Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates

Kaul, Ashok and Klößner, Stefan and Pfeifer, Gregor and Schieler, Manuel

Saarland University, Saarland University, University of Hohenheim, Saarland University

2 March 2015

Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates

Ashok Kaul, Saarland University
Stefan Klößner, Saarland University
Gregor Pfeifer, University of Hohenheim*
Manuel Schieler, Saarland University

This Version: July 29, 2017
(First Version: March 2, 2015)

Abstract

It is becoming increasingly popular in applications of synthetic control methods to include the entire pre-treatment path of the outcome variable as economic predictors. We demonstrate both theoretically and empirically that using all outcome lags as separate predictors renders all other covariates irrelevant. This finding holds irrespective of how important these covariates are for accurately predicting post-treatment values of the outcome, potentially threatening the estimator's unbiasedness. We show that estimation results and corresponding policy conclusions can change considerably when the usage of outcome lags as predictors is restricted, resulting in other covariates obtaining positive weights.

Keywords: Synthetic Control Methods; Economic Predictors; Counterfactuals; Policy Evaluation.

1

1   Introduction

In their seminal papers, Abadie and Gardeazabal (2003) & Abadie et al. (2010) introduced the application of synthetic control methods (SCM) to comparative case studies. Since then, these methods have been applied to diverse research topics, rapidly establishing a new and intuitive alternative for constructing counterfactuals.[1] Athey and Imbens (2017) even state that SCM "is arguably the most important innovation in the policy evaluation literature in the last 15 years".

SCM involve the comparison of outcome variables between a unit representing the case of interest, i.e., a unit affected by an intervention, and otherwise similar but unaffected units reproducing an accurate counterfactual of the unit of interest in absence of the intervention. An algorithm-derived combination of precisely weighted comparison units is supposed to better depict the characteristics of the unit of interest than either any single comparison unit alone or an equally weighted combination of all or several available control units.

It is key for SCM that the synthetic control unit provides a good approximation of how the variable of interest of the treated unit would have developed if no treatment had taken place. In order to achieve this goal, it is clear that the counterfactual pre-treatment values of the outcome variable, provided by the synthetic control unit, must be close to the corresponding actual values ("outer optimization"). In addition, according to Abadie et al. (2010) and Abadie et al. (2015), it is also vital that the synthetic control unit and the treated unit resemble one another with respect to pre-intervention values of *predictors* of the outcome variable

---

[1] After Abadie et al. (2010), which builds on an initial idea in Abadie and Gardeazabal (2003), SCM were applied, i.a., by Nannicini and Billmeier (2011) as well as Billmeier and Nannicini (2013) (economic growth), Coffman and Noy (2012) as well as Cavallo et al. (2013) (natural disasters), Hinrichs (2012) (college affirmative action bans), Hosny (2012) (free trade), Jinjarak et al. (2013) (capital inflows), Kleven et al. (2013) (taxation of athletes), Bauhoff (2014) (school nutrition), Belot and Vandenberghe (2014) (educational attainment), Bohn et al. (2014) (Arizona's 2007 LAWA), Abadie et al. (2015) (Germany's reunification), Bilgel and Galle (2015) (organ donations), Liu (2015) (spillovers from universities), Stearns (2015) (maternity leave), Acemoglu et al. (2016) (political connections), Eren and Ozbeklik (2016) (right-to-work laws), Gobillon and Magnac (2016) (enterprise zones), Kreif et al. (2016) as well as O'Neill et al. (2016) (health improvements due to pay-for-performance schemes), and Kaestner et al. (2017) (effects of ACA medicaid expansions). Eventually, Gardeazabal and Vega-Bayo (2017) find that the SCM estimator performs well as compared to alternative panel approaches, and Klößner and Pfeifer (2017) extend SCM to the forecasting context.

("inner optimization").[2] These predictors consist of lagged values of the outcome, or linear combinations thereof, and other economic variables that have predictive power for explaining the dependent variable (also called "covariates" subsequently).

As stated by Cavallo et al. (2013, p. 1552), it seems an obvious choice to include the entire pre-treatment path of the outcome variable as economic predictors. This is exactly what is done in, e.g., Bilgel and Galle (2015), Billmeier and Nannicini (2013), Bohn et al. (2014), Hinrichs (2012), Kreif et al. (2016), Liu (2015), Nannicini and Billmeier (2011), O'Neill et al. (2016), and Stearns (2015): they use *all* lagged outcome values as separate predictors in addition to several covariates when applying SCM. Potentially, many more researchers (will) do so.

This paper makes two contributions. First, we prove theoretically that if one uses all lagged outcome variables as separate predictors, then all other economic outcome predictors (covariates) become irrelevant. More precisely, we show that including all pre-intervention values of the outcome as economic predictors implies that only the pre-treatment fit with respect to the variable of interest is optimized. This holds true independent of the data-driven procedure used to obtain predictor weights, i.e., both for the so-called nested and regression based approaches. Consequently, in the SCM application we mainly focus on throughout this paper—Billmeier and Nannicini (2013), who analyze the impact of economic liberalization on GDP—the covariates taken from the literature *do not affect* the synthetic control. The authors obtain the very same counterfactual that would have followed if they had used economically meaningless covariates—or even none at all.[3] We further discuss that solely optimizing the pre-treatment fit of the dependent variable and ignoring the covariates can be harmful: the more the covariates are truly influential for future values of the outcome, the larger a potential bias of the estimated treatment effect can become, possibly leading to wrong policy conclusions.

Our second contribution is to show that, allowing for only one pre-treatment measure of

---

[2] Abadie et al. (2010) show that if *both* optimizations are pursued, one obtains asymptotically unbiased estimates of the treatment effect, which under certain conditions are even unbiased in small samples.

