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: March 9, 2018
(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, 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. Monte Carlo studies examine potential bias.

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

*Corresponding author: Gregor Pfeifer, University of Hohenheim, Department of Economics, 520 B, D-70593, Germany; g.pfeifer@uni-hohenheim.de; Phone 0049 711 459 22193. The authors are grateful for valuable remarks from participants of the IAAE and the German Statistical Week. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

1 Introduction

In their seminal papers, Abadie and Gardeazabal (2003) as well as Abadie, Diamond and Hainmueller (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, Diamond and Hainmueller (2010) and Abadie, Diamond and Hainmueller (2015), it is also vital that the synthetic control unit and the treated unit resemble one another with respect to pre-intervention val-

[1] After Abadie, Diamond and Hainmueller (2010), which builds on an initial idea in Abadie and Gardeazabal (2003), SCM were applied, i.a., by Montalvo (2011) (terrorist attacks), 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, Noy and Zheng (2013) (capital inflows), Kleven, Landais and Saez (2013) (taxation of athletes), Bauhoff (2014) (school nutrition), Belot and Vandenberghe (2014) (educational attainment), Bohn, Lofstrom and Raphael (2014) (Arizona's 2007 LAWA), Abadie, Diamond and Hainmueller (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 (2018) extend SCM to the forecasting context.

ues of *predictors* of the outcome variable ("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, Lofstrom and Raphael (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 three 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

[2] Abadie, Diamond and Hainmueller (2010) show that, under certain conditions, one obtains asymptotically unbiased estimates of the treatment effect if *both* optimizations are pursued. Furthermore, they show that under additional conditions, estimates may even be unbiased in small samples.

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

of the estimated treatment effect may become, possibly leading to wrong policy conclusions.

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

As a third contribution, we provide a Monte Carlo study which compares the performance of different estimators for the counterfactual path of the outcome. We find that using all outcome lags as economic predictors results in estimates being more biased and less precise in terms of root mean squared prediction error, as compared to estimators which effectively use the covariates by employing only one outcome-related predictor. Furthermore, we find in line with theory that these results are largely driven by the fact that covariates are fitted rather poorly when all outcome lags are used, introducing a bias that can be substantial even for reasonably long pre-treatment timespans.

The remainder proceeds as follows: Section 2 briefly outlines the synthetic control approach by Abadie, Diamond and Hainmueller (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 economic 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 provides a Monte Carlo study, which shows that using all lagged outcome values as economic predictors may render estimates more biased and less precise. Section 6 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 $k = M + r$) 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 combination 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 nonnegative control unit weights W summing up to unity such that:

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

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, Diamond and Hainmueller (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:

$$(Z_1 - Z_0 W^*(V))' (Z_1 - Z_0 W^*(V)) \xrightarrow{V} \min . \tag{3}$$

Eventually, there is a second data-driven way to determine the predictor weights, the so-called regression based method, used, e.g., by Bohn, Lofstrom and Raphael (2014).[4] This method proceeds as follows: first, for every time t prior to the intervention, the corresponding dependent variable, $Z_{j,t}$ ($j = 1, \dots, J+1$), is regressed on all economic predictors, yielding regression coefficients $\beta_{t,k}$ ($k = 1, \dots, K$). Second, v_k is then set as

$$v_k := \frac{\sum_t \beta_{t,k}^2}{\sum_{k=1\dots,K} \sum_t \beta_{t,k}^2} . \tag{4}$$

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

[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`.

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

$$W^{**} := \underset{W}{\operatorname{argmin}} \, (Z_1 - Z_0W)' \, (Z_1 - Z_0W) \tag{8}$$

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

$$(Z_1 - Z_0W^{**})' \, (Z_1 - Z_0W^{**}) \leq (Z_1 - Z_0W^*(V))' \, (Z_1 - Z_0W^*(V)) \tag{9}$$

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, Lofstrom and Raphael (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 $j = 1$ 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, $Y_{jT} = \widetilde{Y}_{jt} + \theta_t C_j$, is assumed to depend additively on the covariates by $\theta_t C_j$, with $\theta_t \in \mathbb{R}^{1\times r}$, and $\widetilde{Y}$ denoting how the dependent variable would evolve if it was not influenced by the covariates. One such example is the first data generating process considered by Abadie, Diamond and Hainmueller (2010, p. 495, Equation (1)), where $\widetilde{Y}_{jt} = \delta_t + \lambda_t \mu_j + \varepsilon_{jt}$, with

[5] For an empirical affirmation, see Bohn, Lofstrom and Raphael (2014, fn. 11).

