NBER WORKING PAPER SERIES

COMBINING MATCHING AND SYNTHETIC CONTROLS TO TRADE OFF BIASES FROM EXTRAPOLATION AND INTERPOLATION

Maxwell Kellogg
Magne Mogstad
Guillaume Pouliot
Alexander Torgovitsky

Working Paper 26624
http://www.nber.org/papers/w26624

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2020

Research supported by National Science Foundation grant SES-1846832. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Combining Matching and Synthetic Controls to Trade off Biases from Extrapolation and Interpolation
Maxwell Kellogg, Magne Mogstad, Guillaume Pouliot, and Alexander Torgovitsky
NBER Working Paper No. 26624
January 2020
JEL No. C0,H0,J0

ABSTRACT

The synthetic control method is widely used in comparative case studies to adjust for differences in pre-treatment characteristics. A major attraction of the method is that it limits extrapolation bias that can occur when untreated units with different pre-treatment characteristics are combined using a traditional adjustment, such as a linear regression. Instead, the SC estimator is susceptible to interpolation bias because it uses a convex weighted average of the untreated units to create a synthetic untreated unit with pre-treatment characteristics similar to those of the treated unit. More traditional matching estimators exhibit the opposite behavior: They limit interpolation bias at the potential expense of extrapolation bias. We propose combining the matching and synthetic control estimators through model averaging. We show how to use a rolling-origin cross-validation procedure to train the model averaging estimator to resolve trade-offs between interpolation and extrapolation bias. We evaluate the estimator through Monte Carlo simulations and placebo studies before using it to re-examine the economic costs of conflicts. Not only does the model averaging estimator perform far better than synthetic controls and other alternatives in the simulations and placebo exercises. It also yields treatment effect estimates that are substantially different from the other estimators.

Maxwell Kellogg
mdkellogg@uchicago.edu

Magne Mogstad
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
magne.mogstad@gmail.com

Guillaume Pouliot
1307 E 60th St, Chicago, IL 60637
chicago, il 60637
US
guillaume.pouliot@gmail.com

Alexander Torgovitsky
University of Chicago
1126 E. 59th Street
Chicago, IL, 60637
atorgovitsky@gmail.com

1 Introduction

Estimating the causal effect of an intervention (treatment) is a common task across the social sciences. Longitudinal approaches based on difference-in-differences have long been used for this task. However, the credibility of these methods can be strained when the pre-treatment trends or characteristics of the untreated units differ significantly from those of the treated units. This occurs frequently, especially when the units are large aggregates, such as countries or states. For these types of comparative case studies, the synthetic control (SC) method of Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010, 2015) provides an attractive alternative.

The motivation of the SC method is to limit the extrapolation bias that can occur when units with different pre-treatment characteristics are combined using a traditional adjustment, such as a linear regression. Instead, the SC estimator *interpolates* by using a convex weighted average of the untreated units to create a synthetic untreated unit with pre-treatment characteristics similar to those of the treated unit. As observed by Abadie et al. (2010, pp. 495–496), this makes the SC estimator susceptible to interpolation bias. In Section 2, we formalize this observation by showing that the SC estimator will only avoid such bias if the conditional mean of the outcome is linear in pre-treatment characteristics.

As Abadie and L'Hour (2019) observe, the SC estimator belongs to a large class of estimators constructed around the assumption of selection-on-observables. Within this class, its vulnerability to interpolation bias is unique. Most other commonly used estimators, such as nearest-neighbor matching, suffer from the opposite drawback of potentially extrapolating too much when suitable untreated units are unavailable. That is, the SC estimator controls extrapolation bias while being susceptible to interpolation bias, whereas the matching estimator has the opposite properties. This complementarity suggests that an estimator that adaptively combines the SC and matching estimators may be particularly attractive.

In Section 2, we propose the matching and synthetic control (or MASC) estimator as a model averaging estimator that combines the standard SC and matching estimators. We show how averaging these two purposefully-chosen estimators defends against the weaknesses of both while preserving their strengths. In Section 3, we show how to choose the weight assigned to each estimator in the MASC through cross-validation, as in Wolpert (1992), Breiman (1996) and Hansen and Racine (2012). Our cross-validation criterion uses an evaluation concept referred to as rolling-origin recalibration in the forecasting literature (e.g. Tashman, 2000). One attractive feature of the MASC estimator is that its cross-validated weight can be solved for in closed-form, making it only marginally more difficult to implement than the usual SC estimator.

In Sections 4, 5, and 6, we provide evidence that the MASC estimator performs extremely well in practice. In Sections 4 and 5, we use both illustrative and empirical (data-derived) Monte Carlo simulations to demonstrate the concepts of extrapolation and interpolation bias, and how the MASC estimator defends against both. In Section 6, we use the MASC estimator to revisit the Spanish terrorism application of Abadie and Gardeazabal (2003). We evaluate the performance of the matching, SC, penalized SC (Abadie and L'Hour, 2019), and MASC estimators through placebo exercises using the untreated units. Our findings show that the MASC estimator consistently outperforms the others on these exercises. The MASC estimator also yields treatment effect estimates that are substantially different than those using other methods.

Our paper is related to a growing literature on SC (see Abadie, 2019, for a recent survey). The closest work to ours is the paper by Abadie and L'Hour (2019), who propose the penalized SC estimator. The penalized SC and MASC estimators are different, but related in that both assign weights to untreated units while taking into consideration their distance from the treated unit in terms of pre-treatment characteristics. In Section 2, we show that the penalized SC estimator is the solution to a constrained version of the problem implicitly solved by the MASC. Thus, the MASC represents a richer model than the penalized SC. While this does not necessarily mean it will perform better in practice, our simulations and empirical results in Sections 4–6 suggest that it may have an edge in the types of comparative case studies to which SC estimators are often applied.

Also closely related to our work is the paper by Athey, Bayati, Imbens, and Qu (2019), who also consider the benefits of model averaging in the context of comparative case studies. Those authors combine several of the regularized SC and matrix completion estimators developed in Doudchenko and Imbens (2016) and Athey, Bayati, Doudchenko, Imbens, and Khosravi (2018). Our MASC estimator differs from theirs both in details and intent. The purpose of the MASC estimator is to directly guard against the types of interpolation biases that can occur with the SC estimator, and the extrapolation bias that can occur with matching, by adaptively blending them together. Like Athey et al. (2019), we also find that model averaging tends to work quite well, in concordance with a recurring finding of the economic forecasting literature (see e.g. Stock and Watson, 2004, 2006). A contrast with Athey et al. (2019), and much of the forecasting literature, is that the estimators we average are purposefully chosen to be complementary. This is exactly the case when data-driven model averaging should be especially beneficial, see, for example, Breiman (1996) or Elliot (2011).

2 Methodology

2.1 Setup

Suppose that we observe a scalar outcome, Y_{it}, for cross-sectional units denoted by i at times $t = 1, \ldots, T$, as well as a time-invariant binary treatment group indicator, $D_i \in \{0, 1\}$, and a k–dimensional vector of pre-treatment covariates, X_i. Units in the treated group become treated at an event date, $t^\star$, so that treatment status in time t is given by $D_{it} \equiv D_i \mathbb{1}[t \geq t^\star]$. Associated with the outcome and treatment are potential outcomes $Y_{it}(0)$ and $Y_{it}(1)$, which are related to the observed outcome via $Y_{it} = D_{it}Y_{it}(1) + (1 - D_{it})Y_{it}(0)$. Our goal is to estimate the average treatment on the treated (ATT),

$$\text{ATT}_t \equiv \mathbb{E}[Y_{it}(1) - Y_{it}(0)|D_i = 1] = \mathbb{E}[Y_{it}|D_i = 1] - \mathbb{E}[Y_{it}(0)|D_i = 1] \tag{1}$$

where $t \geq t^\star$ is some period after the event date.

Identifying the ATT in (1) is a matter of identifying the mean untreated outcomes for the treated group in the post-period, i.e. $\beta_t \equiv \mathbb{E}[Y_{it}(0)|D_i = 1]$. This quantity is point identified under the following widely-used pair of assumptions.

Assumption 1. (Selection on observables) If x is in the supports of both $X_i|D_i = 0$ and $X_i|D_i = 1$, then $\mathbb{E}[Y_{it}(0)|D_i = 1, X_i = x] = \mathbb{E}[Y_{it}(0)|D_i = 0, X_i = x]$ for all $t \geq t^\star$.

Assumption 2. (Overlap) The support of $X_i|D_i = 1$ is contained in the support of $X_i|D_i = 0$.

Assumption 1 is variously described in the literature as ignorable treatment assignment (Rosenbaum and Rubin, 1983), unconfoundedness (Imbens and Rubin, 2015), or selection on observables (Barnow, Cain, Goldberger et al., 1980; Heckman and Robb, 1985). Together with Assumption 2, it implies that

$$\begin{aligned} \beta_t = \mathbb{E}\left[\mathbb{E}[Y_{it}|D_i = 0, X_i]\Big|D_i = 1\right] &\equiv \mathbb{E}\left[\mu_t(X_i)|D_i = 1\right] \\ \text{where} \quad \mu_t(x) &\equiv \mathbb{E}[Y_{it}|D_i = 0, X_i = x], \end{aligned} \tag{2}$$

so that β_t is point identified by the outcomes for the untreated group, conditional on covariates, after reweighting by the distribution of these covariates in the treated group. For further discussion, see e.g. Heckman, Ichimura, and Todd (1997, 1998), Imbens (2004, 2015), or Imbens and Rubin (2015).

Suppose now that we observe a sample of $n+1$ realizations $\{(y_{i1}, \ldots, y_{iT}, d_i, x_i)\}_{i=1}^{n+1}$ from the distribution of $(Y_{i1}, \ldots, Y_{iT}, D_i, X_i)$. Our focus in this paper is the comparative case study setting considered by Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015), in which there is only a single treated unit. We label this treated unit as $i = 1$, so that $d_1 = 1$, while $d_i = 0$ for all n remaining units $i \geq 2$.

Since we only have a single treated unit, we estimate $\mathbb{E}[Y_{it}|D_i = 1]$ by the realization of Y_{1t} in the post-period. Similarly, since the empirical distribution of X_i given $D_i = 1$ is simply a point mass at x_1, we estimate β_t with an estimator of $\mu_t(x_1)$. Thus, we focus on a class of estimators for the ATT of the form

$$\widehat{\mathrm{ATT}}_t \equiv y_{1t} - \hat{\mu}_t(x_1), \tag{3}$$

where $\hat{\mu}_t(x_1) \equiv \hat{\mu}_t$ is an estimator of $\mu_t(x_1)$, and we suppress the dependence of $\hat{\mu}_t$ on x_1 for notational efficiency. The problem we focus on is how to construct $\hat{\mu}_t$.

