PA

Dynamic Synthetic Controls: Accounting for Varying Speeds in Comparative Case Studies

Jian Cao[1] and Thomas Chadefaux[2]

[1]Department of Economics, Trinity College Dublin, Dublin, Ireland; [2]Department of Political Science, Trinity College Dublin, Dublin, Ireland

Corresponding author: Jian Cao; Email: caoj@tcd.ie

Abstract

Synthetic controls (SCs) are widely used to estimate the causal effect of a treatment. However, they do not account for the different speeds at which units respond to changes. Reactions may be inelastic or "sticky" and thus slower due to varying regulatory, institutional, or political environments. We show that these different reaction speeds can lead to biased estimates of causal effects. We therefore introduce a dynamic SC approach that accommodates varying speeds in time series, resulting in improved SC estimates. We apply our method to re-estimate the effects of terrorism on income (Abadie and Gardeazabal [2003, *American Economic Review* 93, 113–132]), tobacco laws on consumption (Abadie, Diamond, and Hainmueller [2010, *Journal of the American Statistical Association* 105, 493–505]), and German reunification on GDP (Abadie, Diamond, and Hainmueller [2015, *American Journal of Political Science* 59, 495–510]). We also assess the method's performance using Monte Carlo simulations. We find that it reduces errors in the estimates of true treatment effects by up to 70% compared to traditional SCs, improving our ability to make robust inferences. An open-source R package, dsc, is made available for easy implementation.

1. Introduction

Social scientists often seek to estimate the causal effects of treatments, such as events or interventions, on various outcomes of interest. Researchers may, for example, investigate the consequences of introducing minimum wage legislation on unemployment, the influence of terrorist attacks on economic growth, or the effect of electoral redistricting on legislative behavior. Typically, making such causal inferences requires the construction of counterfactuals to infer what would have occurred in the absence of the treatment.

An important tool for constructing counterfactuals is the synthetic control (SC) method, "arguably the most important innovation in the policy evaluation literature in the last 15 years" (Athey and Imbens 2017). This method is designed to provide an accurate representation of the hypothetical outcome for the treated unit without the treatment. It involves combining non-treated units, which, when given appropriate weights, closely resemble the pre-treatment unit. This technique extends the difference-in-difference approach, with the advantage of generating a close match to the unit of interest, even when no single control unit would be appropriate on its own.

However, the SC approach does not account for the potentially different speeds at which units react and adapt to changes. Changes in reactions to an event or a policy may be inelastic or "sticky" and therefore take longer in one unit than in another. Consider the example of public opinion shifts following major policy changes. While some segments of the public may quickly adjust their views in response to

new information or policy announcements, showing a "fast" response, while others might take longer to process the new information or for the effects of the policy change to become apparent in their daily lives. More generally, the speed at which y_t moves reflects how quickly it mirrors changes in its underlying latent process z_t. This speed of reaction to changes may also vary within one unit over time: a previously slow unit may now be fast, and vice versa. In addition, speed may also vary as a function of covariates.[1]

Variations in speed—how quickly units react to changes—have important consequences for our ability to make inferences. We may, for example, conclude that a treatment had a large effect on treated unit $\mathbf{y}_1$ compared to untreated units $\mathbf{y}_2$ and $\mathbf{y}_3$. However, the observed difference in reaction may simply be due to the fact that $\mathbf{y}_2$ and $\mathbf{y}_3$ lag behind $\mathbf{y}_1$—there is in fact no treatment effect, and all we observe is $\mathbf{y}_1$'s reaction to common shocks (not the treatment) ahead of $\mathbf{y}_2$ and $\mathbf{y}_3$. Conversely, we may underestimate the effect on $\mathbf{y}_1$ if $\mathbf{y}_2$ reacts faster to these common shocks. Ignoring these differences in speed creates further problems for SCs, as the donor[2] and the treated units may be out of synchronization, and what appears to be a good pre-treatment match may in fact be due to chance rather than a true alignment. Another problem is that the varying speeds may prevent us from finding a suitable SC at all, as there may be no linear combination of the donors that allows for a close match to the treated unit.

Different units—states, cities, and people—may adjust and react at varying speeds for many reasons. For instance, legal constraints in one state may slow down the speed at which economic decisions are made compared to other states. Öztekin and Flannery (2012), for example, find that the speed with which firms adjust their capital structure correlates with legal and financial norms. Additionally, the speed of convergence—the rate at which a country's per capita income approaches its steady state— has also been shown to vary by country and over time (Barro *et al.* 1991; Canarella and Pollard 2004; Rappaport 2000). Similarly, the effect of a policy may be slower to fully manifest itself in one instance than in another. Thus, Alesina, Cohen, and Roubini (1992) find that the speed at which the deficit is adjusted in election years may be slower than in other years. Moreover, the impact of a shock on, for example, commodity prices, may be almost instantaneous in some markets but drag on in others.

To demonstrate the significance of these varying speeds for our ability to make inferences, suppose that we aim to estimate the causal effect of a policy on an outcome, for example, the impact of cigarette taxes on consumption or the introduction of a minimum wage on employment. We have artificial data for three units (e.g., states) over time, as illustrated in Figure 1. During the observation period, one of the units ($\mathbf{y}_1$) receives a treatment, while the other two ($\mathbf{y}_2$ and $\mathbf{y}_3$) do not, allowing them to be used as potential donor units for counterfactual analysis. However, the researcher is not aware that the true effect of the treatment is zero.[3] In addition to the treatment, which applies exclusively to $\mathbf{y}_1$, we also assume that all three units are subject to universal shocks represented by $\mathbf{z}$. All units react to $\mathbf{z}$ at different speed. As shown in 1, $\mathbf{y}_2$ adjusts slowly when y decreases but recovers fast. In contrast, $\mathbf{y}_3$ decreases fast but recovers slowly. Unit $\mathbf{y}_1$ has a neutral speed.

To generate the synthetic unit, the standard SC approach constructs a weighted average of the two available donor units $\mathbf{y}_2$ and $\mathbf{y}_3$ using weights $\mathbf{w}^* = (w^*_{\mathbf{y}_2}, w^*_{\mathbf{y}_3})$. $\mathbf{w}^*$ is selected such that the resulting SC, which combines the non-treated units $\mathbf{Y}_{-1}$ as $\mathbf{Y}_{-1}\mathbf{w}^*$, closely matches the pre-treatment sequence of the treated unit $\mathbf{y}_1$.[4] Abadie and Gardeazabal (2003) demonstrate that this approach produces a plausible counterfactual for the treated unit.

In the example of Figure 1, however, selecting weights that generate a suitable counterfactual is not straightforward. Indeed, even in this simple case, there is no easily obtainable closed-form solution for finding $\mathbf{w}$. The standard linear approach results in the SC depicted by the blue curve. This curve poorly

---

[1] This concept of speed, detailed in Supplementary Appendix 1, is quantified here through the presence and order of nonzero lag coefficients. A higher order of significant lag coefficients, for example, indicates a slower response rate of the outcome variable to changes in the latent process.