δ_t denoting a common trend, λ_t being an unobservable time-varying confounder or a vector thereof, μ_j being the unobservable j-th unit's corresponding loadings, and ε_{jt} denoting white noise.

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

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

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 $\theta_t \Big(C_1 - \sum_{j=2}^{J+1} w_j C_j \Big)$ 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 θ_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 prediction 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 the pre-treatment timespan growing to infinity (for more details regarding the conditions for asymptotic (un)biasedness of SCM estimators, see Abadie, Diamond and Hainmueller, 2010;

[6] For example, 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., θ_t can be expected to be large.

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. For instance, for the data generating process mentioned above, the approximation of $\lambda_t \mu_1$ by $\lambda_t \sum_{j=2}^{J+1} w_j \mu_j$ might be better when the weights w are determined by ignoring the covariates. Overall, there thus exists a trade-off: (inadvertently) ignoring the covariates typically 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

For illustrating the issue of separately using all pre-treatment values of the outcome variable as economic predictors, we use data from Billmeier and Nannicini (2013), who analyzed the impact of economic liberalization on GDP. To construct synthetic controls, Billmeier and Nannicini (2013) 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 restrictions 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, Diamond and Hainmueller (2015), and Kleven, Landais and Saez (2013).[7] Second, we opt for the *last* pre-treatment value to be included

[7] Note that this is the default of the statistical software package "SYNTH" (see Abadie, Diamond and Hainmueller, 2011).

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, Pinto and Possebom (2017).

Case Study: Economic Liberalization in Guinea-Bissau

For showcasing the consequences of using all pre-treatment outcomes as separate predictors, Figure 1 exhibits the results for the so-called "Type B experiment" for Guinea-Bissau.[10] 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 treatment effect of liberalization change by effectively incorporating the covariates, we now restrict the lags of the dependent variable

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

[10] 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.