2.2 The Synthetic Control Estimator

The synthetic control (SC) estimator proposed by Abadie and Gardeazabal (2003) and later elaborated by Abadie et al. (2010, 2015) is defined as

$$\hat{\mu}_t^{\mathrm{sc}} \equiv y_{0t}'\omega^{\mathrm{sc}} \quad \text{where} \quad \omega^{\mathrm{sc}} \equiv \underset{\omega \in \mathcal{S}}{\arg\min} \left\| x_1 - x_0'\omega \right\|^2, \tag{4}$$

where $y_{0t} \in \mathbb{R}^n$ are the observed outcomes for the untreated units at time t, $x_0 \in \mathbb{R}^{n \times k}$ is a matrix containing the stacked covariate vectors of the untreated units, $\| \cdot \|$ is the Euclidean norm, and

$$\mathcal{S} \equiv \left\{ \omega \in \mathbb{R}^n : \sum_j \omega_j = 1 \quad \text{and} \quad \omega_j \geq 0 \text{ for all } j \right\}$$

is the $(n-1)$–dimensional simplex.[1] The SC weights, ω^{sc}, are chosen so that the weighted average of covariates among the untreated units comes as close as possible to matching the covariate vector of the treated unit, subject to the convexity constraint that they are non-negative and sum to unity. These weights are then used to construct $\hat{\mu}_t^{\mathrm{sc}}$ by simply weighting the observed outcomes for the untreated units in any given post-period, t.

The SC estimator has a number of attractive properties. By construction, it mini-

[1] The Euclidean norm might be weighted by some symmetric, positive semidefinite matrix, but we omit this from the notation for simplicity.

mizes the quantity

$$\text{Ext}(\omega) \equiv \|x_1 - x_0'\omega\|^2, \tag{5}$$

which can be viewed as a measure of extrapolation. Since the weights are constrained to be convex, this ensures that the SC estimator does not extrapolate, at least as long as it is possible to make (5) small. This stands in contrast to linear regression (Abadie, 2019). For example, Imbens (2004, pg. 13) shows that if the treated and untreated groups have very different pre-treatment characteristics, then linear regression adjustment will be quite sensitive to the way it is specified.

Another benefit of the SC estimator is that the weights ω^{sc} are generally sparse, in the sense that they are only non-zero for a few untreated units (Abadie and L'Hour, 2019). This aids in transparency and provides a way for experts to use contextual knowledge to evaluate the plausibility of the resulting estimates. Also, solving for ω^{sc} only requires solving the quadratic program in (4), which is a straightforward convex problem.

2.3 Interpolation Bias

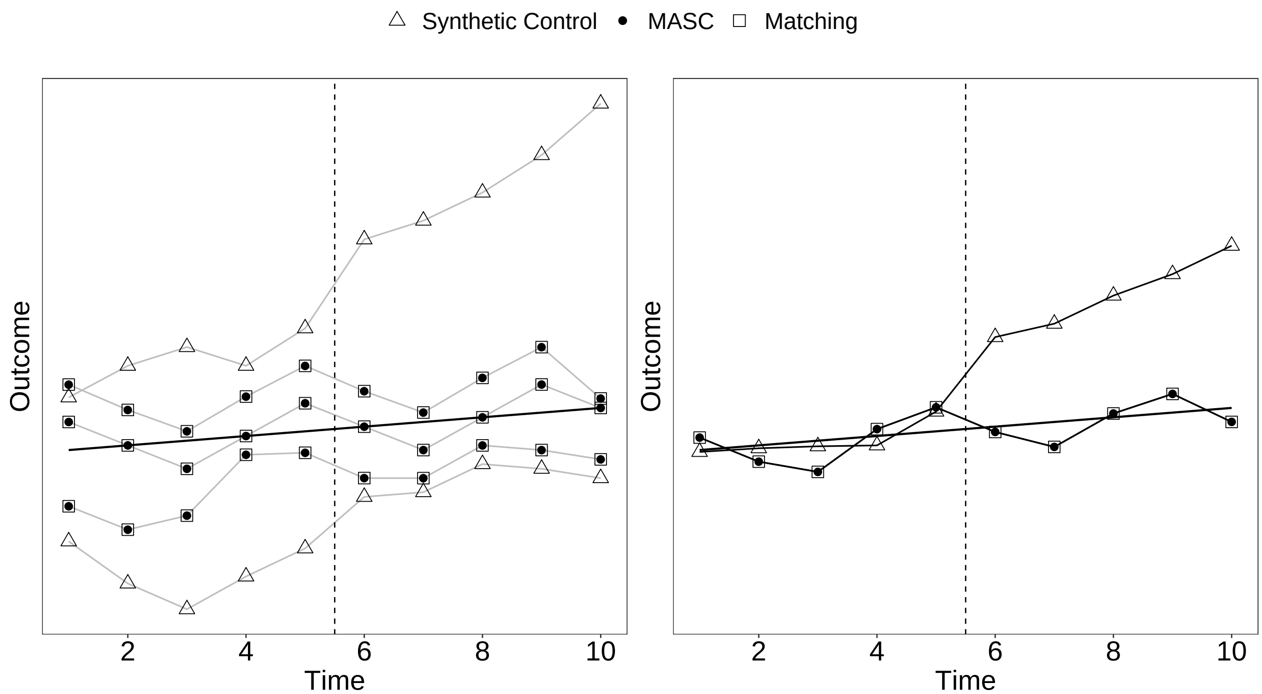
One concern with the SC estimator is that it is susceptible to interpolation bias. This was noted by Abadie et al. (2010, pp. 495–496), and has been discussed more recently by Abadie and L'Hour (2019), although those authors emphasize non-uniqueness issues that occur with many treated or untreated units. Interpolation biases arise when it is possible to reproduce the pre-treatment characteristics of the treated unit by using untreated units with pre-treatment characteristics quite different from the treated unit.

In Figure 1, we illustrate how interpolation biases can arise with the SC estimator. This figure shows a stylized example with untreated outcome paths for a single treated unit and five untreated units. The dashed vertical line indicates the beginning of the treatment period. The outcome paths to the left of the vertical line are observed for both the treated and untreated units. To the right of the vertical line, only the paths for the untreated units are observed, while the path we plot for the treated unit is a potential realization of $Y_{1t}(0)$.

Suppose we follow the recent tradition in the synthetic control literature of taking X_i to include all pre-treatment outcomes and no other covariates (Doudchenko and Imbens, 2016).[2] Then the SC estimator will be comprised solely of the two distant untreated units in Figure 1a. It will put zero weight on the three units whose pre-

[2] We focus on this case throughout the paper both it allows us to examine the methods graphically, and also because it limits specification searching.

Figure 1: *The potential for interpolation bias with the synthetic control estimator*



(a) *Untreated outcome paths for all units. The treated unit is in black.*

(b) *Untreated outcome path for the treated unit against the estimated paths.*

Notes: The vertical dashed line indicates the beginning of the treatment period. Panel (a) depicts all control units in the data as light colored lines. Markers indicate which of these controls are assigned non-zero weight in the different estimators. Panel (a) plots each estimator using the same markers as in panel (b). In both plots, the solid black line indicates the untreated outcome for the treated unit, which is observed as data to the left of the vertical dashed line, and unobserved to the right of it.

period paths oscillate closely around that of the treated unit. This is by design: A convex weighted average of the two distant untreated units gives the best possible fit of the pre-period path of the treated unit. While this choice of weights minimizes extrapolation, it comes at the cost of interpolation.

Whether such interpolation leads $\hat{\mu}_t^{\text{sc}}$ to be biased depends on the structure of the function $\mu_t(x)$. To see this, let $e_{it} \equiv y_{it} - \mu_t(x_i)$ denote the deviation between y_{it} and its conditional mean. Then

$$\hat{\mu}_t^{\text{sc}} \equiv y_{0t}'\omega^{\text{sc}} = \sum_{i\geq 2} \omega_i^{\text{sc}} \left(\mu_t(x_i) + e_{it}\right) = \overbrace{\sum_{i\geq 2} \omega_i^{\text{sc}} \mu_t(x_i)}^{\text{signal}} + \overbrace{\sum_{i\geq 2} \omega_i^{\text{sc}} e_{it}}^{\text{noise}}. \tag{6}$$

In order for the SC estimator to avoid bias due to interpolation, it should be the case that the signal term can approximate $\mu_t(x_1)$. The best possible case is when x_1 lies in the convex hull of x_0, so that $x_1 = x_0'\omega^{\text{sc}} = \sum_{i\geq 2} \omega_i^{\text{sc}} x_i$. However, even when this is so, there is the additional requirement that

$$\sum_{i\geq 2} \omega_i^{\text{sc}} \mu_t(x_i) = \mu_t(x_1) = \mu_t\left(\sum_{i\geq 2} \omega_i^{\text{sc}} x_i\right). \tag{7}$$

We record this necessary condition as the following proposition.

Proposition 1. Suppose that $e_{it} = 0$ for all i, and that $x_0'\omega^{\text{sc}} = x_1$. Then $\hat{\mu}_t^{\text{sc}} = \mu_t(x_1)$ only if (7) holds.

In order for (7) to hold, the function μ_t needs to be effectively linear in x.[3] A sufficient condition for this is that $Y_{it}(0)$ follows a factor structure, as suggested by Abadie et al. (2010, 2015) and further elaborated by Gobillon and Magnac (2016) and Xu (2017). For example, suppose that

$$Y_{it}(0) = \theta_t' X_i + \varphi_t' L_i + U_{it} \quad \text{with} \quad \mathbb{E}[U_{it}|X_i, L_i] = 0, \tag{8}$$

where θ_t is a vector of unknown parameters, φ_t is a vector of unknown time effects, and L_i is a vector of latent factor loadings. Assume further that $\mathbb{E}[L_i|X_i = x] = \Lambda x$ is linear in x. Then

$$\mu_t(x) = \theta_t' x + \varphi_t' \Lambda x,$$

which implies that (7) is satisfied when $x_0'\omega^{\text{sc}} = x_1$, since

$$\sum_{i\geq 2} \omega_i^{\text{sc}} \mu_t(x_i) = \theta_t' \left(\sum_{i\geq 2} \omega_i^{\text{sc}} x_i\right) + \varphi_t' \Lambda \left(\sum_{i\geq 2} \omega_i^{\text{sc}} x_i\right) = \theta_t' x_1 + \varphi_t' \Lambda x_1 = \mu_t(x_1).$$

However, without the linear, additive structure provided by the factor model (8), there is no guarantee that (7) will be satisfied. When it is not, interpolation bias can arise. This is illustrated in Figure 1b, in which the SC estimator interpolates between the two distant untreated units to fit the pre-intervention outcomes of the treated unit. Because μ_t is highly nonlinear, (7) fails, and this interpolation leads $\hat{\mu}_t^{\text{sc}}$ to be a poor estimate of the post-intervention outcomes of the treated unit.