[2] In SC, the untreated units are referred to as donors.

[3] This illustration is simplified and purposefully designed to emphasize the critical role and impact of response speed variability in SC accuracy. More realistic examples follow below.

[4] In other words, we choose $\mathbf{w}^*$ to satisfy $\mathbf{w}^* = \text{argmin}_{\mathbf{w}^*} \|\mathbf{y}_1 - \mathbf{Y}_{-1}\mathbf{w}^*\|_2$. For simplicity, we abstract away any predictors.
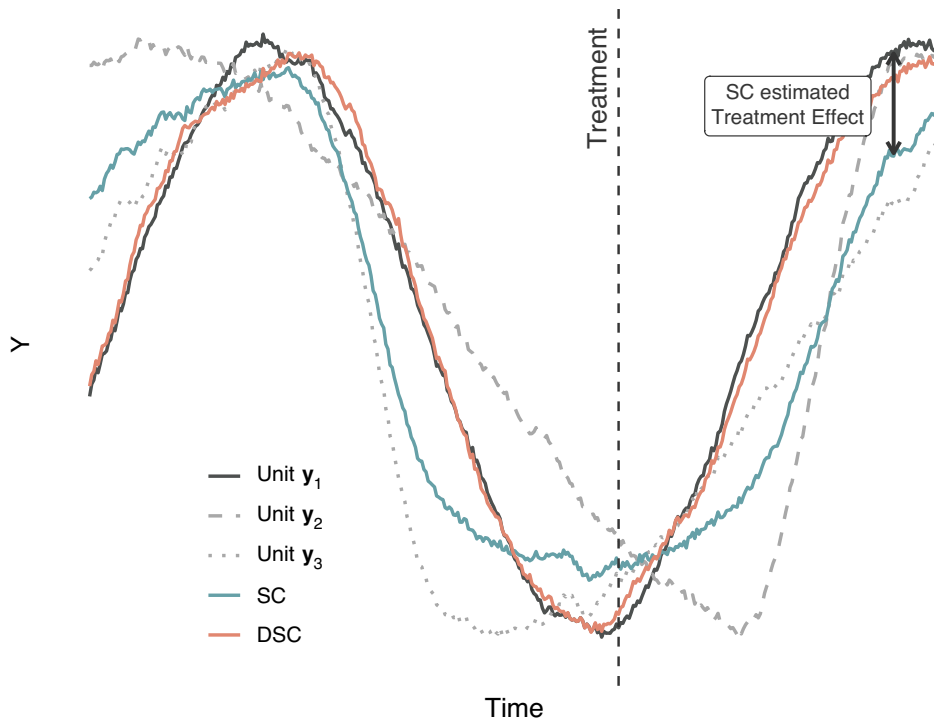
Figure 1. The challenge of varying speeds. Consider a researcher aiming to quantify the effect of a treatment on unit $\mathbf{y}_1$. Unbeknownst to them, no treatment effect actually exists. Employing conventional synthetic control (SC) methods with donor units $\mathbf{y}_2$ (slow) and $\mathbf{y}_3$ (fast), they obtain a post-treatment estimate (blue curve) that diverges markedly from the true outcome (black curve), leading to a significant bias in the estimated treatment effect. In contrast, dynamic synthetic controls, as elaborated below, produce a SC (red curve) that more closely approximates the truth.

matches both the pre- and post-treatment unit $\mathbf{y}_1$, and leads us to erroneously conclude that a treatment effect exists (the post-treatment difference between the blue and black curves) when, in reality, the true treatment effect is zero. Adding nonlinear terms or lags does not improve the situation much either (see Figure A1 in Supplementary Appendix 2, which incorporates multiple lags and polynomial terms).

The approach's challenge to generate a satisfactory approximation of the treated unit using untreated units stems from the challenge of accounting for varying speeds. In this case, $\mathbf{y}_2$, $\mathbf{y}_3$, and $\mathbf{y}_1$ all exhibit cycles that unfold at different rates. All three units display the same patterns, but prior to the treatment, $\mathbf{y}_2$'s cycle is longer (i.e., slower) than $\mathbf{y}_1$'s, and $\mathbf{y}_3$ is faster than $\mathbf{y}_1$. Additionally, the differences in speed fluctuate within each unit over time, so that a single speed adjustment parameter is insufficient. Lastly, variations in speed might depend on unobservable variables or endogenous regressors. Unless we can correctly account for the complex mechanisms and model the speed, the speed difference will cause errors akin to omitted variable bias in SC estimation.[5]

In this paper, we introduce a new method, dynamic synthetic control (DSC), which accounts for varying speeds within and across units. This approach operates by learning the differences in speed between the series during the pre-treatment period. We do this by calculating a dynamic time warping (DTW) path between them. The warping path subsequently provides a measure of speed differentials across units and over time. We then use this warping path to warp the post-treatment donor units to align their speed with that of the treated unit. As a result, we can assess the treatment effect while controlling for the inherent varying speeds. As an illustration of the method's capabilities, the red curve

---

[5]The nature of the omitted variable bias will be discussed in Section 2.

in Figure 1 shows the outcome of applying DSC to this simple case. We observe that warping enables us to more closely match the post-treatment unit compared to a standard SC. Importantly, the estimator only removes speed differences that originate from the pre-treatment period; it does not remove speed differences that are caused by the treatment. This, in turn, allows for a more precise inference of the treatment effect's magnitude.

The remainder of this paper demonstrates that this result extends beyond a simple example. In fact, we show empirically that DSCs serve as a more efficient estimator of the treatment effect compared to standard SCs. We substantiate these findings by replicating leading empirical work in this field and illustrating how the results can be improved using DSC. Additionally, we generalize our results through a Monte Carlo simulation.

2. Synthetic Controls and the Problem of Speed

Many of the questions of interest to social scientists revolve around estimating the effect of treatments such as an event or an intervention (e.g., Becker and Klößner 2016; Brady and McNulty 2011; Card and Krueger 2000; Chattopadhyay and Duflo 2004; Di Tella and Schargrodsky 2004). Difference-in-differences, for example, leverages similarities between a treated unit and an untreated one to deduce the treatment's effect. However, the method potentially suffers from biased control selection and imprecise case similarity—sometimes lacking a suitable comparison altogether.

The SC approach, pioneered by Abadie and Gardeazabal (2003) and refined in Abadie, Diamond, and Hainmueller (2010, 2015), aims to address such limitations by combining a basket of control cases to mimic the pre-treatment scenario for the treated unit. This enables a more robust counterfactual study of the treatment's impact (see Abadie 2021 for a review; a refresher of the method is in Supplementary Appendix 3).

