On the Use of Synthetic Difference-in-differences Approach with (-out) Covariates: The Case Study of Brexit Referendum*

Esther de Brabander[a], Artūras Juodis[a,b], Gabriela Szini[a,b,*]

[a] *Amsterdam School of Economics, University of Amsterdam*
[b] *Tinbergen Institute*

Abstract

The Synthetic Control (SC) method has been a popular and dominant method to evaluate treatment and intervention effects in the last two decades. The method is powerful yet very intuitive to use both for empirical researchers and policy experts, but is not without shortcomings. As a response to this, the new Demeaned SC (DSC) and Synthetic Difference-in-differences (SDID) approaches were introduced in the literature. In this paper, we evaluate the relative benefits of using DSC and SDID using in-sample placebo analysis on the real data on the Brexit referendum, as well as an extensive Monte Carlo study. We show that the SDID estimator minimizes both the interpolation and the extrapolation biases, while the conventional SC and matching estimators only minimize the extrapolation and the interpolation biases, respectively. Overall, using the SDID methodology, we find that the estimated effect of the Brexit referendum on UK GDP at the end of 2018 and 2019 is higher than previously documented in the literature.

Keywords: Brexit, Synthetic Controls, Synthetic Difference-in-differences, Covariates.

JEL: B41, C32, E65, F42.

1. Introduction

As of 2016, many studies have been published on the negative short and long term consequences of the Brexit referendum[1] in 2016 on the UK economy. As opposed to the more descriptive analysis of some early papers, more recent studies focused on addressing the economic effects of this event more formally. For example, using a firm-level micro dataset, Bloom et al. (2019), employed the differences-in-differences (DID) method to investigate the impact of Brexit on firms. They estimated that Brexit gradually reduced investment by about 11% over the three years following the vote and reduced productivity by between 2% and 5%.

Born et al. (2019) was one of the early papers investigating the effect of the UK leaving the EU on GDP while establishing causality without relying on disaggregated data. They argued that the Brexit vote can be considered a natural experiment to conduct policy evaluation regarding the counterfactual non-Brexit trajectory of the UK economy. The quasi-experimental setting advocated by Born et al. (2019) allows for the straightforward application of several modern methodologies to estimate the treatment effect, one being the Synthetic Control (SC) method introduced in a series of papers by Abadie and Gardeazabal (2003), Abadie et al. (2010) and Abadie et al. (2015). Using a particular implementation of the SC approach with covariates, Born et al. (2019) reported 2.4% decrease in GDP due

*We would like to thank Dmitry Arkhangelsky, Jad Beyhum, Irene Botosaru, Hippolyte Boucher, Maurice Bun, Bo Honoré, Marco Ross Fernandes, Andras Lengyel, Timo Schenk, Martin Weidner and conference and seminar participants at the NESG 2022 (Groningen), IPDC 2022 (Bertinoro), IAAE 2022 (London), Baltic Economic Conference 2022 (Kaunas), Universidade Sao Paulo (FEA-USP), Amsterdam Econometrics Workshop 2022, and the PhD Lunch seminar (University of Amsterdam) for comments and suggestions. Financial support from the Netherlands Organization for Scientific Research (NWO) under research grant number $451-17-002$ is gratefully acknowledged by the second and the third authors.

*Corresponding author. Amsterdam School of Economics, University of Amsterdam, The Netherlands, Roetersstraat 11, 1018 WB, Amsterdam, The Netherlands. Email: g.m.m.szini@uva.nl.

Email addresses: eodebrabander@gmail.com (Esther de Brabander), A.Juodis@uva.nl. (Artūras Juodis), g.m.m.szini@uva.nl. (Gabriela Szini)

[1]From now on we refer implicitly use the term"Brexit" to denote the effect of the corresponding referendum, as the time span we consider ends in 2019.

to Brexit at the end of 2018.

Recently proposed extensions of the SC method robustify the approach towards more general underlying Data Generating Processes (DGP). One such suggestion is the Synthetic DID estimator (SDID), proposed by Arkhangelsky et al. (2021). This method additionally allows for both unit fixed effects in the model and time weights in the construction of the counterfactual, automatically bias-adjusting the counterfactual estimators. Arkhangelsky et al. (2021) show (both theoretically and numerically) that the SDID estimator possesses double robustness properties and has the potential to outperform the SC and DID estimators in terms of bias and RMSE. Moreover, Doudchenko and Imbens (2016) and Ferman and Pinto (2021) proposed the Demeaned Synthetic Control (DSC) method, which constitutes a natural middle ground between the SC and the SDID methods. However, the results in both papers are rather limited with regards to the setup where additional covariates are available for matching as in Abadie et al. (2010), and Botosaru and Ferman (2017). This paper fills this gap.

This paper. Overall, the contributions of this paper are two-fold.

1. First of all, using the approach of Kellogg et al. (2021), we decompose the total bias of the SDID estimator into the interpolation and extrapolation biases. We show that the SDID estimator minimizes both components, while the conventional SC and matching estimators only minimize the extrapolation and the interpolation biases, respectively. Hence, while the time weights in the SDID setup were initially included to account for differences between the pre-treatment periods and the intervention period, these weights also minimize the interpolation bias. Moreover, motivated by the common empirical practice of SC methods, we extend the DSC and SDID methods to allow for matching on covariates.

2. Finally, we use the three mentioned methods (with and without covariates) to re-

examine the empirical evidence of Born et al. (2019) on the effect of Brexit on the UK GDP. For this purpose, we introduce several variations on the model/estimator specifications with respect to: (i) treatment date;(ii) different approaches to covariates; (iii) different specification for time weights horizons; (iv) the inclusion of a regularization parameter in the methods. Thus we avoid the "cherry-picking" problem recently emphasized by Ferman et al. (2020).

Our results (overall) indicate stronger effect of Brexit than the one described by Born et al. (2019). We complement the in-sample placebo analysis with a controlled empirical Monte Carlo study. For this purpose, we control the relative strength of the idiosyncratic and the common (factor) components as well as covariates. The results indicate that, as predicted, the SDID approach is mostly beneficial when the common component is non-negligible in comparison with the idiosyncratic component. At the same time, it is almost cost-free when the common component is negligible, as long as the idiosyncratic component is not dominant.

Structure. The remainder of this paper is structured as follows. In Section 2, we introduce the methods. In Section 3, we provide some insights on the SDID approach. The effect of Brexit on the UK GDP is re-evaluated based on procedures discussed in Section 4. Section 5 contains a dedicated empirical Monte Carlo study. Finally, Section 6 concludes. The Supplementary Online Appendix contains additional theoretical discussions, empirical robustness checks and further Monte Carlo results.

Notation. A bold letter refers to a vector and a capital bold letter denotes a matrix. Let $\boldsymbol{\imath}_S$ be the $[S \times 1]$ vector of ones, and $\boldsymbol{W}_S = \boldsymbol{I}_S - S^{-1}\boldsymbol{\imath}_S\boldsymbol{\imath}_S'$ the corresponding within transformation matrix. Moreover, the following definitions will be used later: $\boldsymbol{y}_j = (y_{j,1}, \ldots, y_{j,T_0-1})'$ denotes the $[(T_0-1)\times 1]$ vector of outcomes in all time periods prior to treatment for unit j. Combining all outcomes of control units yields the $[(T_0-1)\times J]$ matrix $\boldsymbol{Y} = (\boldsymbol{y}_2, \ldots, \boldsymbol{y}_{J+1})$.

The matrix $\boldsymbol{X}_j = (\boldsymbol{x}_j^{(1)}, \ldots, \boldsymbol{x}_j^{(P)})$ is a $[(T_0 - 1) \times P]$-dimensional matrix with columns $\boldsymbol{x}_j^{(p)} = (x_{j,1}^{(p)}, \ldots, x_{j,T_0-1}^{(p)})$, this matrix stores the values of all covariates prior to treatment for unit j.

2. Methodology

2.1. The Model

In what follows, we assume that we observe a balanced panel with $j = 1, \ldots, J+1$ units over $t = 1, \ldots, T$ time periods. For each unit j and time period t, the outcome of interest $y_{j,t}$, P explanatory variables $\boldsymbol{x}_{j,t}$, and the binary variable $W_{j,t}$ are observed. The binary variable $W_{j,t}$ equals 1 if the unit j is exposed to treatment at time t, and zero otherwise.

Conforming to the potential outcome framework of Rubin (1974), let the potential outcomes for unit j in period t under control and treatment,respectively, be denoted by $y_{j,t}^0$ and $y_{j,t}^1$. The treated potential outcome is defined as $y_{j,t}^1 = y_{j,t}^0 + W_{j,t}\tau_t$, where τ_t is the treatment effect at t. Without loss of generality, designate the first unit to be the treatment unit ($j = 1$) and the treatment period to be $t = T_0 < T$. Hence, there are J control units and $T - T_0 + 1$ observations post-treatment. The observed outcomes are

$$y_{j,t} = W_{j,t}y_{j,t}^1 + (1 - W_{j,t})y_{j,t}^0. \tag{1}$$

The objective of the analysis is to estimate the treatment effect at any time $t \geq T_0$. Evidently, y_{1,T_0}^0, which is required to compute the treatment effect, is unobserved. All methods discussed in this paper create counterfactual values $\widehat{y}_{1,t}^0$ for treated units using a donor pool of control units, which is then used to estimate the potential outcome of interest in the absence of treatment.

Using $\widehat{y}_{1,t}^0$ the parameter of interest can be naturally estimated as,

$$\widehat{\tau}_t = y_{1,t}^1 - \widehat{y}_{1,t}^0, \quad t = T_0, \ldots, T. \tag{2}$$

In the remainder of this section, we formally introduce the SC estimator (with and without matching on covariates) as originally proposed e.g. in Abadie and Gardeazabal (2003), Abadie et al. (2010) and Abadie et al. (2015). Later on, we discuss the DSC approach of Ferman and Pinto (2021) and Doudchenko and Imbens (2016), followed by the SDID method of Arkhangelsky et al. (2021).

There are certain conditions that need to be satisfied for all approaches to be valid. In particular, it is critical that the setting is such that one can assume that the treatment mechanism satisfies the "no anticipation" and the "no interference" conditions. That is, the intervention has no effect on the outcome before the implementation period, and the treatment does not affect control units before and after treatment. However, in applications, both effects might be present. Abadie (2021) suggests some empirical strategies to mitigate the corresponding effects.

Remark 1. In what follows, we will only consider point estimates $\widehat{\tau}_t$ and completely ignore the topic of inference on such quantities. Inference is generally very hard to conduct for this type of problem. However, recent work by Chernozhukov et al. (2021) opens up some interesting possibilities to attack this problem for a restricted class of Data Generating Processes (DGPs).

2.2. Synthetic Control

The SC method assigns weights to the control units such that the outcome variable for the treated unit and the weighted average of the control units are approximately equal for periods prior to treatment. In particular, the unit weights $\boldsymbol{\omega} = (\omega_2, \ldots, \omega_{J+1})'$ are chosen such that the Mean Squared Prediction Error (MSPE) between the outcome of interest over all periods prior to treatment of the treated unit and of the counterfactual is minimized:

$$\widehat{\boldsymbol{\omega}}^{sc} = \underset{\boldsymbol{\omega} \in W}{arg\,min} \left\{ \sum_{t=1}^{T_0-1} \left(y_{1,t} - \sum_{j=2}^{J+1} \omega_j y_{j,t} \right)^2 \right\}, \tag{3}$$

where $\mathbb{W} := \left\{ \boldsymbol{\omega} \in \mathbb{R}^J : \omega_j \geq 0 \text{ for each } j \in \{2, \ldots, J+1\} \text{ and } \sum_{j=2}^{J+1} \omega_j = 1 \right\}$.

