Cambridge Medium (2020), **37**, 1–48

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Dynamic Synthetic Controls

Accounting for Varying Speeds in Comparative Case Studies

Jian Cao* and Thomas Chadefaux

Department of Political Science, Trinity College Dublin, 2 Clare Street, Dublin 2, Ireland

*Corresponding author. Email: caoj@tcd.ie

Abstract

Synthetic controls 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 synthetic control approach that accommodates varying speeds in time series, resulting in improved synthetic control estimates. We apply our method to re-estimate the effects of terrorism on income (Abadie and Gardeazabal 2003), tobacco laws on consumption (Abadie, Diamond, and Hainmueller 2010), and German reunification on GDP (Abadie, Diamond, and Hainmueller 2015). 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 synthetic controls, improving our ability to make robust inferences. An open-source `R` package, `dsc`, is made available for easy implementation.

Keywords: causal inference, time series, dynamic time warping, difference in differences, comparative case studies

Supplementary Material

Appendix 1. The notion of speed

The notion of "speed" in this analysis is fundamentally about the responsiveness of an outcome variable, denoted by y_t, to fluctuations in a latent driving process, z_t. This concept is illustrated through the relationship between y_t and both current and past values of z_t, encapsulated in a series of lagged coefficients.

Consider the formulation of the outcome variable $y_{1,t}$ as influenced by z_t and its historical values:

$$y_{1,t} = \beta_{1,0} z_t + \beta_{1,1} z_{t-1} + \beta_{1,2} z_{t-2} + \cdots + \beta_{1,t-1} z_1,$$

where each coefficient $\beta_{1,l}$ captures the influence of z_t at lag l on $y_{1,t}$. A nonzero value of $\beta_{1,l}$ signifies that $y_{1,t}$ is affected by z_t not just in its current state but also as it existed l periods ago. The presence of higher-order nonzero lag coefficients indicates a delayed response of $y_{1,t}$ to changes in z_t, typifying a "slower" speed of adjustment.

To differentiate speeds between outcomes, consider another outcome variable $y_{2,t}$ modeled similarly:

$$y_{2,t} = \beta_{2,0} z_t + \beta_{2,1} z_{t-1} + \beta_{2,2} z_{t-2} + \cdots + \beta_{2,t-1} z_1,$$

If we denote the highest-order nonzero lag coefficient for $y_{1,t}$ as β_{1,l_1} and for $y_{2,t}$ as β_{2,l_2}, the comparison of l_1 and l_2 provides a relative measure of speed: if $l_1 < l_2$, it indicates that $y_{1,t}$ responds more rapidly to changes in z_t than does $y_{2,t}$, rendering $y_{1,t}$ "faster" in terms of its reaction time to the latent process dynamics.

This approach allows us to quantitatively assess and compare the speed at which different outcomes adjust to the underlying latent process.

Appendix 2. The speed problem with lags and polynomials

In an attempt to improve the synthetic control's approximation to the treated unit $\mathbf{y}_1$, one might consider enriching the model by adding time lags and polynomial terms. Specifically, in addition to the original control units $\mathbf{y}_2$ and $\mathbf{y}_3$, we extend the set of potential controls to include their lags (up to lag 5) and polynomial terms (up to the 5th degree). These extensions aim to capture additional dynamics and non-linearities in the data.

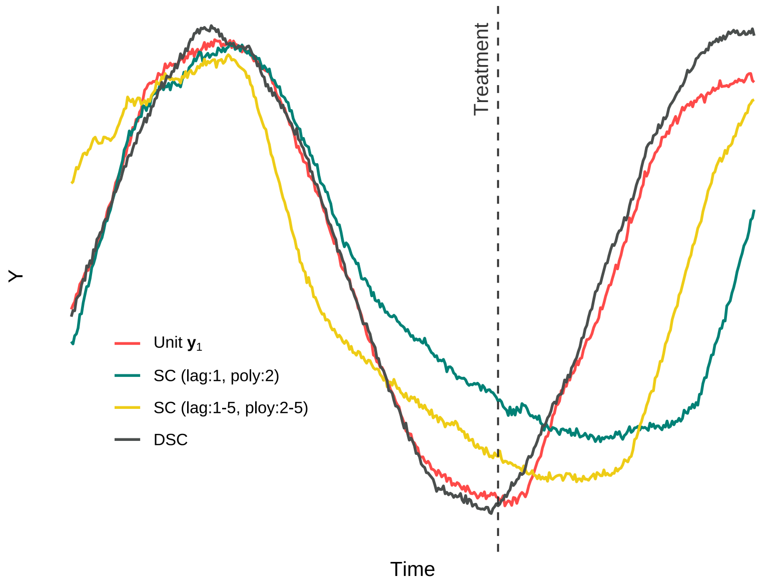

Figure A1. The Challenge of Varying Speeds in Treatment Effect Estimation. A researcher aims to quantify the impact of a treatment on unit $\mathbf{y}_1$. Unbeknownst to them, no treatment effect actually exists. When employing standard synthetic control methods that incorporate control units $\mathbf{y}_2$ (slow) and $\mathbf{y}_3$ (fast) along with their lags (1–5) and polynomial terms (2–5), the estimated post-treatment effect (represented by the blue curve) significantly diverges from the true outcome (indicated by the black curve). In contrast, the Dynamic Synthetic Control method produces an estimated synthetic control that more closely approximates the true trajectory.

As illustrated in Figure A1, however, the inclusion of these additional terms does not appreciably improve the performance of the synthetic control. The blue and yellow curves represent the synthetic control derived from expanded sets of potential controls. The blue curve, for example, solves the following:

$$W^* = \operatorname{argmin}_{W^*} \left\| \mathbf{y}_1 - \left(\mathbf{Y}_{-1}\boldsymbol{\beta}^* + \sum_{l=1}^{5} L^l \mathbf{Y}_{-1}\boldsymbol{\gamma}^* + \sum_{k=2}^{5} \mathbf{Y}_{-1}^k \boldsymbol{\gamma}^* \right) \right\|_2 ,$$

where L^l is the lag operator.

However, both the blue and the yellow curves still deviate from the true trajectory (black curve) of unit $\mathbf{y}_1$. Consequently, one might still erroneously infer a treatment effect where none exists, highlighting the limitations of standard synthetic control methods in handling varying speeds among control units.

Appendix 3. Synthetic Controls: A refresher

Formally, suppose that we have data for $J+1$ units.[24] For time $t \in (1, 2, \ldots, N)$, We observe a target unit $y_{1,t}$ that receives a treatment, and J units $y_{j,t}$, $j \in (2, 3, \ldots, J+1)$ that are untreated and can therefore be used as donor units. Each unit is observed for N periods, and we assume that the treatment takes place at time T. For each unit, we might also observe a set of k predictors $z_{1,j}, z_{2,j} \ldots, z_{k,j}$, although for simplicity we do not include them in the model below.[25] Formally, a synthetic control can be represented by a $J \times 1$ vector of weights, $\mathbf{w} = (w_2, w_3, \ldots, w_{J+1})'$, and the synthetic control estimator of $y_{1,t}^N$ (the potential response in the absence of intervention) is then:

$$\widehat{y}_{1,t}^N = \sum_{j=2}^{J+1} w_j y_{j,t}$$

In other words, synthetic controls are weighted averages of the units in the pool of donors. The estimated effect of intervention is then:

$$\widehat{\tau}_t = y_{1,t} - \widehat{y}_{1,t}^N$$

The key question is how to choose the weights w_j. There are many possibilities, ranging from assigning equal weights to all control units to using a population-weighted unit. A sensible approach is to choose $w_2, w_3, \ldots, w_{J+1}$ such that the resulting synthetic control $\widehat{y}_{1,t}^N$ is as close as possible to the pre-intervention time series for the treated unit as possible (Abadie and Gardeazabal 2003; Abadie, Diamond, and Hainmueller 2010). By minimizing the distance between the trajectory of the treated unit and the combined untreated units, we build a synthetic control that is as close as possible to the (unobservable) counterfactual.