[3] Evidently, the same happens throughout the research done in Bilgel and Galle (2015), Bohn et al. (2014), Hinrichs (2012), Kreif et al. (2016), Liu (2015), Nannicini and Billmeier (2011), O'Neill et al. (2016), and Stearns (2015).

GDP per capita in addition to the set of other covariates used in Billmeier and Nannicini (2013), the estimated impact of economic liberalization on GDP can change drastically. With such a restriction on the outcome lags, other covariates with predictive power for explaining the dependent variable obtain positive weights in order to build the—now differently weighted—synthetic control unit. As a result, we measure significantly different treatment effects for certain cases. While such examples correspond to the data used in Billmeier and Nannicini (2013), the core issue and its implications carry over to several other applications making use of all pre-treatment outcomes as separate economic predictors.

The remainder proceeds as follows: Section 2 briefly outlines the synthetic control approach by Abadie et al. (2010). Section 3 theoretically proves that using all pre-treatment values of the outcome as separate predictors leaves other outcome predictors completely ignored. It further discusses why this might lead to biased estimation results. Section 4 illustrates practical implications of this matter by discussing how economic interpretations of several liberalization examples in Billmeier and Nannicini (2013) change when not all pre-treatment values of the outcome are used as separate predictors and, as a consequence, covariates are accounted for. Section 5 discusses and concludes.

2 SCM Approach

In the SCM approach, the synthetic control unit is created out of a collection of J control units, the so-called donor pool. The comparability of the synthetic control unit to the treated unit in the pre-intervention period is determined by a set of predictors from T pre-treatment periods. Two kinds of predictors are introduced. The first kind is given by M linear combinations of Y in the pre-treatment periods, while the second kind consists of r other covariates with explanatory power for Y. All k predictors (with $k = r + M$) are combined in a $(k \times 1)$ vector X_1 for the treated unit and in a $(k \times J)$ matrix X_0 for all control units.

In one part of the optimization process (the inner optimization), one tries to find a linear combination of the columns of X_0 that represents X_1 best, i.e., one searches for a combina-

tion of the donor units such that the difference of the predictors' values of the treated and the counterfactual becomes as small as possible. The distance metric used to measure this difference is:

$$\|X_1 - X_0W\|_V = \sqrt{(X_1 - X_0W)' V (X_1 - X_0W)}, \tag{1}$$

where the weights used to construct the synthetic control unit are denoted by the vector W and the weights of the predictors are given by the nonnegative diagonal matrix V. The latter takes into consideration that not all predictors have the same predictive power for the outcome variable Y.

The inner optimization is then, for given predictor weights V, the task of finding non-negative control unit weights W summing up to unity such that:

$$\sqrt{(X_1 - X_0W)' V (X_1 - X_0W)} \xrightarrow{W} \min. \tag{2}$$

Denote the solution to this problem by $W^*(V)$.

The other part of the optimization (the outer optimization) deals with finding optimal predictor weights. It often follows a data-driven approach proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010): V is chosen among all positive definite and diagonal matrices such that the mean squared prediction error (MSPE) of the outcome variable Y is minimized over the pre-intervention periods. For convenience, let Z_1 denote the subset of Y for the treated unit over the chosen pre-intervention periods, while Z_0 denotes the analogous matrix for the control units. This results in the outer optimization problem:

$$(Z_1 - Z_0W^*(V))' (Z_1 - Z_0W^*(V)) \xrightarrow{V} \min. \tag{3}$$

Eventually, there is a second data-driven way to determine the predictor weights, the so-called regression based method, used, e.g., by Bohn et al. (2014).[4] This method proceeds as

---

[4] To determine V using the "SYNTH" package for Stata, the regression based option is the default; see the help page of the package, available online at `http://fmwww.bc.edu/RePEc/bocode/s/synth.html`.

follows: first, for every time t prior to the intervention, the corresponding dependent variable, $Z_{j,t}$ $(j = 1, \ldots, J + 1)$, is regressed on all economic predictors, yielding regression coefficients $\beta_{t,k}$ $(k = 1, \ldots, K)$. Second, v_k is then set as

$$v_k := \frac{\sum\limits_{t} \beta_{t,k}^2}{\sum\limits_{k=1\ldots,K} \sum\limits_{t} \beta_{t,k}^2}. \tag{4}$$

Thus, the larger the squared regression coefficients of economic predictor k, the more weight v_k is given to this predictor.

3   Using All Pre-Intervention Outcomes as Predictors

Determining Predictor Weights by Nested Optimization

To demonstrate the effect of using all lagged values of the dependent variable as predictors, it proves helpful to partition the predictor data matrices X_1 and X_0 into

$$X_1 = \begin{pmatrix} C_1 \\ Z_1 \end{pmatrix}, \quad X_0 = \begin{pmatrix} C_0 \\ Z_0 \end{pmatrix}, \tag{5}$$

with C_0 and C_1 containing the values of the covariates of the non-treated and treated units, respectively. We correspondingly partition the predictor weights V:

$$V = \begin{pmatrix} V_C & 0 \\ 0 & V_Z \end{pmatrix}, \tag{6}$$

with V_C containing the weights for the covariates, while V_Z collects the weights of the pre-treatment values of the dependent variable. Using equations (5) and (6), it is easy to see that equation (2) becomes

$$\sqrt{(C_1 - C_0W)' V_C (C_1 - C_0W) + (Z_1 - Z_0W)' V_Z (Z_1 - Z_0W)}. \tag{7}$$