Using the above estimated weights $\widehat{\boldsymbol{\omega}}^{sc}$, the counterfactual (for instance) at the period of treatment $t = T_0$ is then computed as $\widehat{y}_{1,T_0}^{0,SC} = \sum_{j=2}^{J+1} \widehat{\omega}_j^{sc} y_{j,T_0}$. The corresponding treatment effects estimator is then defined as:

$$\widehat{\tau}_{T_0}^{sc} = y_{1,T_0} - \sum_{j=2}^{J+1} \widehat{\omega}_j^{sc} y_{j,T_0}. \tag{4}$$

Alternatively, if one restricts her attention to the treatment period T_0 only, the above estimate can be obtained through the regression function:

$$\left(\widehat{\tau}_{T_0}^{sc}, \widehat{\boldsymbol{\beta}} \right) = \underset{\tau, \boldsymbol{\beta}}{arg\,min} \left\{ \sum_{j=1}^{J+1} \sum_{t=1}^{T_0} \left(y_{j,t} - \beta_t - W_{j,t}\tau \right)^2 \widehat{\omega}_j \right\}, \quad \forall \quad \widehat{\omega}_1 \neq 0. \tag{5}$$

In the original paper of Abadie and Gardeazabal (2003), the main suggested SC implementation involved an extension of Eq. (3), where one matches not only the outcome of interest, but also an additional set of covariates. Next, we discuss this extended approach.

Matching on covariates. Let $\boldsymbol{z}_j$ be a set of observed covariates for any unit $j = 1, \ldots, J+1$. Consider a general setting where we search for the weights $\boldsymbol{\omega}$ such that the pre-treatment fit of K variables for $\boldsymbol{z}_1$ is well approximated by the linear combination of un-treated $\{\boldsymbol{z}_j\}_{j=2}^{J+1}$. In most applications, K is chosen such that $K = P + M$, where one includes P summary statistics of P covariates (one for each covariate for simplicity), and M linear combinations of the pre-treatment outcomes $\boldsymbol{y}_1$. One typical choice is to include each pre-treatment outcome as a predictor separately. Therefore, $M = T_0 - 1$. As for P, it is a common practice (e.g.Born et al. 2019) to consider the mean over all pre-treatment periods for all available regressors. For this standard choice, $\boldsymbol{z}_1 = (\overline{\boldsymbol{x}}_1', \boldsymbol{y}_1')'$, and similarly for the $[K \times J]$ matrix $\boldsymbol{Z}$. In most cases, we would like to weight different elements of $\boldsymbol{z}_1$, as covariates might be of different predictive power. Denote by $\boldsymbol{v}$ a $[K \times 1]$ vector with non-negative entries, and $\boldsymbol{V} = \text{diag}(\boldsymbol{v})$ the corresponding diagonal matrix.

Abadie et al. (2010) suggested estimating weights $\boldsymbol{v}$ jointly with $\boldsymbol{\omega}$ through a nested optimization procedure. This method estimates the weights $\boldsymbol{v}$ that minimize the MSPE

of the pre-treatment outcomes $\boldsymbol{y}_1$ and its counterfactuals over the pre-treatment periods, while for a given $\boldsymbol{V}$ the weights $\boldsymbol{\omega}(\boldsymbol{V})$ minimize the distance between the predictors of the treatment unit and its counterfactuals, weighted by the relative importance of each predictor. More formally:

$$\widehat{\boldsymbol{\omega}}^{sc}(\boldsymbol{V}) = \underset{\boldsymbol{\omega} \in W}{arg\,min} \quad (\boldsymbol{z}_1 - \boldsymbol{Z}\boldsymbol{\omega})' \boldsymbol{V} (\boldsymbol{z}_1 - \boldsymbol{Z}\boldsymbol{\omega}) \tag{6}$$

$$\widehat{\boldsymbol{V}}^{sc} = \underset{\boldsymbol{V} \in V}{arg\,min} \quad (\boldsymbol{y}_1 - \boldsymbol{Y}\widehat{\boldsymbol{\omega}}^{sc}(\boldsymbol{V}))' (\boldsymbol{y}_1 - \boldsymbol{Y}\widehat{\boldsymbol{\omega}}^{sc}(\boldsymbol{V}). \tag{7}$$

Here $\mathbb{V}$ is the set of K-dimensional diagonal positive semi-definite matrices $\boldsymbol{V} = \text{diag}(\boldsymbol{v})$. For a given choice of $\widehat{\boldsymbol{\omega}}^{sc}(\widehat{\boldsymbol{V}}^{sc})$, the counterfactual and the treatment effect estimate are obtained analogously to the case without matching on covariates.

Requirements and properties. The key assumption for the SC method is that the vector of pre-treatment outcomes and covariates (in the case of matching on covariates) of the treated unit is in (or close to) the convex hull of the vectors of the control units. This ensures the existence of a weighted average of the control units that is able (approximately) to track the treated unit. An exact match of the counterfactual to the treated unit in all periods prior to treatment only occurs when $\boldsymbol{z}_1$ is in the convex hull of $\boldsymbol{Z}$. In practice, this is very difficult to achieve, and hence generally weights are found that minimize the distance rather than create an exact match. Abadie et al. (2010) recommends that if the characteristics of the treated unit are poorly matched by the synthetic control, the method should not be used.

Finally, Abadie et al. (2010) provided important results for the bias of the estimator. Namely, provided that there is a perfect pre-treatment fit in terms of outcomes and covariates, then the bias of the treatment effect estimator is controlled by the ratio between the variance of the individual transitory shocks and the number of pre-treatment periods T_0, meaning that the bias converges to zero as $T_0 \to \infty$. Moreover, in terms of bias, Kellogg et al. (2021) show that the SC estimator minimizes exactly the bound on the extrapolation bias

of the treatment effect, such that if this bound can be made zero, the SC estimator has zero extrapolation bias. However, it can still be susceptible to interpolation bias, as noted by Abadie et al. (2015). We discuss further such biases in Section 3.

2.3. Demeaned Synthetic Control

As Ferman and Pinto (2021) propose, the DSC estimator incorporates simple modifications as compared to the SC estimator. As the name suggests, the pre-treatment outcomes are demeaned before the estimation of the weights $\boldsymbol{\omega}$, such that:

$$\widehat{\boldsymbol{\omega}}^{dsc} = \underset{\boldsymbol{\omega} \in W}{arg\,min} \left\{ \sum_{t=1}^{T_0-1} \left((y_{1,t} - \overline{y}_1) - \sum_{j=2}^{J+1} \omega_j (y_{j,t} - \overline{y}_j) \right)^2 \right\}, \tag{8}$$

where $\mathbb{W} := \left\{ \boldsymbol{\omega} \in \mathbb{R}^J : \omega_j \geq 0 \text{ for each } j \in \{2, \ldots, J+1\} \text{ and } \sum_{j=2}^{J+1} \omega_j = 1 \right\}$. Here, the time-series averages of the form $\overline{y}_j$ are taken over the pre-treatment periods only.

In addition, a bias adjustment term is also incorporated in the definition of the estimated treatment effect:

$$\widehat{\tau}_{T_0}^{dsc} = y_{1,T_0} - \underbrace{\sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} y_{j,T_0}}_{\widehat{y}_{1,T_0}^{0,dsc}} - \underbrace{\left(\overline{y}_1 - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} \overline{y}_j \right)}_{\text{bias adjustment term}}. \tag{9}$$

Intuitively, the bias adjustment term removes the average differences between the observed treated unit and its counterfactual over the pre-treatment period, adjusting for level imbalances. When considering the regression function that originates the estimator above, we can see that the bias-adjustment term arises with the inclusion of a unit fixed-effect α_j in Eq. (5):

$$\left(\widehat{\tau}_{T_0}^{dsc}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}} \right) = \underset{\tau, \boldsymbol{\alpha}, \boldsymbol{\beta}}{arg\,min} \left\{ \sum_{j=1}^{J+1} \sum_{t=1}^{T_0} \left(y_{j,t} - \alpha_j - \beta_t - W_{j,t}\tau \right)^2 \widehat{\omega}_j^{dsc} \right\}, \quad \forall \quad \widehat{\omega}_1^{dsc} \neq 0. \tag{10}$$

The DSC estimator without matching on covariates is equivalent to the one concurrently proposed by Doudchenko and Imbens (2016). In essence, they propose to relax several constraints imposed by the standard SC method, one of them being the no-intercept restriction in estimating weights. Namely, it is proposed to allow for an intercept weight ω_0

(equivalent to demeaning) in the SC estimation and for a unit fixed effect in the regression function of the SC estimator. Doudchenko and Imbens (2016) also points out that allowing for this systematic additive difference between the treatment unit and the control units is a key feature of the standard Difference-in-differences (DID) method. When allowing for such terms, we can see the DSC estimator as a generalization of the DID, where the units are weighted by the estimated $\widehat{\boldsymbol{\omega}}$ - instead of fixed weights equal to $\boldsymbol{\omega} = J^{-1}\boldsymbol{\imath}_J$. The units-specific weights turn the DID regression local, while also permitting the lack of parallel trends, a key feature of the SC method. Indeed, this is seen in Eq. (10), which has a DID specification but includes a pre-specified unit weight.

Matching on covariates. Even though the DSC estimator was originally proposed for the case without matching on covariates, we suggest a fairly simple extension to allow for matching on them. Intuitively, one would think of simply demeaning for each unit $\boldsymbol{z}_1$, $\boldsymbol{y}_1$, $\boldsymbol{Z}$ and $\boldsymbol{Y}$. However, we follow the recommendation by Abadie (2021) and only demean the pre-treatment outcomes since multiple covariates of different scales can be included in $\boldsymbol{z}_1$ and $\boldsymbol{Z}$. Denoting by $\ddot{\boldsymbol{z}}_1$ the vector containing the demeaned pre-treatment outcomes and the (average of) covariates for the treated unit, and by $\ddot{\boldsymbol{Z}}$ the stacked analogous vectors for the control units, the nested optimization given by Eqs. (6)-(7) becomes:

$$\widehat{\boldsymbol{\omega}}^{dsc}(\boldsymbol{V}) = \underset{\boldsymbol{\omega}\in W}{arg\,min} \quad (\ddot{\boldsymbol{z}}_1 - \ddot{\boldsymbol{Z}}\boldsymbol{\omega})'\boldsymbol{V}(\ddot{\boldsymbol{z}}_1 - \ddot{\boldsymbol{Z}}\boldsymbol{\omega}) \tag{11}$$

$$\widehat{\boldsymbol{V}}^{dsc} = \underset{\boldsymbol{V}\in V}{arg\,min} \quad (\boldsymbol{y}_1 - \boldsymbol{Y}\widehat{\boldsymbol{\omega}}^{dsc}(\boldsymbol{V}))'\boldsymbol{W}_{T_0-1}(\boldsymbol{y}_1 - \boldsymbol{Y}\widehat{\boldsymbol{\omega}}^{sc}(\boldsymbol{V}). \tag{12}$$

We propose that the computation of the treatment effects is the same as in Eq. (9), with the weights estimated above.