24. This section largely follows Abadie (2021).
25. The predictors can be added after warping for time-constant predictors, or the same algorithm described below can be applied to these predictors for time-varying variables.

Appendix 4. Bias of the synthetic control estimate when the time series have different speeds

In this appendix, we demonstrate that the standard synthetic control estimator may produce biased results when different units respond to shocks at different speeds. Mathematical proofs and models are presented here to support these claims.

Suppose that the target time series $\mathbf{y}_1$ and the donors $\mathbf{y}_j, j \in (2, 3, ..., J+1)$ are of length N. Let T denote the treatment time and let the treatment effect be zero. To obtain weights $\mathbf{w}$, the synthetic control method essentially estimates the following model:

$$y_{1,t} = \sum_{j=2}^{J}(w_j y_{j,t}) + \epsilon_t, t < T$$

if the usual assumptions of the error term hold, then the least square estimate of the weights $\hat{\mathbf{w}}$ is unbiased and efficient, and can produce a synthetic control that is the closest to the target time series $\mathbf{y}_1$.

However, when the time series have different speeds, i.e. each time series has its own delays to responding to the common exogenous shocks $\mathbf{z}$, and the delays vary over time, then the usual assumptions of the error term in model (4) are violated and it introduces bias to $\hat{\mathbf{w}}$.

Taking the different speeds into account, model (11) becomes:

$$\begin{aligned} y_{1,t} &= \sum_{j=2}^{J}[w_j(y_{j,t} + \beta_{1,j,t}y_{j,t-1} + \beta_{2,j,t}y_{j,t-2} + \ldots \beta_{t-1,j,t}y_{j,1})] + \epsilon_t \\ &= (\mathbf{Y}_{-1,t} + \mathbf{Y}_{-1,t-1}\boldsymbol{\beta}_{1,t} + \ldots + \mathbf{Y}_{-1,1}\boldsymbol{\beta}_{t-1,t})\mathbf{w} + \epsilon_t \end{aligned}$$

where $\mathbf{Y}_{-1,t}$ is a vector of donor units at time t, $\mathbf{Y}_{-1,t} = (y_{2,t}, y_{3,t}, ..., y_{J,t})$, and $\mathbf{Y}_{-1,t-1}$, $\mathbf{Y}_{-1,t-2}$, ... , $\mathbf{Y}_{-1,1}$ are all possible lags at time t. $\mathbf{w}$ is the time-independent coefficient. $\boldsymbol{\beta}_{l,t}$ are time-dependent coefficients of the lags. $\epsilon \sim N(0, \sigma^2 I)$ is the error term.

Model (13) allows a varying dependency of the target variable $\mathbf{y}_1$ on lags of the donor variables $\mathbf{Y}_{-1}$. This varying dependency captures the different speeds that $\mathbf{y}_1$ and $\mathbf{Y}_{-1}$ respond to the common exogenous shocks $\mathbf{z}$.

If the time series have different speeds and (13) is the true model, the original synthetic control method would obtain biased estimates of the treatment effect as the lag terms are omitted.

Pre-treatment Period

In traditional synthetic control analysis, the model of interest is,

$$y_{1,t} = \mathbf{Y}_{-1,t}\mathbf{w}_{sc} + \eta_t$$

the new error term η_t is

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

which is the summation of the original error term ϵ_t and the lag terms $\mathbf{Y}_{-1,t-1}\boldsymbol{\beta}_{1,t}\mathbf{w}+\ldots+\mathbf{Y}_{-1,1}\boldsymbol{\beta}_{t-1,t}\mathbf{w}$.

According to Greene 2003, the expectation of $\widehat{\mathbf{w}}_{sc}$ is

$$E[\widehat{\mathbf{w}}_{sc}] = \mathbf{w} + \mathbb{G}\boldsymbol{\beta}_L\mathbf{w}$$

where $\mathbb{G} = (\mathbf{Y}'_{-1}\mathbf{Y}_{-1})^{-1}\mathbf{Y}'_{-1}\mathbf{Y}_{-1,L}$. $\mathbf{Y}_{-1,L}$ is a matrix of lags of $\mathbf{Y}_{-1}$, and $\boldsymbol{\beta}_L$ is matrix of corresponding coefficients.

The expectation of residual

$$\begin{aligned} E[\widehat{\eta}] &= E[\mathbf{Y}_{-1,L}\boldsymbol{\beta}_L + \boldsymbol{\epsilon}] \\ &= E[\mathbf{Y}_{-1,L}\boldsymbol{\beta}_L] + 0 \end{aligned}$$

depends on the existence of lag effects. If lag effects exist, i.e. $E[\mathbf{Y}_{-1,L}\boldsymbol{\beta}_L] \neq 0$, $\widehat{\eta}$ is biased.

The variance of residual

$$\begin{aligned} \mathrm{Var}(\widehat{\eta}) &= \mathrm{Var}(\mathbf{Y}_{-1,L}\boldsymbol{\beta}_L + \boldsymbol{\epsilon}) \\ &= \mathrm{Var}(\mathbf{Y}_{-1,L}\boldsymbol{\beta}_L) + \mathrm{Var}(\boldsymbol{\epsilon}) + 0 \\ &= \mathrm{Var}(\mathbf{Y}_{-1,L}\boldsymbol{\beta}_L) + \sigma^2 I \\ &\geq \sigma^2 I \end{aligned}$$

is at least $\sigma^2 I$. If the lag effects exist, i.e. $E[\mathbf{Y}_{-1,L}\boldsymbol{\beta}_L] \neq 0$, $\mathrm{Var}(\widehat{\eta}) > \sigma^2 I$.

On the contrary, in dynamic synthetic control, the model of interest is:

$$y_t = \mathbf{Y}^*_{-1,t}\mathbf{w}_{dsc} + \epsilon_t$$

where

$$\mathbf{Y}^*_{-1,t} = \mathbf{Y}_{-1,t} + \mathbf{Y}_{-1,t-1}\boldsymbol{\beta}_{1,t} + \ldots + \mathbf{Y}_{-1,1}\boldsymbol{\beta}_{t-1,t}. \tag{A1}$$

The expectation of $\widehat{\mathbf{w}}_{dsc}$ is

$$E[\widehat{\mathbf{w}}_{dsc}] = \mathbf{w}$$

Note the error term is the same as in model (2). so the expectation of residual is just

$$E[\widehat{\epsilon}] = 0$$

and the variance of residual

$$\mathrm{Var}(\widehat{\epsilon}) = \sigma^2 I$$

Therefore, if time series have different speeds, i.e. effects of lags of $\mathbf{Y}_{-1}$ are not zero, in the pre-treatment period, traditional synthetic control method would fail to obtain a synthetic control that closely resembles $\mathbf{y}_1$.