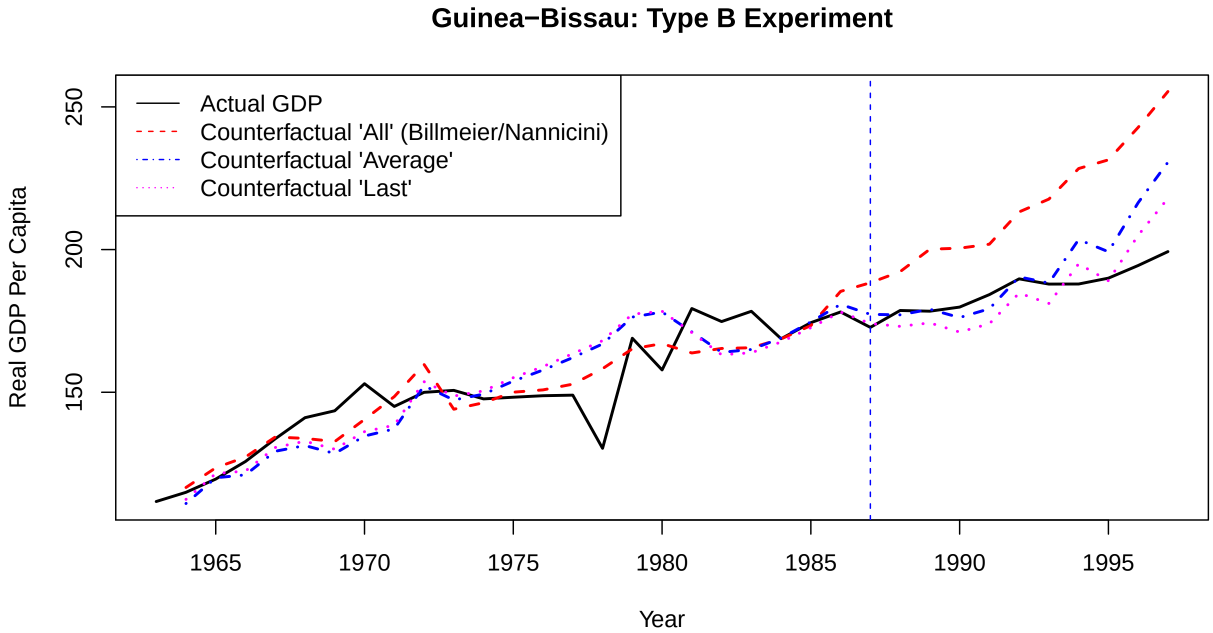

Figure 1: Trends in Per Capita (%) GDP: Guinea-Bissau vs. Synthetic Guinea-Bissau (Type B Experiment)

within the inner optimization to the pre-treatment average of GDP per capita.[11] As a consequence, optimal weights for school enrollment, population growth, investment, and democracy now are given by $4.82 \cdot 10^{-7}\%$, $1.72 \cdot 10^{-5}\%$, 48.20%, and 3.59% 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-Bissau—behaves very similar to the one in Billmeier and Nannicini (2013) (the red dashed

[11] 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	33.3298	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	48.2029	33.3299	0.10	0.13	0.13	0.13
Democracy	0	3.5941	0.1959	0.0103	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.

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.[12]

Our conclusion of no treatment effect is confirmed by looking at our results for the estimator which uses 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.[13] 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, corresponding to the positive predictor weights for these covariates.

[12] If interested in more formal inference-like testing, the SCM literature usually relies on placebo exercises. The appendix provides one more example of the many case studies of Billmeier and Nannicini (2013), the Type A experiment for Barbados, including a particular emphasis on inference.

[13] 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 9.92%, and Malawi 69.15%.

5 Monte Carlo Results

Recall that we have discussed in Section 3 how effectively ignoring the covariates may lead to an additional small-sample bias when estimating the outcome's counterfactual development. Based on the empirical example from the previous section (Guinea-Bissau), we will now present Monte Carlo results which show that such a bias actually arises in relevant situations.[14] To this end, we use the following setup: for the data generating process, we take the factor model of Abadie, Diamond and Hainmueller (2010), where the values of the variable of interest are given by

$$Y_{jt}^N = \delta_t + \theta_t C_j + \lambda_t \mu_j + \varepsilon_{jt}, \tag{12}$$

where j denotes the j-th unit, t denotes time, Y_{jt}^N describes the counterfactual outcome in the absence of the treatment, δ_t is a common trend, θ_t is an unobservable r-dimensional row vector of time-varying factors, the covariates C_j give the r-dimensional column vector of the j-th unit's corresponding factor loadings, λ_t is an unobservable time-varying confounder or a vector thereof, μ_j are the unobservable j-th unit's corresponding loadings, and ε_{jt} denotes noise, uncorrelated with respect to both time and units. While, for the non-treated donor units, Equation (12) describes the actual development of Y both pre- and post-treatment, for the treated unit, Equation (12) is assumed to describe actual pre-treatment and counterfactual post-treatment values.

The pre-treatment values of Y as well as the observed values of the covariates, C, are used to calculate SCM weights w, using as economic predictors either all lags of the dependent variable ('All'), only the last lag ('Last'), or the pre-treatment average ('Average'). Corresponding SCM weights w are then used to calculate the difference of the treated unit's counterfactual and synthetic pre-treatment values

$$Y_{1t}^N - \sum_{j=2}^{J+1} w_j Y_{jt}^N = \theta_t \left(C_1 - \sum_{j=2}^{J+1} w_j C_j \right) + \lambda_t \left(\mu_1 - \sum_{j=2}^{J+1} w_j \mu_j \right) + \varepsilon_{1t} - \sum_{j=2}^{J+1} w_j \varepsilon_{jt}. \tag{13}$$

[14] A more large-scaled Monte Carlo study, varying important quantities like for instance the number of pre-treatment observations or the number of covariates, is beyond the scope of this paper.