Note that the equivalence between the formulations in Doudchenko and Imbens (2016) and Ferman and Pinto (2021) breaks down in the setting with covariates. In particular, the approach we consider in Eqs. (6)-(7) is not equivalent to including an intercept in the inner optimization since the covariates are not taken into account when demeaning.

Requirements and properties. Ferman and Pinto (2021) proposed the DSC to improve the SC in terms of bias since they show that the SC estimator is generally biased if the treatment assignment is correlated with unobserved confounders. To be specific, consider the Data Generating Process (DGP) of the following linear form

$$y_{j,t}^0 = \boldsymbol{\beta}' \boldsymbol{x}_{j,t} + \alpha_j + \boldsymbol{f}_t' \boldsymbol{\gamma}_j + \varepsilon_{j,t}, \tag{13}$$

with the common factor component $\boldsymbol{f}_t' \boldsymbol{\gamma}_j$ as in Bai (2009), Moon and Weidner (2015), Gobillon and Magnac (2016), Juodis (2020), Fernández-Val et al. (2021), Beyhum and Gautier (2019), among others.

In particular, considering imperfect pre-treatment fit and a model with non-diverging common factors[2] and a fixed number of control units J, the estimated SC weights (in general) do not converge in probability to weights that reconstruct the factor loadings of the treated unit ($\boldsymbol{\Gamma}' \widehat{\boldsymbol{\omega}}^{sc}$ do not match $\boldsymbol{\gamma}_1$)[3] even if $T_0 \to \infty$. The SC estimator may also be biased if the counterfactual fails to reconstruct the time-invariant fixed effect of the treated unit α_1. Therefore, the SC can be biased in settings where the DID is not.

As DSC takes into account not only the differences in levels between unit (α_j) but also unit-specific weights ($\widehat{\omega}_j^{dsc}$), it can dominate DID. It is also shown that in settings with both non-diverging and diverging common factors, the diverging elements would not generate asymptotic biases in the DSC estimator (unlike the SC estimator). However, one would need that the treatment assignment is uncorrelated with the non-diverging factors to establish asymptotic unbiasedness.

[2]Ferman and Pinto (2021) defines non-diverging common factors when the pre-treatment average of the first and second moments of the common factors converge in probability to a constant.

[3]Here $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_{J+1})'$ and $\boldsymbol{F} = (\boldsymbol{f}_1, \ldots, \boldsymbol{f}_{T_0-1})'$.

2.4. Synthetic Difference-in-Differences

In addition to the bias-correction term (and, therefore, the inclusion of a unit fixed effect and an intercept weight) outlined above in the DSC, the SDID proposed by Arkhangelsky et al. (2021) introduces time weights in the construction of the counterfactual. In particular, since trends and shocks exist in macroeconomic data not all time periods prior to the intervention are equally representative for the counterfactual in the treatment period.

The SDID estimator is computed by first estimating the unit weights as implemented by the DSC in Eq. (8) and additionally (and independently), estimating the time weights according to:

$$\widehat{\boldsymbol{\lambda}}^{sdid} = \underset{\boldsymbol{\lambda}\in L}{arg\,min} \left\{ \sum_{j=2}^{J+1} \left(y_{j,T_0} - \overline{y}_{T_0} - \sum_{t=1}^{T_0-1} \lambda_t (y_{j,t} - \overline{y}_t) \right)^2 \right\}, \tag{14}$$

where $\mathbb{L} := \left\{ \boldsymbol{\lambda} \in \mathbb{R}^{T_0-1} : \lambda_t \geq 0 \text{ for each } t \in \{1, \ldots, T_0 - 1\} \text{ and } \sum_{t=1}^{T_0-1} \lambda_t = 1 \right\}$. Here $\overline{y}_t$ are cross-sectional averages at each point in time $t = 1, \ldots, T_0$ without $y_{1,t}$ included. Precise estimation of time weights $\boldsymbol{\lambda}$ requires a large cross-sectional donor pool, as formally discussed by Arkhangelsky et al. (2021) where both $T_0 \to \infty$ and $J \to \infty$.

Once both weights are obtained, the treatment effect for the period T_0, for instance, is estimated as:

$$\widehat{\tau}^{sdid}_{T_0} = y_{1,T_0} - \underbrace{\sum_{j=2}^{J+1} \widehat{\omega}^{dsc}_j y_{j,T_0}}_{\widehat{y}^{0,dsc}_{1,T_0}} - \underbrace{\sum_{t=1}^{T_0-1} \widehat{\lambda}^{sdid}_t \left(y_{1,t} - \sum_{j=2}^{J+1} \widehat{\omega}^{dsc}_j y_{j,t} \right)}_{\text{bias adjustment term}}. \tag{15}$$

Note that the bias adjustment term is almost identical to that of the DSC estimator, with the exception that instead of taking a simple average over the pre-treatment periods, a weighted average (with estimated weights $\widehat{\boldsymbol{\lambda}}$) is considered. Hence the approach naturally interpolates between the DiD type of bias-adjustment (with $\lambda_{T_0-1} = 1$) and two-way fixed effects bias adjustment as in Ferman and Pinto (2021), where all pre-treatment λ_t are equal.

Moreover, the regression function that corresponds to the estimate above is similar to its counterpart in Eq. (10), except for the inclusion of the time weights, i.e.:

$$\left(\widehat{\tau}^{sdid}_{T_0}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}\right) = \underset{\tau,\boldsymbol{\alpha},\boldsymbol{\beta}}{arg\,min} \left\{ \sum_{j=1}^{J+1} \sum_{t=1}^{T_0} \left(y_{j,t} - \alpha_j - \beta_t - W_{j,t}\tau\right)^2 \widehat{\omega}^{dsc}_j \widehat{\lambda}^{dsc}_j \right\}, \quad \forall \widehat{\omega}^{dsc}_1 \quad \neq 0. \quad (16)$$

In the Supplementary Online Appendix, we also discuss how the definition for the weights $\boldsymbol{\lambda}$ can be extended to account for the matching on covariates.

2.5. Penalized Estimation of Unit Weights

The literature on SC methods includes several proposals on how to extend the standard approach by including an additional ridge penalty (regularization) in the construction of the estimator. There are several benefits of additional regularization: (i) it increases the dispersion of weights; and (ii) it guarantees uniqueness of the resulting weights $\boldsymbol{\omega}$. Ferman and Pinto (2021) argue that with dispersed weights (in population), even the standard SC estimator does not suffer from the inconsistency problems discussed in Section 2.3. Hence, regularization just ensures that weights are already sufficiently dispersed in finite samples.

In this paper, we follow the penalty term proposed by Arkhangelsky et al. (2021)[4] - in their case, specifically for the SDID estimator - for all methods discussed in this paper. We define the DSC weights (and similarly for SC without demeaning):

$$\widehat{\boldsymbol{\omega}}^{dsc} = \underset{\boldsymbol{\omega}\in W}{arg\,min} \left\{ \sum_{t=1}^{T_0-1} \left((y_{1,t} - \overline{y}_1) - \sum_{j=2}^{J+1} \omega_j (y_{j,t} - \overline{y}_j) \right)^2 \right\} + \zeta^2 (T_0 - 1) \|\boldsymbol{\omega}\|_2^2, \quad (17)$$

where as before $\mathbb{W} := \left\{ \boldsymbol{\omega} \in \mathbb{R}^J : \omega_j \geq 0 \text{ for each } j \in \{2, \ldots, J+1\} \text{ and } \sum_{j=2}^{J+1} \omega_j = 1 \right\}$, while the regularization parameter is set as:

$$\zeta = \left(T_{\text{post}}\right)^{1/4} \sqrt{\frac{1}{J\left(T_0 - 2\right)} \sum_{j=2}^{J+1} \sum_{t=1}^{T_0-2} \left(\Delta_{jt} - \bar{\Delta}\right)^2}. \quad (18)$$

[4]Although the initial idea of penalized weights goes back to (at least) Doudchenko and Imbens (2016).

Here $\Delta_{j,t} = y_{j,t+1} - y_{j,t}$ and $\bar{\Delta} = \frac{1}{J(T_0-2)} \sum_{i=2}^{J+1} \sum_{t=1}^{T_0-2} \Delta_{j,t}$, and T_{post} denotes the number of post-treatment periods.

Note that T_{post} differs depending whether one is interested in one or multiple post-treatment periods. In particular, for the in-sample place tests different T_{post} are applicable: in cases (i) and (ii), it is equal to one, since we match only in one post-treatment period; in case (iii), up to the period of interest when the treatment effects are being assessed).

Remark 2. We also consider the same regularization approach for the remaining methods discussed in this paper. For the nested optimizations considered when matching on covariates, this term is introduced in the inner optimization, where it is optimized over the weights $\boldsymbol{\omega}$. Moreover, as Arkhangelsky et al. (2021), we do not include a ridge penalization for the estimation of the time weights. In the preliminary version of the paper, we also considered penalized time weights $\boldsymbol{\lambda}$, but in the context of our empirical applications, it plays no role.

3. Discussion

3.1. Bias-reducing Properties of the SDID

Recall the DGP as in Eq. (13). As pointed out by Ferman and Pinto (2021), for the estimator DSC to be unbiased the difference between the counterfactual and treated components of common shocks, $\boldsymbol{f}_t'(\boldsymbol{\gamma}_1 - \boldsymbol{\Gamma}'\widehat{\boldsymbol{\omega}}^{dsc})$, should be small for all t. However, since the goal is to well approximate the entire unobserved factor component at the post-treatment periods only, the relevant biases will also be decreased when $\boldsymbol{F}'\widehat{\boldsymbol{\lambda}}^{sdid}$ approximates $\boldsymbol{f}_{T_0}$. In other words, even if one set of weights fails to remove the biases from the correlation of treatment assignment to the unobserved components, the combination of both weights might still guarantee a decrease in biases. This indicates the double robustness property of SDID. Therefore, there are three bias reduction components in the SDID: (i) the double differencing - similar to the DID; (ii) the unit weights - similarly to SC; and (iii) the time weights.

3.2. Interpolation and Extrapolation Biases

In this section, we analyze and compare the biases of the SC, the DSC, and the SDID estimators. To be specific the bias is defined as the difference $Bias_{1,T_0} = y^0_{1,T_0} - \widehat{y}^0_{1,T_0}$ for any given estimator of $\widehat{y}^0_{1,T_0}$. In our analysis, we follow Kellogg et al. (2021) by decomposing the bias of the synthetic control estimator into interpolation and extrapolation bias. To simplify the notation, we assume that $y^0_{j,t}$ is some general deterministic function of observed and unobserved covariates $\boldsymbol{x}_{j,t}$, i.e. $y^0_{j,t}[\boldsymbol{x}_{j,t}]$. This way we also accommodate the common factor components $\boldsymbol{f}_t'\boldsymbol{\lambda}_j$ in the definition of $\boldsymbol{x}_{j,t}$.