Post-treatment Period

In post-treatment period, both synthetic control methods estimates the treatment effect using post-treatment $\mathbf{y}_1, \mathbf{Y}_{-1}$ and pre-treatment $\widehat{\mathbf{w}}_{sc}, \widehat{\mathbf{w}}_{dsc}$. For cleaner notation, let A denote the pre-treatment period and B denote the post-treatment period.

The estimated treatment effect from traditional synthetic control is:

$$\widehat{\boldsymbol{\tau}}_{sc} = \mathbf{y}_1^B - \mathbf{Y}_{-1}^B\widehat{\mathbf{w}}_{sc}^A$$

The expectation of $\widehat{\boldsymbol{\tau}}_{sc}$ is

$$\begin{aligned}
E[\widehat{\boldsymbol{\tau}}_{sc}] &= E[\mathbf{y}_1^B - \mathbf{Y}_{-1}^B \widehat{\mathbf{w}}_{sc}^A] \\
&= E[(\mathbf{Y}_{-1}^B \mathbf{w} + \mathbf{Y}_{-1,L}^B \boldsymbol{\beta}_L^B) - \mathbf{Y}_{-1}^B \widehat{\mathbf{w}}_{sc}^A] \\
&= E[\mathbf{Y}_{-1,L}^B \boldsymbol{\beta}_L^B + \mathbf{Y}_{-1}^B (\mathbf{w} - \mathbf{w} - \mathbb{G}^A \boldsymbol{\beta}_L^A \mathbf{w})] \\
&= E[\mathbf{Y}_{-1,L}^B \boldsymbol{\beta}_L^B - \mathbf{Y}_{-1}^B \mathbb{G}^A \boldsymbol{\beta}_L^A \mathbf{w}]
\end{aligned}$$

Whether the expectation of $\widehat{\boldsymbol{\tau}}_{sc}$ equals to zero depends on $\mathbf{Y}_{-1}$, $\mathbf{Y}_{-1,L}$, $\boldsymbol{\beta}_L$, and $\mathbf{w}$. There are two sources of bias. One is omitted lag effects, $\mathbf{Y}_{-1,L}^B \boldsymbol{\beta}_L^B$. The other is biased estimate of weights, $-\mathbf{Y}_{-1}^B \mathbb{G}^A \boldsymbol{\beta}_L^A \mathbf{w}$. The two sources of bias are both zeros only when the lag effects are zero, i.e. $E[\mathbf{Y}_{-1,L} \boldsymbol{\beta}_L] = 0$. Otherwise, as the two sources of bias are only marginally correlated ($\mathbf{Y}_{-1}$ and $\mathbf{Y}_{-1,L}$ have some common terms), it is very rare that the two parts have different signs and add up to zero, which means that the estimated treatment effect from traditional synthetic control is highly likely biased.

The variance of $\widehat{\boldsymbol{\tau}}_{sc}$

$$\begin{aligned}
\mathrm{Var}[\widehat{\boldsymbol{\tau}}_{sc}] &= \mathrm{Var}(\mathbf{y}_1^B - \mathbf{Y}_{-1}^B \widehat{\mathbf{w}}_{sc}^A) \\
&= \mathrm{Var}[(\mathbf{Y}_{-1}^B \mathbf{w} + \mathbf{Y}_{-1,L}^B \boldsymbol{\beta}_L^B + \boldsymbol{\epsilon}) - \mathbf{Y}_{-1}^B \widehat{\mathbf{w}}_{sc}^A] \\
&= \mathrm{Var}[\mathbf{Y}_{-1,L}^B \boldsymbol{\beta}_L^B - \mathbf{Y}_{-1}^B \mathbb{G}^A \boldsymbol{\beta}_L^A \mathbf{w} + \boldsymbol{\epsilon}] \\
&= \mathrm{Var}[\mathbf{Y}_{-1,L}^B \boldsymbol{\beta}_L^B - \mathbf{Y}_{-1}^B \mathbb{G}^A \boldsymbol{\beta}_L^A \mathbf{w}] + \sigma^2 I
\end{aligned}$$

is larger or equal to $\sigma^2 I$. The equal sign holds only when $\widehat{\boldsymbol{\tau}}_{sc}$ is not biased, i.e. (25) equals zero.

Instead, the estimated treatment effect from dynamic synthetic control is

$$\widehat{\boldsymbol{\tau}}_{dsc} = \mathbf{y}_1^B - \mathbf{Y}_{-1}^{*B} \widehat{\mathbf{w}}_{dsc}^A$$

The expectation of $\widehat{\boldsymbol{\tau}}_{dsc}$ is

$$\begin{aligned} E[\widehat{\boldsymbol{\tau}}_{dsc}] &= E[\mathbf{y}_1^B - \mathbf{Y}_{-1}^{*B}\widehat{\mathbf{w}}_{dsc}^A] \\ &= E[\mathbf{y}_1^B - \mathbf{Y}_{-1}^{*B}\mathbf{w}] \\ &= E[\boldsymbol{\epsilon}^B] \\ &= 0 \end{aligned}$$

The variance of $\widehat{\boldsymbol{\tau}}_{dsc}$ is

$$\begin{aligned} \mathrm{Var}[\widehat{\boldsymbol{\tau}}_{dsc}] &= \mathrm{Var}(\mathbf{y}_1^B - \mathbf{Y}_{-1}^{*B}\widehat{\mathbf{w}}_{dsc}^A) \\ &= \mathrm{Var}[(\mathbf{Y}_{-1}^B\mathbf{w} + \mathbf{Y}_{-1,L}^B\boldsymbol{\beta}_L^B + \boldsymbol{\epsilon}) - \mathbf{Y}_{-1}^B\widehat{\mathbf{w}}_{sc}^A] \\ &= \mathrm{Var}(\mathbf{y}_1^B - \mathbf{Y}_{-1}^{*B}\mathbf{w}) \\ &= \mathrm{Var}(\boldsymbol{\epsilon}^B) \\ &= \sigma^2 I \end{aligned}$$

Therefore, when time series have different speeds, dynamic synthetic control method can produce unbiased and efficient estimate of the treatment effect while the traditional synthetic control method would most likely produce biased estimates and inflated variances.

Appendix 5. Discussion of identification threats

This section addresses some of the challenges of identifying causal effects.

Unobserved Heterogeneity

First, we examine the impact of an unobserved characteristic, $\boldsymbol{\varphi}_j$, which influences both the likelihood of treatment and the dynamics before and after treatment, thereby affecting the interpretation of the results. We discuss here how this unobserved factor affects the estimation of treatment effects and implications for empirical analysis.