However, in a wide class of situations, there is in fact no easy way to generate a good counterfactual using standard SC method. Suppose that we observe a time series $y_{1,t}$ exposed to a treatment at time T. We also observe a basket of time series $y_{j,t}, j \in (2,3,\cdots,J+1)$ to be used as donors. All time series are of length N. Assuming that the target time series $y_{1,t}$ depends not only on current $y_{j,t}$, but also on lags $y_{j,t-l}, l \in (1,2,\cdots,t-1)$, then the model of interest for the behavior of $\mathbf{y}_1$ over time is

$$y_{1,t} = \sum_{j=2}^{J+1} [w_j(y_{j,t} + \beta_{1,j,t}y_{j,t-1} + \beta_{2,j,t}y_{j,t-2} + \cdots + \beta_{t-1,j,t}y_{j,1})] + \epsilon_t, \tag{1}$$

where $\beta_{l,j,t}$ is a time-dependent coefficient for the lag term $y_{j,t-l}$, l refers to the order of the lag, j to the donor unit, and t to the time. w_j is a constant weight for donor $\mathbf{y}_j$ and its lags, and ϵ_t is the error term. We assume that all classical assumptions about ϵ_t apply. This model explores the speed problem by allowing a varying number of lags of $y_{j,t}$ to influence the target time series $y_{1,t}$. The higher order of lags that have nonzero coefficients, the "slower" $y_{1,t}$ is relative to $y_{j,t}$—that is, the more the past drags on.

Consider now the standard SC approach. As it only includes the current donor time series $y_{j,t}$ but not the lags $y_{j,t-l}$, the model used in estimating the weights $\mathbf{w}_{sc}$ is not the one of Equation (2.1), but rather

$$y_{1,t} = \sum_{j=2}^{J+2} (w_j y_{j,t}) + \eta_t = \mathbf{Y}_{-1,t}\mathbf{w}_{sc} + \eta_t,$$

and the new error term η_t is

$$\eta_t = \mathbf{Y}_{-1,L}\boldsymbol{\beta}_L\mathbf{w} + \epsilon_t,$$

where $\mathbf{Y}_{-1,L}$ are lag terms and $\boldsymbol{\beta}_L$ are lag coefficients. Clearly, ignoring the lagged effects of the donor time series results in omitted variable bias. This subsequently leads to a biased estimate of the treatment effect and inflated standard errors (see Supplementary Appendix 4 for a proof).

Various extensions of the SC method have been introduced to tackle the problems of poor pre-treatment fits, but none are suitable for the varying speed problem. For example, Ben-Michael, Feller, and Rothstein (2022) examine the SC method in scenarios where policies are implemented by different units at distinct times. Other than SC methods, Goodman-Bacon (2021) discusses how improved difference-in-difference estimators address the bias from time-varying treatment effects. While many newly developed methods (Ben-Michael *et al.* 2022; Ferman, Pinto, and Possebom 2020; Goodman-Bacon 2021) strive to enhance causal inference outcomes when the data requirements outlined in Abadie and Cattaneo (2021) are not met, none specifically address the varying speed problem which involves estimating the *JNN* lag coefficients.[6]

In a related paper, Becker and Klößner (2018) have improved the accuracy of the SC method. They do so by accommodating varying time resolutions and introducing time-specific weights for donors and predictors. This has broadened the applicability of the method in analyzing complex economic phenomena with multiple dependent variables and predictors of differing temporal granularities. Our methodology complements Becker and Klößner (2018) by addressing the challenge of differential response speeds among units—an aspect of temporal variability not directly tackled by the focus on time resolution in the work by Becker and Klößner (2018). By employing a warping technique on the outcome time series y_j, we enhance the alignment of treated and control units, beyond observable characteristics or the temporal granularity of the data. This process effectively mitigates bias arising from the varying speeds at which units respond to latent processes or external shocks, ensuring the SC more accurately mirrors the counterfactual scenario. While the framework by Becker and Klößner (2018) handles varying time resolutions and introduces time-specific weights for donors and predictors, it does not explicitly address pre-treatment alignment of outcome series to manage differences in unit response speeds. Our warping approach specifically fills this gap.[7]

In this paper, we address the speed problem by relying on DTW, a non-parametric method derived from speech recognition (Vintsyuk 1968), which allows us to recover speed differences between sequences. The DTW algorithm is widely used to find the optimal alignment between two time series that may vary in speed. It obtains the alignment by warping the sequences in the time dimension, such that the Euclidean distance between the warped and the target sequences is minimized. Determining the best way to warp the series is equivalent to estimating the lag coefficient βs.

After estimating the lag coefficient βs, we combine the lag terms into a single time series

$$y_{j,t}^w = \sum_{l=t-N}^{t-1} \hat{\beta}_{l,j,t} y_{j,t-l},$$

which we refer to as a "warped" time series. Then when applying SC, we replace the original donor time series $\mathbf{y}_j$ with the warped time series $\mathbf{y}_j^w$ to mitigate the impact of the speed problem.[8] We provide details of DSC method below.

---

[6](J donor time series) $\times$ (N time periods) $\times$ ($N/2$ backward lags + $N/2$ forward lags). Please see Supplementary Appendix 6 for a discussion of forward lags.

[7]More specifically, Becker and Klößner (2018) employ a generalized model that allows multiple dependent variables and covariates at different time resolution, time aggregates, and lags. Importantly, it allows for time-specific weights for dependent variables and covariates so that more weight can be put on the data closer to the treatment time. This reduces the errors caused in estimating the weights of covariates (inner optimization) and thus yields better weights of donors (outer optimization) and generates more accurate SC. By reducing the dimensionality of the optimization space by identifying "sunny"/"shady" donors, Becker and Klößner's (2018) method is potentially able to estimate the time-specific weights for *all* possible lags of the donors ($\mathbf{Y}_{-1,L}$). However, for most applications, it would not be computationally efficient to do so, as each donor comes with N lagged units. In addition, estimating all JNN coefficients is typically not feasible given limited data.

[8]DSC uses a non-parametric method, DTW, to estimate the time-specific coefficients for the lagged donor units the covariates. By warping the donor units and the covariates, it reduces errors caused by speed difference in both inner and outer optimizations. Since the method finishes the correction before the SC starts, it can be used with SC extensions such as MSCMT (Becker and Klößner 2018), and further reduces bias caused by poor-performance optimizers. Together, these methods generate better SCs.

3. Accounting for Speed: Dynamic Synthetic Controls

The speed issue arises from two possible sources: those caused by the treatment itself and those resulting from inherent speed differences (see Figure 1 and the previous section). Here, we introduce a method designed to mitigate the inherent speed differences between treated and untreated units while maintaining the effects induced by the treatment—whether in level or speed. As a result, we can attribute any post-treatment differences between the two groups exclusively to the treatment. This approach eliminates the influence of pre-existing speed disparities in the time series.

The DSC method introduced below addresses the issue of varying speeds across time and units. It begins by warping the untreated units to minimize the speed differences between them and the treated unit. The algorithm extracts information on speed differences from the pre-treatment period, and adjusts the speed of the entire donor time series to align with the treated time series.[9] This indirect warping approach allows the remaining difference between the treated and donor units to be attributed exclusively to treatment differences rather than inherent speed differences. After adjusting the series' speeds, the standard SC method is applied.