Denoting by

$$W^{**} := \operatorname*{argmin}_{W} (Z_1 - Z_0 W)' (Z_1 - Z_0 W) \tag{8}$$

the weights W that produce the best fit with respect to the outcome variable alone, then by the definition of W^{**}, we have:

$$(Z_1 - Z_0 W^{**})' (Z_1 - Z_0 W^{**}) \leq (Z_1 - Z_0 W^*(V))' (Z_1 - Z_0 W^*(V)) \tag{9}$$

for all V. On the other hand, if we choose V^* with V_C^* equal to zero and V_Z^* equal to the identity matrix, then inspection of equation (7) entails

$$W^*(V^*) = W^{**}. \tag{10}$$

In the outer optimization, therefore, it will be optimal to choose the predictor weights such that the covariates are annihilated and only the lagged values of the outcome variable are taken into account. Put differently, although the inner objective function (2) explicitly takes the covariates into account, the synthetic control resulting after optimizing the outer objective function (3) will be calculated by ignoring the covariates completely. In the literature, many authors have failed to recognize this and thus inadvertently ignored their carefully chosen covariates, inter alia Bilgel and Galle (2015), Billmeier and Nannicini (2013), Bohn et al. (2014), Hinrichs (2012), Kreif et al. (2016), Liu (2015), Nannicini and Billmeier (2011), O'Neill et al. (2016), and Stearns (2015). There is a reason why researchers probably did not recognize that they unwittingly rendered the covariates irrelevant: in SCM applications, it is often the case that there exist several different combinations of predictor weights V minimizing the outer objective function (3), see Klößner et al. (2017). Thus, solutions for V with positive weights for the covariates may be reported, seemingly indicating that the covariates are being taken into account, although they are actually ignored.

We now turn to the consequences of using all lagged values of the dependent variable as separate economic predictors when the regression based method is used to form V.

Regression Based Predictor Weights

When $Z_{j,t}$ $(j = 1, \ldots, J+1)$ is regressed on all economic predictors, then one of the regressors for the lagged values of the dependent variable is the regressand itself. Therefore, the regression coefficients $\beta_{t,k}$ will equal unity if k corresponds to the respective lagged dependent variable, and vanish for all other economic predictors. Then, v_k will be zero for all true covariates, while $v_k = \frac{1}{T}$ for all k corresponding to lags of the dependent variable. Consequently, determining the predictors' weights by using the regression based option will also lead to the covariates being irrelevant, while lagged values of the dependent variable are the only economic predictors that matter. Thus, when using regression-based predictor weights, the resulting synthetic control unit will again be determined by solely optimizing the fit with respect to the outcome's pre-treatment values, an approach which is called "constrained regression" by Doudchenko and Imbens (2016).

Potential Consequences of Ignoring the Covariates

Above, we have proven that using all pre-treatment values of the outcome variable as separate predictors inevitably leads to every single covariate being ignored, i.e., the synthetic control unit will be determined by minimizing the outer objective function (3). This finding holds no matter what the covariates actually are and how important and helpful these might be in order to predict post-treatment values of the outcome variable. What is more, it holds regardless of the data-driven method used to determine the predictor weights.[5] When ignoring relevant covariates, the statistician's principle of using all available data is violated and synthetic controls are not applied as they are intended to be (Gardeazabal and Vega-Bayo, 2017): "the synthetic control is primarily designed to use any covariates that help explain the outcome variable as predictors, and not only pre-treatment values of the outcome variable".

To address the implications of disregarding the covariates, we introduce some additional notation: we denote by Y_{jt} the dependent variable's value at time t for unit j, with $j = 1$

---

[5] For an empirical affirmation, see Bohn et al. (2014, fn. 11).

for the treated unit and $j = 2, \ldots, J + 1$ for the donor units. While $t \leq T$ indicates pre-treatment values, $t > T$ refers to post-treatment values, with Y_{1t} for $t > T$ being understood as the counterfactual values that would have been observed in absence of the intervention. Usually, $Y_{jT} = \widetilde{Y}_{jt} + \alpha_t C_j$, is assumed to depend additively on the covariates by $\alpha_t C_j$, with $\alpha_t \in \mathbb{R}^{1 \times r}$, and $\widetilde{Y}$ denoting how the dependent variable would evolve if it was not influenced by the covariates. Given weights $W = (w_2, \ldots, w_{J+1})'$ for synthesizing, the difference between treated unit and its synthetic control with respect to the outcome then is

$$Y_{1t} - \sum_{j=2}^{J+1} w_j Y_{jt} = \widetilde{Y}_{1t} - \sum_{j=2}^{J+1} w_j \widetilde{Y}_{jt} + \alpha_t \left(C_1 - \sum_{j=2}^{J+1} w_j C_j \right). \tag{11}$$

If the covariates obtain positive weights during the inner optimization, irrespective of the actual method used to determine V, then $C_1 - \sum_{j=2}^{J+1} w_j C_j$ will be close to or ideally even exactly zero. Therefore, if the estimator W is such that $C_1 - \sum_{j=2}^{J+1} w_j C_j$ is always zero, then the additional term $\alpha_t \left(C_1 - \sum_{j=2}^{J+1} w_j C_j \right)$ appearing in (11) vanishes. However, if the estimator W—which is a random variable as it depends on Y—ignores the covariates, then $C_1 - \sum_{j=2}^{J+1} w_j C_j$ may be arbitrarily large, and its expectation might be significantly different from zero. Therefore, ignoring the covariates introduces an additional uncontrolled small-sample bias to the estimation which is likely to be significant, especially when α_t takes large values.[6] Correspondingly, as can be seen from the Monte Carlo studies of O'Neill et al. (2016), the SCM estimator ignoring the covariates performs quite poorly as compared to other non-SCM estimators that make good use of covariates, with rather large values for relative bias and root mean squared error.

