholding, possibly accompanied by matrix completion (Amjad et al. 2018; Athey et al. 2020), reduce overfitting by reducing noise in the predictors. Abadie and L'Hour (2020) modify the SCM so that a small number of control units can be selected in cases in which the original version of the SCM cannot.

Other research proposes extensions that reduce possible biases in the SCM. Although not always followed, Abadie, Diamond, and Hainmueller (2015) do not recommend the use of the SCM if the outcomes of the synthetic control poorly fit the outcomes of the treated unit in the pretreatment period.[2] Abadie and L'Hour (2020) and Ben-Michael and colleagues (2020) independently offer one solution to the consequent bias. Ferman and Pinto (2021) offer an alternative solution that is less general but can be implemented by applying the original SCM to modified data.

Interpolation bias can also occur because the SCM implicitly assumes a linear relationship between the outcomes and the predictors. Abadie and L'Hour's (2020) approach reduces interpolation bias by penalizing the use of control units whose predictor values are far from those of the treated unit. Other research, such as by Kellog and colleagues (2020), suggests methods that reduce interpolation bias but allow for another problem, extrapolation bias.

Another improvement has been to model selection and inference. Because the SCM is transparent, it is more difficult to select a model that yields a specific result. But Ferman, Pinto, and Possebom (2020) suggest that results can still be cherry-picked to some degree, and they offer some guidance for reducing the scope of the problem. Firpo and Possebom (2018) extend inference in several ways.[3] First, they create a confidence set by inverting the p-values at each point in the posttreatment set. Second, they describe how to use null hypotheses beyond "no effect." Finally, they adapt a method for creating confidence intervals from p-values to the SCM.

A new but important extension of the SCM is the ability to cover multiple treated units. Numerous papers have contributed to this approach. Athey and colleagues (2020) and Xu (2017) fit SCM and the difference-in-difference method within a larger framework of models. Arkhangelsky and colleagues (2020) create new estimators that blend the SCM and the standard difference-in-differences estimator. Some, such as Donohue, Aneja, and Weber (2019) and Abadie and L'Hour (2020), estimate the SCM separately on different treated units and average the results across treated units. Ben-Michael and colleagues (2019) propose to combine that approach with one that estimates a set of weights for the pool of all treated units.

In this report, we update McClelland and Gault (2017) by reviewing selected articles on some of these improvements. We also update the step-by-step guidance for implementing the SCM described in McClelland and Gault (2017) to incorporate this new research. Our report shares some similarities with Abadie (2020) because both discuss recent updates and extensions of the SCM. To ensure our report's accessibility to a wider audience, we focus more on verbal and graphical intuition than mathematical explanation. Further, we do not cover the extension to multiple treated units because it is unclear how they complement the case study methodology. We also do not review methods that focus on extrapolation bias because those methods allow control units to have negative weights, or combined weights that do not have to equal one. Although they can reduce interpolation bias, they no longer result in a synthetic control created from a small number of control units, which reduces the transparency of the SCM. Finally, singular value decomposition and thresholding

methods are both new to the literature and complex, making it difficult to maintain the simplicity and transparency of the SCM. As such, we do not cover these two topics in this report.

THE SYNTHETIC CONTROL METHOD

As described in McClelland and Gault (2017), the SCM can be used when there is only one treated unit, such as a municipality or a state that raises the minimum wage, raises or lowers taxes, or enacts gun control laws. Many articles using the SCM describe its use to study rare or unique events. Abadie, Diamond, and Hainmueller (2015) examine the economic cost of the reunification of East and West Germany. Hankins (2020) examines Nebraska's unique switch from a bicameral house legislature to a unicameral house. In both cases, the treatment is completely unique, and in the case of the reunification, it is difficult to think of treatment applying to any other nation. Other studies, such as McClelland and Iselin (2019), study relatively rare changes in policy.

The SCM combines the clarity and transparency of a case study with the statistical inference that comes from econometrics methods such as regression analysis. It is an ideal companion to a case study because it can provide an empirical component performed by knowledgeable analysts. These analysts can choose predictors that are closely associated with the outcomes and, for transparency, list these predictors in their study. The SCM provides a list of units used to build the synthetic control and the share they contributed to the synthetic control. Those same subject matter experts can review the list and determine if the donor units are sufficiently similar to the treated unit, and all of this information can be made available in any report produced. Moreover, the SCM provides both graphical information about the relative strength of the treatment and a form of testing for statistical significance.

The basic idea of the SCM is to create a variable that represents the outcomes of the treated unit in the posttreatment periods, but in a scenario in which the treatment never occurred. Ideally, we would like to compare the outcomes of the treated unit in all of the posttreatment periods with the outcomes of the treated unit had the treatment not been applied. The SCM creates a synthetic control, whose posttreatment outcomes can be viewed as a stand-in for the outcomes of the treated unit had the treatmentnever occurred. Then, comparing the actual outcomes with the outcomes of the synthetic control allows us to estimate the treatment's effect. If the outcomes are similar, the treatment does not appear to have affected the outcome. If, on the other hand, the outcomes are substantially different, the treatment presumably caused the difference.

The SCM creates a synthetic control unit by matching predictors from a set of donor units chosen from a pool of potential candidates. The predictors from the donor units can be outcomes from some or all of the observations in the pretreatment, or they can include nonoutcome predictors from the pretreatment period (which we refer to here as covariates). Kaul and colleagues (2016) describe the SCM approach as a two-step optimization procedure with an "outer" and "inner" layer. In the outer layer, weights for the predictors are chosen to minimize the mean square prediction error (MSPE) between the predictors of the synthetic control and the predictors of the treated unit. Articles aimed at improving the SCM usually focus on the inner layer. In

it, weights for the donor units are chosen to minimize the MSPE between pretreatment outcomes of the synthetic control and the treated unit. In most studies, predictors include at least some of the outcomes in the pretreatment period, and in many studies, all of the outcomes in the pretreatment period are used. Kaul and colleagues (2016) has shown that in the latter case the outer optimization will leave all covariates unweighted, so analysts must choose between using all pretreatment outcomes and including covariates.

Conditions for Using the SCM

Several conditions are necessary for the method to accurately estimate the effect of a treatment. First, no unit in the donor pool can have a similar policy change. In principle, this restriction can dramatically reduce the donor pool. But as described below, studies usually have too many rather than too few potential donor units.

Second, the treatment cannot affect the outcome in the pool of donor units. This is the equivalent of the stable unit treatment value assumption for the case in which a single unit received treatment. For example, in the Abadie, Diamond, and Hainmueller (2010) study on the effect of an increase of cigarette taxes on cigarette purchases per capita, cigarette purchases in border states, such as Nevada or Arizona, are assumed not to change. If potential control units are suspected of being affected by the treatment, they should be removed from the donor pool. It is still possible to estimate the amount of spillover by using the SCM on the potentially contaminated unit as though it were the treated unit. If there are several potentially contaminated units, the SCM can be run on each individually and the results averaged.

Third, the values of the predictors for the treated unit should be near or inside the convex hull of the values for the donor pool. This is necessary because in the standard SCM, the weights used to construct the synthetic control must be nonnegative and sum to one. In Figure 1, blue points C_1 through C_5 represent the values of predictors A and B for five units in the pool of donors. The convex hull of the points are the dotted lines around points C_1 through C_4. The predictors for treated unit T_1 are represented by a red point. They are near the convex hull and so satisfy the conditions necessary for using the SCM. A synthetic treated unit $\hat{T}_1$ represented by a white point has predictors that make up a linear combination of C_1 and C_2 and do a good job approximating the predictors of the treated unit. Moving along the dotted line of C_1 and C_2 represents different combinations of weights that form the synthetic treated unit $\hat{T}_1$. The red point representing treated unit T_2 is entirely inside the convex hull, so it satisfies the necessary conditions, but as discussed in the appendix, it can make the SCM less useful. The red point representing predictors for T_3 has a value for predictor B that is quite near C_2 but it has a higher value for predictor A than any of the control units. If weights must be nonnegative and sum to one, then we are limited to the dotted line between C_2 and C_3, and it is not possible to closely approximate this treated unit. Thus, it is not a good candidate for estimating treatment with the SCM.