First, consider the SC estimator. For this setup, the bias can then be decomposed as:

$$\begin{aligned} y^0_{1,T_0} - \widehat{y}^{0,sc}_{1,T_0} &= y^0_{1,T_0}[\boldsymbol{x}_{1,T_0}] - \sum_{j=2}^{J+1} \widehat{\omega}^{sc}_j y^0_{j,T_0}[\boldsymbol{x}_{j,T_0}] \\ &= \underbrace{\left(y^0_{1,T_0}[\boldsymbol{x}_{1,T_0}] - y^0_{1,T_0}\left[\sum_{j=2}^{J+1} \widehat{\omega}^{sc}_j \boldsymbol{x}_{j,T_0}\right]\right)}_{\text{extrapolation bias}} + \underbrace{\left(y^0_{1,T_0}\left[\sum_{j=2}^{J+1} \widehat{\omega}^{sc}_j \boldsymbol{x}_{j,T_0}\right] - \sum_{j=2}^{J+1} \widehat{\omega}^{sc}_j y^0_{j,T_0}[\boldsymbol{x}_{jT_0}]\right)}_{\text{interpolation bias}}. \end{aligned} \tag{19}$$

By definition, as it is the minimizer of the MSPE, the SC estimator minimizes the extrapolation bias. More specifically, when the treated unit falls within the convex hull of control units such that a perfect fit of the counterfactual can be constructed, the extrapolation bias is, by definition, zero. However, Kellogg et al. (2021) show that the SC will avoid interpolation bias only if $y^0_{j,t}[\boldsymbol{x}_{j,t}]$ is a linear function in $\boldsymbol{x}_{j,t}$. More specifically, assuming that $\boldsymbol{x}_{1,T_0}$ falls in the convex hull of $(\boldsymbol{x}_{2,T_0}, \ldots, \boldsymbol{x}_{J+1,T_0})'$, then the SC estimator avoids interpolation bias if and only if:

$$\sum_{j=2}^{J+1} \widehat{\omega}^{sc}_j y^0_{j,T_0}[\boldsymbol{x}_{j,T_0}] = y^0_{j,T_0}[\boldsymbol{x}_{1,T_0}] = y^0_{j,T_0}\left[\sum_{j=2}^{J+1} \widehat{\omega}^{sc}_j \boldsymbol{x}_{j,T_0}\right], \tag{20}$$

where we assume that the functional form of $y^0_{j,T_0}[\cdot]$ for $j = 2, \ldots, J+1$ and $y^0_{1,T_0}[\cdot]$ are the same. Not very surprisingly, the bias is larger when the paths of the dependent variable

are highly nonlinear in $\boldsymbol{x}_{j,t}$ (as documented by the finite sample evidence in Kellogg et al. 2021).

The SDID method poses a solution to the susceptibility of the synthetic control method to interpolation bias. For simplicity, assume that the treated unit falls within the convex hull of the control units and hence the extrapolation bias is zero, then the bias simplifies to:

$$y^0_{1,T_0} = \widehat{y}^{0,sc}_{1,T_0} + \left(y^0_{1,T_0} \left[\sum_{j=2}^{J+1} \omega_j^{sc} \boldsymbol{x}_{j,T_0} \right] - \sum_{j=2}^{J+1} \omega_j^{sc} y^0_{j,T_0}[\boldsymbol{x}_{j,T_0}] \right). \tag{21}$$

Conforming with the notation of Kellogg et al. (2021), the SDID estimator defined in Eq. (15) is given by:

$$\widehat{y}^{0,sdid}_{1,T_0} = \widehat{y}^{0,dsc}_{1,T_0} + \sum_{t=1}^{T_0-1} \widehat{\lambda}_t^{sdid} \left(y_{1,t}[\boldsymbol{x}_{1,t}] - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} y_{j,t}[\boldsymbol{x}_{j,t}] \right). \tag{22}$$

The SDID estimator is therefore very similar to the expression given in Eq. (21).[5] First note that under the assumption of zero extrapolation bias, it holds that $\sum_{j=2}^{J+1} \omega_j^{sc} \boldsymbol{x}_{j,t} = \boldsymbol{x}_{1,t}$ for all t and hence $y^0_{1T_0}[\sum_{j=2}^{J+1} \omega_j^{sc} \boldsymbol{x}_{j,T_0}] = y^0_{1,T_0}[\boldsymbol{x}_{1,T_0}]$. Moreover, the time weights $\widehat{\boldsymbol{\lambda}}$ are optimized such that the weighted average of pre-treatment periods is most representative of the actual treatment period $t = T_0$ (up to a constant). Therefore, while the unit weights minimize the extrapolation bias, the time weights introduced in the SDID estimator minimize the interpolation bias.

The interpolation bias of the SDID estimator is zero if the weights $\widehat{\boldsymbol{\lambda}}$ exist such that the difference in the outcome of the treated unit and the counterfactual in the treatment period falls within the convex hull of the difference between the treated unit and the

[5] One difference between the two equations is the fact that we use the DSC weights for the SDID estimate, but the same argument holds for the DSC estimator. Namely, inspecting Eq. (9), the DSC estimator would only minimize interpolation biases if a simple average of the pre-treatment periods well approximates the post-treatment period.

counterfactual in the periods prior to treatment up to a constant. More generally, similar to the construction of the unit weights for the counterfactual, if a perfect fit does not exist, time weights are found such that the distance is minimized. The SDID estimator is able to control for the interpolation bias of the SC estimator.

Other estimators proposed in the literature also aim to reduce the interpolation bias, e.g., in Kellogg et al. (2021) and Abadie and LHour (2021). The matching and synthetic control (MASC) estimator of Kellogg et al. (2021) relies on a model averaging of the SC estimator and matching estimators. Since the latter minimizes interpolation and the former extrapolation biases, the resulting estimator carries both properties. The penalized SC of Abadie and LHour (2021), even though it focuses more on providing a unique solution for the estimated unit weights, also reduces the interpolation bias by penalizing covariate discrepancy between the treated unit and each unit that contributes to the counterfactual (and not only the discrepancy between the treated unit and the counterfactual as a whole, as in the standard SC). Therefore, it is claimed that it reduces interpolation biases by assigning more weight to units that are close to the treated unit in the space of matching covariates. The disadvantage of such estimators is that they rely on tuning parameters, requiring cross-validation techniques.

Another estimator that focuses on reducing biases is the augmented synthetic control (ASCM) of Ben-Michael et al. (2021). However, it mainly corrects for biases arising from to imperfect pre-treatment fit and allows for more extrapolation. It relies on an outcome model to estimate such bias and uses it to de-bias the SC estimate. The drawback of this method is that it relies on a correct specification for the outcome model.

4. The Effect of the Brexit Referendum on the UK GDP Re-evaluated

4.1. The Dataset

The dataset we use comprises quarterly data for the GDP and six covariates for the 36 OECD countries, including the UK, for 1995:Q1-2020:Q4. The six covariates considered are: *Consumption, Investment, Exports, Imports, Labour productivity growth,* and *Employment share*.[6] All variables are standardized with respect to the mean and log-transformed to adjust for differences in size between the units. After removing all countries with missing data, 23 potential control/donor units are left, as in Born et al. (2019).

Despite the minor assumptions required to perform synthetic control estimation, the identification strategy of Born et al. (2019) is built upon the two key foundations:

1. First, the treatment date is set to be the date of the referendum in 2016:Q3.[7] More importantly, as the outcome of the vote was largely unexpected, it is reasonable to consider it a natural experiment.

2. Second, the voting behaviour was largely driven by political considerations (i.e. , re-claiming full freedom for political decision-making) rather than (macro-) economic reasons. Even though voting behaviour differed systematically among various socio-economic groups, the factors among which the voting behaviour differed (such as educational attainment, demography and regional industry structure) were claimed to be unlikely to impact macroeconomic performance.

[6]The variable Labour productivity growth is constructed by taking the log difference of quarterly real GDP and quarterly total employment, while the Employment share is the ratio between total employment and the working-age population.

[7]Notice that in Born et al. (2019) the authors mention that the treatment date is set to 2016:Q2, however, in their definition the treatment is materialized only in the period after the treatment date, that is, 2016:Q3. In our setting, we define that the treatment effect materializes in the treatment period.

These two assumptions are essential for the identification of the counterfactual trend that the UK's GDP would have followed if the referendum had not been held (or if it had failed).

4.2. The Zoo of Specifications and Methods

Methods. We consider the SC approach of Born et al. (2019) (SC(B)) as the benchmark for your comparative analysis. Their method differs with respect to the standard SC with matching on covariates since they consider both in the inner and the outer optimization in Eqs. (6)-(7) vectors and matrices stacking the pre-treatment outcomes and the average of (or last) covariates for the treated and control units, respectively. It can be argued, that the approach of Born et al. (2019)ensures that (by construction) the covariates are non-redundant. Importantly, this points out to the fact that the theoretical predictions of Kaul et al. (2021) do not hold for the SC(B) estimation method, as already acknowledged by Born et al. (2019).

In this paper, we also consider penalized (ridge) optimization procedures for $\boldsymbol{\omega}$ and $\boldsymbol{\lambda}$. As mentioned originally by Doudchenko and Imbens (2016), the inclusion of such a penalty leads to a higher dispersion, ensures the uniqueness of the weights, and delivers better interpretability of the unit weights. For the sake of simplicity, we adopt the penalization approach of Arkhangelsky et al. (2021) for unit weights in all estimators. Further discussion and the corresponding empirical result are provided in the Supplementary Online Appendix.

We estimate the SC(B), SC, and DSC methods using the concomitant *synth* and *synthdid* packages in *R*; see Abadie et al. (2010) and Arkhangelsky et al. (2021) for more details. Note that the package *synthdid* has one key difference with respect to the implementation of Born et al. (2019), related to the initialization for the $\boldsymbol{v}$ weights. While the latter uses the relative standard deviation of variables for the initialization, the former uses the initialization that provides the lowest MSPE between a regression-based initialization and

equal weights. This justifies why our estimates for the SC(B) method might differ slightly compared to the results in the original paper.

Specifications. In this paper, we extend the specifications considered by Born et al. (2019) in several ways:

i) We consider two possible treatment periods - 2016:Q2 and 2016:Q3. This is done for two reasons: due to the ambiguity of the date of the Brexit vote (at the end of 2016:Q2) and to consider possible anticipation effects.

ii) For the specifications with covariates, we also allow for the use of only the single most-recent period for matching. This is mostly done to better account for the unit-root properties of the included covariates.[8]

iii) When constructing $\boldsymbol{\lambda}$ weights for SDID, we allow for the possibility of matching on the first period of treatment (case (i)); on the average of post-treatment periods until the periods in which the treatment effect is being evaluated, that is, 2018:Q4 and 2019:Q4 (case (ii)); and on the outcomes in the periods in which the treatment effect is being evaluated only (case (iii)).

iv) Given the results of Kaul et al. (2021) (on the irrelevance of covariates with all pre-treatment outcomes included in the matching algorithm), we consider different matching schemes with all pre-treatment periods (the standard approach), as well as half pre-treatment periods and the last pre-treatment period (one pre-treatment period). These specifications are used to investigate the relevance of matching on covariates.

[8]However, we note that this specification should be mostly seen as an illustration for the possibilities discussed in Kaul et al. (2021), rather than reflecting the actual practise of using covariates in SC.