The warping process comprises three steps (see Figure 2 and Algorithm A1 in Supplementary Appendix 7). First, we match the pre-treatment parts of the target (black) and the donor (red) time series; second, we match the post-treatment and the pre-treatment parts of the donor time series; finally, we combine the time alignments obtained in the previous two steps and create a warped donor time series (blue) that minimizes the inherent speed difference between itself and the target time series (black). We now discuss each step in detail.

Step 1. Matching Pre-Treatment Time Series

The DSC algorithm first estimates the speed relationship between the pre-treatment segments of the target time series $\mathbf{y}_1$ and the donor time series $\mathbf{y}_j$. This is achieved by employing the DTW method to align the pre-treatment portion of $\mathbf{y}_1$ with $\mathbf{y}_j$ and storing the warping path matrix.

The warping path matrix stores the results of DTW alignments and is formally defined as follows:

$$\mathbf{P}_j = [p_{j,v,t}], \quad v,t \in [1,N],$$

$$p_{j,v,t} = \begin{cases} 1, & \text{if } y_{j,v} \text{ matches } y_{1,t}, \\ 0, & \text{otherwise.} \end{cases}$$

As illustrated in Figure 3, the points of $\mathbf{y}_1$ are matched to the points of $\mathbf{y}_j$ by DTW in a manner that minimizes the total distance between the matched points.[10] The warping path matrix $\mathbf{P}_j$ captures the speed difference between $\mathbf{y}_1$ and $\mathbf{y}_j$.

We first match the pre-treatment target time series $\mathbf{y}_{1,1:T}$ to the donor time series $\mathbf{y}_j$ (see the first part of Figure 2). The point $y_{1,T}$ is matched to $y_{j,C}$.[11] Let $\mathbf{y}_{j,pre} = \mathbf{y}_{j,1:C}$ denote the pre-treatment part of $\mathbf{y}_j$ and $\mathbf{y}_{j,post} = \mathbf{y}_{j,C:N}$ the post-treatment portion. Similarly, we define $\mathbf{y}_{1,pre} = \mathbf{y}_{1,1:T}$ and $\mathbf{y}_{1,post} = \mathbf{y}_{1,T:N}$. Using DTW, we obtain a warping path from $\mathbf{y}_{j,pre} \to \mathbf{y}_{1,pre}$, which is stored in a $C \times T$ matrix $\mathbf{P}_{j,pre}$.

The warping path $\mathbf{P}_{j,pre}$ is used in step three to warp the time axis of $\mathbf{y}_{j,pre}$ with the goal of minimizing the speed difference between $\mathbf{y}_{j,pre}$ and $\mathbf{y}_{1,pre}$. Additionally, it is employed along with the warping path derived from step two to adjust the speed of $\mathbf{y}_{j,post}$.

Step 2. Matching Pre- and Post-Treatment Donor Time Series

Learning the warping path between the post-treatment segments $\mathbf{y}_{j,post}$ and $\mathbf{y}_{1,post}$ is more challenging. We cannot directly align the two sequences, because their differences are due not only to their different

---

[9]Importantly, we avoid inferring speed differences from the post-treatment period, as it could conflate the effects of inherent speed differences and treatment, leading to biased estimates.

[10]Formally, $\mathbf{P}_j = \operatorname{argmin}_{\mathbf{P}_j} \left(\sum_{t=1}^{N} \sum_{v=1}^{N} \frac{p_{j,v,t}|y_{1,t} - y_{j,v}|}{\sum_{v=1}^{N} p_{j,v,t}} \right)$.

[11]To ensure optimal matching, we do not impose an end rule. Consequently, the matched time series $\mathbf{y}_{1,1:T}$ and $\mathbf{y}_{j,1:C}$ may have different lengths.
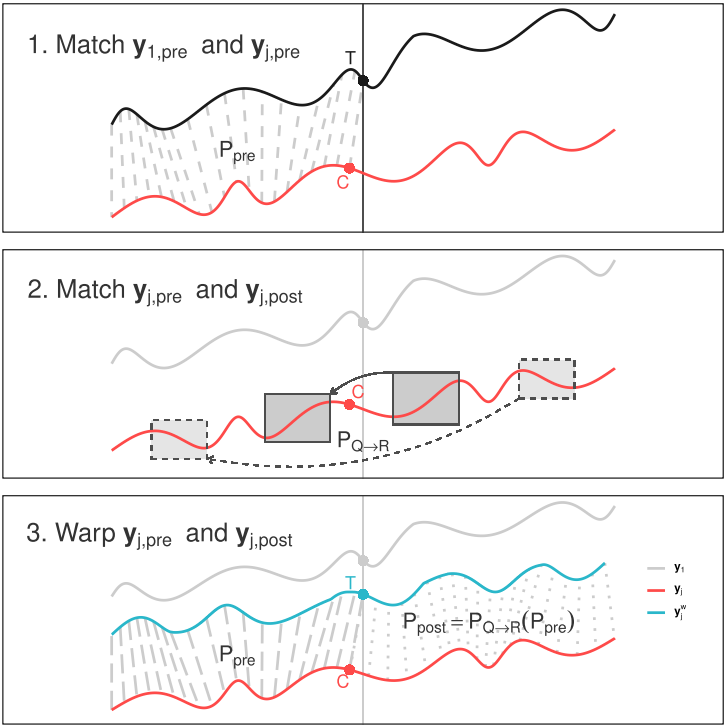
Figure 2. Dynamic synthetic control (DSC). The warping process of the DSC algorithm operates in three key steps. First, it matches the pre-treatment segments of $\mathbf{y}_j$ and $\mathbf{y}_1$ to derive the warping path $\mathbf{P}_{pre}$. Second, it aligns the pre- and post-treatment segments of $\mathbf{y}_j$, yielding $\mathbf{P}_{Q \to R}$. Finally, $\mathbf{y}_j$ is warped using both $\mathbf{P}_{pre}$ and $\mathbf{P}_{post}$ to produce the time-warped series $\mathbf{y}_j^w$.
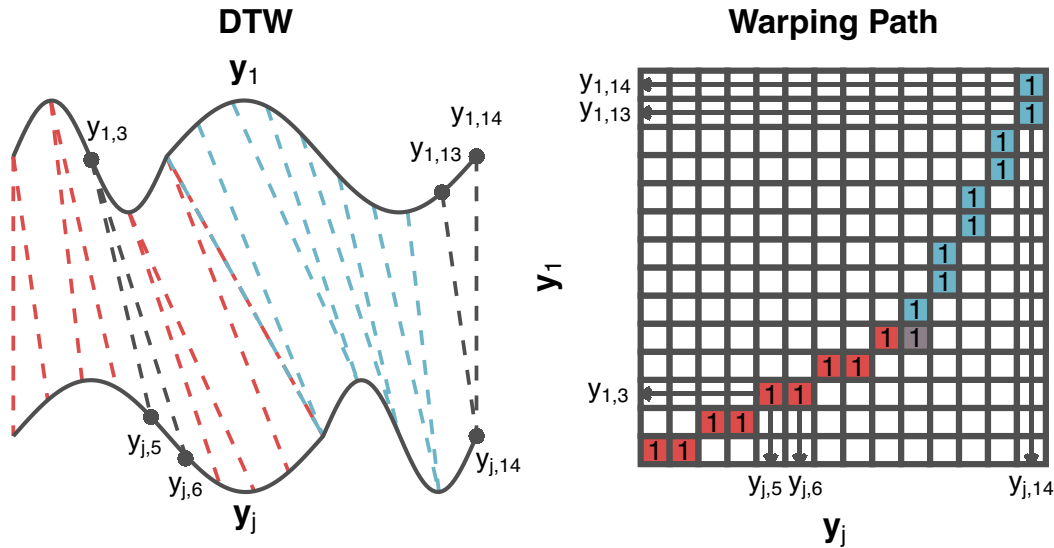
Figure 3. Warping path. The left figure shows data points matched in DTW, connected by dashed lines. The right figure displays the corresponding warping path matrix, where only matched pairs (ones) are shown. The time series $\mathbf{y}_j$ initially progresses at a rate $2\times$ slower (indicated in red) than $\mathbf{y}_1$ but later becomes $2\times$ faster (in blue) than $\mathbf{y}_1$.