The outcome $\mathbf{y}_j$ can be written as:

$$\mathbf{y}_j = f_{\mathbf{y}}(\mathbf{s}_j, \mathbf{X}_j, \mathbf{z}) + \pi_j \boldsymbol{\tau}_j,$$

where $f_{\mathbf{y}}(\mathbf{s}_j, \mathbf{X}_j, \mathbf{z})$ is the outcome time series, which is determined by the universal shocks $\mathbf{z}$ that applied to all units, speed profile $\mathbf{s}_j$, and predictors $\mathbf{X}_j$. The second part $\pi_j \boldsymbol{\tau}_j$ is the random treatment, where π_j is the probability of receiving a treatment and $\boldsymbol{\tau}_j$ is the treatment effect.

Assume the treatment also changes the speed profile. The new speed profile $\mathbf{s}_j^{'}$ can be simplified as a combination of the original speed and the changes in speed:

$$\mathbf{s}_j^{'} = \mathbf{s}_j + \Delta \mathbf{s}_j$$

Given the speed changes, the treatment effect $\boldsymbol{\tau}_j$ consists of two parts:

$$\boldsymbol{\tau}_j = f_{\mathbf{y}}(\Delta \mathbf{s}_j, \mathbf{X}_j, \mathbf{z}) + f_{\mathbf{y}}(\mathbf{s}_j + \Delta \mathbf{s}_j, \mathbf{X}_j, \boldsymbol{\theta}),$$

The first part $f_{\mathbf{y}}(\Delta \mathbf{s}_j, \mathbf{X}_j, \mathbf{z})$ is caused by changes in speed. The second part $f_{\mathbf{y}}(\mathbf{s}_j + \Delta \mathbf{s}_j, \mathbf{X}_j, \boldsymbol{\theta})$ is caused by the treatment $\boldsymbol{\theta}$.

Assume there is an unobserved characteristic $\boldsymbol{\varphi}_j$ affects both the speed profiles and the probability of receiving a treatment:

$$\mathbf{s}_j = f_{\mathbf{s}}(\boldsymbol{\varphi}_j)$$
$$\Delta\mathbf{s}_j = f_{\Delta\mathbf{s}}(\boldsymbol{\varphi}_j)$$
$$\pi_j = f_{\pi}(\boldsymbol{\varphi}_j),$$

If the research focus on the expected treatment effect for all units, it is of the form:

$$\frac{1}{J+1}\sum_{j=1}^{J+1}\pi_j\boldsymbol{\tau}_j$$

The unobserved characteristic $\boldsymbol{\varphi}_j$ affects the result in the way that, for units who have larger probability to receive treatments, their reaction to the treatment, e.g. slow/fast response speed, reduced/amplified response level, and etc., will highly influence the expected treatment effect.

If the research focus only on the treatment effect on the treated units, the model is reduced to a conditional form:

$$\mathbf{y}_1 = f_{\mathbf{y}}(\mathbf{s}_1, \mathbf{X}_1, \mathbf{z}) + \boldsymbol{\tau}_1$$
$$\mathbf{y}_j = f_{\mathbf{y}}(\mathbf{s}_j, \mathbf{X}_j, \mathbf{z}),$$

conditional on $\mathbf{y}_1$ received the treatment and the others did not.

To estimate the treatment effect, we first warp the donor units $\mathbf{y}_j$ to make them align to the speed of $\mathbf{y}_1$:

$$\mathbf{y}_j^{w} \approx f_{\mathbf{y}}(\mathbf{s}_1, \mathbf{X}_j, \mathbf{z})$$

Then we estimate the weights and generate a synthetic control that closely resembles unit $\mathbf{y}_1$ prior to the treatment.

$$\mathbf{y}_1^{dsc} = \mathbf{y}_{-1}^{w}\mathbf{w} \approx f_{\mathbf{y}}(\mathbf{s}_1, \mathbf{X}_1, \mathbf{z})$$

The estimated treatment effect is:

$$
\begin{aligned}
\widehat{\boldsymbol{\tau}}_1 &= \mathbf{y}_1 - \mathbf{y}_1^{dsc} \\
&\approx f_{\mathbf{y}}(\Delta \mathbf{s}_j, \mathbf{X}_j, \mathbf{z}) + f_{\mathbf{y}}(\mathbf{s}_j + \Delta \mathbf{s}_j, \mathbf{X}_j, \boldsymbol{\theta}).
\end{aligned}
$$

The unobserved characteristic $\boldsymbol{\varphi}_1$ affects the result through post-treatment dynamics $\mathbf{s}_j$ and $\Delta \mathbf{s}_j$.

Spillover Effects

We will now examine how spillovers from the treatment to the donor units might influence the results. These effects are a significant concern for causal inference setups, such as Difference-in-Differences and synthetic controls, because they imply that the control units are also impacted by the treatment. Consequently, we generally expect spillover effects to result in an underestimation of the treatment effect size and inflated standard errors.

Suppose the treatment effect is τ_t and that the spillover effect on donor unit j is, for simplicity, an additive term of the form $p_j \tau_t, p_j \in (0, 1)$. Then the affected donor unit j would be:

$$
y'_{j,t} = y_{j,t} + p_j \tau_t
$$

Since the patterns in the post-treatment donor unit are affected by the spillover, there is a risk that they may now be matched to the wrong pre-treatment sequences. There are two cases to consider:

Case 1: The matched sequence is correct. When warping the donor unit, both original donor unit and the spillover effect are warped:

$$
y^{w,'}_{j,t} = y^{w}_{j,t} + p_j \tau^{w}_t.
$$

The addition of the warped spillover effect $p_j \tau^w_t$ will cause bias. If the spillover effect is warped faster, i.e. the effect unfolds in a shorter period of time, then the warped spillover effect would be larger, i.e. $|p_j \tau^w_t| > |p_j \tau_t|$. Otherwise, if the spillover effect is warped slower, then the effect is smaller $|p_j \tau^w_t| < |p_j \tau_t|$. However, notice that no matter whether the spillover effect is warped faster or slower, its consequence is still to dampen our estimate of the treatment effect. The warping does not affect the direction of the bias. The bias is always toward underestimation. What the warping does is affect whether the underestimation is more or less severe—more severe when the spillover effect is sped up, less severe when it is slowed down.

Case 2: If, on the other hand, the matched sequence is incorrect, then the warped donor unit would be:

$$y_{j,t}^{w,'} = (y_{j,t}^{w} + \epsilon_{j,t}) + p_j \tau_t^{w},$$

where errors $\epsilon_{j,t}$ are introduced because the donor unit is incorrectly warped. The errors $\epsilon_{j,t}$ could be either positive or negative. If the combination of errors and warped spillover effect is in the same direction of the treatment effect (have the same sign), i.e., if:

$$(\epsilon_{j,t} + p_j \tau_t^{w}) \times \tau_t^{w} > 0,$$

the treatment effect will be underestimated. If the combination is on the opposite direction of the treatment effect (have different signs), the treatment effect will be overestimated.