On the other hand, observed covariates that are ignored are not different from unobserved confounders, and the synthetic control method is—under certain conditions—known to be asymptotically unbiased even in the presence of the latter, where "asymptotically" refers to

---

[6] Exemplarily, Stearns (2015) decides on a specification using all pre-treatment outcomes as separate predictors. Additionally, she includes certain demographic covariates which she considers particularly important for capturing the true effect of interest, i.e., α_t can be expected to be large.

the pre-treatment timespan growing to infinity (see Abadie et al., 2010; Ferman and Pinto, 2017). Thus, ignoring the covariates and optimizing only the fit with respect to many lags of the outcome may to some part be beneficial, as by fitting with respect to lagged outcomes alone, the synthetic control might do a better job in capturing the unobserved confounders. Overall, there exists a trade-off: (inadvertently) ignoring the covariates reduces efficiency of the synthetic control due to introducing a small-sample bias, but improves efficiency due to taking more care of unobserved confounders. Therefore, to what extent ignoring the covariates is harmful depends on the length of the pre-treatment timespan as well as the unobserved confounders' and observed covariates' importance for explaining the outcome.

4 Empirical Illustration

In the following, we use the data from Billmeier and Nannicini (2013) to illustrate the issue of separately using all pre-treatment values of the outcome variable as economic predictors. Even though we could have just as well used data from one of the several other studies also using every single pre-treatment observation of the dependent variable as a separate economic predictor, we make use of the specific data from Billmeier and Nannicini (2013). This is because, when the authors analyze whether liberalizing the economy of a country leads to higher GDP per capita, they are able to examine treatment effects of a huge battery of different units of interest.

Billmeier and Nannicini (2013) use two different selections of control units: in about 30 so-called "Type A experiments", they restrict the donor pool to eligible countries in the macroregion of those treated; in the analogous "Type B experiments", they allow the donor pool to consist of all eligible countries. To construct synthetic controls, the authors use every single value of GDP per capita before the treatment as predictors. Furthermore, they rely on the following set of covariates (if available): investment as a share of GDP, population growth, secondary school enrollment, average inflation rate, and a democracy dummy.

We contrast Billmeier and Nannicini (2013)'s results against two types of rigorous restric-

tions on the pre-treatment outcome lags. First, we apply a widely used SCM approach—solely including the *average* of the pre-intervention outcomes in addition to the covariates, as do, e.g., Abadie and Gardeazabal (2003), Abadie et al. (2015), and Kleven et al. (2013).[7] Second, we opt for the *last* pre-treatment value to be included in addition to the set of covariates, by arguing that it is especially important to achieve a good fit at the treatment cutoff.[8] Certainly, one might also opt for using more than one linear combination of the outcome, but we refrain from doing so as with increasing number of pre-treatment outcomes, it becomes more and more probable to run into the case discussed above, particularly if the number of covariates is small.[9] For a detailed discussion on specifications regarding pre-treatment outcome usage, including the case of all values and only one single value, respectively, see Ferman et al. (2017).

Example 1: Guinea-Bissau

As a first example for the consequences of using all pre-treatment outcomes as separate predictors, Figure 1 shows the Type B Experiment for Guinea-Bissau. While the black solid line depicts the actual GDP-per-capita trajectory of Guinea-Bissau, the red dashed line shows the synthetic control as calculated by Billmeier and Nannicini (2013). A comparison of these two lines seemingly reveals a negative, and therefore counterintuitive, treatment effect of liberalization in Guinea-Bissau. This result relies on the following donor units: Burkina Faso (16.3%), Burundi (40.8%), China (15.8%), Congo (0.7%), Malawi (25.8%), and Pakistan (0.6%). However, in line with our theoretical results of the previous section, the corresponding weighting scheme is characterized by zero weights for all predictors except the pre-treatment observations of GDP per capita. As a consequence, the chosen synthetic control unit's covariates, i.e., school enrollment, population growth, investment, and democracy do not compare well to the covariates of the treated unit (see Table 1). To illustrate how policy conclusions about the

---

[7] Note that this is the default of the statistical software package "SYNTH" (see Abadie et al. (2011)).

[8] One could also think about this as an "Ashenfelter's dip"-type argument. A treatment might be more likely to be triggered in period t if the preceding period $(t-1)$ was somewhat an outlier. Hence, it might be important to capture that particular lag. A similar approach is taken by Montalvo (2011), who uses the last two outcome lags as economic predictors.

[9] The case of utilizing only very few covariates—while considering all available outcome lags as separate

11

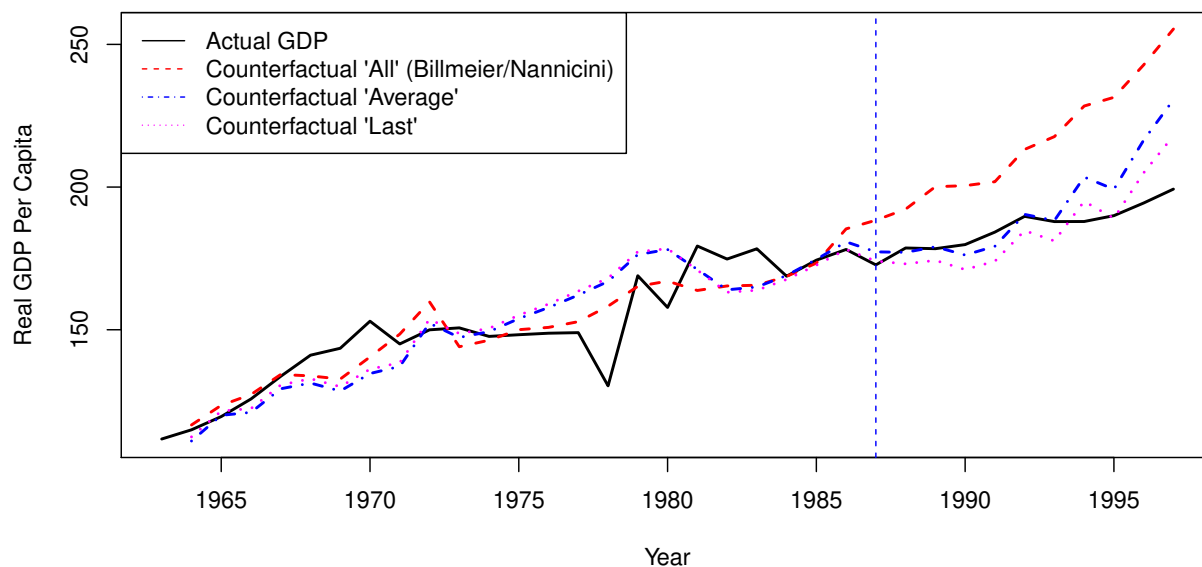## Guinea–Bissau: Type B Experiment

Figure 1: Trends in Per Capita (%) GDP: Guinea-Bissau vs. Synthetic Guinea-Bissau (Type B Experiment)