The extended set of specifications is mainly considered to avoid the cherry-picking problem recently highlighted by Ferman et al. (2020).

Despite the discussion in Section 2.4 on SDID with matching on covariates for $\boldsymbol{\lambda}$, in practice, we only match on the outcome of interest. The reasons are twofold: (i) the estimates indicate that matching or not on covariates, the time weights are almost always entirely assigned to the last pre-treatment period; (ii) for the unit weights there is not a well-defined "last covariate" value.

4.3. Empirical Results

Our main results are summarized in Tables 1 - 3. Here, we report the estimated percentage differences in GDP (counterfactual minus the UK) at the end of 2018 and 2019.

Equivalent to the original result of Born et al. (2019), the SC(B) with covariates (including all the pre-treatment periods and considering mean of covariates with treatment date 2016:Q3) estimated a reduction of 2.4% at the end of 2018. This number is smaller than the corresponding numbers from the competing procedures.

Overall, under the remaining specifications, the same pattern is observed - i.e., the remaining methods points to a larger effect in magnitude compared to the SC(B) method - with only two (non-consequential) exceptions. Importantly, considering the initial estimate of Born et al. (2019) of 2.4%, the SDID estimation with or without covariates point out to a much bigger economic loss due to Brexit under all specifications (except one in Table 3).

Other important conclusions from Tables 1 - 3 are: (i) irrespective of the method and/or specification considered the estimated gap in the GDP is increasing over time, suggesting a clear structural shift in the growth of the GDP; (ii) for the SDID approach (with or without covariates), the estimated effects are bigger when considering 2016:Q2 as treatment period compared to 2016:Q3. Therefore, the initial choice of treatment period is not innocuous.

All in all, there are several ways in which slightly different specifications compared to the one taken into account by Born et al. (2019) may lead to a more sizeable treatment effect.

Table 1: The estimated percentage differences in GDP between the counterfactuals and the UK including all pre-treatment periods as predictors.

	SC(B)	**SC**	**SC cov.**	**DSC**	**DSC cov.**	**SDID (i)**	**SDID cov. (i)**	**SDID (ii)**	**SDID cov. (ii)**	**SDID (iii)**	**SDID cov. (iii)**
Mean of covariates - Treatment period 2016:Q2											
2018:Q4	2.40	3.12	3.09	3.04	3.33	3.17	3.32	3.05	3.26	3.05	3.26
2019:Q4	3.60	4.28	4.23	4.18	4.56	4.31	4.54	4.19	4.48	4.19	4.48
Mean of covariates - Treatment period 2016:Q3											
2018:Q4	2.43	3.06	3.11	2.98	2.90	2.76	2.75	2.79	2.78	2.79	2.78
2019:Q4	3.61	4.20	4.28	4.12	4.00	3.89	3.85	3.92	3.89	3.92	3.89
Last covariates - Treatment period 2016:Q2											
2018:Q4	1.64	3.12	2.77	3.04	3.33	3.17	3.41	3.05	3.32	3.05	3.32
2019:Q4	2.53	4.28	4.05	4.18	4.49	4.31	4.58	4.19	4.49	4.19	4.49
Last covariates - Treatment period 2016:Q3											
2018:Q4	2.31	3.06	1.75	2.98	3.25	2.76	2.95	2.79	2.96	2.79	2.96
2019:Q4	3.37	4.20	2.87	4.12	4.41	3.89	4.10	3.92	4.12	3.92	4.12

Note: The results represent losses in percentages. The methods without covariates, SC, DSC, SDID (i), SDID (ii) and SDID (iii) naturally show the same results regardless if mean or last covariates are considered.

Table 2: The estimated percentage differences in GDP between the counterfactuals and the UK including half pre-treatment periods as predictors.

	SC(B)	**SC**	**SC cov.**	**DSC**	**DSC cov.**	**SDID (i)**	**SDID cov. (i)**	**SDID (ii)**	**SDID cov. (ii)**	**SDID (iii)**	**SDID cov. (iii)**
Mean of covariates - Treatment period 2016:Q2											
2018:Q4	1.33	3.12	1.34	3.04	3.15	3.17	3.31	3.05	3.33	3.05	3.33
2019:Q4	2.43	4.28	2.44	4.18	4.41	4.31	4.57	4.19	4.58	4.19	4.58
Mean of covariates - Treatment period 2016:Q3											
2018:Q4	1.33	3.06	1.37	2.98	2.80	2.76	2.60	2.79	2.61	2.79	2.61
2019:Q4	2.44	4.20	2.48	4.12	3.97	3.89	3.77	3.92	3.78	3.92	3.78
Last covariates - Treatment period 2016:Q2											
2018:Q4	1.55	3.12	2.53	3.04	2.91	3.17	3.14	3.05	3.03	3.05	3.03
2019:Q4	2.41	4.28	3.37	4.18	3.98	4.31	4.21	4.19	4.10	4.19	4.10
Last covariates - Treatment period 2016:Q3											
2018:Q4	2.42	3.06	2.46	2.98	3.12	2.76	2.74	2.79	2.75	2.79	2.75
2019:Q4	3.49	4.20	3.52	4.12	4.24	3.89	3.86	3.92	3.86	3.92	3.86

Note: For methods without covariates the results are equivalent to those in Table 1. See Table 1 for more discussion.

Table 3: The estimated percentage differences in GDP between the counterfactuals and the UK including one pre-treatment periods as predictor.

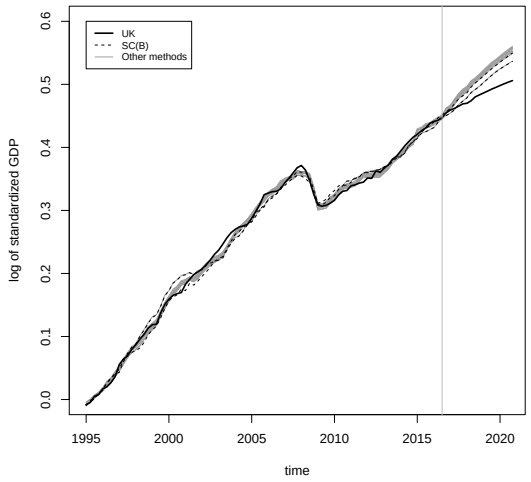
	SC(B)	**SC**	**SC cov.**	**DSC**	**DSC cov.**	**SDID (i)**	**SDID cov. (i)**	**SDID (ii)**	**SDID cov. (ii)**	**SDID (iii)**	**SDID cov. (iii)**
Mean of covariates - Treatment period 2016:Q2											
2018:Q4	1.24	3.12	0.41	3.04	0.98	3.17	2.15	3.05	2.20	3.05	2.20
2019:Q4	2.18	4.28	1.39	4.18	2.05	4.31	3.22	4.19	3.27	4.19	3.27
Mean of covariates - Treatment period 2016:Q3											
2018:Q4	0.99	3.06	1.49	2.98	1.34	2.76	1.71	2.79	1.75	2.79	1.75
2019:Q4	1.91	4.20	2.48	4.12	2.44	3.89	2.81	3.92	2.85	3.92	2.85
Last covariates - Treatment period 2016:Q2											
2018:Q4	2.74	3.12	3.11	3.04	2.88	3.17	3.14	3.05	2.83	3.05	2.83
2019:Q4	4.15	4.28	4.12	4.18	3.97	4.31	4.22	4.19	3.91	4.19	3.91
Last covariates - Treatment period 2016:Q3											
2018:Q4	2.48	3.06	2.88	2.98	2.58	2.76	2.18	2.79	2.18	2.79	2.18
2019:Q4	3.58	4.20	3.90	4.12	3.59	3.89	3.18	3.92	3.18	3.92	3.18

Note: For methods without covariates the results are equivalent to those in Table 1. See Table 1 for more discussion.

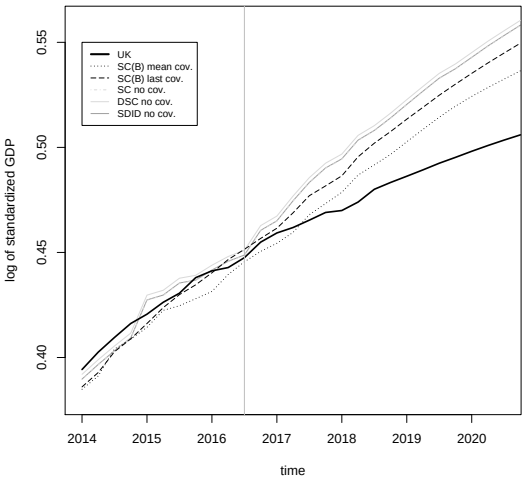
Figure 1 depicts some of the patterns we discussed above. Overall, all methods generate counterfactual trends that are able to replicate the GDP trend of the UK prior to the treatment date. In addition, all methods predict that the GDP of the UK would have been higher if they had not left the EU, but there are some key discrepancies between the methods. In the specifications considering the mean of covariates, the SC(B) and SC seem to underestimate the UK GDP, and it carries over to post-treatment periods. The DSC and SDID estimates are nearly identical to the SC (in terms of trajectory), but in general (with some exceptions when considering matching on the last covariates), shifted upwards, as a result of the constant bias adjustment over time. Notably, when the average of covariates is used in estimation, the bias adjustment seems to be much bigger in magnitude, as compared to other specifications. Hence, the inclusion of covariates and/or time weights does not have a homogeneous effect on results.

We also note that for numerical reasons, the theoretical prediction of Kaul et al. (2021) does not always hold across all specifications. In particular, even when considering all pre-treatment periods, there are still differences in the estimated effects with or without covariates. However, we can see that, especially for the DSC and the SDID methods the differences in the estimated treatment effects are larger when comparing matching or not matching on covariates for the case where one pre-treatment period is considered. In the Supplementary Online Appendix we show how estimated $\boldsymbol{v}$ and the corresponding $\boldsymbol{\omega}$ weights impact these conclusions.

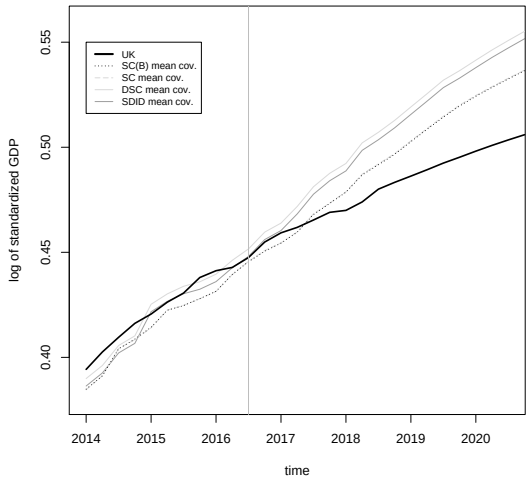
For the sake of illustration, in Table 4, we present estimated weights for one of our specifications. In particular, we see that for the SC, DSC and SDID methods, the estimated unit weights are quite similar when considering no covariates or matching on covariates with all pre-treatment periods. However, when considering only half or one pre-treatment period, the resulting $\boldsymbol{\omega}$ weights differ substantially. In the latter cases, the most striking result is that a much bigger weight is assigned to the U.S. As we discuss in the Supplementary



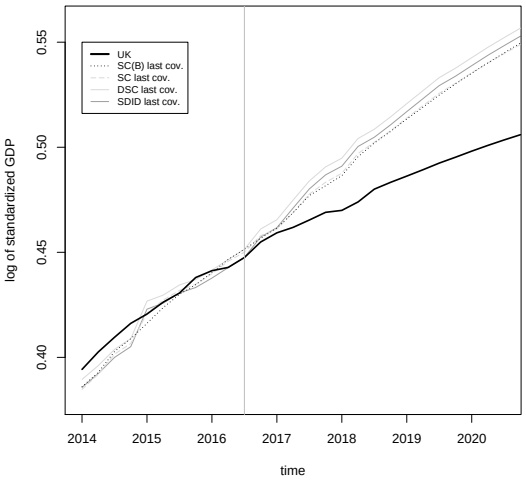
(a) SC(B) vs. all other methods.