FIGURE 1
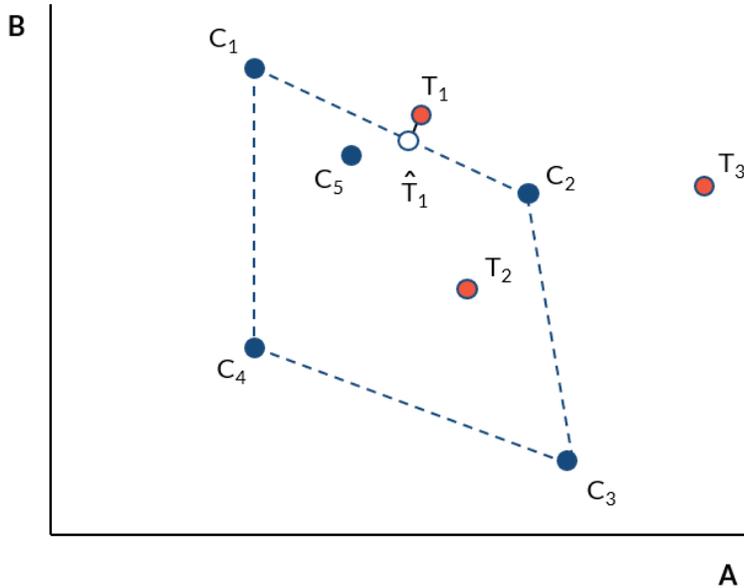
In addition, outcomes must be available for periods before the treatment and at least one period after treatment for both the treated unit and the pool of potential donor units. Abadie, Diamond, and Hainmueller (2010) recommend that the SCM not be used if only a few pretreatment periods are available. This is particularly important if the units in the donor pool have relatively large idiosyncratic shocks or if the donor pool is large relative to the number of pretreatment time periods. As we describe below, it is more likely that the model will be overfit in that the selected donors could be chosen because of transitory shocks rather than because the underlying values or trends in the predictors match the treated unit in the pretreatment period.

Further, the policy change must have no effect before it is enacted. This may not always be the case. For example, California smokers might hoard packs of cigarettes prior to a cigarette tax increase, or alcohol drinkers in Illinois might stock up on bottles of alcohol before a liquor tax increase. This is usually a small concern because if researchers suspect this, treatment can be assumed to have happened in an earlier period, and the SCM reestimated using the earlier date.

The treated state's counterfactual outcome can be approximated by a fixed combination of donor states. Similar to when a small number of pretreatment periods exist, Abadie, Diamond, and Hainmueller (2015) recommend that the SCM be avoided if the synthetic control does not closely match the treated unit in the pretreatment period. Ben-Michel and colleagues (2020) and Ferman and Pinto (2020) extend the SCM to explicitly handle instances in which the model fit is poor.

Finally, the predictors must have values for the donor pool regions that are similar to those of the affected region. This reduces the potential for interpolation bias, which can arise if donor units are chosen with outcomes far from the treated unit. Kellogg and colleagues (2020) show that some degree of interpolation bias exists unless the conditional mean of the outcomes is linear in pretreatment predictors. As we discuss below, Abadie and L'Hour (2020) describe a modification of the SCM that reduces interpolation bias.

The Abadie, Diamond, and Hainmueller (2010) Study

The quintessential use of the SCM is in Abadie, Diamond, and Hainmueller's (2010) study, as described by McClelland and Gault (2017):

> In that article, the authors examine how California's tobacco control program under Proposition 99, implemented in 1988, affected smoking by creating a synthetic control version of California. They estimate that by 2000, per capita sales of cigarette packs had fallen 26 packs because of the program. Here, we describe the steps in analyzing a policy change with the SCM using the Proposition 99 example when possible. We also use the original data in Abadie, Diamond, and Hainmueller (2010) to demonstrate the results' sensitivity to various modeling choices.

> California voters approved Proposition 99 in 1988, an initiative that raised the cigarette excise tax by 25 cents a pack and implemented a large-scale antitobacco media campaign. The tax raised $100 million annually, and the revenue was initially directed toward antismoking efforts, including antismoking education budgets. Those efforts were substantially larger than efforts in other states, but the California assembly passed Assembly Bill 99 in 1991, which diverted a large share of the tax revenue for other purposes (Glantz and Balbach 2000). Beyond the tax increase, Abadie, Diamond, and Hainmueller (2010) report that Proposition 99 led to numerous local ordinances prohibiting smoking in indoor spaces such as restaurants and workplaces, and by 1993 almost two-thirds of employees in California worked in smoke-free environments. They also note that tobacco lobbyists in California responded by spending 10 times more in 1991 and 1992 than in 1985 and 1986.

Abadie, Diamond, and Hainmueller (2010) determine the impact of Proposition 99 on per capita cigarette packs using data from 1970 through 1988 as the pretreatment period and 1989 through 2000 as the posttreatment period. They use four covariates averaged over the pretreatment years: the average retail price of cigarettes, the share of the population between the ages of 15 and 24, per capita beer consumption, and the log of per capita gross domestic product (GDP). They also use per capita annual sales of cigarette packs for the years 1975, 1980, and 1988. Abadie, Diamond, and Hainmueller (2010) start with the 49 states other than California, dropping the District of Columbia. After eliminating states that also raised cigarette taxes by at least 50 cents before or after treatment, they were left with a pool of 38 potential donors.

The results are shown in figure 2. Cigarette sales per capita in the synthetic control carefully match cigarette sales in actual California until 1988, the year cigarette taxes were increased. From that point forward, the two paths diverge. Although both the synthetic control and actual California show a continued decline in cigarette sales per capita, the decline is faster in actual California, with the difference providing an estimate of the effect of Proposition 99.

Synthetic California Per Capita Cigarette Sales (ADH 2010)

Before and after Proposition 99 passage in 1988
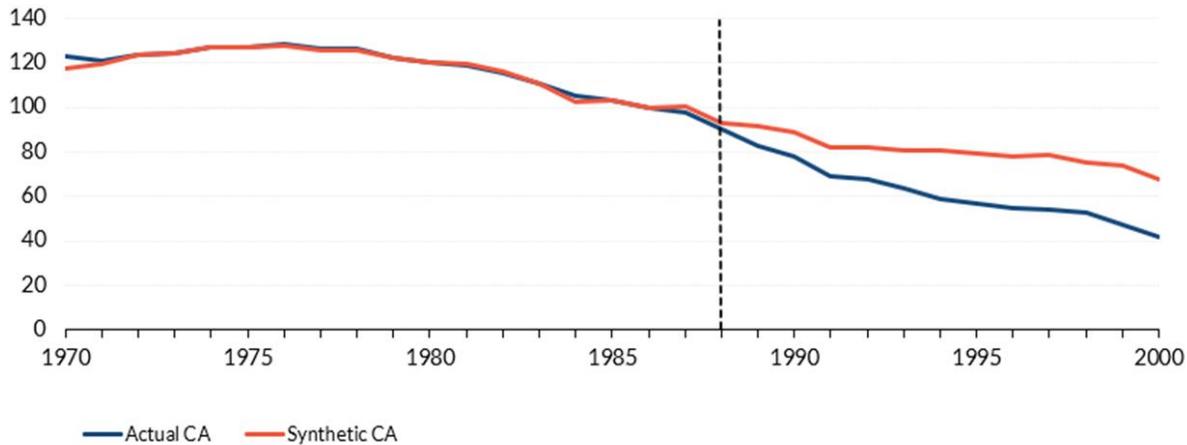
Cigarette pack sales

Source: Authors' calculations using the Synth package and data from ADH to replicate ADH's analysis.
Note: ADH = Abadie, Diamond, Hainmueller (2010).