Building on the Guinea-Bissau example discussed above, we take the data on covariates C as given, i.e. we use the original data on secondary school enrollment, population growth, investment as a share of GDP, and the democracy dummy. We assume that each of these covariates drives an idiosyncratic change of the GDP per capita's level and slope:

$$\theta_t = (a_1 + b_1 t, a_2 + b_2 t, a_3 + b_3 t, a_4 + b_4 t). \tag{14}$$

The time-specific additive constants, δ_t ($t = 1, \ldots, T$), as well as the parameters of the covariates' influence on GDP per capita, a_k, b_k ($k = 1, \ldots, 4$), are determined from the actual pre-treatment data by regressing the outcome on the linear trend and the covariates' terms, i.e., by minimizing[15]

$$\sum_{t=1,\ldots,T} \sum_{j=1,\ldots,J+1} \left(Y_{jt}^N - (\delta_t + \theta_t C_j) \right)^2 . \tag{15}$$

For the unobserved confounder, λ_t, we assume that it takes the form $100 \cdot (1 - \frac{1}{t})$, describing a level shift in GDP per capita of \$100 that takes time to be completely realized.[16] The unobserved values of μ, which determine each unit's share of this shift, are drawn from the uniform distribution on $[0, 1]$. Eventually, for simulating ε, we use Gaussian white noise with a standard deviation of \$100.

Table 2 shows the bias and root mean squared prediction error (RMSPE) of $\sum_{j=2}^{J+1} w_j Y_{jt}^N$ when estimating Y_{1t}^N over a post-treatment horizon of ten years, based on $10,000$ MC replications.[17] Additionally, it also shows the part of the bias that is caused by the critical quantity $\theta_t(C_1 - \sum_{j=2}^{J+1} w_j C_j)$. The estimator using all lags of the outcome as economic predictors performs significantly worse than those that effectively use the covariates: this holds true for all forecast horizons and both in terms of bias and RMSPE. Concerning the estimators that do not

[15] Note that it is not necessary to determine post-treatment values of δ_t, as these cancel out from the difference $Y_{1t}^N - \sum_{j=2}^{J+1} w_j Y_{jt}^N$, see Equation (13).

[16] We have also tried several other timelines yielding essentially unchanged results, which are available on request. In particular, we tried quadratic growth ($\lambda_t = \frac{t^2}{10}$), slow, sublinear growth ($\lambda_t = 20\sqrt{t}$), a slowly decaying shock ($\lambda_t = \frac{100}{t}$), and complete absence of confounders ($\lambda_t = 0$).

[17] Calculations were carried out using R (see R Core Team, 2017) and the algorithms of Becker and Klößner (2018*a*) as supplied by package MSCMT (see Becker and Klößner, 2018*b*).

include all outcome lags as predictors, 'Last' turns out to be slightly better than 'Average', again both in terms of bias and RMSPE.

Table 2: Bias and RMSPE of Different Estimators

Horizon	Bias			Bias by Covariates			RMSPE		
	'All'	'Av.'	'Last'	'All'	'Av.'	'Last'	'All'	'Av.'	'Last'
1	61.28	41.05	29.06	62.52	42.83	30.16	130.67	130.19	127.48
2	67.52	46.41	33.49	68.12	46.34	33.54	133.50	132.53	127.93
3	73.78	49.17	37.31	73.73	49.86	36.93	135.57	130.94	126.74
4	77.74	50.34	38.43	79.33	53.37	40.31	140.06	134.56	130.62
5	84.41	55.13	42.46	84.94	56.89	43.70	143.58	136.36	131.61
6	90.97	59.74	47.30	90.55	60.40	47.08	147.22	137.04	131.67
7	94.78	60.79	48.28	96.15	63.92	50.46	151.19	138.58	133.64
8	100.65	65.16	53.12	101.76	67.43	53.85	154.27	140.03	135.00
9	107.76	71.05	57.42	107.36	70.95	57.23	159.51	143.20	137.34
10	111.89	72.75	59.45	112.97	74.46	60.61	162.28	143.77	138.34

Note: Bias and RMSPE of $\sum_{j=2}^{J+1} w_j Y_{jt}^N$ *as an estimator of* Y_{1t}^N *over different forecast horizons. The columns labeled 'Bias by Covariates' display that part of the bias that is due to* $\theta_t(C_1 - \sum_{j=2}^{J+1} w_j C_j)$. Example based on case study Guinea-Bissau Type B experiment from Billmeier and Nannicini (2013).