treatment effect of liberalization change by effectively incorporating the covariates, we now restrict the lags of the dependent variable within the inner optimization to the pre-treatment average of GDP per capita.[10] As a consequence, optimal weights for school enrollment, population growth, investment, and democracy now are given by $1.000212 \cdot 10^{-06}\%$, $3.570276 \cdot 10^{-05}\%$, 84.9756%, and 7.4437% respectively. Basically, the investment share and the democracy indicator are the predictors gaining noticeable weights. These new weights lead to a completely new mix of donor units representing the new synthetic Guinea-Bissau: Burundi (19.18%), Chad (1.04%), China (12.08%), and Malawi (67.70%). Additionally, the predictor values of our new synthetic Guinea-Bissau are mostly closer to the values of actual Guinea-Bissau. As depicted in Table 1, this is especially true for the important, i.e. highly weighted, predictors "investment share" and "democracy", which are fitted exactly, but also for "secondary school".

With regard to the empirical consequences of this issue, first note that the fit of GDP per capita *before* the treatment—the blue dashed-dotted line of the new synthetic Guinea-

---

predictors—occurs, e.g., in Hinrichs (2012).

[10] Note that we constructed corresponding v-weights by using the nested optimization, not by the regression based approach.

Table 1: v Weights and Covariates for Guinea-Bissau (Type B Experiment)

	'All' (v)	'Av.' (v)	'Last' (v)	'All'	'Av.'	'Last'	Actual
Secondary School	0	0.0000	0.4221	8.79	7.92	7.25	7.25
Population growth	0	0.0000	0.0000	2.19	2.53	2.55	1.90
Investment share	0	84.9756	99.3814	0.10	0.13	0.13	0.13
Democracy	0	7.4437	0.1959	0.02	0.00	0.00	0.00

Note: Calculations are based on data from Billmeier and Nannicini (2013). The first column reports the four covariates, columns two to four report predictor weights v (in percent) for the three estimators (explained below), columns five to seven report respective covariate values for the pre-treatment period, and the last column reports analogous covariate values for the *Actual* Guinea-Bissau. The counterfactual *All* is calculated using the four covariates appearing in the table as well as all lagged values of GDP per capita as separate economic predictors—as do Billmeier and Nannicini (2013). The counterfactual *Av.* is calculated using the four covariates appearing in the table and only the average pre-treatment value of GDP per capita as economic predictors. The counterfactual *Last* is calculated using the four covariates appearing in the table and only the last pre-treatment value of GDP per capita as economic predictors. The Type B Experiment is constructed using a donor pool of all eligible countries.

Bissau—behaves very similar to the one in Billmeier and Nannicini (2013) (the red dashed line). However, it becomes evident that there is a remarkable change in the estimated treatment effect. The difference in Figure 1 between the black solid line and the blue dashed-dotted line *post* treatment now reveals a rather non-existent effect of economic liberalization on GDP.

Our conclusion of no treatment effect is confirmed by looking at our results for the estimator where we make use of the last pre-treatment observation of the dependent variable as additional economic predictor (see the magenta dotted timeline in Figure 1). As one might expect, we observe the W weights of the two corresponding synthetic control units being very close to each other.[11] With respect to the pre-treatment fit of the covariates, Table 1 shows again that predictor values of the new synthetic Guinea-Bissau are almost all closer to those of actual Guinea-Bissau than the ones by Billmeier and Nannicini (2013), with even three out of four covariates fitted exactly.

Example 2: Barbados

When there is a notable and meaningful, i.e. positive, effect of liberalizing a respective country's economy on GDP per capita, the question arises whether this estimated treatment effect is, in fact, significant. In this respect, we look at the case of the Type A Experiment for Barbados, which is given in Figure 2. As becomes evident from visual inspection, treatment effect

---

[11] For the "average estimator" we have Burundi 19.18%, Chad 1.04%, China 12.08%, and Malawi 67.70%, for the "last estimator" we have Burundi 18.91%, Chad 2.01%, China 9.93%, and Malawi 69.15%.