(b) Counterfactuals (other than SC(B)) without covariates.



(c) Counterfactuals with the mean of covariates.



(d) Counterfactuals with the last value of covariates.

Figure 1: Estimated counterfactuals for treatment period 2016:Q3 and the half of pre-treatment outcomes.

Online Appendix, large weights of the United States and Hungary are not specific to the exact specification considered and are a general feature of this empirical problem.

	SC(B) all	SC(B) half	SC(B) one	SC no cov	SC all	SC half	SC one	DSC no cov	DSC all	DSC half	DSC one
Australia	0.0000	0.0000	0.0000	0.0052	0.0005	0.0000	0.0002	0.0075	0.0000	0.0000	0.0000
Austria	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.0000	0.0000	0.0000	0.0000
Belgium	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0036	0.0000	0.0000	0.0000	0.0001
Canada	0.0000	0.0001	0.0000	**0.1612**	**0.1734**	0.0000	0.0004	**0.1916**	**0.2103**	0.0705	0.0001
Finland	0.0000	0.0000	0.0000	0.0021	0.0002	0.0000	0.0003	0.0031	0.0103	0.0000	0.0000
France	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000
Germany	0.0458	**0.1145**	0.0000	0.0026	0.0000	**0.1012**	0.0004	0.0000	0.0000	0.0000	0.0379
Hungary	**0.1078**	0.0001	0.0000	**0.2186**	**0.2262**	0.0000	0.0011	**0.2311**	**0.2258**	**0.1697**	0.0000
Iceland	0.0089	0.0682	0.0000	0.0000	0.0076	0.0740	0.0006	0.0000	0.0000	0.0350	0.0569
Ireland	0.0114	0.0001	0.0000	0.0543	0.0462	0.0000	0.0002	0.0503	0.0469	0.0305	0.0000
Italy	**0.1744**	**0.1631**	**0.1667**	0.0353	0.0013	**0.1673**	**0.2098**	0.0334	0.0373	0.0471	**0.1773**
Japan	0.0000	0.0000	0.0000	**0.1773**	**0.1821**	0.0000	0.0000	**0.1842**	**0.1838**	**0.1884**	0.0000
Korea	0.0000	0.0000	0.0000	0.0030	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000
Luxembourg	0.0000	0.0483	0.0310	0.0000	0.0000	0.0371	0.0535	0.0000	0.0000	0.0000	0.0382
Netherlands	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0013	0.0000	0.0001	0.0000	0.0000
New Zealand	**0.1432**	0.0001	0.0586	0.0000	0.0000	0.0000	**0.1183**	0.0000	0.0144	0.0001	0.0001
Norway	0.0001	0.0000	0.0000	**0.1256**	**0.1133**	0.0000	0.0001	**0.1208**	**0.1042**	0.0000	0.0000
Portugal	0.0000	0.0001	0.0000	0.0123	0.0407	0.0000	0.0024	0.0037	0.0011	0.0000	0.0000
Slovak Republic	0.0000	0.0000	0.0000	0.0031	0.0000	0.0000	0.0005	0.0000	0.0000	0.0000	0.0000
Spain	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000
Sweden	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000
Switzerland	0.0000	0.0000	**0.1283**	0.0000	0.0000	0.0000	0.0169	0.0000	0.0000	0.0000	0.0001
United States	**0.5083**	**0.6052**	**0.6153**	**0.1994**	**0.2083**	**0.6203**	**0.5893**	**0.1740**	**0.1655**	**0.4587**	**0.6891**

Table 4: Estimated $\boldsymbol{\omega}$ weights considering 2016:Q3 as treatment period and mean of covariates. Weights > 0.1 are highlighted in Bold.

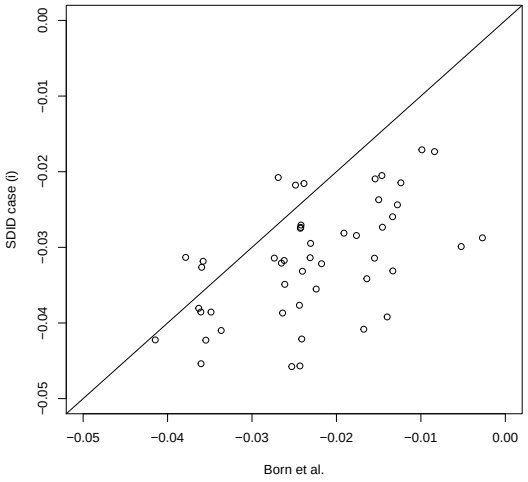
In this application, the time weights for SDID are always completely assigned to the last pre-treatment period. This outcome can be explained by the random-walk behaviour of the GDP series, see Online Supplementary Appendix for some visual evidence. This result can be seen as a pitfall of the SDID estimator when considering outcome levels. One might consider, alternatively, an empirical exercise with data in changes. However, a key feature of SC method is that the counterfactual estimate should actually match pre-exposure trends. Therefore, it is more intuitive and, in practice, more acceptable to match the levels of the series. We do not deviate from this practice.

The scatter plots in Figure 2 summarize the findings of this *Zoo of specifications and methods* (this includes all the specifications considered in the main text and in the Supplementary Online Appendix). Overall, we find that, in general, the estimated treatment effects are bigger in magnitude when: applying the SDID method as opposed to the SC(B) or SC methods; no covariates are taken into account when matching; and no penalty regularization is introduced.

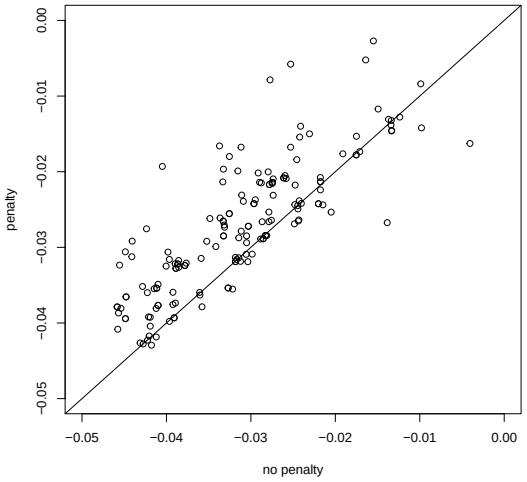
4.4. In-sample Placebo Analysis

Besides pointing out the differences among the estimators, no overall conclusions about the relative performance of the methods can be drawn from the above analysis. In what follows, we evaluate the relative performance of the considered procedures using the in-sample (time-series) placebo analysis as in Born et al. (2019).

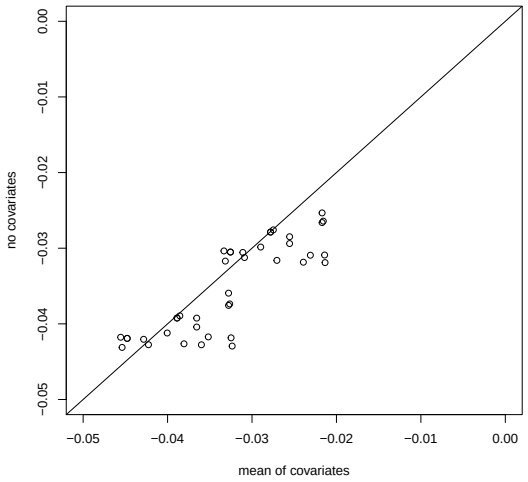
The idea of the in-sample placebo analysis is to advance the treatment date to $t = T_0' < T_0$ and construct a counterfactual using the data up to that date. The GDP level of the counterfactual at the artificial treatment date is then computed ($\widehat{y}_{1,T_0'}$) and compared to the actual realization ($y_{1,T_0'}$). Due to the no anticipation assumption for $T_0' < T_0$, the GDP level of the estimated counterfactual should be sufficiently close to the actual (observed) GDP value. This exercise is repeated for all time periods using the expanding window approach



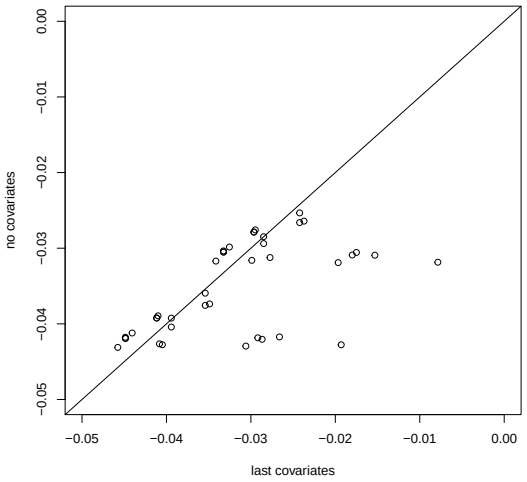
(a) SC(B) vs. SDID.



(b) With vs. without the penalty.



(c) No covariates vs. mean of covariates in estimation.



(d) No covariates vs. last covariate in estimation.

Figure 2: Scatter plots of estimated treatment effects, while keeping the remaining specifications/methods constant.

in the interval 2010:Q1-2014:Q4 one period at the time.

Note that for this placebo study we allow for the previously mentioned three different cases on how one considers post-treatment periods in the construction of the time weights $\boldsymbol{\lambda}$ for SDID. More specifically:

i) The first treatment period and evaluating the treatment effect in this period.

ii) The fourth quarter after the treatment period and computing the treatment effect for the end of the fourth quarter.

iii) The average over four quarters after the treatment period and computing the treatment effect for the end of the fourth quarter.

We consider these three settings since it can be that the time weights would not always be completely assigned to the last pre-treatment period. Thus, the cases above may lead to different estimates of treatment effects.

The results in Table 5 are not unexpected. In particular, the SC(B) generally performs the worst, while the SDID method (with or without covariates) for each specification always performs better. Importantly, the performance of the SDID is almost invariant to the inclusion of covariates. Strikingly, when considering half of the pre-treatment periods, the inclusion of covariates might even be detrimental for some of the methods. Finally, while the inclusion of the bias correction term in the DSC is sometimes beneficial, the inclusion of time effects (SDID) seems to have a larger effect an improving the estimates.

Table 5: In-sample placebo analysis across periods for 2010:Q1-2014:Q4.