In line with the theory outlined above, comparing corresponding columns of Table 2 also shows that any estimator's bias is completely driven by the covariate-induced quantity $\theta_t \left(C_1 - \sum_{j=2}^{J+1} w_j C_j \right)$. As argued earlier, this quantity is on average much less close to zero when all outcome lags are used as economic predictors, because fitting the covariates then is completely ignored. In contrast, the corresponding quantities become much smaller when the covariates are accounted for, either by using the average or the last of the pre-treatment values.

6 Discussion

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, Diamond and Hainmueller (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,[18] 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. 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. As evaluated by our Monte Carlo study, this indeed makes a notable difference regarding respective bias and RMSPE. Similar issues can be expected for various other pieces of research.

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 results than the use of the last pre-treatment value as additional economic predictor, such circumstances exist:[19] there, estimation results seem to be pretty sensitive to the specific inclusion of certain pre-treatment values of the outcome. One therefore cannot simply infer a stable economic conclusion since, depending on the respective estimator, results differ drastically, reflecting an unstable weighting mix of the synthetic control. This is similar in

[18] 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".

[19] One example is the Type B experiment for Gambia for which full results are available upon request.

spirit to the issue addressed by Ferman, Pinto and Possebom (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. 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.

Moreover, there also appear cases where there is no clear difference between the solution shown by Billmeier and Nannicini (2013) and the two alternative estimators outlined above. In such cases, one might indeed consider using solely all outcome lags as predictors, since the covariates seem to be rather unimportant for predicting the dependent variable and, hence, can possibly be ignored in order to improve efficiency (see, e.g., Kaestner et al., 2017). Further, 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, 2018).

However, if SCM are used as originally intended by Abadie and Gardeazabal (2003) as well as Abadie, Diamond and Hainmueller (2010), and as done in most of the previously mentioned studies throughout this 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, 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. 2011. "SYNTH: Stata module to implement Synthetic Control Methods for Comparative Case Studies." *Statistical Software Components, Boston College Department of Economics*.

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." *The American Economic Review*, 93(1): 113–132.

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton. 2016. "The Value of Connections in Turbulent Times: Evidence from the United States." *Journal of Financial Economics*, 121(2): 368–391.

Athey, Susan, and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives*, 31(2): 3–32.

Bauhoff, Sebastian. 2014. "The Effect of School District Nutrition Policies on Dietary Intake and Overweight: A Synthetic Control Approach." *Economics and Human Biology*, 12: 45–55.

Becker, Martin, and Stefan Klößner. 2018*a*. "Fast and Reliable Computation of Generalized Synthetic Controls." *Econometrics and Statistics/The Annals of Computational and Financial Econometrics*, 5: 1–19.

Becker, Martin, and Stefan Klößner. 2018*b*. "MSCMT: Multivariate Synthetic Control Method Using Time Series." R package version 1.3.2.

Belot, Michèle, and Vincent Vandenberghe. 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, Fırat, and Brian Galle. 2015. "Financial incentives for kidney donation: A comparative case study using synthetic controls." *Journal of Health Economics*, 43: 103–117.

Billmeier, Andreas, and Tommaso Nannicini. 2013. "Assessing Economic Liberalization Episodes: A Synthetic Control Approach." *The Review of Economics and Statistics*, 95(3): 983–1001.

Bohn, Sarah, Magnus Lofstrom, and Steven Raphael. 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, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano. 2013. "Catastrophic Natural Disasters and Economic Growth." *The Review of Economics and Statistics*, 95(5): 1549–1561.

Coffman, Makena, and Ilan Noy. 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, Nikolay, and Guido W. Imbens. 2016. "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis." NBER Working Paper No. 22791.

Eren, Ozkan, and Serkan Ozbeklik. 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, Bruno, and Cristine Pinto. 2017. "Revisiting the Synthetic Control Estimator." Working Paper.

Ferman, Bruno, Cristine Pinto, and Vitor Possebom. 2017. "Cherry Picking with Synthetic Controls." Working Paper.