To further demonstrate the fit, Abadie, Diamond, and Hainmueller (2010) provide two tables. In the first, they compare the predictors of the actual California with its synthetic control. As shown in table 1, the average values for the synthetic control's predictors closely match those of California. If the predictors affect the outcome, this provides evidence that the synthetic control mimics the behavior of the treated unit in the absence of treatment. Table 2 shows the weights of each state used in the synthetic control. Nevada and Utah contribute the most, constituting more than 50 percent of the total. Montana makes up an additional 20 percent, and Colorado and Connecticut make up about 22 percent combined.

Actual and Synthetic California Predictor Means (ADH 2010)

Variable	Years	Actual CA	Synthetic CA
Beer consumption per capita	1984–88	24.28	24.21
Log state per capita GDP	1980–88	10.08	9.86
Retail price of cigarettes	1980–88	89.42	89.41
Share of state population ages 15-24	1980–88	0.17	0.70
Cigarette sales per capita, 1988	1988	90.1	91.64
Cigarette sales per capita, 1980	1980	120.2	120.45
Cigarette sales per capita, 1975	1975	127.1	127.06

Source: McClelland and Gault (2017) using the Synth package and data from ADH to replicate ADH's analysis.
Notes: ADH = Abadie, Diamond, and Hainmueller (2010). Units are gallons for beer consumption per capita, cents for retail price of cigarettes, and packs for cigarette sales per capita.

Figure 3 plots the estimated effect of Proposition 99 as the difference between the per capita cigarette sales in actual California and the synthetic control, along with the analogous difference for 37 placebos.[1] Before 1988, the distance between actual California and the synthetic california is smaller than the analogous gaps for the placebos. This may happen because the variables were chosen to fit best for California, not the other states. The line at the very bottom of the figure represents New Hampshire, which Abadie, Diamond, and Hainmueller (2010) note has the highest per capita cigarette sales in each year before 1988. After 1988, the estimated effect for California is greater than the estimate effect for most placebo states. This strongly suggests that the estimated effect for California is, in fact, the result of Proposition 99. Abadie, Diamond, and Hainmueller (2010) refine this approach by only including states with a pretreatment MSPE that is 20 times that of California, five times that of California, and twice that of California.

TABLE 2
Synthetic California Donor State Weights (ADH 2010)

State	Weight
Colorado	0.161
Connecticut	0.068
Montana	0.201
Nevada	0.235
Utah	0.335
Sum	**1.000**

Source: McClelland and Gault (2017) using the Synth package and data from ADH to replicate ADH's analysis.
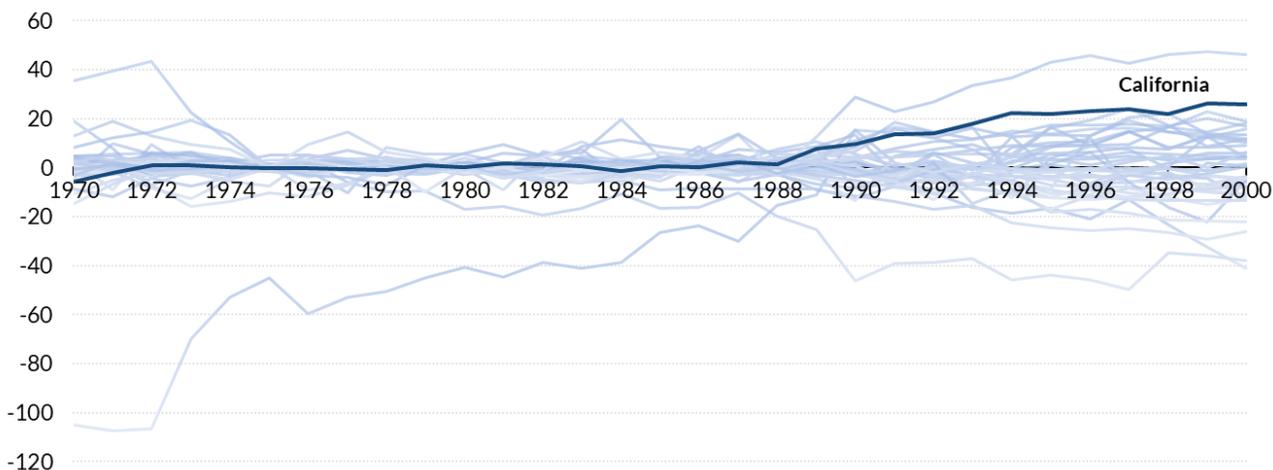**Note:** ADH = Abadie, Diamond, and Hainmueller (2010).

---

[1] Utah cannot be estimated as a synthetic control because it has the lowest per capita sales of cigarettes.

FIGURE 3

Difference in Synthetic and Actual Per Capita Cigarette Sales for California and 37 Placebo States

Before and after Proposition 99 passage in 1988

Cigarette pack sales

Source: Authors' estimations using the Synth package and data from Abadie, Diamond, and Hainmueller (2010).
**Note:** The Synth package cannot generate a synthetic placbeo for Utah, one of the 38 donor pool states.

In addition to visually inspecting plots such as figure 3, Abadie, Diamond, and Hainmueller (2010) use the ratio of the posttreatment MSPE to the pretreatment MSPE as a measure of the effectiveness of Proposition 99. A large difference between the output of the actual California and its synthetic control suggest that the treatment was effective, and normalizing by the pretreatment MSPE ensures that a large MSPE is not simply because of a poor fit. A larger MSPE ratio for the treated unit than the ratio for the placebos is further evidence that the treatment was effective. California's ratio is about 130, while the largest ratio for a placebo is less than 100, and most ratios are less than 20. In an informal measure of statistical significance, Abadie, Diamond, and Hainmueller (2010) note that if the treatment were randomly assigned, the probability of obtaining a ratio as large as California's is 1/39, or about 2.6 percent.

Here we introduce a new method to help assess the value of predictors used to create the synthetic control. In figure 4, we plot for each predictor the difference between the value for the states and the value for California. To make the predictors comparable to each other, the differences of each predictor have been adjusted to have a unit standard deviation. The blue dot at zero for each predictor represents California. The gray dots represent the difference between the predictor of a state in the donor pool and California, and the red dots are the values of the predictors for states selected as donors.

Differences between Average Predictor Values of Control Units and the Treated Unit in Abadie, Diamond, and Hainmueller (2010)
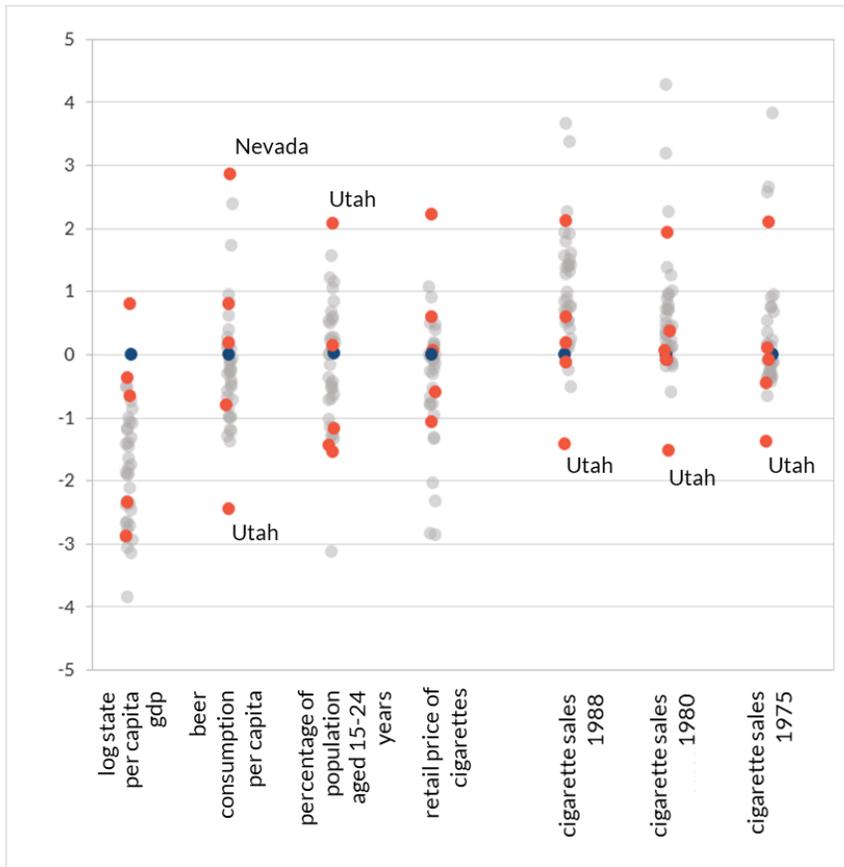
Source: Authors' calculations using the Synth package and data from ADH to replicate Abadie, Diamond, and Hainmueller (2010).