Appendix 6. The forward lag coefficients

In model 1, $y_{1,t}$ can be influenced not only by past values of $y_{j,t}$, but also by its future values. This does not suggest that the future impacts the past. Instead, suppose that both $\mathbf{y}_j$ and $\mathbf{y}_1$ are functions of a latent variable $\mathbf{z}$—a variable representing true underlying shocks—while the observed time series for units $\mathbf{y}_j$ and $\mathbf{y}_1$ are the manifestations of the shocks. In this situation, the observed time series are always slowed-down versions of $\mathbf{z}$, but at different rates. For example, the time series for $\mathbf{y}_1$ may reflect changes in $\mathbf{z}$ immediately, whereas there may be a delay for $\mathbf{y}_j$. As a result, we would observe that $y_{1,t}$ is a function of the future of $y_{j,t}$, simply because $\mathbf{y}_1$ reacts faster to changes in $\mathbf{z}$ than $\mathbf{y}_j$ does. This phenomenon, for example, occurs in the US financial markets, where the reactions of the bond and stock markets to sudden changes in the Federal Reserve's monetary policies often occur at different speeds (Fleming and Remolona 1999). After the announcement of an interest rate hike, the bond market reacts to the news instantaneously, while the stock market requires active trading to adjust prices. In such cases, we observe a strong correlation between current bond prices and stock prices shortly thereafter. This means that, in order to correct for the relative speeds of $\mathbf{y}_j$ and $\mathbf{y}_1$, we need not only consider past lags of $\mathbf{y}_j$, but also forward lags. The forward lags have non-zero coefficients when $\mathbf{y}_1$ is faster than $\mathbf{y}_j$. However, note that since $\mathbf{y}_1$ and $\mathbf{y}_j$ are always slower than the latent common exogenous shocks $\mathbf{z}$, the forward lags of $\mathbf{y}_j$ are associated with exogenous shocks that have already occurred, not future events. We do not suggest that the future affects the past, but rather simply that $\mathbf{y}_j$ and $\mathbf{y}_1$ react at different speeds to $\mathbf{z}$, and hence that $\mathbf{y}_1$ may *appear* to be a function of future $\mathbf{y}_j$. This implies that there are really a total of JNN potential coefficients to estimate ((J donor time series) $\times$ (N time periods) $\times$ ($N/2$ backward lags + $N/2$ forward lags)).

Appendix 7. The Dynamic Synthetic Control algorithm

Algorithm A1 Dynamic Synthetic Control (DSC)

Input: $\mathbf{y}_1 = \{y_{1,1}, y_{1,2}, ..., y_{1,N}\}$: time series with treatment.
$\mathbf{y}_j = \{y_{j,1}, y_{j,2}, ..., y_{j,N}\}$: J time series without treatment.
T: treatment time.
C_j: cutoff time of $\mathbf{y}_j$.
k: size of the sliding window.
m: margin of the sliding window.
n_Q: increment for sliding the target window.
n_R: increment for sliding the reference window.
θ: threshold of distance in window matching.

Output: synthetic control $\mathbf{y}_1^N$.

Process:

$j \leftarrow 1$
for $j = 2$ to $J + 1$ **do**
  Step 1: Match $\mathbf{y}_{1,1:T}$ and $\mathbf{y}_j$;
    DTW match $\mathbf{y}_{1,1:T}$ and $\mathbf{y}_j$ with $open.end = TRUE$,
    save $\mathbf{y}_{j,1:C_j}$ that matches $\mathbf{y}_{1,1:T}$ as $\mathbf{y}_{j,pre}$,
    save $\mathbf{y}_{j,C_j:N}$ as $\mathbf{y}_{j,post}$, $\mathbf{y}_{1,1:T}$ as $\mathbf{y}_{1,pre}$, and $\mathbf{y}_{1,T:N}$ as $\mathbf{y}_{1,post}$
    save warping path $\mathbf{y}_{j,pre} \rightarrow \mathbf{y}_{1,pre}$ as $\mathbf{P}_{j,pre}$,
  Step 2: Match $\mathbf{y}_{j,post}$ and $\mathbf{y}_{j,pre}$ (double sliding window);
    $u \leftarrow 1$; $\mathbb{P}_i \leftarrow NULL$
    while $u \leq length(\mathbf{y}_{j,post}) - k$ **do**
      locate target window $Q_u = \mathbf{y}_{j,post}[u : (u + k)]$,
      $i \leftarrow 1$; $costs \leftarrow NULL$
      while $i \leq length(\mathbf{y}_{j,pre}) - k - m$ **do**
        locate reference window $R_i = \mathbf{y}_{j,pre}[i : (i + k + m)]$,
        DTW match Q_u and R_i with $open.end = TRUE$,
        save DTW distance to $costs$,
        $i \leftarrow i + n_R$.
      end while
      if $min(costs) \leq \theta$ **then**
        find i^* that minimizes $costs$,
        locate reference window $R^* = \mathbf{y}_{j,pre}[i^* : (i^* + k + m)]$,
        subset $\mathbf{P}_{j,pre}$ and obtain warping path $\mathbf{P}_{j,R^*}$,
        save warping path $Q_u \rightarrow R^*$ as $\mathbf{P}_{j,Q_u \rightarrow R^*}$,
        obtain warping path $\mathbf{P}_{j,Q_u} = \mathbf{P}_{j,Q_u \rightarrow R^*}(\mathbf{P}_{j,R^*})$,
        store $\mathbf{P}_{j,Q_u}$ in list $\mathbb{P}_j$,
      end if
      $u \leftarrow u + n_Q$.
    end while
    merge all $\mathbf{P}_{j,Q_u}$ in $\mathbb{P}_j$, obtain $\mathbf{P}_{j,post}$.
  Step 3: Warp $\mathbf{y}_j$ according to $\mathbf{y}_1$
    $\mathbf{y}_j^w = [\mathbf{P}_{j,pre}(\mathbf{y}_{j,pre}), \mathbf{P}_{j,post}(\mathbf{y}_{j,post})]$.
end for
Step 4: use $\mathbf{y}_j^w, j \in (2, \ldots, J + 1)$ to construct synthetic control $\mathbf{y}_1^N$.
return $\mathbf{y}_1^N$

Appendix 8. Deriving post-treatment warping paths from pre-treatment warping paths

To derive the warping path $\mathbf{P}_{j,post}$ from $\mathbf{P}_{j,pre}$, we first match the short-term patterns in $\mathbf{y}_{j,post}$ to the patterns in $\mathbf{y}_{j,pre}$, and for each matched pattern in $\mathbf{y}_{j,pre}$, we obtain a sub-warping path matrix by subsetting $\mathbf{P}_{j,pre}$; then we adjust the sub-warping path matrix according to the DTW matching from the pattern in $\mathbf{y}_{j,post}$ to the pattern in $\mathbf{y}_{j,pre}$; lastly, we combine the adjusted sub-warping path and obtain $\mathbf{P}_{j,post}$.

To find the similar patterns in $\mathbf{y}_{j,post}$ and $\mathbf{y}_{j,pre}$, we use a double-sliding windows approach as shown in the second part of Figure 2. The first sliding window is in $\mathbf{y}_{j,post}$ and the second in $\mathbf{y}_{j,pre}$. Specifically, for any target window Q_u in the post-treatment time series $\mathbf{y}_{j,post}$, the method finds an optimal reference window R^* in the pre-treatment time series $\mathbf{y}_{j,pre}$ that minimizes the DTW distance between Q_u and R_i. Pattern R^* is therefore the closest match to pattern Q_u:

$$
\begin{aligned}
R^* &= \text{argmin}_{R_i}[DTW(Q_u, R_i)], \\
Q_u &= \mathbf{y}_{j,u:(u+k)}, u \in [C, N-k], k \in \mathbb{Z}^+ \\
R_i &= \mathbf{y}_{j,i:(i+k+m)}, i \in [1, C-k-m], m \in \mathbb{Z}^0
\end{aligned}
$$

where k is the default size of the sliding windows, and m is an extra right boundary of the reference window R_i for the open-end DTW, i.e. allowing the the target window Q_u to freely match to a pattern that is within R_i, which enables the DSC algorithm to consider patterns in different lengths. m is chosen by the user to ensure it is large enough to have the matching achieved within R_i, while as small as possible to reduce the computational burden.