speeds but also to the treatment effect. Aligning them would artificially remove that treatment effect. So the post-treatment warping path $\mathbf{P}_{j,post}$, that is, the inherent differences in speed, must be learned from the pre-treatment path $\mathbf{P}_{j,pre}$.

To extract similar short-term sequences in $\mathbf{y}_{j,post}$ and $\mathbf{y}_{j,pre}$ and therefore infer the warping path $\mathbf{P}_{j,post}$ from $\mathbf{P}_{j,pre}$, we use a double-sliding window approach.[12] In essence, the double-sliding window serves as a dynamic "lens"—the first window scans through the post-treatment time series to identify segments requiring alignment, while the second window sifts through the pre-treatment time series to find the most similar segments, which facilitates optimal matching and warping of patterns between the two time series. As illustrated in the second part of Figure 2, we slide a target window Q_u in $\mathbf{y}_{j,post}$ and a reference window R_i in $\mathbf{y}_{j,pre}$. For each short-term sequence Q_u in $\mathbf{y}_{j,post}$, we find the best-matching sequence R^* from $\mathbf{y}_{j,pre}$, and record the warping path $Q_u \to R^*$ as $\mathbf{P}_{j,Q_u \to R^*}$. We also extract the corresponding pre-treatment warping path $\mathbf{P}_{j,R^*} : R^* \to \mathbf{y}_1$ from $\mathbf{P}_{j,pre}$. Next, $\mathbf{P}_{j,R^*}$ is adjusted based on $\mathbf{P}_{j,Q_u \to R^*}$ to account for the differences between Q_u and R^*. The resulting adjusted warping path is $\mathbf{P}_{j,Q_u}$. Once the sliding target window Q_u reaches the right boundary $y_{j,N}$, we merge all the resulting warping paths $\mathbf{P}_{j,Q_u}$ to obtain the final warping path $\mathbf{P}_{j,post}$ for the post treatment donor time series.[13]

Step 3. Warping Donor Time Series

In the third step, the DSC algorithm uses the previously estimated warping path $\mathbf{P}_{j,pre}$ to warp the pre-treatment donor series $\mathbf{y}_{j,pre}$. Similarly, it uses $\mathbf{P}_{j,post}$ from the second step to warp $\mathbf{y}_{j,post}$.[14] The resulting time series are combined into a single warped donor time series $\mathbf{y}_j^w$:

$$\mathbf{y}_j^w = [\mathbf{y}_{j,pre}^w, \mathbf{y}_{j,post}^w]$$
$$= [\mathbf{P}_{j,pre}(\mathbf{y}_{j,pre}), \mathbf{P}_{j,post}(\mathbf{y}_{j,post})].$$

After being warped, the inherent speed differences between $\mathbf{y}_j^w$ and $\mathbf{y}_1$ are minimized, while the differences caused by the intervention are unchanged. Finally, using the warped donor time series $\mathbf{y}_j^w, j \in (2,3,\cdots,J+1)$, we apply the SC method to construct a counterfactual of $\mathbf{y}_1$ to minimize the impact of the speed problem.[15]

All methods discussed in this paper have been implemented in an accompanying R package.[16]

4. Evaluating the Method

We showcase the advantages of our proposed method using two approaches. First, we create synthetic data using a Monte Carlo simulation. This lets us design and know the treatment effect, in contrast to real-world data where the true treatment effect is never known. This allows us to show that, across a

---

[12] Details on the double-sliding window approach can be found in Supplementary Appendix 8.

[13] To ensure that the resulting warping path $\mathbf{P}_{j,post}$ has the best accuracy, a threshold θ is applied to the DTW distances to filter out any less desired matches. If for a target window Q_u, we cannot find any matches that meet the standard θ, the target window will be unchanged, that is, no warping will be given.

[14] To warp the time series using a warping path matrix $\mathbf{P}_j$, we first obtain estimates of lag coefficients $\hat{\beta}_{v-t,j,t} = \frac{p_{j,v,t}}{\sum_{v=1}^{N} p_{j,v,t}}$, then combine the lag terms and obtain the warped time series $y_{j,t}^w = \sum_{l=t-N}^{t-1} \hat{\beta}_{l,j,t} y_{j,t-l}$.

[15] The DSC method, in line with traditional SC and difference-in-differences approaches, operates under the assumption that treatment effects can be distilled into an additive-separable scalar. This simplification facilitates the estimation of these effects by comparing the outcomes of the treated unit to those of an SC constructed from donor units. However, we acknowledge that this approach implicitly assumes the treatment's exogeneity and does not systematically account for the endogeneity of treatment or the potential complexity of treatment effects that may not be adequately captured by a single scalar measure. Furthermore, the DSC method assumes negligible spillover effects from the treatment to the donor units. This assumption is critical, as significant spillover effects would likely bias the SC methods, especially if the donor units that are most similar to the treated unit are the ones affected. These spillover effects tend to lead to an underestimation of the true treatment effect—and to inflated standard errors—because they imply that the control units are also impacted by the treatment. These considerations reflect broader challenges within causal inference methodologies. Other threats to identification in our model are discussed in Supplementary Appendix 5.

[16] The R package can be found in Supplementary Appendix 13.

wide range of parameters, our estimate of the treatment effect is more efficient than the one obtained using the standard SC approach.

Second, we apply the DSC method to data from three seminal articles on SCs: Abadie and Gardeazabal's (2003) data on the effect of terrorist attacks on GDP; Abadie *et al.*'s (2010) data on changes in tobacco consumption in California; and Abadie *et al.*'s (2015) study of the economic impact of the 1990 German reunification. In each case, we demonstrate that our estimates of the placebos are more efficient, such that the statistical test of the effect of the treatment has greater power.

In each case, we make two main arguments. First, we argue that our method generates a better counterfactual than the standard SC. Second, we demonstrate that our method enhances precision by reducing uncertainty in the treatment effect estimate. A key issue in SC is the possibility that observed effects occur by chance. Existing literature often shows that while SCs for untreated units (placebo units) generally yield less extreme results than the treated unit, the confidence intervals remain wide. Our DSC method effectively narrows these intervals, thereby improving estimator precision.