Figure 4 shows that log state per capita GDP may not be a good choice for a predictor. Ideally, states with values similar to California's values exist, as well as larger and smaller values. But in this case, no other states with values similar to California exist, and only one state with a value higher than California's value is present. The other predictors appear to be better choices because California's value lies among those of the donor pool. But notice that the trend in cigarette sales indicates that prior to 1988, cigarette sales in California were falling relative to the sales in other states before the policy change. Although it bears further investigation, it does not imply that no states followed trends similar to California.

Figure 5 shows that the trends in cigarette sales for two of the selected states, Nevada and Utah, match California's trend. In the figure, sales are indexed to equal 1 in 1975, the first of the three years in which cigarette sales are used as predictors. The trends in the two states bracket California's trend, and by 1988, all three states experienced about a 30 percent decline in cigarette sales. Connecticut, in contrast, has the same average sales in the pretreatment period, but those sales increase over a period and show only a modest decline by 1988.
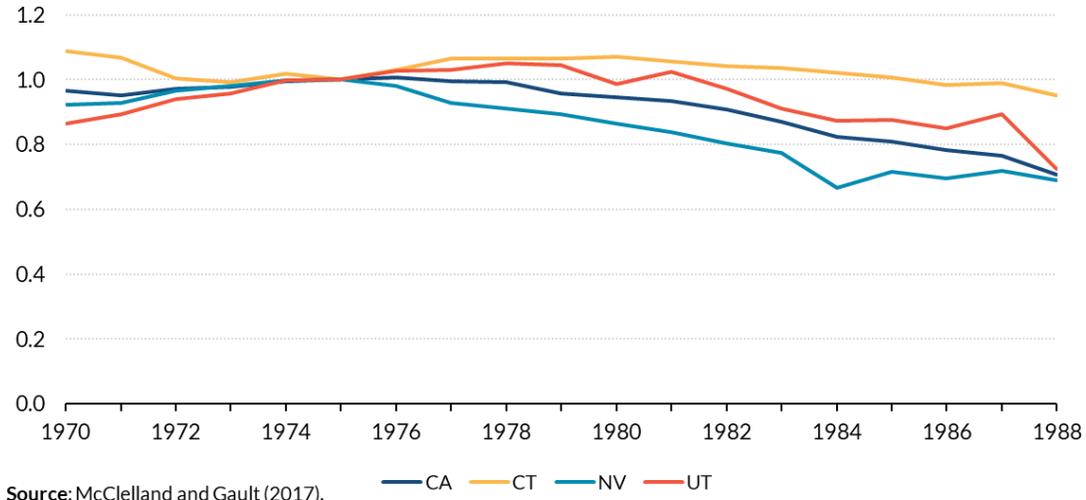
FIGURE 5

Per Capita Cigarette Sales in California and Selected Donor States Indexed to 1 in 1975

Before Proposition 99 passage in 1988

Cigarette pack sales indexed to 1 in 1975

Source: McClelland and Gault (2017).

Legend: CA CT NV UT

SELECTED RECENT ADVANCES

Reducing Interpolation Bias

Interpolation bias can arise if a nonlinear relationship between the predictors and the outcome exists. The outcome of the synthetic control is formed as a weighted average of outcomes from donor units in which the weights are chosen by minimizing the distance between a weighted average of their predictors and the predictors of the treated unit. Abadie, Diamond, and Hainmueller (2010) show that when the outcome and the predictors are related through a linear factor model, the bias of the synthetic control is bounded by a function that goes to zero as the number of pretreatment periods increase. But if the relationship is nonlinear, the proof does not hold and interpolation bias increases with the distance between donor's predictors and those of the treated state. Both biases are explained in more detail in the appendix.

Figure 4 demonstrates how the SCM can select extreme values, which can exacerbate interpolation bias. As noted, Connecticut is one of the states used to form a synthetic control for California, and the Connecticut value for log per capita GDP is the largest positive distance from California's value. For beer consumption per capita, states with the largest positive and negative distances, Nevada and Utah respectively, were used in the synthetic California. Utah has the largest positive distance for percentage of the population between 15 and 24 years of age. Connecticut again has the largest positive distance for the retail price of cigarettes. Utah has the largest negative distance for cigarette sales in 1988, 1980, and 1975. Despite their extreme values, Nevada and Utah had the two largest donor state weights within the synthetic control.

It has often been advised (Abadie, Diamond, and Hainmueller 2010, Abadie, Diamond, and Hainmueller 2015, Abadie 2020) to restrict the donor pool to those units whose predictors are similar to those of the treated unit. Abadie, Diamond, and Hainmueller (2010) suggests that "researchers trying to minimize biases caused by interpolating across regions with very different characteristics may restrict the donor pool to regions with similar characteristics to the region exposed to the event or intervention of interest."

Figure A.2 extends figure 4 to discuss a graphical method for evaluating the effect of deleting units with predictors that are far from the treated unit. The idea is to display the results of progressively limiting the pool of potential donors to those with, for example, predictors less than 3 standard deviations from the treated unit, 2.5 standard deviations from the treated unit, and so on. In the section on inferences and testing, we describe how reducing the pool too far can adversely affect how p-values are constructed.

Acting on a suggestion in Abadie, Diamond, and Hainmueller (2015), Abadie and L'Hour (2020) develop a solution that does not reduce the pool of donor units. Instead, it penalizes the use of control units whose predictor values are far from those of the treated unit.[4] As described above, the SCM chooses weights for the donor units to minimize the MSPE between the predictors of the synthetic control and the predictors of the

treated unit. Abadie and L'Hour (2020) add another term to the minimization problem that represents the weighted square of the difference between each unit's predictors and those of the treated unit. As a consequence, a unit that sharply lowers the MSPE between the synthetic control and treated unit's predictors may nevertheless receive no weight if its predictors are too far from those of the treated unit. The trade-off between the two objectives is controlled by a parameter λ, where a value of zero leads to the original SCM, while as λ approaches infinity, the traditional SCM becomes unimportant, and the problem reduces to selecting the nearest neighbor. The appendix discusses methods for selecting λ.

Shrinking a Very Large Pool of Donor Units

Although for many estimation methods more information is better than less, in the SCM a large number of donors can cause problems. First, if pretreatment outcomes are used as predictors, a large pool of donor units relative to the length of the pretreatment period increases the chance that a donor unit is selected because idiosyncratic shocks make the outcome of the donor appear to resemble the treated unit. This form of overfitting can lead to biased results because donor units that are selected may not have outcomes that match the outcomes of the treated unit.

Several factors can reduce the risk of overfitting. Increasing the length of the pretreatment period decreases overfitting for two reasons. First, although overfitting can also occur among covariates, those predictors are typically averaged over a number of years. Increasing the length of time over which the predictor is averaged reduces the impact of a shock in any one year. Second, more pretreatment periods allow for more outcome predictors. As the number of predictors increases, any one predictor becomes less important in the selection of the donor units. This is particularly true if all of the pretreatment outcomes are used as predictors because in that case no other predictors can be used and all of the predictors are equally weighted.