Of course, Figure 1 is a stylized example, which we have constructed to show how interpolation bias can arise. The extent to which it actually arises in applications is an

[3] We say *effectively* linear because technically this condition is only required when considering points in the empirical support of X_i.

empirical question. In Section 5, we characterize the situations in which interpolation bias is a problem by using data-driven Monte Carlo simulations that are chosen to roughly reproduce the data in Abadie and Gardeazabal (2003). In Section 6, we provide evidence of interpolation bias in the original data used by Abadie and Gardeazabal (2003).

2.4 The Matching Estimator

Local nonparametric smoothing estimators are a classical way to estimate $\mu_t(x_1)$. In general, these estimators can be written as

$$\hat{\mu}_t^{\mathrm{lo}} \equiv \sum_{i\geq 2} \kappa\left(\|x_i - x_1\|\right) y_{it} \equiv y_{i0}' \omega^{\mathrm{lo}}, \tag{9}$$

where κ is a kernel function that determines the weight applied to each untreated observation. For such an estimator to be local, the function κ should be decreasing, so that untreated units with predetermined characteristics more distant from the treated unit are given less weight. Local smoothing estimators do not require the linearity condition (7) that was required for the SC estimator. Instead, they rely only on μ_t being sufficiently smooth in its continuous components (e.g. Fan and Gijbels, 1992).

Unlike the SC estimator, local smoothing estimators do not necessarily have sparse, convex weights. However, the specific class of k–nearest neighbors estimators (Cover, 1968) does have weights with these properties. Estimators based on the nearest neighbors idea are widely used for causal inference problems under Assumptions 1 and 2, in which case they are commonly described as matching estimators (e.g. Dehejia and Wahba, 1999; Abadie and Imbens, 2006).

The matching estimator we consider is defined by choosing an integer $m \geq 1$ and setting

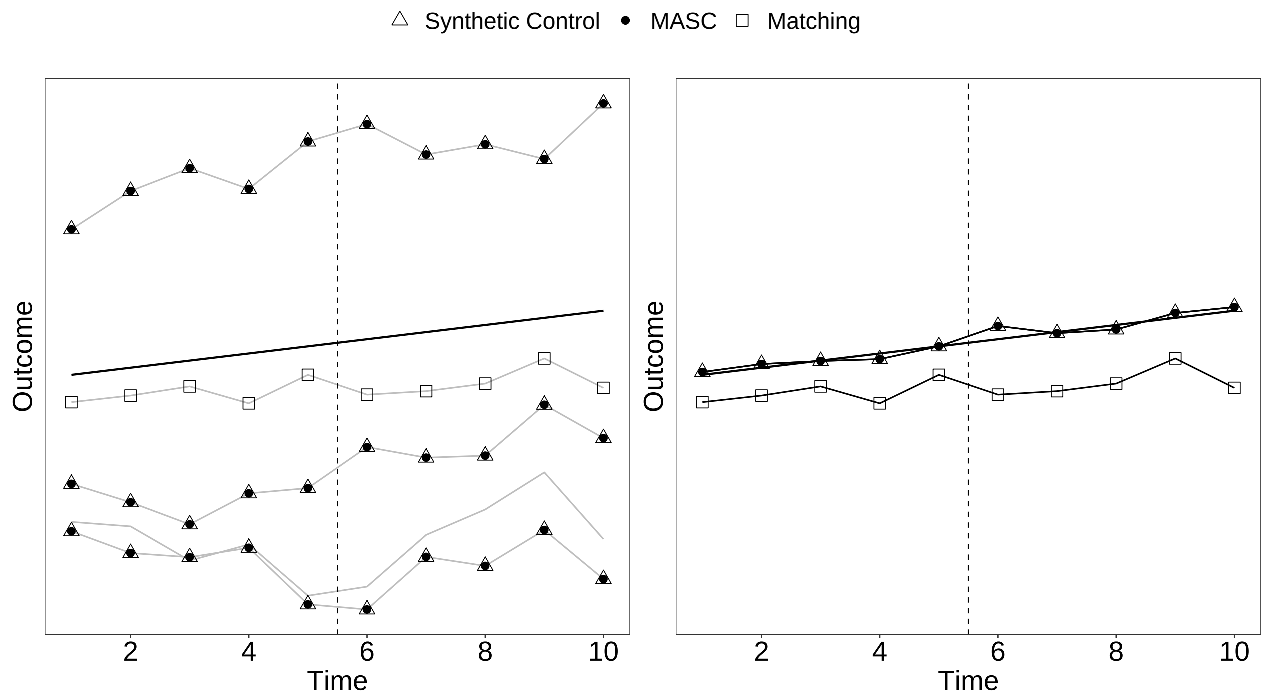
$$\kappa(v) = \begin{cases} 1/m, & \text{if } m \geq \sum_{j\geq 2} \mathbb{1}[v \geq \|x_j - x_1\|] \\ 0, & \text{otherwise,} \end{cases}$$

where, for simplicity, we are assuming there are no ties. Letting $\mathcal{M} \equiv \{i \geq 2 : \kappa(\|x_i - x_1\|) = 1/m\}$ denote the set of untreated units for which this weighting function is non-zero, we can write the matching estimator more concisely as

$$\hat{\mu}_t^{\mathrm{ma}}(m) \equiv \frac{1}{m} \sum_{i\in\mathcal{M}} y_{it} = \sum_{i\geq 2} \frac{1}{m} \mathbb{1}[i \in \mathcal{M}] y_{it} \equiv y_{0t}' \omega^{\mathrm{ma}}(m). \tag{10}$$

Intuitively, the matching estimator just takes the average outcomes across the m un-

Figure 2: *The potential for extrapolation bias with matching*



(a) *Untreated outcome paths for all units. The treated unit is in black.*

(b) *Untreated outcome path for the treated unit against the estimated paths.*

Notes: The vertical dashed line indicates the beginning of the treatment period. Panel (a) depicts all control units in the data as light colored lines. Markers indicate which of these controls are assigned non-zero weight in the different estimators. Panel (a) plots each estimator using the same markers as in panel (b). In both plots, the solid black line indicates the untreated outcome for the treated unit, which is observed as data to the left of the vertical dashed line, and unobserved to the right of it.

treated units that have pre-period characteristics closest to the treated unit. Like the SC estimator, it is a sparse, convex weighted average of the post-period outcomes of the untreated units.

There is an alternative way of expressing the weights for the matching estimator that facilitates comparison with the SC estimator:

$$\omega^{\mathrm{ma}}(m) = \operatorname*{arg\,min}_{\omega \in \mathcal{S}} \sum_{i \geq 2} \omega_i \|x_1 - x_i\|^2 \quad \text{s.t.} \quad \omega_i \leq \frac{1}{m} \quad \text{for all } i \geq 2. \tag{11}$$

This formulation shows that, in contrast to the SC estimator, the matching estimator aims to minimize a measure of *interpolation*:

$$\mathrm{Int}(\omega) \equiv \sum_{i \geq 2} \omega_i \|x_1 - x_i\|^2. \tag{12}$$

For example, with $m = 3$, the matching estimator would equally weight the three units in Figure 1a that oscillate around the treated unit, since their pre-period outcomes are close to those of the treated unit.[4] As a consequence, the matching estimator is less susceptible to interpolation bias than the SC estimator, and in this example provides a better estimate of the post-intervention outcomes for the treated unit.

However, the matching estimator is more vulnerable to extrapolation bias than the SC estimator. To see this, consider Figure 2. In this example, the matching estimator uses the single untreated unit that is closest to the treated unit, even though the two units are not actually that close, resulting in considerable bias. In contrast, the SC estimator weights three of the other untreated units in a way that provides an excellent fit to the untreated outcome path of the treated unit throughout the pre- and post-period. The reason is that μ_t is close to linear in this example, so that (7) is close to satisfied, and the SC estimator has little interpolation bias.

2.5 Model Averaging with the MASC Estimator

Both the SC and matching estimators share a number of appealing properties in common. As illustrated in Figures 1 and 2, however, their drawbacks are different and diametrically opposed: The SC estimator controls extrapolation bias but not interpolation bias, while the matching estimator does the opposite. This complementarity suggests that a model averaging estimator will be able to harness the best properties of both the matching and SC estimators.[5]

With this motivation, we define the matching and synthetic control (MASC) estimator as

$$\hat{\mu}_t^{\text{masc}} \equiv \phi \hat{\mu}_t^{\text{ma}}(m) + (1 - \phi)\hat{\mu}_t^{\text{sc}} \equiv y_{0t}' \omega^{\text{masc}}$$

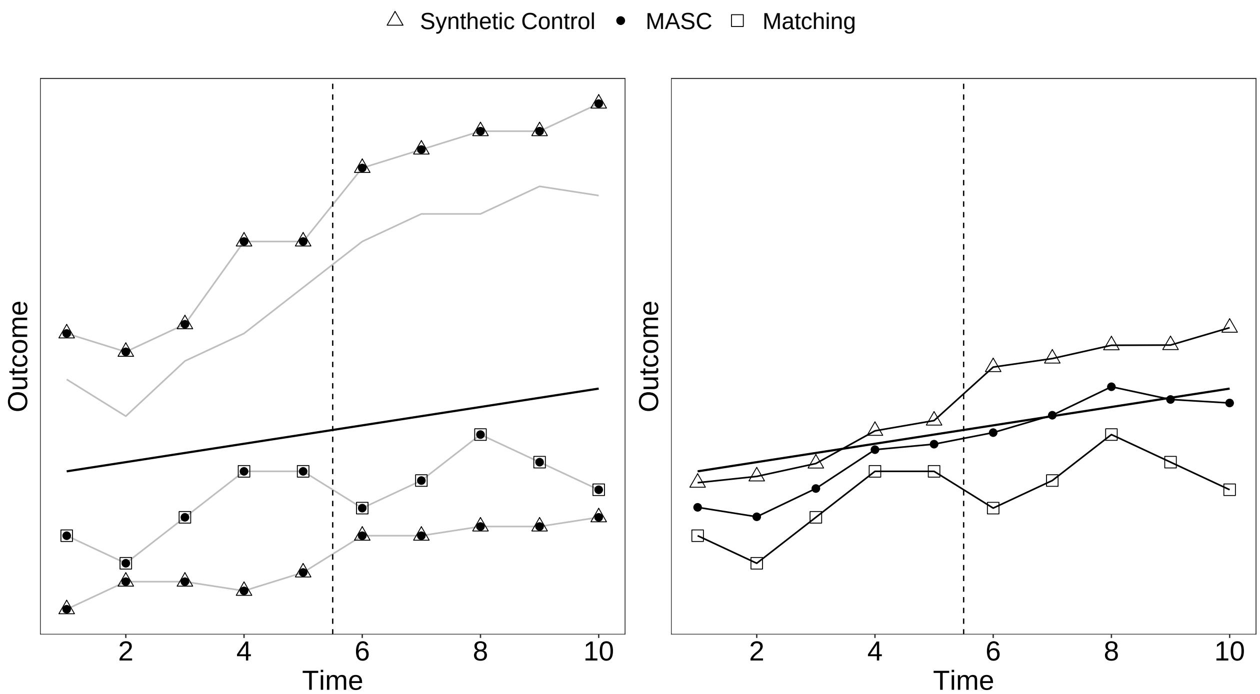
where $\phi \in [0, 1]$ is a tuning parameter, and $\omega^{\text{masc}} \equiv \phi \omega^{\text{ma}}(m) + (1-\phi)\omega^{\text{sc}}$. In Section 3, we provide a cross-validation procedure for choosing ϕ and m. This allows the MASC to control both interpolation and extrapolation biases in a data-driven way. When interpolation is the chief concern, the procedure makes the MASC estimator assign more weight to the matching estimator. In Figure 1, it sets $\phi = 1$, so that the MASC exactly coincides with the matching estimator. On the other hand, when extrapolation is the concern, the procedure assigns more weight to the SC estimator. For example, in Figure 2, it sets $\phi = 0$, so that the MASC exactly coincides with the SC estimator.