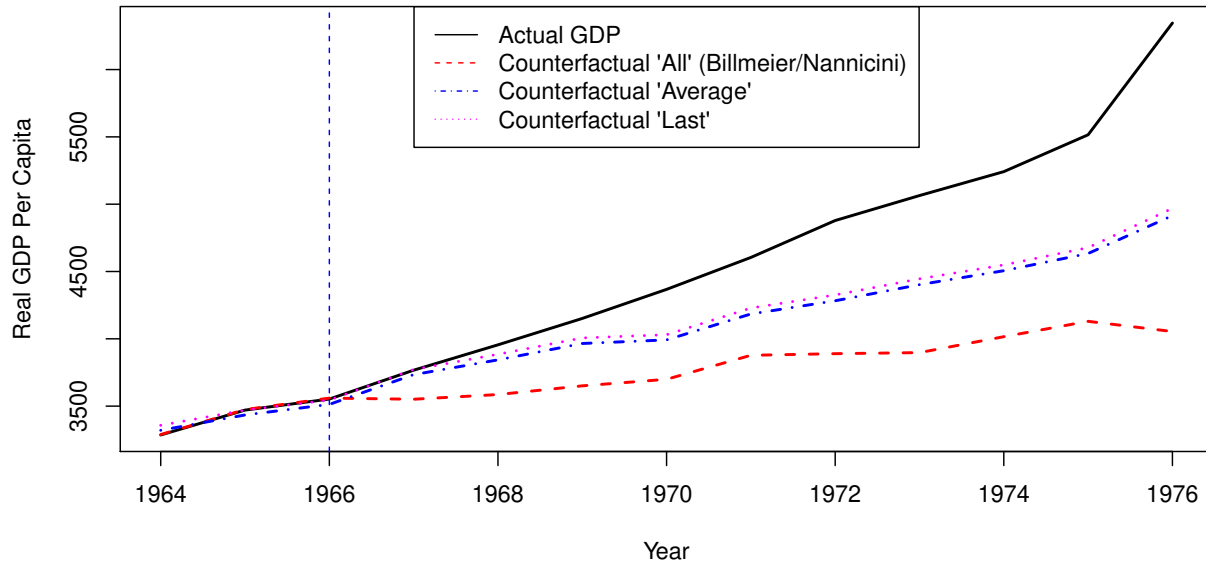**Barbados: Type A Experiment**

Figure 2: Trends in Per Capita (%) GDP: Barbados vs. Synthetic Barbados (Type A Experiment)

estimation results can be significantly inflated by using all pre-treatment observations of GDP separately versus using its average or its last value in addition to the set of covariates. Again, our two alternative estimators provide quite resembling solutions based on a very similar W weighting mix.[12]

For simplicity, we therefore exclusively focus on the first estimator (using the average) when comparing the significance of the estimated effect to the results from Billmeier and Nannicini (2013) using a standard SCM placebo check.[13] More precisely, in this placebo study, one reassigns the treatment to a comparison unit so that the treated unit moves into the donor pool while one of the control units is synthesized instead. Applying this idea to each country from the original donor pool allows to compare the estimated effect for Barbados to the distribution of placebo effects obtained for other countries.

As a consequence, when checking the effect's significance by comparing the post-treatment

---

[12] The "average estimator" uses Argentina with 26.73% and Trinidad & Tobago with 73.27%, while the "last estimator" makes use of the same countries with 24.89% and 75.11% respectively.

[13] Corresponding results for the estimator using the last outcome lag as additional predictor are available upon request.

Barbados: Type A Experiment – Placebo Study
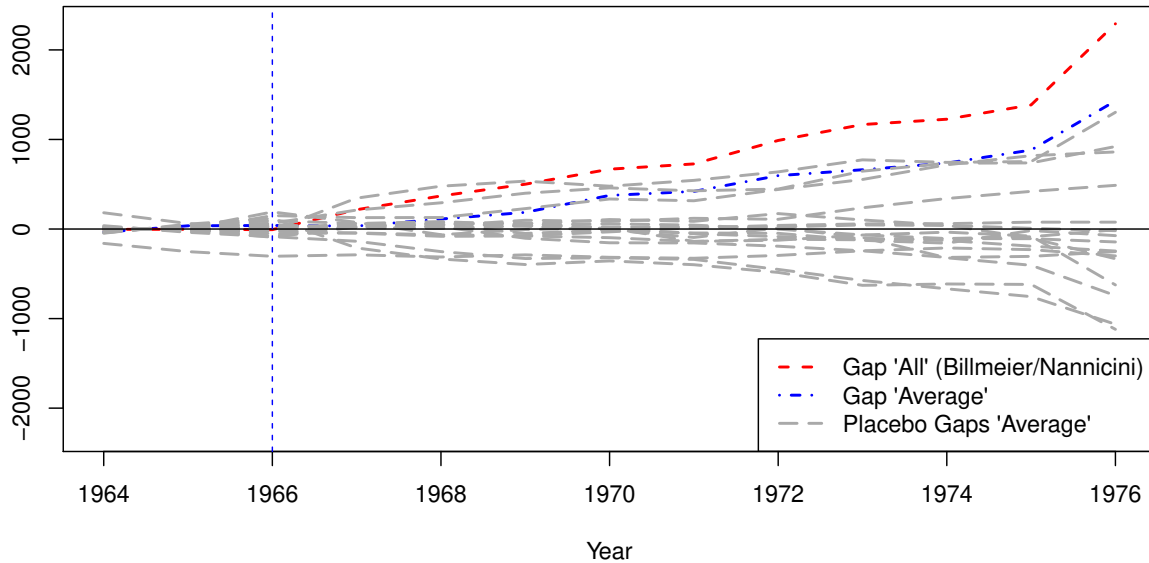

Figure 3: Per Capita (%) GDP Gaps for Barbados (Type A Experiment) and Placebo Countries

difference of the actual and the synthetic trajectory (the "gap") between Barbados and other control units in Figure 3, we come to a conclusion very different from Billmeier and Nannicini (2013). While solely relying on lags of the dependent variable throughout the inner optimization, the authors find a significant treatment effect, since the red dashed gap timeline clearly stands out of the mass of placebos. Billmeier and Nannicini (2013) call it a "positive and robust impact of economic liberalization", since their placebo tests confirm that "none of the fake experiments in the potential controls is above the effect in the treated country".[14] However, when one restricts the numbers of outcome lags, the corresponding blue dashed-dotted gap trajectory mainly lies within the "confidence band", not revealing a significant effect of economic liberalization on GDP for Barbados.