A second problem is that it increases the probability that the predictors of the synthetic control exactly match the predictors of the treated unit. Fully satisfying the optimization problem causes difficulties, which is counterintuitive, especially because the proof of the tight bounds on SCM's bias in Abadie, Diamond, and Hainmueller (2010) requires this condition. But Abadie (2020) explains that if there is an exact match, the set of donor units used to form the synthetic control are almost certainly not unique and will use many or all of the potential donor units. That reduces the value of having subject matter experts evaluate the appropriateness of the selected donor units because many or all will be used and many different combinations of weights can be generated. A graphical explanation for this is given in the appendix.

Although a perfect fit is relatively rare, Abadie (2020) points out that the risk of it occurring can be reduced by limiting the pool of donors to units with predictors most like those of the treated unit. In other words, the graph-based procedure described above that reduces interpolation bias also reduces the risk of a perfect fit.

Reducing Bias from a Poorly Fitting Synthetic Control

Abadie, Diamond, and Hainmueller (2015) recommended that the SCM not be used if the outcomes of the synthetic control poorly fit the outcomes of the treated unit in the pretreatment period. Although this step is often skipped, fit can be determined from tables of the predictors, as in table 1. If only pretreatment outcomes are used as predictors, the fit can be determined visually or by calculating the MSPE of the synthetic control in the pretreatment period. A large MSPE suggests that another method should be used, and although no agreed upon benchmark exists, MSPE should presumably be small compared to the overall variance of the synthetic control in the pretreatment period.

Figures analogous to figure 3 provide a large amount of visual information that can be used to assess the pretreatment fit of the synthetic control. Ideally, the difference between the actual treated unit and the synthetic unit hovers near zero. The paths of the placebo units should also hug the zero line, although because predictors are chosen to improve the fit of the treated unit, the paths of the placebo units may not be as close to zero as the path of the treated unit. If the difference varies systematically from zero, or if the difference appears to be trending, the SCM may be inappropriate. Tables such as table 1 also can be used to assess the fit of predictors, especially covariates.

However, a number of articles provide methods for addressing this. Duodenchenko and Imbens (2017) provide a general method, while Abadie and L'Hour (2020), Ben-Michel and colleagues (2020), and Ferman and Pinto (2021) provide special cases.

Abadie and L'Hour (2020) and Ben-Michel and colleagues (2020) independently develop a method for handling the bias caused by poorly fitting synthetic controls. Recall that the purpose of the synthetic control is to mimic the outcomes of the treated unit had it not received treatment. Assuming that predictors are related to posttreatment outcomes, if the predictors of the synthetic control are far from the treated unit's predictors, the synthetic outcomes will be far from the outcomes of the treated unit without treatment. In that case, the difference between the outcomes of the treated unit and the synthetic control are a biased measure of the treatment. The methods use regression analysis to estimate the relationship between the predictors and the posttreatment outcomes then modifies the outcomes so that they do not depend on the predictors. One advantage is that, while code is available, analysts can use any regression package along with the original SCM code. It also maintains the restriction that the weights are positive and sum to one. It should be noted that the weights are applied to transformed variables, and those restrictions do not necessarily hold on implied weights of the untransformed variables. The methods are explained in more detail in the appendix.

Ferman and Pinto (2021) independently suggest subtracting the pretreatment mean of all variables, which results in the same solution as using a constant in the regression-based methods above. One advantage of this approach is that it is easy to preprocess variables by subtracting the pretreatment mean from all variables prior to using the standard SCM. As with the other solutions, the implicit weights on the untransformed variables are no longer bound to be nonnegative and sum to one, so the weights must be interpreted with some caution.

Improving Model Selection

Although the SCM can be transparent compared to other methods, Ferman, Pinto, and Possebom (2020) point out that it is still possible to select predictors that push results toward or away from finding a significant effect from treatment. This includes not just covariates, but which pretreatment outcomes to use. Ferman, Pinto, and Possebom (2020) show using both Monte Carlo simulations and simulations on real data that there can be scope to choose the result by choosing from a set of previously used models.

To mitigate this problem, they recommend that analysts take the following steps:

1. Always report the results from using all pretreatments periods to avoid concerns about cherry-picking.
2. Only include covariates if it is important to match on those predictors. If all pretreatment outcomes are used as predictors, covariates receive no weight. Therefore, if covariates are to be used, some pretreatment outcomes must be excluded. In this case, use only specifications that allow the set of outcome predictors to increase as the set of pretreatment periods increase. For example, include an outcome from every other pretreatment period, rather than the three periods preceding the treatment. In these specifications, the choice of covariates becomes irrelevant as the number of pretreatment outcomes increases to infinity.
3. Show results for several specifications. Because this could complicate inference, it is alternatively appropriate to choose one specification by using an objective standard, such as the specification that minimizes the MSPE in the pretreatment period. Dube and Zipperer (2015) suggest another criterion: minimizing the MSPE for the placebos in the posttreatment periods.

Extending Inference Tests

The most common method for inference and model evaluation is the ratio of the root mean square error after treatment to the root mean square error before treatment. A large ratio for a treated unit implies that the difference between the synthetic control and the treated unit is larger after treatment, and a larger ratio for the treated unit than the donor units suggests that the treatment was effective. A form of statistical inference is found by calculating the share of units with lower ratios. For example, in Abadie, Diamond, and Hainmueller (2010), California has the highest ratio of the 39 states (California plus a pool of 38 potential donors). The ratio of 1:39 (0.026) is then interpreted as a p-value. Note that reducing the pool of control units reduces the granularity of available p-values. With only 25 units in a pool, the smallest possible p-value is 0.04. With only 15 units, it is 0.066.

Aside from the issue with granularity, these p-values can be problematic because using a ratio of root mean square error after and before treatment is only valid if treatment is randomly assigned, which is generally not true in cases in which the SCM is used. Ferman and Pinto (2017) refer to the advice of Abadie, Diamond, and Hainmueller (2010): that the SCM should only be used if the pretreatment fit is good predisposes the

denominator of the ratio towards zero. This inflates the value of the ratio, and using the Fisher inverse approach can lead to overrejection of the null hypothesis.

Nevertheless, Firbo and Possebaum (2018) compare that ratio with similar statistics and find that it works well in terms of size, power, and robustness. They show that the ratio can be adapted to test other hypotheses beyond simple effectiveness of the treatment and suggest testing whether the treatment has a simple additive effect or follows linear, quadratic, or exponential functions. Rather than calculate the MSPE of the difference between the outcomes of the treated unit and the synthetic control, it is calculated on the difference between the treated unit and the sum of the synthetic control and the hypothesized function. Then, the calculation of the ratios and the p-value proceeds as before. Confidence intervals for these functions can also be constructed by inverting the p-values. Because this construction can be computationally burdensome, they recommend it only for simple functions, such as the treatment having an additive or linear effect. In each case, constructing a 1-α confidence interval requires locating the parameters such that α percent of the MSPE ratios are below the ratio created by the hypothesized treatment.

CONCLUSION

The widespread interest in the SCM has led to its use in investigating various policies. Recent analyses include decriminalizing prostitution (Cunningham and Shah 2018), carbon pricing (Andersson 2019), right-to-carry laws (Donohue et al. 2019), the effectiveness of face masks in reducing COVID-19 transmission (Mitze et al. 2020), and Nebraska's move to a unicameral house (Hankins 2020). Although each application brings its own challenges, they all use the same underlying method.

Although McClelland and Gault (2017) provide guidance for the use of the SCM, numerous improvements have been made since its publication. Some of these have extended its use beyond the case of a single treated unit or have otherwise reduced the transparency of the method. In this updated guide, we have incorporated improvements that still allow the SCM to work as an effective and clear supplement to case studies. Rather than overwhelm readers with mathematical notation, our goal is to make these topics more accessible by discussing them and how they apply to research with new visualizations and simplified intuitive frameworks.

STEPS IN USING THE SYNTHETIC CONTROL METHOD

Analysts can use the SCM to study a policy treatment's effect in a state, region, or locality on a particular outcome by comparing the treatment outcome with that of the synthetic control.

Step 1: Determine the pretreatment period (often based on the availability of data on the outcome variable)

Longer pretreatment periods reduce overfitting.

The pretreatment period for outcomes should be as long as is possible, but it should not include prior treatments.