4.1. Monte Carlo simulation

We first use a Monte Carlo simulation to replicate the types of empirical challenges faced with real data. We generate hundreds of artificial panel datasets, each containing time series for ten units observed over 100 periods. In each sample of 10 series, one time series is designated as the "treated" unit, while the remaining nine serve as the donor pool to construct the SC (i.e., $J = 9$, $N = 100$).[17]

All 10 time series follow a common autoregressive integrated moving average (ARIMA) process but exhibit differing speeds. A time series' speed is either random or a function of the time series direction (increasing or decreasing). The idea behind it is to capture the possibility that speed may vary as a function of the direction of the underlying series. Economic crashes (i.e., a decreasing series), for example, may unfold faster than recoveries (Reinhart and Rogoff 2014). A dataset-specific parameter $\psi \in (0, 1)$ determines the extent to which this occurs. For instance, $\psi = 0$ implies that the speed will be entirely governed by a random normal noise. Conversely, if $\psi = 1$, the speed will completely depend on the direction of the time series.

For each dataset, we implement DSC and standard SC to construct counterfactuals for the treated unit. A key advantage of the Monte Carlo simulation, as opposed to real observational data, is that we know the true treatment effect and thus the true counterfactual. Consequently, we can assess how well counterfactuals formed using different methods approximate the post-treatment series. In accordance with common practice in the literature, we evaluate the performance of each method over the 10 periods following the treatment, that is, from $t = 61$ to $t = 70$.

To evaluate the quality of the SC generated by each method, we compute the 10-period post-treatment mean squared error (MSE) for the estimated treatment effects. Specifically, the MSE for dataset d is defined as: $\text{MSE}_d = \frac{1}{10} \sum_{t=61}^{70} (\widehat{\tau}_{t,d} - \tau_t)^2$, where $\widehat{\tau}_{t,d}$ denotes the estimated treatment effects using one of the two SC methods. A small MSE indicates that the estimated treatment effect closely fits the true effect imposed during the data generation process. For each simulation, we calculate the MSE for the standard SC approach—MSE_{SC}—and the MSE for our DSC approach—MSE_{DSC}. The log ratio of these two values yields a measure r representing the relative performance of each method, that is, $r = \log(\text{MSE}_{DSC}/\text{MSE}_{SC})$. Negative log ratios suggest superior performance of our method via lower MSE. We compute the log ratio r for each simulated dataset and conduct a t-test on the resulting 100 log ratios to assess whether the ratio is significantly different from 0 ($p < 0.01$).

The estimated treatment effects for all datasets are illustrated in Figure 4. We see that, in the 10 periods following the treatment, the average estimated treatment effects from the two methods are close to the true effects. However, the new method, DSC, produces a significantly smaller error area than the standard method. Specifically, the average logged MSE for the standard SC method ($\overline{\log(\text{MSE}_{SC})}$)

---

[17]Detailed data generation method is shown in Supplementary Appendix 10.
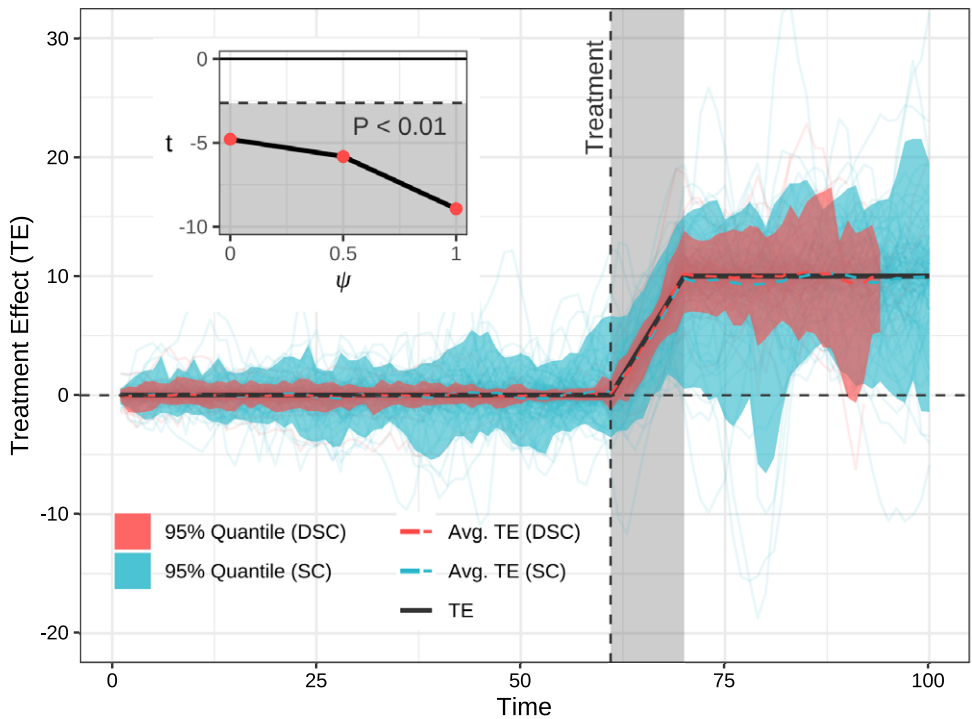
Figure 4. Results from the simulation study with 95% confidence intervals. The main graph showcases results drawn from Monte Carlo simulations where $\psi = 1$. The gray-shaded region delineates the period over which performance is estimated. Red and blue lines represent the distribution of estimated treatment effects for the dynamic synthetic control and synthetic control methods, respectively. The true treatment effect is in black. An inset in the top-left corner demonstrates that larger ψ values lead to improved performance—as evidenced by more negative t-values.

is 0.90, while the value for the DSC method ($\overline{log(\text{MSE}_{DSC})}$) is −0.50.[18] Moreover, our approach outperforms the standard method 77% of the time (Wilcox test < 0.001). These results suggest a strong expected benefit to using this method and are in line with our theory: Section 2 shows that the estimated treatment effect from SC is potentially biased and has a larger variance. In the Monte Carlo study, we do not directly observe bias because the biases around the true values tend to average out to zero. But it is clear in the Monte Carlo results that the SC estimator has a higher variance, and hence that the DSC estimator of the treatment effect is more efficient.

Finally, we also show that the results hold for all values of ψ—the parameter which controls the extent to which speed varies as a function of the shape of the series (see the subfigure of Figure 4), although larger ψs lead to more significance (more negative t value). When $\psi = 1$, the average log ratio of MSEs is $\bar{r} = -1.39$. This value indicates that, on average, the DSC method reduces the MSE of the estimated treatment effect by 75.19% when compared to the standard SC method ($1 - e^{-1.39} \approx 75.19\%$).