Example 3: Gambia

Even though, throughout the many empirical applications in Billmeier and Nannicini (2013), we rarely find cases where the use of the pre-treatment average leads to markedly different

---

[14] Note that the mass of placebos look nearly identical for the "All' and the "Average" case.
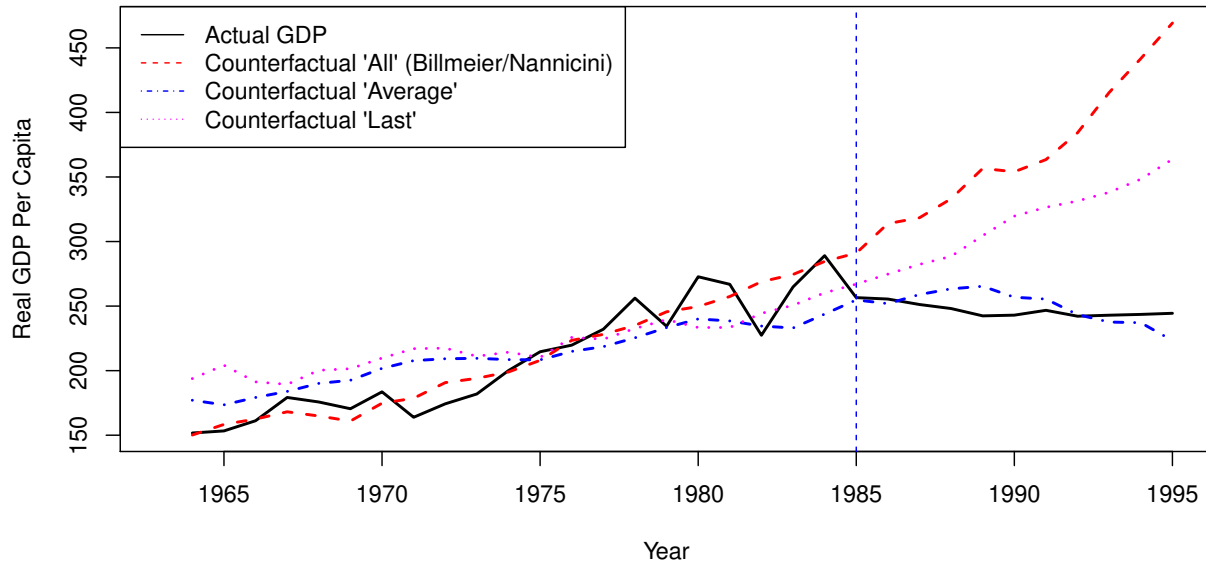
Gambia: Type B Experiment

Figure 4: Trends in Per Capita (%) GDP: Gambia vs. Synthetic Gambia (Type B Experiment)

results than the use of the last pre-treatment value as additional economic predictor,[15] such circumstances exist. As Figure 4 demonstrates, corresponding W weights can be distinctly disparate, leading to treatment effect estimation results which significantly differ across the three alternatives. Since Billmeier and Nannicini (2013) only rely on lags of the dependent variable throughout the inner optimization, they conclude that Gambia experienced a huge negative impact on GDP from economic liberalization. As for the case of Guinea-Bissau, this is at odds with economic theory. Moreover, this result cannot be confirmed by using an estimator relying on only the average of outcome lags as additional predictor. Here, we do not observe any difference between actual and synthetic Gambia. Using the last value of pre-treatment GDP, though, another synthetic Gambia arises lying somewhere in the middle between the estimator used by Billmeier and Nannicini (2013) and the estimator using the average of pre-treatment outcomes. Hence, when looking at this example, estimation results seem to be

---

[15] Certainly, there are also several cases where there is no clear difference between the solution shown by Billmeier and Nannicini (2013) and the two estimators outlined in this section. In such cases, one might even think about choosing the alternative of using solely all outcome lags as predictors, since the covariates seem not to play a role in predicting the dependent variable and, hence, can possibly be ignored in order to improve efficiency.

pretty sensitive to the specific inclusion of certain pre-treatment values of the outcome. In the case of Gambia, we therefore cannot infer any stable economic conclusion since, depending on the respective estimator, results differ drastically, reflecting an unstable weighting mix of the synthetic control.[16] Consequently, one might want to reconsider the data at hand, economic theory, specific covariates, or even the method of identification, but should not simply proceed with any of the many possible, and different, SCM estimation results.

5   Conclusion

When applying SCM, it is important to consider pre-treatment observations of the dependent variable throughout the outer optimization to ensure a good fit prior to the intervention. However, the prime objective of SCM is to build a synthetic control that properly reflects how the treated unit would have evolved *after* the intervention if the latter had not taken place. To achieve this goal, covariates with predictive power for the variable of interest should be matched, too. As stated by Abadie et al. (2015), "it is of crucial importance that synthetic controls closely reproduce the values that variables with a large predictive power on the outcome of interest take for the unit affected by the intervention." Thus, it is important that the explicitly chosen covariates are allowed to influence the estimated synthetic control.

As shown theoretically, however, using all pre-treatment outcomes as separate predictors leads to optimizing the pre-treatment fit of the outcome only, rendering all covariates irrelevant. The upside of this, i.e., the achievement of an optimal pre-treatment fit of the outcome,[17] comes at the cost of ignoring the entire set of covariates, leading to a potentially biased estimator. Empirically, then, when comparing results from Billmeier and Nannicini (2013), who did not restrict the use of pre-treatment outcomes as predictors, to results when using only one pre-treatment value of the outcome, estimated treatment effects can change drastically.

---

[16]This is similar in spirit to the issue addressed by Ferman et al. (2017), who analyze to what extent results using synthetic control methods may be prone to "cherry picking" by searching for a specification that delivers a significant treatment effect.

[17] For instance, Nannicini and Billmeier (2011, fn. 21) even declare that each observation of the pre-treatment outcome "is used as a separate predictor to improve the fit".