Intermediate cases can also arise, as in Figure 3. In this case, the outcome paths

[4] Our cross-validation procedure, which we discuss ahead in Section 3, selects $m = 3$ in this example.

[5] For example, see the discussion surrounding Theorem 1 of Breiman (1996).

Figure 3: MASC adapts to control both extrapolation and interpolation bias



(a) Untreated outcome paths for all units. The treated unit is in black.

(b) Untreated outcome path for the treated unit against the estimated paths.

Notes: The vertical dashed line indicates the beginning of the treatment period. Panel (a) depicts all control units in the data as light colored lines. Markers indicate which of these controls are assigned non-zero weight in the different estimators. Panel (a) plots each estimator using the same markers as in panel (b). In both plots, the solid black line indicates the untreated outcome for the treated unit, which is observed as data to the left of the vertical dashed line, and unobserved to the right of it.

are moderately non-linear, so the SC estimator suffers from interpolation bias. At the same time, there are no untreated units that closely match the pre-period path of the treated unit, so the matching estimator suffers from extrapolation bias.[6] In contrast, the cross-validation procedure chooses $\phi \approx .5$, which allows the MASC estimator to mix the SC estimator with the matching estimator, mitigating both sources of bias.

2.6 The Penalized Synthetic Control Estimator

A related, but much different estimator has recently been proposed by Abadie and L'Hour (2019). Those authors start with the SC estimator and add a penalty that discourages choosing units far from the treated unit. Their penalized SC estimator is

[6] In this case, the cross-validation procedure selects $m = 1$ for the matching estimator.

defined as

$$\hat{\mu}_t^{\text{pen}} \equiv y_{0t}'\omega^{\text{pen}}$$

$$\text{where} \quad \omega^{\text{pen}} \equiv \underset{\omega\in\mathcal{S}}{\arg\min}\ (1-\pi)\|x_1 - x_0'\omega\|^2 + \pi\left(\sum_{i\geq 2}\omega_i\|x_i - x_1\|^2\right), \tag{13}$$

where $\pi \in [0,1]$ is a tuning parameter that controls the penalty incurred by weighting untreated units with pre-treatment characteristics different from the treated unit. When $\pi = 0$, the penalized SC estimator reduces to the usual SC estimator, $\hat{\mu}_t^{\text{sc}}$, while for $\pi = 1$, it is equal to $\hat{\mu}_t^{\text{ma}}(m)$ with $m = 1$.[7]

The optimization problem solved by the penalized SC estimator is a constrained version of the one implicitly solved by the MASC estimator. This is because (13) can also be written as

$$\omega^{\text{pen}} = \underset{\omega^a,\omega^b\in\mathcal{S}}{\arg\min}\ (1-\pi)\|x_1 - x_0'\omega^a\|^2 + \pi\left(\sum_{i\geq 2}\omega_i^b\|x_i - x_1\|^2\right) \quad \text{s.t.} \quad \omega^a = \omega^b,$$

whereas ω^{masc} is the solution to this program (with π replaced by ϕ) when $m = 1$ and the constraint $\omega^a = \omega^b$ is dropped. While the MASC estimator takes a convex combination of the SC and matching estimators—which respectively minimize extrapolation and interpolation bias—the penalized SC estimator solves a constrained problem which can lead it to choose an entirely different set of weights.

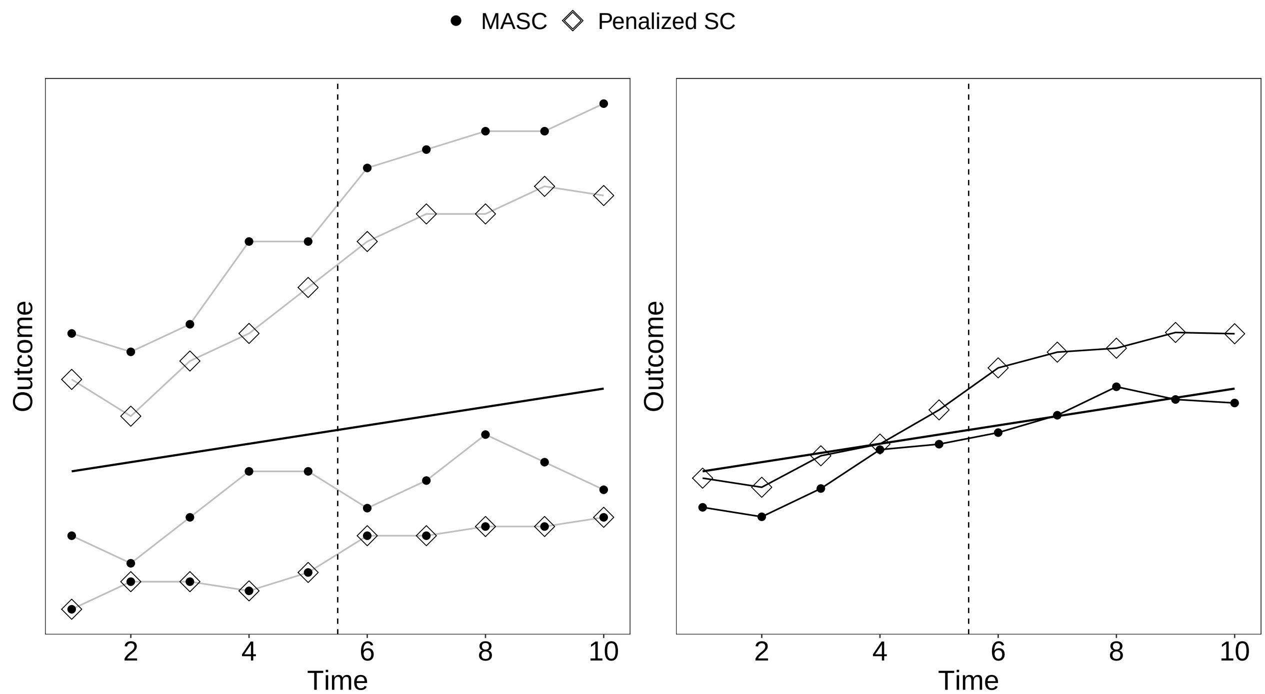
Figure 4 demonstrates the implications this can have in practice, using the same scenario as in Figure 3. Recall that in this example, the MASC estimator is a roughly equal combination of the SC and matching estimators. In contrast, the penalized SC estimator represents neither the SC estimator nor the matching estimator. It puts roughly half of its weight on the untreated unit whose path is most below the treated unit, which is a unit that gets weight in the SC estimator. The other half of the penalized SC weight is placed on a unit that is given zero weight in both the SC and matching estimators. Intuitively, the penalized SC estimator ends up being the average

[7] Note that Abadie and L'Hour (2019) parameterize their criterion function slightly differently as

$$\underset{\omega\in\mathcal{S}}{\arg\min}\ \|x_1 - x_0'\omega\|^2 + \tilde{\pi}\left(\sum_{i\geq 2}\omega_i\|x_i - x_1\|^2\right),$$

for $\tilde{\pi} \geq 0$. This can be made comparable to (13) by dividing by $(1-\pi)$, so that $\tilde{\pi} \equiv \frac{\pi}{(1-\pi)}$. Then, $\tilde{\pi} = 0$ corresponds to the SC estimator, while the matching estimator with $m = 1$ is recovered as $\pi \to 1$ so that $\tilde{\pi} \to \infty$, just as in Abadie and L'Hour (2019).

Figure 4: *Contrasting the behavior of penalized SC with MASC*



(a) *Untreated outcome paths for all units. The treated unit is in black.*

(b) *Untreated outcome path for the treated unit against the estimated paths.*

Notes: The vertical dashed line indicates the beginning of the treatment period. Panel (a) depicts all control units in the data as light colored lines. Markers indicate which of these controls are assigned non-zero weight in the different estimators. Panel (a) plots each estimator using the same markers as in panel (b). In both plots, the solid black line indicates the untreated outcome for the treated unit, which is observed as data to the left of the vertical dashed line, and unobserved to the right of it.

of a suboptimal SC estimator and a suboptimal matching estimator. In this example, its bias is comparable to that of the SC estimator, and significantly greater than that of the MASC estimator.

It is important to observe that we are ignoring a primary motivation provided by Abadie and L'Hour (2019) for the penalized SC estimator, which is its ability to solve the non-uniqueness problem that can arise when solving the SC problem (4). As Abadie and L'Hour (2019) discuss, this problem is usually not an issue when there is a single treated unit, which is the case we consider here. It becomes much more likely to be problematic with multiple treated units. In such settings, one could modify the MASC so that it averages between the matching and *penalized* SC estimators. We expect that the resulting estimator would behave similar to the way the MASC behaves when there is a single treated unit.

3 Cross-Validation

In this section, we propose a cross-validation procedure for choosing the tuning parameters for the estimators discussed in the previous section. As in Abadie et al. (2015), our procedure is based on optimizing the fit of the treated unit's outcome series in the pre-treatment period. Whereas those authors used a single training-validation split, our procedure uses a series of one-step ahead forecasts, each of which is estimated using data only from periods prior to the forecast date. This is called rolling-origin recalibration in the forecasting literature (e.g. Tashman, 2000; Bergmeir and Benítez, 2012).[8]

We define our folds, f, as consisting of all pre-period data running from period $t = 1$ up to $t = f$. Let $\hat{\mu}_{f+1}(\tau)$ denote a generic estimator of the outcome in period $f + 1$ based on data in fold f, where τ is a generic vector of tuning parameters. Our cross-validation procedure chooses τ to minimize

$$Q(\tau) \equiv \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} \left(y_{1,f+1} - \hat{\mu}_{f+1}(\tau)\right)^2, \tag{14}$$

where $\mathcal{F}$ is a subset of time periods taken from the pre-period, $\{1, \ldots, t^\star - 1\}$.