Step 2: Identify predictors of the outcome variable

Choose predictor variables that should affect outcomes in units both before and after treatment.

If outcomes from all pretreatment periods are used, covariates cannot be used.

If covariates are used, determine the range of pretreatment years over which the predictors will be averaged. A longer pretreatment year range is better than a short one, and covariates do not need to be rejected if some years in the pretreatment period are unavailable.

Most studies using covariates also include outcomes from several pretreatment years. Choose values that highlight the trend of the outcome before treatment.

Step 3: Identify a pool of donor units

Smaller pools reduce the chance of overfitting, but they reduce the granularity of possible p-values, so it is possible to have a pool that is too small.

Exclude any units that enacted policy treatments of similar or larger size during the selected period. Relatively small treatments do not necessarily disqualify a unit from being in the donor pool.

Exclude units that may have been affected by spillovers from the treated unit.

If the number of units is still large relative to the number of pretreatment periods, limit the pool to those with predictor values close to the values of the treated unit before treatment. The values of each predictor in the treated unit must be neither the largest nor smallest and ideally are closely surrounded by values from donor units.

Visual inspection of a figure similar to figure 4 may be useful. Ideally, the predictors of the treated unit should be surrounded by the predictors of the potential donors.

Alternatively, if some members of the donor pool have predictor values far from those of the treated unit, consider using the method of Abadie and L'Hour (2020) to reduce potential interpolation bias.

Step 4: Estimate the synthetic control

See the appendix for a list of code supporting articles discussed in this paper.

Step 5: Assess the pretreatment period goodness of fit of the synthetic control

Compare the values of the predictors of synthetic control with the values from the treated unit. If the values differ for one or more predictors, investigate using a figure similar to figure 4.

Evaluate how closely the outcome path of synthetic control during the pretreatment period follows that of the treated unit using the MSPE or visual inspection.

If the synthetic control poorly matches the treated unit using either approach, either do not use the SCM or consider a bias correction method such as one proposed by Ben-Michel and colleagues (2020), Ferman and Pinto (2020), or equivalent methods.

Step 6: Conduct additional analyses

Review predictor weights to determine the selected predictor variables' strengths in explaining the outcome. If all of the pretreatment outcomes are used as predictors, predictor weights are all equal, and this step can be skipped.

Conduct an in-time placebo test by reestimating the model but assuming that the treatment takes place at an earlier period. The outcome of the synthetic control should match the outcome of the treated unit at least until the actual treatment takes place.

Test the fragility of the results to the use of specific donors by sequentially omitting each donor from the pool and reestimating the synthetic control.

If covariates are used, reestimate the SCM using all pretreatment outcomes and report both results.

Show the results from several sets of predictors.

Review the selected units and their weights to assess the reasonableness of their use in the synthetic control. It can be more important for outcomes of donor units to have a trend similar to that of the treated unit than for the units to have a similar average.

Step 7: Conduct significance tests of the results

Conduct placebo tests on units in the donor pool to evaluate the significance of the results.

Visually inspect the posttreatment difference between the treated state and its synthetic control. If, as in figure 3, the difference between the treated unit and the control unit is larger than the difference for most of the placebo units, there is evidence that the treatment had an effect.

For the treated unit and each donor, calculate the ratio of the MSPE after treatment to the MSPE before treatment. Evidence of an effective treatment is found if the ratio for the treated unit is much higher than for the donors.

Calculate a p-value by ranking the treated unit's ratio relative to the ratios of the control units.

Other hypotheses can be tested, and confidence sets formed using the methods from Firbo and Possebaum (2018).

GRAPHICAL DESCRIPTION OF INTERPOLATION BIAS, EXTRAPOLATION BIAS, AND CONVEX HULLS

Interpolation and extrapolation bias, which have received attention in several articles on the SCM, are caused by unaccounted nonlinearities. Interpolation bias can occur in the original SCM, while extrapolation bias cannot. Some extensions to the SCM address interpolation bias but open the door for extrapolation bias.

Both types of bias can be seen in figure 1. In the original SCM, the weights for the control units must be nonnegative and sum to one. That necessarily means that the predictor values for the synthetic control must lie in the convex hull of the points in figure 1. For treated unit T_1, the predictor values are $\hat{T}_1$. This is on the convex hull, lying between C_1 and C_2. Those two units will receive positive weights, while the other units in the donor pool will be unweighted. It is not a coincidence that in this example there are two predictors and two units with positive weights. In general, the number of positively weighted units is limited by the number of predictors, not the number of units in the donor pool.

The possibility for interpolation bias stems from the fact that C_1 and C_2 are chosen from the donor pool, even though control unit C_5 is much closer to T_1. Although it is closer, C_5 cannot be selected because it is inside the convex hull of the control units, so that any combination of it with another unit will produce a worse fit than $\hat{T}_1$. With a linear relationship between outcomes and the predictors, the distance between the control units C_1 and C_2 does not matter because if $\hat{T}_1$ is close to T_1 then the synthetic outcome formed using the same weights as $\hat{T}_1$ will be similarly close to the treated unit's outcome. But if the relationship is nonlinear, the outcome of the synthetic control may not match the outcome of the treated unit. The amount of mismatch will be a function of both the distance of the control unit's predictors from the predictors of the treated unit and the amount of nonlinearity. If the synthetic control's outcome substantially varies from the treated unit's outcome in the lack of treatment, comparing the synthetic control to the treated unit will misrepresent the effect of treatment.

Attempting to mitigate interpolation bias can lead to extrapolation bias. If the weights used to form the synthetic control are not limited to be nonnegative and sum to one, it is possible for C_5 to be used to form the synthetic control. Thus, removing the restrictions on the weights can allow much closer points to be used. But removing the restrictions also removes the limit on the number of control units used to form the synthetic control. C_3 and C_4 could also now be used in the synthetic control. As before, if a linear relationship between outcomes and predictors exist, the distance between control units and treated unit does not matter. If the relationship is nonlinear, extrapolating from far away units can cause the synthetic control's outcome to badly miss the outcome of the treated outcome. C_4, in particular, has much smaller values for both predictors than T_1. Moreover, if the treated unit has predictor T_3, all of the potential control units have substantially lower values of predictor A than the treated unit. A relationship between outcome and predictors may appear to be sufficiently

linear in the domain of the control units, with the nonlinearity arising when extrapolating the relationship out to T_3.

If the predictors of the treated unit lie inside the convex hull, the SCM becomes less transparent to use because the SCM will use many or all of the potential control units and in any of an infinite number of combinations. An analyst or subject matter expert therefore would have less of an ability to judge.

On the other hand, the synthetic predictors will perfectly match the treated unit's predictors, producing the best possible synthetic control for the treated unit. These ideas can be seen in Figure A.1. To see how a weighted average of C_1 through C_4 can perfectly match T_2, imagine that the weighting takes place in two stages. In the first stage, point P_1 is created as a weighted average of C_1 and C_4 and point P_6 is created as a weighted average of C_2 and C_4. The solid line connecting P_1 and P_6 shows that a weighted average of the two points exactly equals T_2. Thus, all four control units are used, and they exactly match the treated unit. More generally, all control units in the donor pool can be used to exactly match the treated unit. But other points, such as P_2 and P_5, could have been created instead and can be averaged to exactly match T_2. Similarity exists with P_3 and P_4. It should be clear that there are infinitely many possible combinations of weights that can exactly match the

FIGURE A.1
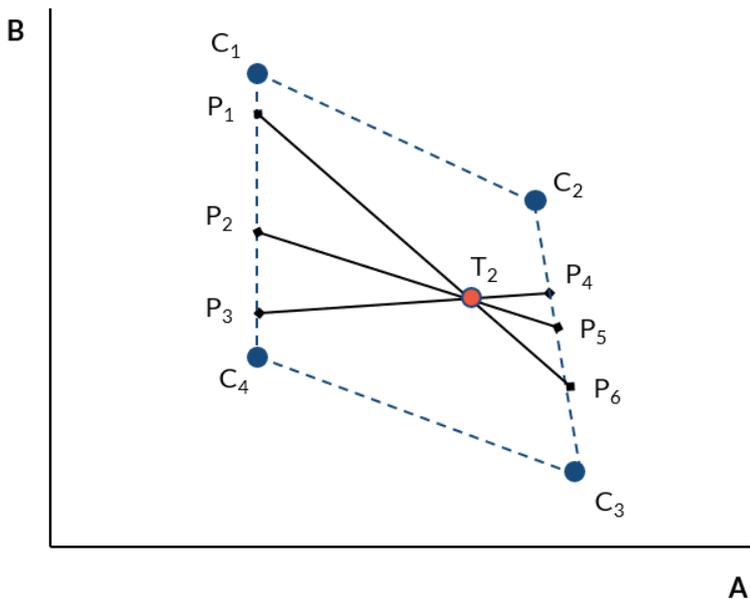
treated unit.