This is due to the fact that when restricting the usage of outcome lags within the inner optimization, other economic predictors eventually attain positive weights. Consequently, the algorithm finds a differently weighted set of control units for synthesizing a respective unit of interest, including a better fit with respect to the (observed) covariates. Similar issues can be expected for various other pieces of research.

Nevertheless, one might still calculate the version that uses all lagged values of the outcome as separate predictors. If this estimate is very close to corresponding specifications with restrictions on the outcome lags, the covariates seem to be rather unimportant for predicting the outcome, and ignoring them might lead to a variance reduction (see, e.g., Kaestner et al., 2017). Moreover, researchers could also deliberately decide to construct their synthetic control unit based on past values of the outcome variables *alone*, because, e.g., there is no theoretical foundation for using covariates or simply because there are no covariates available due to data restrictions. Then, as in a univariate time series approach, SCM rely exclusively on pre-intervention outcomes to predict the post-intervention path (see, e.g., Klößner and Pfeifer, 2017).

However, if SCM are used as originally intended by Abadie and Gardeazabal (2003) & Abadie et al. (2010), and as done in all previously mentioned studies throughout the paper, namely where covariates are (deemed) important and should be taken into account, it should not be a first-best solution to simply optimize the pre-treatment fit by using all pre-treatment values of the outcome as separate economic predictors.

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: Stata module to implement synthetic control methods for comparative case studies. Statistical Software Components, Boston College Department of Economics.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *The American Economic Review*, 93(1):113–132.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics*, 121(2):368–391.

Athey, S. and Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

Bauhoff, S. (2014). The effect of school district nutrition policies on dietary intake and overweight: A synthetic control approach. *Economics and Human Biology*, 12:45–55.

Belot, M. and Vandenberghe, V. (2014). Evaluating the "threat" effects of grade repetition: Exploiting the 2001 reform by the french-speaking community of Belgium. *Education Economics*, 22(1/2):73–89.

Bilgel, F. and Galle, B. (2015). Financial incentives for kidney donation: A comparative case study using synthetic controls. *Journal of Health Economics*, 43:103–117.

Billmeier, A. and Nannicini, T. (2013). Assessing economic liberalization episodes: A synthetic control approach. *The Review of Economics and Statistics*, 95(3):983–1001.

Bohn, S., Lofstrom, M., and Raphael, S. (2014). Did the 2007 Legal Arizona Workers Act reduce the state's unauthorized immigrant population? *The Review of Economics and Statistics*, 96(2):258–269.

Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013). Catastrophic Natural Disasters and Economic Growth. *The Review of Economics and Statistics*, 95(5):1549–1561.

Coffman, M. and Noy, I. (2012). Hurricane Iniki: measuring the long-term economic impact of a natural disaster using synthetic control. *Environment and Development Economics*, 17(2):187–205.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. NBER Working Paper No. 22791.

Eren, O. and Ozbeklik, S. (2016). What do right-to-work laws do? Evidence from a synthetic control method analysis. *Journal of Policy Analysis and Management*, 35(1):173–194.

Ferman, B. and Pinto, C. (2017). Revisiting the synthetic control estimator. Working Paper.

Ferman, B., Pinto, C., and Possebom, V. (2017). Cherry picking with synthetic controls. Working Paper.

Gardeazabal, J. and Vega-Bayo, A. (2017). An empirical comparison between the synthetic control method and Hsiao et al.'s panel data approach to program evaluation. *Journal of Applied Econometrics*, pages n/a–n/a.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *The Review of Economics and Statistics*, 98(3):535–551.

Hinrichs, P. (2012). The effects of affirmative action bans on college enrollment, educational attainment, and the demographic composition of universities. *The Review of Economics and Statistics*, 94(3):712–722.

Hosny, A. S. (2012). Algeria's trade with GAFTA countries: A synthetic control approach. *Transition Studies Review*, 19(1):35–42.

Jinjarak, Y., Noy, I., and Zheng, H. (2013). Capital controls in Brazil – stemming a tide with a signal? *Journal of Banking and Finance*, 37(8):2938–2952.

Kaestner, R., Garrett, B., Chen, J., Gangopadhyaya, A., and Fleming, C. (2017). Effects of ACA medicaid expansions on health insurance coverage and labor supply. *Journal of Policy Analysis and Management*, 36(3):608–642.

Kleven, H. J., Landais, C., and Saez, E. (2013). Taxation and international migration of superstars: Evidence from the European football market. *The American Economic Review*, 103(5):1892–1924.

Klößner, S., Kaul, A., Pfeifer, G., and Schieler, M. (2017). Comparative politics and the synthetic control method revisited: A note on Abadie et al. (2015). *Swiss Journal of Economics and Statistics*, pages n/a–n/a.

Klößner, S. and Pfeifer, G. (2017). Outside the box: Using synthetic control methods as a forecasting technique. *Applied Economics Letters*, pages n/a–n/a.

Kreif, N., Grieve, R., Hangartner, D., Turner, A. J., Nikolova, S., and Sutton, M. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Economics*, 25:1514–1528.

Liu, S. (2015). Spillovers from universities: Evidence from the land-grant program. *Journal of Urban Economics*, 87:25–41.

Montalvo, J. G. (2011). Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections. *The Review of Economics and Statistics*, 93(4):1146–1154.

Nannicini, T. and Billmeier, A. (2011). Economies in transition: How important is trade openness for growth? *Oxford Bulletin of Economics and Statistics*, 73(3):287–314.

O'Neill, S., Kreif, N., Grieve, R., Sutton, M., and Sekhon, J. S. (2016). Estimating causal effects: considering three alternatives to difference-in-differences estimation. *Health Services and Outcomes Research Methodology*, 16:1–21.

Stearns, J. (2015). The effects of paid maternity leave: Evidence from temporary disability insurance. *Journal of Health Economics*, 43:85–102.