We recommend using slope-constrained step patterns such as *symmetricP2* or *asymmetricP2* in DTW to avoid extreme warping paths. In addition, all windows are normalized before DTW for better matching results.

If no similar pattern can be found in $\mathbf{y}_{j,pre}$, i.e. $min(costs)$ is too large, R^* can not provide enough information to help build the warping path $\mathbf{P}_{j,post}$. Instead, including poor matches would introduce noise to the warping path. To avoid this problem, we use a threshold θ to filter out the poor matches.

For each pattern Q_u that has a close match R^* whose DTW distance does not exceed the threshold θ, the warping path $Q_u \rightarrow R^*$ is stored in a $k \times (k+m)$ matrix $\mathbf{P}_{j,Q_u}$. Although Q_u and R^* are close matches, they are not identical. We can not directly use warping path $\mathbf{P}_{j,R^*}$ (i.e. sub-matrix of $\mathbf{P}_{j,pre}$

that warps R^* towards $\mathbf{y}_{1,pre}$) to warp Q_u. We need to adjust $\mathbf{P}_{j,R^*}$ and take the difference between Q_u and R^* into account.

To help transform $\mathbf{P}_{j,R^*}$ using $\mathbf{P}_{j,Q_u}$, we can transform the warping path matrix $\mathbf{P}_{j,pre}$ into a Speed Profile (SP):

$$\phi_{j,pre} = \{\mathbf{P}_{j,pre}[diag(\mathbf{I}_{1\times C}\mathbf{P}_{j,pre})^{-1}]\}\mathbf{I}_{T\times 1}$$

where $\mathbf{I}_{1\times C}$ is a row of ones of length C, and similarly, $\mathbf{I}_{T\times 1}$ is a column of ones of length T.

A speed profile can also be seen as a one-dimensional version of the warping path matrix. It generates the same warped time series as the warping path matrix does. The difference between a speed profile and a warping path matrix is that the speed profile only supports single-direction warping, while the latter supports both directions. Since we are not interested in warping $\mathbf{y}_{1,pre} \rightarrow \mathbf{y}_{j,pre}$, in our case, an one-dimensional speed profile works as well as a warping path matrix. And most importantly, it greatly simplifies the process of deriving the warping path $\mathbf{P}_{j,post}$, as we can use a sub-speed profile to warp the corresponding window of time series and ignore the remaining speed profile without losing any information. In our method, a reference window $R_i = \mathbf{y}_{j,i:(i+k+m)}$, $i \in [1, C-m-k]$ can be warped towards $\mathbf{y}_{1,pre}$ using the sub-speed profile $\phi_{R_i} = \phi_{j,pre}[i:(i+k+m)]$.

The DSC algorithm then uses the warping path $\mathbf{P}_{j,Q_u}: Q_u \rightarrow R^*$, to transform the sub-speed profile ϕ_{j,R^*} and obtains a new speed profile ϕ_{j,Q_u} for $Q_u \rightarrow \mathbf{y}_{1,post}$. The speed profile ϕ_{j,Q_u} on one hand preserves the speed relationship between $\mathbf{y}_{j,pre}$ and $\mathbf{y}_{1,pre}$, on the other hand, makes adjustments and makes sure the difference between Q_u and R^* is taken into account. Specifically, without losing generality, in matching target window Q_u and its closest reference window R^*, if one data point $Q_u[s]$ matches one or more data points $R^*[p:q]$, we define the weight of $Q_u[s]$ in speed profile is the average of the weights of $R^*[p:q]$. Formally,

$$\phi_{j,Q_u}[s] = \frac{\sum_{i\in[p,q]}\phi_{j,R^*}[i]}{q-p+1}, s \in [1,k], 1 \leq p \leq q \leq (k+m)$$

The transformed speed profile ϕ_{j,Q_u} can be obtained through the following process:

$$\phi_{j,Q_u} = \{[diag(\mathbf{P}_{j,Q_u}\mathbf{I}_{(k+m)\times 1})^{-1}]\mathbf{P}_{j,Q_u}\}\phi_{j,R^*}$$

The resulting speed profiles ϕ_{j,Q_u}, $u \in [C, N - k]$ are stacked in a $(N - C - k + 2) \times (N - C + 1)$ matrix $\boldsymbol{\phi}_j$:

$$\boldsymbol{\phi}_j = [\phi_{j,\xi,\upsilon}], \xi \in [1, N - C - k + 2], \upsilon \in [1, N - C + 1]$$
$$\phi_{j,\xi,\upsilon} = \phi_{j,Q_u}[\upsilon - \xi + 1]$$

And the estimate of the speed profile $\phi_{j,post}$ is the column mean of the stacked weight matrix $\boldsymbol{\phi}_j$:

$$\phi_{j,post} = \mathbf{I}_{1 \times (N-C-k+2)} \boldsymbol{\phi}_j$$

Please see Appendix 9 for the method to warp time series using a speed profile.

Appendix 9. Warp time series using speed profile

As discussed in Appendix 8, the speed profile ϕ is an one-dimensional version of a warping path matrix $\mathbf{P}$:

$$\phi = \{\mathbf{P}[diag(\mathbf{I}_{1\times nrow(\mathbf{P})}\mathbf{P})^{-1}]\}\mathbf{I}_{ncol(\mathbf{P})\times 1}$$

It is used in the DSC algorithm to simplify the double-sliding window approach and can also be used to warp time series.

Given a speed profile ϕ, we can warp the corresponding time series $\mathbf{y}$ through the following transformation:

$$\mathbf{y}^{w} = \mathbf{y}(\phi) = \mathbf{H}\mathbf{y}$$

where $\mathbf{H}$ is a transformation matrix determined by the speed profile ϕ. The process for obtaining $\mathbf{H}$ from ϕ is shown in Algorithm A2.