Gardeazabal, Javier, and Ainhoa Vega-Bayo. 2017. "An Empirical Comparison Between the Synthetic Control Method and Hsiao et al.'s Panel Data Approach to Program Evaluation." *Journal of Applied Econometrics*, 32(5): 983–1002.

Gobillon, Laurent, and Thierry Magnac. 2016. "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls." *The Review of Economics and Statistics*, 98(3): 535–551.

Hinrichs, Peter. 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, Amr Sadek. 2012. "Algeria's Trade with GAFTA Countries: A Synthetic Control Approach." *Transition Studies Review*, 19(1): 35–42.

Jinjarak, Yothin, Ilan Noy, and Huanhuan Zheng. 2013. "Capital Controls in Brazil – Stemming a Tide with a Signal?" *Journal of Banking and Finance*, 37(8): 2938–2952.

Kaestner, Robert, Bowen Garrett, Jiajia Chen, Anuj Gangopadhyaya, and Caitlyn Fleming. 2017. "Effects of ACA Medicaid Expansions on Health Insurance Coverage and Labor Supply." *Journal of Policy Analysis and Management*, 36(3): 608–642.

Kleven, Henrik Jacobsen, Camille Landais, and Emmanuel Saez. 2013. "Taxation and International Migration of Superstars: Evidence from the European Football Market." *The American Economic Review*, 103(5): 1892–1924.

Klößner, Stefan, and Gregor Pfeifer. 2018. "Outside the Box: Using Synthetic Control Methods as a Forecasting Technique." *Applied Economics Letters*, 25(9): 615–618.

Klößner, Stefan, Ashok Kaul, Gregor Pfeifer, and Manuel Schieler. 2017. "Comparative Politics and the Synthetic Control Method Revisited: A Note on Abadie et al. (2015)." *Swiss Journal of Economics and Statistics*, n/a–n/a.

Kreif, Noémi, Richard Grieve, Dominik Hangartner, Alex James Turner, Silviya Nikolova, and Matt Sutton. 2016. "Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units." *Health Economics*, 25: 1514–1528.

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

Montalvo, José 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, Tommaso, and Andreas Billmeier. 2011. "Economies in Transition: How Important Is Trade Openness for Growth?" *Oxford Bulletin of Economics and Statistics*, 73(3): 287–314.

O'Neill, Stephen, Noémi Kreif, Richard Grieve, Matthew Sutton, and Jasjeet S. Sekhon. 2016. "Estimating causal effects: considering three alternatives to difference-in-differences estimation." *Health Services and Outcomes Research Methodology*, 16: 1–21.

R Core Team. 2017. "R: A Language and Environment for Statistical Computing." Vienna, R Foundation for Statistical Computing.

Stearns, Jenna. 2015. "The effects of paid maternity leave: Evidence from Temporary Disability Insurance." *Journal of Health Economics*, 43: 85–102.

Appendix: Economic Liberalization in 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 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.[20]

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.[21] 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 difference of the actual and the synthetic trajectory (the "gap") between Barbados and the set of 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".[22] 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.

[20] 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.

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

[22] Note that the mass of placebos look virtually identical for the "All' and the "Average" case.