Figure 5 illustrates the structure of the rolling-origin cross-validation procedure. In this example, the treatment date is $t^\star = 21$, so that there are 20 pre-treatment periods. Fold $f = 19$ uses all data from $t = 1, \ldots, 19$ to construct a forecast of the treated unit's outcome in period $f + 1 = 20$. Fold $f = 18$ uses data from $t = 1, \ldots, 18$ to forecast at $t = 19$, and so on. The criterion is constructed by averaging together the squared prediction errors from a choice of folds, $\mathcal{F}$.

The largest that $\mathcal{F}$ can be is of course $\{1, \ldots, t^\star - 1\}$. In practice, we use fewer folds than this, and prefer folds that are longer. The bias-variance trade-offs that drive this choice are natural. Folds closer to the treatment date are likely to be more relevant to the post-treatment period. They are also larger, so that the estimators use more data. On the other hand, we expect that having more folds will decrease the variance of $Q(\tau)$. These trade-offs are also present in more common applications of cross-validation with independent and identically distributed data (e.g. Hastie, Tibshirani, and Friedman, 2009, pg. 242–243). The added complication here is that not all folds are equally valuable, so we prefer ones that are closer to the actual treatment date.

The parameters τ differ by estimator. The synthetic control estimator has no tuning parameters.[9] The matching parameter has the number of matches, m. The

[8] A similar evaluation concept is the rolling-window considered by Swanson and White (1997).

[9] As we mentioned in footnote 1, the Euclidean norm defining the synthetic control or matching estimators

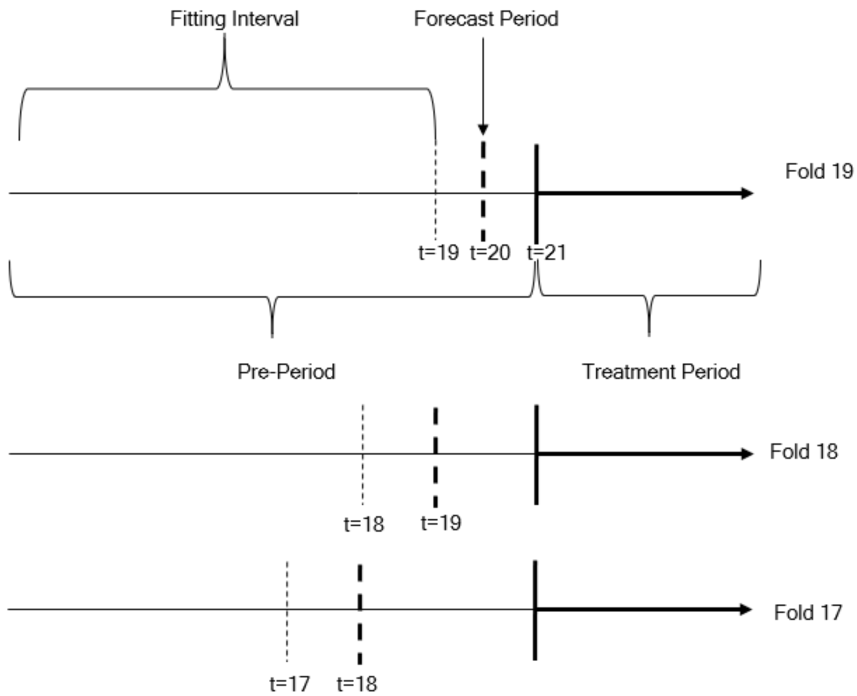


Figure 5: *Cross-validation based on rolling-origin recalibration.*

MASC estimator has both m and the model average parameter, ϕ. The penalized synthetic control estimator has the penalty parameter, π.

For the MASC estimator, it is straightforward to find the unconstrained minimum of $Q(\phi, m)$ in ϕ for any fixed m. Using least squares algebra, the solution is

$$\phi^\star(m) \equiv \frac{\sum_{f\in\mathcal{F}}(\hat{\mu}^{\text{ma}}_{f+1}(m) - \hat{\mu}^{\text{sc}}_{f+1})(y_{1,f+1} - \hat{\mu}^{\text{sc}}_{f+1})}{\sum_{f\in\mathcal{F}}(\hat{\mu}^{\text{ma}}_{f+1}(m) - \hat{\mu}^{\text{sc}}_{f+1})^2}. \tag{15}$$

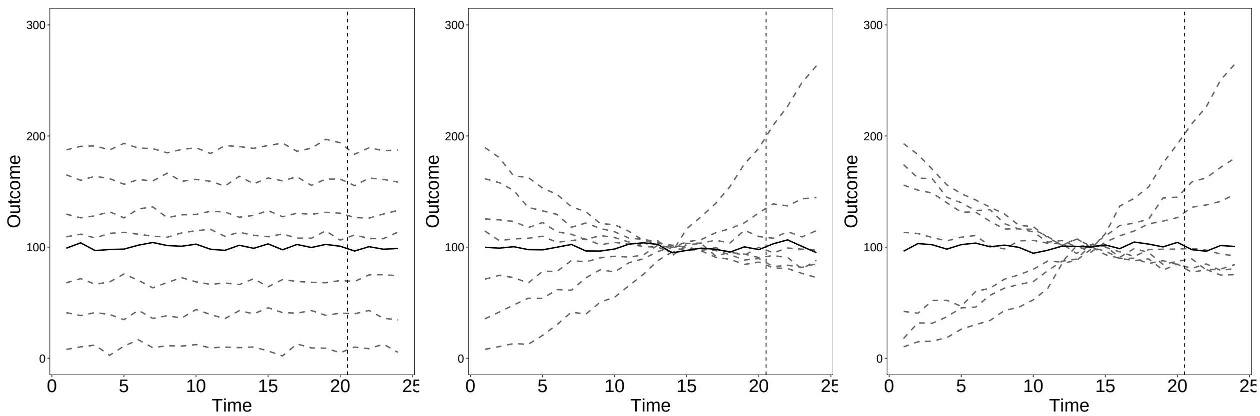
This means that cross-validating the MASC is extremely easy computationally. First, compute $\phi^\star(m)$ for a set of potential matches, m. Then for each m, set

$$\hat{\phi}(m) \equiv \begin{cases} 0, & \text{if } \phi^\star(m) \leq 0 \\ 1, & \text{if } \phi^\star(m) \geq 1 \\ \phi^\star(m) & \text{otherwise} \end{cases}$$

Finally set $\hat{m} \equiv \arg\min_m Q(\hat{\phi}(m), m)$, and set $\hat{\phi} \equiv \hat{\phi}(\hat{m})$. The cross-validated MASC

could be weighted. Abadie et al. (2010, 2015) view the weights as tuning parameters and choose them using cross-validation. We could do this as well with our criterion (14), but we have elected not to in the current paper because optimizing over the weights introduces computational issues that, while solvable, are not the main focus of our paper (Becker and Klößner, 2017, 2018).

Figure 6: *Typical draws in the illustrative Monte Carlo*



(a) *No time trend* ($\alpha = 1$) ***(b)*** *Nonlinear* ($\alpha = .95$) ***(c)*** *Increased dispersion*

Notes: The vertical dashed line indicates the beginning of the treatment period. The treated unit is shown in black.

estimator is a weighted average of $\hat{\mu}^{\text{sc}}$ and $\hat{\mu}^{\text{ma}}(\hat{m})$ with weights $(1-\hat{\phi})$ and $\hat{\phi}$, respectively.

For the penalized SC estimator, $Q(\pi)$ is not necessarily convex in π, which makes it harder to find the global minimum. In the results ahead, we use a grid search to cross-validate both the MASC and penalized SC estimators, so that we can focus on the statistical differences instead of computational artifacts. In practice, one should cross-validate the MASC estimator analytically, as described in the previous paragraph.

4 An Illustrative Monte Carlo

We first illustrate the behavior of different estimators in a fabricated simulation. This simulation is meant to be expository, not necessarily realistic; we consider a Monte Carlo based on real data in the next section.

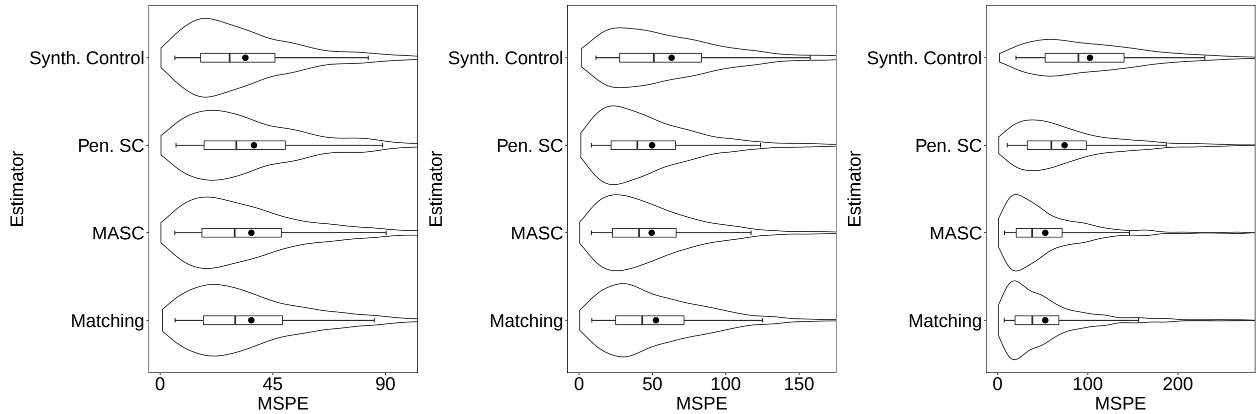
The data is generated by taking draws over $T = 24$ periods from

$$Y_{it} = \text{trend}_{it}(\alpha) + V_{it},$$

where $\text{trend}_{it}(\alpha)$ is a unit-specific time trend that depends on a scalar parameter, α, and V_{it} is normally distributed for all units with mean 0 and standard deviation 5. The full expression for $\text{trend}_{it}(\alpha)$ is given in Appendix A; the important part is that its nonlinearity over time decreases with α for each of the seven untreated units. As shown in Figures 6a and 6b, for $\alpha = 1$ there is no trend, and for $\alpha = .95$ the untreated units range from also have no trend to having a severely nonlinear trend.