Algorithm A2 Obtain transformation matrix H from speed profile ϕ

Input: $\phi = \{\phi_1, \phi_2, \ldots, \phi_N\}$: speed profile for $\mathbf{y}$.
N: length of ϕ.
Output: H: transformation matrix used to warp $\mathbf{y}$.
Process:

$i \leftarrow 1; j \leftarrow 1; H \leftarrow NULL; r \leftarrow \phi[1]; c \leftarrow 1$
while $i \leq N$ **do**
    $H[j, i] = min(r, c)$
    if $r > c$ **then**
        $j = j + 1$
        $r = r - c$
        $c = 1$
    end if
    if $r < c$ **then**
        $i = i + 1$
        $c = c - r$
        $r = \phi[i]$
    end if
    if $r == c$ **then**
        $i = i + 1$
        $j = j + 1$
        $c = 1$
        $r = \phi[i]$
    end if
end while
return H

For example, a speed profile $\phi = (1.5, 1, 0.5)$ gives a transformation matrix **H**:

$$\phi = \begin{bmatrix} 1.5 & 1 & 0.5 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{bmatrix} = H$$

Notice that the transformation matrix is just a matrix of lag coefficients $\beta_{l,t}$ in equation 1:

$$H = \begin{bmatrix} \beta_{0,1} = 1 & \beta_{-1,1} = 0 & \beta_{-2,1} = 0 \\ \beta_{1,2} = 0.5 & \beta_{0,2} = 0.5 & \beta_{-1,2} = 0 \\ \beta_{2,3} = 0 & \beta_{1,3} = 0.5 & \beta_{0,3} = 0.5 \end{bmatrix}.$$

And process **Hy** is equivalent to combining the lag terms into a new time series in equation A1.

Appendix 10. Data generation in Monte Carlo study

Each dataset contains ten time series. They all follow a common Autoregressive integrated moving average (ARIMA) process but exhibit differing speeds. To simulate these varying speeds, each time series j is warped by a matrix $\mathbf{P}_j$, which is either random or a function of the time series direction (increasing or decreasing). More formally,

$$\begin{aligned}
\mathbf{y}_1 &= \mathbf{P}_1(\mathbf{z}) + \boldsymbol{\tau} + \boldsymbol{\epsilon}_1, \\
\mathbf{y}_j &= \mathbf{P}_j(\mathbf{z}) + \boldsymbol{\epsilon}_j, \\
\mathbf{P}_j &= f[\psi \mathbf{s}_j + (1-\psi)\boldsymbol{\lambda}_j] \\
\boldsymbol{\lambda}_j &\sim N(\overline{\mathbf{s}_j}, var(\mathbf{s}_j)),
\end{aligned} \tag{A2}$$

where $\mathbf{z}$ denotes a 100 observation-long ARIMA(1,1,0) process.[26] This process is warped by a warping path $\mathbf{P}_j$, which varies across units. The warping path $\mathbf{P}_j$ is a function of variable $\mathbf{s}_j$, which takes on value δ_j (a random variable specific to the unit) if $\mathbf{z}$ is increasing, and δ_j^{-1} otherwise. I.e.:

$$s_{j,t} = \begin{cases} \delta_j & \text{if } \mathbf{z}_t' > 0 \\ \frac{1}{\delta_j} & \text{otherwise,} \end{cases} \qquad \delta_j \sim U(a, b)$$

The idea behind $\mathbf{s}_j$ 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).

The other term in the warping path expression, $\psi \in (0, 1)$, determines the extent to which this occurs. For instance, $\psi = 0$ implies that the warping path will be entirely governed by a random normal variable λ. In this case, the speed at each observation will randomly differ from the next, and $\mathbf{s}_j$ will play no role. Conversely, if $\psi = 1$, the speed profile will completely depend on the direction of the time series (although the speed is still specific to each series, as the random draw δ_j is unit-specific).

26. We also generated data using ARIMA (0,1,1) and ARIMA (1,1,1) models. Our findings indicate that the results across these models are similar.

Finally, τ_t denotes the value of the true treatment effect at time t. For the treated unit,

$$\tau_t = \begin{cases} 0 & \text{if } t \leq T \\ \frac{t-T}{10}\kappa & \text{if } t \in [T, T+10) \\ \kappa & \text{if } t \geq T+10, \end{cases}$$

where κ refers to the magnitude of the treatment effect, and $T = 60$ here. The chosen form of τ_t represents a shock that unfolds over several periods, capturing the gradual impact often observed in real-world scenarios. This modeling choice, not central to our approach, is simply intended to enhance the representation of actual shocks, whose effects may take time to materialize. An illustrative sample dataset is displayed in Figure A2.

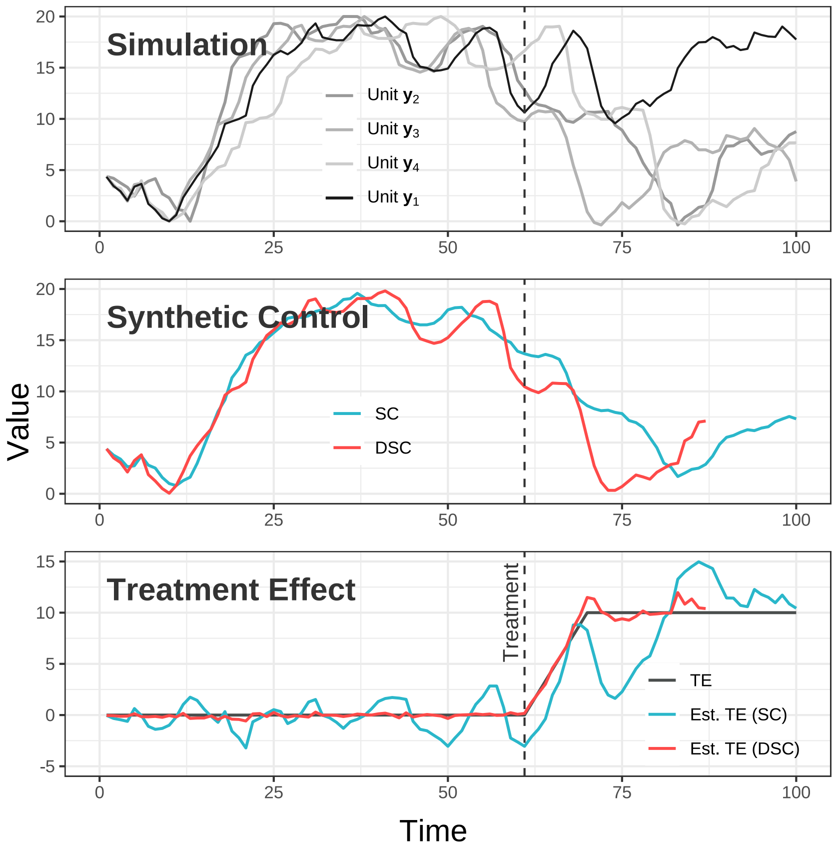

Figure A2. Illustration of an artificial dataset generated using the Model in equation A2. Only a subset of units is displayed for readability. Top: Units $\mathbf{y}_2$–$\mathbf{y}_4$ are used to construct counterfactuals approximating Unit $\mathbf{y}_1$ during the pre-treatment period ($t < 60$). The middle figure depicts synthetic controls: standard SC in blue and DSC in red. The lower figure shows that the estimated treatment effects from DSC are more accurate in capturing the true effect compared to those from the standard SC method.

Appendix 11. Empirical replication with warped covariates

We replicate the placebo tests in Figure 5 with warped covariates. In each analysis, we use the same covariates from the original applications, while these covariates are warped using the same warping path as we used in warping the donor units. This is under the assumption that the covariates have the same speed with the donor unit. The results are shown in Figure A3.