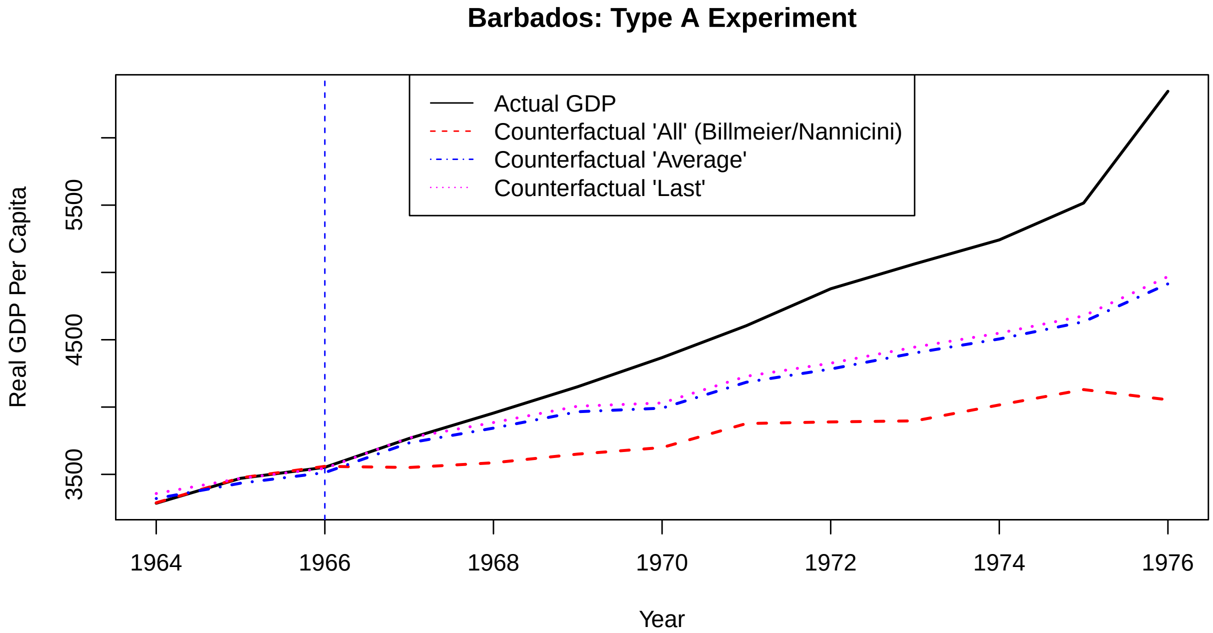

Figure 2: Trends in Per Capita (%) GDP: Barbados vs. Synthetic Barbados (Type A Experiment)

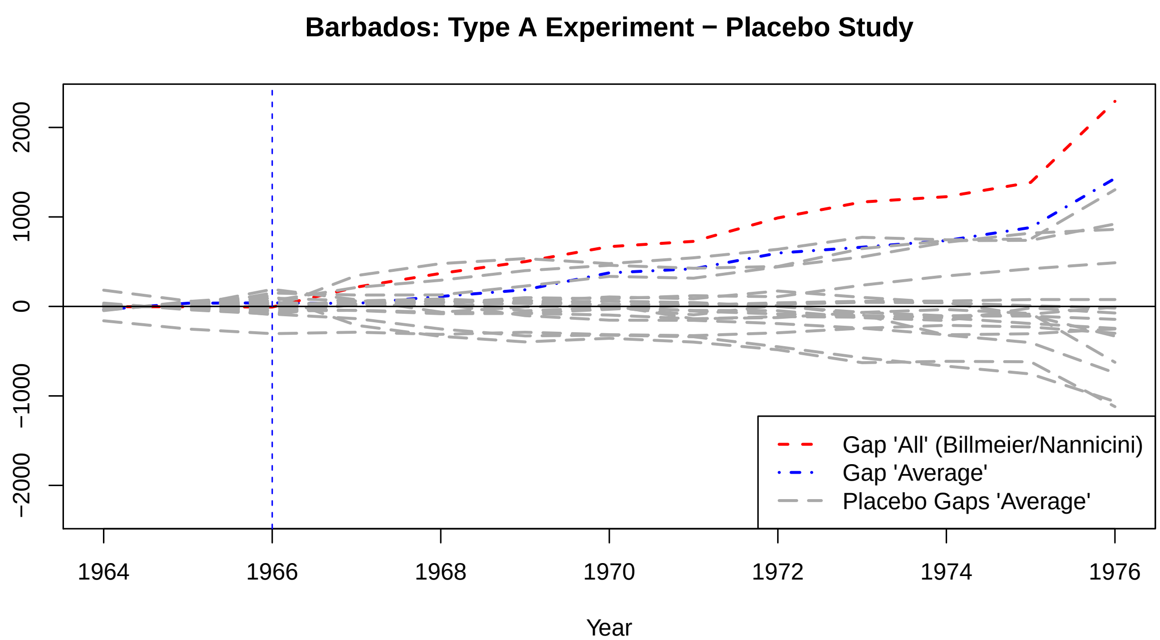

Figure 3: Per Capita (%) GDP Gaps for Barbados (Type A Experiment) and Placebo Countries