	SC(B)	**SC**	**SC cov.**	**DSC**	**DSC cov.**	**SDID (i)**	**SDID cov.(i)**	**SDID (ii)**	**SDID cov.(ii)**	**SDID (iii)**	**SDID cov.(iii)**
Mean of covariates - All pre-treatments											
RMSE	0.0101	0.0089	0.0092	0.0087	0.0106	0.0067	0.0063	0.0134	0.0131	0.0134	0.0132
MAB	0.0090	0.0072	0.0076	0.0070	0.0087	0.0037	0.0040	0.0111	0.0114	0.0111	0.0115
MedAB	0.0096	0.0055	0.0071	0.0052	0.0074	0.0016	0.0021	0.0103	0.0112	0.0107	0.0112
Last covariates - All pre-treatments											
RMSE	0.0117	0.0089	0.0076	0.0087	0.0098	0.0067	0.0068	0.0134	0.0140	0.0134	0.0141
MAB	0.0098	0.0072	0.0057	0.0070	0.0078	0.0037	0.0041	0.0111	0.0123	0.0111	0.0123
MedAB	0.0093	0.0055	0.0048	0.0052	0.0055	0.0016	0.0020	0.0103	0.0113	0.0107	0.0115
Mean of covariates - Half of the pre-treatments											
RMSE	0.0108	0.0089	0.0107	0.0087	0.0104	0.0067	0.0056	0.0134	0.0100	0.0134	0.0100
MAB	0.0098	0.0072	0.0092	0.0070	0.0091	0.0037	0.0038	0.0111	0.0081	0.0111	0.0082
MedAB	0.0099	0.0055	0.0101	0.0052	0.0088	0.0016	0.0015	0.0103	0.0066	0.0107	0.0067
Last covariates - Half of the pre-treatments											
RMSE	0.0131	0.0089	0.0066	0.0087	0.0083	0.0067	0.0059	0.0134	0.0117	0.0134	0.0118
MAB	0.0106	0.0072	0.0053	0.0070	0.0067	0.0037	0.0035	0.0111	0.0096	0.0111	0.0096
MedAB	0.0101	0.0055	0.0045	0.0052	0.0055	0.0016	0.0017	0.0103	0.0083	0.0107	0.0087
Mean of covariates - One of the pre-treatments											
RMSE	0.0085	0.0089	0.0071	0.0087	0.0096	0.0067	0.0043	0.0134	0.0075	0.0134	0.0075
MAB	0.0072	0.0072	0.0059	0.0070	0.0089	0.0037	0.0035	0.0111	0.0060	0.0111	0.0061
MedAB	0.0077	0.0055	0.0065	0.0052	0.0090	0.0016	0.0028	0.0103	0.0051	0.0107	0.0051
Last covariates - One of the pre-treatments											
RMSE	0.0089	0.0089	0.0057	0.0087	0.0059	0.0067	0.0040	0.0134	0.0095	0.0134	0.0095
MAB	0.0076	0.0072	0.0049	0.0070	0.0045	0.0037	0.0028	0.0111	0.0076	0.0111	0.0077
MedAB	0.0064	0.0055	0.0038	0.0052	0.0036	0.0016	0.0020	0.0103	0.0050	0.0107	0.0053

Remark 3. In the Supplementary Online Appendix, we provide additional robustness checks with a restricted set of donor countries. Overall, we find that our conclusions remain unaffected.

5. Monte Carlo Simulations

In this section, we supplement the in-sample placebo analysis with a dedicated empirical Monte Carlo study. For the data generating process (DGP), we consider the factor model used by Abadie et al. (2010) and Kaul et al. (2021), given by:

$$y_{j,t}^0 = \delta_t + \boldsymbol{\theta}_t' \boldsymbol{c}_j + \gamma_t \mu_j + \varepsilon_{j,t}. \tag{23}$$

Here $y_{j,t}^0$ denotes the outcome in the absence of treatment, δ_t is the common time effects component, $\boldsymbol{\theta}_t$ is a vector of P-dimensional time-varying coefficients, one for each covariate given by the vector $\boldsymbol{c}_j$, $\gamma_t \mu_j$ is a typical interactive unobserved fixed-effect term, and $\varepsilon_{j,t}$ is the idiosyncratic error term with zero mean and variance σ^2, where we consider $\sigma \in \{0.25, 1\}$.

The values of δ_t and $\boldsymbol{\theta}_t$ are calibrated based on our empirical setting, see e.g. Ferman and Pinto (2021), Kaul et al. (2021), Arkhangelsky et al. (2021), who also consider real GDP per capita as the outcome of interest in their empirical Monte Carlo studies. In particular, we take the values of $\boldsymbol{c}_j$ as the time-averages of the observed covariates over the period 2010:Q2 and 2016:Q1, which is also the time span of the placebo study in the empirical section. The common factor γ_t is generated deterministically as $\gamma_t = t\gamma/T_0$, [9] for $\gamma \in \{0, 0.25, 1, 1.5\}$. As for the factor loadings, for all we set for $j = 1, \ldots, J+1$ we set $\mu_j \sim U[0, 1]$. Note that the larger the value of σ and/or γ, the lower the relative importance

[9] In the preliminary version of the paper we also considered a setup with simple linear trend $\gamma_t = t$, but as the results are qualitatively similar we only report the case of "bounded" common factors considered here.

of observed covariates.

The number of Monte Carlo replications is set to $M = 1000$ for all specifications.

5.1. RMSE Decomposition Details

In what follows, we adapt the approach of Kaul et al. (2021) and decompose the SC estimation error into three components: (i) error originating from covariates; (ii) error originating from the common factor component; (iii) from the presence of the idiosyncratic error term. In particular, observe that:

$$\widehat{\tau}_{T_0}^{sc} = \underbrace{\boldsymbol{\theta}_{T_0}' \left(\boldsymbol{c}_1 - \sum_{j=2}^{J+1} \widehat{\omega}_j^{sc} \boldsymbol{c}_j \right)}_{(i) covariates} + \underbrace{\gamma_{T_0} \left(\mu_1 - \sum_{j=2}^{J+1} \widehat{\omega}_j^{sc} \mu_j \right)}_{(ii) common factor} + \underbrace{\varepsilon_{1,T_0} - \sum_{j=2}^{J+1} \widehat{\omega}_j^{sc} \varepsilon_{j,T_0}}_{(iii) idiosyncratic}. \tag{24}$$

Evidently, similar decomposition can be also considered for the DSC estimator:

$$\begin{aligned} \widehat{\tau}_{T_0}^{dsc} = & \left(\boldsymbol{\theta}_{T_0} - \frac{1}{T_0 - 1} \sum_{t=1}^{T_0 - 1} \boldsymbol{\theta}_t \right)' \left(\boldsymbol{c}_1 - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} \boldsymbol{c}_j \right) \\ & + \left(\boldsymbol{\gamma}_{T_0} - \frac{1}{T_0 - 1} \sum_{t=1}^{T_0 - 1} \boldsymbol{\gamma}_t \right) \left(\mu_1 - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} \mu_j \right) \\ & + \varepsilon_{1,T_0} - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} \varepsilon_{j,T_0} - \frac{1}{T_0 - 1} \sum_{t=1}^{T_0 - 1} \left(\varepsilon_{1,t} - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} \varepsilon_{j,t} \right). \end{aligned} \tag{25}$$

As well as the SDID estimator:

$$\begin{aligned} \widehat{\tau}_{T_0}^{sdid} = & \left(\boldsymbol{\theta}_{T_0} - \sum_{t=1}^{T_0 - 1} \widehat{\lambda}_t^{sdid} \boldsymbol{\theta}_t \right)' \left(\boldsymbol{c}_1 - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} \boldsymbol{c}_j \right) \\ & + \left(\gamma_{T_0} - \sum_{t=1}^{T_0 - 1} \widehat{\lambda}_t^{sdid} \boldsymbol{\gamma}_t \right) \left(\mu_1 - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} \mu_j \right) \\ & + \varepsilon_{1,T_0} - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} \varepsilon_{j,T_0} - \sum_{t=1}^{T_0 - 1} \widehat{\lambda}_t^{sdid} \left(\varepsilon_{1,t} - \sum_{j=2}^{J+1} \widehat{\omega}_j^{dsc} \varepsilon_{j,t} \right). \end{aligned} \tag{26}$$

5.2. The Results

In what follows, we discuss the Monte Carlo results as summarized by means of the RMSE in Table 6. We also report contributions of the idiosyncratic, covariates, and the common factors parts for the total RMSE in Tables 7-9.

Table 6 shows that overall, when $\sigma = 1$, the SC(B) taking into account all covariates, the SC without covariates and the DSC also without covariates perform the best. However, while those methods are quite sensitive to the inclusion of covariates (in particular, the SC and the DSC methods perform worse when covariates are included), the results obtained with SDID are less sensitive.

For $\sigma = 0.25$, the performance of all methods is more comparable, and the results are less sensitive to covariates. Moreover, it is noticeable that as γ increases, the SC(B) and the SC methods perform worse, while the DSC and the SDID methods are less sensitive. This reinforces the idea that the latter estimators provide a better match in terms of the unobserved common factors.

Thus, once σ decreases, i.e. the relative importance of the idiosyncratic component is lower, the performance of the SDID methods seems to approach that of SC. This observation is supported by the RMSE decomposition results in Tables 7-9. From those tables, it is clear that for these designs, the noise associated with $\varepsilon_{j,t}$ dominates all other components, especially for $\sigma = 1$. As such, additional noise from (associated with the estimation uncertainty of the time-weights) $\varepsilon_{j,t}$ for the SDID outweighs the relative benefits of SDID in terms of the lower RMSE associated with the other two components. However, it is clear that for $\sigma = 0.25$, the additional noise from the idiosyncratic component in the SDID is more comparable to that of other methods. As a result, the remaining components of the RMSE that display a better performance for this estimator are not dominated.

In particular, for the part of the error term stemming from matching on covariates, when $\sigma = 1$, the DSC and the SDID, taking into account one covariate perform the best. We note that both the SC(B) and the SC methods improve if only one pre-treatment period is considered, while for the DSC and the SDID, the effect is moderate. When $\sigma = 0.25$, the inclusion of covariates is less important for all methods. Moreover, the SDID method provides the lowest RMSE in this case, especially for larger γ.

Results in Table 8 confirm our prior expectations that DSC and SDID methods improve upon the SC methods in terms of the error associated with the common factor components. In some cases, the improvement is as large as 50% in the (component) RMSE terms. When $\sigma = 1$, it is clear that all estimators perform worse as γ increases. However, as expected, the SDID shows a smaller impairment in terms of the RMSE (in comparison to other levels of σ). Surprisingly, for $\sigma = 0.25$, the performance of SDID does not deteriorate for larger values of γ, in contrast to all other methods.

Turning our attention to the question of the relative benefits of matching on covariates, we first consider the results in Table 6. We do not document benefits of matching on covariates, supporting our in-sample placebo results, if the total RMSE values are analyzed. However, Table 7 shows that for the SC, DSC, and SDID, in general, matching on one covariate leads to the smallest RMSE for the part of the error stemming from matching on covariates (even if the improvement is reduced for the SDID and the DSC opposed to the SC and SC(B) methods).

Therefore, the superior overall performance of the methods without covariates (in Table 6) stems primarily from the lower RMSE associated with the common factor component, especially for higher values of σ (Table 8). This conclusion, is in line with one of the conjectures by Kaul et al. (2021) regarding the same observed phenomenon.