4.2. Re-Evaluating Empirical Findings

We now show that these results are not limited to artificial data but extend to real-world empirical data. Specifically, we apply our method to three seminal articles in the SC literature: Abadie and Gardeazabal's (2003) analysis of the economic costs of terrorism in the Basque country; Abadie *et al.*'s (2010) assessment of the effect of Proposition 99—a large-scale tobacco control program implemented

---

[18]We use the log of the MSEs because outliers can take on too much weight in non-pairwise mean comparisons.

in California in 1988; and Abadie *et al.*'s (2015) evaluation of the economic impact of the 1990 German reunification on West Germany.[19]

One challenge, of course, is that we do not know the true treatment effect for the main unit of interest. This effect must be estimated and, unlike in the Monte Carlo study, there is no way to argue that a particular result is more accurate or less biased than another. However, we do know that there is no treatment effect in the other (non-treated) units. Thus, we can assess the performance of the estimators by comparing how well their respective synthetic controls approximate the true post-treatment period for non-treated units. For example, we cannot know the true effect of the German reunification on West Germany's GDP, but we do know that there was no treatment in Canada, and can thus attempt to match Canada's post-1990 trajectory—as well as each of the untreated units.[20] Due to the limited set of untreated units, however, this is not enough to yield a sufficiently large set of estimates for comparison. We therefore use jacknife resampling to generate more datasets in which we randomly remove one of the countries.

We now review each of the three datasets and the results obtained using each method.

4.2.1. Terrorism and GDP per Capita in the Basque Country

Abadie and Gardeazabal (2003) find that the outbreak of terrorism in the late 1960s significantly affected per capita GDP in the Basque Country. When compared to an SC region without terrorism, the Basque Country's GDP declined by about 10 percentage points.

Our estimate of the effect on the Basque country is similar to the standard SC estimate. However, when we build SCs for the untreated units themselves, we find that they are closer to the true trajectory than is the standard SC. Figure 5 (top) displays the distribution of our estimated treatment effects for all untreated units. Since these units did not receive a treatment, our SC should ideally be as close as possible to the post-treatment values of the time series. In other words, the average difference between the post-treatment series and the SC should deviate as little as possible from zero. Visually, we observe that the band for our approach is narrower than the one for standard SC.

However, this plot does not capture the full extent of the true improvement, as we should be comparing the pairwise performance of each algorithm (instead of the pooled comparison shown here). To more formally demonstrate the improvement, we calculate the MSE for each SC compared to the unit of interest. We then calculate a t-test of $\log(\text{MSE}_{DSC}/\text{MSE}_{SC})$. A negative value indicates that our MSE is smaller than the one obtained with the standard approach. We find that this is indeed the case. The log ratio of the MSEs is significantly less than zero ($t = -7.91, p < 0.0001$).[21]

In terms of our ability to make inferences, the observed reduction in MSE when employing the DSC method suggests its superior efficiency compared to traditional SC methods. In short, our DSC approach more closely approximates the true treatment effects and therefore enhances the robustness and reliability of our causal inference.

4.2.2. The Effects of Proposition 99 on Tobacco Sales in California

Abadie *et al.* (2010) study the impact of Proposition 99, a large-scale tobacco control program implemented in California in 1988. They show that by 2000, California's per-capita cigarette sales were 26 packs fewer than would have been expected without Proposition 99.

---

[19]The role of covariates, $x_{j,k}$, raises important considerations regarding their speed relative to the outcome variable, y_j. While covariates matching the outcome's speed may not affect our conclusions significantly, it may be possible to mitigate the speed differential in y_j by weighting predictors more heavily, potentially bypassing the need for y_j entirely (Becker and Klößner 2018; Kaul *et al.* 2022). Our analysis here relies on un-warped covariates. However, we show in Supplementary Appendix 11 that the results with warped covariates are very similar. We thank an anonymous reviewer for this suggestion.

[20]That is, we estimate the "treatment effect" for Canada using Japan, France, the United States, and so forth; we estimate the treatment effect for France using Japan, Canada, Japan, and so on for all untreated units.

[21]The raw MSEs are as follows: $\overline{log}(\text{MSE}_{SC}) = 10.06$, $\overline{log}(\text{MSE}_{DSC}) = 9.88$.
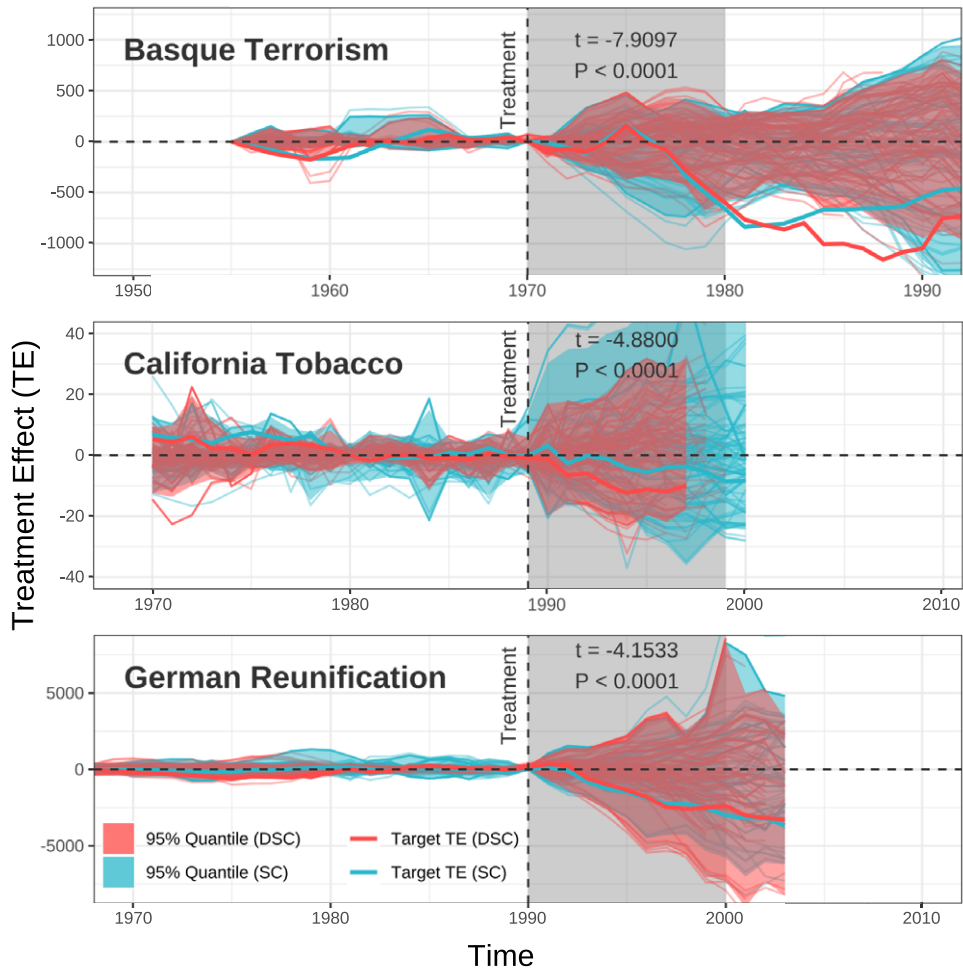
Figure 5. Placebo tests, real data. We revisited the placebo tests reported in Abadie and Gardeazabal (2003), Abadie *et al.* (2010), and Abadie *et al.* (2015). The plots report the placebo tests for each of these studies, using standard synthetic controls (SCs) (blue) and dynamic synthetic control (red). In addition, the estimated treatment effects for the treated units—Basque Country, California State, and West Germany—are shown as thick, brighter lines. For each study, find that our placebo estimates exhibit smaller variance than those using standard SCs, which do not account for variations in speed.