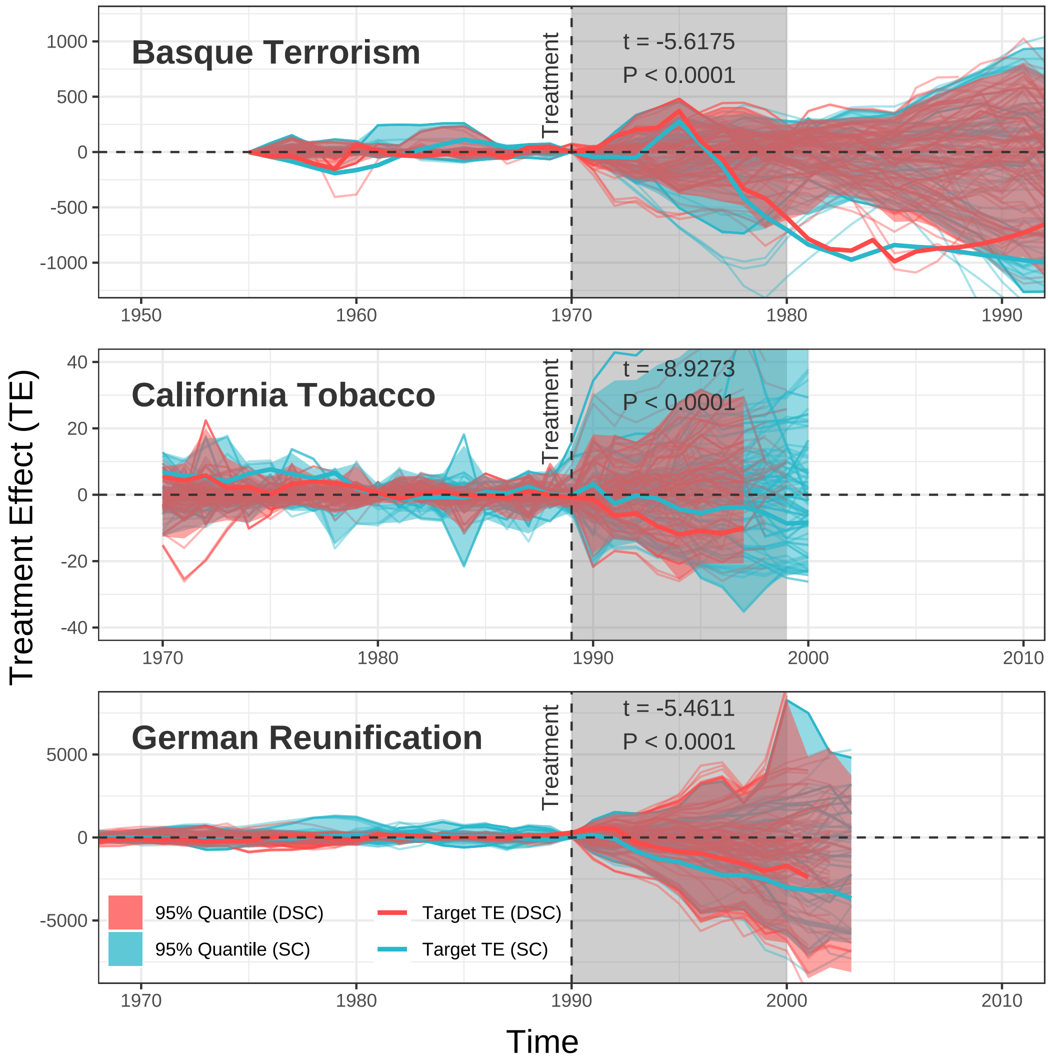

Figure A3. Placebo tests, real data. We revisited the placebo tests reported in Abadie and Gardeazabal (2003), Abadie, Diamond, and Hainmueller (2010) and Abadie, Diamond, and Hainmueller (2015). The covariates are warped. The plots report the placebo tests for each of these studies, using standard synthetic controls (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 synthetic controls, which do not account for variations in speed.

Appendix 12. Length of warped time series

The length of the warped time series $\mathbf{y}_j^w$ often differs from the target time series $\mathbf{y}_1$ for two reasons: (i) the cutoff time C_j and the treatment time T are not necessarily the same; (ii) the warping of the post-treatment time series $\mathbf{y}_{j,post}$ is not based on a direct DTW matching between $\mathbf{y}_{j,post}$ and $\mathbf{y}_{1,post}$. While $\mathbf{y}_{j,pre}^w$ is guaranteed to be of length T, the length of $\mathbf{y}_{j,post}^w$ could be shorter or longer than $\mathbf{y}_{1,post}$ depending on whether C_j is smaller than T or the warping path $\mathbf{P}_{j,post}$ accelerates the average speed of $\mathbf{y}_{j,post}$, thus shortening the warped time series. The difference in length is normal, as the speeds of $\mathbf{y}_1$ and $\mathbf{y}_j$ can vary and we do not impose a strict end-rule that forces $y_{j,N}$ to match $y_{1,N}$. If $\mathbf{y}_{j,post}^w$ is shorter than $\mathbf{y}_{1,post}$, we lose some data points at the end of the time series but gain better speed alignment immediately after the treatment at T. This improved alignment is beneficial for comparative studies.

Changes in the length of the warped donor time series have significant implications for the implementation of the DSC method. In instances where the post-treatment time series are very short, if the DSC method yields a shorter synthetic control, there might be insufficient post-treatment data points to deduce a solid causal inference. Take, for instance, a scenario where the post-treatment segment consists of 12 periods. If, after applying the DSC method, the resulting synthetic control has only 8 periods, then conducting a causal effect study over 10 periods becomes unfeasible.

To avoid this issue, we recommend the adoption of slope-constrained step patterns like *symmetricP2* or *asymmetricP2* (Sakoe and Chiba 1978) in DSC so that the time series are not overly warped. Depending on the chosen step pattern, users can determine the minimum periods of the warped time series. For instance, after warping using the *symmetricP2* or *asymmetricP2* step patterns, the potential length of the warped time series lies within the $[\frac{2}{3}N, \frac{3}{2}N]$ range. This suggests that maintaining a buffer of $\frac{1}{3}N$ can guarantee that the warped time series retains adequate periods for subsequent analysis.

Appendix 13. R package

For the convenience of researchers and practitioners interested in employing the methods discussed in this paper, we have developed an accompanying R package. This package is publicly available at github.com/conflictlab/dsc. The package includes functions to implement all described methods, and it comes with comprehensive documentation to guide users through the installation and application process.

To install the package, one can execute the following R command:

```
devtools::install_github("conflictlab/dsc")
```

Additional documentation and examples for using the package can be found at https://github.com/conflictlab/dsc/blob/main/README.md.

Researchers are encouraged to refer to the package when utilizing the methods described in this paper, and a citation format for acknowledging the package is provided in the package documentation.

References

Abadie, A. 2021. Using synthetic controls: feasibility, data requirements, and methodological aspects. *Journal of Economic Literature* 59 (2): 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 (490): 493–505.

———. 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.

Fleming, M. J., and E. M. Remolona. 1999. Price formation and liquidity in the us treasury market: the response to public information. *The journal of Finance* 54 (5): 1901–1915.

Greene, W. H. 2003. *Econometric analysis.* Pearson Education India.

Reinhart, C. M., and K. S. Rogoff. 2014. Recovery from financial crises: evidence from 100 episodes. *American Economic Review* 104 (5): 50–55.

Sakoe, H., and S. Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing* 26 (1): 43–49.