Finally, we observe that for the results with covariates, inclusion of half pre-treatment periods is a preferred empirical strategy as it results in a lower total RMSE. This is also in line with our conclusions from the empirical in-sample placebo analysis.

Table 6: Monte Carlo Results: Total RMSE.

σ	γ	**SC(B) all**	**SC(B) half**	**SC(B) one**	**SC**	**SC cov. half**	**SC cov. one**	**DSC**	**DSC cov. half**	**DSC cov. one**	**SDID**	**SDID cov. half**	**SDID cov. one**
1	0	1.0645	1.1101	1.1909	1.068	1.0907	1.1375	1.074	1.0911	1.1505	1.1596	1.1675	1.1832
1	0.25	1.1213	1.1740	1.2135	1.1201	1.1405	1.1790	1.136	1.1637	1.1960	1.2304	1.2510	1.2588
1	1	1.1553	1.1817	1.2479	1.1528	1.1641	1.2250	1.1511	1.1687	1.2134	1.2193	1.2303	1.2397
1	1.5	1.1567	1.2039	1.2976	1.1552	1.1810	1.2292	1.1523	1.1733	1.1976	1.2600	1.2769	1.2586
0.25	0	0.068	0.0696	0.0772	0.0693	0.0694	0.0716	0.0686	0.0698	0.073	0.0748	0.0756	0.0771
0.25	0.25	0.0791	0.0782	0.0963	0.0779	0.0797	0.0828	0.0770	0.0784	0.0814	0.0821	0.0833	0.0833
0.25	1	0.0894	0.0871	0.1138	0.0870	0.0875	0.0936	0.0785	0.0792	0.0870	0.0819	0.0822	0.0851
0.25	1.5	0.0931	0.0903	0.1160	0.0894	0.0902	0.0975	0.0817	0.0831	0.0902	0.0866	0.0875	0.0901

Table 7: Monte Carlo Results: RMSE component due to covariates.

σ	γ	**SC(B) all**	**SC(B) half**	**SC(B) one**	**SC**	**SC cov. half**	**SC cov. one**	**DSC**	**DSC cov. half**	**DSC cov. one**	**SDID**	**SDID cov. half**	**SDID cov. one**
1	0	0.129	0.1347	0.0628	0.1407	0.1210	0.0567	0.0260	0.0228	0.0104	0.0271	0.0241	0.0106
1	0.25	0.1298	0.1381	0.0626	0.1407	0.1228	0.0567	0.0247	0.0225	0.0122	0.0246	0.0224	0.0123
1	1	0.1358	0.1357	0.0630	0.1482	0.1274	0.0536	0.0250	0.0222	0.0088	0.0247	0.0214	0.0087
1	1.5	0.1374	0.1433	0.0681	0.1493	0.1302	0.0596	0.0249	0.0226	0.0126	0.0240	0.0212	0.0120
0.25	0	0.0159	0.0179	0.0331	0.0186	0.0186	0.0171	0.0167	0.0160	0.0105	0.0081	0.0079	0.0058
0.25	0.25	0.0377	0.0448	0.0369	0.0440	0.0448	0.0428	0.0214	0.0201	0.0136	0.0084	0.0081	0.0060
0.25	1	0.0457	0.0614	0.0501	0.0530	0.0542	0.0525	0.0237	0.0201	0.0122	0.0039	0.0040	0.0034
0.25	1.5	0.0509	0.0672	0.0539	0.0577	0.0586	0.0570	0.0275	0.0230	0.0129	0.0034	0.0037	0.0032

Table 8: Monte Carlo Results: RMSE component due to common factors.

σ	γ	**SC(B) all**	**SC(B) half**	**SC(B) one**	**SC**	**SC cov. half**	**SC cov. one**	**DSC**	**DSC cov. half**	**DSC cov. one**	**SDID**	**SDID cov. half**	**SDID cov. one**
1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0.25	0.0787	0.0789	0.0854	0.0787	0.0796	0.0813	0.0398	0.0404	0.041	0.0406	0.0414	0.0419
1	1	0.2789	0.2741	0.3251	0.2786	0.2782	0.3053	0.1526	0.1553	0.1604	0.1507	0.1528	0.1605
1	1.5	0.3826	0.3761	0.4863	0.3823	0.3811	0.4398	0.2287	0.2337	0.2509	0.2169	0.2215	0.2351
0.25	0	0	0	0	0	0	0	0	0	0	0	0	0
0.25	0.25	0.0633	0.0650	0.0658	0.0657	0.0666	0.0700	0.0313	0.0314	0.0346	0.0146	0.0146	0.0163
0.25	1	0.0843	0.0911	0.0981	0.0879	0.0884	0.0961	0.0409	0.0388	0.0469	0.0095	0.009	0.0108
0.25	1.5	0.0924	0.1006	0.1057	0.0941	0.0944	0.1001	0.0437	0.0407	0.0493	0.0077	0.0072	0.0087

Table 9: Monte Carlo Results: RMSE component due to $\varepsilon_{j,t}$.

σ	γ	**SC(B) all**	**SC(B) half**	**SC(B) one**	**SC**	**SC cov. half**	**SC cov. one**	**DSC**	**DSC cov. half**	**DSC cov. one**	**SDID**	**SDID cov. half**	**SDID cov. one**
1	0	1.0536	1.0948	1.1837	1.0553	1.0782	1.1329	1.0735	1.0908	1.1502	1.1594	1.1674	1.1831
1	0.25	1.1141	1.1665	1.2116	1.1119	1.1349	1.1753	1.1351	1.1632	1.1952	1.2292	1.2498	1.2579
1	1	1.1156	1.1470	1.2011	1.1117	1.1254	1.1741	1.1410	1.1579	1.2001	1.2103	1.2210	1.2267
1	1.5	1.0913	1.1494	1.2107	1.0898	1.1188	1.1515	1.1306	1.1508	1.1732	1.2377	1.2541	1.2365
0.25	0	0.0664	0.0683	0.0711	0.0668	0.0673	0.0702	0.0673	0.0679	0.0721	0.0734	0.0741	0.0764
0.25	0.25	0.0716	0.0716	0.0731	0.0711	0.0727	0.0742	0.0718	0.0733	0.075	0.0797	0.0809	0.0807
0.25	1	0.0721	0.0729	0.0753	0.0715	0.0728	0.0739	0.0711	0.0724	0.0758	0.0797	0.0803	0.0827
0.25	1.5	0.0758	0.0757	0.0771	0.0743	0.0755	0.0765	0.075	0.0766	0.0778	0.0853	0.0863	0.0878

6. Concluding Remarks

In this paper, we extensively review the recently suggested extensions to the popular Synthetic Control (SC) method. In particular, we consider the Demeaned SC (DSC) and the Synthetic Difference-in-differences (SDID) approaches. We argue that both the DSC and SDID methods have certain desirable finite-sample bias reduction properties. In particular, the SDID method is targeted at minimizing the interpolation bias of the treatment effect.

In the empirical section, we re-investigate the effects of the Brexit referendum on UK GDP. Our results (overall) indicate a stronger effect of Brexit than the one described by Born et al. (2019). Contrary to the original study of Born et al. (2019), we do not advocate the inclusion of covariates if one is solely interested in the GDP series. In particular, the inclusion of covariates has a detrimental effect on the precision of counterfactual estimates as measured via the in-sample placebo analysis. We find that our theoretical predictions in terms of the interpolation bias translate into the superior properties of the SDID approach. Moreover, we also show how one can easily generate a *Zoo* of different specifications and methods in a fairly simple setting we consider.

Our empirical Monte Carlo study confirms the superiority of the DSC approach over the standard SC estimator. On the other hand, we found that the relative benefits of the SDID over the DSC are marginal at best; as for the trend specification considered in this paper, the benefits of the SDID do not outweigh the costs of estimating an additional large dimensional vector of time weights. Similarly to the in-sample placebo analysis, we do not document any finite sample benefits of using covariates for estimating unit-specific weights

Finally, as neither the placebo nor the dedicated simulation-based studies are perfect selection tools for the best methods for an empirical data at hand, we follow Ferman et al. (2020) and Advani et al. (2019) and recommend that empirical researchers consider multiple specification and report multiple results (as it is done in this paper).

References

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

Abadie, A., A. Diamond, and J. 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, 493–505.

——— (2015): "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59, 495–510.

Abadie, A. and J. Gardeazabal (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93, 113–132.

Abadie, A. and J. LHour (2021): "A penalized synthetic control estimator for disaggregated data," *Journal of the American Statistical Association*, 116, 1817–1834.

Advani, A., T. Kitagawa, and T. Sloczynski (2019): "Mostly Harmless Simulations? Using Monte Carlo Studies for Estimator Selection," *Journal of Applied Econometrics*, 34, 893–910.

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2021): "Synthetic Difference-in-Differences," *American Economic Review*, 111, 4088–4118.

Bai, J. (2009): "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

Ben-Michael, E., A. Feller, and J. Rothstein (2021): "The Augmented Synthetic Control Method," *Journal of the American Statistical Association*, 116, 1789–1803.

Beyhum, J. and E. Gautier (2019): "Square-root Nuclear Norm Penalized Estimator for Panel Data Models with Approximately Low-rank Unobserved Heterogeneity," ArXiv.

Bloom, N., P. Bunn, S. Chen, P. Mizen, P. Smietanka, and G. Thwaites (2019): "The Impact of Brexit on UK Firms," Tech. rep., National Bureau of Economic Research.

Born, B., G. Müller, M. Schularick, and P. Sedláček (2019): "The Costs of Economic Nationalism: Evidence from the Brexit Experiment," *The Economic Journal*, 129, 2722–2744.

Botosaru, I. and B. Ferman (2017): "On the Role of Covariates in the Synthetic Control Method," *Econometrics Journal*, 22, 117–130.

Chernozhukov, V., K. Wüthrich, and Y. Zhu (2021): "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls," *Journal of the American Statistical Association*, 116, 1849–1864.

Doudchenko, N. and G. W. Imbens (2016): "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis," Working Paper 22791, National Bureau of Economic Research.

Ferman, B. and C. Pinto (2021): "Synthetic Controls with Imperfect Pretreatment Fit," *Quantitative Economics*, 12, 1197–1221.

Ferman, B., C. Pinto, and V. Possebom (2020): "Cherry Picking with Synthetic Controls," *Journal of Policy Analysis and Management*, 39, 510–532.

Fernández-Val, I., H. Freeman, and M. Weidner (2021): "Low-Rank Approximations of Nonseparable Panel Models," *Econometrics Journal*, C40–C77.

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

Juodis, A. (2020): "This Shock is Different: Estimation and Inference in Misspecified Two-Way Fixed Effects Regressions," Mimeo.

Kaul, A., S. Klössner, G. Pfeifer, and M. Schieler (2021): "Standard Synthetic Control Methods: The Case Of Using All Preintervention Outcomes Together With Covariates," *Journal of Business & Economic Statistics*, (forthcoming).

Kellogg, M., M. Mogstad, G. Pouliot, and A. Torgovitsky (2021): "Combining Matching and Synthetic Control to Trade off Biases from Extrapolation and Interpolation," *Journal of the American Statistical Association*, 116, 1804–1816.

Moon, H. R. and M. Weidner (2015): "Linear Regression for Panel With Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83, 1543–1579.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.