Our analysis suggests that the effect might in fact be greater, with an estimated reduction of about 31 packs in cigarette sales. Although it is impossible to definitively determine which estimate is more accurate, as the true treatment effect remains unknown, we do find that our estimates of the post-treatment behavior of states other than California is closer to the true path than is the standard SC, with smaller MSE ratios ($t = -4.88$, $p < 0.001$).[22]

4.2.3. The Effects of the German Reunification on West Germany's GDP per Capita

Finally, Abadie *et al.* (2015) seek to estimate the economic impact of the 1990 German reunification on West Germany. They find that the per-capita GDP of West Germany was reduced by on average about

---

[22]$\overline{log}(\text{MSE}_{SC}) = 4.08$, $\overline{log}(\text{MSE}_{DSC}) = 4.01$. Note that the SCs generated by our method are shorter than those of the standard SC. This is due to the warping adjustments made to account for varying speeds. Some of the time series are warped to a shorter length due to their slower speeds relative to the target time series (see Supplementary Appendix 12 for more detail).

1,600 USD per year over the 1990–2003 period, which is approximately 8% of the 1990 level. In 2003, the last year of their study period, the estimated per-capita GDP in the synthetic data is about 12% higher than in the real data.

In this replication, our estimate of the treatment effect on West Germany is similar to the one of Abadie and Gardeazabal (2003), and visually the 95% quantile areas of the estimated treatment effect of two SC methods on the untreated countries are also close. But in the pairwise comparison, we find that the DSC method generally exhibits a reduced MSE in estimating treatment effects for control countries, compared to the standard SC approach ($t = -4.15$, $p < 0.001$).[23]

Overall, our findings on all three empirical studies indicate that the DSC method offers significant advantages when compared to the standard SC approach. In particular, DSC achieves closer approximations to the true trajectories of non-treated units, as evidenced by the observed reduction in MSE. As a result, the DSC method provides a more accurate approximation of the true treatment effects. It is a more efficient estimator of the treatment effects and mitigates the risk of biased estimates.

5. Conclusion

In this study, we tackle the issue of varying reaction speeds across different units when estimating causal effects of treatments. This issue arises from the fact that units—such as states, cities, or people—may adjust and react at different speeds due to a multitude of factors, including legal constraints, institutional differences, and the nature of the treatment itself. Ignoring speed variations can bias estimates and conclusions, weakening the validity of treatment effect analyses.

We introduce the DSC method, which extends the SC approach by integrating a DTW algorithm to adjust for speed differences. By doing so, the DSC method enables researchers to construct counterfactuals that more accurately represent the hypothetical outcomes for treated units without the treatment, with improved precision and efficiency of treatment effect estimates compared to the standard SC method.

Through Monte Carlo simulations and real-world datasets, we show that DSC outperforms standard SC in treatment effect estimation. It reduces uncertainty and boosts test power, minimizing the risk of false conclusions. These results highlight the value of our approach in addressing the speed problem and improving the accuracy and precision of treatment effect estimates.

Future research will be needed to explore the applications of DSC in various settings, such as assessing the impact of policies, interventions, or shocks across multiple dimensions and over different time horizons. Additional research could also assess DSC's sensitivity to unobservable variables and endogenous regressors.

Data Availability Statement. Replication code for this article has been published in Code Ocean, a computational reproducibility platform that enables users to run the code, and can be viewed interactively at https://codeocean.com/capsule/7670710/tree/v1. A preservation copy of the same code and data can also be accessed via Dataverse at https://doi.org/10.7910/DVN/DIUPUA (Cao and Chadefaux 2024).

Supplementary Material. For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.2024.14.

References

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

Abadie, A., and M. D. Cattaneo. 2021. "*Introduction to the Special Section on Synthetic Control Methods.*" *Journal of the American Statistical Association* 116 (536): 1713–1715

---

[23]$\overline{log(\mathrm{MSE}_{SC})} = 13.89$, $\overline{log(\mathrm{MSE}_{DSC})} = 13.78$.

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 (490): 493–505.

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

Abadie, A., and J. Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93 (1): 113–132.

Alesina, A., G. D. Cohen, and N. Roubini. 1992. "Macroeconomic Policy and Elections in OECD Democracies." *Economics and Politics* 4 (1): 1–30.

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

Barro, R. J., X. Sala-i-Martin, O. J. Blanchard, and R. E. Hall. 1991. "Convergence across States and Regions." *Brookings Papers on Economic Activity* 22 (1): 107–182.

Becker, M., and S. Klößner. 2016. "Estimating the Economic Costs of Organized Crime by Generalized Synthetic Control Methods."

Becker, M., and S. Klößner. 2018. "Fast and Reliable Computation of Generalized Synthetic Controls." *Econometrics and Statistics* 5: 1–19.

Ben-Michael, E., A. Feller, and J. Rothstein. 2022. "Synthetic Controls with Staggered Adoption." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 84 (2): 351–381.

Brady, H. E., and J. E. McNulty. 2011. "Turning Out to Vote: The Costs of Finding and Getting to the Polling Place." *American Political Science Review* 105 (1): 115–134.

Canarella, G. and S. Pollard. 2004. "Parameter Heterogeneity in the Neoclassical Growth Model: A Quantile Regression Approach." *Journal of Economic Development* 29: 1–32.

Card, D., and A. B. Krueger. 2000. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply." *American Economic Review* 90 (5): 1397–1420.

Chattopadhyay, R., and E. Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409–1443.

Di Tella, R., and E. Schargrodsky. 2004. "Do Police Reduce Crime? Estimates Using the Allocation of Police Forces after a Terrorist Attack." *American Economic Review* 94 (1): 115–133.

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

Goodman-Bacon, A. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics* 225 (2): 254–277.

Kaul, A., S. Klößner, G. Pfeifer, and M. Schieler. 2022. "Standard Synthetic Control Methods: The Case of Using All Preintervention Outcomes Together with Covariates." *Journal of Business & Economic Statistics* 40 (3): 1362–1376.

Öztekin, Ö., and M. J. Flannery. 2012. "Institutional Determinants of Capital Structure Adjustment Speeds." *Journal of Financial Economics* 103 (1): 88–112.

Rappaport, J. 2000. "Is the Speed of Convergence Constant?" FRB of Kansas City Research Working Paper No. 00-10.

Reinhart, C. M., and K. S. Rogoff. 2014. "Recovery from Financial Crises: Evidence from 100 Episodes." *American Economic Review* 104 (5): 50–55.

Vintsyuk, T. K. 1968. "Speech Discrimination by Dynamic Programming." *Cybernetics* 4 (1): 52–57.