NOTES ON IMPLEMENTING METHODS DISCUSSED IN THE PAPER

A GRAPHICAL VIEW OF REDUCING INTERPOLATION BIAS

A simple method for reducing the potential for interpolation bias is to restrict the number of control units based on their distance from the treated unit. The effect of this can be seen on plots such as figure 4. A subset of donor units, such as those with predictors within three standard deviations of the treated unit, can be chosen and the data replotted. Figure A.2 shows the result of this process for four distances. Because the use of log state per capita GDP dramatically reduces the number of available donor units, it is omitted. As the first panel shows, restricting the donor pool to those with predictors that are all within three standard deviations of California's predictors slightly reduces the number of potential donor units from 38 to 35 states. The second panel shows that reducing the distance to two standard deviations further shrinks the number of states to 29. The third panel shows that 23 states have all predictors within 1.5 standard deviations. Finally, only seven states have all predictors within one standard deviation.

The reduced opportunity for interpolation bias should be compared to the difference in fit. For each subset of donors, the SCM can be reestimated and the fit in the pretreatment period can be visually inspected or the MSPEs from the different subsets can be compared.

CHOOSING THE PENALTY PARAMETER λ IN ABADIE AND L'HOUR (2020)

The authors suggest two methods for choosing λ. In the first method, an analyst would start with some value of λ and in a placebo-type exercise separately apply the SCM to each unit in the donor pool, calculating the MSPE after treatment. The MSPEs are then averaged, and the calculation is repeated for various values of λ until the minimum average MSPE is located. Because none of the control units receive treatment, the outcomes of the synthetic and actual control units should track each other after treatment. If they do, there is no evidence of interpolation bias and a low value of λ will ultimately be selected. If they diverge, λ is increased to reduce the interpolation bias and if the divergence is reduced, a high value for λ will ultimately be selected. The second method focuses on the treated unit. In it, the pretreatment period is divided into a "training" block which that precedes a "validation" block that includes the periods immediately before treatment. The synthetic control is formed using data in the training block, and the difference between the outcomes of the synthetic control and the treated unit are calculated in the validation block. The λ parameter is chosen to minimize an aggregate

measure of the difference, such as the sum of the squared differences, or the square of the summed differences.

BIAS REMOVAL IN ABADIE AND L'HOUR (2020) AND BEN-MICHEL AND COLLEAGUES (2020)

A poorly fit synthetic control can result in a biased estimate of the treatment effect. Abadie and L'Hour (2020) suggest a fix that removes the bias individually for each period after treatment in several steps. First, estimate the synthetic controls in the usual way, noting the weights that are applied to the donor units. Then, using the pool of potential donors, regress the outcome in one posttreatment period on the pretreatment predictors. This estimates the relationship between the predictors and the outcome for that period. The outcome for each control state is then modified by subtracting this relationship. In other words, the residual from the regression is substituted for the outcome. By construction, this version of the outcome is uncorrelated with the predictors. Because the treated unit was not used in the regression, it does not have a residual. Instead, the coefficients from the regression are used to predict the outcome for the treated unit, and that prediction is then subtracted from the actual outcome. This process is repeated for every outcome in the posttreatment period. The weights from the original synthetic control estimate are applied to the modified outcomes for the donor and treated units.

Although linear regression is the simplest method for estimating the outcome, it is not the only one. Ben-Michael and colleagues (2020) loosen the constraint that weights must be nonnegative and sum to one and reduce extrapolation bias by using a ridge regression that applies a penalty the sum of the squared coefficients in the regression.[5] In figure 1, this would be attempting to penalize the use of C_3 and C_4. Chernozhukov and colleagues (2021) use a lasso function, which penalizes the sum of the absolute values of the coefficients. Doudchenko and Imbens (2017) loosen the weight restriction but suggest using an elastic net penalty function, which is a combination of ridge and lasso penalties. When calculating the treatment effect, they also estimate and add a constant to the difference between the treated unit and the synthetic control unit. This effectively shifts the synthetic control up or down by a fixed amount, and it is useful when pretreatment predictors and outcomes are outliers.

Ben-Michael and colleagues (2020) assume that only pretreatment outcomes are used but extend their results to the use of covariates. If covariates are used, they make two recommendations. First, if there is a moderately large number of those predictors, they suggest standardizing all of the predictors to have a unit standard deviation and then proceed as before using a ridge regression. But if there are few of those predictors compared to the number of potential control units, they suggest that the outcomes pre- and posttreatment be "residualized" against the covariates and then the process described above be applied, using a ridge regression.
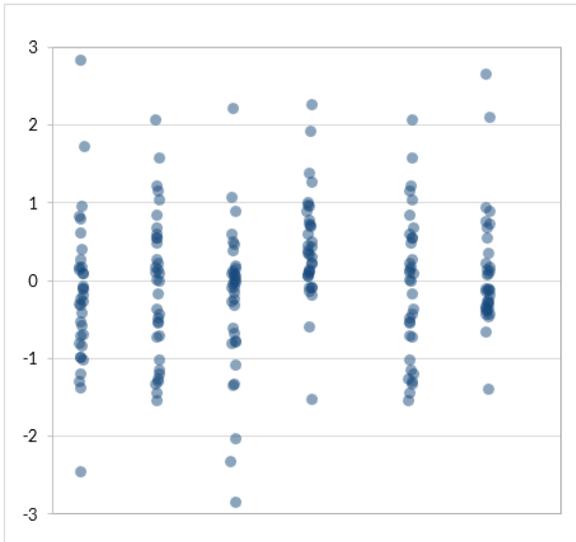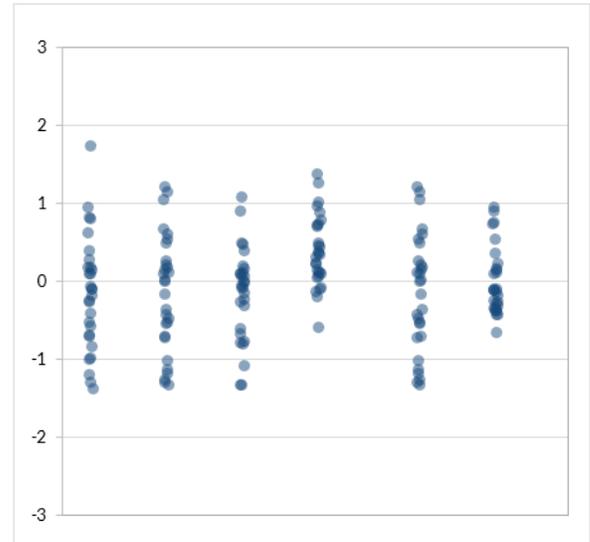
Differences between Average Predictor Values of Control Units and the Treated Unit in Abadie, Diamond, and Hainmueller (2010), Sample Reduced by Proximity to Treated Unit
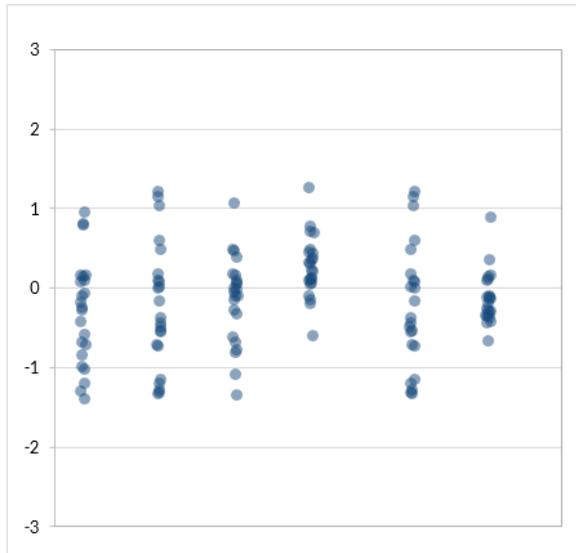