We measure performance in terms of the mean squared prediction error averaged

Figure 7: *Performance in the illustrative Monte Carlo*



(a) *No time trend* ($\alpha = 1$) ***(b)*** *Nonlinear* ($\alpha = .95$) ***(c)*** *Increased dispersion*

Notes: These violin plots show the density of MSPE over 1,000 simulation draws. The internal boxplots indicate the 5th and 95th percentiles at the end of the whiskers, the 25th and 75th percentiles at the hinges, and the median at the line splitting the box. The black dot is the mean.

over the post-period, which starts in $t^\star = 21$. For a generic estimator $\hat{\mu}$, this is defined as

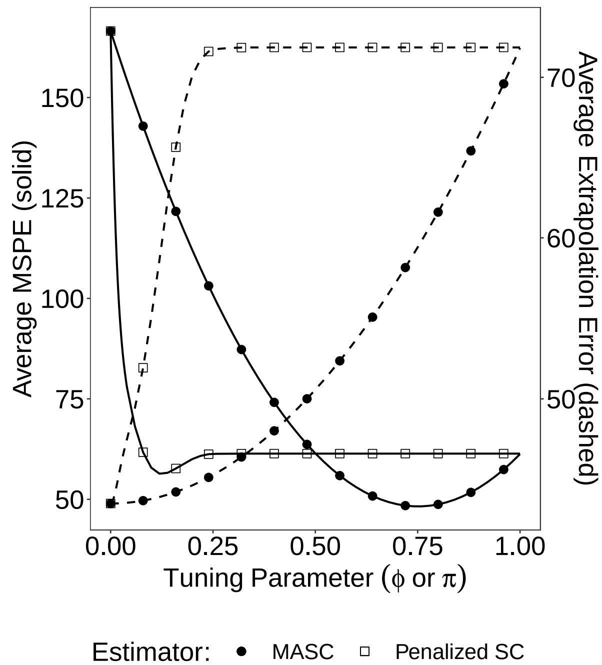
$$\text{MSPE} \equiv \frac{1}{4} \sum_{t=21}^{24} \left(y_{1t} - \hat{\mu}_t\right)^2 .$$

Figures 7a and 7b report violin plots of MSPE across simulations for $\alpha = 1$ and $\alpha = .95$.[10] For the linear case with $\alpha = 1$, all estimators have similar MSPE distributions, although the SC estimator is a bit more concentrated towards 0. For the nonlinear case with $\alpha = .95$, the SC estimator starts to suffer from interpolation bias with both larger average MSPE and larger, more frequent poor-performing outlying draws, while the other estimators—including both the MASC and penalized SC estimators—continue to perform about equally well.

In Figure 6c, we make the interpolation bias even more severe by maintaining $\alpha = .95$, but increasing the dispersion in the initial conditions of the untreated units. One untreated unit remains a close neighbor, while the others have moved further away in the early parts of the sample. The violin plot in Figure 7c shows that both the matching and MASC estimators continue to perform well in this case, because they pick up this close neighbor. In contrast, the SC and penalized SC estimators perform much worse.

[10] Our Monte Carlo simulations use 1,000 replications. The tuning parameters in the matching, penalized SC, and MASC estimators are chosen through cross-validation with $\mathcal{F} = \{12, \ldots, 19\}$ consisting of 8 folds. For the matching and MASC estimators, we choose the number of matches from $m \in \{1, 2, 3, 4, 5\}$.

Figure 8: *Decreasing MSPE by allowing for extrapolation error*



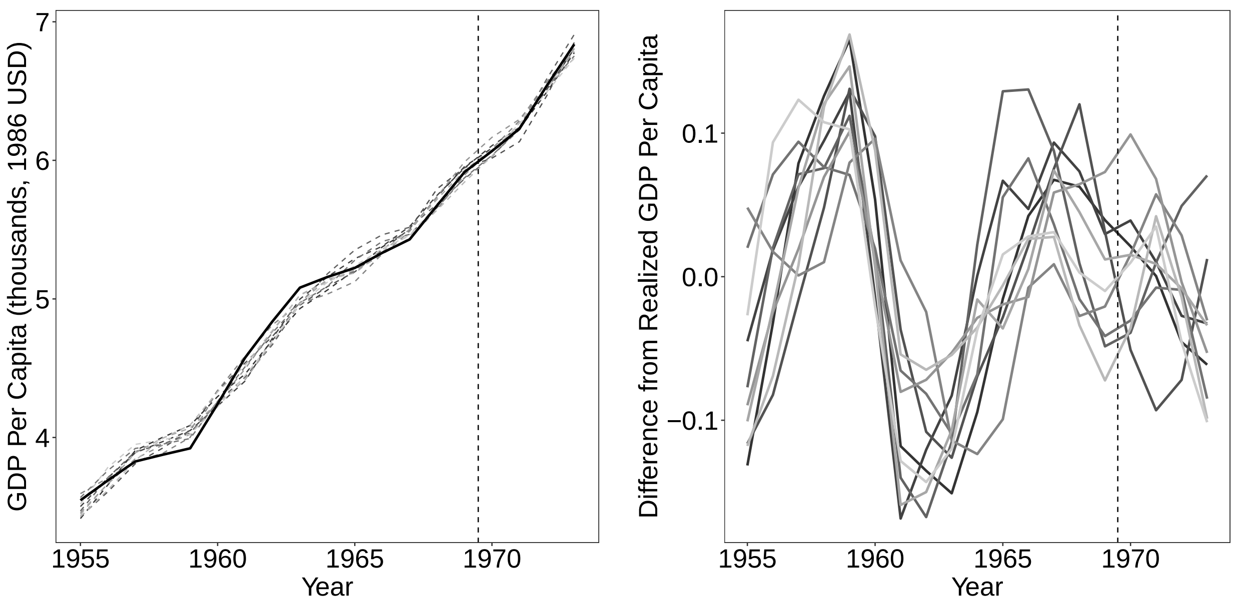
Note: The results are for the third design, which is depicted in Figure 6c. The MASC estimator uses $m = 1$.

Figure 8 shows why this occurs. The figure plots both the average MSPE and the average extrapolation error, as measured by $\text{Ext}(\omega)$, as functions of ϕ and π for both the MASC and penalized SC estimators. When $\phi = \pi = 0$, both estimators correspond to the standard SC estimator which, while it minimizes extrapolation error, has a very large average MSPE in this case. Increasing ϕ or π increases extrapolation error, but reduces average MSPE due to reduced interpolation error. After a certain point, the trade-off reverses, and average MSPE starts increasing again until reaching the matching estimator at $\phi = \pi = 1$. Both the MASC and penalized SC estimators capture this trade-off, although in this case the MASC does so more efficiently due the wider range over which it outperforms both the SC and matching estimators.

5 An Empirical Monte Carlo

In the next section, we reexamine Abadie and Gardeazabal's (2003) seminal application of the synthetic control method to estimating the effect of terrorism on per capita GDP in Spain. In this section, we first implement a Monte Carlo simulation designed to mimic their data by simulating from a model that fits their data. The purpose of this simulation is to study the different estimators in a setting that is constructed

Figure 9: Ten draws of Catalonia in the empirical Monte Carlo



(a) Outcome paths *(b) Difference between draw and Catalonia*

Notes: The vertical dashed line is drawn halfway between 1969 and 1970 to indicate the beginning of the placebo treatment period in 1970. The solid black line in panel (a) indicates the original times series for Catalonia.

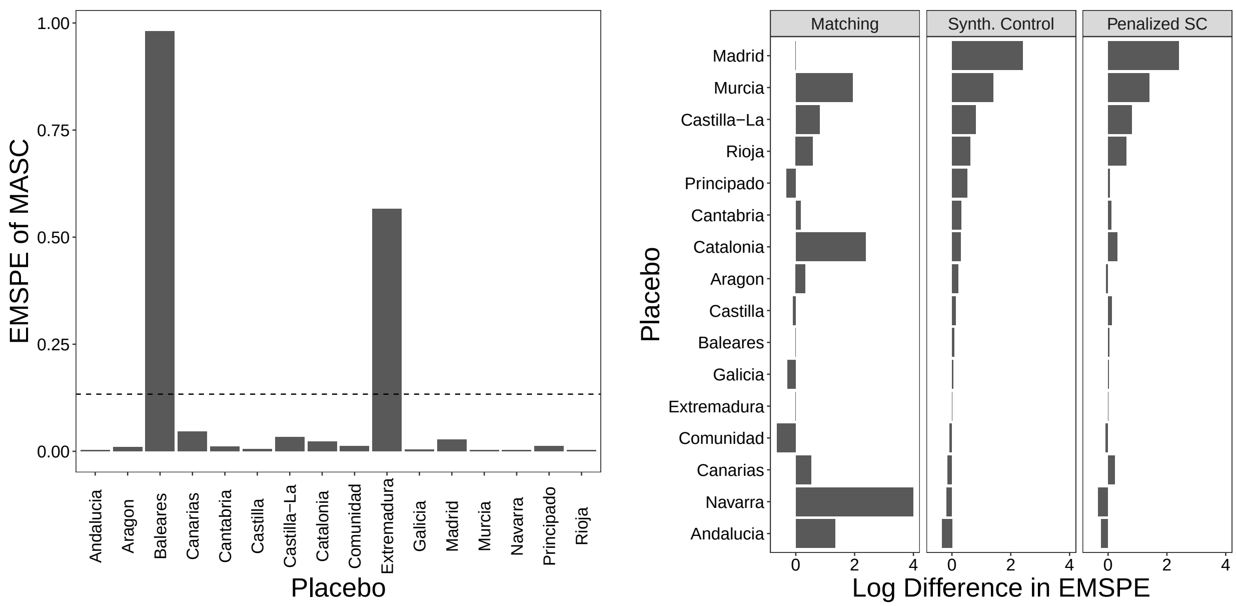
to closely resemble real data. The original data consists of time series on per-capita GDP running from 1955–1997 for 17 regions in Spain. The treated unit is the Basque Country, and the treatment (the onset of separatist terrorism) begins in 1970. Here, we use only the period from 1955–1973, and we conduct a placebo study that uses only the 16 untreated units.

We construct a data generating process by fitting a spatial autoregressive model with normally distributed errors.[11] Then, we generate simulated data by taking draws from this model. Figure 9 shows the original data for one untreated unit (Catalonia), as well as ten sample draws from the estimated model. The perturbations from the data are relatively minor, which is our intent, since we want to preserve the qualitative features of the data. The model does good job at faithfully reproducing the patterns in the Spanish data with a bit of sampling error added. It does not mechanically favor either the SC or matching estimator.

We use this data generating process to compare estimators through a placebo exer-

[11] The model is similar to that in Blanchard and Katz (1992) or Acemoglu, Naidu, Restrepo, and Robinson (2019). First, we regress the raw outcome paths against a set of time dummies as well as a unit-specific cubic trend. Then, we fit the detrended paths with an AR(2) process with normally distributed innovations that are allowed to be arbitrarily correlated across regions.

Figure 10: *Results of the Spanish placebo Monte Carlo simulation*



(a) *Average MSPE for MASC* ***(b)*** *Average MSPE relative to MASC*

Note: The dashed line in panel (a) indicates the mean square treatment effect estimated for the Basque Country for scale. Panel (b) reports the proportional (log) difference in average MSPE between each estimator and MASC.

cise. For each untreated unit, we fit an estimator using all other untreated units, and dropping the Basque Country. Then we look at the distribution of its four-year MSPE over 1970–1973 across simulation draws. We expect this distribution to be clustered close to zero if the estimator is working well.

Figure 10a shows the average MSPE for the MASC estimator in each region's placebo study.[12] Figure 10b shows the proportional (log) difference in average MSPE between the MASC estimator and the other four estimators. The MASC estimator tends to outperform the other estimators, including the penalized SC estimator, which in this case looks close to the standard SC estimator. The MASC adapts to regions such as Madrid, where matching performs well but SC does poorly. It also adapts to regions such as Catalonia, where matching does poorly, but SC does well. In regions like Murcia, where neither matching nor SC perform well, the MASC outperforms both. For the two regions where the MASC does poorly, Balearas and Extremadura, the other estimators perform equally poorly.[13]

[12] In this simulation we cross-validate using $\mathcal{F} = \{1962, \ldots, 1968\}$ and choose the number of matches from all integers between 1 and 10.

[13] For Extremadura, all estimators turn out to be identical. This is because both its outcome path lies substantially and uniformly below that of Castilla-La Mancha, which itself lies substantially and uniformly below the outcome paths of all other regions. All estimators thus place all of their weight on Castilla-La

Figure 11: *Cross-validation performance in the Spanish placebo Monte Carlo Simulation*

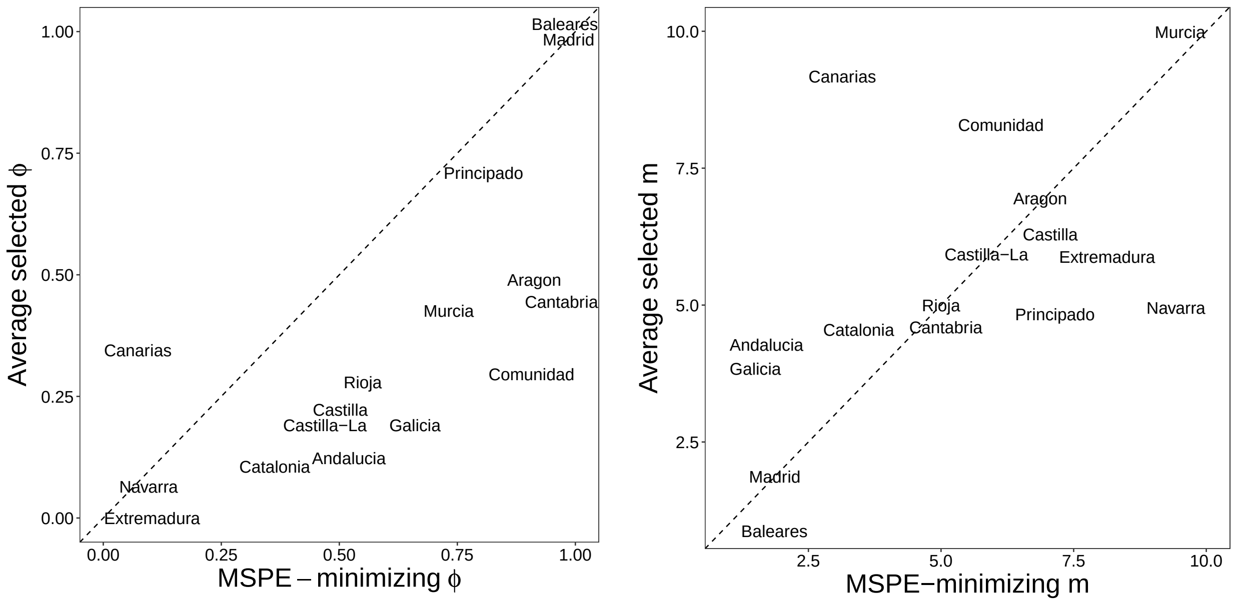


Figure 11 plots the average values of ϕ and m chosen for the MASC by the rolling-origin cross-validation procedure against the infeasible values that would minimize the unknown expected MSPE. For most regions, the average selected values of ϕ and m are reasonably close to the 45 degree line. To the extent that the average values deviate, the deviations tend to be towards smaller ϕ's and larger m's. This suggests that, if anything, the cross-validated MASC is underfitting, since the cross-validation procedure appears biased towards weighting the SC estimator and using a larger number of matches.

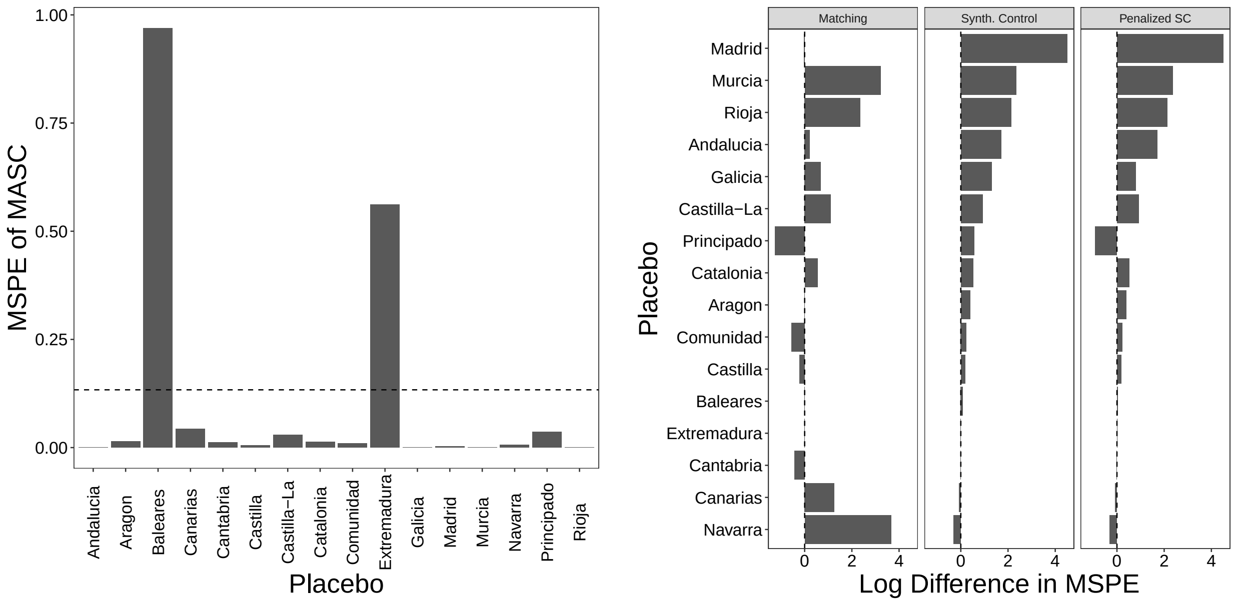
6 Re-Examining the Economic Costs of Conflict

In this section, we use the entire Spanish data set to re-examine the estimates from Abadie and Gardeazabal (2003) using the MASC estimator.

We begin by conducting the same type of placebo exercise as in the previous section, but now using the real data. Abadie and Gardeazabal (2003) performed this analysis using Catalonia as the placebo region, since Catalonia is a similar region with lower exposure to terrorism, and the one that received the most weight in their original application of the SC estimator. They found that the SC estimator reproduced the actual per capita GDP for Catalonia quite well, at least up to the late 1980s. They interpreted this as evidence in support of their estimates for the Basque Country.

Mancha in the Extremadura placebo.

Figure 12: *Results of the placebo estimates*



(a) *MSPE for MASC* ***(b)*** *MSPE relative to MASC*

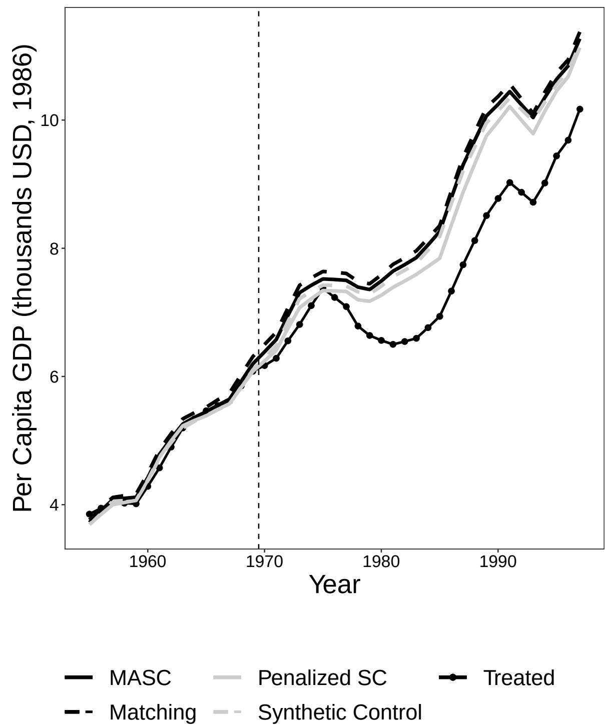
Note: The dashed line in panel (a) indicates the mean square treatment effect estimated for the Basque Country for scale. Panel (b) reports the proportional (log) difference in MSPE between each estimator and MASC.

We apply a similar placebo exercise to each of the 16 untreated region, including Catalonia. For each one, we exclude both the region in question and the Basque Country when constructing the various estimators. Then we use each estimator to compute treatment effect estimates over the pre-terrorism years 1970–1973.[14] This iterative procedure yields a distribution of MSPE across regions where no intervention took place. As in the previous section, we expect the distribution to cluster near 0 if the estimator is working well.

Figure 12 displays the results. Consistent with the simulation results, the MASC estimator performs much better than either matching, SC, or penalized SC estimators. On average, these estimators have MSPEs that are, respectively, 24, 27, and 21 percent higher than the MASC. Figure 10b suggests that the reason is the same as we found in the placebo simulation exercise in the previous section. The MASC estimator is able to adapt to regions where matching performs well, and to regions where the SC estimator performs well, while blending the two successfully in regions where both do poorly. In contrast, the penalized SC estimator tends to behave quite similarly to the standard SC estimator. This echoes its interpretation as a restricted version of the

[14] As in Section 5, we continue to cross-validate all estimators using $\mathcal{F} = \{1962, \ldots, 1968\}$, and we choose the number of matches for the matching and MASC estimators from all integers between 1 and 10.