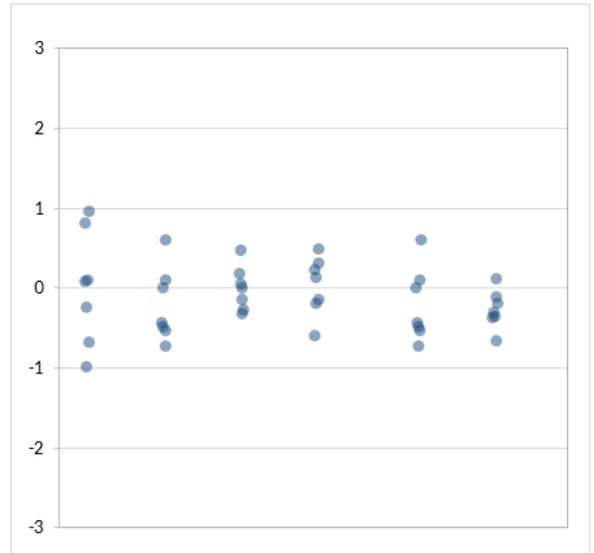35 states with predictors within 3 standard deviations of CA's predictors

29 states with predictors within 2 standard deviations of CA's predictors

23 states with predictors within 1.5 standard deviations of CA's predictors

7 states with predictors within 1 standard deviation of CA's predictors

Source: Authors' calculations using the Synth package and data from ADH to replicate Abadie, Diamond, and Hainmueller (2010).

Note: Predictors featured in each figure include (from left to right) log state per capita GDP, beer consumption per capita, percent of state population 15 - 24 years of age, retail price of cigarettes, 1998 cigarette sales, 1980 cigarette sales, 1975 cigarette sales.

[1] Stata, R, and MATLAB code can be found at https://web.stanford.edu/~jhain/synthpage.html. For additional relevant code see the following: Extensions to the synthetic control significance tests and plots and calculating average treatment effects for multiple treated states. Tidysynth, a tidy implementation of the synthetic control method in R Studio. A key benefit of a tidy implementation is that the entire preparation process for building the synthetic control can be accomplished in a single pipe.

[2] Replication code (in R) can be found at https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/24714.

[3] Stata and R code can be found at https://codeocean.com/capsule/5800231/tree/v1.

[4] Stata, R, and MATLAB code can be found at https://github.com/jeremylhour/pensynth.

[5] R code can be found at https://github.com/ebenmichael/augsynth.

Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects" *Journal of Economic Literature* 59 (2): 391-425.

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

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–32.

Abadie, Alberto, and Jeremy L'Hour. 2021. "A Penalized Synthetic Control Estimator for Disaggregated Data." *Journal of the American Statistical Association* 116 (536): 1817–34. https://doi.org/10.1080/01621459.2021.1971535.

Amjad, Muhammad, Devavrat Shah, and Dennis Shen. 2018. "Robust Synthetic Control." *Journal of Machine Learning Research* 19: 1–50.

Andersson, Julius J. "Carbon Taxes and CO 2 Emissions: Sweden as a Case Study." *American Economic Journal: Economic Policy* 11.4 (2019): 1–30.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. "Synthetic Difference in Differences." *American Economic Review* 111 (12): 4088–118).

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2020. "Matrix Completion Methods for Causal Panel Data Models." Working paper. Ithaca, NY: Cornell University. https://arxiv.org/abs/1710.10251.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2020. "The Augmented Synthetic Control Method." *Journal of the American Statistical Association* 116 (536): 1789–1803. https://10.1080/01621459.2021.1929245.

Cunningham, Scott, and Manisha Shah. 2018. "Decriminalizing Indoor Prostitution: Implications for Sexual Violence and Public Health." *Review of Economic Studies* 85 (3): 1683–1715.

Donohue, John J. III, Abhay Aneja, and Kyle D. Weber. 2019. "Right-to-Carry Laws and Violent Crime: A Comprehensive Assessment Using Panel Data and a State-Level Synthetic Controls Analysis." *Journal of Empirical Legal Studies* 16 (2): 198–247.

Dube, Arindrajit, and Ben Zipperer. 2015. "Pooling Multiple Case Studies Using Synthetic Controls: An Application to Minimum Wage Policies." Working paper IZA DP No. 8944.

Duodenchenko, Nikolay, and Guido Imbens. 2017. "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis." Manuscript. Available at https://arxiv.org/abs/1610.07748

Ferman, Bruno, and Cristine Pinto. 2021. "Synthetic Controls with Imperfect Pre-treatment Fit." *Quantitative Economics* 12 (4): 1197–221.

Ferman, Bruno, and Cristine Pinto. 2017. "Placebo Tests for Synthetic Controls." MPRA paper 78079. University Library of Munich, Germany.

Ferman, Bruno, Cristine Pinto, and Vitor Possebom. 2020. "Cherry Picking with Synthetic Controls." *Journal of Policy Analysis and Management* 39 (2): 510–532.

Firpo, Sergio, and Vitor Possebaum. 2018. "Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets." *Journal of Causal Inference* 6 (2): 1–26.

Glantz, Stanton A., and Edith D. Balbach. 2000. "Tobacco War: Inside the California Battles." Berkeley: University of California Press.

Hankins, William B. 2020. "Finally, Nebraska: A Synthetic Control Analysis of Legislative Structure." *State Politics & Policy Quarterly* 20 (1): 3–22.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social and Biomedical Sciences: An Introduction.* 1st edition. Cambridge, England: Cambridge University Press.

Kellogg, Maxwell, Magne Mogstad, Guillaume Pouliot, and Alexander Torgovitsky. 2020. "Combining Matching and Synthetic Controls to Trade off Biases from Extrapolation and Interpolation." Working Paper 26624. Cambridge, MA: National Bureau of Economic Research.

Mitze, Timo, Reinhold Kosfeld, Johannes Rode, and Klaus Wälde. 2020. "Face Masks Considerably Reduce COVID-19 Cases in Germany." *Proceedings of the National Academy of Sciences* 117 (51): 32293–301.

Rosenbaum, Paul R. 2002. *Observational Studies.* New York: Springler Science + Business Media.

Xu, Yiqing. 2017. "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." *Political Analysis* 25: 57–76.

ABOUT THE AUTHORS

Robert McClelland is a senior fellow in the Urban-Brookings Tax Policy Center. Previously, he worked in the tax analysis division of the Congressional Budget Office (CBO), where he examined the impact of federal tax policy on charitable giving and bequests, the realization of capital gains, labor supply, and small businesses. He worked for the CBO from 1999 to 2005 and from 2011 to 2016, and in between, he directed the division of price and index number research at the Bureau of Labor Statistics. McClelland is a member of the Conference on Research in Income and Wealth. He has taught econometrics at Johns Hopkins University, where he won an Excellence in Teaching award in 2006. McClelland received a BA in economics and environmental studies from the University of California, Santa Cruz, and a PhD in economics from the University of California, Davis.

Livia Mucciolo is a research assistant in the Urban-Brookings Tax Policy Center. She graduated magna cum laude from Brown University and holds a BA in economics and public health.

TPC

The Tax Policy Center is a joint venture of the
Urban Institute and Brookings Institution.

URBAN INSTITUTE

BROOKINGS

For more information, visit taxpolicycenter.org
or email info@taxpolicycenter.org