Figure 13: Counterfactuals for the Basque Country by estimator



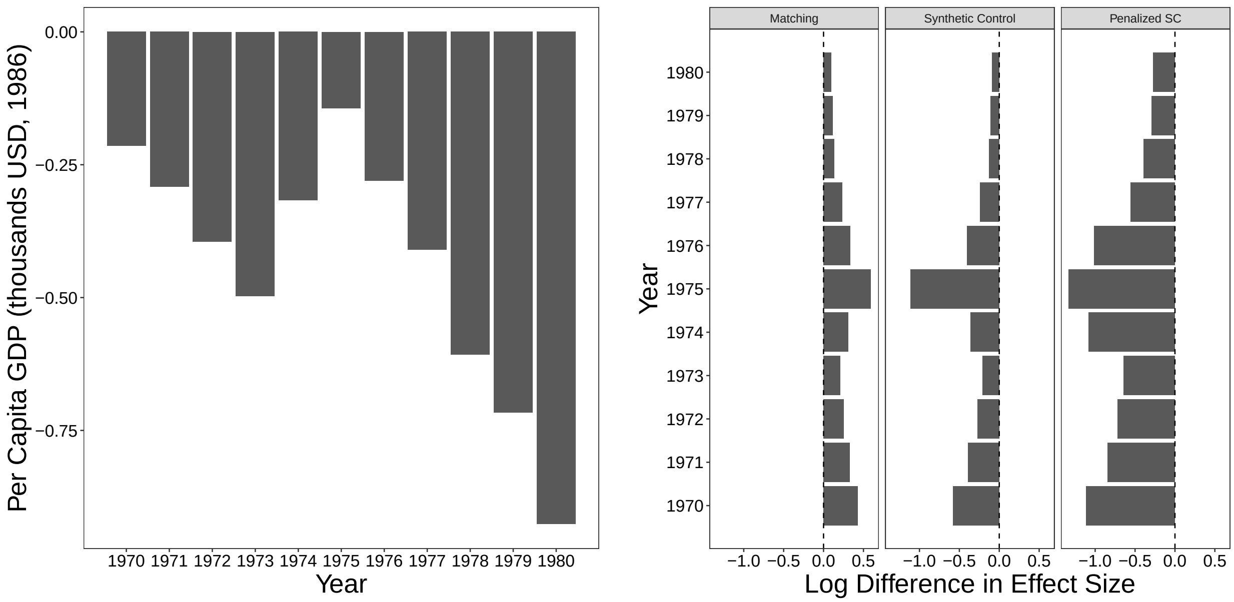
Note: The dashed line is drawn halfway between 1969 and 1970 to indicate the beginning of the treatment period in 1970.

MASC estimator. At least in the current setting, it appears that the extra flexibility of the MASC estimator provides substantial benefits over the penalized SC estimator.

Having established that the MASC performs better on the placebo exercise, we now turn to treatment effect estimates for the Basque Country. Figure 13 shows the counterfactual estimates for the Basque Country in the absence of terrorism. The SC and penalized SC estimators track the actual Basque Country series well up until 1974, which is the year that separatist terrorism started to really ramp up (Abadie and Gardeazabal, 2003, Table 1). The matching and MASC estimators track the series less well, despite the fact that the MASC estimator performed much better than the rest on the placebo exercises. This demonstrates the point (already well-known in the literature on SC) that pre-period fit alone should not be used to evaluate the credibility of an estimator.

Figure 14a shows the yearly treatment effect estimates for the MASC estimator in the period 1970–1980. The MASC estimates suggest that separatist terrorism caused an economically significant decrease in per capita GDP in the Basque Country. By

Figure 14: *Treatment effects for the Basque Country by estimator*



(a) *Treatment Effects for MASC* ***(b)*** *Effect size relative to MASC*

Note: Panel (b) reports the log difference in treatment effect size between each estimator and MASC.

1975, for example, we estimate that terrorism had reduced GDP per capita by 144 US dollars (a 1.9 percent reduction). By way of comparison, the SC estimate is -47 US dollars (a 0.6 percent reduction), and the penalized SC estimate is 37.9 US dollars (a 0.5 percent increase).[15] Figure 14b shows the proportional (log) difference in the estimated treatment effects between the MASC estimator and the other estimators. Here we see that across the entire post-period, the matching estimator suggests considerably larger effect estimates than MASC, while the penalized and standard SC estimators suggest much smaller effect sizes.

7 Conclusion

One of the major impacts of the synthetic control method has been to recast longitudinal comparative case studies as prediction problems. In this paper, we made use of two tools from the machine learning and economic forecasting literature: Model averaging and rolling-origin forecast evaluation. By examining the weakness of the synthetic control (SC) to interpolation bias, and the weakness of the matching estimator to extrapolation bias, we showed how to use these tools to build a third estimator, the matching and synthetic control (MASC) estimator, which is able to effectively avoid

[15] These are 1986 dollars, as in Abadie and Gardeazabal (2003).

both sources of bias. Using both simulated and empirical placebo studies, we showed that the MASC performs much better than either the matching, SC, or penalized SC estimators. We used the MASC estimator to re-examine Abadie and Gardeazabal's (2003) application to the economic costs of conflict in the Basque Country and found significantly larger effects than with SC.

Appendix

A The Data Generating Process for the Illustrative Monte Carlo

The trend terms are unit-specific autoregressive (AR) sequences defined recursively as

$$\text{trend}_{it} = \rho_i \text{trend}_{i(t-1)}^{\alpha}.$$

For $\alpha = 1$, this is an AR(1), while for $\alpha = .95$ it becomes nonlinear, as in Figures 6b and 6c. The autoregressive coefficient for the treated unit, ρ_1, is taken to be $\rho_1 = 100^{1-\alpha}$. For the seven untreated units with $i \geq 2$, it is set by

$$\log(\rho_i) = \left(\frac{1-\alpha}{1-0.95}\right)\left(\frac{13}{\sum_{s=0}^{12} 0.95^s}\right)\log\left(\rho_1 \frac{\text{trend}_{11}}{\text{trend}_{i1}}\right).$$

We specify the initial conditions as either

$$\begin{pmatrix} \text{trend}_{1t} \\ \text{trend}_{2t} \\ \text{trend}_{3t} \\ \text{trend}_{4t} \\ \text{trend}_{5t} \\ \text{trend}_{6t} \\ \text{trend}_{7t} \\ \text{trend}_{8t} \end{pmatrix} = \begin{pmatrix} 100 \\ 10 \\ 40 \\ 70 \\ 110 \\ 130 \\ 160 \\ 190 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \text{trend}_{1t} \\ \text{trend}_{2t} \\ \text{trend}_{3t} \\ \text{trend}_{4t} \\ \text{trend}_{5t} \\ \text{trend}_{6t} \\ \text{trend}_{7t} \\ \text{trend}_{8t} \end{pmatrix} = \begin{pmatrix} 100 \\ 10 \\ 25 \\ 40 \\ 110 \\ 160 \\ 175 \\ 190 \end{pmatrix},$$

where the first set of values is used for Figures 6a and 6b, and the second set is used when we fan out the initial conditions in Figure 6c.

References

Abadie, A. (2019): "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*, forthcoming. 3, 6

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic Control Methods for

Comparative Case Studies: Estimating the Effect of Californias Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505. 2, 5, 6, 8, 16

——— (2015): "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59, 495–510. 2, 5, 8, 15, 16

Abadie, A. and J. Gardeazabal (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country," 93, 113–132. 2, 3, 5, 9, 19, 22, 24, 25, 26

Abadie, A. and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235–267. 9

Abadie, A. and J. L'Hour (2019): "A Penalized Synthetic Control Estimator for Disaggregated Data," . 2, 3, 6, 12, 13, 14

Acemoglu, D., S. Naidu, P. Restrepo, and J. A. Robinson (2019): "Democracy Does Cause Growth," *Journal of Political Economy*, 127, 47–100. 20

Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2018): "Matrix Completion Methods for Causal Panel Data Models," Working Paper 25132, National Bureau of Economic Research. 3

Athey, S., M. Bayati, G. Imbens, and Z. Qu (2019): "Ensemble Methods for Causal Effects in Panel Data Settings," *AEA Papers and Proceedings*, 109, 65–70. 3

Barnow, B. S., G. G. Cain, A. S. Goldberger, et al. (1980): *Issues in the Analysis of Selectivity Bias*, University of Wisconsin, Inst. for Research on Poverty. 4

Becker, M. and S. Klössner (2017): "Estimating the Economic Costs of Organized Crime by Synthetic Control Methods," *Journal of Applied Econometrics*, 32, 1367–1369. 16

——— (2018): "Fast and Reliable Computation of Generalized Synthetic Controls," *Econometrics and Statistics*, 5, 1–19. 16

Bergmeir, C. and J. M. Benítez (2012): "On the Use of Cross-Validation for Time Series Predictor Evaluation," *Information Sciences*, 191, 192–213. 15

Blanchard, O. J. and L. F. Katz (1992): "Regional Evolutions," *Brookings Papers on Economic Activity*, 1992, 1–75. 20

Breiman, L. (1996): "Stacked Regressions," *Machine Learning*, 24, 49–64. 2, 3, 11

Cover, T. (1968): "Estimation by the Nearest Neighbor Rule," *IEEE Transactions on Information Theory*, 14, 50–55. 9

Dehejia, R. H. and S. Wahba (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062. 9

Doudchenko, N. and G. W. Imbens (2016): "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis," Working Paper 22791, National Bureau of Economic Research. 3, 6

Elliot, G. (2011): "Averaging and the Optimal Combination of Forecasts," *Working Paper*. 3

FAN, J. AND I. GIJBELS (1992): "Variable Bandwidth and Local Linear Regression Smoothers," *The Annals of Statistics*, 20, 2008–2036. 9

GOBILLON, L. AND T. MAGNAC (2016): "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 98, 535–551. 8

HANSEN, B. E. AND J. S. RACINE (2012): "Jackknife Model Averaging," *Journal of Econometrics*, 167, 38–46. 2

HASTIE, T., R. TIBSHIRANI, AND J. H. FRIEDMAN (2009): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, New York, NY: Springer, 2nd ed ed. 15

HECKMAN, J. J., H. ICHIMURA, AND P. TODD (1998): "Matching as an Econometric Evaluation Estimator," *The Review of Economic Studies*, 65, 261–294. 4

HECKMAN, J. J., H. ICHIMURA, AND P. E. TODD (1997): "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies*, 64, 605–654. 4

HECKMAN, J. J. AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and B. Singer, Cambridge University Press. 4

IMBENS, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *The Review of Economics and Statistics*, 86, 4–29. 4, 6

——— (2015): "Matching Methods in Practice: Three Examples," *Journal of Human Resources*, 50, 373–419. 4

IMBENS, G. W. AND D. B. RUBIN (2015): *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press. 4

ROSENBAUM, P. R. AND D. B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55. 4

STOCK, J. H. AND M. W. WATSON (2004): "Combination Forecasts of Output Growth in a Seven-Country Data Set," *Journal of Forecasting*, 23, 405–430. 3

——— (2006): "Chapter 10 Forecasting with Many Predictors," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann, Elsevier, vol. 1, 515–554. 3

SWANSON, N. R. AND H. WHITE (1997): "Forecasting Economic Time Series Using Flexible versus Fixed Specification and Linear versus Nonlinear Econometric Models," *International Journal of Forecasting*, 13, 439–461. 15

TASHMAN, L. J. (2000): "Out-of-Sample Tests of Forecasting Accuracy: An Analysis and Review," *International Journal of Forecasting*, 16, 437–450. 2, 15

WOLPERT, D. H. (1992): "Stacked Generalization," *Neural Networks*, 5, 241–259. 2

XU, Y. (2017): "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models," *Political Analysis*, 25, 57–76. 8